Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

Christian Moe Sletten

Automatic Detection of Mental Stress Responses from Electroencephalogram Signals

Master's thesis in Cybernetics and Robotics
Supervisor: Marta Molinas
Co-supervisor: Andres Soler
June 2023

NTNU
Norwegian University of
Science and Technology

Christian Moe Sletten

Automatic Detection of Mental Stress Responses from Electroencephalogram Signals

Master's thesis in Cybernetics and Robotics
Supervisor: Marta Molinas
Co-supervisor: Andres Soler
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Norwegian University of
Science and Technology

Preface

This Master's Thesis was completed during my final semester in the Cybernetics and Robotics Master program at the Norwegian University of Science and Technology (NTNU). The project was submitted to fulfil the requirements for the degree at the Department of Engineering Cybernetics at NTNU.

I extend my gratitude to my supervisor, Prof. Marta Molinas, for her continuous guidance, invaluable feedback and the significant amount of time she devoted to ensure my success in this project. Her generosity in lending the necessary equipment and providing office space was instrumental in the completion of this work.

Likewise, I express my sincere appreciation for the guidance and feedback provided by Postdoc Andres Soler throughout this project. His expertise and patience has been invaluable in navigating the complexities of this project.

Furthermore, I extend my gratefulness to Anne Joo Yun Marthinsen, Ida Marie Andreassen, Ivar Tesdal Galtung and Øystein Stavnes Sletta for their cooperation during the data collection period. Their effort and work was essential to successfully perform the data collection.

Christian Sletten
Trondheim, June 2023

Abstract

This thesis presents a detailed study of the effectiveness and limitations of Electroencephalography (EEG)-based mental stress detection in humans using an interpersonal model. Three classification algorithms, AdaBoost, Random Forest and EEGNet, were explored and evaluated. The EEGNet classifier, when used with an 80:20 train-test split, provided the highest accuracy of 71% but achieved only 50% sensitivity. Conversely, the AdaBoost classifier yielded the best sensitivity of 67%, with an accuracy of 63%, under the same split conditions. However, evaluation using Leave-One-Subject-Out (LOSO) cross-validation suggests that the EEGNet classifier is struggling to generalise to new data. Furthermore, the effectiveness of the various EEG frequency bands was evaluated. The delta, theta, alpha, and beta frequency bands have been identified as the most effective for mental stress detection, while the gamma frequency band exhibited less effectiveness. The study also explored the significance of electrode placement and the selection and rejection of channels. The findings suggest that the right-side frontal, central, and temporal regions of the brain have the greatest change in activity during mental stress. Channel selection, performed in a simple way by removing less significant channels, was found to increase sensitivity for the Random Forest classifier while reducing sensitivity for the AdaBoost classifier. These findings suggest that future research should employ more channels and more advanced channel selection techniques. Despite the achievements in the study, the results were limited by the size, balance and quality of the collected dataset. Future studies should focus on collecting high-quality, balanced and comprehensive datasets

for further advancements in the field of EEG-based mental stress detection.

Sammendrag

Denne avhandlingen presenterer en detaljert studie av effektiviteten og begrensningene av Elektroencefalografi (EEG)-basert psykisk stressdeteksjon hos mennesker ved bruk av en interpersonlig modell. Tre klassifiseringsalgoritmer, AdaBoost, Random Forest og EEGNet, ble utforsket og evaluert. EEGNet-modellen, kombinert med 80:20 trening/test-splitt, ga den høyeste nøyaktigheten på 71%, men oppnådde en sensitivitet på kun 50%. Til forskjell oppnådde AdaBoost-modellen den beste sensitiviteten på 67%, med en nøyaktighet på 63%, ved de samme splitt-betingelsene. Derimot viste en evaluering ved bruk av Leave-One-Subject-Out (LOSO) kryssvalidering at EEGNet-modellen sliter med å generalisere til ny data. Videre ble effektiviteten til de ulike EEG-frekvensbåndene evaluert. Delta-, theta-, alfa- og beta-frekvensbåndene ble identifisert som de mest effektive for psykisk stressdeteksjon, mens gamma-frekvensbåndet viste seg å være mindre effektivt. Studien undersøkte også betydningen av elektrodeplassering og effekten av å fjerne kanaler. Resultatene viser at den frontale, sentrale og temporale regionen på høyre side av hjernen har den største endringen i aktivitet under psykisk stress. Valg av kanaler, som ble utført på en enkel måte ved å fjerne de minst betydningsfulle kanelene, viste at sensitiviteten økte for Random Forest men gikk ned for AdaBoost. Disse resultatene tyder på at fremtidig forskning bør benytte flere kanaler og utforske mer avanserte teknikker for valg av kanaler. Til tross for prestasjonene i denne studien, har resultatene vært begrenset av størrelsen, balansen og kvaliteten til det innsamlede datasettet. Fremtidige studier bør fokusere på innsamling av større, mer balanserte datasett av høyere kvalitet for å

oppnå fremskritt innenfor EEG-basert psykisk stressdeteksjon.

Acronyms

CNN convolutional neural network. 3, 16, 44

DFT discrete Fourier transform. 11

EEG Electroencephalography. ii, iii, iv, v, x, xi, xii, 1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 43, 48, 51, 52, 53, 54, 55, 56, 57, 58, 59, 63, 64, 65, 69, 70, 71, 72, 73, 74, 75, 76, 79, 80

LOSO Leave-One-Subject-Out. ii, iv, x, 45, 53, 60, 61, 63, 64, 65, 73, 74, 80

MEG magnetoencephalogram. 35

NTNU Norwegian University of Science and Technology. i, 1, 24

PCG Phonocardiography. xi, 1, 23, 24, 25, 26, 27, 29, 30, 31, 32, 75

PSD Power Spectral Density. xi, xii, 1, 8, 11, 12, 13, 35, 37, 38, 39, 40, 51, 52, 71

SSA Subjective Stress Assessment. 13, 24, 30

STAI-Y State-Trait Anxiety Inventory Y. 13, 24, 25, 29, 30, 31, 44

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Problem Description

The aim of this thesis is twofold. The first aim is to collect a multimodal dataset consisting of Electroencephalography (EEG) and Phonocardiography (PCG) data from participants in a non-stressed and stressed state. Collecting a balanced dataset consisting of recordings from a substantial number of participants while minimising environmental noise are key targets for this stage of the project. Data collection in this project has been carried out in cooperation with Anne Joo Yun Marthinsen, Ida Marie Andreassen, Ivar Tesdal Galtung and Øystein Stavnes Sletta, fellow students at the Norwegian University of Science and Technology (NTNU).

The second aim is to create an automated stress detection system using EEG signals. Even though the PCG data is collected, it will serve the purpose of future research where multimodal stress detection can be explored. This single-modal system aims to set a groundwork for comparison with such future multimodal systems. The proposed stress detection system will consist of several stages, including preprocessing of the raw EEG data, feature extraction using specific frequency bands from the Power Spectral Density (PSD) of EEG data and training and tuning of multiple classification models that can accurately distinguish between non-stressed and stressed

states in participants.

1.2 Motivation

In the modern age, technology, while offering numerous benefits, has become a double-edged sword. Technology has given us the ability to effortlessly connect with people around the world from the comfort of our smartphones. On the other hand, the smartphone has tethered us to an ever-growing heap of work emails and a constant stream of social media interactions. This constant connection to the outside world means that many do not get the chance to disconnect and unwind, which is vital to maintain a healthy mental well-being. Technology is one of the contributing factors to the fast-paced and stressful life many now take for granted.

Although stress in the short term can be beneficial in enhancing focus and productivity, chronic stress can produce severe health consequences. The study by Cohen et al. (2007) outlines the correlations between chronic mental stress and diseases and disorders such as depression, cardiovascular disease, HIV/AIDS and cancer. This is only the tip of the iceberg, as other studies have found chronic stress to trigger anxiety disorders, contribute to obesity, and lead to sleep disturbances (Selye (1965); Dohrenwend and Dohrenwend (1974); Brown and Harris (1978)).

The impact of chronic stress on mental and physical health highlights the importance of early detection and monitoring of stress. Although, methods for measuring stress exist, such as self-reported measures or questionnaires, they are vulnerable to personal biases and inaccuracies. Creating a system that can objectively measure stress would therefore have a significant impact on preventing and managing chronic stress.

To further this goal, this study explores the utilisation of EEG signals, a non-invasive method for measuring electrical potentials on the scalp, as a tool for stress detection. EEG-based stress detection remains an open research problem, despite significant advances in machine learning and neural network models. With the development of more accurate algorithms and a deeper understanding of the correlation between EEG signals and stress, we hope to unlock the potential of EEG-based stress

detection, paving the way towards more effective stress management strategies, and ultimately, better health outcomes.

1.3 Literature Review

Recent research has demonstrated the capacity of EEG to detect and quantify stress levels. Hosseini et al. (2010) illustrated that stress-induced changes in the brain can be reflected in the EEG signals. Particularly, changes in EEG frequency bands, especially the theta, alpha, and beta bands, are often indicative of stress response. For example, an increase in beta band activity is often associated with a heightened stress level (Seo et al. (2010)). A comprehensive literature review by Katmah et al. (2021) corroborated this finding, identifying these bands as the most effective across various studies. Hence, these frequency bands are valuable features that can be utilised by machine learning algorithms for stress detection models.

The role of machine learning in EEG data analysis has led to remarkable progress in the research field, as evidenced by the successful implementation of algorithms such as AdaBoost, Random Forest, and EEGNet.

AdaBoost, for instance, has been applied successfully in a variety of EEG-based applications, including detecting driver fatigue, diagnosing epilepsy and stress detection (Hu (2017); Prabhakar and Rajaguru (2018); Liu et al. (2022)).

Likewise, the Random Forest classifier has proven effective in EEG data analysis, helping in the classification of sleep stages, early detection of seizures and identification of driving-induced stress (Fraiwan et al. (2012); Donos et al. (2015); Halim and Rehan (2020)).

EEGNet, a deep learning convolutional neural network (CNN) architecture specifically designed for EEG data, has shown promising potential. According to Lawhern et al. (2018), EEGNet outperformed traditional machine learning algorithms in several EEG classification tasks, including epilepsy diagnosis and event-related potential identification. Recently, Bhatnagar et al. (2023) successfully employed EEGNet to detect stress in patients exposed to music.

While promising results have been reported, several challenges remain. The in-

dividual variability in stress response makes it challenging to create an interpersonal stress detection model. Additionally, the EEG frequencies that have the largest change in response during mental stress is still an open research problem. Lastly, the lack of interpretability of deep learning models such as EEGNet complicates the physical understanding of the stress response in the brain.

However, several challenges persist in this area of research. One notable challenge is the individual variability in stress response, which complicates the development of universal stress detection models. This variability could stem from personal factors such as the perception of stressors. Additionally, determining the EEG frequencies that have the largest change in response during mental stress remains an open research problem. This challenge highlights the need for more research to discover the precise neurophysiological markers of stress. Lastly, the lack of interpretability of deep learning models like EEGNet inhibits the physical understanding of the stress response in the brain, making the explanation and validation of these models challenging.

In conclusion, the use of machine learning algorithms, particularly Random Forest, AdaBoost, and EEGNet, for EEG-based stress detection represents a promising yet challenging research area. Addressing these challenges may potentially lead to more accurate and reliable stress detection models in the future.

1.4 Outline

The thesis is structured as follows:

- Chapter 2: Presents the theoretical background of the study, including an explanation and overview of EEG, the association between EEG and mental stress, preprocessing methods, feature extraction techniques and classification algorithms.
- Chapter 3: Details the materials and methods used in the study, including data collection, preprocessing methods, feature extraction and classification approaches.
- Chapter 4: Presents the findings of the study.

- Chapter 5: Provides an analysis of the results and a discussion of their implications.
- Chapter 6: Provides a summary of the main findings of the study and suggestions for future work in the field.

Chapter 2

Theory

2.1 Electroencephalography

Electroencephalography (EEG) is a widely used technique for non-invasive measurement of the electrical activity of the brain. EEG is commonly used in the field of neuroscience to study brain function and diagnose a range of neurological diseases (Niedermeyer and da Silva (2005)). One such application is stress detection (Sanei and Chambers (2007)).

EEG is based on the principle that activity in the brain's neurons generates small changes in electrical potential that can be measured using electrodes on the scalp (Sanei and Chambers (2007)).

Noise and artefact reduction is a critical aspect of EEG analysis, as it enables a more precise understanding of brain activity. Common sources of noise include muscle movements, electrical interference and electrode-related issues (Lai et al. (2018)). Through the application of preprocessing techniques, it becomes possible to remove or minimise these noise sources, leading to a more accurate representation of the brain's activity.

Note: Portions of this section are borrowed or adapted from Sletten (2022). For more in-depth information on EEG, please refer to Sletten (2022).

2.2 Electroencephalography and Mental Stress

The Power Spectral Density (PSD) of an EEG signal offers valuable insights into brain activity, particularly during stress. Research by Hosseini et al. (2010), Khosrowabadi et al. (2011) and Lim and Chia (2015) found a strong correlation between the PSD of EEG signals and stress. This relationship suggests the potential of EEG as a reliable tool for detecting changes in stress levels.

An EEG signal can be decomposed into several frequency bands, namely delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (30-100 Hz). Each band represents a particular range of frequencies thought to reflect different types of brain activity. The cognitive and physiological functions associated with these frequency bands will be further detailed in Section 2.4.

The alpha band has been reported to show decreased activity during periods of stress, according to a variety of studies explored by Katmah et al. (2021). The review of these studies also found that the accuracy achieved when classifying stress based on the alpha band was found to be the highest among the frequency bands. This suggests that alpha oscillations might have a key role in stress detection since they likely reflect the individual's active engagement with the stressor and a disruption in their relaxed mental state.

The beta band, which is generally associated with active, busy or anxious thinking and active concentration, has also shown a significant correlation with stress. Specifically, Seo et al. (2010) found that stress is positively correlated with beta EEG power. This is backed up by the studies explored by Katmah et al. (2021), where the beta frequency band was found to result in the second-best accuracy when used for stress classification, following the alpha band.

Gamma activity has demonstrated varied results, based on a variety of studies explored by Katmah et al. (2021). The review found that decreased gamma activity can be related to both relaxed and stressful situations, suggesting that gamma oscillations may not be particularly sensitive to variations in stress levels. This is consistent with the results presented by Katmah et al. (2021), where the gamma frequency band has a notably worse performance compared with the alpha and beta frequency bands.

Theta band activity showed the lowest accuracy among the EEG frequency bands according to the studies reviewed by Katmah et al. (2021). However, the literature review showed that the frequency band has an accuracy that varies significantly across studies

Lastly, the delta band demonstrated high classification accuracy, as reported by the studies explored by Katmah et al. (2021). However, only one of these studies used the delta band alone for stress classification, meaning that the result should be interpreted carefully.

In conclusion, each frequency band contributes uniquely to the understanding and detection of stress via EEG signals. While some bands, such as alpha and beta, demonstrated superior performance in the reviewed studies, the results across all bands showed considerable variability. This variability, combined with the unique findings related to each band, motivates a comprehensive investigation of all frequency bands in this study's dataset.

2.3 Preprocessing

EEG signals often contain noise and artefacts from various sources, such as power sources, eye blinks, heartbeats and muscle movements (Lai et al. (2018)). Although ideally, one would prevent artefacts from occurring, some noise sources are challenging to eliminate or simply unavoidable (Urigüen and Garcia-Zapirain (2015)). Therefore, preprocessing plays a significant role in enhancing the quality of the acquired EEG signals.

Notch filtering and bandpass filtering are essential preprocessing techniques for improving the quality of EEG signals (Sanei and Chambers (2007)). In this section, we will explore the principles behind these filters and discuss how they are commonly used to preprocess EEG signals.

Notch filtering is a technique that attenuates specific, narrow frequency bands from a signal. A notch filter has a high-pass and a low-pass filter connected in parallel, which results in a frequency response with a dip at the targeted frequency. The idea behind notch filtering is to suppress a particular frequency without affecting the

neighbouring frequencies significantly. The design of the notch filter can be adjusted by controlling the quality factor and the bandwidth of the filter to achieve a more selective or broader attenuation (Oppenheim (1999)). In the context of EEG signal processing, notch filters are primarily used to remove power line interference (50 Hz or 60 Hz, depending on the region) and their harmonics, which can corrupt the EEG signal (Sanei and Chambers (2007)).

Bandpass filtering, on the other hand, is designed to allow only a specific range of frequencies to pass through while attenuating frequencies outside this range. A bandpass filter is essentially a combination of a high-pass filter and a low-pass filter connected in series. The high-pass filter removes low-frequency components, while the low-pass filter eliminates high-frequency components. The resulting output signal contains only the frequencies within the specified range (Oppenheim (1999)). When filtering EEG signals, the frequency range is commonly chosen so that the desired frequency bands, such as delta, theta, alpha, beta and gamma are isolated. This helps eliminate noise from external or physiological sources that lie outside the specified range (Sanei and Chambers (2007)).

2.4 Feature Extraction

Feature extraction is an important part of EEG signal classification since the extracted features have a significant impact on the performance of the classifier and the computational cost. It is essential to choose features that maximise class separability to help ensure that the classifier correctly separates each class. A good feature is one that minimises the intraclass distance and maximises the interclass distance (Rajoub (2020)).

In this study, I have chosen to focus on frequency bands as the features of choice. Frequency bands are useful in EEG signal analysis due to the rich information they provide about the underlying processes in the brain. The frequency bands of an EEG signal can be categorised into various groups, each associated with specific cognitive and physiological functions. The main frequency bands include delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz) and gamma (30-100 Hz) (Sanei and Chambers

(2007)). It is worth mentioning that the literature presents several definitions of the frequency ranges associated with each band.

To compute the power frequency band features, it is necessary to obtain the PSD of the EEG signal. The PSD can be calculated by computing the discrete Fourier transform (DFT) of the signal. There are several methods for computing the DFT of a signal, with the Fast Fourier Transform and Welch's method among the most common techniques.

The PSD is an important measure as it reveals the distribution of the signal's power across different frequency bands. By calculating the PSD of each frequency band, we can extract features that represent the power distribution within each band (Welch (1967)).

2.4.1 Delta Band (0.5-4 Hz)

The delta band resides within the frequency range of 0.5 to 4 Hz. The band is associated with deep sleep and unconscious cognitive activity. It is the slowest brainwave frequency and is predominantly observed during non-rapid eye movement sleep, infancy and in individuals who have suffered brain injuries (Sanei and Chambers (2007)).

2.4.2 Theta Band (4-8 Hz)

The theta band, operating within the range of 4 to 8 Hz, is associated with states of drowsiness, relaxation and the early phases of sleep. The band is most prominently active during cognitive tasks that require creative thinking, meditation and working memory. Furthermore, theta oscillations are believed to play a vital role in memory formation, spatial navigation, and emotional regulation (Sanei and Chambers (2007)).

2.4.3 Alpha Band (8-13 Hz)

The alpha band, spanning from 8 to 13 Hz, is associated with relaxation, mental calmness and alertness. It is often observed in restful states such as when individuals have their eyes closed. Alpha waves are thought to be involved in cognitive processes

such as attention, memory and information processing and are the most prominent rhythms in the brain (Sanei and Chambers (2007)).

2.4.4 Beta Band (13-30 Hz)

The beta band, which ranges from 13 to 30 Hz, is associated with active thinking, focus, problem-solving and decision-making. The presence of beta oscillations is more prominent during tasks that require concentration and cognitive processing. Additionally, beta waves have been linked to motor functions and sensory processing (Sanei and Chambers (2007)).

