BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF COMPUTER SYSTEMS
ÚSTAV POČÍTAČOVÝCH SYSTÉMŮ

DIAGNOSING ANXIETY AND DEPRESSION FROM BRAIN ELECTROENCEPHALOGRAM (EEG) SIGNALS

DIAGNOSTIKA ÚZKOSTI A DEPRESE ZE SIGNÁLŮ MOZKOVÉHO ELEKTROENCEFALOGRAMU (EEG)

MASTER'S THESIS
DIPLOMOVÁ PRÁCE

AUTHOR **BC. MARTIN OSVALD**
AUTOR PRÁCE

SUPERVISOR **doc. AAMIR SAEED MALIK, Ph.D.**
VEDOUCÍ PRÁCE

BRNO 2024

BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Master's Thesis Assignment

153393

Institut:	Department of Computer Systems (DCSY)
Student:	**Osvald Martin, Bc.**
Programme:	Information Technology and Artificial Intelligence
Specialization:	Machine Learning
Title:	**Diagnosing anxiety and depression from brain electroencephalogram (EEG) signals**
Category:	Biocomputing
Academic year:	2023/24

Assignment:

1. Study and learn about anxiety & depression and how they affect the brain in terms of brain signals and images.
2. Get acquainted with signal & image processing methods as well as machine learning techniques and their application to the brain EEG signals and images.
3. Find out the challenges for diagnosing anxiety & depression from brain signals and images as well as the limitations of the existing methods.
4. Design an algorithm for diagnosing anxiety & depression from brain EEG signals and images.
5. Implement the designed algorithm.
6. Create a set of benchmark tasks to evaluate the quality of anxiety & depression diagnostics as well as the corresponding computational performance and memory usage.
7. Conduct critical analysis and discuss the achieved results and their contribution.

Literature:

- According to supervisor's advice.

Requirements for the semestral defence:

- Items 1 to 4 of the assignment.

Detailed formal requirements can be found at https://www.fit.vut.cz/study/theses/

Supervisor:	**Malik Aamir Saeed, doc., Ph.D.**
Head of Department:	Sekanina Lukáš, prof. Ing., Ph.D.
Beginning of work:	1.11.2023
Submission deadline:	17.5.2024
Approval date:	30.10.2023

Abstract

Mental disorders represent inevitable emotions in our society. These psychological states affect the cognitive, emotional and behavioural functioning of individuals. Common mental disorders fall into two main diagnostic categories: depressive disorders and anxiety disorders. The aim of this work is to find a new method for detecting whether a given patient suffers from anxiety or depression using EEG classification. In this work, we use combination of genetic algorithms and models from deep learning.

Abstrakt

Duševné poruchy predstavujú širokú škálu emócií v našej spoločnosti. Tieto psychické stavy významne ovplyvňujú kognitívne, emocionálne a behaviorálne fungovanie jednotlivcov. Bežné duševné poruchy sa vzťahujú na dve hlavné diagnostické kategórie: depresívne poruchy a úzkostné poruchy. Cielom tejto práce je nájsť novú metódu na detekciu či daný pacient trpí úzkosťou alebo depresiou pomocou klasifikácie EEG. V tejto práci používame kombináciu genetických algoritmov a modelov z hlbokého učení.

Keywords

EEG, Major Depressive Disorder, Anxiety, genetics algorithm, Machine Learning, SVM, CNN, Bert, Transformers

Klíčová slova

EEG, depresia, úzkosť, genetické algoritmy, Strojové učenie, SVM, CNN, Bert, Transformers

Reference

OSVALD, Bc. Martin. *Diagnosing anxiety and depression from brain electroencephalogram (EEG) signals.* Brno, 2024. Master's thesis. Brno University of Technology, Faculty of Information Technology. Supervisor doc. Aamir Saeed Malik, Ph.D.

Diagnosing anxiety and depression from brain electroencephalogram (EEG) signals

Declaration

I hereby declare that this Master's thesis was prepared as an original work by the author under the supervision of Mr. Malik. I have listed all the literary sources, publications and other sources, which were used during the preparation of this thesis.

.....
Bc. Martin Osvald
May 17, 2024

Acknowledgements

I would like to thank to Mr. Malik for his expertise and professionalism and friendly approach as well as guidance.

Contents

List of Figures

Chapter 1

Introduction

Mental disorders are a wide range of emotions in our society. These psychological conditions significantly affect individuals' cognitive, emotional, and behavioural functioning. Common mental disorders refer to two main diagnostic categories: depressive disorders and anxiety disorders [35].

The World Health Organization estimates that 264 million people worldwide suffered from anxiety disorder and 322 million from depressive disorder in 2015, corresponding to prevalence rates of 3.6 [35].

Through the COVID-19 pandemic, many people suffer financial loss, loneliness, or even the death of close friends. Furthermore, mental disorders are correlated with cardiovascular disease, as well as with the death rate in the population[78]. Quality of life (QoL) is a subjective measure of happiness, although groups with mental disorders automatically have a lower QoL than others[35]. Furthermore, mental disorders reduce the quality of life over time[35].

The primary purpose of this Master's thesis is to use various techniques of machine learning and genetics algorithms to classify if a patient suffers an anxiety or Major Depressive Disorder when the subject is correctly connected to the Brain-Computer Interface.

Another purpose of this thesis is to help psychiatrists and specialists with hard evidence of data based on the classification of actual measured data. Most of the time, diagnosing Anxiety or Major Depressive Disorder starts with a physical exam and questionnaire [81]. Diagnosis using an EEG device-based classification could improve or even speed up a process, preventing more health damage or even preventing suicides. Automating the process of detecting mental disorders is necessary.

Chapter 2

Anxiety and Depression

Anxiety and Depression are known as common mental illnesses, although there is a significant difference. Feelings of nervousness, worry, and fear cause anxiety. Anxiety feels like you are constantly on edge or scared.

Depression includes sadness, hopelessness, and fatigue. Depression significantly affects daily functioning. Treatment for anxiety usually involves talking to a therapist about being afraid, treating by mixing medications and talking with therapists and psychiatrists. If anxiety is not properly treated at the right time, it can develop into major depressive disorder [2]. Furthermore, the prevalence of anxiety and depression among college students is a growing concern, with a significant percentage reporting these mental health disorders [33].

2.1 Anxiety Disorders

Experiencing anxiety is a health concern that involves feelings of stress, panic or fear that can greatly affect ones life [63]. It's important to understand that anxiety is distinct, from fear even though people often use these terms interchangeably. Individuals with anxiety disorders may have thoughts or concerns that lead them to avoid situations as a way of coping. This avoidance behaviour is an indicator of anxiety. Physical symptoms of anxiety can include a heart rate, tightness in the chest, sweating, trembling, dizziness and a rapid heartbeat [57]. In addition to these signs, anxiety also manifests itself psychologically. These mental signs may involve thoughts of losing control or the desire to escape from challenging situations.

The causes of anxiety can vary significantly between individuals due to the contributing factors involved in its nature. Some common triggers for anxiety include imbalances in brain chemistry, traumatic experiences, chronic illnesses and substance abuse [57]. There are types of anxiety disorders, such as Generalised Anxiety Disorder (GAD), Social Anxiety Disorder (SAD), Panic Disorder, and specific phobias [3]. Managing anxiety requires a variety of treatments, including therapy sessions, medication regimens, and self-care practices. Dealing with anxiety effectively involves addressing both the psychological aspects since it is a complex condition that impacts individuals holistically.

Anxiety is a health concern marked by feelings of stress, panic or fear, which can significantly impact an individual's life.

Anxiety symptoms can differ, ranging from signs like a heartbeat and sweating to mental cues, such as intrusive thoughts and avoidance behaviours. People often mix up anxiety

with fear, though they are experiences. The reasons behind anxiety vary, with triggers being unique to each individual. Dealing with and addressing anxiety requires a mix of approaches, including therapy, medication and self-care routines. Due to the availability of relevant datasets online, my thesis is specifically interested and restricted to classifying Social Anxiety Disorder.

2.2 Research of Social Anxiety Disorder

This research study [51] of Social Anxiety Disorder (SAD) discusses examining Electroencephalogram (EEG) recordings, especially indicators such as ERN, which stands for error-related negativity or correct-response negativity CRN. An extensive analysis of 66 articles underscores the value of ERN in regulation as a strong cross-diagnostic characteristic indicative of clinical anxiety. Nevertheless, future investigations are needed for individuals of older ages (60 and above) to confirm this indicator and its clinical application. The research encompassed in the analysis employed emotional attentional control tasks, underscoring the necessity for further exploration in this domain with broader clinical anxiety populations like those with specific phobias and SAD to validate ERN clinical relevance as a cross-diagnostic trait marker for clinical anxiety overall.

Another study [12] is about classifying several anxiety levels using Convolutional Neural Networks (CNN), Long short-term memory (LSTM) and CNN+LSTM models on preprocessed EEG signals, which are converted then to a Pairwise Channel Difference (PCD) matrix. The PCD is used here as a source of normalisation of EEG signals.

$$PDC_{ij} = \frac{A_{ij}(f)}{\sqrt{a_j - H(f)\overline{a_j(f)}}} \tag{2.1}$$

where:

- PDC_{ij} represents the directional influence and intensity of the information flow from channel j to channel i at a frequency of f.
- $A(F) = \sum_{r=1}^{p} \mathbf{A}_r e^{-i2\pi fr}$

The calculation of the PDC matrix for each frequency band was obtained using continuous 3-second segments of EEG recording. Since the duration of the EEG signal for each subject is 180 seconds, a total of 60 (8×8)-PDC matrices are obtained for each frequency band. Therefore, the PDC matrices for each subject from 5 frequency bands are equal to 300 [12]. The study has results better in using a combination of CNN+LSTM with performation of **93%** accuracy. However, in this study, we don't see a comparison between different segments of time for the calculation of PCD metrics. In this study, we can see that a combination of CNNs and RNNs has better results overall in Table 2.1. We can also clearly see that LSTM preform worse than CNN in Every category. It's worth noting that the Sensitivity of the label Moderate reaches **100%** without the combination of LSTM.

Deep Model	Eval. Metrics	Severe	Moderate	Mild	Control	Average
LSTM	Sensitivity	57.14	71.43	71.43	85.71	71
	Specificity	85.71	90.48	90.48	95.24	90
	Precision	57.14	71.43	71.43	85.71	71
	Accuracy	78.57	85.71	85.71	92.86	**86**
CNN	Sensitivity	71.43	100.00	75.00	85.71	83
	Specificity	90.48	95.45	95.00	95.24	94
	Precision	71.43	85.71	85.71	85.71	82
	Accuracy	85.71	96.43	89.29	92.86	**91**
CNN+LSTM	Sensitivity	85.71	85.71	87.50	83.33	86
	Specificity	95.24	95.24	100.00	90.91	95
	Precision	85.71	85.71	85.71	71.43	86
	Accuracy	92.86	92.86	96.43	89.29	**93**

Table 2.1: Performance of Deep Models by SAD Level from [12].

This paper [53] suggests that biomarkers of anxiety also depend on the location of the subject, not only on EEG recordings. This is understandable because subjects can feel differently in a work or school environment than at home. This suggests a slight bias in every EEG recording session. Subjects can definitely feel anxious and fearful when they visit the doctor. The research involved 60 subjects using Apple iPhones and Oura Rings. Data was collected through the Delphi data acquisition app. The study suggests there is a need for further investigation, which can help to predict and diagnose the anxiety levels of the subject.

		Depression		Anxiety		Stress	
Category	**Feature**	**r**	**p**	**r**	**p**	**r**	**p**
GPS features	Location variance	-0.31	0.035*	-0.26	0.110	-0.28	0.077
	Total distance	-0.26	0.111	-0.28	0.077	-0.17	0.572
	Location entropy	-0.30	0.035*	-0.17	0.572	-0.22	0.251
	Norm. loc. entropy	-0.26	0.100	-0.13	0.788	-0.20	0.298
	Homestay	0.12	0.788	0.13	0.788	0.07	0.788
Smartphone	Usage time	0.09	>0.99	0.05	>0.99	0.07	>0.99
	Usage frequency	0.15	0.716	0.24	0.079	0.12	0.971
Wearable device	Steps	-0.23	0.516	-0.16	>0.99	-0.19	>0.99
	Metabolic equi.	-0.21	0.99	-0.07	>0.99	-0.06	>0.99
	Total sleep time	0.15	>0.99	0.04	>0.99	0.12	>0.99
	Sleep onset latency	0.08	>0.99	0.06	>0.99	0.04	>0.99
	Waking up	0.20	>0.99	0.14	>0.99	0.16	>0.99
	Time in bed	0.15	>0.99	0.03	>0.99	0.11	>0.99
Mood	Arousal	-0.30	0.003*	-0.35	0.001*	-0.36	0.001*
	Valence	-0.48	<0.001*	-0.44	<0.001*	-0.48	<0.001*

Table 2.2: Correlation of Features with Depression, Anxiety, and Stress from [53].

In Table 2.2, we can see that the Usage time of Smartphones correlates with Depression and Anxiety and also with Stress. Also, sleep time is correlated with Depression, Anxiety and as well Stress.

Another paper [71] for research in Social Anxiety Disorder uses Stacked Sparse Autoencoders models for classifications of award-winning Dasps dataset [15]. The results they achieved are accuracies of **83.93%** and **70.25%** using Stacked Sparse Autoencoders and Decision Tree for two-class anxiety classification. They were also able to reduce feature redundancy. The problem with this paper is that they compare just the results with binary labelled data, even though the Dasps dataset has four labels. Authors of [15] mainly use Support vector machine classifiers and k-Neirest-Neighbours classifiers, which are compared in this paper with Stacked Sparse Autoencoders.

2.3 Major Depressive Disorder

Major Depressive Disorder (MDD) is a mental health condition. It is characterised by persistent sadness, chaotic sleep cycles, loss of interest and dramatic changes in weight [67]. It also correlated with thoughts on death and suicide [67]. Many subjects who were diagnosed with MDD have a loss of interest in things and the presence of negative feelings as indicators of depression. Also, poverty and unemployment are risk factors for developing this condition.

Brain Regions Impacted

The critical difference between Major Depressive Disorder and Social Anxiety Disorder is that MDD physiologically change brain itself [21]. The most consistent findings in the affected brain regions are reduced hippocampus volumes and cortical volume in the medial and superior temporal regions. These changes are often seen with Major depressive disorder, Alzheimer and Schizophrenia [21]. We can clearly see the brain damage in figure 2.1.

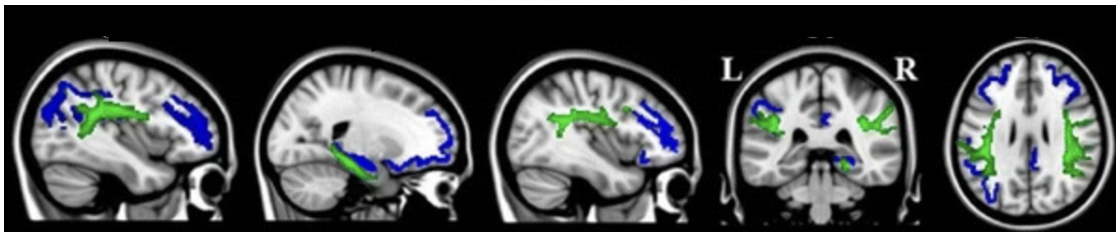

Figure 2.1: Blue: Regional Core thinking. Green White matter tract damage from [21].

2.4 Research in Major Depressive Disorder

Automated classification of Major Depressive Disorder

This paper [59] discussed the detection of MDD using deep learning models. The proposed research suggests that the classification of MDD is better performing on eyes-closed subjects than on eyes-open subjects. Also, study discussed the idea of removing the channels based on their significance with a backwards-elimination algorithm [36]. Overall, **91.67%** accuracy with the full set of channels and **87.5%** after channel reduction.

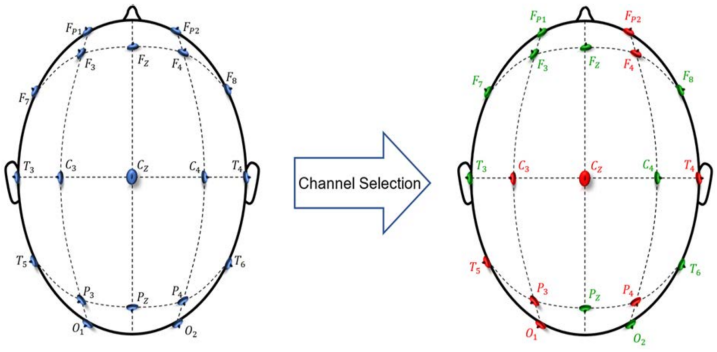

Figure 2.2: Channel selection from [59].

EEG properties of MDD

In a research study [38], an EEG analysis was performed to compare differences in brain activity between groups of individuals. The study involved 25 patients with major depressive disorder (MDD), 26 individuals, and 19 patients with depression. Interestingly, it was observed that those in the depression group had **less alpha activity** (indicating higher activation) in the right anterior hemisphere compared to the left. It should be noted that participants with right and left hands were included in this study.

Using EEG features to the classification of MDD

This paper [68] compares classical models of Machine Learning like SVM, k-NN, etc., to the classification of depression using EEG features instead of recorded EEG signal or PCD matrix. The paper, however, doesn't explain why they picked the represented feature. In chapter number 7, we introduce a possible solution for selecting the features based on genetics algorithms.

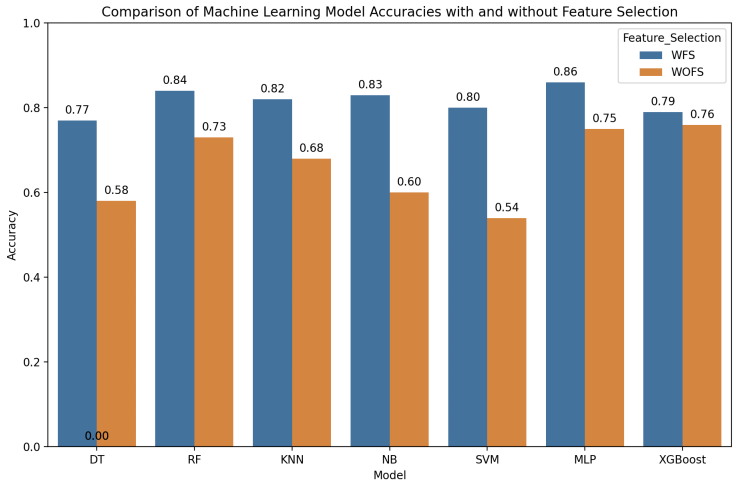

Figure 2.3: Bar graph of performance from [68].

2.5 Relation between SAD & MDD

This paper [45] analyses similarities between Social Anxiety Disorder and Major Depressive Disorder. Using network analyses, the study aimed to examine the symptoms that may play a role in the co-occurrence of SAD and MDD. The clinical subjects diagnosed were picked, specifically 130 women. The research lacks of male subjects. The study showed that the main symptoms of social fear and depression were found at opposite ends of the network and were only slightly connected. Bridges linking these symptoms were formed through other factors, especially feelings of worthlessness.

Another paper [9] is trying to analyse the subject who suffers from Social Anxiety Disorder with Major Depressive Disorder. The analysis consists of $N = 20123$ subjects who suffer from MDD or MDD-SAD combination or MDD with one anxiety disorder or more than two anxiety disorders. In Table 2.3, we can see that younger people aged between 18 and 29 suffer less than other people in the population. Also, the paper suggests that the male gender has a more significant factor in developing MDD-SAD. Still, the female gender has a more significant factor in developing more than two anxiety disorders.

	MDD alone	**MDD-SAD**	**MDD-1 ANX**	**MDD-≥2 ANX**
Total	4.71 (4.27–5.19)	0.77 (0.55–1.08)	3.71 (3.33–4.12)	7.43 (6.58–8.37)
Gender				
Male	4.36 (3.79–5.02)	0.88 (0.55–1.40)	2.71 (2.19–3.35)	4.92 (4.16–5.83)
Female	5.01 (4.47–5.63)	0.67 (0.49–0.93)	4.61 (4.04–5.25)	9.68 (8.60–10.88)
Age				
18–29	3.92 (3.07–4.98)	0.89 (0.62–1.28)	3.70 (3.05–4.48)	7.28 (6.11–8.65)
30–44	5.38 (4.63–6.25)	0.98 (0.58–1.65)	4.14 (3.26–5.24)	8.61 (7.43–9.95)
45–59	5.34 (4.49–6.34)	0.52 (0.31–0.85)	4.14 (3.46–4.94)	9.07 (7.26–11.29)
60+	3.80 (2.95–4.88)	0.66 (0.33–1.30)	2.55 (1.94–3.34)	3.80 (2.86–5.03)

Table 2.3: Prevalence Rates by Category

Although numerous studies have shown a high rate of comorbidity between Social Anxiety Disorder (SAD) and Major Depressive Disorder (MDD), our understanding remains limited regarding how individuals transition from experiencing one set of symptoms to the other[45].

