Depression Classification with MDD (Major Depressive Disorder) using Signal Processing and Machine Learning

by

Shahriar Saleque
19341016
Gul-A-Zannat Spriha
16301089
MD Rasheeq Ishraq Kamal
16101074
Rafia Tabassum Khan
16301081

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
April 2020

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Shahriar Saleque
19341016

Gul-A-Zannat Spriha
16301089

MD Rasheeq Ishraq Kamal
16101074

Rafia Tabassum Khan
16301081

Approval

The thesis titled "Depression Classification with MDD (Major Depressive Disorder) using Signal Processing and Machine Learning" submitted by

1. Shahriar Saleque (19341016)
2. Gul-A-Zannat Spriha (16301089)
3. MD Rasheeq Ishraq Kamal (16101074)
4. Rafia Tabassum Khan (16301081)

Of Summer, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on April 7, 2020.

Examining Committee:

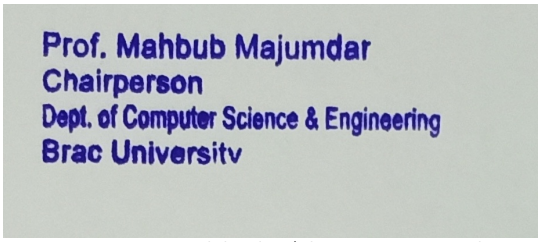
Supervisor:
(Member)

Dr. Amitabha Chakrabarty
Associate Professor
Department of Computer Science and Engineering
Brac University

Co-supervisor:
(Member)

Dr. Mohammad Zavid Parvez
Assistant Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)



Dr. Mahbub Alam Majumdar
Professor
Department of Computer Science and Engineering
Brac University

Abstract

Depression is accorded as one of the leading causes to all the problems related to mental health in the Global disease burden study (GBD). Major depressive disorder (MDD) is when this depression reaches to a larger extent, when depression persists for two weeks or more. Sadly, many individuals of our society tend to neglect depression and refuse to label it as a mental disease and has a tendency to not seek medical help. Not only this, they are being curbed because of the few or very limited biological indicators for MDD and depression identification. Our main objective is to develop a non-intrusive approach that will detect and differentiate brain signals of patients with MDD from healthy patients. We were able to obtain an optimized model with an accuracy of (82%). Primarily we obtained a raw EEG data-set upon research and since it matched our requirements, we performed noise removal on them. Afterwards we extracted relevant features for depression detection, such as one feature was Absolute delta power. Finally, we entered these features into three classification algorithms; Logistic Regression (LR), Support Vector Machine (SVM) and Naïve-Bayes (NB). To check the accuracy and precision, we performed a ten-fold cross validation on them. Hopefully, our results will encourage and motivate people suffering from this to seek the proper and effective medical help and to eradicate the negative stigma around it.

Keywords: MDD; Absolute delta power; LR; SVM; NB; EEG.

Dedication

We would like to dedicate this thesis to our loving parents and all amazing faculties we encountered and learnt from in the course of pursuing our Bachelor's degree.

Acknowledgement

First and foremost, praises and thanks to Allah, the Lord of the entire universe, for showering His blessings on us, giving us the perseverance, throughout our research work so that we could complete it successfully.
We would like to express our deep and sincere gratitude to our supervisor Dr. Amitabha Chakrabarty for giving us the opportunity to do this study and thesis and for his continuous support and guidance all over the thesis work. Be it his patience, enthusiasm, motivation, vision or insightful comments, he enormously inspired us to learn more and know more. Then we would like to thank our co-supervisor Dr. Mohammad Zavid Parvez, who never hesitated to give his precious time to us whenever we failed to understand anything. It was a great honor and privilege to work under their direction. Furthermore, we would like to show our appreciation to the Department of Computer Science and Engineering for providing an amazing working atmosphere that excites learning and also for providing resources that aided our project.
Last but not the least, we would express our love and regards to our parents and friends, who were our pillar of never-ending moral-support. We would not have come this far and be proficient in it without any of them.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI Artificial Intelligence