2.4.5 Gamma Band (30-100 Hz)

The gamma band, ranging from 30 to 100 Hz, is the frequency band containing the highest frequencies. The band is connected to complex cognitive processes, such as perception, attention and learning. Gamma oscillations are believed to play a role in combining information from different brain regions and enabling rapid neural communication (Sanei and Chambers (2007)).

In addition to frequency bands, other feature extraction methods have been explored in the literature for EEG signal classification. These methods can be broadly categorised into time-domain, frequency-domain, time-frequency domain and non-linear methods. Time-domain methods focus on extracting features directly from the EEG signal and include features such as mean, variance, skewness and kurtosis. Frequency-domain methods, like the one used in this study, involve transforming the EEG signal into the frequency domain and extracting features from the PSD of the signal. Time-frequency domain methods combine both time and frequency information to provide a more comprehensive representation of the signal, using techniques such as the Wavelet Transform. Non-linear methods extract features that describe the complexity and irregularity of EEG signals, with fractal features being a common choice (Stancin et al. (2021)).

The choice of frequency bands as the primary features for this study was motivated by the desire to use features that enable clear visualisation of differences be-

tween the two classes. By comparing the PSD of recordings from stressed and non-stressed samples it is possible to effectively determine the differences between these classes in the frequency domain. Another advantage of using frequency bands as features are their strong connection to neurophysiological processes, as established through extensive prior research (Katmah et al. (2021)). Each frequency band is associated with specific cognitive and physiological functions, which allows for a more comprehensive understanding of brain activity during stress.

2.5 Classification Algorithms

Classification is the process of separating entities into distinct classes. When the classification is guided by known class assignments, it is referred to as supervised classification. In contrast, unsupervised classification takes place when class assignments are unknown. Classification tasks can be binary, where data is divided into two classes, or multi-class, where data is divided into multiple classes (Nisbet et al. (2018)).

In this project, the goal is to develop a reliable pipeline for classifying samples as stressed or non-stressed, which constitutes a binary classification task. Since the dataset contains ground truth information in the form of State-Trait Anxiety Inventory Y (STAI-Y) scores and Subjective Stress Assessment (SSA) scores, the classification is supervised.

The classification algorithms that have been chosen for this study are AdaBoost, Random Forest and EEGNet. In this section, we will discuss the fundamental theory behind each algorithm and how they relate to the objective of the study.

2.5.1 AdaBoost

AdaBoost (Adaptive Boosting) is a supervised learning method for classification and regression analysis, which employs an ensemble learning approach. Ensemble learning methods use multiple classifiers to improve generalisation and overall prediction accuracy, and AdaBoost is a prominent example in this category (Polikar (2012)). The method was first introduced by Freund and Schapire in 1996 (Freund et al. (1996)).

As the name of the method implies, AdaBoost is a learning method that makes use of boosting to improve the performance of the model. Boosting works by repeatedly using weak classifiers (e.g., decision trees) that perform slightly better than random guessing to classify different subsets of the data. For each iteration, misclassified samples are weighted more heavily, which ensures that the classifier focuses on these instances. This iterative refinement improves the accuracy of the classifier over time (Freund et al. (1996)).

The weak classifiers are combined using a weighted sum, where the accuracy of each classifier determines its weight. The final prediction is determined by a weighted vote of the individual classifier predictions (Freund et al. (1996)).

Figure 2.1 shows a flowchart of the main steps of the AdaBoost algorithm, showcasing how weak classifiers are trained using a subset of the data and combined using weighted sums of their outputs.

AdaBoost has been used successfully for a variety of EEG-based applications. Recent studies have demonstrated its effectiveness in detecting fatigue in drivers, diagnosing epilepsy and stress detection (Hu (2017); Prabhakar and Rajaguru (2018); Liu et al. (2022)). Based on these promising results, AdaBoost was identified as a suitable classifier for this study.

AdaBoost has several desirable properties that makes it well-suited for EEG analysis. One of these properties is the ability to reduce overfitting, which is particularly important when working with small and imbalanced datasets, which is often the case with many EEG datasets. The resampling method increases the diversity of the classifiers by using different subsets of the training data for each weak classifier. This helps prevent the model from overfitting to any specific subset of the data. Additionally, the increased weighting of misclassified samples ensures that the classifiers in the ensemble focus more on the samples that are harder to classify, preventing the model from relying too heavily on easily learned samples (Freund et al. (1996)).

Another advantage of AdaBoost is its ability to handle high-dimensional feature spaces, which is relevant for EEG analysis where multiple features are often computed across a large number of channels. This can result in high-dimensional data, which can be challenging to analyse with other machine learning algorithms. AdaBoost can

effectively handle high-dimensional feature spaces by using an ensemble of classifiers that each only focus on a subset of the features (Freund et al. (1996)).

2.5.2 Random Forest

Random Forest is a supervised learning method for classification and regression analysis, which employs an ensemble learning approach. The model combines multiple decision trees to improve the accuracy and generalisation of the classification and regression. The model was first introduced by Breiman and Cutler in 2001 as an extension to the decision tree algorithm (Breiman (2001)). The key principles of Random Forest are bagging and feature randomness, which combine to create a diverse ensemble of decision trees (Breiman (2001)).

Bagging, an abbreviation of bootstrap aggregation, involves generating several random subsets of the training data with replacement. Each of the training sets are used to build an individual decision tree. The final prediction of the Random Forest model is obtained by aggregating predictions from each decision tree through a majority vote (Breiman (1996)).

The Random Forest algorithm extends upon the bagging technique by using a technique known as feature randomness. In addition to sampling a random subset of the training data, each decision tree uses a random subset of the features at each split of the tree. The combination of randomness at the data level and feature level increases the diversity of the decision trees, aiming to improve accuracy and generalisation performance (Breiman (2001)).

Random Forest has several properties that make it highly suitable for EEG analysis. One of these properties is its ability to prevent overfitting. As mentioned in Subsection 2.5.1, preventing overfitting is vital for small and imbalanced datasets, a common characteristic of EEG datasets. Random Forest prevents overfitting by employing randomness at both the data and feature level, preventing the model from relying too heavily on specific features or subsets of the data (Breiman (2001)).

Another key property of Random Forest is its ability to handle high-dimensional data, a trait it shares with AdaBoost, as described in Subsection 2.5.1. This capability

is important for EEG data, where high-dimensional feature spaces are often encountered. The algorithm's feature selection process contributes to the algorithm's ability to tackle high-dimensional data. The feature selection process uses a random subset of the features at each split of the decision trees, instead of the complete feature set, which helps manage the complexity of high-dimensional data in an effective manner (Breiman (2001)).

The Random Forest classifier has been demonstrated as a successful approach for analysing EEG data in various applications, as shown in several previous studies. For instance, Random Forest has achieved good results in applications such as classifying sleep stages, early-stage seizure detection and driving-induced stress detection (Fraiwan et al. (2012); Donos et al. (2015); Halim and Rehan (2020)). These results, combined with the properties mentioned previously, suggest that Random Forest can be an effective tool for analysing EEG data.

Figure 2.2 shows a flowchart of the main steps of the Random Forest algorithm, showcasing how decision trees are trained on a subset of the data and their predictions aggregated to generate the final output.

2.5.3 EEGNet

In recent years, the use of convolutional neural networks (CNNs) has become an increasingly common choice for classifying EEG signals. CNNs have gained popularity partly due to their ability to automatically extract features, eliminating the need for manual feature extraction.

EEGNet, a CNN specifically engineered for use with EEG data, has shown remarkable performance across a variety of EEG classification tasks, including stress detection, epilepsy diagnosis and event-related potential identification as shown in recent studies (Bhatnagar et al. (2023); Shoji et al. (2021); Lawhern et al. (2018)). The versatility of EEGNet combined with the accuracy the model has achieved constitute some of the main reasons for choosing this model in the study.

In Figure 2.3, the model architecture of EEGNet is depicted, while Table 2.1 provides a detailed description of the architecture.

The first block in the model makes use of two convolutional steps in sequence, a 2D convolutional layer and a depthwise convolutional layer. The 2D convolutional layer is designed to capture frequency information from the EEG data, with the kernel size chosen to be half of the sampling frequency of the EEG recording. This kernel size ensures the capture of frequency information at 2 Hz and above. The number of temporal filters is denoted F_1. The depthwise convolutional layer aims to learn spatial filters for each of the F_1 feature maps generated by the 2D convolutional layer. The depth parameter, D, determines the number of spatial filters learned from each feature map (Lawhern et al. (2018)).

The second block in the model contains a separable convolutional layer, which is a depthwise convolution followed by a pointwise convolution (Chollet (2017)). This design reduces the number of trainable parameters while also learning temporal features specific to each feature map and then combining them optimally. The number of pointwise filters is determined by the parameter F_2 (Lawhern et al. (2018)).

The final block of the EEGNet model is responsible for classifying the input data. Since this study focuses on solving a binary classification problem, the sigmoid activation function is used instead of the softmax function, as was originally proposed by the authors in Lawhern et al. (2018).

For further information regarding EEGNet, readers are encouraged to consult the original paper by Lawhern et al. (2018).

Block	Layer	Size	# Params	Output	Activation	Options
1	Input	-	-	(C, T)	-	
	Reshape	-	-	(C, T, 1)	-	
	Conv2D	(1, 125)	$125 \cdot F_1$	(C, T, F_1)	Linear	mode=same
	BatchNorm	-	$4 \cdot F_1$	(C, T, F_1)	-	-
	DepthwiseConv2D	(C, 1)	$C \cdot D \cdot F_1$	(1, T, $D \cdot F_1$)	Linear	mode=valid, depth=D, max norm=1
	BatchNorm	-	4	(1, T, $D \cdot F_1$)	-	-
	Activation	-	-	(1, T, $D \cdot F_1$)	eLU	-
	AveragePool2D	(1, 8)	-	(1, $T//8$, $D \cdot F_1$)	-	-
	Dropout	-	-	(1, $T//8$, $D \cdot F_1$)	-	p=0.25
2	SeparableConv2D	(1, 32)	$32 \cdot D \cdot F_1 + F_2 \cdot D \cdot F_1$	(1, $T//8$, F_2)	Linear	mode=same
	BatchNorm	-	$2 \cdot D$	(1, $T//8$, F_2)	-	-
	Activation	-	-	(1, $T//8$, F_2)	eLU	-
	AveragePool2D	(1, 16)	-	(1, $T//128$, F_2)	-	-
	Dropout	-	-	(1, $T//128$, F_2)	-	p=0.25
	Flatten	-	-	($F_2 \cdot T//128$)	-	-
3	Dense	-	$F_2 \cdot T//128 + 1$	(N)	Sigmoid	max norm= 0.25

Table 2.1: EEGNet architecture, where C is the number of channels, T is the length of the time series for each sample, D is the depth multiplier, F_1 is the number of temporal filters, F_2 is the number of pointwise filters and N is the number of classes.

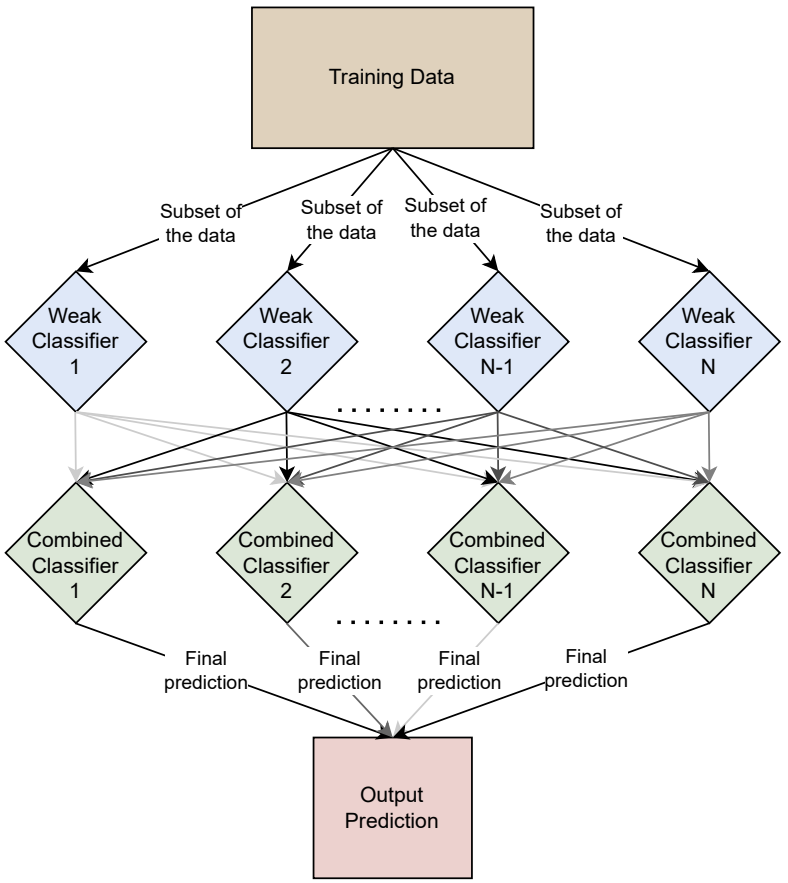


Figure 2.1: Flowchart depicting the AdaBoost algorithm, highlighting the most important stages in training and prediction. The top rectangle, coloured orange, represents the input data. The subsequent layer consists of diamond-shaped blue elements, representing the weak classifiers. The weak classifiers connect to the combined classifiers in the layer below, depicted as diamond-shaped green elements. The combined classifiers are then linked to the output prediction layer in the final step. Arrows between the boxes illustrate the flow of data and decisions, with varying arrow colour intensities indicating the weights assigned to each classifier. Darker arrow colours correspond to heavier weights, while lighter arrow colours correspond to lighter weights.

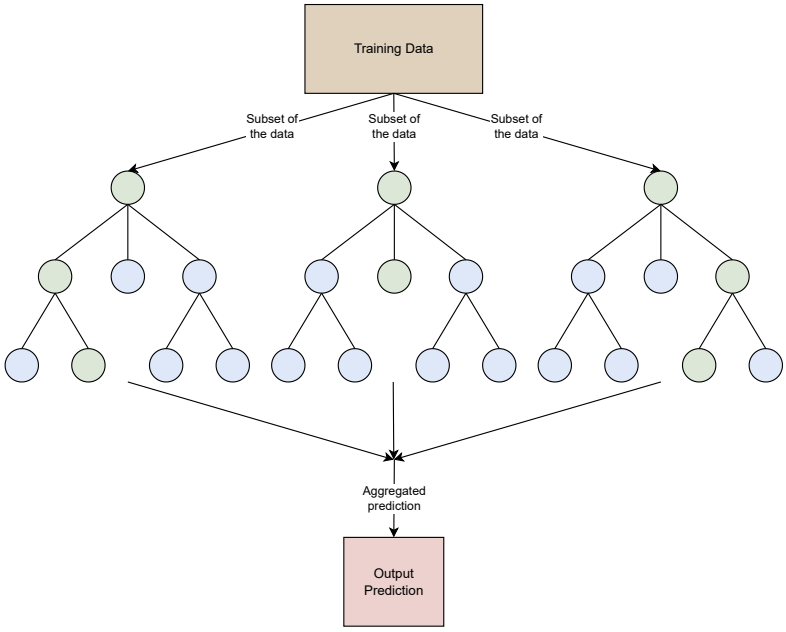


Figure 2.2: Flowchart depicting the Random Forest algorithm, highlighting the most important stages in training and prediction. The key stages shown in the figure include data bootstrapping, decision trees with random features at each split and aggregation of individual tree predictions to produce the final prediction. The figure shown uses an ensemble of 3 decision trees.

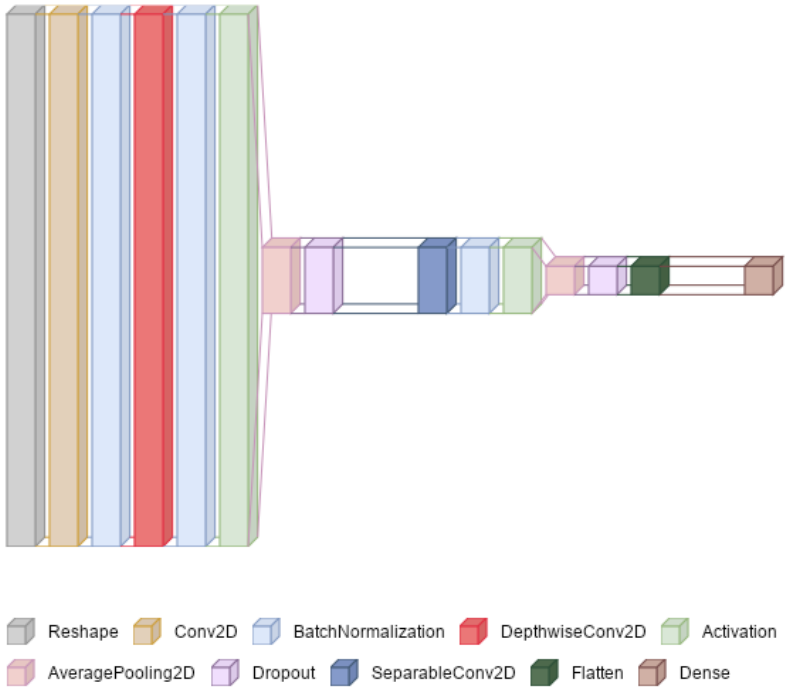


Figure 2.3: The layered diagram showcases the EEGNet architecture, with each layer colour-coded for clarity: reshape layers in light grey, 2D convolutional layers in orange, batch normalisation layers in light blue, depthwise convolutional layers in red, activation layers in light green, average pooling layers in pink, dropout layers in light purple, separable convolutional layers in dark blue, flatten layers in dark green and dense layers in light brown. The division into distinct blocks is visually highlighted using increased padding between layers. The first block employs a temporal convolution to learn temporal filters and a depthwise convolution to learn frequency-specific spatial filters. The second block uses a separable convolution, combining depthwise and pointwise convolution, to capture the spatial relationship between different frequency bands (Lawhern et al. (2018)).

Chapter 3

Materials and Methods

3.1 Data Collection

3.1.1 Experiment Overview

The aim of the data collection is to collect data that can be used to investigate the potential of using EEG and Phonocardiography (PCG) signals as modalities for stress detection, both independently and in combination. The overarching goal is to develop more accurate and effective stress detection systems.

The data was collected from a sample of 28 healthy participants aged between 20 and 28 years, with an average age of 23.1 years. There were two sessions in which PCG and EEG signals were collected. The initial session was conducted before the participants' exam, while the second session took place after their return from holiday. The purpose of the two sessions is to capture a baseline condition and exam-related stress for each participant.

During each session, two runs were performed. The first run was designed to induce a relaxed state in the participants, while the second was designed to induce a stressed state by subjecting the participant to an arithmetic test. Markers were recorded automatically each time the participant responded to an arithmetic question.

In addition to collecting EEG and PCG signals, participants were asked to complete the STAI-Y form and to rate their stress levels on a scale from 1 to 10 to gauge their level of mental stress. The 1-10 self-assessment scale, with 1 being the least stressed and 10 being the most stressed, will be referred to as the SSA. Both STAI-Y and SSA are widely used to assess mental stress levels and have been shown to have good reliability and validity in previous studies (Bergua et al. (2012); Silverstein and Kritz-Silverstein (2010)). The STAI-Y form is presented in Section A.2.