Chapter 3

Brain & EEG

The brain is the most complex part of the human body. It weighs around 1.4 kg. This organ is the centre of the nervous neural system in our bodies. We can categorise the responsibilities of the human brain in cognitive function, sensory processing, motor control, emotional regulation, language processing, homeostatic functions, autonomous control, awareness, sleep regulation, learning, and adaptation.[4] The brain is also a vital part of the memory system for storing facts, ideas, and personal experiences.

3.1 Anatomy of Neuron

The main building block of the brain is the neurone. Neurones are structural, functional, and central to the nervous system. They transmit information through chemical and electrical signals [46]. The central part of the neuron is the cell body called *Soma*. *Soma* acts as the nucleus of the neurone. *Soma* contains all genetic material essential for neuronal function. *Dendrites* act as an input receiver. It has a tree structure and accepts information from other cells.

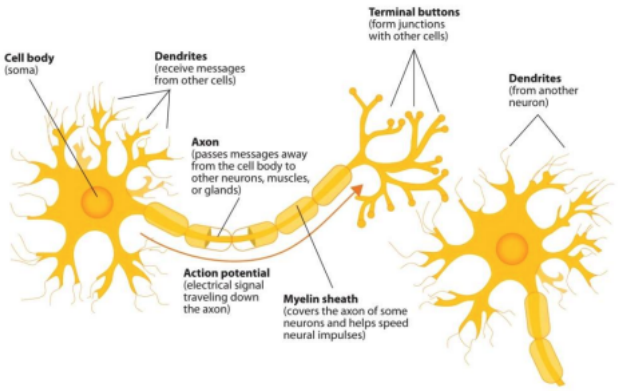

Figure 3.1: The components of neuron from [46].

The pivotal juncture for decision-making in a neuron occurs at the *Axon* hillock, where the *Axon* begins from the cell body. If electrical inputs from *dentrites* exceed a certain threshold, an electrical impulse is created that is shown as an action potential. These spikes are then measured and recorded on an EEG.

The *Axon* is a cable-like structure which connects *Dentrites* from two cells. It transmits action potentials over a distance between these two cells. This process of transferring an action potential ensures information transmission between neurones.

3.2 Anatomy of the Brain

Knowing the brain's structure is essential for caring for it and understanding health treatment techniques. The brain consists of two parts: the left and right cerebral hemispheres.

The frontal lobe is associated with complex cognitive functions such as thought, planning, problem-solving, and aspects of personality and emotional makeup. The right frontal lobe excelled in mathematical processing and planning, and the left frontal lobe excelled in cognitive flexibility and mental flexibility.[22].

The perilateral lobe processes sensory information from the body. Then, it helps to understand spatial orientation and navigation. The occipital lobe is the centre of visual processing in the brain. The temporal lobe is essential for processing auditory information and involves memory and emotion. The prefrontal cortex is considered the executive part of the brain, it is responsible for decision making, social behaviour, and personality expression.

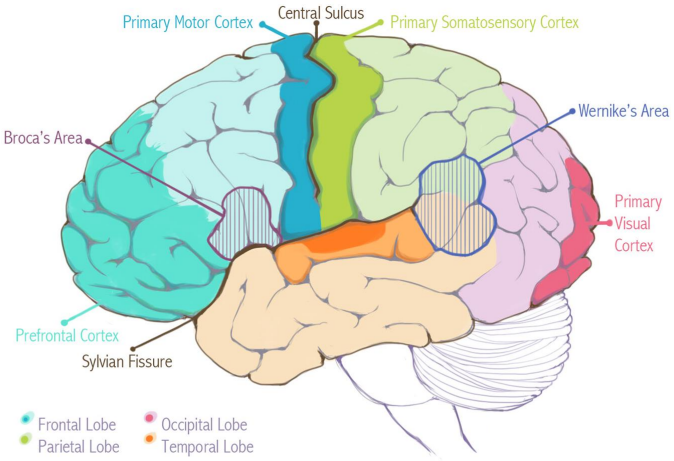

Figure 3.2: Brain structure from [1].

The Primary Motor Cortex is located just in front of the central sulcus. This area is responsible for executing voluntary movements. The Central Sulcus is a prominent landmark of the brain. It separates the parietal lobe from the frontal lobe and the primary motor cortex from the somatosensory cortex.

Primary Somatosensory Cortex (yellow-green area): Positioned just behind the central sulcus in the parietal lobe, it processes sensory information from the body.

Broca's area is usually found in the left hemisphere, and this region is involved in speech production and language processing. The Wernicke area is typically located in the left temporal lobe and is essential for language comprehension.

The primary visual cortex is located in the occipital lobe at the back of the brain, where it receives and processes visual information. The Sylvian fissure separates the temporal lobe from the frontal and parietal lobes.

3.3 EEG

EEG stands for electroencephalogram, which collects brainwaves from the brain using small metal discs or electrodes attached to patients' scalp [31]. Electroencephalography (EEG) was discovered in 1929 by the German psychiatrist Hans Berger and was a historical breakthrough providing a new neurologic and psychiatric diagnostic tool at the time [77]. Most of the time, the examination of patients is not noninvasive [50]. It is a standard for diagnosing patients with dementia, head injuries, concussions, and sleep disorders.

Electroencephalography (EEG) has become widely used in recording brain activity because it is an invasive and affordable method. One area of focus within EEG-based Brain-Computer Interfaces (BCI) is motor imagery (MI), which involves generating responses through the imagination of movements, such as moving the left or right hand. Understanding these activities during MI allows people with motor neurone disorders (such as stroke or Parkinson's disease) to regain some motor skills with the help of devices. In addition to rehabilitation, EEG-based MI also finds applications in controlling wheelchair operating arms and controlling quadcopters [84].

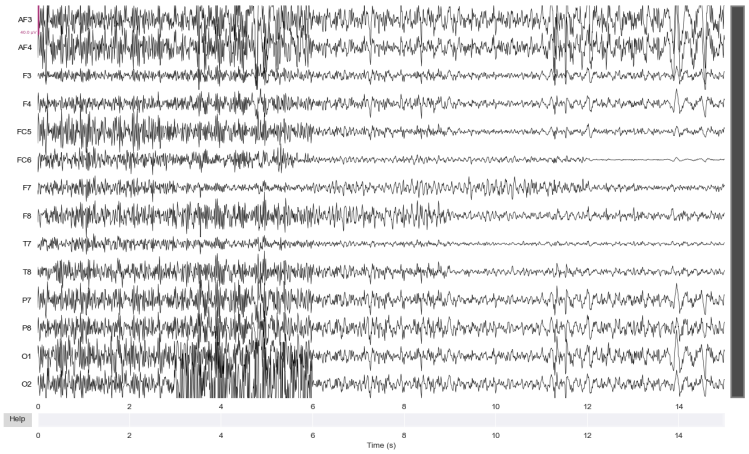

Figure 3.3: Example of EEG graph of subject 4 from Dasps[16] dataset.

How it works

The electroencephalogram is a record of oscillations of brain electric potentials acquired from electrodes on the human scalp [49]. We can measure this electric potential by attaching a voltage meter to any two points on the ionic surface. There are two types of EEG devices: those that utilise gel for electrode conductivity and those that do not require gel application.

3.4 Types of EEG waves

EEG-based signals are nonstationary. The nonstationary signals have properties that **variance** and **mean change with time**. EEG waves can be separated into five leading bands: δ delta (0.5-4 Hz), θ theta (4-8Hz), α alpha (8-12 Hz),β beta (12-30 Hz), γ gamma ($>$ 30 Hz). We can see that picture in 3.5. Different literature shows ranges like (8-13 Hz) for alpha.

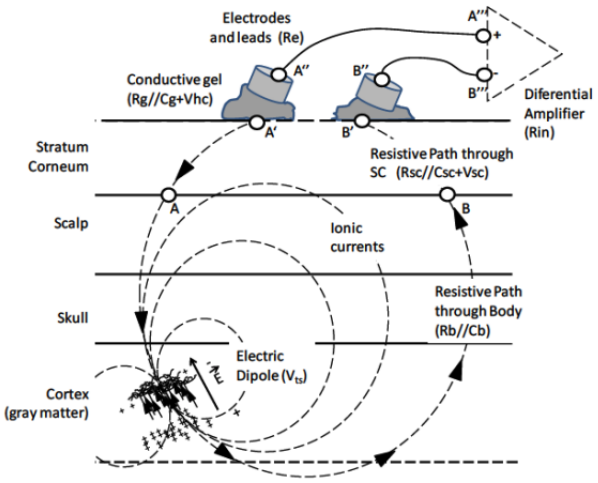

Figure 3.4: Measuring electronic potential schema from [49].

Delta Band

Delta waves are in the range of 0.5 Hz -4 Hz. They are generally seen when the subject is sleeping. In addition, they are seen in subjects with brain injuries, coma, and brain trauma [50]. A possible explanation for why they are seen under these conditions is the slow neural activity of the subjects.

Theta band

Theta waves are in the range of 4Hz to 8Hz. Like delta waves, Theta waves are non-sinusoidal. They are generally seen in subjects with a lack of presence or attention deficit disorders.

Alpha band

Alpha waves occur in awake and relaxed subjects but have closed eyes [8]. They are at a maximum in the occipital area [50]. The measurement of alpha waves is critical for classifying an anxiety disorder. If they are measured in large amounts of diffuse, this would indicate an anxiety disorder. Measurements like coherence, synchrony, and asymmetry can also be used as biofeedback for an utterly normal subject [28].

Beta band

Beta waves are commonly associated with alertness, thinking, and maintaining focus. They usually fall within about 14 to 30 cycles per second (referred to as Hertz or Hz). Beta waves are often the predominant feature in EEG recordings during tasks such as problem-solving exercises or mental concentration periods.

Gamma band

Gamma waves are the fastest group in the EEG frequency bands. Thanks to the frequency, Gamma waves are smaller than other bands. They are believed to be involved in the integration of sensor input information, such as visual or audio stimulus [18].

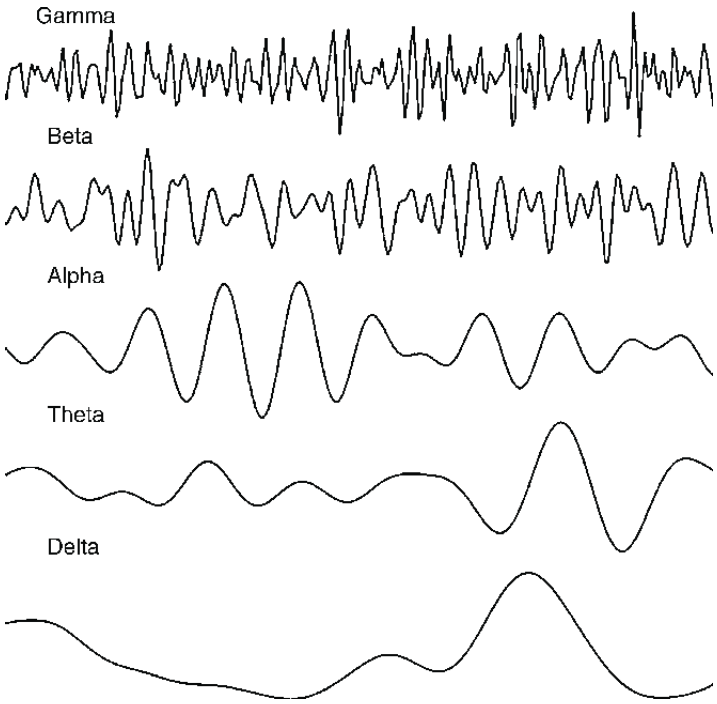

Figure 3.5: EEG bands [58]

EEG artifacts

EEG artifacts are signals found in EEG recordings that do not come from brain activity itself but from other sources. They are divided into physiological artifacts and external artifacts. These unwavering signals can significantly impact EEG analysis and the accuracy of machine learning models.

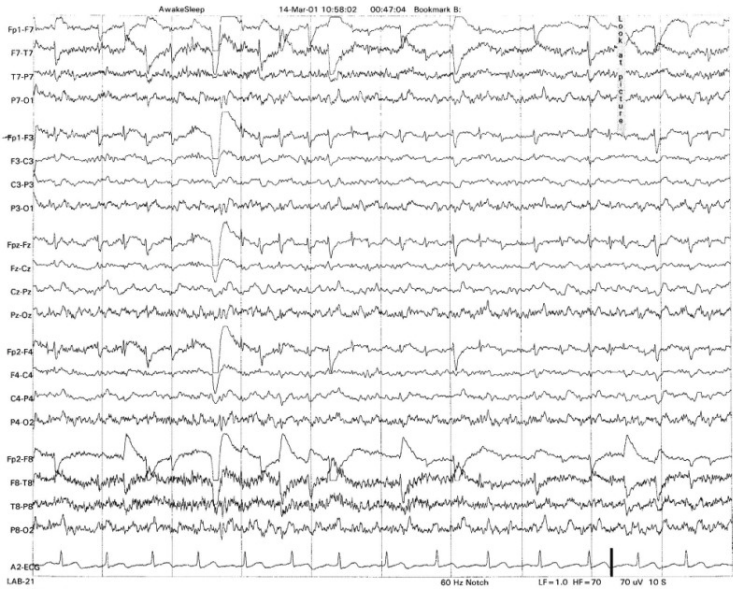

Figure 3.6: Rapid eye movement artifacts [52].

Physiological artifacts in EEG include:

- **Ocular artifacts**, which are generated by eye movements and blinking
- **Muscle artifacts**, made by movement of muscles.
- **Cardiac artifacts**, made by movement of heart. Heartbeats create a small electrical signal that EEG electrodes can pick.

External artifacts are:

- **Electrical Interference**: Electrical devices can cause an interference with EEG recording devices.
- **Electrode Popping**: Bad contact with the skull.
- **Movement artifacts**: Walking and moving the body can disturb the electrode placement.

EEG montage

An EEG montage, or an electroencephalogram montage, refers to how the electrodes are placed on a patient's scalp during an EEG test to record and interpret brain activity accurately. Different montages are used for this purpose. The International 10 to 20 system is the EEG montage. It ensures placement across different patients by using specific locations based on percentages of distances between anatomical landmarks on the scalp (nasion, inion and preauricular points). This system provides a view of brain activity.

The 10 to 10 system is an extension of the 10 to 20 system. It offers electrodes for more detailed spatial coverage and is commonly used in cases.

A bipolar montage represents the difference in potential between two electrodes in each channel. This setup helps identify how activity propagates in seizure disorders.

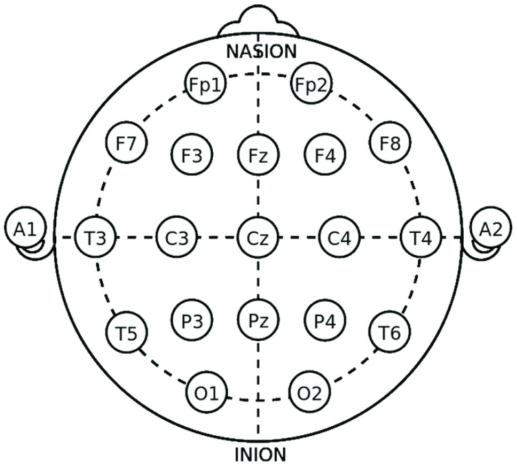

Figure 3.7: International system of EEG electrode placement from [62]

3.5 Signal analysis and EEG features

Analysis of signal can be complex, and EEG recording is even more difficult because instead of having noise interference from a microphone, for example, analysis of speakers, we have all-new categories of physiological artefacts. So, the analysis of EEG signals originates from the analysis of the signal itself. However, there are specific techniques that are applied only to EEG signal analysis (such as Event-Related Potential Analysis (ERP) [26]). In this section, we will discuss some of the most critical features, which also resulted from our genetic algorithm.

EEG-recorded data contains time information (such as voltage amplitude) and frequency information. Therefore, we can divide the EEG analysis into three domains: **Time Domain Analysis**, **Frequency Domain Analysis** and combined **Time-Frequency Analysis**.

Time Domain Analysis

Time Domain Analysis analyses signals and how they change and vary over time. Statistical features such as mean, median, variance, standard deviation, skewness, kurtosis, and similar are also used in the frequency domain [72].

Mean

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.1}$$

The mean value of the EEG signal reflects the activity observed throughout the recording session.

Kurtosis

Kurtosis measures the relative flatness of a distribution. Distributions with positive kurtosis are termed „super-Gaussian“, while those with negative kurtosis are „sub-Gaussian“ [80].

$$\text{Kurtosis} = \frac{N(N+1)}{(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{s}\right)^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{3.2}$$

- N: This represents the dataset's number of observations or samples. In the context of an EEG signal, it would be the number of signal measurements or data points you have collected.
- x_i: This variable represents the EEG signal's individual observations or data points.
- $\overline{x}$: is mean of these observations.

Frequency Domain Analysis

Frequency domain analysis examines the signal's frequency content. Here are relevant features which came from the genetic algorithm for the Support Vector Machine.

Spectral Slope

Spectral Slope can help researchers and clinicians identify specific frequency ranges that are more or less active in the brain, which can indicate various neurological conditions or states, such as epilepsy or sleep stages [41]. It was also used to detect multiple levels of anxiety in this article [47].

$$\text{Spectral Slope} = \frac{\log(P_f) - \log(P_i)}{\log(f_f) - \log(f_i)} \tag{3.3}$$

- P_f is Power spectral density at f_f
- P_i is Power spectral density at f_i
- f_f represents the final frequency point.
- f_i represents the initial frequency point.

Hjorth Mobility

Hjorth Mobility represents the average rate of change of the EEG signal, giving an idea of the signal's frequency content without requiring a Fourier transform [42].

$$\text{Mobility} = \sqrt{\frac{\text{Var}\left(\frac{dS(n)}{dn}\right)}{\text{Var}(S(n))}} \tag{3.4}$$

- S: is signal
- Var: is Variance
- $dS(n)$: is its first derivative

Hurst Exponent

The Hurst exponent estimates the degree of self-similarity and predictability of a time series, which, under this non-linear statistical model, can adopt two opposing tendencies regarding how these data series are mobilised over time [27].

$$H = \frac{\log(\frac{R(n)}{S(n)})}{\log(n)} \tag{3.5}$$

- $R(n)$: is the range of the cumulative deviations from the series' mean, rescaled by the standard deviation.
- $S(n)$: is standard deviation of the time series.

- n: is the number of data points in the time series

Time-Frequency Analysis

The Teager-Kaiser Energy Operator

The Teager-Kaiser Energy Operator is used to compute a signal's energy and is a popular feature for Machine Learning methods [19].

$$\Psi(x[n]) = x[n]^2 - x[n+1]x[n-1] \tag{3.6}$$

- $\Psi(x[n])$: represents the energy of the signal at sample.
- $x[n]$: is the signal value at the current sample.
- $x[n-1]$ and $x[n+1]$ are neighbour signals.

The method operates by squaring the signal value and deducting the result of the neighbouring signal values. This approach highlights the strength of elements that vary rapidly, which proves especially effective for signals displaying linear and non-stationary traits, such as EEG signals.

SVD Entropy

The Singular Value Decomposition is a technique to analyse the signals matrix, including EEG signals [32]. The SVD algorithm treats EEG as a linear combination of multiple features, each characterised by its spatial and temporal distribution and amplitude. This approach facilitates transparent methodologies for sampling, data reduction, normalisation, and statistical significance calculation, advantages that are less apparent when analysis is confined to a single domain.

$$H = -\sum_{i=1}^{r} p_i \log(p_i) \tag{3.7}$$

- p_i: is the normalized singular value, calculated as $p_i = \frac{\sigma_i}{\sum_{j=1}^{r} \sigma_j}$.
- σ_i: are the singular values obtained from the SVD of the EEG signal matrix.
- r is a rank of the EEG matrix.

Phase Locking Value

Phase Locking Value is widely used for EEG correlation between electrodes in EEG signals. These characteristic spatial and temporal shifts in synchronisation are strongly related to emotional activity [23].

$$PLV = \left| \frac{1}{N} \sum_{n=1}^{N} e^{i\theta_n} \right| \tag{3.8}$$

- N is the number of samples where PLV is calculated
- θ_n represents the phase difference between two EEG signals at n.
- $e^{i\theta_n}$ is complex exponetial using imaginary i.