AUC Area under the curve

CNS Central nervous system

DFA Detrended Fluctuation Analysis

EC Eyes Closed

ECG Electrocardiography

$ECOG$ Eastern Cooperative Oncology Group

$ECoG$ Electrocorticography

EDF European Data Format

EEG Electroencephalography

EO Eyes Open

FA Fractional anisotropy

$fMRI$ Functional magnetic resonance imaging

FPR False positive rate

GP Grassberger and Procaccia

H Healthy Controls

ICA Independent Component Analysis

IDE Integrated development environment

KNN K-nearest neighbors

LR Logistic Regression

MDD Major Depressive Disorder

MRI Magnetic resonance imaging

NB Naïve-Bayes

PCA Principle Component Analysis

$PDDI$ Persistent depressive disorder

$PMDDI$ Premenstrual dysphoric disorder

PMS Premenstrual syndrome

PNS Peripheral nervous system

PSD Power Spectral Density

RF Beck depression inventory

RF Random forest

ROC Receiver operating characteristic

SAD Seasonal affective disorder

$sEEG$ Stereo-electrocenphalography

SVM Simple Vector Machine

TN True non-cases

TP True cases

TPR True positive rate

WHO World Health Organization

$WOSA$ Weighted overlapped segment averaging

$YASA$ Yet Another Spindle Algorithm

Chapter 1

Introduction

This chapter introduces the topic,addresses the problem and provides a brief overview of the research, the purpose, goal and method of solving it through this analysis.

Depression is a common yet serious mood disorder, which might cause continual feeling of sadness, loss of interest and energy, deteriorating one's memory and concentration. Even if we are suffering mentally, as long as the pain does not appear externally people dodge it. Major depressive disorder occurs when there is a frequent emotional dysregulation with the existence of continual sorrow, a feeling of losing your own confidence and self-worth. This also includes feeling somewhat fidgety and gradually losing enthusiasm in your daily activities for a long period [39]. As a result it hampers any sort of decisions, choices and attention at any sort of work [44]. World Health Organization (WHO) provides statistics that indicate over 350 million people are victims of major depression [9] and many of them have a tendency to lead to suicidal reactions. Although MDD is alarming for one's mental health, psychiatrists can only take the aid of psychotropic drugs and rely on hypnotherapists to heal an MDD patient [29], [31]. Adding to this, there is a lack of availability of any proper laboratory analysis [39]. So WHO predicted how by 2030 depression is going to be the leading cause of the disease burden [35] and it is condemnatory to treat depression and save lives [48]. If we look at intermediate phenotype indicators, we can figure out how MDD and abnormality in someone's brain structure are interconnected. Resolutely, it is an effective step to measure brain activity which will help us to elaborate MDD into more comparable markers to find a constructive cure for it [33], [69]. A study by Hossain et al talks about how the growth of mental disorders of Bangladeshi children has risen from 13.40% to 22.9% between 1998 to 2004 [51], [59]. Around an estimate of 4.6% is the frequency of occurrence for the depressive disorder [25]. To add to this, a research at Dhaka university [70] says this depression can destroy one's hopes and will to live completely (especially for the adolescents) and ends up in suicidal activities. Even then, there are only 210 psycho-analysts for a population of 210 million people [20], [89]. Therefore, due to lack of mental health facilities, they fail to receive any therapy. This paper is principally engrossed on showing differences in brain signals emitted from Healthy people and MDD patients and hence classifying depression as a disease as real and concrete as cancer.

1.1 Motivation

Generally, any and every indicator acts like a probable pointer to recognize the countermeasure of ill-health when they anticipate it. However, there has not been one such indicator related to depression that has entirely exemplified their caliber to distinguish what specifically is irregular in a depressed patient's brain and how they can collaborate it with their medical resources to boost their medical handlings [78]. Moreover, it is found in the WHO 2012 report that when an individual decides to go through with the suicide, at least 20 more follow them to end their precious lives as well [42]. Surprisingly women are said to be more likely to be depressed when they are expectant. As a result, the depression is going to traumatize both the unborn child and their mothers. During 2005, in a rural sub-district of Bangladesh, some people carried out a public analysis of around 110,000 pregnant ladies and narrowed down 361 depressed conceiving women [24]. 14% of these shortlisted women even had dreadful thoughts about self-harm during this phase [24]. Therefore, depression is no longer a trivial issue and must be taken care of with proper measures. Ultimately, we decided on using the EEG signals of the brain to detect these explicit problems in the brain that can sort out the MDD patients from the control patients.

1.2 Thesis Overview

EEG signals are progressively substantial in any neurocyte related diseases concerning any sort of brain impairment. By the aid of these signals' psychiatrists can carry out an appropriate treatment of those diseases [72]. Instead of concentrating on the typical symptom-based methods of judging depression we therefore used EEG signals to identify it. Initially, we obtained data from both healthy and MDD candidates with their eyes opened and closed. EEG data must be conducted in a closed room with no noise or disturbance. Hence, the data had to undergo, Principle Component Analysis (PCA) and Independent Component Analysis (ICA) to get rid of the noise and interferences. Secondly, we had to gather extensive and accurate medical domain as to which features that can be extracted from the brain signals and of them the ones which are relevant to our particular field of interest, i.e. depression detection. We used two features: Alpha-Interhemispheric Asymmetry, which in simple terms is the modulated difference between the band-powers of the signals emitted from the left and right hemisphere and Absolute power for each frequency band (alpha, beta and gamma), focusing on the latter feature extensively. These relevant features are given as input to the classification algorithms; Logistic Regression, Simple Vector Machine (SVM) and Naïve-Bayes. We had to use a ten-fold cross validation to determine how accurate each algorithm used were and if they were valid for our purpose. The aim of our research work is to help reduce the dread surrounding depression by precisely showing differences in brain signals emitted from Healthy people and MDD patients. We also encourage development of a positive mindset in our current society by developing an extensive and effective way of diagnosing MDD. Consequently, our results will encourage and motivate people suffering from this disease to seek the proper and effective medical help and thereby aid our society as a whole to advance towards a brighter and enlightened future.