To gain a thorough understanding of the participants' stress levels, the STAI-Y form was completed at the beginning and the end of each session. In addition, participants were asked to rate their mental stress level after each run using the SSA scale. Combining two measures of mental stress can help provide a more complete picture of the participants' stress levels.

The collected dataset provides a foundation for measuring the effectiveness of EEG and PCG signals for stress detection and assists in advancing our understanding of the multimodal approach to stress detection. Investigating the use of these modalities both independently and in combination could contribute to the development of more accurate and effective stress detection systems.

3.1.2 Participants

A cohort of 28 participants aged between 20 and 28 years, with an average age of 23.1 years was recruited for the experiment. The sample was comprised of 16 males and 12 females, resulting in a roughly equal gender distribution. All participants were students enrolled at the Norwegian University of Science and Technology (NTNU). Before being included in the study, the participants completed a screening questionnaire to confirm that they were in good health and did not have any known cardiac, neurological, or mental health disorders that could impact the study results. Additionally, the participants were screened to ensure that they were not taking any medications that could potentially affect brainwaves or interfere with the study. The questionnaire used to screen the participants is presented in Section A.3.

3.1.3 Equipment and Software

Data collection for this study was conducted using the Mentalab Explore EEG device from Mentalab and the Eko DUO ECG + Digital Stethoscope from Eko. The Mentalab Explore EEG device features 8 channels and support for wet and dry electrodes. The EEG signals were sampled at a frequency of 250 Hz and dry electrodes were used due to their ease of use and reduced preparation time. The Eko DUO ECG + Digital Stethoscope from Eko was used to record PCG signals, which were sampled at a frequency of 22050 Hz. To optimise EEG signal quality, we used a range of tools to prepare the scalp and electrodes. Specifically, we cleaned the scalp with isopropyl alcohol and applied an electrical conducting gel using Q-tips. We also secured the reference node on the ear lobe with tape and a clothespin. Additionally, to ensure high-quality PCG recordings, a tie-down strap was used to keep the stethoscope securely in place.

For the purpose of streaming EEG data, PCG data and generating markers, three computers were used. The ExplorePy software from Mentalab (*ExplorePy* (2019)) was utilised to stream the EEG data, while AudioCapture (*AudioCapture* (2018)) was employed to stream the PCG data. The markers were generated using PsychoPy (*PsychoPy* (2015)). To ensure synchronisation of the data streams, we employed Lab Streaming Layer (*Lab Streaming Layer* (2011)).

A photo taken during one of the recording sessions, providing a visual representation of the experiment setup, is shown in Figure 3.1.

3.1.4 Data Collection Protocol

After screening out participants based on the exclusion criteria, eligible participants were informed via email and given the option to select a convenient time slot for the data collection. The participants were also instructed to avoid hair products and wear a t-shirt on the day of the measurements to prevent interference with the recordings.

Upon arrival, participants were provided with detailed information about the data collection protocol and offered the opportunity to review and sign an informed consent form. The informed consent form is presented in Section A.1. After signing the informed consent form, participants were asked to complete the STAI-Y form, which

Figure 3.1: A photograph taken during a recording session. The participant is seen wearing the Mentalab Explore EEG device, identifiable by the blue cap. The Eko DUO PCG device is securely fastened to the participant's chest using the buckle strap shown on the participant's back. The monitor placed in front of the participant is used to display the arithmetic questions.

will be used to label the recorded data.

Next, the participants were fitted with the Mentalab Explore EEG device. To ensure optimal conductivity, the electrodes were carefully cleaned using isopropyl alcohol prior to placement. Afterwards, an electrical conductivity gel was applied to each electrode to further improve signal quality. During the setup process, the electrical impedance of each electrode was continuously monitored with help from the ExplorePy software (*ExplorePy* (2019)). The `impedance` command was used to visualise the electrical impedance in a browser dashboard.

The placement of the EEG channels was determined based on previous analysis conducted by Marthinsen (2022). In the study, a genetic algorithm was utilised to select the optimal combination of 8 EEG channels which demonstrated the highest accuracy based on a dataset consisting of 32-channel EEG signals. A detailed description of the process used to determine the optimal set of EEG channels can be found on page 30 of the publication by Marthinsen (2022). The electrode placements used in this study, as found to be the best in the dataset used by Marthinsen (2022), are shown in Figure 3.2, used with permission from the author of Marthinsen (2022). The chosen positions are highlighted with bold outlines.

The PCG device was placed on the participant's chest, above the location of the heart and on top of their t-shirt. To minimise motion artefacts, a tie-down strap was used to securely fasten the PCG device in place. The mobile application Eko: Digital Stethoscope, ECG, which connects to the PCG device via Bluetooth, was used to ensure that the signal strength from the PCG device was satisfactory (*Eko: Digital Stethoscope, ECG* (2019)).

To verify that data from both the EEG device and the PCG device were being accurately captured and synced, an initial test was performed where data was collected for approximately 10 seconds and plotted for analysis. This verification allowed for quick identification of potential issues that may have arisen during the setup process.

During the first run of data collection, participants were instructed to maintain a relaxed state with their eyes open, minimise muscle movement and refrain from speaking. The first run of data collection lasted for five minutes. After completing the recording, the participants were asked to provide a self-assessment of their mental

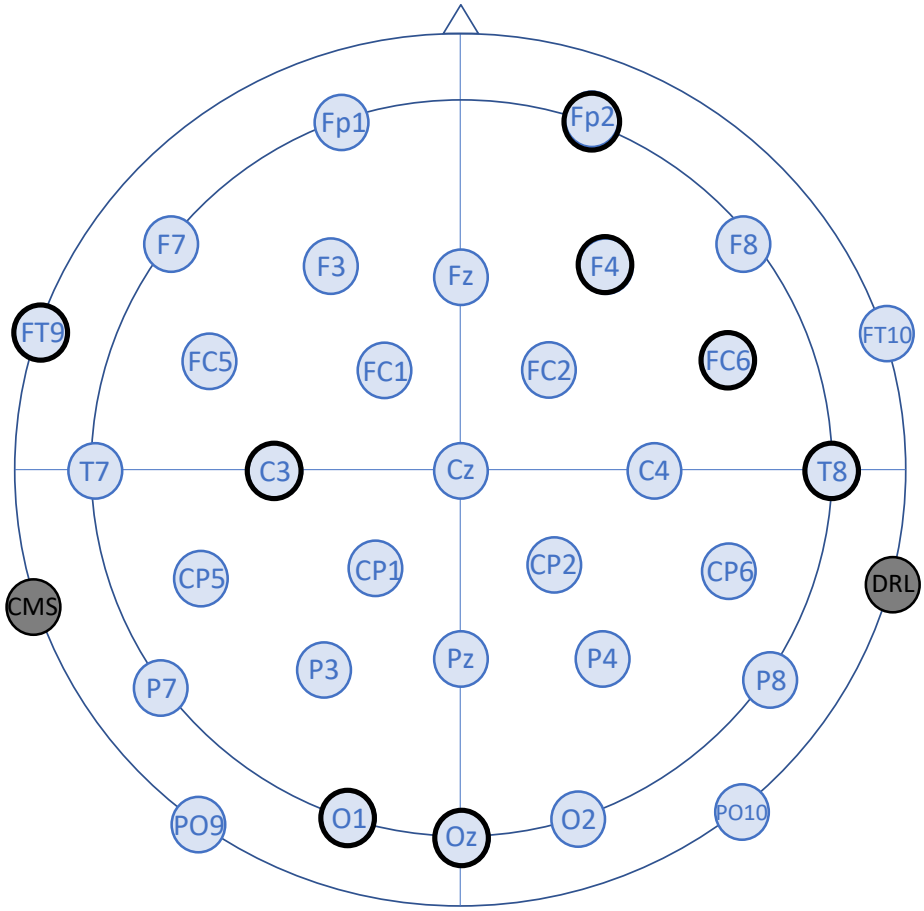


Figure 3.2: The best electrode placements when using 8 electrodes, as determined by a genetic algorithm applied to a dataset analysed by Marthinsen (2022). The chosen positions, F4, Fp2, C3, FC6, O1, Oz, FT9 and T8, are outlined in bold.

stress level on a scale ranging from 1 to 10.

During the second run of data collection, participants were presented with arithmetic problems designed to induce stress. The problems involved basic arithmetic operations, such as addition, subtraction, multiplication and division, and used numbers smaller than 100. No decimal numbers were included in the problems. Although there was no time pressure imposed on the participants, the nature of the arithmetic test was still expected to create a stressed state.

Feedback on the participants' performance was not provided during the test. After completing the recording, participants were given feedback on their performance only when requested. This approach ensured that the stress experienced during the test was primarily induced by the arithmetic problems themselves and not by the feedback or time constraints.

The participants were instructed to minimise any muscle movements while using the keyboard placed in front of them to confirm the validity of the arithmetic equality presented on the computer monitor. An example of an arithmetic problem used for stress induction is shown in Equation 3.1. During the arithmetic test, markers were recorded for each keystroke made by the participant in response to an arithmetic question. The second run of data collection also lasted five minutes per participant.

$$6 \times 7 - 8 \times 9 = -32 \tag{3.1}$$

After completing the second recording, the participants were again asked to provide a self-assessment of their mental stress level on a scale ranging from 1 to 10. Following the completion of both data collection runs, the participants were asked to complete the STAI-Y form again. After completing the STAI-Y form, the EEG and PCG equipment was gently removed from the participants. Any excess conductive gel was removed from the hair of the participants by using damp paper towels. After completing the study protocol, the participants were asked to confirm their well-being and were thanked for their participation in the study. As a token of appreciation for their participation, participants were compensated with a gift card worth 250 NOK after the completion of the second session.

The data collection protocol is illustrated in Figure 3.3 through a flowchart.

3.1.5 Data Storage and File Structure

The collected data was stored in a specific file structure to facilitate easy organisation and retrieval. The data files were organised as follows:

- A root folder named `multimodal_data` was created to store all data files related to the study.
- Inside the `multimodal_data` folder, separate folders were created for each type of data: raw EEG data, filtered EEG data, raw PCG data, marker data and labels. These folders were named `eeg_raw`, `eeg_filtered`, `pcg`, `markers` and `labels`, respectively.
- The file names in the `eeg_raw`, `eeg_filtered`, `pcg` and `markers` folders follow the format `P0XX_S00X_00X.npy`, where `P0XX` represents the participant number, `S00X` indicates the session number and `00X` corresponds to the run number.
- The `labels` folder contains a single file named `labels.pkl`.

The data was stored in the following file formats:

- EEG data: stored in the `.npy` format
- PCG data: stored in the `.npy` format
- marker data: stored in the `.npy` format
- STAI-Y and SSA scores: stored as a Pandas DataFrame in the `.pkl` format

The EEG data, PCG data and markers were stored in the `.xdf` file format during the experiment. To improve read performance and lower memory usage the EEG and PCG data were converted from the `.xdf` format to the more efficient `.npy` format. Similarly, the markers were converted from the `.xdf` format to the more efficient and less memory intensive `.pkl` format. The conversion was performed using the `pyxdf` package in Python (*pyXDF* (2019)).

Figure 3.3: Flowchart of the experimental protocol for data collection using EEG and PCG signals. The preliminary steps, participant screening, recruitment and time slot selection are shown in light grey. Sessions 1 and 2 are represented in light blue and light red, respectively, each incorporating: signing of informed consent, STAI-Y completion, setup, two recording stages (without and with a stressor), another STAI-Y completion and debriefing. The recording stages, Run 1 and Run 2, are emphasised with darker shades of blue and red. The arrows indicate the sequence of steps, and the blocks represent each step in the protocol.

3.1.6 Data Exclusions and Multimodal Dataset

Since the dataset is multimodal, both EEG and PCG is required for each. During the data collection process, certain PCG samples were lost due to recording issues. To maintain consistency within the multimodal data, the corresponding EEG data has also been excluded. Additionally, two participants were absent from the second session, leading to missing data for those specific runs.

Out of the initially planned 112 runs, 25 runs were excluded due to PCG recording issues and 4 runs were lost due to participant absence, resulting in a total of 29 excluded runs. It is important to note that the dataset only consists of runs which have both EEG and PCG data. This ensures that the dataset is consistent for further research and analysis. The participants with missing data due to absence have not been excluded from the dataset as a whole, and their data from the other session remains part of the analysis.

Only the EEG data from the multimodal dataset will be used in this study. This means that certain EEG recordings will not be used, which will potentially lead to reduced classification performance since the models have less data to train on. The primary objective of using the multimodal dataset is to establish a baseline for a single-modal system, which can be compared to future multimodal systems. This will allow future research to focus on potential performance improvements of a multimodal classification system that incorporate both EEG and PCG data.

Lastly, a number of recordings were impacted by considerable background noise, primarily caused by ongoing nearby construction activities during the data collection process. Unfortunately, we did not catalogue which specific recordings were exposed to the increased level of background noise. While noise reduction techniques have been employed to counteract these effects, it is vital to recognise that these strategies may not completely remove all noise-induced artefacts. As a result, this could introduce some degree of variability in the analysis results that should be accounted for during data interpretation.

3.1.7 Ethical Considerations

Several ethical considerations were taken to ensure the comfort, safety and privacy of the participants. These considerations were based on the principles of the Helsinki Declaration (Association. (2001)), an established set of ethical principles for studies involving human subjects. The principles include obtaining informed consent, ensuring the well-being of participants and protecting their privacy and confidentiality.

All participants were provided with informed consent forms to sign and were informed about the purpose and procedures of the study, as well as any benefits or risks. Participants were assured that their participation was voluntary and that they could withdraw from the study at any time without penalty.

Throughout the study, the comfort and well-being of the participants was prioritised. Great care was taken to avoid discomfort when applying the EEG electrodes, and the participants were given the opportunity to request adjustments to the electrodes or other elements of the setup during the study.

Proper data management was a top priority throughout the data collection period and after the completion of the study. While the data collection was ongoing the data was stored locally on one of the computers used for the experiment. After the study was completed, the data was transferred to an access-controlled Microsoft Sharepoint folder to allow for simpler collaboration, while also preventing unauthorised access. At all times, the data was stored in an anonymised way by assigning each participant a unique identification number.

For the purpose of data analysis, a spreadsheet mapping the identification numbers to the participants' names and contact details was temporarily stored in the same access-controlled SharePoint folder. A decision on how the data will be stored and managed is being actively evaluated to ensure compliance with ethical guidelines.

3.2 Preprocessing

The main objective of preprocessing the raw EEG data is to enhance the quality and reliability of the signals by removing noise and artefacts. This is achieved by using a

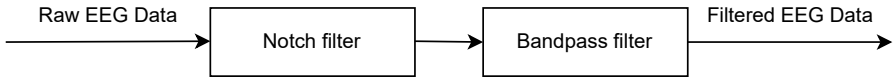


Figure 3.4: Flowchart of the preprocessing pipeline. Raw EEG data is first subjected to notch filtering, which is employed to remove the common mode noise, while bandpass filtering retains frequency information within a specified range.

combination of notch filtering and bandpass filtering. The code for the filtering can be found in the project's GitHub repository, as referenced by Sletten (2023).

Notch filtering was applied to attenuate specific frequencies from the original signal. In this study, the frequency of the power grid, which is 50 Hz in Norway, is attenuated by the notch filter. The signal interference from the power grid is known as common mode noise. The notch filter was implemented using the `mne.io.Raw.notch_filter` function from the `mne` package (*MNE-Python* (2011)).

Bandpass filtering of the raw EEG data aimed to isolate specific frequency ranges of interest while attenuating frequencies outside the filter's frequency band. In this study, the frequency range between 0.5 Hz and 70 Hz was selected for bandpass filtering. The choice of this range was based on several considerations. First, the range includes the frequencies most vital for EEG analysis, which typically ranges from 0.5 Hz to 100 Hz, as explained in Section 2.4. Second, the range was chosen to optimise the performance of the classifiers in this study, namely AdaBoost, Random Forest and EEGNet. It was observed that the models' performance improved when the bandpass filter was applied within the range of 0.5 Hz to 70 Hz, as opposed to 0.5 Hz to 100 Hz. Bandpass filtering was employed using the `mne.io.Raw.filter` function from the `mne` package.

Figure 3.4 provides an overview of the filtering pipeline, which includes the application of notch filtering followed by bandpass filtering.

A time-domain comparison of the EEG signals before and after filtering is shown in Figure 3.5 for participant P024 during session 2, run 1. The plot represents the average of the 8 collected channels throughout the recording period. Examining the raw data plot reveals that the EEG signal appears to be affected by drift, as indicated by its

gradual change across the duration of the recording. Several factors could potentially contribute to the drift observed in Figure 3.5, including unstable contact between the electrode and the scalp, or a drift in the recording equipment (Sanei and Chambers (2007)). However, the filtered data in Figure 3.5 displays a significant reduction in the signal drift, implying that the drift could be caused by frequency components outside the bandpass filter's frequency range. Figure 3.5 demonstrates the effectiveness of the preprocessing methods employed on the raw EEG signals in the time domain.

A comparison of the PSD of the EEG signals before and after filtering is presented in Figure 3.5 for participant P024 during session 2, run 1, focusing on a specific channel: channel F4. Only a single channel is shown to provide a clearer visual representation of the PSD. The raw EEG data shows three prominent peaks in the frequency spectrum, located approximately at 0 Hz, 50 Hz and 100 Hz. The peak in the amplitude of the raw EEG signal around 0 Hz could be attributed to DC offset, a low-frequency component present even in the absence of brain activity. Factors such as electrode polarisation, skin impedance or contact resistance between the electrode and the scalp can cause this DC offset. The peaks observed at 50 Hz and 100 Hz likely from the electrical power supply and its harmonics. The harmonics of the electrical power supply are multiples of its frequency, such as 100 Hz, 150 Hz and 200 Hz.

3.3 Feature Extraction

Feature extraction for this project is implemented using the Python package `MNE-features` which contains a large set of functions designed for computing features from magnetoencephalogram (MEG) and EEG data. The package documentation can be found at the provided citation *MNE-Features* (2018).

The first stage of the feature extraction process was focused on finding the frequency bands that most effectively discriminated the data from the classes. In this stage, the performance of the frequency bands detailed in Section 2.4, delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-70 Hz), was assessed. Note that the range of the gamma frequency band has been limited to the range 30 to 70 Hz, compared to the range 30 to 100 Hz presented in Section 2.4, be-

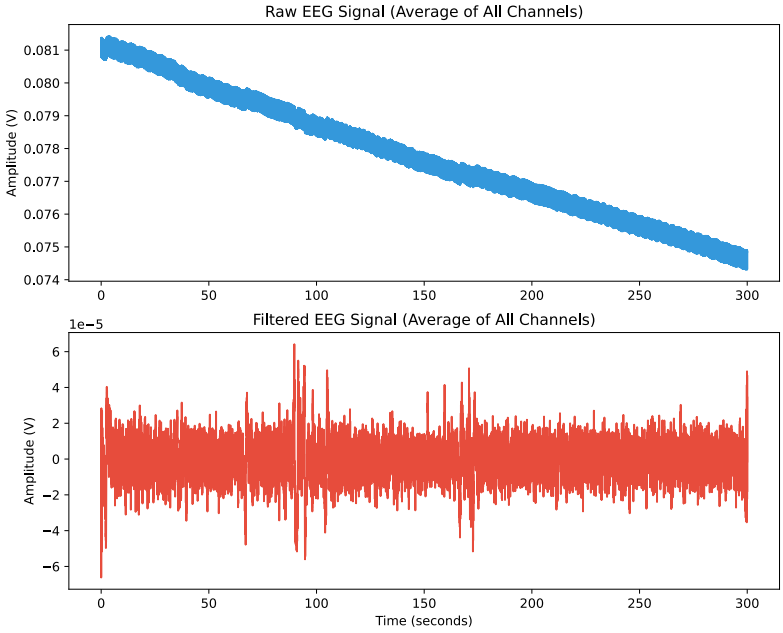


Figure 3.5: Comparison of the raw and filtered EEG signals for participant P024, session 2, run 1, averaged across all channels. The top plot shows the raw EEG data, and the bottom plot shows the filtered data.