Chapter 4

Machine learning techniques for anxiety and depression

In this chapter, we will explain the current status of how Social anxiety disorder and Major depressive disorder are classified. As well we will explain methods used for algorithm design discussed in chapter 7. Generally, as a machine learning method for EEG classifying, the most popular methods are SVMs, CNN with LSTM layers, and Random Forest. They have results as shown. However, Q learning methods are lacking. In this study[54], machine learning (ML) methods often face scalability challenges due to the reliance on crafted features. However, deep learning (DL) approaches although lack interpretability. Genetic algorithms (GAs) have emerged as a promising alternative to address this issue. GAs are gaining attention for their stochastic search capabilities in scenarios, which make them capable of identifying suitable solutions from a wide range of possibilities. In today's era of data growth across activities and advancements in data processing and storage capabilities, there is a growing need to select an optimal subset of features from large datasets for research purposes[54]. In this regard, GAs have shown promise in the literature for their ability to identify nearly optimal features by removing unnecessary or redundant elements from high-dimensional feature spaces.

4.1 Support Vector Machine

The Support Vector Machine is a linear classifier. It is famous for classification in the EEG domain [60]. Unlike CNN, it can not produce a probability of classes for specific inputs. However, probabilities can be achieved by making a soft score for a particular class and converting it using a Softmax function. The objective function is closer to the recognised maximum count of actual positive classes than to maximising the probability that everything is classified correct[17].

The fundamental idea of how SVM works is to find the best separation hyperplane that separates datapoints of different classes. The criterion for finding the best hyperplane is to maximise the margin between datapoint and hyperplane.

Not every problem is linearly separable, especially in EEG features domain. To solve it, SVM uses a method called **Kernel trick** [17]. The choice of kernel function, such as linear kernels, polynomial kernels, radial basis function (RBF) kernels, or sigmoid kernels, depends on data characteristics and the specific use case [5].

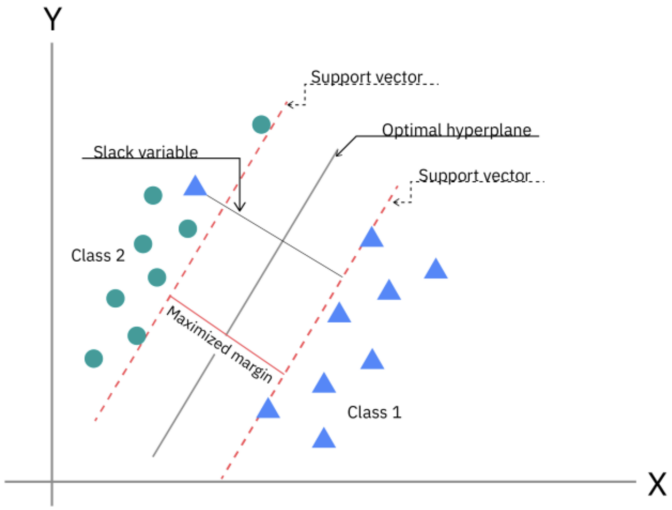

Figure 4.1: Visualition of SVM from [5].

In the default version, SVM can handle only binary data, but nowadays, it is used with a one-vs-one scheme. One-vs-one splits a multiclass classification dataset into binary classification problems. It splits the dataset into one dataset for each class versus every other class. For example, three classes: Normal, SAD and MDD will produce three binary classification problems.

- Normal vs SAD.
- Normal vs MDD.
- SAD vs MDD.

4.2 Convolutonal Neural Networks

Convolutional neural networks revolutionised image classifications, especially AlexNet in 2012 [44]. Convolution as a mathematical operation is a vital part of convolutional neural networks.

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m]g[n-m] \tag{4.1}$$

CNNs are nowadays a standard part of Deep Learning classification algorithms. CNN are usually a forward network, but nowadays, there are also acyclic implementations. CNNs consist of layers, which can be divided by type into Convolutional layers, Activation layers, Pooling layers and Fully connected layers.

Convolutional layers have a learning parameter called kernel, also called a filter. The kernel is a learnable parameter of the convolutional layer. The equation for this filter is simple and defined in 4.2.

$$I(u,v) = \sum_{i=-a}^{b} \sum_{j=-c}^{d} h(i,j)I(u+i, v+j) \tag{4.2}$$

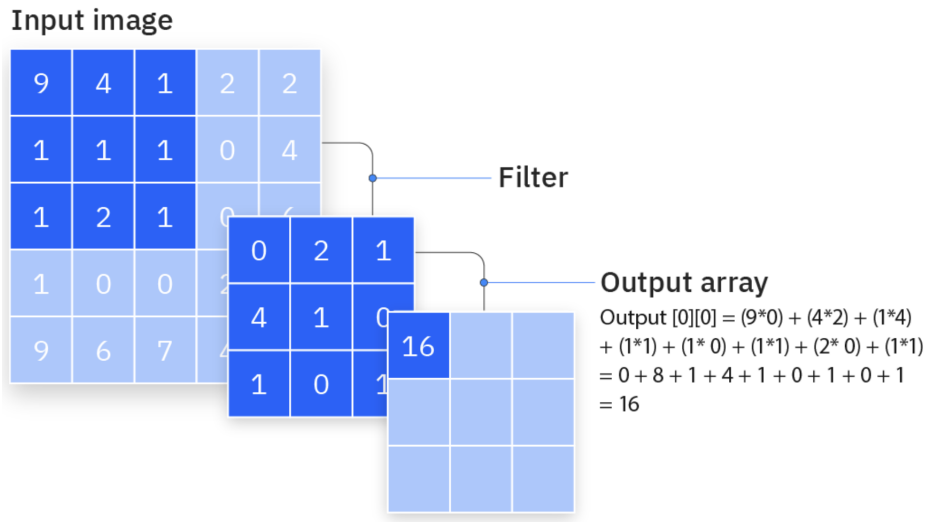

Figure 4.2: Convolutional operations in CNNs from[6].

Padding and stride are also parameters that affect mathematical operation. These parameters can significantly affect the network's performance and efficiency.

Padding is a technique of adding zero values around the corner of the matrix. This is typically done before the convolutional operation. The purpose of the padding is size preservation and border information. After each convolution, the size of the result matrix is reduced. The padding can preserve the size of the matrix if that is a desired goal.

The stride is defined as a movement of the kernel in the matrix. Stride one means that the filter is moving by one pixel. If we want to down-sampling the input, we just increase a stride number.

Activation layers introduce non-linear properties to the model, which allows them to learn more complex patterns in the data. There is no manual, which network should use which function [70]. The activation functions can be Binary Step Function, Linear, Sigmoid, Tanh, ReLU, Leaky ReLU, Parametrized ReLU, Exponential Linear Unit, Swish and SoftMax. Some of them are in 4.3.

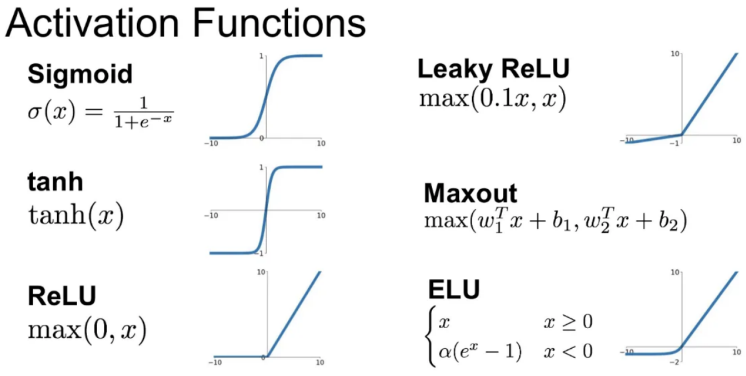

Figure 4.3: Types of activation functions from [39].

The pooling layer is created using a pooling operator to gather information from each small area of the input feature channels and then down-sample the aggregated results.

Although hand-crafted pooling operations are commonly used for this aggregation, they do not necessarily minimise the training error. [75]. The usual pooling operator is Max , another possible operators are average or min pooling.

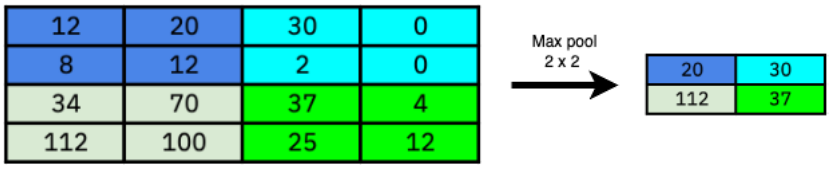

Figure 4.4: Max pool operation.

4.3 Recurrent Neural Networks

Recurrent neural networks are types of neural networks which are designed to analyse sequential data [66]. They are typically used in natural language processing and speech recognition. They are distinguished from CNNs by their "memory" as they take information from prior inputs to influence the current input and output. [7].

The first Recurrent neural network was Jordan's sequential network in 1986 [69], which introduced the cycle in the neural networks graphs. Backpropagation Through Time (BPTT) adapts the backpropagation algorithm for RNNs [65]. We can calculate the loss by:

$$\mathcal{L}(\mathbf{O}, \mathbf{Y}) = \sum_{t=1}^{T} \ell_t(\mathbf{O}_t, \mathbf{Y}_t) \tag{4.3}$$

The loss $\mathcal{L}(\mathbf{O}, \mathbf{Y})$ is the sum of all losses in T where ℓ_t can be specified loss function depending on the problem. Like CrossEntropy or MeanSquareError loss.

RNNs have one fundamental problem: vanishing and exploding gradients. This problem motivated the introduction of the long short-term memory units (LSTMs) to mainly handle the vanishing gradient problem [65].

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{hh}} = \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \mathbf{O}_t} \cdot \frac{\partial \mathbf{O}_t}{\partial \phi_o} \cdot \mathbf{W}_{ho} \sum_{k=1}^{t} \frac{\partial \mathbf{H}_t}{\partial \mathbf{H}_k} \cdot \frac{\partial \mathbf{H}_k}{\partial \mathbf{W}_{hh}} \tag{4.4}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{xh}} = \sum_{t=1}^{T} \frac{\partial \ell_t}{\partial \mathbf{O}_t} \cdot \frac{\partial \mathbf{O}_t}{\partial \phi_o} \cdot \mathbf{W}_{ho} \sum_{k=1}^{t} \frac{\partial \mathbf{H}_t}{\partial \mathbf{H}_k} \cdot \frac{\partial \mathbf{H}_k}{\partial \mathbf{W}_{xh}} \tag{4.5}$$

In equations for weight derivation in 4.4 and 4.5 they introduce a $\frac{\partial \mathbf{H}_k}{\partial \mathbf{W}_{xh}}$ and $\frac{\partial \mathbf{H}_k}{\partial \mathbf{W}_{hh}}$. These matrix multiplications can be done in a very long sequence. If they are minimal values (less than one) in matrix multiplication, this causes the gradient to decrease with each layer (or time step) and finally vanish [65]. This work vice versa, if the values are tremendous, the gradient can explode [65].

4.4 Transformers

Transformers have become a trendy AI architecture. They replaced all LSTM-RNN model architectures in NLP, thanks to better performance [48]. Also, this breakthrough is due

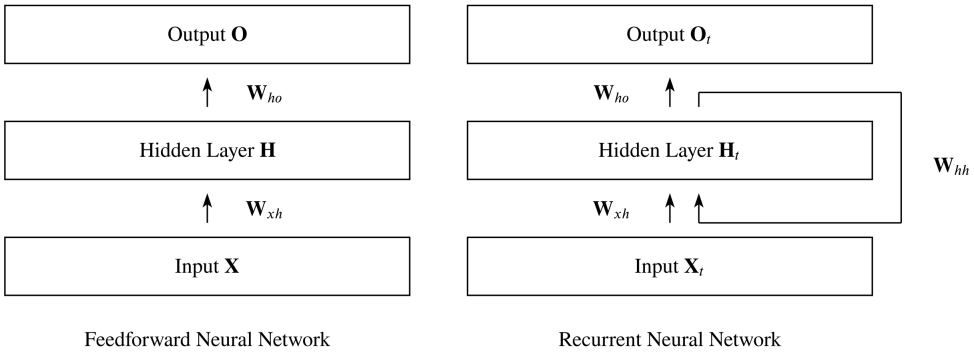

Figure 4.5: Differences between Feedfoward NNs und Recurrent NNs from [65].

to the popularised AI chatbot for generating text like ChaptGPT4 [10], which is based on transformer architecture.

Transformers introduce a new concept, which is called self-attention. Self-attention is a mechanism which allows transformers to process input data (like text) by enabling each part of the input to interact with and weigh the importance of other parts, all within a single representation. In practical terms, we calculate the attention function for a group of queries simultaneously. The result is then packed into a matrix Q. The keys and values are also packed into matrices K and V [79]. Matrix of outputs (or we can say score) is then:

$$\mathrm{Q} = W^Q x \quad \mathrm{K} = W^K x \quad \mathrm{V} = W^V x \tag{4.6}$$

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{4.7}$$

The input is converted to embeddings, vector representations of text or other data. After converting the input to input embeddings, the positional encodings are applied. An attention function can be described as mapping a query and a set of key-value pairs to an output, where the query, keys, values, and production are all vectors [79].

Transformers are also auto-encoder architectures. In the original implementation, the encoder consists of N identical layers. Every layer consists of Multi-Head Attention and FeedForward Neural Network. In the Decoder, there is another new sub-layer type addition which is called Masked Multi-Head Attention.

The final linear layer and softmax operation transforms the float vector produced by the decoder stack into a word. In our implementation, we do not use Transformers as is stated in this version, but we use architecture Bert [24], which is based on the transformer.

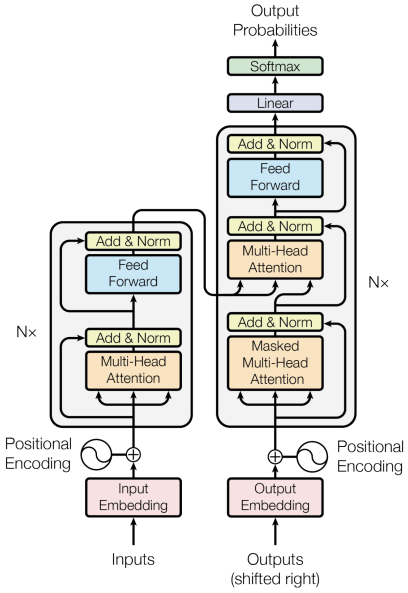

Figure 4.6: Transformer model architecture from [79].

4.5 Bert Transformer

BERT, or Bidirectional Encoder Representations from Transformers, is a transformer-based machine learning model introduced by Google researchers in 2018 [24]. It's designed to deeply understand the nuances of language by processing words in the context of all the other words in a sentence—rather than one direction at a time. This bidirectional approach allows BERT to capture the full context of a sentence, making it robust for tasks like language understanding and contextual word relationships [24].

The architecture is based on Transformer, which we discussed in the previous section. BERT uses WordPiece tokenisation, breaking down words into smaller subwords or characters. This allows BERT to handle out-of-vocabulary words and capture finer linguistic details. In our implementation of Transformer, we do not use tokenisation but put our custom embeddings into Bert.

BERT consists of multiple layers of Transformers. The number of layers can vary depending on the specific variant of BERT (e.g., BERT-base, BERT-large). Each layer refines the representation of the input text by processing it through self-attention mechanisms and feed-forward neural networks.

BERT is built using several layers of Transformers, with the exact number depending on the variant, such as BERT-base or BERT-large. The input text is refined through self-attention mechanisms and feed-forward neural networks to enhance its representation in each layer. The critical difference between original Transformer from [79], is using only the encoder part. Another key difference is that BERT, for the representation of each word, is informed by the entire sentence context (bidirectional). In contrast, in the original transformer, the decoder processes words in a left-to-right sequence when generating text.

Layer normalisation is used in BERT to help stabilise the learning process. It normalises the inputs across the features instead of the batch dimension:

$$\text{LayerNorm}(x) = \gamma \left(\frac{x - \mu}{\sigma} \right) + \beta \tag{4.8}$$

- μ and σ where are the mean and standard deviation of the features, and
- γ and β are learnable parameters, for normalisation of the data

4.6 Genetic algorithms

The genetic algorithm is a metaheuristic inspired by the natural selection process. Genetic algorithms are commonly used to generate high-quality solutions to optimisation and search problems by relying on bioinspired operators such as mutation, crossover, and selection [82]. Genetic algorithms are computational models inspired by Darwin's theory of evolution[64].

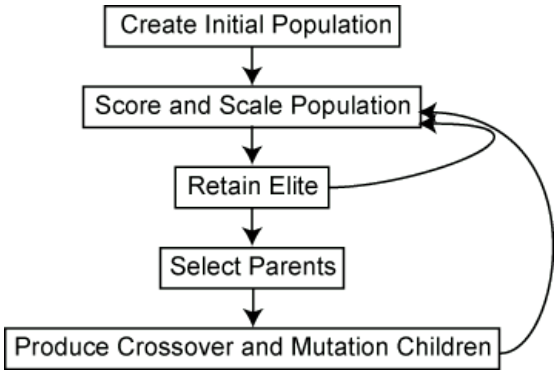

Figure 4.7: Concept behind GA algorithms [1].

The genetics algorithms are a type of evolution algorithm. Evolution algorithms include a wide variety of algorithms. Evolution algorithms are classified as state space search algorithms. To find a solution, they used a population of candidates. Using just a parallelism, getting a quality of genetics algorithm is impossible. The Differential Evolution (DE) algorithm is an evolution-based algorithm proposed by Storn [73] in 1996. This algorithm implements mutation, crossover, and selection as operators in its structure. In this thesis, the Differential Evolution algorithm is used. Other popular algorithms are Particle Swarm Optimization [43] and Ant Colony Optimization (ACO) [29]. The PSO and ACO were not applied in this thesis, becauseour problem's characteristic isn'te easily interpreted and coded in these two algorithms.

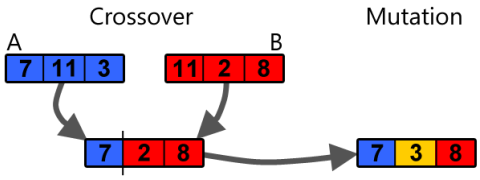

Figure 4.8: GA operations on binary strings.

4.7 AI trends in EEG

The paper [14] provides a comprehensive review of machine learning and deep learning methods for diagnosing depression. The study categorised the models into classification and deep learning. The models they inspected were state-of-the-art [14]. Among the survey,

the Support Vector Machine was the most used classifier for EEG anxiety depression. SVM are also very resistant to overfitting in the State-of-the-art models [14]. The accuracies of SVM classifiers are mostly greater than 75%. The paper doesn't mention using EEG-based Transformers for classifying depression.

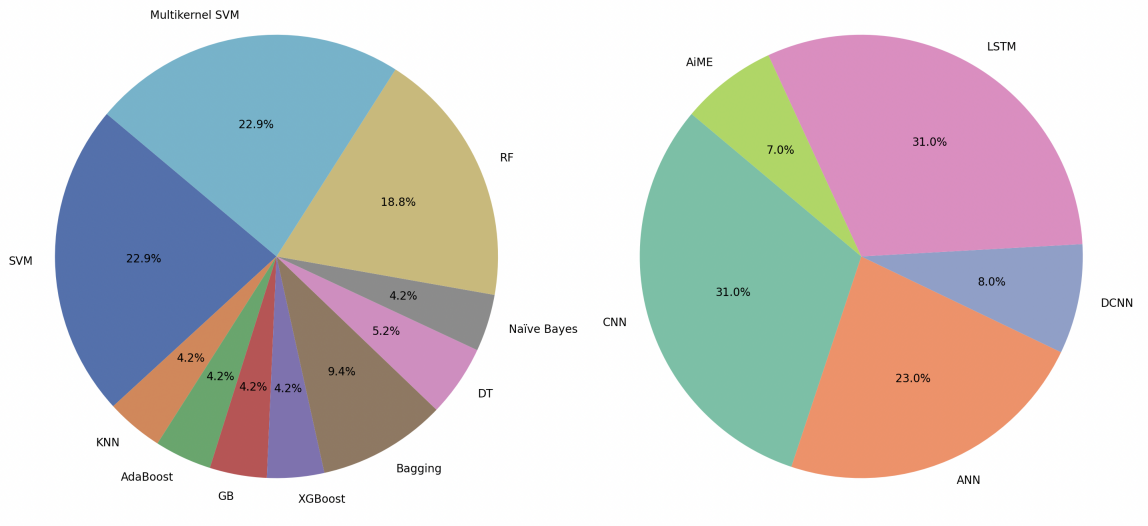

Figure 4.9: Trends in EEG Machine learning vs Deep Learning from [14].

However, Transformers are used in EEG classification of fatigue or attention, specifically model named *EEG-deformer* from [25]. They introduce an architecture where the input, instead of a tokeniser of the word for the transformer, applied a CNN layer as an input for this transformer. Also, instead of using EEG features, they used EEG recordings, which are then used in CNN. Problems with attention or fatigue can be a symptom of some mental disorders. This paper suggests that models based on Transformers can be successful with classifications of Mental Disorders when fatigue and attention were accurate with the accuracy of **82.72 %**, **79.32 %** and **73.18 %**.