1.3 Thesis Orientation

The following chapters of this paper are arranged in the following order consecutively. First in chapter 2, we talk about the preceding workings and examinations based on this method and other similar attempts, also their drawbacks based on which we came up with our approach. Chapter 3 elaborately deals with the contextual information that we gathered up to proceed with our work. Based on those comes chapter 4 where we talk about our selected model and how we set up the working environment comprehensively. Chapter 5 deals with the results we achieved and the accuracy limits of it and findings interrelated to it. Chapter 6 sums up the paper with our final thoughts over the topic and how we will use it in the future.

Chapter 2

Literature Review

This chapter talks about other papers and studies which are directly or indirectly related to our research topic. It also includes the overall limitations of the former studies which led us to the reason of choosing our topic.

Previously, there has been multiple studies, which have attempted to integrate the EEG signal with machine learning to distinguish healthy patients from the depressed ones [43], [56]. Some talk about the dimensional difference in the EEG spectrum between the two groups [26], [32]. Hosseinifard et al did a study about accuracies in the linear discriminant analysis of the brain bands of alpha, beta, delta and theta with good precision rates of 73%, 70%, 67% and 70% respectively [46]. Furthermore, he performed an identification method on the basis of these different features: EEG spectral power, DFA, Higuchi fractal dimension, Grassberger and Procaccia (GP) algorithm-based correlation dimension, and Lyapunov exponent to group healthy and depressed patients, and the GP algorithm based correlation outshined all the other features with a accuracy of 83.3% [46]. Blum on the other hand examined the early stage of depression using kernel region of interest [23]. The neuroimaging proved how a depressed person has lack of interest compared to a controlled person. Neuro imaging also highlights the abnormal brain structures in a depressed patient which was proven by Vederine et al. (2011) that how a depressed person has an abnormal front fornix system [37]. The idea was further extended by Kwaasteniet (Kwaasteniet et al., 2013) [47] who deep-rooted this by making sure how the nervous tissue in the brain which connects the pre frontal with the limbic system was dysfunctional for a depressed victim by performing fMRI and FA on them. Also in 2011 Zhang with his colleagues worked with fMRI on how a non-habitual first episodic MDD victim has a distorted brain connectivity [40]. Therefore, MDD can damage the global regional anatomy of one's entire brain network. Again, Zeng worked with functional connectivity of the brain MRI, in order, to reach an objective indicator for MDD, but the limitation was he used an unsupervised classification [55]. In another paper we found out how two other approaches where combined with machine learning singleton and dual to detect depression [86]. However, they used a different classifier technique- Random forest(RF) with two threshold values and another one had 2 separate classifier, one focusing on to figure out who were depressed and the other to find the non-depressed ones [86]. However, the features they attempted to use were based on textual and syntactical data but proved to weak in comparison. Feature selection methods should be carefully shortlisted and chosen relevantly, since