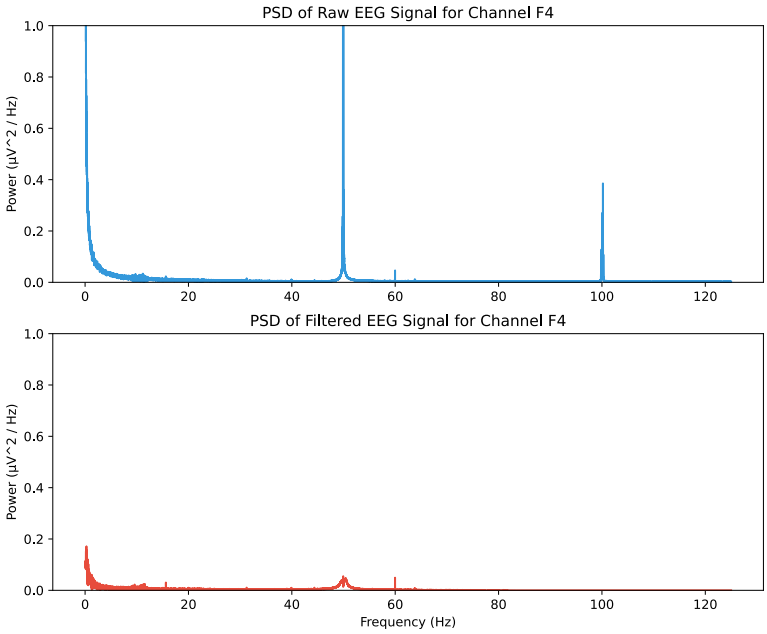


Figure 3.6: Comparison of the PSD of the raw and filtered EEG signals for participant P024, session 2, run 1, for channel F4. The top plot shows the PSD of the raw EEG data, and the bottom plot shows the PSD of the filtered data. The PSD plot is a way to visualise the frequency content of the EEG signal, with higher power values indicating greater signal activity at a given frequency.

cause the EEG signals have been bandpass filtered in the range 0.5 to 70 Hz and, consequently, any information above 70 Hz has been removed from the signal. The power frequency bands were computed using the `compute_pow_freq_bands` function from the `MNE-features` package. The computed power frequency bands were then classified using the AdaBoost and Random Forest classifiers, with their implementations discussed in detail in Section 3.4. The same classification pipeline described in Subsection 3.4.2 was used, incorporating an 80-20 train-test split. Data segmentation was performed using a 2-second epoch length.

Figure 3.7 presents a comparison of the AdaBoost and Random Forest classifier accuracies for the delta, theta, alpha, beta and gamma frequency bands. The classification results suggest that the delta, theta, alpha and beta power frequency bands are best at discriminating the stressed and non-stressed samples. Consequently, the figure shows a decline in the classification accuracy when the gamma power frequency band is used.

To further compare the discriminative ability of the frequency bands, Figure 3.8 displays the PSD of a stressed and non-stressed recording, with individual subplots representing each channel. The recording channels used in this study were F4, Fp2, C3, FC6, O1, Oz, FT9 and T8, and an illustration of the electrode positions is shown in Figure 3.2. These recordings are from participant P024 during session 2; the non-stressed recording is taken from run 1, while the stressed recording is from run 2. The most pronounced differences between the PSDs of the non-stressed and stressed EEG data can be observed in channels F4, Fp2, C3, FC6 and T8. The differences between the non-stressed and stressed PSDs are most apparent within the 0.5 to 10 Hz frequency range, which encompasses the delta (0.5-4 Hz), theta (4-8 Hz) and alpha (8-12 Hz) frequency bands. This observation is largely consistent with the findings in Figure 3.7, where lower-frequency bands yield superior classification accuracy.

To gain a more comprehensive understanding of why certain channels have a better discriminative effect for certain frequency ranges, Figure 3.9 plots a topographical map of the PSD for the delta, theta, alpha, beta and gamma frequency bands, using the same non-stressed and stressed recordings used in Figure 3.8.

In the delta frequency band's topographic map, the stressed sample demonstrates

increased activity near the Fp2 electrode compared to the non-stressed sample. This observation aligns with previous findings for Figure 3.8, where the Fp2 channel exhibits the most significant difference between the PSDs of non-stressed and stressed data. The increased neural activity close to the Fp2 electrode is also apparent for the theta and alpha frequency bands. Similarly, the F4 electrode position shows increased activity for the stressed sample in the delta, theta and alpha bands.

Channels FC6 and T8 display variations in electrical activity between non-stressed and stressed recordings within the delta frequency band. This corresponds well to the results presented in Figure 3.8, where these channels show their best discriminative ability in the low-frequency range below 4 Hz. This finding is in line with the results shown in Figure 3.8, where these channels demonstrate their highest discriminative capabilities in the low-frequency range below 4 Hz. Channel C3 reveals a marginal increase in activity in the delta frequency band, corroborating the outcomes displayed in Figure 3.8. For the remaining channels, discerning differences in electrical activity between non-stressed and stressed topographic maps proves challenging.

To further improve the classification accuracy, channel selection was performed with the aim of removing redundant features. To explore which combinations of channels achieved the best classification accuracy, a simple feature selection technique was employed.

The channels that showed the least effective discriminative ability were systematically removed in a sequential manner. The ability of these channels to distinguish between classes was determined by referring to 3.8 and 3.9, which provide detailed information on the PSD and topographic maps, respectively. By examining these figures it was possible to identify and eliminate the channels that displayed minimal differences in the PSD between stressed and non-stressed states across the frequency spectrum.

The following list presents the order in which channels were removed during the channel selection process:

1. Oz
2. O1

3. FT9
4. C3
5. FC6
6. T8
7. F4

The performance of the AdaBoost classifier and the Random Forest classifier across the channel combinations and feature bands are shown in Table 3.1 and Table 3.2, respectively. The accuracy, sensitivity and specificity of the models are measured.

Table 3.1, presenting the performance of the AdaBoost classifier, shows that the model has low specificity for most of the combinations of channels and frequency bands. In general, the table suggests that the specificity increases and the sensitivity decreases as the number of channels decreases. The chosen combination of channels and frequency bands should be based on all the performance metrics shown in the graph. As a result, the channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8 combined with the alpha frequency bands appears to be the best combination of these performance metrics, achieving an accuracy of 70%, sensitivity of 20% and specificity of 90%.

Table 3.2, presenting the performance of the Random Forest classifier, also shows that the model has low specificity for most of the combinations of channels and frequency bands. The accuracy of the classifier decreases on average as the number of channels decreases. In contrast with the AdaBoost classifier, the sensitivity increases on average as the number of channels decreases. Based on the aim of achieving high accuracy and a good sensitivity/specificity trade-off, the channels F4, Fp2 and T8 were chosen together with the beta frequency band, achieving an accuracy of 68%, sensitivity of 17% and specificity of 89%.

Note that the method used for channel selection is very simple and has several drawbacks compared to other, more sophisticated methods. This channel selection method does not consider all possible channel combinations, and the evaluation of which channels have the best discriminative ability is only based on PSD plots and

Channels/Frequency Bands		Delta	Theta	Alpha	Beta	Gamma
F4, Fp2, C3, FC6, O1, Oz, FT9, T8	Accuracy	70%	70%	70%	66%	60%
	Sensitivity	8%	7%	20%	18%	47%
	Specificity	94%	95%	90%	85%	65%
F4, Fp2, C3, FC6, O1, FT9, T8	Accuracy	69%	70%	69%	67%	58%
	Sensitivity	5%	7%	17%	16%	47%
	Specificity	94%	95%	90%	87%	62%
F4, Fp2, C3, FC6, FT9, T8	Accuracy	70%	70%	68%	71%	64%
	Sensitivity	3%	6%	14%	11%	31%
	Specificity	96%	95%	90%	95%	78%
F4, Fp2, C3, FC6, T8	Accuracy	70%	70%	69%	70%	64%
	Sensitivity	6%	2%	13%	11%	30%
	Specificity	96%	97%	92%	94%	78%
F4, Fp2, FC6, T8	Accuracy	71%	70%	70%	72%	68%
	Sensitivity	2%	4%	12%	8%	13%
	Specificity	99%	96%	93%	97%	90%
F4, Fp2, T8	Accuracy	71%	70%	71%	72%	67%
	Sensitivity	2%	3%	7%	11%	11%
	Specificity	99%	97%	97%	96%	90%
F4, Fp2	Accuracy	71%	71%	70%	72%	69%
	Sensitivity	1%	1%	5%	9%	5%
	Specificity	100%	99%	95%	98%	95%
Fp2	Accuracy	71%	71%	71%	71%	71%
	Sensitivity	0%	0%	1%	0%	0%
	Specificity	100%	100%	99%	100%	100%

Table 3.1: Channel selection performance for the AdaBoost classifier. The table demonstrates the classification accuracy, sensitivity and specificity achieved using different combinations of channels, which were systematically removed based on their discriminative ability. The channels were removed according to the order presented in the list in Section 3.3, with the goal of optimising classification performance across the delta, theta, alpha, beta and gamma frequency bands. The classification was carried out using filtered data with an epoch length of 2 seconds.

Channels/Frequency Bands		Delta	Theta	Alpha	Beta	Gamma
F4, Fp2, C3, FC6, O1, Oz, FT9, T8	Accuracy	63%	71%	71%	69%	35%
	Sensitivity	12%	0%	0%	2%	96%
	Specificity	84%	99%	100%	96%	11%
F4, Fp2, C3, FC6, O1, FT9, T8	Accuracy	66%	71%	70%	70%	40%
	Sensitivity	11%	0%	1%	2%	96%
	Specificity	88%	100%	98%	98%	18%
F4, Fp2, C3, FC6, FT9, T8	Accuracy	63%	71%	66%	71%	72%
	Sensitivity	14%	3%	8%	5%	3%
	Specificity	83%	99%	90%	97%	99%
F4, Fp2, C3, FC6, T8	Accuracy	59%	67%	69%	70%	60%
	Sensitivity	23%	9%	7%	5%	30%
	Specificity	73%	90%	94%	95%	72%
F4, Fp2, FC6, T8	Accuracy	62%	65%	67%	69%	59%
	Sensitivity	29%	16%	13%	5%	40%
	Specificity	75%	85%	88%	95%	67%
F4, Fp2, T8	Accuracy	62%	63%	60%	68%	52%
	Sensitivity	18%	29%	15%	17%	52%
	Specificity	79%	77%	78%	89%	52%
F4, Fp2	Accuracy	63%	58%	59%	61%	53%
	Sensitivity	27%	35%	33%	22%	46%
	Specificity	77%	67%	70%	76%	55%
Fp2	Accuracy	57%	55%	57%	55%	55%
	Sensitivity	40%	32%	36%	38%	37%
	Specificity	64%	64%	65%	62%	62%

Table 3.2: Channel selection performance for the Random Forest classifier. The table demonstrates the classification accuracy, sensitivity and specificity achieved using different combinations of channels, which were systematically removed based on their discriminative ability. The channels were removed according to the order presented in the list in Section 3.3, with the goal of optimising classification performance across the delta, theta, alpha, beta and gamma frequency bands. The classification was carried out using filtered data with an epoch length of 2 seconds.

topographical maps of two recordings. Additionally, the channel selection method has only been performed for an epoch length of 2 seconds. While the method may only capture a limited effect of removing channels, it serves its purpose as a time-efficient channel selection technique.

3.4 Classification

3.4.1 Choice of Classification Algorithms

The selection process for the classifiers used in this study was based on their intrinsic characteristics, their success in previous EEG-related classification research and their performance in a preliminary evaluation of classifiers in the `scikit-learn` (*scikit-learn* (2007)) module using `lazypredict` (*Lazy Predict* (2019)).

In Section 2.5, it was explained how AdaBoost, Random Forest and EEGNet all have desirable properties that make them suitable for this study. As discussed in Section 2.5, previous research has demonstrated the effectiveness of AdaBoost, Random Forest and EEGNet in various EEG-related classification problems.

In addition to the intrinsic characteristics and previous research, a preliminary evaluation of all classifiers in the `scikit-learn` module was carried out. The classifier `LazyClassifier` from the module `lazypredict.Supervised` was used to quickly assess the performance of a large number of classifiers with minimal effort. This evaluation found that AdaBoost and Random Forest consistently ranked among the top-performing classifiers across different segment lengths and feature extraction methods, further justifying their inclusion. EEGNet was selected independently of the preliminary evaluation due to its specific design for EEG data classification and its success in the literature, even though it was not among the models tested by the `lazyclassifier`

3.4.2 Classification Pipeline

The classification process begins with deciding whether to use raw or filtered data. Both raw data and filtered data were used in this study to explore the potential effects

of filtering on classification performance.

For the AdaBoost and Random Forest classifiers, the feature extraction method described in Section 3.3 was applied to the data to generate a feature vector. For EEGNet, no features were manually extracted as the CNN automatically captures features from the data.

All the recordings in the dataset were trimmed to ensure that their length did not exceed the intended duration of five minutes. Additionally, the recordings were segmented into shorter lengths to increase the sample size, with various non-overlapping epoch lengths explored in this study. The study focused on epoch lengths in the range of 1 to 12 seconds. This range was chosen to strike a balance between capturing sufficient temporal information from the data while maintaining an adequate number of samples for effectively training the model.

The labels for this study were derived from the STAI-Y forms completed by the participants. Given the study's objective of measuring short-term stress, only Form Y-1, which assesses the participant's current state of anxiety, was utilised for calculating each participant's STAI-Y score. This choice is in contrast to Form Y-2, which measures trait anxiety. The STAI-Y score was calculated as a weighted sum of the participants' responses, following the method outlined in Section A.2. Each participant's STAI-Y score falls within the range of 20 to 80.

To enhance the distinction between the classes, recordings with moderate state anxiety levels were excluded from the dataset. According to the paper Spielberger et al. (1983), the creator of the STAI-Y assessment method defines the cutoff range for moderate state anxiety levels to lie between 40 and 60. However, an analysis of the dataset's distribution showed that applying this range resulted in a ratio of 49:2 between non-stressed and stressed samples, which is heavily imbalanced. To address the class imbalance, the cutoff range was adjusted to 37 to 45. The adjusted range provided a ratio of 46:24 between non-stressed and stressed samples.

Initially, the dataset was divided into a training set (80% of the data) and a test set (the remaining 20%). The data is split in a stratified manner to ensure a representative sample while avoiding participant overlap. The 80:20 train-test split serves as a baseline evaluation method for the classifiers. However, to further explore the classifiers'

ability to generalise to unseen data, Leave-One-Subject-Out (LOSO) cross-validation is employed as an alternative evaluation method. In the LOSO approach, each participant's data is used as a test set while the remaining participants' data form the training set. This evaluation method is more akin to the real-world classification scenario, where a new participant is evaluated as being stressed or non-stressed by the existing model. To maintain the independence of the training and test sets, avoiding participant overlap is essential. This reduces the risk of overfitting and helps ensure that the evaluation of the classifiers is unbiased. The function `StratifiedGroupKFold` from the `sklearn.model_selection` package is employed to implement the 80:20 train-test split.

To obtain a more accurate and reliable evaluation of the classifiers' performance, five-fold cross-validation is used for all classifiers. Similar to the train-test split, the `StratifiedGroupKFold` function is employed for implementing the five-fold cross-validation process. This approach helps maintain a balanced representation of the classes and prevents participant overlap across the different folds, consequently contributing to a more accurate and reliable evaluation of the classifiers' performance. The model that achieved the best performance across the five folds was chosen as the best model.

The macro F1-score was used as the main performance measure for the AdaBoost and Random Forest classifiers, while Cohen's Kappa was used for EEGNet. Both of these performance measures were used instead of accuracy, which is generally considered to be unsuitable for imbalanced datasets. As previously described, the class balance after removing moderate samples 46 non-stressed samples to 24 stressed samples. Using accuracy alone as the main performance metric can be misleading in this case since the classifier could achieve an accuracy of $\frac{46}{46+24} = 65.7\%$ simply by predicting the majority class. The macro F1-score, which takes precision and recall into account, offers a more balanced perspective. Similarly, Cohen's Kappa adjusts for the chance agreement between observers, making it a robust measure for this scenario.

After training and validating the classifiers, their performance was assessed using the `sklearn.metrics` module. The `classification_report` function from this module was used to provide a comprehensive summary of the accuracy, precision, re-

call and macro F1-score of the model, based on its performance against the test set. In addition, the `sklearn.metrics.confusion_matrix` function was used to generate the confusion matrix of the model, which is a valuable tool for further analysing the performance of the model. Specifically, the confusion matrix effectively highlights classification errors as well as any potential imbalances in the model's predictions.

The implementation of the AdaBoost, Random Forest and EEGNet classifiers, as well as other related code, can be found in the GitHub repository referenced as Sletten (2023).

The following subsections will discuss the design and implementation of the classifiers used in this study.

3.4.3 AdaBoost

The AdaBoost classifier was implemented using the `sklearn.ensemble.AdaBoostClassifier` class. Before applying the classifier to the dataset, the data was preprocessed to ensure that the features have a similar scale. Scaling is an important part of the classification process as it prevents large features from dominating the model. The dataset was scaled using the `sklearn.preprocessing.MinMaxScaler` class.

In order to find the best hyperparameters for the AdaBoost classifier, a grid search with cross-validation was employed. Grid search is an exhaustive search technique that tests all the possible combinations of the specified hyperparameter values and identifies the combination that achieves the best performance based on an evaluation metric. Cross-validation is integrated into the grid search process to ensure a more accurate and reliable evaluation of the model performance. As previously stated, the macro F1-score is the chosen evaluation metric when validating the model.

The `sklearn.model_selection.GridSearchCV` function was used to perform the grid search and cross-validation. The grid search was conducted over a grid of possible values for the number of estimators, the learning rate and the choice of algorithm. The number of estimators refers to the number of base estimators in the ensemble, the learning rate is the weight applied to each base estimator and the choice of algorithm determines which boosting algorithm to use. In this study, the range of

values for the number of estimators was [50, 100, 200], the learning rate range was [0.1, 0.5, 1] and the algorithm choices were `SAMME` and `SAMME.R`. After selecting the optimal hyperparameters, the model was fitted with the training data and the chosen hyperparameters.

The implementation of the AdaBoost classifier can be found in the GitHub repository referenced as Sletten (2023).

3.4.4 Random Forest

The Random Forest classifier was implemented using the `sklearn.ensemble.RandomForestClassifier` class. Before applying the classifier to the dataset, the data was preprocessed to ensure that the features have a similar scale. Scaling is an important part of the classification process as it prevents large features from dominating the model. The dataset was scaled using the `sklearn.preprocessing.RobustScaler` class.