Genetics algorithms are also used in EEG classification. This paper [55] uses a non-dominated sorting genetic (NSGA-II) evolutionary algorithm to find the best features and best frequencies. They achieved on MDD dataset 93.3%, specificity of 93.4% and accuracy of **93.5%**. The key idea was to find specific frequencies for classification by selecting channels in evolution.

4.8 Limitations

The EEG classifications tasks are overall hard due to numerous factors. Artifacts in EEG recordings cause noise, which is not easily eliminated as in audio task. Another problem is that, EEG devices are not widely popular as smartphones or smartwatches, usually their cost with 14 channels and more is around 1000$. Also, in EEG classification, we need to instead of dividing the data to seen and unseen to seen subjects and unseen subjects to eliminate a bias. Lastly one more factor is critical and that is, datasets for anxiety disorders as well as Major Depressive Disorder are not that many publicly available. With factors above, is hard to examine a lot of people diagnosed with social anxiety disorders or Major Depressive Disorders.

Chapter 5

Datasets

5.1 DASPS dataset

It's a databases of subjects with levels of anxiety disorders based on stimuli response. The levels of anxiety is Normal, Light, Mild and Severe. To obtained label data, the technique Psychological Stimulation [15] was used. The protocol of this dataset was discussed with psychotherapists with use of exposure therapy. The number of subjects is 23 participants. For every subject, 6 situations are made before and after stimulation. In total, there are 276 samples annotated by the hamilton1 score and the hamilton2 score. The paper to produce the results does not mention which labelled part was used for classifying (Hamilton1 or Hamilton2).

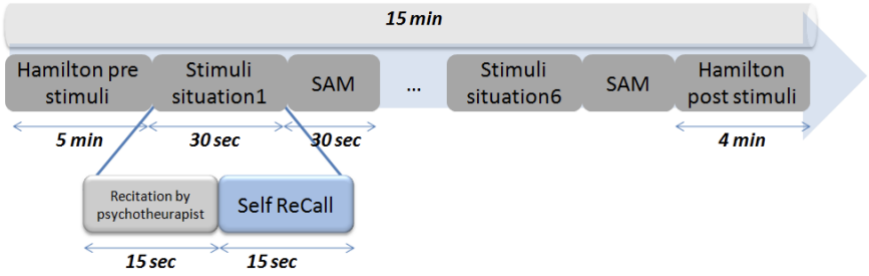

Figure 5.1: Methodology from [15]

The therapist requests that the participant assess the intensity of each symptom using a four-point scale to indicate its severity. This data is then utilised to calculate a score reflecting the anxiety severity level experienced by the individual [15]. After that, the subject is prepared to begin the experiment with eyes closed to minimise muscle movement so that there are fewer artefacts in the EEG graph. The psychotherapist starts the stimulation by talking to the subject to imagine it. This phase is divided into two stages, before and after talking, where there is a recall.

The number of EEG channels is 14 with the names AF3, AF4, F3, F4, FC5, FC6, F7, F8, T7, T8, P7, P8, O1, O2. The SPS (Samples per second) is 128.

For the DASPS data set and generally for medical data, a lot of time when there are not enough data, kernel machine learning methods like GMM, SVM K-NN are more suitable solution than Neural Networks. Most of the time the problem of not enough data is solved

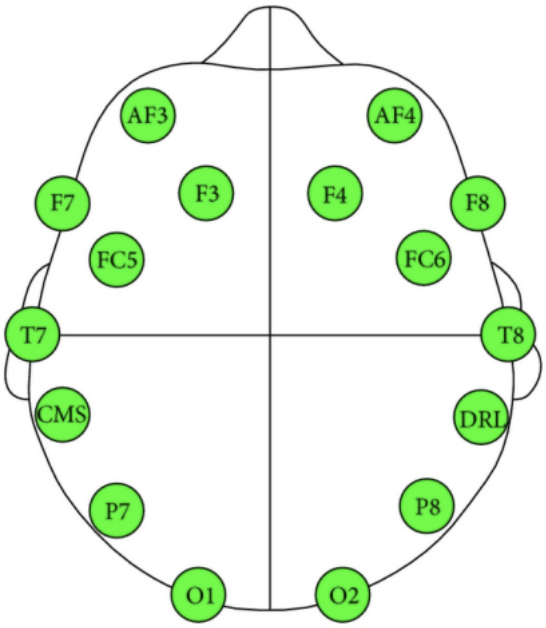

Figure 5.2: Emotiv EPOC electrodes placement [15].

by data augmentation or data generator. We can see in tables 5.1 and 5.2 that the shorter the duration of the trial (more data is split and more trains) the better the results are given. It is also because of how the features are computed differently, but the idea is the same. The best results are given for the binary classification of k closest neighbours **81.40%** accuraccy.

Table 5.1: 4 LEVELS [15]

Trial	Feature	#Feat	SVM	k-NN
15s	Hjorth	42	56.20	56.50
	qEEG	25	56.50	56.50
	HHT	10	57.00	56.80
	Power	56	58.30	57.60
	RMS	56	59.10	56.50
5s	Hjorth	42	57.40	58.80
	qEEG	25	56.80	56.40
	HHT	9	56.90	56.30
	Power	56	62.00	63.20
	RMS	56	65.30	64.30
1s	Hjorth	42	60.10	57.00
	qEEG	25	58.30	56.40
	HHT	7	56.60	56.30
	Power	56	64.40	68.00
	RMS	56	70.20	**73.60**

Table 5.2: 2 LEVELS [15]

Trial	Feature	#Feat	SVM	k-NN
15s	Hjorth	42	66.30	63.80
	qEEG	25	64.10	63.80
	HHT	10	64.10	64.10
	Power	56	66.30	66.30
	RMS	56	66.30	67.00
5s	Hjorth	42	72.90	64.90
	qEEG	25	64.00	63.60
	HHT	9	64.00	64.10
	Power	56	73.10	70.50
	RMS	56	72.90	73.40
1s	Hjorth	42	67.40	**81.40**
	qEEG	25	64.00	63.50
	HHT	7	64.00	63.60
	Power	56	76.00	74.90
	RMS	56	77.40	80.30

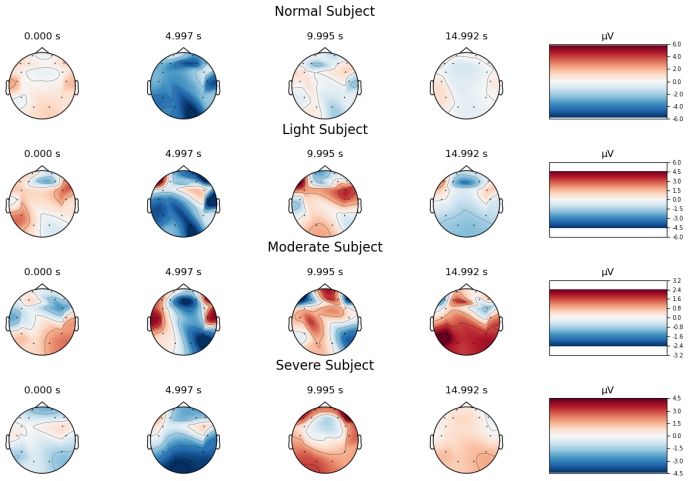

Figure 5.3: Topomaps of subjects during the time in DASPS.

5.2 MDD Patients and Healthy Controls EEG dataset

In this study [56] two groups of subject were involved. 33 patients with Major Depressive Disorder with age $\overline{x} = 40.33$, and 30 healthy subjects with age with $\overline{x} = 38.227$. Participants were recruited from the outpatient clinic of hospital Universiti Sains Malaysia (HUSM), Malaysia [56]. The subjects in MDD group were diagnosed with MDD by senior psychiatrists in the psychiatric clinic. Subjects with psychotic symptoms, pregnant patients, alcoholics, smokers and patients with epilepsy were excluded. This was necessary to exclude any other bias in the MDD group. The subjects were recorded with 19-channel BCI with channels Fp1, Fp2, F3, F4, F7, F8, Fpz, T3, T4, T5, T6, P3, P4, P7, P8, O1, O2, C3, C4 with a sample rate of 256 samples per second.

In the paper [56], the features were ranked according to criterion i.e., receiver operating characteristics (ROC). They iteratively ran the simulation of the top 5,10,15 and 19 features based on z-score values to obtain the best features. In result, they switched the SVM classifier for Linear Regression classifier with accuracy **98.33%** achieved. One of the key problem of this dataset is that the labels are binary, which contributes to better results in the end.

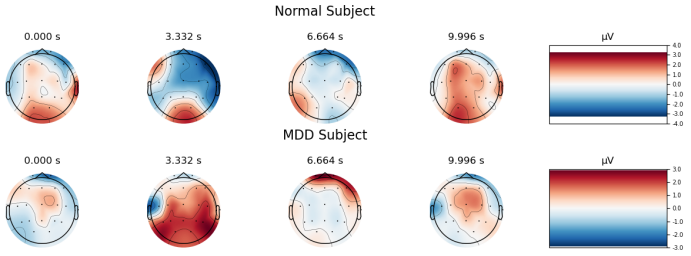

Figure 5.4: Topomaps of subjects during the time in MDD.

5.3 Social Anxiety Disorder dataset

The last dataset used in this thesis is the Social Anxiety Disorder dataset from paper [11]. It consists of eighty-eight individuals from 502 group, who self filled the Social Interaction Anxiety Scale (SIAS) questionnaire[1]. The participants were from the range of 18 to 25, and it consisted of 36 with age mean $\overline{x} = 21.97$ females and 52 males with mean $\overline{x} = 22.73$. BCI recording device name is not mention but it was recorded in 2048 Hz and reduced to 256 Hz. The channels name are FP1, FP2, FPz, F7, F8, F3, F4, FC5, FC1, FC2, FC6, Fz, Cz, T8, P7, P8, C3, C4, C3, CP2, CP4, CP1, CP6, CP5, P3, P4, Pz, O1, O2, and Poz. They were located on cerebral cortex with referred to CPz [11]. The raw EEG data were **already processed** to remove unnecessary and noisy segments. Artifacts caused by eye movements, breathing, power interference, and cardiac movements were visually inspected, removed, or corrected using spatial filters.

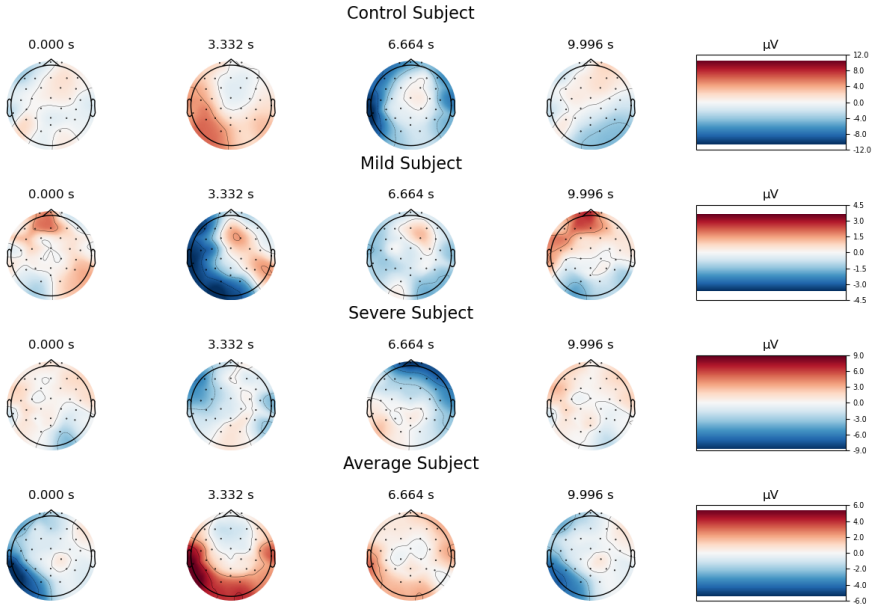

Figure 5.5: Topomaps of subjects during the time in SAD.

Artifacts resulting from eye movements, respiration, power interference, and cardiac activity were visually examined and removed or corrected using spatial filters. To improve signal quality and remove high-frequency artifacts, signal noise, and low-frequency deflections, a bandpass filter was used to isolate segments with optimal signal quality between 0.4 and 40 Hz. The montage is 10 to 20 system. In addition, for cleaning, the independent component analysis (ICA) negatively affects the PDC estimation [11]. In this thesis, the ICA was not applied in the cleaning process to this dataset.

[1] https://en.wikipedia.org/wiki/Social_Interaction_Anxiety_Scale

Chapter 6

Preprocessing techniques

6.1 Filters

Finite impulse response filter (FIR) and infinite impulse response filter (IIR) are essential in signal processing [20]. The filters are mainly used to noise removal in Audio processing. These techniques are also applied to EEG signal processing[1]. In frequency domain, the filter is defined:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_M z^{-M}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N}} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 + \sum_{k=1}^{N} a_k z^{-k}} \tag{6.1}$$

In time domain , the filter is defined:

$$y(n) = b_0 x(n) + \ldots + b_M x(n-M) - a_1 y(n-1) - \ldots - a_N y(n-N) \tag{6.2}$$

$$y(n) = \sum_{k=0}^{M} b_k x(n-k) - \sum_{k=1}^{N} a_k y(n-k) \tag{6.3}$$

FIR (Finite Impulse Response) and IIR (Infinite Impulse Response) filters each offer unique advantages and drawbacks as proposed in [20]. FIR filters can be designed to have linear phase, meaning they provide consistent time delays across all frequencies, a feature not achievable with causal IIR filters. Typically, FIR filters also exhibit superior phase and group delay characteristics. On the other hand, IIR filters can achieve sharper cutoffs compared to FIR filters of the same order, making them effective for applications requiring strict frequency separation with fewer calculations. However, IIR filters tend to be less numerically stable than FIR filters. This instability often results from the recursive calculations they perform, which can accumulate errors over time. In EEG preprocessing the FIR filters with additional implementation FIR windowed (FIRWIN) filters are used.

[1] https://mne.tools/stable/auto_tutorials/preprocessing/25_background_filtering.html#disc-filtering

6.2 Independent Component Analysis

Independent Component Analysis is a statistical signal-processing method. The main objective of Independent Component Analysis (ICA) is to reduce the statistical dependency among the elements of a random vector, effectively eliminating additive background noise and distinguishing between blended signals.[13]. Criterion which has the following multivariate probability density function of **u** factorizes given as:

$$f_u(u) = \prod_{i=1}^{N} f_{u_i}(u_i) \tag{6.4}$$

In the context of EEG analysis, ICA is applied to solve the problem of signal preprocessing, source localization, and cancellation of artifacts. In practice, real EEG signals either include super-Gaussian's noise, especially for ERPs, or sub-Gaussian background such as EOG or other power frequency [74].

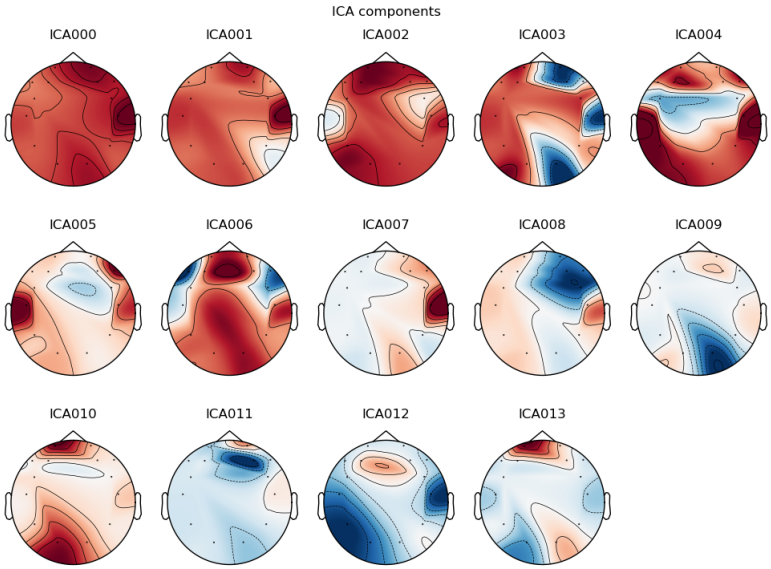

Figure 6.1: ICA Compoments of subject S02 in DASPS using MNE library[2].

6.3 Normalization

The Normalization is statistical technique to adjust or scale the data. In this thesis, normalization is used to eliminate differences between DASPS, SAD and MDD datasets. This is crucial to eliminate different biases from BCI devices, which were different in all three dataset. There are numerous techniques to normalize data like Z-score Normalization, Min-Max normalization. In this thesis, the z-score normalization algorithm was re-implemented from EEGLAB[3] for MATLAB[4]. This 6.5 normalization was applied per subject, after preprocessing techniques like FIRWIN and ICA. Also, in the pipeline, common channels of all datasets were used.

$$x'_{i,j} = \frac{x_{i,j}}{\sigma_i} \tag{6.5}$$

where:

- σ_i is the standard deviation of the channel
- $x_{i,j}$ is old point
- $x'_{i,j}$ is new rewritten point

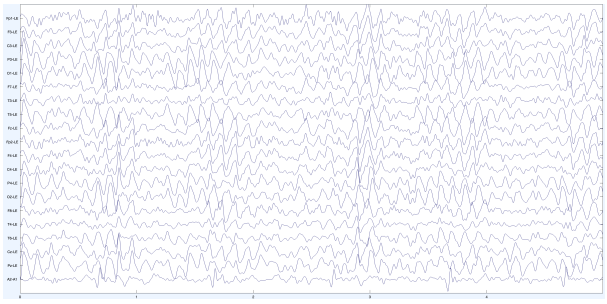

Figure 6.2: MDD subject before normalization.

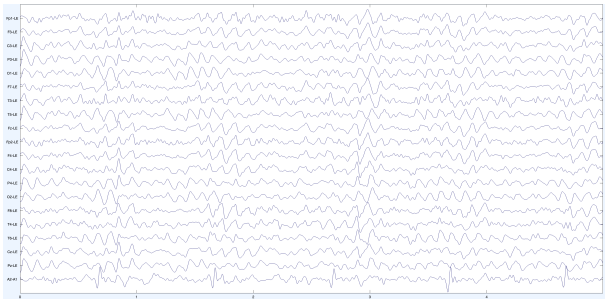

Figure 6.3: MDD subject after normalization.

[3] https://eeglab.org/
[4] https://www.mathworks.com/products/matlab.html

6.4 AutoReject

Autoreject[5] is EEG library for repairing or removing bad epochs of EEG recordings. This tool is a helpful to utilize more information from brain activity and removing noise from EEG recordings. The implementation is on cross-validation in conjunction with a robust evaluation metric to estimate the optimal peak-to-peak threshold – a quantity commonly used for identifying bad trials in M/EEG [40].

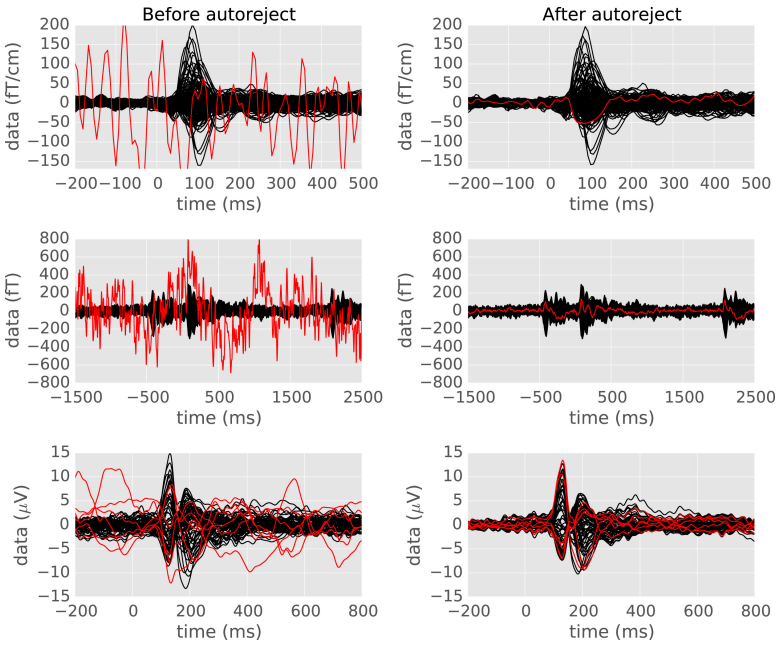

Figure 6.4: Effect of Autoreject from [40].

By automating the detection and correction of epochs, Autoreject helps to improves the quality and dependability of EEG data analysis. This leads to a understanding of the brains electrical activity crucial, for applications, like clinical diagnoses, neuroscientific studies and brain computer interfaces.