In order to find the best hyperparameters for the Random Forest classifier, a grid search with cross-validation was employed. Grid search is an exhaustive search technique that tests all the possible combinations of the specified hyperparameter values and identifies the combination that achieves the best performance based on an evaluation metric. Cross-validation is integrated into the grid search process to ensure a more accurate and reliable evaluation of the model performance. As previously stated, the macro F1-score is the chosen evaluation metric when validating the model.

Class weighting was used to combat the class imbalance of the dataset. By weighting samples corresponding to the underrepresented class more heavily during training, the model is encouraged to give equal importance to both classes. Class weighting was implemented using the function `compute_class_weight` from the package `sklearn.utils.class_weight`.

The `sklearn.model_selection.GridSearchCV` function was used to perform the grid search and cross-validation. The grid search was conducted over a grid of possible values for the number of estimators and the max depth. The number of estimators refers to the number of base estimators in the ensemble and the max depth

refers to the maximum depth of the trees in the random forest. In this study, the range of values for the number of estimators was [100, 200, 300] and the max depth range was [5, 10, 20, None]. After selecting the optimal hyperparameters, the model was fitted with the training data and the chosen hyperparameters.

The implementation of the Random Forest classifier can be found in the GitHub repository referenced as Sletten (2023).

3.4.5 EEGNet

The EEGNet model was implemented with the Keras (*Keras* (2015)) and Tensorflow (*Tensorflow* (2015)) packages using a `keras.Sequential` object. The architecture specified in Table 2.1 was added as layers to the `keras.-Sequential` object. The model was fit using the Adam optimiser, minimising the binary cross-entropy loss function. This choice of optimiser and loss function aligns with the original creators' recommendations in Lawhern et al. (2018), with the exception that the binary cross-entropy loss function is employed instead of the categorical cross-entropy loss function, given the binary classification problem addressed in this project. The learning rate of the optimiser was chosen by performing a grid search, with the learning rate range [0.001, 0.0001, 0.0001]. The default batch size of 32 is maintained while the number of epochs is set to 100. Similar to the implementation of Random Forest, class weighting was used to combat class imbalance for the EEGNet classifier. Class weighting was implemented using the function `compute_class_weight` from the package `sklearn.utils.class_weight`.

The original EEGNet model was specifically designed for EEG data sampled at 128 Hz. For this study, however, we've adapted the model to accommodate our data's sampling rate of 250 Hz, following the method recommended in the EEGNet GitHub repository Lawhern (2018). This involved scaling the lengths of the temporal kernels and the average pooling size in blocks 1 and 2 proportionately to the increase in sampling rate. The model's original creator cautions that the impacts of sampling rates other than 128 Hz haven't been fully explored. The specific implementation details, including the scaled kernel sizes and average pooling blocks, are outlined in Table 2.1.

To improve the speed at which the model is trained and mitigate overfitting, early stopping has been employed. Early stopping stops the training of the model when the validation loss does not improve over a predefined number of epochs, a number referred to as *patience*. In this project, patience has been set to 5. This value was chosen to balance training time and model performance. This approach means that if the validation loss fails to improve for 5 consecutive epochs, the training stops. This prevents the model from learning noise and specific details of the training data, mitigating overfitting.

Raw data was used for classification, as opposed to filtered data. The model's convolutional layers inherently perform bandpass filtering, reducing the need for preprocessed data. However, scaling of the dataset was performed to prevent large features from dominating the model. The dataset was scaled using the `sklearn.preprocessing.RobustScaler` class.

The implementation of the EEGNet model can be found in the GitHub repository referenced as Sletten (2023).

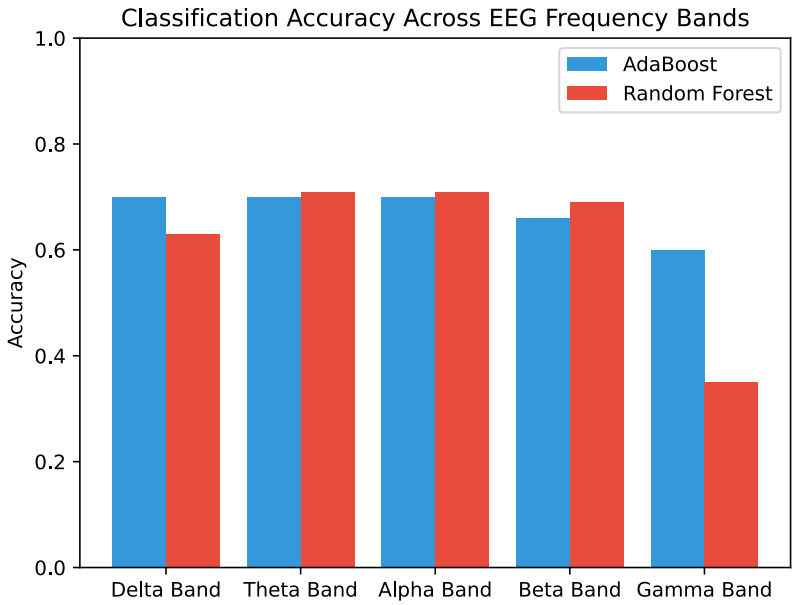


Figure 3.7: Comparative performance of AdaBoost and Random Forest classifiers across distinct frequency bands. The accuracy of the classifiers was evaluated using the delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-70 Hz) power frequency bands. The AdaBoost classifier accuracies are shown in blue, while the Random Forest classifier accuracies are illustrated in red. The classification was carried out using filtered data with an epoch length of 2 seconds.

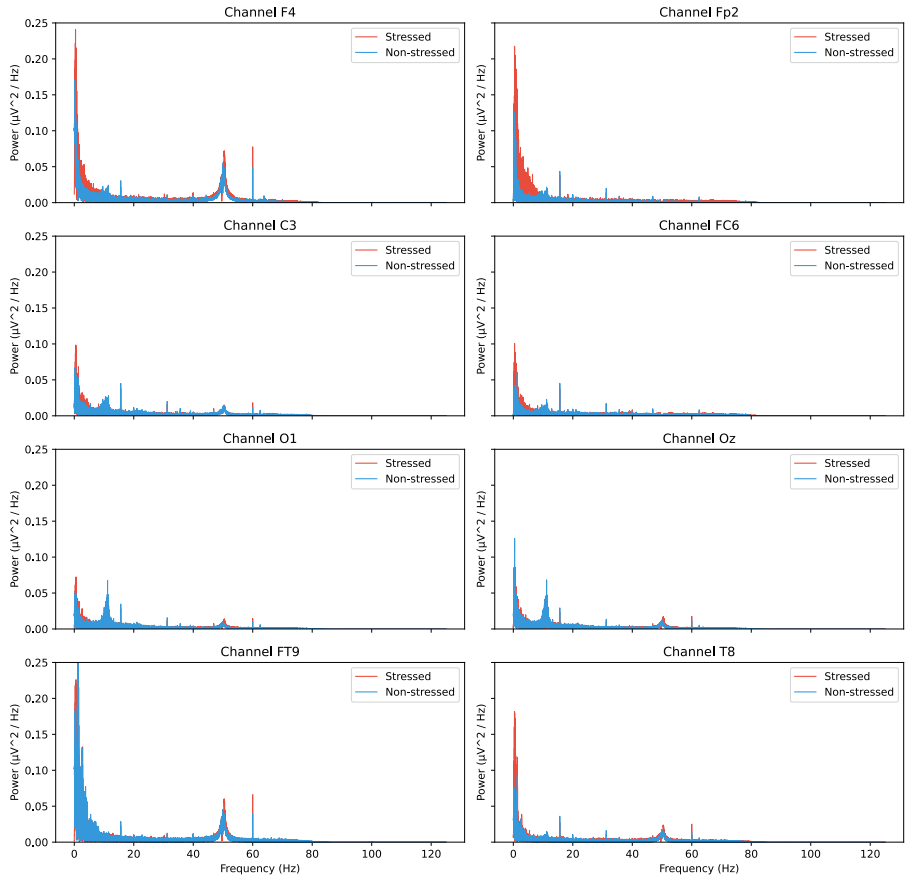


Figure 3.8: Comparison of PSDs for non-stressed (blue) and stressed (red) EEG signals across the 8 channels (F4, Fp2, C3, FC6, O1, Oz, FT9 and T8). Both the stressed and non-stressed recording belongs to participant P024 during session 2, however, the non-stressed recording is from run 1 while the stressed recording is from run 2. The PSDs were computed using filtered data with an epoch length of 2 seconds.

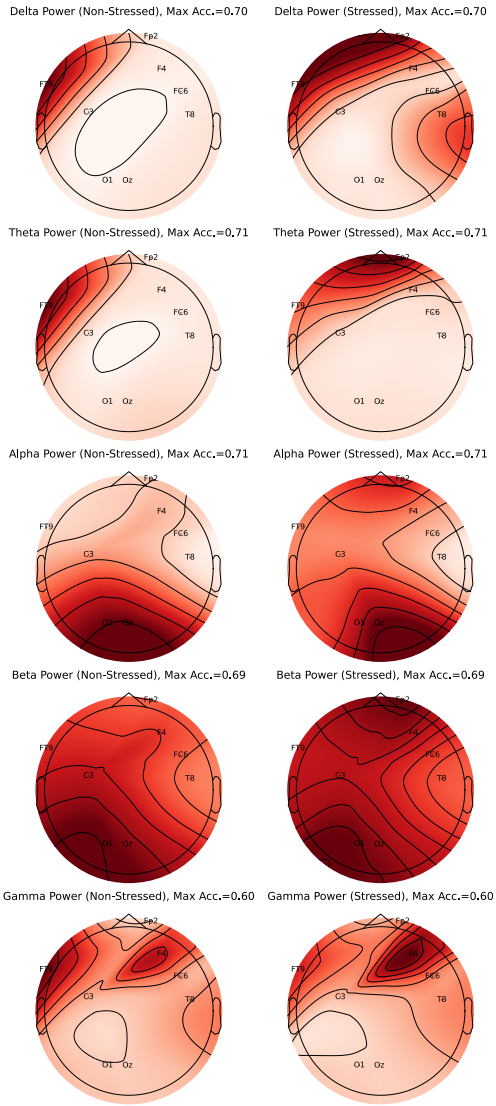


Figure 3.9: The figure illustrates the topographic distribution of EEG power across five distinct frequency bands (delta, theta, alpha, beta and gamma) for two subject conditions (Non-Stressed and Stressed). Each row represents a separate frequency band, while each column corresponds to a different subject condition. In the topographical maps, the colour-coded activity levels at various electrode positions can be observed, with darker red hues indicating higher activity. The electrode positions F4, Fp2, C3, FC6, O1, Oz, FT9 and T8 are explicitly marked on the topographical maps for reference. The maximum classification accuracy attained using the power features from each frequency band is displayed in the title of each subplot. The PSDs were computed using filtered data with an epoch length of 2 seconds.

Chapter 4

Results

In the previous chapter, the methods used to gather and analyse the EEG data were presented. In this chapter, the results of the study will be presented. Section 4.1 explores the epoch length of the EEG data's impact on the classification performance. In Section 4.2 and Section 4.3, the obtained results for all classifiers using an 80:20 train-test split and LOSO cross-validation, respectively, will be presented. Lastly, Section 4.4, will present the best-performing hyperparameters for the AdaBoost, Random Forest and EEGNet classifiers.

The metrics used to evaluate the classification performance include accuracy, macro F1-score, sensitivity, specificity and Cohen's Kappa, and their respective equations are shown in Equation 4.1, Equation 4.2, Equation 4.4, Equation 4.5 and Equation 4.6.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of predictions}} \tag{4.1}$$

$$\text{Macro F1-score} = \frac{1}{2}(\text{F1-score}_{\text{class1}} + \text{F1-score}_{\text{class2}}) \tag{4.2}$$

$$\text{F1-score} = 2 \cdot \frac{\frac{\text{True positives}}{\text{True positives + False positives}} \cdot \frac{\text{True positives}}{\text{True positives + False negatives}}}{\frac{\text{True positives}}{\text{True positives + False positives}} + \frac{\text{True positives}}{\text{True positives + False negatives}}} \tag{4.3}$$

$$\text{Sensitivity} = \frac{\text{True positives}}{\text{True positives + False negatives}} \tag{4.4}$$

$$\text{Specificity} = \frac{\text{True negatives}}{\text{True negatives + False positives}} \tag{4.5}$$

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{4.6}$$

Note: Portions of this section are borrowed or adapted from Sletten (2022).

4.1 Epoch Length Analysis

This section presents the performance metrics accuracy, sensitivity, and specificity for the AdaBoost, Random Forest, and EEGNet classifiers across varying epoch lengths, ranging from 1 to 12 seconds. The results are shown in Figure 4.1, Figure 4.2 and Figure 4.3, respectively. The choice of epoch length is important as the EEG data needs to facilitate good classification accuracy while maintaining a balance between sensitivity and specificity.

The performance of the AdaBoost classifier was evaluated using the alpha frequency band computed from the filtered data and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. This specific combination of frequency band and EEG channels was established as the most effective for an epoch length of 2 seconds for the AdaBoost classifier, as detailed in Section 3.3.

When examining Figure 4.1, it is clear that as the epoch length increases from 1 to 8 seconds, the specificity declines and the sensitivity improves. The balance between sensitivity and specificity is significantly better for an epoch range of 8 seconds compared to a 1-second epoch length. However, there appears to be a trade-off, as longer epoch lengths are associated with a downward trend in accuracy.

Despite the overall decrease in accuracy with increasing epoch lengths, the impact is not drastic. For instance, with an epoch length of 8 seconds, an accuracy of 63% is achieved, which is only 7 percentage points lower than the peak accuracy of 70%

observed at a 2 and 3-second epoch length. Based on this reasoning, an 8-second epoch length appears to be the best choice to obtain a balance between sensitivity and specificity while maintaining a satisfactory accuracy for the AdaBoost classifier.

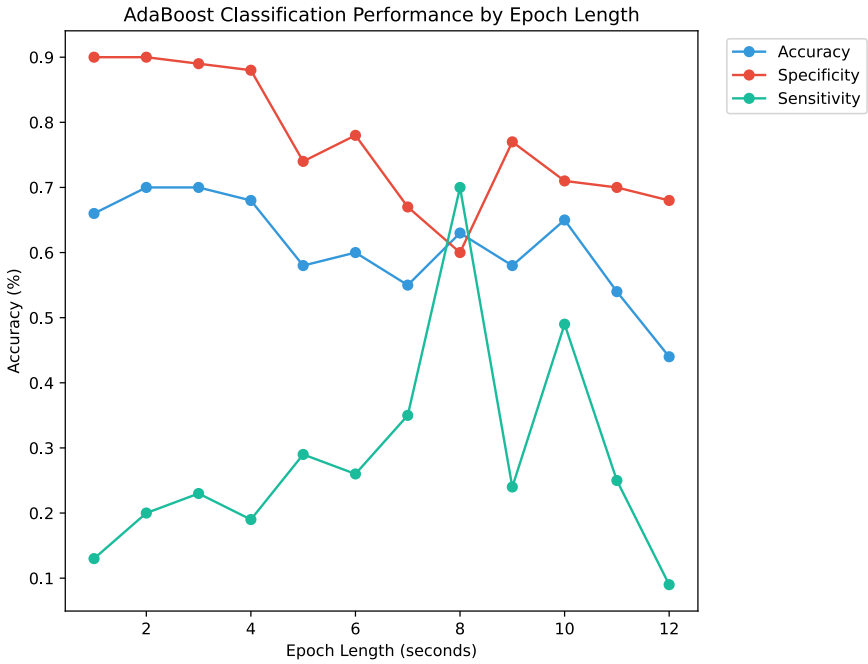


Figure 4.1: The figure illustrates the variation in the performance metrics accuracy, specificity and sensitivity of the AdaBoost classifier on all filtered data, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8, across epoch lengths ranging from 1 to 12 seconds. The plot shows the impact of varying epoch lengths on the classifier's performance, particularly highlighting the trade-off between accuracy, specificity and sensitivity.

For the Random Forest classifier, performance evaluation was carried out using the beta frequency band computed from the filtered data and the EEG channels F4, Fp2 and T8. This specific combination of frequency band and EEG channels was established as the most effective for an epoch length of 2 seconds for the Random Forest classifier,

as detailed in Section 3.3.

A close examination of Figure 4.2 reveals that the accuracy, sensitivity and specificity of the Random Forest classifier remain relatively stable for epoch lengths in the range from 1 to 12 seconds. However, the ratio between the specificity and sensitivity is unbalanced for the majority of the epoch length, with an average specificity of 94% and an average sensitivity of 7%. The performance for the 2-second epoch length appears to be the choice of epoch length that best balances specificity and sensitivity while maintaining a satisfactory accuracy. The accuracy for 2-second long epochs is 68%, which is only 2 percentage points lower than the best-achieved accuracy, which was 70% using 10-second long epochs.

The EEGNet classifier's performance was evaluated using raw data and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

A close examination of Figure 4.3 reveals a noteworthy trend among the three performance metrics: accuracy, specificity, and sensitivity. The accuracy and specificity seem to have a positive correlation, meaning they simultaneously increase or decrease. However, sensitivity appears to be inversely related, decreasing when accuracy and specificity increase, and vice versa.

A prime example is the performance with 6-second epochs, which shows a spike in accuracy and specificity but a corresponding drop in sensitivity. Given the aim to maintain satisfactory accuracy and a balanced sensitivity and specificity, 3-second epochs appear to be the best choice. This length yields an accuracy of 71%, only marginally lower than the highest achieved accuracy of 74% using 6, 10 and 11-second long epochs. Additionally, this choice of epoch length appears to produce the smallest difference between sensitivity and specificity.

4.2 80:20 Train-Test Split

This section presents the results of the AdaBoost, Random Forest, and EEGNet classifiers in terms of accuracy, specificity, sensitivity, and confusion matrices when using an 80:20 train-test split. The performance of the classifiers was evaluated using the epoch lengths determined to yield the best performance, as discussed in Section 4.1.

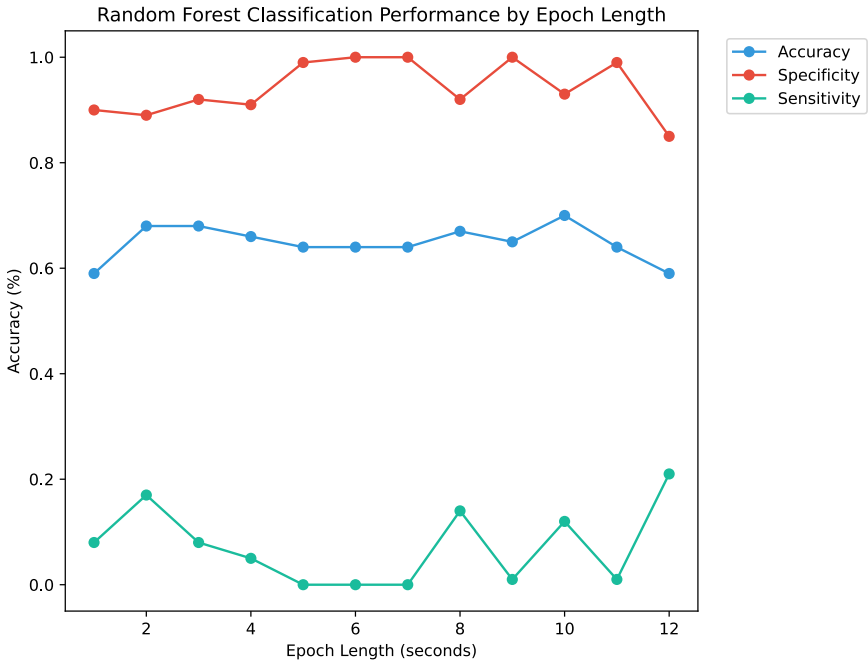


Figure 4.2: The figure illustrates the variation in the performance metrics accuracy, specificity and sensitivity of the Random Forest classifier on all filtered data, computed from the beta frequency band and EEG channels F4, Fp2, and T8, across epoch lengths ranging from 1 to 12 seconds. The plot shows the impact of varying epoch lengths on the classifier's performance, particularly highlighting the trade-off between accuracy, specificity and sensitivity.

Table 4.1 provides a comparison of classification accuracy, specificity and sensitivity for the classifiers. The AdaBoost classifier demonstrated an accuracy of 63%, with a sensitivity of 67% and a specificity of 61%. Random Forest performed moderately better in terms of accuracy and specificity, achieving values of 68% and 89% respectively, but fell behind AdaBoost in sensitivity with a lower score of 17%. The EEGNet classifier surpassed both AdaBoost and Random Forest in accuracy, achieving a score of 71%, and exhibited high specificity at 88%. However, it showed a sensitivity rate of

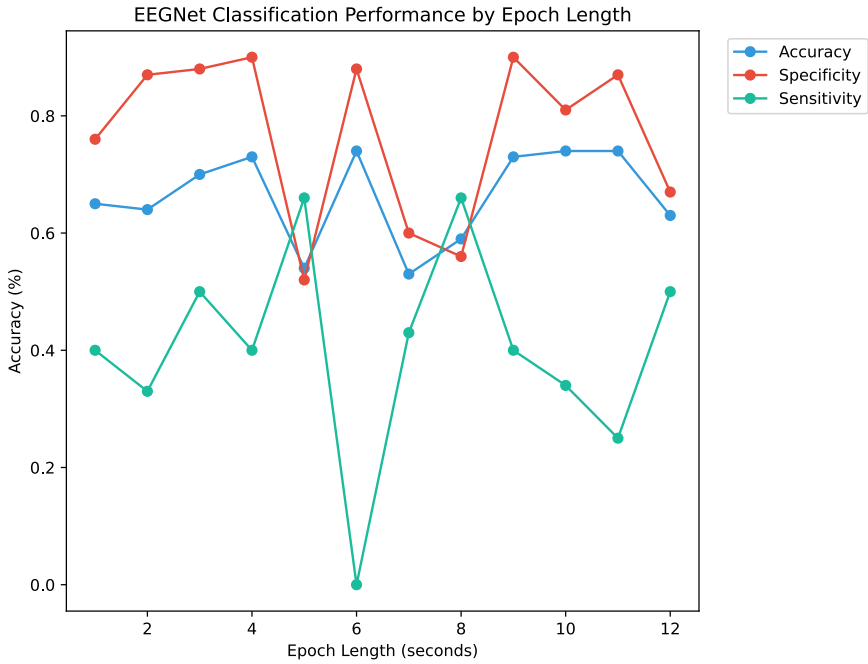


Figure 4.3: The figure illustrates the variation in the performance metrics accuracy, specificity and sensitivity of the EEGNet classifier on all raw data, using the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8, across epoch lengths ranging from 1 to 12 seconds. The plot shows the impact of varying epoch lengths on the classifier's performance, particularly highlighting the trade-off between accuracy, specificity and sensitivity.

50%, which was considerably better than Random Forest but still behind AdaBoost.

Table 4.2 displays the confusion matrices for the classifiers. The confusion matrix for the AdaBoost classifier shows 203 true negatives, 132 false positives, 49 false negatives, and 99 true positives. The Random Forest classifier achieved 1332 true negatives, 173 false positives, 498 false negatives, and 103 true positives. The EEGNet classifier achieved 700 true negatives, 100 false positives, 301 false negatives, and 300 true positives.

Classifier	Accuracy	Sensitivity	Specificity
AdaBoost	63%	67%	61%
Random Forest	68%	17%	89%
EEGNet	71%	50%	88%

Table 4.1: The table shows the comparison of classification accuracy, sensitivity and specificity for AdaBoost, Random Forest, and EEGNet classifiers using an 80:20 train-test split. The AdaBoost classifier was evaluated on all filtered data using 8-second epochs, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. The Random Forest classifier was evaluated on all filtered data using 2-second epochs, computed from the beta frequency band and EEG channels F4, Fp2 and T8. The EEGNet classifier was evaluated on all filtered data using 3-second epochs and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

Classifier	Confusion Matrix
AdaBoost	$\begin{bmatrix} 203 & 132 \\ 49 & 99 \end{bmatrix}$
Random Forest	$\begin{bmatrix} 1332 & 173 \\ 498 & 103 \end{bmatrix}$
EEGNet	$\begin{bmatrix} 700 & 100 \\ 301 & 300 \end{bmatrix}$

Table 4.2: The table shows the comparison of confusion matrices for AdaBoost, Random Forest, and EEGNet classifiers using an 80:20 train-test split. Each matrix contains four entries: true negatives (top left), false positives (top right), false negatives (bottom left), and true positives (bottom right). The AdaBoost classifier was evaluated on all filtered data using 8-second epochs, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. The Random Forest classifier was evaluated on all filtered data using 2-second epochs, computed from the beta frequency band and EEG channels F4, Fp2 and T8. The EEGNet classifier was evaluated on all filtered data using 3-second epochs and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

These results demonstrate the performance of the AdaBoost, Random Forest, and EEGNet classifiers using an 80:20 train-test split. The EEGNet classifier achieved the

highest accuracy and demonstrated improved performance compared to the other classifiers. The confusion matrices provide additional insights into the classification performance of each classifier by showing the distribution of true and false predictions.

4.3 Leave-One-Subject-Out Cross-Validation

In this section, we present the results obtained from applying the LOSO cross-validation method to our three chosen classifiers: AdaBoost, Random Forest, and EEGNet. This method provides a robust evaluation of the classifiers' performance as it assesses their ability to generalise to unseen data, simulating a realistic application of these algorithms.

As seen in Table 4.3, the average accuracy across all participants for each classifier was similar, approximately 54%. An accuracy of 54% is only slightly above chance, meaning that in general, the classifiers are struggling to generalise to unseen data. However, substantial variations were observed when considering individual participant data. The AdaBoost classifier's accuracy varied from 34% (participant P007) to 84% (participant P019). Similar variations were observed for the Random Forest classifier (32% to 84%) and the EEGNet classifier (24% to 75%). The inter-subject variation in accuracy could have multiple causes, including differences in the quality of the collected data and imprecise labelling of the recordings.

Participant P014 displayed notable variation across classifiers with an accuracy of 77% for the AdaBoost classifier, declining to 43% and 33% for the Random Forest and EEGNet classifiers, respectively. Furthermore, some participants, such as P007 and P019, showed consistent accuracies across all classifiers. This pattern suggests that certain participant data might have characteristic features that are consistently interpreted by different classifiers.

Importantly, there were several participants (P001-P003, P005-P006, P010-P012, P015, P018, P020, P022-P023, P025-P026) whose data did not include both non-stressed and stressed samples, thus no meaningful accuracy could be computed.

Evaluating the performance of classifiers with LOSO cross-validation becomes

more insightful when confusion matrices are employed, as they reveal potential imbalances in predictions.

Table 4.4 and Table 4.5 display the confusion matrices for three classifiers - AdaBoost, Random Forest, and EEGNet. In each matrix, four components are presented: the true negatives (top left), false positives (top right), false negatives (bottom left), and true positives (bottom right). If a participant's data lacks samples from both stressed and non-stressed classes, a '-' symbol is displayed, indicating that generating a confusion matrix is not feasible.

Upon close inspection of Table 4.4 and Table 4.5, a concerning pattern emerges for the EEGNet classifier. For a majority of participants, 10 out of the 14 participants who have samples from both classes, the EEGNet classifier predicts only one class. This result casts doubt on the LOSO accuracies for EEGNet discussed earlier in this subsection, because a classifier that predominantly predicts one class, irrespective of the actual class distribution, is not demonstrating good classification performance. The imbalance in predictions suggests that the model might be overfitting to the majority class in the training data, thereby failing to generalise well to unseen participants. This suggests that EEGNet might not be suitable for this type of classification task with this dataset.

The AdaBoost and Random Forest classifiers, in contrast, do not exhibit this behaviour. They predict more than one class for all participants. However, a pattern of a larger false negative rate over the false positive rate is observed for both classifiers. This imbalance is more pronounced with the Random Forest classifier. Ideally, a stress detection system should balance false positive and negative rates. In this context, a larger false positive rate is preferable because incorrectly classifying a non-stressed subject as stressed is less problematic than misclassifying a stressed subject. The observed higher false negative rate might be attributed to class imbalance.

4.4 Hyperparameter Tuning

This section showcases the optimal hyperparameter combinations for the AdaBoost, Random Forest, and EEGNet classifiers, as determined by the hyperparameter tuning

approach outlined in Section 3.4. The classifiers were tuned using the same combination of features and channels mentioned in earlier sections, paired with the epoch lengths that were proven to provide the best performance in Section 4.1.

For the AdaBoost classifier, the `SAMME.R` algorithm was found to perform slightly better than `SAMME`. Additionally, higher learning rates generally resulted in higher F1-scores, while the number of estimators did not significantly affect the performance. This is evident in Figure 4.4, where the heatmaps represent macro F1-scores for different combinations of the number of estimators (`n_estimators`) and learning rate (`learning_rate`). The highest macro F1-score was achieved using the `SAMME.R` algorithm, a learning rate of 1, and 100 estimators.

The performance of the Random Forest classifier was noticeably better with increased maximum tree depth, as shown in Figure 4.5. The number of estimators (`n_estimators`) did not significantly impact the results, as a similar performance was observed across various numbers of estimators. The highest macro F1-score was achieved using a max depth of None and 200 estimators.

For the EEGNet classifier, as shown in Figure 4.6, a lower learning rate resulted in better performance. The heatmap in the figure presents Kappa scores for different learning rates, with the highest Kappa score achieved at a learning rate of 1e-3. The Kappa score is a statistical measure that ranges from -1 to 1. A score of 1 implies perfect agreement between the model's predictions and the actual labels, while a score of -1 implies complete disagreement. A score close to zero would imply a level of agreement equivalent to random chance. Therefore, the highest Kappa score of 0.149 achieved by our classifier, as seen in Figure 4.6, suggests that there is some level of agreement between the predictions and the labels.

While the tuning results presented here suggest optimal hyperparameters for our specific experiment setup, these results represent only a single combination of data, features, channels and epoch length for each classifier. Other combinations might yield different results. A more exhaustive evaluation across various combinations of data, features, channels, and epoch lengths could potentially lead to different optimal hyperparameters.

Participant	AdaBoost	Random Forest	EEGNet
P001	-	-	-
P002	-	-	-
P003	-	-	-
P004	39%	55%	50%
P005	-	-	-
P006	-	-	-
P007	34%	32%	24%
P008	48%	58%	33%
P009	53%	70%	67%
P010	-	-	-
P011	-	-	-
P012	-	-	-
P013	61%	66%	67%
P014	77%	43%	33%
P015	-	-	-
P016	45%	39%	67%
P017	40%	32%	74%
P018	-	-	-
P019	84%	66%	75%
P020	-	-	-
P021	52%	51%	50%
P022	-	-	-
P023	-	-	-
P024	51%	61%	67%
P025	-	-	-
P026	-	-	-
P027	55%	57%	50%
P028	59%	54%	50%
AVERAGE	54%	53%	54%

Table 4.3: The table shows the comparison of classification accuracy for AdaBoost, Random Forest, and EEGNet classifiers using the LOSO cross-validation method for generating holdout sets. The entries '-' indicate that the corresponding participant's data did not include both non-stressed and stressed samples, meaning that no meaningful accuracy could be computed. The final row of the figure highlights the average classification accuracy across all participants for each classifier. The AdaBoost classifier was evaluated on all filtered data using 8-second epochs, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. The Random Forest classifier was evaluated on all filtered data using 2-second epochs, computed from the beta frequency band and EEG channels F4, Fp2 and T8. The EEGNet classifier was evaluated on all filtered data using 3-second epochs and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

Participant	AdaBoost	Random Forest	EEGNet
P001	-	-	-
P002	-	-	-
P003	-	-	-
P004	$\begin{bmatrix} 22 & 15 \\ 31 & 7 \end{bmatrix}$	$\begin{bmatrix} 128 & 23 \\ 112 & 40 \end{bmatrix}$	$\begin{bmatrix} 0 & 100 \\ 0 & 101 \end{bmatrix}$
P005	-	-	-
P006	-	-	-
P007	$\begin{bmatrix} 34 & 3 \\ 98 & 17 \end{bmatrix}$	$\begin{bmatrix} 142 & 8 \\ 412 & 54 \end{bmatrix}$	$\begin{bmatrix} 100 & 0 \\ 310 & 0 \end{bmatrix}$
P008	$\begin{bmatrix} 45 & 29 \\ 29 & 8 \end{bmatrix}$	$\begin{bmatrix} 215 & 85 \\ 103 & 47 \end{bmatrix}$	$\begin{bmatrix} 100 & 100 \\ 100 & 0 \end{bmatrix}$
P009	$\begin{bmatrix} 55 & 19 \\ 33 & 4 \end{bmatrix}$	$\begin{bmatrix} 263 & 38 \\ 96 & 55 \end{bmatrix}$	$\begin{bmatrix} 200 & 0 \\ 100 & 0 \end{bmatrix}$
P010	-	-	-
P011	-	-	-
P012	-	-	-
P013	$\begin{bmatrix} 63 & 11 \\ 32 & 5 \end{bmatrix}$	$\begin{bmatrix} 246 & 53 \\ 100 & 49 \end{bmatrix}$	$\begin{bmatrix} 199 & 0 \\ 99 & 0 \end{bmatrix}$
P014	$\begin{bmatrix} 32 & 5 \\ 23 & 51 \end{bmatrix}$	$\begin{bmatrix} 121 & 29 \\ 227 & 73 \end{bmatrix}$	$\begin{bmatrix} 100 & 0 \\ 200 & 0 \end{bmatrix}$

Table 4.4: This table presents the first part of the comparison of the confusion matrices for the AdaBoost, Random Forest, and EEGNet classifiers, obtained via the LOSO cross-validation method. The data for the remaining participants can be found in Table 4.5. Each matrix contains four entries: true negatives (top left), false positives (top right), false negatives (bottom left), and true positives (bottom right). The entries '-' indicate that the corresponding participant's data did not include both non-stressed and stressed samples, meaning that no confusion matrix could be calculated. The AdaBoost classifier was evaluated on all filtered data using 8-second epochs, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. The Random Forest classifier was evaluated on all filtered data using 2-second epochs, computed from the beta frequency band and EEG channels F4, Fp2 and T8. The EEGNet classifier was evaluated on all filtered data using 3-second epochs and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

Participant	AdaBoost	Random Forest	EEGNet
P015	-	-	-
P016	$\begin{bmatrix} 31 & 6 \\ 55 & 19 \end{bmatrix}$	$\begin{bmatrix} 122 & 28 \\ 245 & 55 \end{bmatrix}$	$\begin{bmatrix} 0 & 100 \\ 0 & 200 \end{bmatrix}$
P017	$\begin{bmatrix} 32 & 2 \\ 87 & 24 \end{bmatrix}$	$\begin{bmatrix} 138 & 13 \\ 394 & 57 \end{bmatrix}$	$\begin{bmatrix} 0 & 100 \\ 4 & 297 \end{bmatrix}$
P018	-	-	-
P019	$\begin{bmatrix} 105 & 6 \\ 18 & 19 \end{bmatrix}$	$\begin{bmatrix} 372 & 78 \\ 126 & 24 \end{bmatrix}$	$\begin{bmatrix} 300 & 0 \\ 100 & 0 \end{bmatrix}$
P020	-	-	-
P021	$\begin{bmatrix} 70 & 4 \\ 67 & 7 \end{bmatrix}$	$\begin{bmatrix} 269 & 33 \\ 263 & 38 \end{bmatrix}$	$\begin{bmatrix} 0 & 201 \\ 0 & 200 \end{bmatrix}$
P022	-	-	-
P023	-	-	-
P024	$\begin{bmatrix} 55 & 19 \\ 35 & 2 \end{bmatrix}$	$\begin{bmatrix} 263 & 37 \\ 137 & 13 \end{bmatrix}$	$\begin{bmatrix} 100 & 100 \\ 0 & 100 \end{bmatrix}$
P025	-	-	-
P026	-	-	-
P027	$\begin{bmatrix} 69 & 5 \\ 62 & 12 \end{bmatrix}$	$\begin{bmatrix} 286 & 14 \\ 246 & 55 \end{bmatrix}$	$\begin{bmatrix} 0 & 200 \\ 0 & 200 \end{bmatrix}$
P028	$\begin{bmatrix} 35 & 2 \\ 28 & 9 \end{bmatrix}$	$\begin{bmatrix} 121 & 29 \\ 109 & 41 \end{bmatrix}$	$\begin{bmatrix} 0 & 100 \\ 0 & 100 \end{bmatrix}$

Table 4.5: This table is a continuation from Table 4.4, displaying the confusion matrices of the final 14 participants for AdaBoost, Random Forest, and EEGNet classifiers. Each matrix, generated via the LOSO cross-validation method, contains four entries: true negatives (top left), false positives (top right), false negatives (bottom left), and true positives (bottom right). The entries '-' indicate that the corresponding participant's data did not include both non-stressed and stressed samples, meaning that no confusion matrix could be calculated. The AdaBoost classifier was evaluated on all filtered data using 8-second epochs, computed from the alpha frequency band and EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8. The Random Forest classifier was evaluated on all filtered data using 2-second epochs, computed from the beta frequency band and EEG channels F4, Fp2 and T8. The EEGNet classifier was evaluated on all filtered data using 3-second epochs and the EEG channels F4, Fp2, C3, FC6, O1, Oz, FT9 and T8.