6.5 Result of preprocessing

These techniques in Sections 6.1, 6.2, 6.3, and 6.4 were applied to the EEG recordings per subject in DASPS, MDD and SAD dataset. In figures 6.7 we can clearly see a differences on last channels. Also movement artifacitfacts were repaired in channels F4.

[5] https://autoreject.github.io/stable/index.html

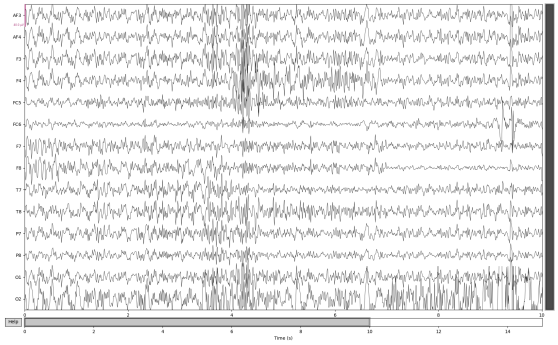

Figure 6.5: EEG graph of S03 before cleaning.

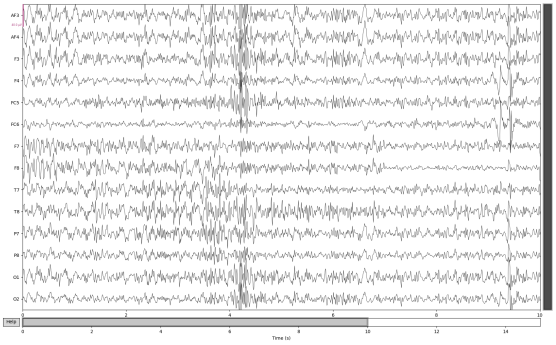

Figure 6.6: EEG graph of S03 after cleaning.

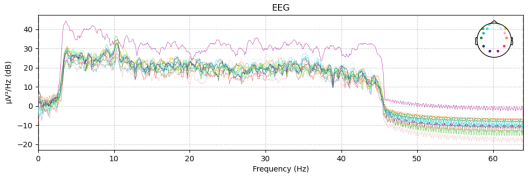

Figure 6.7: Power Spectral Density graph of S03 before cleaning.

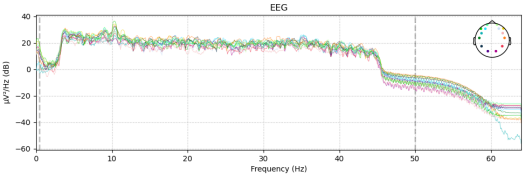

Figure 6.8: Power Spectral Density graph of S03 after cleaning.

Chapter 7

Algorithm Design

The main idea of algorithm design is to combine all promising techniques from chapter 2. The main goal of this thesis is to classify patients who have Social Anxiety Disorder as well as Major Depressive Disorder. The dataset, which includes healthy subjects, subjects diagnosed with Social Anxiety Disorder and Major Depressive Disorder, did not exist at the time of implementing and writing this thesis. To solve this issue, combining datasets together with the normalization step is necessary to solve this issue. In addition, the merging of these three datasets will give us more data, which is generally a positive factor to train neural networks and not overfit the data. Another key point of the algorithm design is to divide the segments into different time lengths, as experiments from Tables 5.1 and 5.2 suggest different accuracies across time.

7.1 Conceptual Framework

The algorithm pipeline consists of several steps. Preprocessing, normalizing, splitting, getting the best features and using a more robust classifier to obtain a result. This experimental pipeline is combining the techniques from papers [55], [25], [68] and [12]. Instead of removing the channels by evolutionary algorithm, we used a genetic algorithm to find a best features of our eeg signals for classification. The problem of a lot of papers is that the mathematical findings, formulas and results can't be easily replicated [37]. To be sure of the correct implementation of EEG features, the standard library called *mne_feauters*[1] is used for extraction of EEG recordings. These extracted features from EEG recordings are then given as input to CNN, CNN GA and Bert Transformer with removed tokenizer. The loading and preprocessing of EEG recordings is done by applying with techniques discussed in chapter 6. The pipeline is shown in Figure 7.1.

[1] https://mne.tools/mne-features/

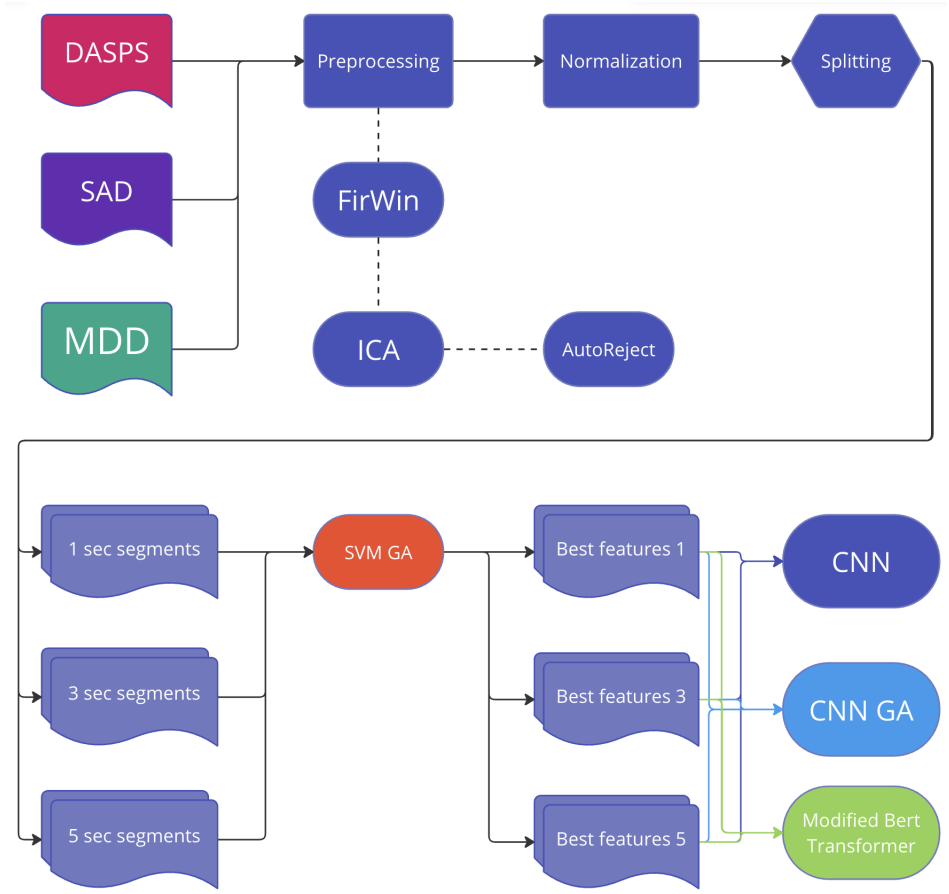

Figure 7.1: Algorithm design flow graph.

7.2 Finding best features

To find the best features for our dataset, we used an evolutionary algorithm, specifically a genetic algorithm. The genome of our candidate solution is coded in binary strings consisting of 1 or 0. The genome's length is the count of all features, which is 31 features in total. Specifically: App entropy, Decorr time, Higuchi fd, Hjorth complexity, Hjorth complexity spect, Hjorth mobility, Hjorth mobility spect, Hurst Exp, Katz FD, Kurtosis, Line Length, Mean, Pow Freq Bands, Ptp Amp, Quantile, RMS, Samp Entropy, Skewness, Spect Edge Freq, Spect Entropy, Spect Slope, Std, SVD Entropy, SVD Fisher Info, Teager Kaiser Energy, Variance, Wavelet Coef Energy, Zero Crossings, Max Cross Corr, Nonlin Interdep, Phase Lock Val, Spect Corr, Time Corr. We used all features available in *mne_features* library. The fitness function of GA is the accuracy of the SVM algorithm. We used the SVM classifier because in most papers discussed in section 4.7, SVM is the state-of-the-art model. SVM is also used for the genetic algorithm in this paper [54]. This classification is not as precise as CNN or Transformers, but it can give us a faster evaluation of the features than CNN or Transformers. However, different features give us different vector lengths. To not implement a custom architecture for every input without padding, the SVM classifier is the best option for that.

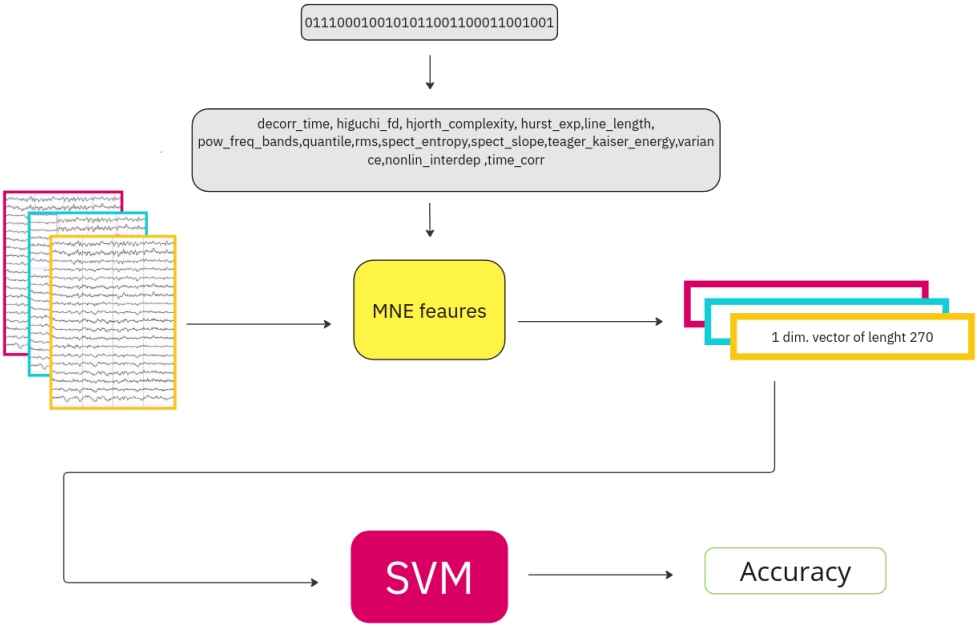

Figure 7.2: Evaluation of one candidate in SVM GA.

7.3 Finding best architecture

The next step was to find the best architecture optimized for a given best-feature dataset. The best architecture would be based on an algorithm called CNN-GA [83]. The algorithm will build the architecture based on the genome on a given maximum length and depth. This algorithm was modified so that instead of merely classifying images, it will classify a given feature vector. CNNs often perform better than SVMs on large and complex datasets. To verified this, CNN and CNN-GA was implemented.

CNN

This convolutional neural network was implemented to analyse 1-dimensional data. Usually, 1-dimensional data are time-series signals or audio signals. In this implementation, the input is a 1-dimensional data vector of features, which were given as a result of the SVM classifier. The network architecture begins in a 1D convolutional layer with 16 filters. The kernel size is 5. Padding is set to 2 to preserve the dimensionality of the input through this layer. Then, the activation function ReLu is applied. The pooling layer applies MaxPool with kernel size 2 and padding is 2. The second convolutional layer doubles the filter size to 32 then with same Relu and pooling layers. The result from these layers is then flatten and used to Linear fully connected layers which they gives a output for one hot encoding label data[61]. This simple Convolutional Neural Network is not complicated enough to overfit the data. The architecture is shown in Figure 7.3.

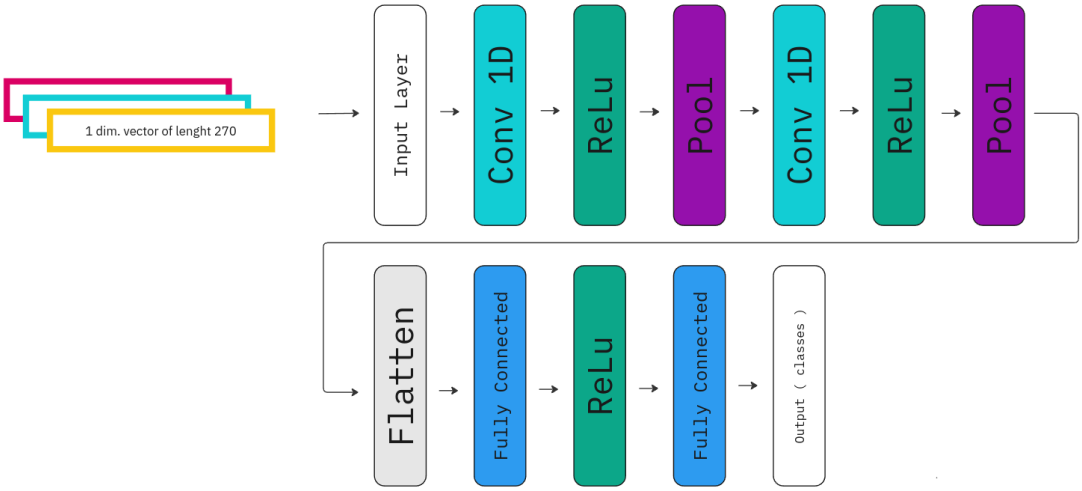

Figure 7.3: Architecture of simple CNN.

CNN-GA

The problem of finding a best architecture, which is not design by people, is a space state search problem [76]. To get valuable results, we needed to limit the search space. To generate the best architecture, we limited how big the architecture can be. The architecture is inspired from [76], but is simplified and re-implemented to our problem. Again, to represent a candidate, the binary number is used. The binary number is split into two numbers, which represent count of layers. For example, the binary number 110001 is split into 110 and 001, which corresponds to 6 convolutional layers blocks and 1 linear blocks. Also the architecture can't be big as Resnet-56 [34], because there is not enough vector data after that much convolutions. Also, in the early stages of implementation, LSTM block was used as paper[12] suggest a better accuracy. In my experiments, these CNN-LSTM-GA would not learn above accuracy 50%, so they were excluded from this thesis.

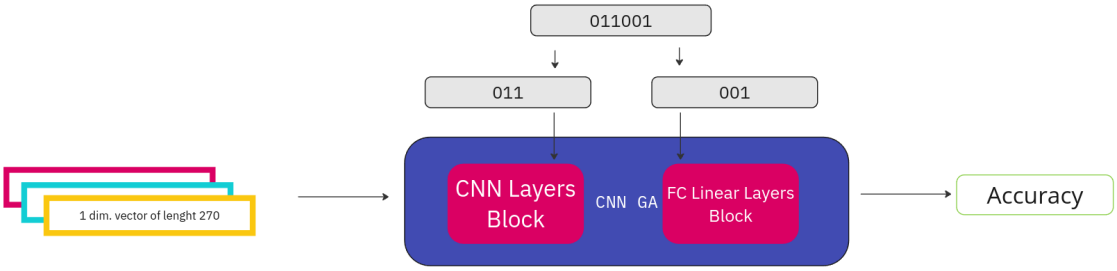

Figure 7.4: CNN GA evaluation of one candidate.

7.4 Experimental Bert Transformer implementation

The design is based on BERT model, which is typically known for its strong performance in natural language processing tasks as discussed in Section 4.5. In this thesis instead of using BERT tokenizer for words, we applied a Convolutional Layers as a input to extract the spatial information from features vectors. We decided to try this approach instead of using *EEG-deformer*[25], to be able to explain, which features are put as an input into Neural Network as well as explain what numbers behind the input means, because we can by mathematically express every number from math formula, which we discussed in Section 3.5. Also this gives a control how big the input vector is, to reduce a memory performance. The final touches of the architecture can be seen in Figure 7.5.

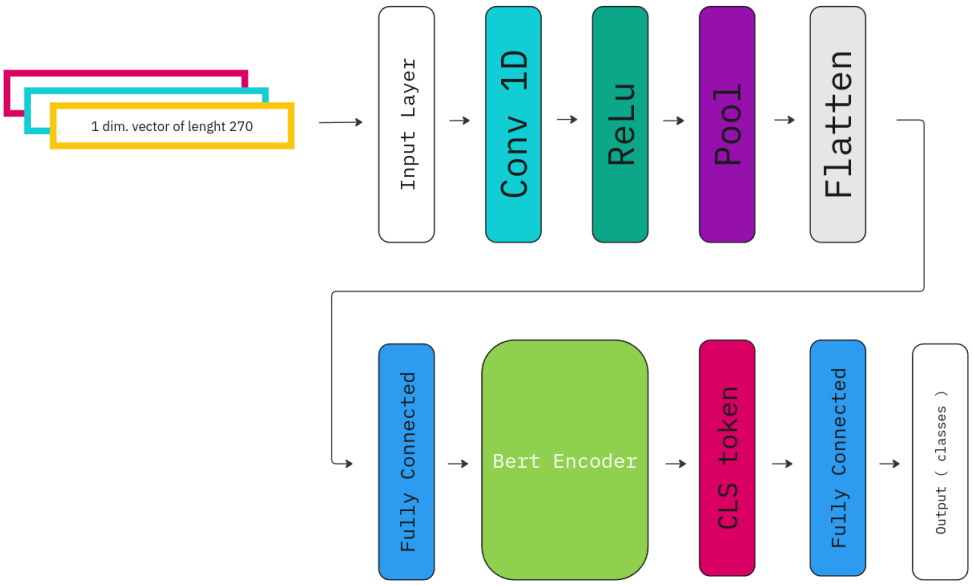

Figure 7.5: The experimental Bert Architecture.

7.5 Implementation

The implementation was done by using by Python 3.10 using Anaconda[2] environment. The library for loading and cleaning the EEG recording is called MNE library[30]. The Support Vector Machine, which we discussed in Section 4.1, is implemented from standard library Scikit[3]. The implementation of Convolutional Neural Networks is done in the library called Pytorch[4]. The bone of BertArchitecture is from library called HugginFace [5]. The algorithm for generating genomes candidates, with cross-over implementation is from library called Pygad[6]. To replicated results for every step in the pipeline can be run also with Makefile[7].

[2] https://www.anaconda.com/
[3] https://scikit-learn.org/stable/modules/svm.html
[4] https://pytorch.org/
[5] https://huggingface.co/
[6] https://pygad.readthedocs.io/en/latest/
[7] https://en.wikipedia.org/wiki/Make_(software)

Distribution of the data

The DASPS 5.1 dataset have labels light, normal, modern and severe. The SAD 5.3 dataset have labels Control, Average, Mild and Severe. The MDD 5.2 dataset have labels H which stands for normal and MDD, which stands for Major Depresive Disorder.

After normalization, the models separate these labels into four groups. The light and normal from DASPS with Control and Average from SAD to the class called Normal. The moderate from DASPS and Mild from SAD is in class called Mild Moderate. At last, the labels severe and Severe from DASPS and SAD are in class Severe.

After re-labeling, the data was distributed in 70 to 15 to 15 ratio for training, validation and testing. We can see that the Figure 7.6. Which specific subjects were used, we can see in the Table 7.1. It's worth noting that even if we mixed three datasets, 15 subjects validation is not a representative sample of population. This is a general problem of EEG classification as well as classification in medical data.

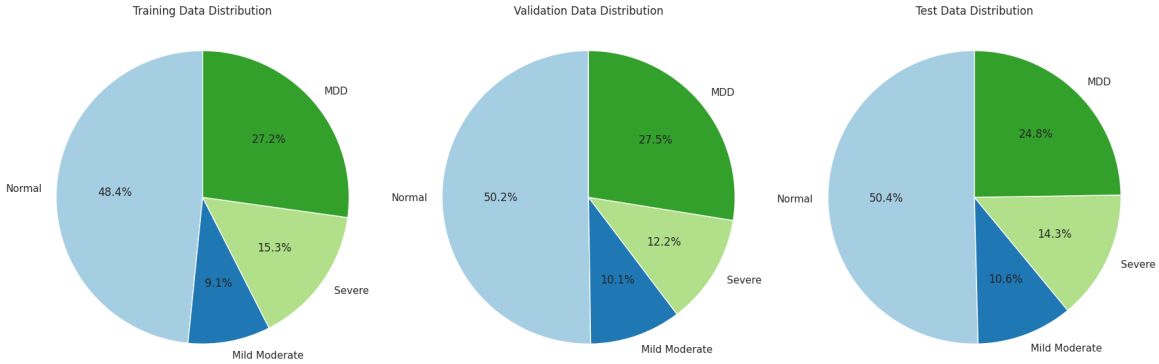

Figure 7.6: Distribution of classes in Mixed dataset.

Dataset	Training Subjects	Validation Subjects	Testing Subjects
MDD	S27, S6, S28, S17, S23, S2, S3, S10, S19, S30, S14, S15, S9, S7, S11, S13, S29, S4, S18	S1, S21, S24, S20, S8, S32	S25, S5, S22, S26, S16, S34
DASPS	S09, S17, S10, S03, S16, S02, S12, S19, S04, S01, S06, S11	S23, S18, S14, S22, S07, S13	S15, S08, S05, S20, S21
SAD	C8, C9, C1, C17, C15, C11, C18, C3, C6, C7, C13, C12	C2, C4, C10	C5, C16, C14

Table 7.1: Training, Validation, and Testing Sets for Each Dataset.