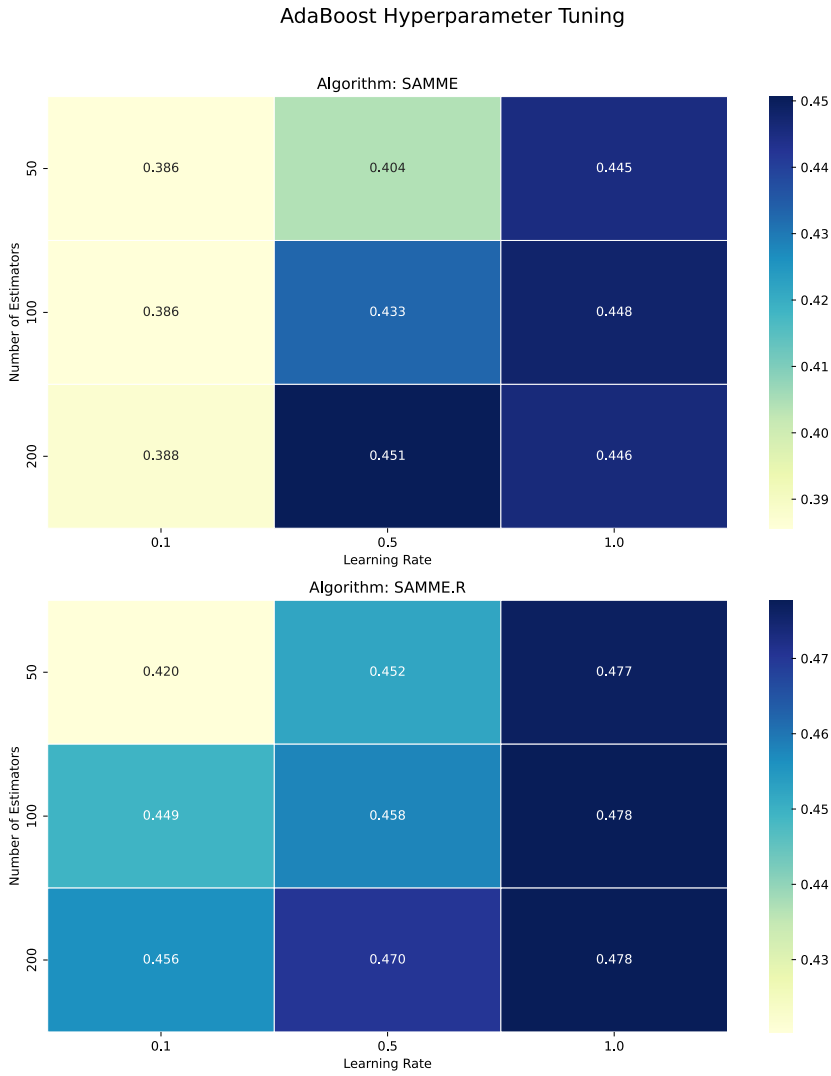


Figure 4.4: The figure presents the results of hyperparameter tuning for the AdaBoost classifier, with each heatmap reflecting macro F1-scores for combinations of the number of estimators (`n_estimators`) and learning rate (`learning_rate`) hyperparameters. Each subplot corresponds to a different algorithm (`SAMME` or `SAMME.R`). The colour gradient in the heatmap represents the highest F1-score obtained in the cross-validation process, with darker colours indicating higher scores. The best macro F1-score was achieved using the `SAMME.R` algorithm, using a learning rate of 1 and 100 estimators. These scores were obtained from the filtered data using 8-second epochs.

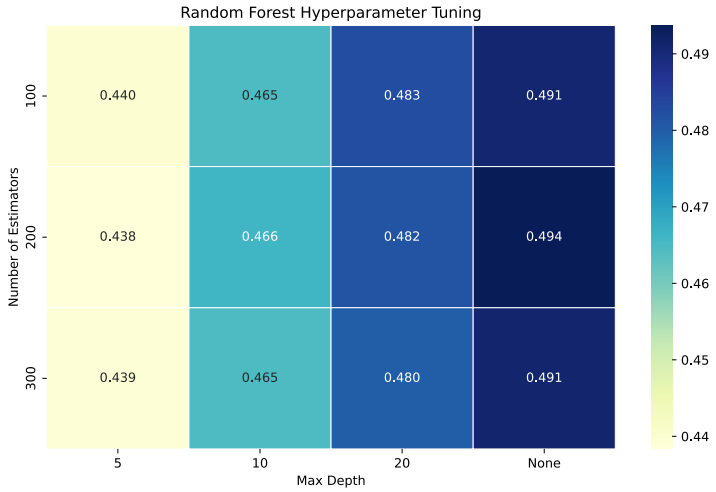


Figure 4.5: The figure presents the results of hyperparameter tuning for the Random Forest classifier, with a heatmap reflecting macro F1-scores for combinations of the number of estimators (`n_estimators`) and maximum depth (`max_depth`) hyperparameters. The colour gradient in the heatmap represents the highest F1-score obtained in the cross-validation process, with darker colours indicating higher scores. The best macro F1-score was achieved using a max depth of None and 200 estimators. These scores were obtained from the filtered data using 2-second epochs.

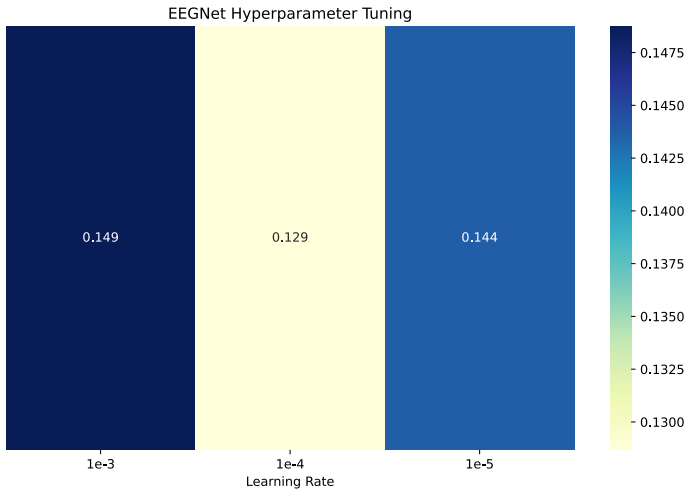


Figure 4.6: The figure presents the results of hyperparameter tuning for the EEGNet classifier, with a heatmap reflecting Kappa scores for different learning rates. The colour gradient in the heatmap represents the highest Kappa score obtained in the cross-validation process, with darker colours indicating higher scores. The best Kappa score was achieved using a learning rate of 1e-3. These scores were obtained from the raw data using 3-second epochs.

Chapter 5

Discussion

5.1 Frequency Band and Channel Selection Evaluation

Section 3.3 presented a performance evaluation of the delta, theta, alpha, beta and gamma frequency bands. The findings, shown in Figure 3.7, revealed that delta, theta, alpha, and beta bands demonstrated high accuracy in classifying stress. In contrast, the gamma band showed significantly lower accuracy, suggesting it might not be as effective for EEG-based stress detection. These results align with the theory and literature referenced in Section 2.2 and the studies reviewed in Katmah et al. (2021).

However, an interesting discrepancy surfaced with the theta frequency band. Despite being presented as the least effective in Section 2.2, the findings in this study place it among the highest performers. This difference could be attributed to the variability in the accuracy of the reviewed studies as indicated by Katmah et al. (2021), suggesting that the effectiveness of the theta band may be more context-dependent than previously assumed.

In Section 3.3 an evaluation of the performance of the delta, theta, alpha, beta and gamma frequency bands was presented. Figure 3.7 presented the accuracy of the

AdaBoost and Random Forest classifiers across the chosen frequency bands using all the available EEG channels. The figure showed that the delta, theta, alpha and beta frequency bands obtained the highest accuracy, while a significantly lower accuracy was obtained using the gamma frequency band.

Section 3.3 also presented a comparison of topographic maps of non-stressed and stressed recordings for the delta, theta, alpha, beta and gamma frequency bands, shown in Figure 3.9. It was determined that the Fp2, F4, T8, and FC6 channels, situated on the right side of the scalp within the frontal, central, and temporal regions, showed the largest discriminative ability of the available channels. For reference, the electrode positions are shown in Figure 3.2. These findings suggest that the changes in brain activity during periods of mental stress are primarily located in these regions.

Given these observations, it is advisable for future EEG-based stress detection studies to focus on increasing the number of electrodes in these areas. This could potentially amplify the discriminative effect and enhance the accuracy of stress detection.

However, these results should be considered with caution. The analysis using topographic maps only compares a single non-stressed and stressed recording. The findings for these recordings might not be applicable to other recordings. Therefore, future research should compare topographic maps of the remaining pairs of recordings to gain a more comprehensive understanding of the discriminative effect of the individual channels.

The final part of Section 3.3 evaluated the effect of reducing less significant channels on the performance of classifiers. The results indicated that the AdaBoost classifier did not improve upon removing these channels. Instead, as channels were reduced, sensitivity dropped to 0%, and specificity rose to 100%. Essentially, this pattern indicates that the model defaulted to predicting only one class, with accuracy reflecting the underlying class distribution. One plausible explanation for this behaviour is that the already limited dataset was further compressed, leading to a lack of diversity in the feature set that hindered the model's learning ability.

On the other hand, the Random Forest classifier demonstrated a different trend. The specificity generally improved with fewer channels, although accuracy decreased.

The divergence between AdaBoost and Random Forest when reducing channels is likely explained by the underlying mechanisms of the classifier, where Random Forest appears to be better suited to model datasets with smaller feature sets.

It is however important to recognise the limitations of the channel selection method employed in this study. As explained in Section 3.3, the study employed a straightforward channel selection method that did not take all combinations of channels into account. Also, the sequence of channel removal was solely based on the PSD comparison of two recordings, which may not capture the dataset's full diversity.

In addition, a larger number of recorded channels might have led to more successful channel selection because it could provide a more exhaustive view of brain activity patterns related to stress. Therefore, future investigations should aim to implement a more comprehensive channel selection process using the dataset collected in this study. This could potentially lead to a more refined understanding of EEG-based stress detection and improved classifier performance.

This study found that the best-performing frequency bands after performing channel selection were the alpha frequency band for the AdaBoost classifier and the beta frequency band for the Random Forest classifier. These findings align with the theory presented in Section 2.2, where the alpha frequency band was found to obtain the best performance, followed by the beta frequency band. The reason why the best-performing frequency band is different for the AdaBoost and Random Forest classifiers is likely the underlying mechanisms of the classifiers.

5.2 Epoch Length Analysis

Section 4.1 presents the performance of the classifiers when varying the epoch length of the EEG samples. For the AdaBoost classifier, a notable increase in sensitivity was observed when the epoch length was extended from 1 second to 8 seconds. This suggests that AdaBoost may extract more meaningful features from longer data segments, thereby providing a broader context for EEG signals and enhancing the distinction between stressed and non-stressed states. Interestingly, for the selected epoch length of 8 seconds, sensitivity was higher than specificity. This might be preferable in a stress

detection system because false positives (mistakenly classifying a non-stressed subject as stressed) are less harmful than false negatives (failing to identify a stressed subject). False negatives could lead to missed opportunities for stress intervention and management, which is why it is important to avoid them.

In contrast, the Random Forest classifier displayed significantly lower sensitivity than specificity across all tested epoch lengths. Given that high sensitivity is crucial in a reliable stress detection system, this suggests that the Random Forest classifier may not be the most suitable choice.

As for the EEGNet classifier, its performance demonstrated a significant variability across different epoch lengths. However, at the chosen epoch length of 3 seconds, a compromise was found between sensitivity and specificity. As previously stated, a desirable property of the model used for stress detection is that the sensitivity is larger than the specificity. The results showed that this could only be achieved while reducing the accuracy significantly. These results suggest that additional tuning might be necessary to create a reliable stress detection system using EEGNet.

5.3 80:20 Train-Test Split

The results of the study when using an 80:20 train-test split reveal that the EEGNet classifier surpassed both AdaBoost and Random Forest classifiers in terms of accuracy and specificity, with an overall accuracy of 71%. The improved performance of the EEGNet classifier may be attributed to its ability to extract and process temporal and spatial characteristics of the input EEG data efficiently, which could have contributed to its high accuracy and specificity.

However, the AdaBoost classifier outperformed the other two classifiers in terms of sensitivity, with a value of 67%. EEGNet and Random Forest, on the other hand, achieved a sensitivity of 50% and 17%, respectively. Sensitivity is a key performance measure in situations where it is critical to correctly identify positive instances. The higher sensitivity of the AdaBoost classifier implies that it was more capable of correctly detecting true positive cases. Even though it exhibited lower accuracy and specificity as compared to EEGNet and Random Forest, its superior sensitivity might

make it the preferred choice for stress detection due to the aim of reducing false negatives.

5.4 Leave-One-Subject-Out Cross-Validation

The study's findings highlight several important aspects of applying the LOSO cross-validation method to evaluate the performance of the AdaBoost, Random Forest, and EEGNet classifiers. Overall, the classifiers demonstrated an average accuracy of approximately 54%. This accuracy is marginally higher than random choice and indicates that the classifiers are struggling to generalise to unseen data. However, the accuracy substantially varied across individual participants, emphasising the complex nature of EEG data and the number of individual differences between subjects. The inter-subject variability may be attributed to multiple factors such as inconsistent data quality, imprecise labelling, or inherent inter-subject variations. Therefore, future studies should consider these individual differences when designing and training classification models.

An additional concern arose from the confusion matrices of the classifiers. The EEGNet classifier predicted a single class for the majority of participants, which suggests that the model is overfitting. Conversely, the AdaBoost and Random Forest classifiers displayed more balanced predictions. However, it was found that both classifiers had a higher false negative rate than false positive rate, which has negative implications when creating a reliable stress detection system. These observations underscore the importance of considering the confusion matrices in combination with the accuracy measures. The findings underline the necessity of balancing false positive and negative rates in classifier design, especially in stress detection scenarios where misclassification could have significant repercussions. In summary, the results indicate that the EEGNet may be unsuitable as it does not appear to generalise to unseen data. While the AdaBoost and Random Forest classifiers demonstrated potential for EEG-based stress detection, the findings also point to areas that need improvement and the importance of considering individual variations in EEG data.

The level of overfitting observed for the EEGNet classifier suggests that the tech-

niques implemented to prevent overfitting have not been sufficient. Despite implementing measures such as class weighting, stratified 5-fold cross-validation, early stopping, and dropout, as described in Subsection 3.4.5, these strategies did not sufficiently mitigate overfitting. It is apparent that new techniques need to be explored to prevent overfitting in future studies.

5.5 Factors Influencing Classifier Performance

A significant difference in performance is observed when using LOSO cross-validation as compared to an 80:20 train-test split for the EEGNet classifier. One potential reason for this discrepancy is that the specific holdout set selected in the 80:20 train-test split might closely mirror the distribution of the training set. This could artificially inflate the estimated performance of the classifier, as it might be overly tailored to the characteristics of that specific split. It is important that future research acknowledges this potential issue. As a result, the LOSO may in some cases offer a more realistic evaluation of the model's performance in a real-world scenario.

An important factor influencing the analysis is that not all participants provided samples from both stressed and non-stressed classes. Only half of the participants (14 out of 28) had samples from both classes. This implies that the LOSO cross-validation performance evaluation was performed only for 14 of the participants, leading to an incomplete view of the LOSO results. Future studies should aim to collect both stressed and non-stressed samples from all participants, to ensure a more comprehensive evaluation of classifier performance.

The drop in performance when moving from the 80:20 train-test split to LOSO cross-validation is more drastic for the EEGNet classifier compared to the AdaBoost and Random Forest classifiers. One potential reason is the lack of manual filtering for the EEGNet classifier. As discussed in Subsection 3.4.5, the convolutional layers of the EEGNet model perform bandpass filtering, but the automatic filtering might not sufficiently suppress noise in the EEG data.

The performance drop is more noticeable for the EEGNet classifier when shifting from an 80:20 train-test split to LOSO cross-validation compared to AdaBoost

and Random Forest classifiers. This discrepancy may stem from the lack of manual filtering for the EEGNet classifier, as the model was used with raw EEG data. The rationale for this decision, previously explained in Subsection 3.4.5, is that the convolutional layers of the EEGNet model inherently perform bandpass filtering. It is possible that this automatic approach does not adequately suppress the noise in the EEG data, leading to reduced performance.

The model's suitability for sampling rates other than 128 Hz, without data downsampling, might also be an issue. As detailed in Subsection 3.4.5, the model's adaptability to other sampling rates has not been fully explored by its creators. In particular, adjusting the model to support a 250 Hz sampling rate without downsampling could impact both data preprocessing and feature extraction. Future studies should explore the impact of downsampling the EEG data to 128 Hz and using the original EEGNet model for classification, to better understand these performance variations.

Despite using various models, our study did not achieve an accuracy above 71%, significantly lower than the average 90% accuracy reported in related studies that used the alpha frequency band as a feature (Katmah et al. (2021)). Several factors could contribute to this gap in accuracy.

The available amount of data for this study might have influenced the accuracy. A sufficiently large and representative dataset is a critical component when creating reliable stress detection systems that can accurately model the large inter-subject variations found in EEG signals. However, the study used a relatively small dataset with only 83 recordings. The exclusion of 29 EEG recordings to establish a single-modal system baseline removed a significant amount of the available data. This exclusion aimed at preparing for future multimodal systems requiring each recording to consist of both EEG and PCG modalities. Additionally, the removal of an additional 17 moderate samples meant that the final dataset only consisted of 70 recordings.

Class imbalance in the dataset is another significant factor. Of the recordings, 46 represented non-stressed samples, and only 24 represented stressed ones. This 1.91:1 ratio creates a bias towards non-stressed samples, potentially affecting the classifiers' learning process. Classifiers trained on imbalanced datasets often favour the majority class, making it more challenging to achieve high sensitivity.

The number of electrodes utilised for EEG recording in the study also likely influenced the accuracy. We limited the study to eight electrodes, which might not have fully captured the diverse brain activity associated with stress. More electrodes could provide a more comprehensive understanding of neural changes related to stress, capturing subtle changes across different brain regions and frequencies.

Lastly, substantial background noise during several recording sessions might have affected the accuracy, even with the applied filtering techniques. This noise could distort the original EEG signals, complicating feature extraction and stress classification.

In future studies, addressing these factors can improve the accuracy of EEG-based stress detection systems. Considerations include ensuring sufficient and balanced data, using an optimal number of electrodes, and recording data in less noisy environments.

5.6 Hyperparameter Tuning

The results from hyperparameter tuning suggested that the number of estimators used by both the AdaBoost and Random Forest classifiers had little effect on the classifiers' performance. In the case of the AdaBoost classifier, underfitting may occur with too few estimators, while too many estimators may lead to overfitting. The results indicate that the explored range of estimators does not cause underfitting or overfitting. For the Random Forest classifier, the improvement in accuracy typically starts to plateau after a certain number of trees. This suggests exploring a different range of the number of estimators that also includes a smaller number of estimators.

However, the AdaBoost classifier's learning rate and the Random Forest classifier's maximum depth had a significant impact on performance. A larger learning rate for AdaBoost means that each new estimator added to the model contributes more substantially to the final decision, which potentially corrects prior errors more effectively. A higher learning rate enables the AdaBoost model to reach optimal performance more quickly.

For the Random Forest classifier, a larger maximum depth allows each individual decision tree within the forest to capture more complex patterns in the data, which

potentially improves the model's accuracy. However, increasing the maximum depth should be approached with caution, as too high values can lead to overfitting.

For the EEGNet classifier, a learning rate of 1e-3 was determined to be optimal. The learning rate influences the speed at which the model adjusts its weights during training, with lower values leading to slower learning and higher values leading to quicker, but potentially less accurate, learning. A balance is needed, and in this case, a learning rate of 1e-3 appears to achieve that balance.

It is important to note that the results of the hyperparameter tuning presented in this study are based on one combination of data, features, and epoch length for each classifier. Consequently, these findings may not be representative of other combinations. As a result, conclusions drawn from these results should be interpreted with this consideration in mind.

Chapter 6

Conclusions and Future Work

The analysis presented in this thesis yielded several important insights into the effectiveness and the limitations of EEG-based stress detection. The study demonstrated that the delta, theta, alpha and beta frequency bands are the most effective for detecting stress. The gamma frequency band proved to be less effective for detecting stress.

Electrode placement, particularly in the right side of the scalp within the frontal, central, and temporal regions, was also observed to significantly influence the discriminative ability of the classifiers. As a result, future research should focus on enhancing electrode coverage in these areas to improve the accuracy of stress detection systems.

The study also indicated that removing less significant channels had different effects on the classifiers. For the AdaBoost classifier, this led to a drop in sensitivity and a rise in specificity. Conversely, the Random Forest classifier demonstrated improved sensitivity with reduced channels, although with decreased accuracy. Future studies should consider additional electrodes and more advanced channel selection to further explore performance improvements from channel selection.