Chapter 8

Results

8.1 SVM GA

The results were achieved with parameters of 20 generations with 20 candidates with adaptive mutation. The trials were done for 1, 3 and 5 seconds segments to achieved the best features for specific segments.

For 1 seconds segments is accuracy **0.68**. List the best features identified by the SVM model for 1 seconds segment: decorr time, higuchi fd, hjorth complexity, hjorth complexity spect, hjorth mobility, hurst exp, mean, pow freq bands, spect entropy, spect slope, time corr.

Table 8.1: Val. Class Report for 1 second

Class	Precision	Recall	F1-Score
Normal	0.72	0.71	0.72
Mild Moderate	0.50	0.01	0.01
Severe	0.30	0.42	0.35
MDD	0.83	0.99	0.90
Accuracy		0.68	

Table 8.2: Test Class Report 1 second

Class	Precision	Recall	F1-Score
Normal	0.64	0.62	0.63
Mild Moderate	0.92	0.04	0.08
Severe	0.41	0.54	0.47
MDD	0.63	0.82	0.72
Accuracy		0.60	

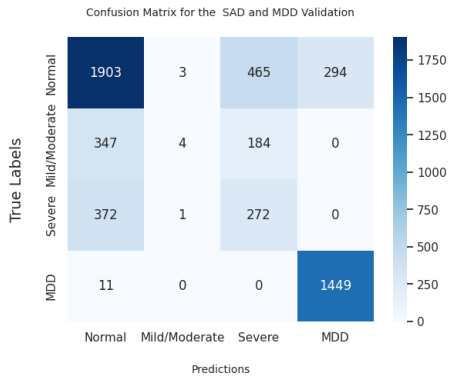

Figure 8.1: Confusion Matrix for Validation Data in Mixed 1 sec.

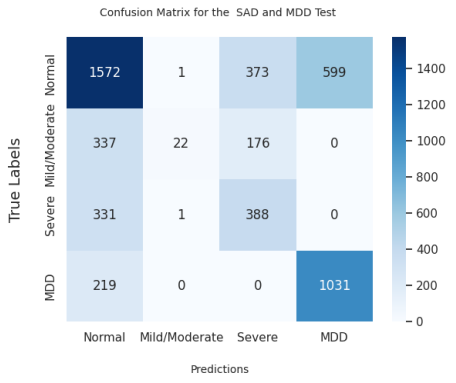

Figure 8.2: Confusion Matrix for Test Data in Mixed 1 sec.

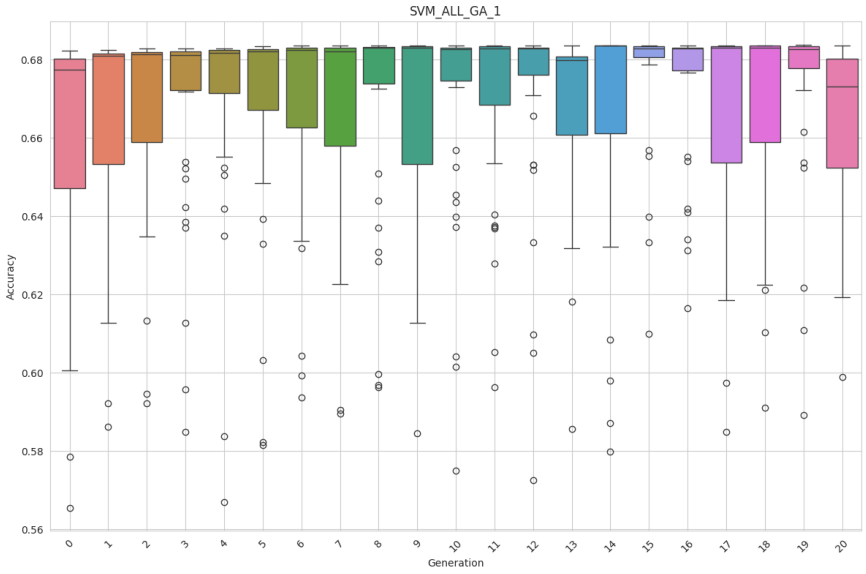

Figure 8.3: Box plot of SVM GA 1 during generations.

In Figure 8.3, we can clearly see that the SVM GA classification suffers to find a better solution than generation 4, which would be better and the algorithm tries a different combination of features, which in most cases has worse results than in generation 4.

For mixed 3 seconds is accuracy **0.64**. List the best features identified by the SVM model for 1 seconds segment: higuchi fd, hjorth complexity spect, hjorth mobility spect, katz fd, mean, pow freq bands, spect slope, std, svd fisher info, nonlin interdep, phase lock val, time corr. The results and graphs can be seen in Appendix A.

For mixed 5 seconds is accuracy **0.68**. List the best features identified by the SVM model for 5 seconds segment: app entropy, decorr time, higuchi fd, hjorth complexity,hjorth complexity spect, hurst exp, katz fd, line length, mean, pow freq bands, ptp amp, rms, spect entropy, spect slope, std, svd entropy, svd fisher info, teager kaiser energy, variance, wavelet coef energy, nonlin interdep, phase lock val, time corr. The results and graphs can be seen in Appendix A.

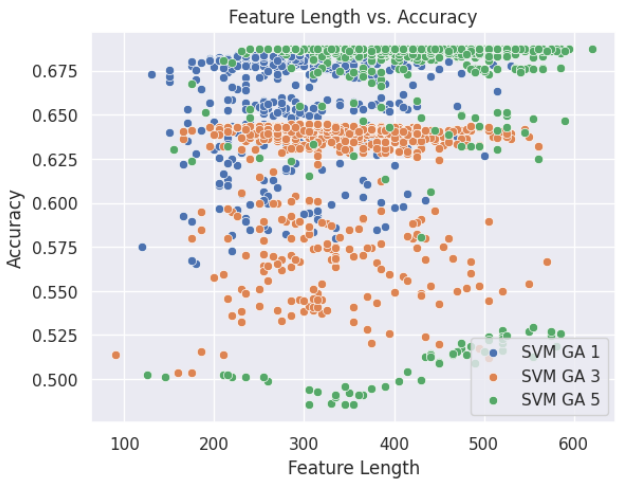

Figure 8.4: Accuracy based on length of features vector for SVM GA.

The feature vector length for specific features was recorded with accuracies. In Figure 8.4, we can clearly see that 1 and 5 seconds segments are overall better than 3 seconds segment. Also, we can have a same accuracy with smaller vector than with longer vectors in SVM classification.

8.2 CNN

Training CNN was done with Adam Optimizer and with early stop callback to prevent a overfitting. On Best features from a SVM GA results in Section 8.1 was preform next training. Using a CNN as a classifier, we can see a dramatical improvement by **12%**. The validation accuracy during training with validation loss can be seen in Figure 8.5 and in Figure 8.6. Because of early stop, we can also see that 1 second best features preforming better over training. The explanation of this result can be interpreted as 1 second features segments have more data, which CNN benefits from.

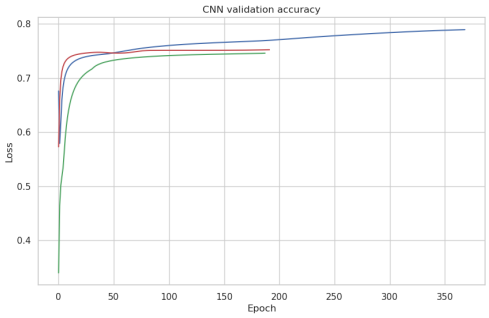

Figure 8.5: CNN val. acc. during training.

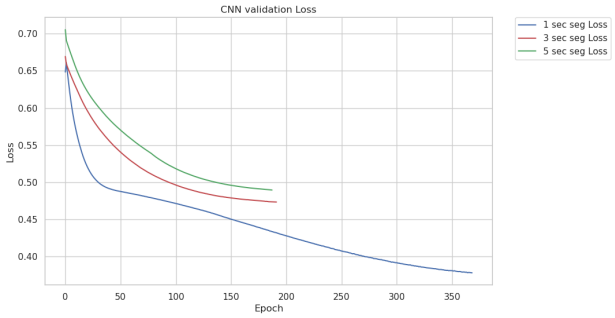

Figure 8.6: CNN val. loss during training.

To preform more detailed inspection on how CNN classify, The AUC and DET graphs were made. They can be seen in 8.7 and 8.8. The 3 and 5 second CNN suffer to classify correctly Anxiety classes. Otherwise for classyfing Major Depressive Disorder we can see a **97%** AUC value. We can say, that CNN are more than suitable to identifying a Major Depressive Disorder.

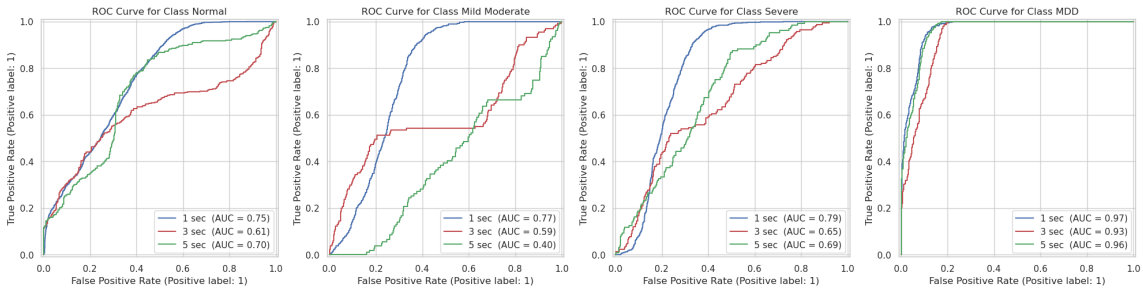

Figure 8.7: AUC graph for Normal vs Mild Moderate vs Severe vs MDD.

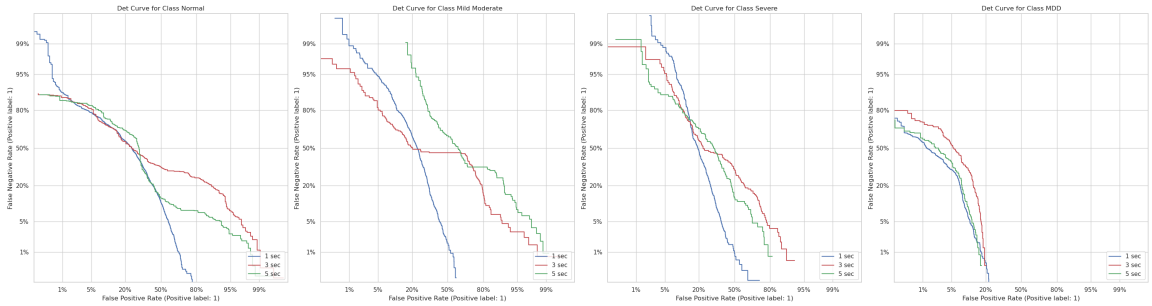

Figure 8.8: DET graph for Normal vs Mild Moderate vs Severe vs MDD

Even if AUC and DET graphs suggests, that by 1 second segment can be tune by different thresholds, the actual trained CNN have statistics, where almost all Anxiety labels were classified as Normal label. We can see in confusion matrices for validation in Figure 8.9 and in Figure 8.10. As consequence, Precision , Recall and F1-score are 0 in Mild Moderate class. It's worth noting that in that Recall values for MDD are **0.87%** and **0.77%**. The classification report tables can be seen in Table 8.3 and Table 8.4.

Table 8.3: Val. Class Report for 1 second

Class	Precision	Recall	F1-Score
Normal	0.68	0.47	0.56
Mild Moderate	0.00	0.00	0.00
Severe	0.04	0.01	0.02
MDD	0.77	0.87	0.82
Accuracy		0.81	

Table 8.4: Test Class Report for 1 second

Class	Precision	Recall	F1-Score
Normal	0.67	0.60	0.63
Mild Moderate	0.00	0.00	0.00
Severe	0.62	0.14	0.22
MDD	0.64	0.77	0.70
Accuracy		0.81	

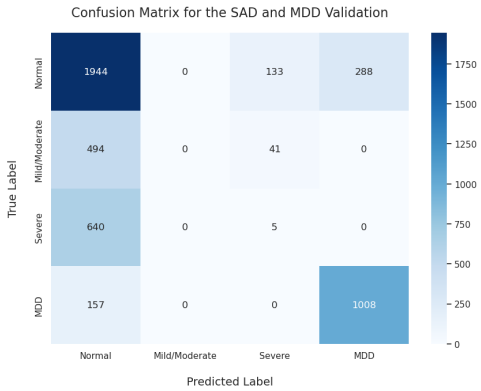

Figure 8.9: Val Confusion Matrix 1 second.

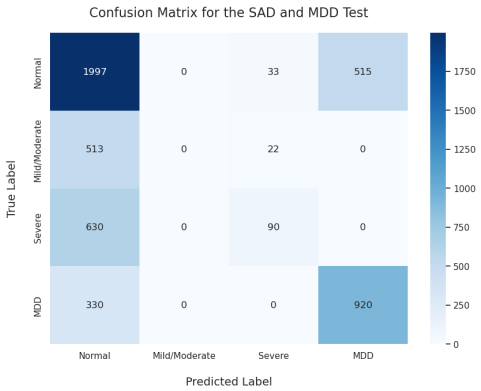

Figure 8.10: Test Confusion Matrix 1 second.

8.3 CNN GA

CNN GA was ran in 5 generation cycle with 5 individuals for every type of segment. The actual number of trained neural networks of this section is 75. Every CNN GA use own best features which came as a result from Section 8.1. We can see in the Figures 8.11, that using Evolutionary algorithm for CNN GA did enhanced the generation overall accuracy, but it never outperformed the CNN results discussed in Section 8.2.

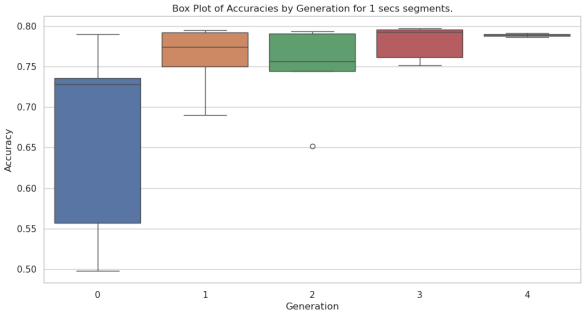

Figure 8.11: Box plot of CNN GA 1 during generations.

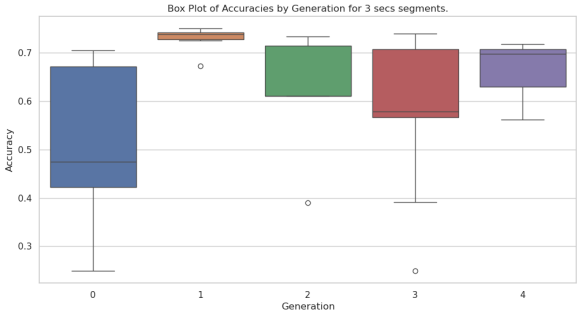

Figure 8.12: Box plot of CNN GA 3 during generations.

To inspect more, second generation of AUC graph is shown in Figure 8.13. We can from this best generation conclude that the our implementation of CNN GA didn't preform better than CNN implementation. We also need to discussed a architecture efficiency to

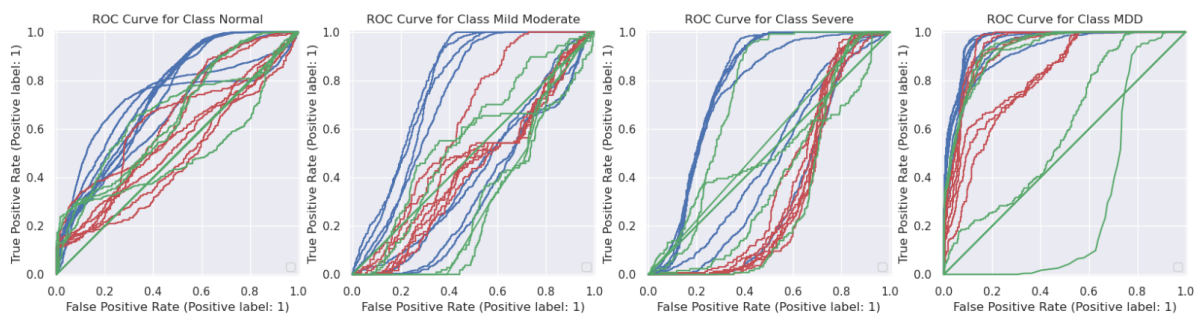

Figure 8.13: Second Best generation of CNN by accuracy.

accuracy. In Figure 8.14, we can see that architectures of CNN GA preforms better when they have less learnable parameters. Also the 1 seconds segments preforms better than the rest of the architectures. For CNN GA 3 we can see also a outliers with $3.5 \times 1e7$ but didn't preform better then the rest of the architectures. CNN and CNN GA can classify Healthy subjects and subjects with Major Depressive Disorder but it struggle with different Anxiety levels.

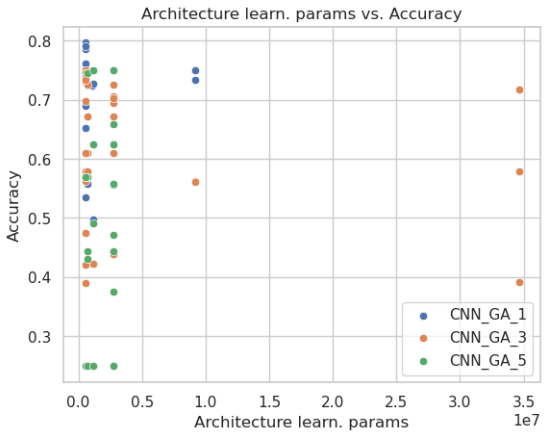

Figure 8.14: Architecture efficiency to accuracy.

8.4 Experimental Bert Transformer

The last experiment is using our custom Bert implementation. Transformers are generally used for sequential data and in general, they outperform the CNN in this fields. With our results, we can confirm it. The optimizer for training was Adam, same as in CNN for comparison purpose. We run the experiments on 100 epochs without Early stopping. There are two main differences than on SVM and CNN category. First, in training accuracy we hit more than **90%** in 1 seconds segments and 5 second segments. Second, in validation accuracy for 5 seconds we hit **85%** accuracy. The training and validation are shown in Figure 8.15.

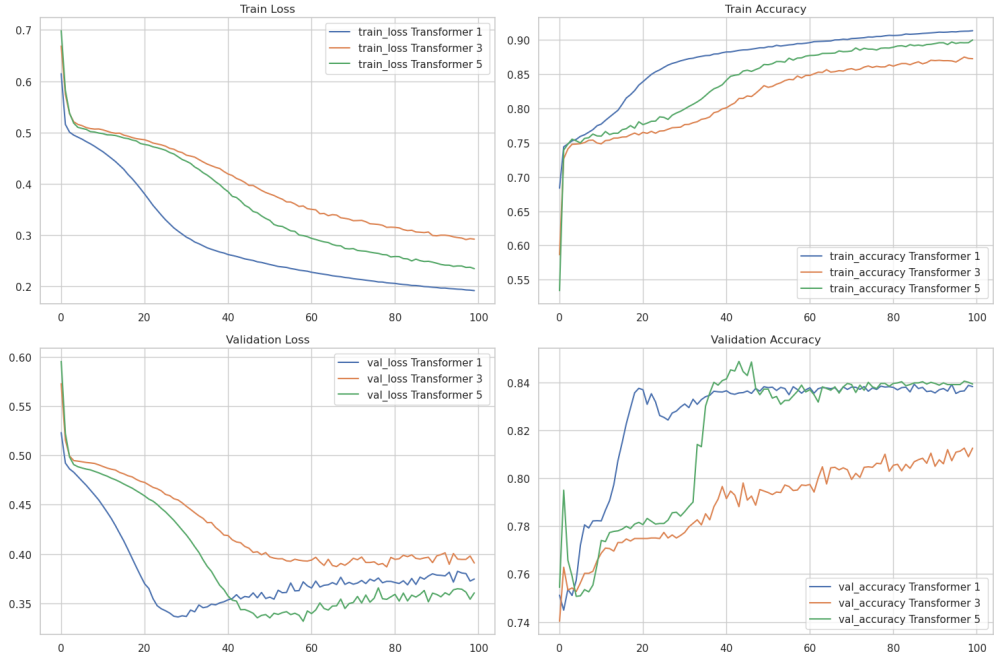

Figure 8.15: Train Loss, Train Accuracy, Validation Loss, and Validation Accuracy over 100 epochs.