In terms of epoch length, longer data segments appeared beneficial for the AdaBoost and EEGNet classifier, but less so for the Random Forest classifier.

The best accuracy was obtained using the EEGNet classifier and an 80:20 train-test split. However, improved sensitivity was achieved using the AdaBoost classifier. The low sensitivity of the Random Forest classifier suggests that the model may not be suitable for stress detection with the existing dataset. When comparing the results from an 80:20 train-test split with the LOSO cross-validation method, the LOSO results reveal that the EEGNet classifier is only predicting one class. This suggests that the EEGNet classifier is not generalising well, and that further hyperparameter tuning may be needed.

Given the challenges of creating an EEG model with high accuracy and sensitivity, the study points to several areas for improvement related to data collection. Future studies should aim to obtain sufficient and balanced data from participants, use an optimal number of electrodes and work to reduce environmental noise during recording sessions.

Lastly, hyperparameter tuning presented minimal impact on the performance of both the AdaBoost and Random Forest classifiers in terms of the number of estimators. However, the learning rate of the AdaBoost classifier and the maximum depth of the Random Forest classifier demonstrated a significant influence on performance. For EEGNet, a learning rate of 1e-3 appeared to strike a balance between learning speed and accuracy.

In conclusion, while this study provides valuable insights into the importance of various frequency bands for EEG-based stress detection, it also highlights the challenge of creating high-quality, balanced datasets. Moving forward, a concerted effort to address these issues through methodological refinement, comprehensive data collection, and continuous fine-tuning of models will be crucial in advancing the reliability and effectiveness of EEG-based stress detection.

Appendix A

Appendix

A.1 Informed Consent.pdf

Department of Engineering Cybernetics

DATA ACQUISITION CONSENT FORM

You are being invited to participate in a research study, which the Norwegian Center for Research Data (NSD) has reviewed and approved for conduction by the investigators named here. This form is designed to provide you - as a human subject - with information about this study. The investigator or his/her representative will describe this study to you and answer any of your questions. You are entitled to a copy of this form. If you have any questions or complaints about the informed consent process of this research study or your rights as a subject, please contact the PI or Co-PI (marta.molinas@ntnu.no, +47 94287670, andres.f.soler.guevara@ntnu.no).

Project Title: FlexEEG in Mental Health

Principal Investigators: Marta Molinas

Co-investigator: Andres Soler & Mohit Kumar

Thank you for agreeing to participate in this research project. This study involves research aimed at detecting the presence of psychological stress in the human body based on the analysis of EEG and PCG signals. You will participate in two separate data collection sessions. The first session will take place in the exam period of nov-dec 2022, and the second will take place after the holidays, early 2023. Before each session we will ask you to answer a self-evaluation questionnaire called 'State-Trait Anxiety Inventory'. This questionnaire will be used to determine whether you are stressed or not. During both sessions, you will be recorded twice: one five-minute period with no stressor, and one five-minute period with an Arithmetic stressor. You will be asked to rate your stress level on a scale from 1-10 after each recording. The Arithmetic stressor consists of different arithmetic statements presented on a screen. Your task will be to calculate each task in your head and click "T" on the keyboard if the statement is True, and "F" if it is False. This task is supposed to induce stress so please keep this in mind. Each session will last about 30 minutes. 10 of these minutes are for recording of EEG and PCG signals using Mentalab EEG and EkoDuo stethoscope. We will clean the areas of the scalp where the electrodes are placed with isopropyl alcohol. Electrode cap gel will be applied to the areas, but it is easily washed out with water and shampoo.

Participation in this study will take approximately 60 minutes of your time. We warn that the set-up of the EEG cap can lead to some discomfort, and the tasks you are given will (hopefully) induce some stress response. Your participation in this study is completely voluntary. Should you decide to discontinue participation or decline to answer any specific part of the study, you may do so without penalty.

Your participation in this study may help you understand the manifestations of stress on EEG signals. We are not asking you to place your name anywhere on the experimental booklet, so your participation is anonymous. None of your answers can be directly traced back to you. Should you have any further questions, please feel free to contact the study's principal investigator or co-PI, Marta Molinas and Andres Soler at the Department of Engineering Cybernetics. Her office is at Elektro D+B2 room D244, her phone number is +47 94287670, and her e-mail address is marta.molinas@ntnu.no.

By signing below, I confirm that:

- I give my consent to participate in the research study entitled “FlexEEG in Mental Health”.
- I hereby confirm that I have read the above information and have been informed about the content and purpose of the research.
- I fully understand that I may withdraw from this research project at any time without prejudice or effect on my standing with NTNU.
- I also understand that I am free to ask questions about techniques or procedures that will be undertaken.
- I give my consent for the collection and use of all data of the research “EEG and PCG in mental health” for use in research and teaching purposes.
- I give my consent to use my data for scientific purposes, its documentation and publications (including any exhibitions and further publications)
- I hereby declare that I am currently not diagnosed by a with any heart disease, or neurological disease
- I am also not on any medications affecting heart rate and/or brain wave function
- I hereby declare that I am not officially diagnosed with any mental illness

Date and place: _____ and _____

Participant’s signature: _____

First and last name: _____

Date of Birth and current Age: _____and _____

I hereby certify that I have given an explanation to the above individual of the contemplated study and its risks and potential complications.

29/11/2022

_____ _____

Principal Investigator’s signature Date

A.2 STAI-Y Form.pdf

State-Trait Anxiety Inventory for Adults

Self-Evaluation Questionnaire

STAI Form Y-1 and Form Y-2

Developed by Charles D. Spielberger
in collaboration with R.L. Gorsuch, R. Lushene, P.R. Vagg, and G.A. Jacobs

Published by Mind Garden
1690 Woodside Road Suite 202, Redwood City, CA 94061 USA 650-261-3500
www.mindgarden.com

A.3 Participant Screening.pdf

Participation form
Mental health EEG and PCG data collection

We need participants for collection of EEG and PCG data for stress detection. All participants will receive a **Midtbyen gift card** valued **200 NOK** as a thanks. The data will be used in our master's thesis'. The data will of course be anonymized, and will not be traceable back to you.

The experiment will take about 30 minutes including set-up and recording. The recording will be done using non-invasive methods: an electrode-cap for EEG and a digital stethoscope for PCG. We will need two recordings of each participant: One during the exam period of nov-dec 2022 (stressed state), and one after the holidays (non-stressed state).

Please sign up via the form if you're interested in participating.

* Required

* This form will record your name, please fill your name.

1. Name *

2. Email *

3. Telephone number *

4. Gender *

- ◯ Female
- ◯ Male
- ◯ Other/prefer not to say

5. I am not diagnosed with any heart and/or neurological disease *

- ◯ True
- ◯ False

6. I am not on any medications affecting heart rate and/or brain wave function *

- ◯ True
- ◯ False

7. I am not diagnosed with any mental illness *

- ◯ True
- ◯ False

8. I consent to being contacted for participation in this study *

- ◯ Yes
- ◯ No

References

Association., W. M. (2001). World medical association declaration of helsinki. ethical principles for medical research involving human subjects., *Bulletin of the World Health Organization* **79**(4): 373 – 374.

AudioCapture (2018). `https://github.com/labstreaminglayer/App-AudioCapture`. [Online; accessed 10-May-2023].

Bergua, V., Meillon, C., Potvin, O., Bouisson, J., Le Goff, M., Rouaud, O., Ritchie, K., Dartigues, J.-F. and Amieva, H. (2012). The stai-y trait scale: psychometric properties and normative data from a large population-based study of elderly people, *International Psychogeriatrics* **24**(7): 1163–1171.

Bhatnagar, S., Khandelwal, S., Jain, S. and Vyawahare, H. (2023). A deep learning approach for assessing stress levels in patients using electroencephalogram signals, *Decision Analytics Journal* p. 100211.

Breiman, L. (1996). Bagging predictors, *Machine learning* **24**: 123–140.

Breiman, L. (2001). Random forests, *Machine learning* **45**: 5–32.

Brown, G. W. and Harris, T. (1978). Social origins of depression: a reply, *Psychological medicine* **8**(4): 577–588.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.

Cohen, S., Janicki-Deverts, D. and Miller, G. E. (2007). Psychological stress and disease, *Jama* **298**(14): 1685–1687.

Dohrenwend, B. S. and Dohrenwend, B. P. (1974). *Stressful life events: Their nature and effects.*, John Wiley & Sons.

Donos, C., Dümpelmann, M. and Schulze-Bonhage, A. (2015). Early seizure detection algorithm based on intracranial eeg and random forest classification, *International journal of neural systems* **25**(05): 1550023.

Eko: Digital Stethoscope, ECG (2019). `https://apps.apple.com/no/app/eko-digital-stethoscope-ecg/id1599911580`. [Online; accessed 11-Feb-2023].

ExplorePy (2019). `https://mentalab.com/eeg-software/`. [Online; accessed 10-May-2023].

Fraiwan, L., Lweesy, K., Khasawneh, N., Wenz, H. and Dickhaus, H. (2012). Automated sleep stage identification system based on time–frequency analysis of a single eeg channel and random forest classifier, *Computer methods and programs in biomedicine* **108**(1): 10–19.

Freund, Y., Schapire, R. E. et al. (1996). Experiments with a new boosting algorithm, *icml*, Vol. 96, Citeseer, pp. 148–156.

Halim, Z. and Rehan, M. (2020). On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning, *Information Fusion* **53**: 66–79.

Hosseini, S. A., Khalilzadeh, M. A., Naghibi-Sistani, M. B. and Niazmand, V. (2010). Higher order spectra analysis of eeg signals in emotional stress states, *2010 Second international conference on information technology and computer science*, IEEE, pp. 60–63.

Hu, J. (2017). Automated detection of driver fatigue based on adaboost classifier with eeg signals, *Frontiers in computational neuroscience* **11**: 72.

Katmah, R., Al-Shargie, F., Tariq, U., Babiloni, F., Al-Mughairbi, F. and Al-Nashash, H. (2021). A review on mental stress assessment methods using eeg signals, *Sensors* **21**(15): 5043.

Keras (2015). `https://keras.io`. [Online; accessed 10-May-2023].

Khosrowabadi, R., Quek, C., Ang, K. K., Tung, S. W. and Heijnen, M. (2011). A brain-computer interface for classifying eeg correlates of chronic mental stress, *The 2011 International Joint Conference on Neural Networks*, pp. 757–762.
URL: *https://ieeexplore.ieee.org/abstract/document/6033297*

Lab Streaming Layer (2011). `https://labstreaminglayer.org/`. [Online; accessed 10-May-2023].

Lai, C. Q., Ibrahim, H., Abdullah, M. Z., Abdullah, J. M., Suandi, S. A. and Azman, A. (2018). Artifacts and noise removal for electroencephalogram (eeg): A literature review, *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pp. 326–332.

Lawhern, V. J. (2018). ARL-EEGModels, `https://github.com/vlawhern/arl-eegmodels`. [Online; accessed 10-May-2023].

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P. and Lance, B. J. (2018). Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces, *Journal of neural engineering* **15**(5): 056013.

Lazy Predict (2019). `https://github.com/shankarpandala/lazypredict`. [Online; accessed 10-May-2023].

Lim, C.-K. A. and Chia, W. C. (2015). Analysis of single-electrode eeg rhythms using matlab to elicit correlation with cognitive stress, *International Journal of Computer Theory and Engineering* **7**(2): 149.

Liu, L., Ji, Y., Gao, Y., Li, T. and Xu, W. (2022). A novel stress state assessment method for college students based on eeg, *Computational Intelligence and Neuroscience* **2022**.

Marthinsen, A. J. Y. (2022). Detection of mental stress from eeg data using artificial intelligence.

MNE-Features (2018). `https://mne.tools/mne-features/`. [Online; accessed 10-May-2023].

MNE-Python (2011). `https://mne.tools/stable/index.html`. [Online; accessed 10-May-2023].

Niedermeyer, E. and da Silva, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*, Lippincott Williams & Wilkins.

Nisbet, R., Miner, G. and Yale, K. (2018). Chapter 9 - classification, *in* R. Nisbet, G. Miner and K. Yale (eds), *Handbook of Statistical Analysis and Data Mining Applications (Second Edition)*, second edition edn, Academic Press, Boston, pp. 169–186.
URL: *https://www.sciencedirect.com/science/article/pii/B9780124166325000098*

Oppenheim, A. V. (1999). *Discrete-time signal processing*, Pearson Education India.

Polikar, R. (2012). Ensemble learning, *Ensemble machine learning: Methods and applications* pp. 1–34.

Prabhakar, S. K. and Rajaguru, H. (2018). Adaboost classifier with dimensionality reduction techniques for epilepsy classification from eeg, *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017, Thessaloniki, Greece, 18-21 November 2017*, Springer, pp. 185–189.

PsychoPy (2015). `https://www.psychopy.org`. [Online; accessed 10-May-2023].

pyXDF (2019). `https://github.com/xdf-modules/pyxdf/blob/main/pyxdf/pyxdf.py`. [Online; accessed 10-May-2023].

Rajoub, B. (2020). Chapter 2 - characterization of biomedical signals: Feature engineering and extraction, *in* W. Zgallai (ed.), *Biomedical Signal Processing and Artificial Intelligence in Healthcare*, Developments in Biomedical Engineering and Bioelectronics, Academic Press, pp. 29–50.
URL: *https://www.sciencedirect.com/science/article/pii/B9780128189467000020*

Sanei, S. and Chambers, J. (2007). *Introduction to EEG*, John Wiley & Sons, Ltd, chapter 1, pp. 1–34.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470511923.ch1*

scikit-learn (2007). `https://scikit-learn.org/stable/`. [Online; accessed 10-May-2023].

Selye, H. (1965). The stress syndrome, *The American Journal of Nursing* **65**(3): 97–99.
URL: *http://www.jstor.org/stable/3453119*

Seo, S.-H., Lee, J.-T. and Crisan, M. (2010). Stress and eeg, *Convergence and hybrid information technologies* **27**.

Shoji, T., Yoshida, N. and Tanaka, T. (2021). Automated detection of abnormalities from an eeg recording of epilepsy patients with a compact convolutional neural network, *Biomedical Signal Processing and Control* **70**: 103013.

Silverstein, S. T. and Kritz-Silverstein, D. (2010). A longitudinal study of stress in first-year dental students, *Journal of dental education* **74**(8): 836–848.

Sletten, C. (2022). *Automated Stress Detection using Electroencephalogram Signals*, PhD thesis, Norwegian University of Science and Technology (NTNU).

Sletten, C. (2023). Eeg stress detection master, `https://github.com/wavesresearch/eeg_stress_detection_master/`.

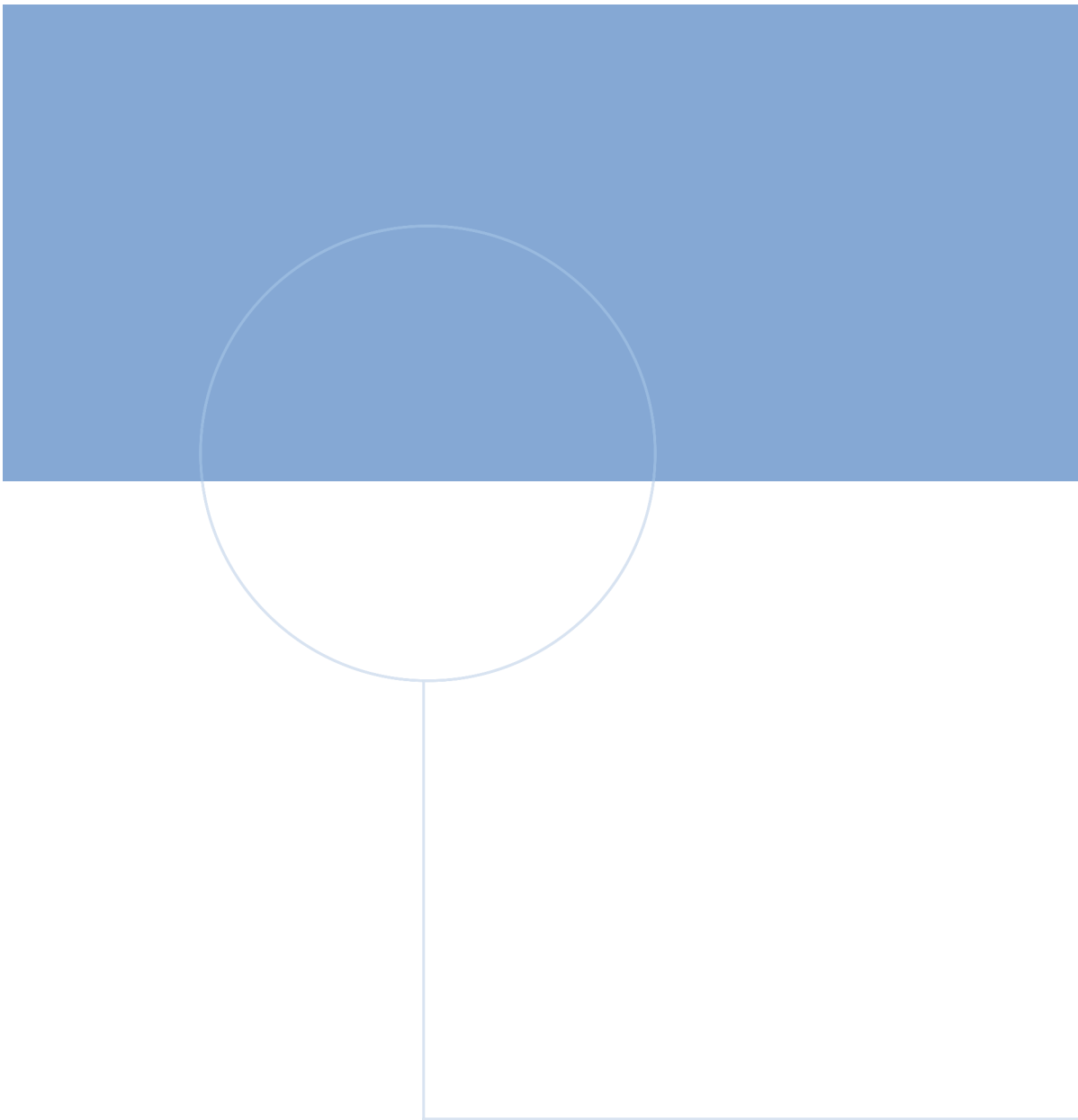
Spielberger, C., Goruch, R., Lushene, R., Vagg, P. and Jacobs, G. (1983). Manual for the state-trait inventory stai (form y), *Mind Garden, Palo Alto, CA, USA* .

Stancin, I., Cifrek, M. and Jovic, A. (2021). A review of eeg signal features and their application in driver drowsiness detection systems, *Sensors* **21**(11).
URL: *https://www.mdpi.com/1424-8220/21/11/3786*

Tensorflow (2015). `https://www.tensorflow.org`. [Online; accessed 10-May-2023].

Urigüen, J. A. and Garcia-Zapirain, B. (2015). Eeg artifact removal—state-of-the-art and guidelines, *Journal of Neural Engineering* **12**(3): 031001.
URL: *https://dx.doi.org/10.1088/1741-2560/12/3/031001*

Welch, P. (1967). The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Transactions on audio and electroacoustics* **15**(2): 70–73.



NTNU

Norwegian University of
Science and Technology