To evaluate closely, classification report for validation and test dataset is shown in Tables 8.5 and 8.6. One second segments are displayed as in previous Sections 8.1 and 8.2 . Transformers in some cases was able to dissociate the some cases of SEVERE and Mild Moderate of Anxiety levels. Also is more certain if subject have Major Depressive Disorder with Recall **0.98%**.

Table 8.5: Val. Class Report for 1 second

Class	Precision	Recall	F1-Score
Normal	0.80	0.59	0.68
Mild Moderate	0.22	0.00	0.01
Severe	0.27	0.37	0.31
MDD	0.81	0.98	0.89
Accuracy		0.85	

Table 8.6: Test Class Report for 1 second

Class	Precision	Recall	F1-Score
Normal	0.70	0.52	0.59
Mild Moderate	0.39	0.01	0.03
Severe	0.42	0.53	0.47
MDD	0.62	0.83	0.71
Accuracy		0.80	

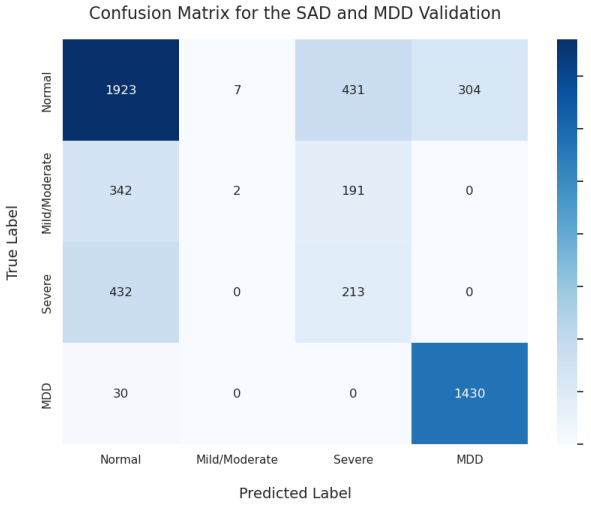

Figure 8.16: Transformer Validation Confusion Matrix 1 second.

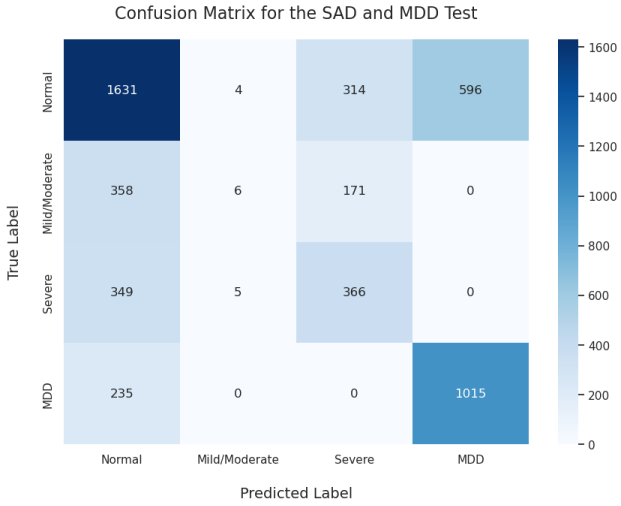

Figure 8.17: Transformer Test Confusion Matrix 1 second.

To summarize, for our dataset which was correctly splitted to 70 to 15 to 15 ration, we applied different technique of Machine Learning as well as Genetics algorithm. The result speaks in that SVM machines can be used and a evaluation technique to find best features. These selected features by genetic algorithm can be in shorter segments improved by stronger classifier like CNN about 10%. The bigger architectures of CNN in this field doesn't performer dramatically better. In the end the most accurate models were Transformers, specifically Bert Transformer with 85% accuracy for 5 seconds segments, with features: app entropy, decorr time, higuchi fd, hjorth complexity, hjorth complexity spect, hurst exp, katz fd, line length, mean, pow freq bands, ptp amp, rms, spect entropy, spect slope, std, svd entropy, svd fisher info, teager kaiser energy, variance, wavelet coef energy, nonlin interdep, phase lock val and time corr. The graph which sumarize all validation results can be seen in Figure 8.18.

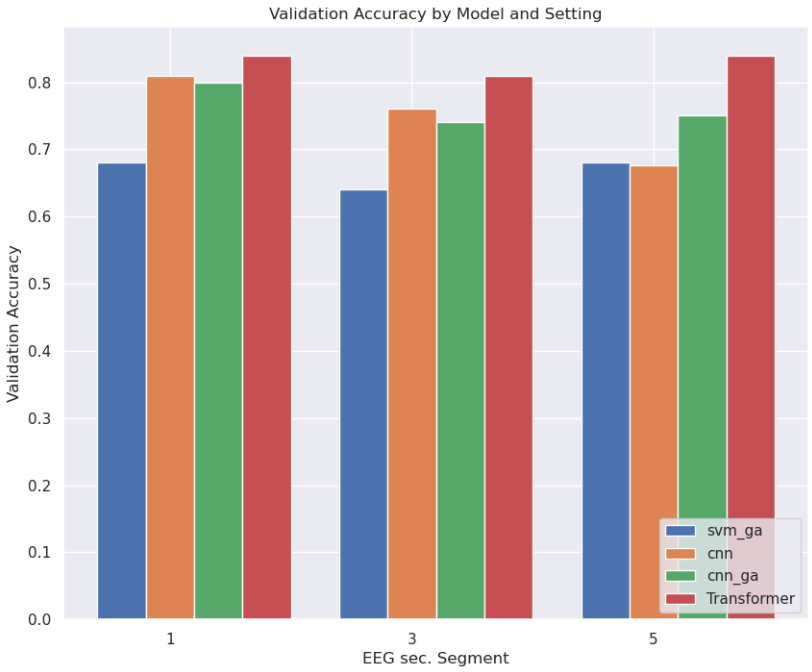

Figure 8.18: Results of applied techniques of different time segments.

Chapter 9

Conclusion

The purpose of this thesis was to investigate and develop classification methods that could classify healthy subjects, subjects with anxiety disorder, and subjects with major depressive disorder. In Chapter 2, we explain what anxiety disorders and depression are and what the relationships are between them. We also discuss different methods for classifying mental disorders and their long-term effects. In Chapter 3, I researched the anatomy of the brain and what an electroencephalogram is. Next, we discussed signal analysis of EEG and the properties of EEG signals. In Chapter 4, I explain and explore current popular machine learning methods, which were also used in developing our algorithm. We discussed current trends in EEG classification as well as the limitations of our classification of Anxiety and Depression. In Chapter 5, we discussed datasets used for classifying Social Anxiety Disorder and Major Depressive Disorder. In Chapter 6, we discussed the actual preprocessing techniques, specifically FIR filters, Independent Component Analysis and Normalization, which were then used to clean our dataset for classification. In Chapter 7, we deeply discuss every step of our algorithm and the solution for this thesis. We propose a solution that uses a genetic algorithm to select the best features for our dataset. After finding significant features, we developed three architectures to enhance our accuracy for classification. This solution allows us to optimize input vectors for classification and the number of trainable parameters. We also developed an experimental Transformer architecture based on Bert[24] Transformer. In Chapter 8, we discussed our achieved results. Best accuracies were achieved when the EEG recordings were split into 1-second segments. The classification was improved overall using transformer-based models by **16%**, and the best achieved accuracy was **85%**. The results were evaluated using benchmark statistics F1, Recall, Precision, Confusion Matrix AUC, and DET graphs. In our results, we provide an analysis of memory usage by how the vector length affects accuracy as well as how the count of trainable parameters affects the score. The contribution of this thesis is a model which can classify Healthy subjects, subjects with Anxiety and subjects with Major Depressive Disorder. Also, we provide a solution based on a genetics algorithm and deep learning models.

Bibliography

[1] Available at: https://step1.medbullets.com/neurology/113013/cerebral-cortex.

[2] *The Critical Relationship Between Anxiety and Depression*. Available at: https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2020.20030305.

[3] *Anxiety Disorders*. January 1995. Available at: https://www.nimh.nih.gov/health/topics/anxiety-disorders.

[4] *Major Structures and Functions of the Brain - Discovering the Brain - NCBI Bookshelf*. January 2001. Available at: https://www.ncbi.nlm.nih.gov/books/NBK234157/.

[5] Mar 2024. Available at: https://www.ibm.com/topics/support-vector-machine.

[6] Apr 2024. Available at: https://www.ibm.com/topics/convolutional-neural-networks.

[7] Jan 2024. Available at: https://www.ibm.com/topics/recurrent-neural-networks.

[8] Abo Zahhad, M.; Ahmed, S. and Seha, S. N. A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals. *International Journal of Intelligent Systems and Applications (IJISA)*, may 2015, vol. 07, p. 48–54.

[9] Adams, G. C.; Balbuena, L.; Meng, X. and Asmundson, G. J. When social anxiety and depression go together: A population study of comorbidity and associated consequences. *Journal of Affective Disorders*, 2016, vol. 206, p. 48–54. ISSN 0165-0327. Available at: https://www.sciencedirect.com/science/article/pii/S0165032716303020.

[10] Agrawal, A. *Illuminate: A novel approach for depression detection with explainable analysis and proactive therapy using prompt engineering*. 2024.

[11] Al Ezzi, A.; Kamel, N.; Al Shargabi, A. A.; Al Shargie, F.; Al Shargabi, A. et al. Machine learning for the detection of social anxiety disorder using effective connectivity and graph theory measures. *Frontiers in Psychiatry*, 2023, vol. 14. ISSN 1664-0640. Available at: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1155812.

[12] Al Ezzi, A.; Yahya, N.; Kamel, N.; Faye, I.; Alsaih, K. et al. Severity Assessment of Social Anxiety Disorder Using Deep Learning Models on Brain Effective Connectivity. *IEEE Access*, 2021, vol. 9, p. 86899–86913.

[13] Albera, L.; Kachenoura, A.; Comon, P.; Karfoul, A.; Wendling, F. et al. ICA-Based EEG denoising: a comparative analysis of fifteen methods. *Polska Akademia Nauk.Bulletin of the Polish Academy of Sciences*, december 2012, vol. 60, no. 3, p. 407. Available at: https://www.proquest.com/scholarly-journals/ica-based-eeg-denoising-comparative-analysis/docview/1331136924/se-2. Copyright - Copyright Versita Dec 2012; Last updated - 2016-10-08.

[14] Aleem, S.; Huda, N. u.; Amin, R.; Khalid, S.; Alshamrani, S. S. et al. Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. *Electronics*, 2022, vol. 11, no. 7. ISSN 2079-9292. Available at: https://www.mdpi.com/2079-9292/11/7/1111.

[15] Baghdadi, A.; Aribi, Y.; Fourati, R.; Halouani, N.; Siarry, P. et al. DASPS: A Database for Anxious States based on a Psychological Stimulation. *CoRR*, 2019, abs/1901.02942. Available at: http://arxiv.org/abs/1901.02942.

[16] Baghdadi, A.; Aribi, Y.; Fourati, R.; Halouani, N.; Siarry, P. et al. *DASPS database*. 2021. Available at: https://dx.doi.org/10.21227/barx-we60.

[17] Bishop, C. M. *Pattern recognition and machine learning*. New York : Springer, [2006] ©2006, [2006]. Available at: https://search.library.wisc.edu/catalog/9910032530902121. Textbook for graduates.;Includes bibliographical references (pages 711-728) and index.

[18] Boonstra, T. W.; Stins, J. F.; Daffertshofer, A. and Beek, P. J. *Effects of sleep deprivation on neural functioning: an integrative review*. March 2007. Available at: https://doi.org/10.1007/s00018-007-6457-8.

[19] Boudraa, A.-O. and Salzenstein, F. Teager–Kaiser energy methods for signal and image analysis: A review. *Digital Signal Processing*, 2018, vol. 78, p. 338–375. ISSN 1051-2004. Available at: https://www.sciencedirect.com/science/article/pii/S1051200418300927.

[20] Burrus, C. and Parks, T. Time domain design of recursive digital filters. *IEEE Transactions on Audio and Electroacoustics*, 1970, vol. 18, no. 2, p. 137–141.

[21] Canu, E.; Kostić, M.; Agosta, F.; Munjiza, A.; Ferraro, P. M. et al. Brain structural abnormalities in patients with major depression with or without generalized anxiety disorder comorbidity. *Journal of Neurology*, May 2015, vol. 262, no. 5, p. 1255–1265. ISSN 1432-1459. Available at: https://doi.org/10.1007/s00415-015-7701-z.

[22] Cho, R.; Zaman, M.; Cho, K. T. and Hwang, J. *Analyzing Brain Activity During Learning Tasks with EEG and Machine Learning*. 2024.

[23] Cui, G.; Li, X. and Touyama, H. Emotion recognition based on group phase locking value using convolutional neural network. *Scientific Reports*, Mar 2023, vol. 13, no. 1, p. 3769. ISSN 2045-2322. Available at: https://doi.org/10.1038/s41598-023-30458-6.

[24] Devlin, J.; Chang, M.-W.; Lee, K. and Toutanova, K. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019.

[25] Ding, Y.; Li, Y.; Sun, H.; Liu, R.; Tong, C. et al. *EEG-Deformer: A Dense Convolutional Transformer for Brain-computer Interfaces*. 2024.

[26] Donoghue, T. and Voytek, B. Automated meta-analysis of the event-related potential (ERP) literature. *Scientific reports*. England: Nature Publishing Group, 2022, vol. 12, no. 1, p. 1867–1867. ISSN 2045-2322.

[27] Díaz M., H. A. and Córdova, F. On the meaning of Hurst entropy applied to EEG data series. *Procedia Computer Science*, 2022, vol. 199, p. 1385–1392. ISSN 1877-0509. Available at: https://www.sciencedirect.com/science/article/pii/S1877050922001764. The 8th International Conference on Information Technology and Quantitative Management (ITQM 2020 2021): Developing Global Digital Economy after COVID-19.

[28] Feldman, R. *Parent-infant synchrony and the construction of shared timing; physiological precursors, developmental outcomes, and risk conditions.* March 2007. Available at: https://acamh.onlinelibrary.wiley.com/doi/10.1111/j.1469-7610.2006.01701.x.

[29] Gan, B. S.; Hara, T.; Han, A.; Alisjahbana, S. W. and As'ad, S. Evolutionary ACO Algorithms for Truss Optimization Problems. *Procedia Engineering*, 2017, vol. 171, p. 1100–1107. ISSN 1877-7058. Available at: https://www.sciencedirect.com/science/article/pii/S1877705817304770. The 3rd International Conference on Sustainable Civil Engineering Structures and Construction Materials - Sustainable Structures for Future Generations.

[30] Gramfort, A.; Luessi, M.; Larson, E.; Engemann, D. A.; Strohmeier, D. et al. MEG and EEG Data Analysis with MNE-Python. *Frontiers in Neuroscience*, 2013, vol. 7, no. 267, p. 1–13.

[31] Gu, W.; Yang, B. and Chang, R. *Machine Learning-based EEG Applications and Markets.* 2022.

[32] Harner, R. N. Singular Value Decomposition—A general linear model for analysis of multivariate structure in the electroencephalogram. *Brain Topography*, Sep 1990, vol. 3, no. 1, p. 43–47. ISSN 1573-6792. Available at: https://doi.org/10.1007/BF01128860.

[33] Hateli, B. The effect of non-directive play therapy on reduction of anxiety disorders in young children, june 2021. Available at: https://scite.ai/reports/10.1002/capr.12420.

[34] He, K.; Zhang, X.; Ren, S. and Sun, J. *Deep Residual Learning for Image Recognition.* 2015.

[35] Hohls, J. K.; König, H.-H.; Quirke, E. and Hajek, A. Anxiety, depression and quality of life-A systematic review of evidence from longitudinal observational studies. *Int. J. Environ. Res. Public Health.* MDPI AG, november 2021, vol. 18, no. 22, p. 12022.

[36] Ibrahim, N.; Hamid, H.; Rahman, S. and Fong, S. Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy. *Pertanika Journal of Science and Technology*, january 2018, vol. 26, p. 329–340.

[37] Inouye, K. and Mills, D. Fear of the academic fake? Journal editorials and the amplification of the 'predatory publishing' discourse. *Learned Publishing*, 2021, vol. 34, no. 3, p. 396–406. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/leap.1377.

[38] Ionescu, D. F.; Niciu, M. J.; Mathews, D. C.; Richards, E. M. and Zarate Jr, C. A. NEUROBIOLOGY OF ANXIOUS DEPRESSION: A REVIEW. *Depression and Anxiety*, 2013, vol. 30, no. 4, p. 374–385. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/da.22095.

[39] Jadon, S. Introduction to different activation functions for deep learning. *Medium, Augmenting Humanity*, 2018, vol. 16.

[40] Jas, M.; Engemann, D. A.; Bekhti, Y.; Raimondo, F. and Gramfort, A. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 2017, vol. 159, p. 417–429. ISSN 1053-8119. Available at: https://www.sciencedirect.com/science/article/pii/S1053811917305013.

[41] Jw, B.; Lc, F.; Jl, H.; Korb, P.; Mz, K. et al. *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. January 2016. Available at: https://doi.org/10.5698/978-0-9979756-0-4.

[42] Kaboli, M.; De La Rosa T, A.; Walker, R. and Cheng, G. In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors. In: *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. 2015, p. 1155–1160.

[43] Kennedy, J. and Eberhart, R. Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. 1995, vol. 4, p. 1942–1948 vol.4.

[44] Krizhevsky, A.; Sutskever, I. and Hinton, G. *ImageNet classification with deep convolutional neural networks*. NEW YORK: ACM, 2017. ISSN 0001-0782.

[45] Langer, J. K.; Tonge, N. A.; Piccirillo, M.; Rodebaugh, T. L.; Thompson, R. J. et al. Symptoms of social anxiety disorder and major depressive disorder: A network perspective. *Journal of Affective Disorders*, 2019, vol. 243, p. 531–538. ISSN 0165-0327. Available at: https://www.sciencedirect.com/science/article/pii/S0165032718305573.

[46] Leon, F. *A Review of Findings from Neuroscience and Cognitive Psychology as Possible Inspiration for the Path to Artificial General Intelligence*. 2024.

[47] Li, Z.; Wu, X.; Xu, X.; Wang, H.; Guo, Z. et al. The Recognition of Multiple Anxiety Levels Based on Electroencephalograph. *IEEE Transactions on Affective Computing*, 2022, vol. 13, no. 1, p. 519–529.

[48] Liu, H.; Dai, Z.; So, D. R. and Le, Q. V. Pay Attention to MLPs. *CoRR*, 2021, abs/2105.08050. Available at: https://arxiv.org/abs/2105.08050.

[49] Lopez Gordo, M. A.; Sanchez Morillo, D. and Pelayo Valle, F. Dry EEG electrodes. *Sensors*. BASEL: Mdpi, 2014, vol. 14, no. 7, p. 12847–12870. ISSN 1424-8220.

[50] Malik, A. S. *EEG-based experiment design for major depressive disorder : machine learning and psychiatric diagnosis*. London ; San Diego: Elsevier ; Academic Press, 2019. ISBN 978-0-12-817420-3.

[51] Michael, J. A.; Wang, M.; Kaur, M.; Fitzgerald, P. B.; Fitzgibbon, B. M. et al. EEG correlates of attentional control in anxiety disorders: A systematic review of error-related negativity and correct-response negativity findings. *Journal of Affective Disorders*, 2021, vol. 291, p. 140–153. ISSN 0165-0327. Available at: https://www.sciencedirect.com/science/article/pii/S0165032721003815.

[52] Misulis, K. E. *Atlas of EEG, Seizure Semiology, and Management*. Second edition.th ed. United Kingdom: Oxford University Press, 2013. ISBN 9780199985913.

[53] Moshe, I.; Terhorst, Y.; Opoku Asare, K.; Sander, L. B.; Ferreira, D. et al. Predicting Symptoms of Depression and Anxiety Using Smartphone and Wearable Data. *Frontiers in Psychiatry*, 2021, vol. 12. ISSN 1664-0640. Available at: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.625247.

[54] Mrazek, V.; Jawed, S.; Arif, M. and Malik, A. S. Effective EEG Feature Selection for Interpretable MDD (Major Depressive Disorder) Classification. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. New York,

NY, USA: Association for Computing Machinery, 2023, p. 1427–1435. GECCO '23. ISBN 9798400701191. Available at: https://doi.org/10.1145/3583131.3590398.

[55] Mrazek, V.; Jawed, S.; Arif, M. and Malik, A. S. Effective EEG Feature Selection for Interpretable MDD (Major Depressive Disorder) Classification. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. New York, NY, USA: Association for Computing Machinery, 2023, p. 1427–1435. GECCO '23. ISBN 9798400701191. Available at: https://doi.org/10.1145/3583131.3590398.

[56] Mumtaz, W.; Xia, L.; Ali, S. S. A.; Yasin, M. A. M.; Hussain, M. et al. Electroencephalogram (EEG)-based computer-aided technique to diagnose major depressive disorder (MDD). *Biomedical Signal Processing and Control*, 2017, vol. 31, p. 108–115. ISSN 1746-8094. Available at: https://www.sciencedirect.com/science/article/pii/S1746809416300866.

[57] Ndetei, D. M.; Khasakhala, L.; Mbwayo, A. and Mutiso, V. *Epidemiological Patterns of Anxiety Disorders in Kenya*. August 2011. Available at: https://doi.org/10.5772/19184.

[58] Pineda, J.; Juavinett, A. and Datko, M. Rationale for Neurofeedback Training in Children with Autism. In:. January 2014, p. 439–460. ISBN 978-1-4614-4787-0.

[59] Rafiei, A.; Zahedifar, R.; Sitaula, C. and Marzbanrad, F. Automated Detection of Major Depressive Disorder With EEG Signals: A Time Series Classification Using Deep Learning. *IEEE Access*, 2022, vol. 10, p. 73804–73817.

[60] Richhariya, B. and Tanveer, M. EEG signal classification using universum support vector machine. *Expert Systems with Applications*, 2018, vol. 106, p. 169–182. ISSN 0957-4174. Available at: https://www.sciencedirect.com/science/article/pii/S0957417418302100.

[61] Rodríguez, P.; Bautista, M. A.; Gonzàlez, J. and Escalera, S. Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing*, 2018, vol. 75, p. 21–31. ISSN 0262-8856. Available at: https://www.sciencedirect.com/science/article/pii/S0262885618300623.

[62] Rojas, G.; Alvarez, C.; Montoya Moya, C.; Iglesia Vaya, M. de la; Cisternas, J. et al. Study of Resting-State Functional Connectivity Networks Using EEG Electrodes Position As Seed. *Frontiers in Neuroscience*, march 2018, vol. 12.

[63] Roy-Byrne, P.; Davidson, K. W.; Kessler, R. C.; Asmundson, G. J.; Goodwin, R. D. et al. Anxiety disorders and comorbid medical illness. *General hospital psychiatry*, may 2008, vol. 30, no. 3, p. 208–225. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0163834307002599.

[64] Sainte Marie, M.; Meunier, J.-G.; Payette, N. and Chartier, J.-F. *The concept of evolution in the Origin of Species: a computer-assisted analysis*. May 2011. Available at: https://doi.org/10.1093/llc/fqr019.

[65] Schmidt, R. M. Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. *CoRR*, 2019, abs/1912.05911. Available at: http://arxiv.org/abs/1912.05911.

[66] Schmidt, R. M. *Recurrent Neural Networks (RNNs): A gentle Introduction and Overview*. January 2019. Available at: https://arxiv.org/abs/1912.05911.

[67] Schmitt, A.; Malchow, B.; Hasan, A. and Falkai, P. *The impact of environmental factors in severe psychiatric disorders*. January 2014. Available at: https://doi.org/10.3389/fnins.2014.00019.

[68] Seal, A.; Bajpai, R.; Karnati, M.; Agnihotri, J.; Yazidi, A. et al. Benchmarks for machine learning in depression discrimination using electroencephalography signals. *Applied Intelligence*, May 2023, vol. 53, no. 10, p. 12666–12683. ISSN 1573-7497. Available at: https://doi.org/10.1007/s10489-022-04159-y.

[69] Servan Schreiber, D.; Cleeremans, A. and Mcclelland, J. Graded State Machines: The Representation of Temporal Contingencies in Simple Recurrent Networks. *Machine Learning*, september 1991, vol. 7, p. 161–193.

[70] Sharma, S.; Sharma, S. and Athaiya, A. Activation functions in neural networks. *Towards Data Sci*, 2017, vol. 6, no. 12, p. 310–316.

[71] Shikha; Agrawal, M.; Anwar, M. A. and Sethia, D. Stacked Sparse Autoencoder and Machine Learning Based Anxiety Classification Using EEG Signals. In: *Proceedings of the First International Conference on AI-ML Systems*. New York, NY, USA: Association for Computing Machinery, 2021. AIMLSystems '21. ISBN 9781450385947. Available at: https://doi.org/10.1145/3486001.3486227.

[72] Stancin, I.; Cifrek, M. and Jovic, A. A Review of EEG Signal Features and their Application in Driver Drowsiness Detection Systems. *Sensors (Basel)*. Switzerland: [b.n.], may 2021, vol. 21, no. 11.

[73] Storn, R. and Price, K. Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, january 1997, vol. 11, p. 341–359.

[74] Sun, L.; Liu, Y. and Beadle, P. Independent component analysis of EEG signals. In: *Proceedings of 2005 IEEE International Workshop on VLSI Design and Video Technology, 2005.* 2005, p. 219–222.

[75] Sun, M.; Song, Z.; Jiang, X.; Pan, J. and Pang, Y. Learning Pooling for Convolutional Neural Network. *Neurocomputing*, 2017, vol. 224, p. 96–104. ISSN 0925-2312. Available at: https://www.sciencedirect.com/science/article/pii/S0925231216312905.

[76] Sun, Y.; Xue, B.; Zhang, M.; Yen, G. G. and Lv, J. Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification. *IEEE Transactions on Cybernetics*. Institute of Electrical and Electronics Engineers (IEEE), september 2020, vol. 50, no. 9, p. 3840–3854. ISSN 2168-2275. Available at: http://dx.doi.org/10.1109/TCYB.2020.2983860.

[77] Tudor, M.; Tudor, L. and Tudor, K. Hans Berger (1873-1941)–povijest elektroencefalografije [Hans Berger (1873-1941)–the history of electroencephalography]. *Acta Med Croatica*, 2005, vol. 59, no. 4, p. 307–313.

[78] Tully, P. J.; Harrison, N. J.; Cheung, P. and Cosh, S. Anxiety and Cardiovascular Disease Risk: a Review. *Current Cardiology Reports*, Oct 2016, vol. 18, no. 12, p. 120. ISSN 1534-3170. Available at: https://doi.org/10.1007/s11886-016-0800-3.

[79] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L. et al. *Attention Is All You Need*. 2023.

[80] Wang, G.; Shepherd, S. J.; Beggs, C. B.; Rao, N. and Zhang, Y. The use of kurtosis de-noising for EEG analysis of patients suffering from Alzheimer's disease. *Biomed Mater Eng*. Netherlands: [b.n.], 2015, 26 Suppl 1, p. S1135–48.

[81] Weinberg, A. and Creed, F. *Stress and psychiatric disorder in healthcare professionals and hospital staff*. February 2000. Available at: https://doi.org/10.1016/s0140-6736(99)07366-3.

[82] Xie, L. and Yuille, A. L. Genetic CNN. *CoRR*, 2017, abs/1703.01513. Available at: http://arxiv.org/abs/1703.01513.

[83] Youme, O.; Dembele, J. M.; Ezin, E. C. and Cambier, C. Evolution under Length Constraints for CNN Architecture design. In: *Proceedings of the 2023 7th International Conference on Digital Signal Processing*. ACM, February 2023. ICDSP 2023. Available at: http://dx.doi.org/10.1145/3585542.3585546.

[84] Zhang, C.; Kim, Y.-K. and Eskandarian, A. *EEG-Inception: An Accurate and Robust End-to-End Neural Network for EEG-based Motor Imagery Classification*. 2021.

Appendix A

Results and Graphs

For mixed 1 seconds is accuracy **0.6838831291234684**. List the best features identified by the SVM model for 1 seconds segment: **decorr_time**, **higuchi_fd**, **hjorth_complexity**, **hjorth_complexity_spect**, **hjorth_mobility**, **hurst_exp**, **mean**, **pow_freq_bands**, **spect_entropy**, **spect_slope**, **time_corr**.

Class	Precision	Recall	F1-Score	Support
0	0.72	0.71	0.72	2665
1	0.50	0.01	0.01	535
2	0.30	0.42	0.35	645
3	0.83	0.99	0.90	1460
Accuracy		0.68		
Macro Avg	0.59	0.53	0.50	5305
Weighted Avg	0.68	0.68	0.65	5305

Table A.1: Validation Class Report for 1 second SVM$_G A$

Class	Precision	Recall	F1-Score	Support
0	0.64	0.62	0.63	2545
1	0.92	0.04	0.08	535
2	0.41	0.54	0.47	720
3	0.63	0.82	0.72	1250
Accuracy		0.60		
Macro Avg	0.65	0.51	0.47	5050
Weighted Avg	0.63	0.60	0.57	5050

Table A.2: Test Class Report 1 second SVM GA.

Class	Precision	Recall	F1-Score	Support
0	0.72	0.64	0.68	597
1	0.00	0.00	0.00	131
2	0.30	0.55	0.38	167
3	0.84	1.00	0.91	292
Accuracy		0.64		
Macro Avg	0.46	0.55	0.49	1187
Weighted Avg	0.61	0.64	0.62	1187

Table A.3: Validation Class Report for 3 second segment SVM GA.

Class	Precision	Recall	F1-Score	Support
0	0.63	0.57	0.60	557
1	0.00	0.00	0.00	131
2	0.40	0.62	0.49	192
3	0.63	0.82	0.72	250
Accuracy		0.57		
Macro Avg	0.42	0.50	0.45	1130
Weighted Avg	0.52	0.57	0.54	1130

Table A.4: Test Class Report for 3 second segment SVM GA.

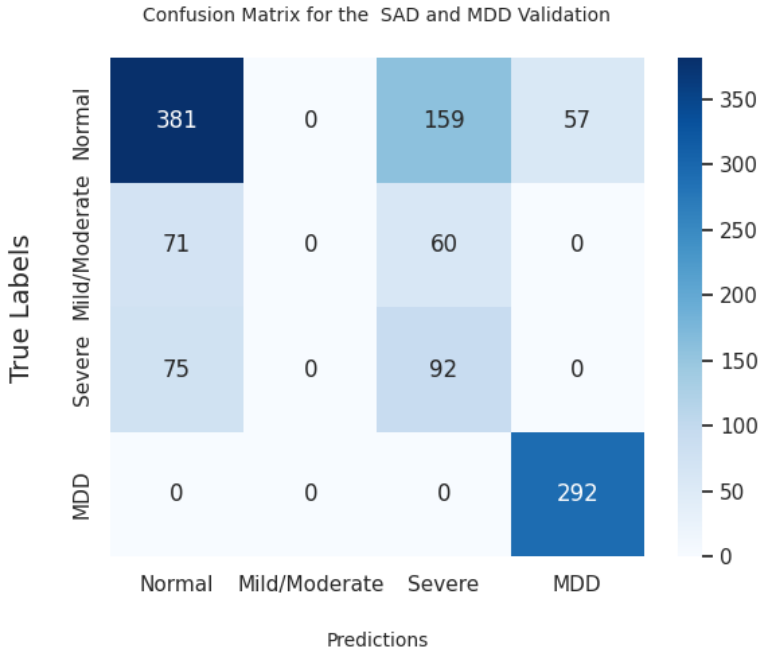

Figure A.1: Confusion Matrix for Validation Data in Mixed 3 sec SVM GA.

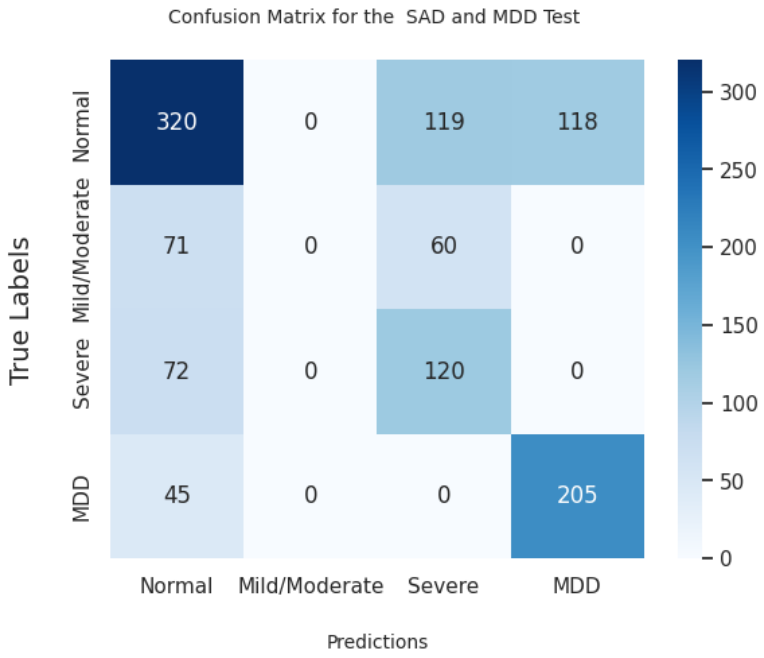

Figure A.2: Confusion Matrix for Test Data in Mixed 3 sec SVM GA.

Table A.5: Validation Class Report

Class	Precision	Recall	F1-Score	Support
0	0.73	0.71	0.72	533
1	0.00	0.00	0.00	107
2	0.30	0.44	0.36	129
3	0.84	1.00	0.91	292
Accuracy		0.69		
Macro Avg	0.47	0.54	0.50	1061
Weighted Avg	0.63	0.69	0.66	1061

Table A.6: Test Class Report

Class	Precision	Recall	F1-Score	Support
0	0.63	0.63	0.63	509
1	0.00	0.00	0.00	107
2	0.40	0.50	0.45	144
3	0.63	0.81	0.71	250
Accuracy		0.59		
Macro Avg	0.42	0.48	0.45	1010
Weighted Avg	0.53	0.59	0.56	1010

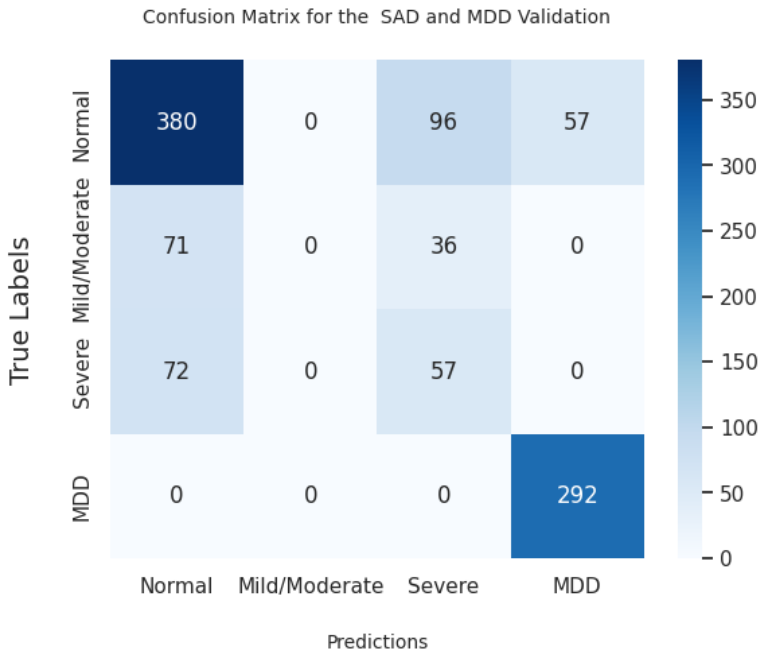

Figure A.3: Confusion Matrix for Validation Data in Mixed 5 sec.

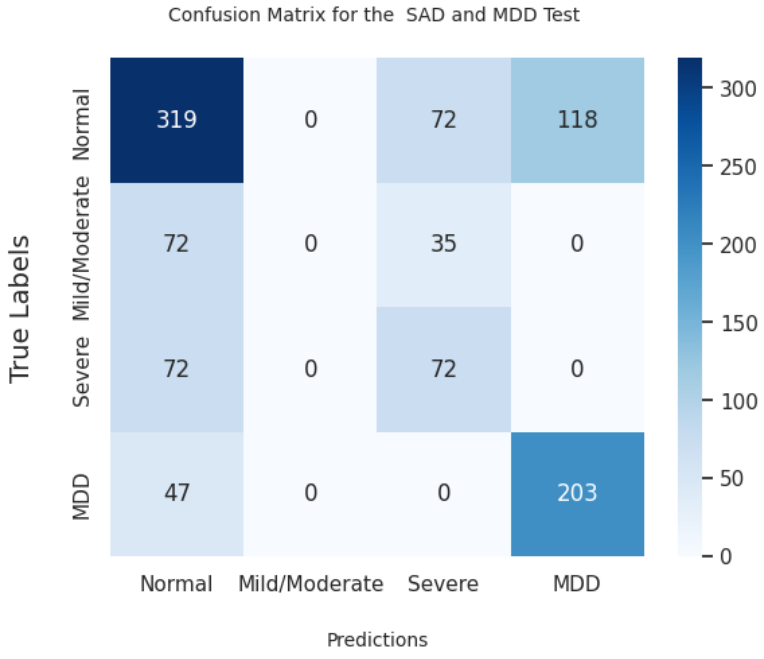

Figure A.4: Confusion Matrix for Test Data in Mixed 5 sec.

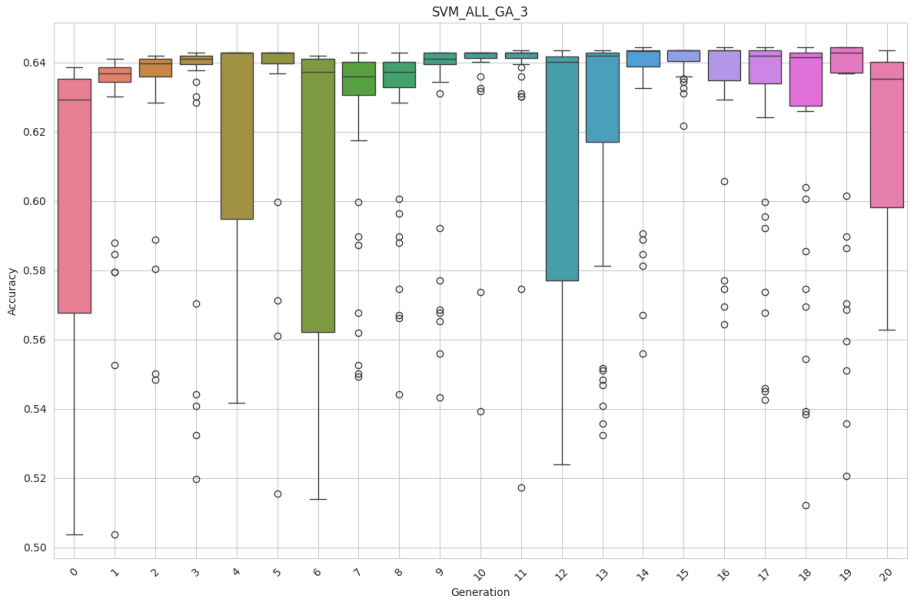

Figure A.5: SVM_ALL_GA_3

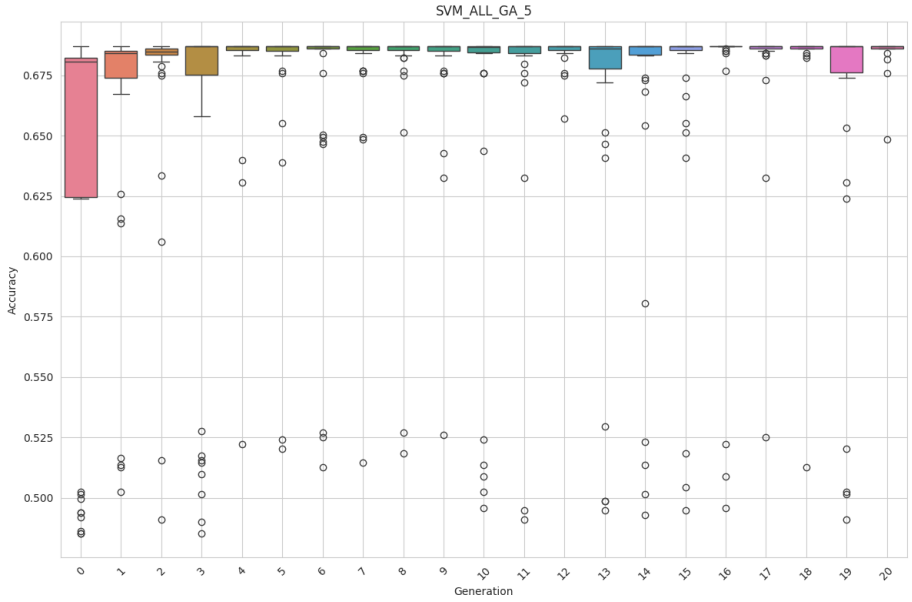

Figure A.6: SVM_ALL_GA_5

Appendix B

SD card structure