they forecast the future performance and guide us with pattern recognition [49].
Some people used other unique analysis methods like Lee et al, who performed a detrended fluctuation analysis (DFA) [19]. The statistical self-affinity of a signal is calculated by this method. This method comes handy for time series with a long memory process and taking advantage of this fact he took 11 control candidates and 11 MDD candidates and found out a higher DFA for the MDD patients [19]. Li et al. took out 20 types of EEG features out of 128 channels and shortlisted the best 20 features out of the total 2560 channels, and performed k-nearest neighbor classification method on them (KNN) [60]. He found a good accuracy of 99.1% [60] but his data-set was limited to only clinically nine depressive and nine clinically non depressive patients, so that was a drawback. Depression even causes brain signal imbalance. Deslandes et al. evaluated the EEG signal asymmetry of a depressed individual and a normal individual and how has depression destroyed one's quality of living in general [21]. Furthermore, Nandrino et al. shows how the complications in a depressed patients brain activity affects his communication with the environment [8].
Other papers have also stated how detecting MDD at the initial stages can prove to come in useful [15]. Halfin's study proved this how identifying the depression in early stages can not only lessen the financial and mental load of a patient but also help in the cure of this disease [18]. Rost et al talked about how early screening of depression can help employers with their work boosting their productivity and quality in the work field [15]. Picardi et al also noticed progress in the symptoms of depression and also in the lives of the subjects in general when the depression had been pinpointed in the preliminary stages [71].
There were quiet a few studies regarding the early identification of depression. Ophir et al directed on the coding scheme and found some depression signals on the juvenile subjects, the users of Facebook, Inc to measure depression in the earlier stages [77].De Choudhury et al with the help of Twitter, Inc users involvement patterns and semantics markers found a tool for detecting MDD by doing a comparison on the scores of Center for Epidemiologic Studies Depression Scale [2] and BDI [12] and gained an accuracy of 70% [47].These study successfully found out a lot of terms that surrounds depression and goes on the minds of a depressed subject, the emotional breakdowns, the isolation, the day to day negativity caused in them because of this disorder. However, there were some drawbacks in this study. Firstly, it only was based on the cases where the subjects themselves reported their issue. Unfortunately, there are a lot of people who are yet uncertain that they are depressed and this study fails to target those people [75]. Secondly, and finally this study did not diagnose the depression in the initial stages.
Nowadays, technology has advanced a lot and almost everyone is glued to the social media. Therefore, it will be a smarter approach to use the social networks as a tool to figure out the abnormalities of the brain and to specify the prevalence of the mental disorders. Prieto et all came up with the idea to use Twitter, Inc app to dynamically detect the occurrence of the mental disorders [52]. Since a vast number of people use this application, during 2010, Chunara et al formed a chain of tweets relating to the Haitian cholera outbreak of the initial 100 days [41]. Chew and Eysenbach did a investigation on 2 million tweets on the basis of the emotional aspect of the subjects to create a Syndromic Surveillance [30]. Aladag et al has noticed a similar post pattern with similar hashtags regarding the suicidal attempts

of the youth [79]. In order to verify this point Rice et all highlighted how much effective with is when there is the involvement of a huge mass of people in a cost friendly platform, like the social platforms [53]. Also, studies guarantee that the youth starting these sorts of thread can give favorable results.
Not only this, other miscellaneous findings support the claim of how these social media platforms anticipate mental dysfunctions like depression [57], [64]. De Choudhury et al has proposed an arithmetical method determining unique markers related to suicide from Reddit, Inc which acts as a model of what to expect beforehand, regarding mental health problems [65].Birnbaum on the other hand, found social media markers related to schizophrenia combining his knowledge of machine learning with medical equipment [74].
We also came across studies where they specifically focused on depression not mental disorders, but not specifically on MDD. For instance:- De Choudhury et al created a social media depression index after interconnecting various social activity of emotional perspective and the language signals from Twitter, Inc [65]. Using topic modeling and rule-based methods, there was a task performed at the Computational Linguistics and Clinical Psychology Workshop 2015, regarding depression and mental health and it achieved really good feedback [58], [62].
Though there has been a lot of former works done on MDD with different classification methods, all of them has a lot of space for improvement. The major limitations of these are how much long time we have to invest for the validation of EEG signals and the lengthy preparation time for the recording sites. Not only this, we then have to perform noise reduction and other filtering processes. Longer EEG recording sessions will make our subject feel monotonous and he may fall asleep which will add up to the problem.
Our main advantage of the study is that we focus on MDD. MDD in itself is more severe than any depression in the prior stages. So, the topic is more narrowed down. Also, the recording time used in this data set [68] was for 5 minutes which is comparatively a short time, giving the signals less room for noise and interference. Also, to filter the noises we used an advanced technique (ICA and PCA).Usually, in other papers they generally use BESA software for noise reduction. Although, BESA software is user-friendly but it is very much outmoded and nowadays no one uses it due to its lack of flexibility. Whereas, ICA and PCA techniques are very useful and gives precise results. ICA deals with separating unconventional signals from the complicated ECG signals and PCA works with dimensionality reduction techniques to appropriately perform feature extraction [34]. Not only this the classification methods that we used outshines other old classification methods due to its usefulness.Support Vector Machine classifies the best when the classes are appropriately distinguised with a clear border, giving its best performance in spaces which are high in magnitude [91]. SVM will exceedingly surpass other techniques, when number of samples are less than the dimension values being good with memory [91]. LR was chosen since it is easier to understand and can be trained effectively and can be executed without any complexity. Also when there are high dimensional values we can overfit it [108]. Also, it tells us if the values are positive or negative along with giving its relevancy to us [108]. Finally we worked with Naive Bayes classifier because it is very much extensible with the capability to make probabilistic prognosis [101]. It can even deal with missing values and is super fast. SVM can work more efficiently compared to other algorithms if the assumption of independence is

true and therefore we decided to work with it [101].
Overall, our target is to attain a higher accuracy with shorter recording time and few EEG recording sites. We hope to achieve that by our classification methods used in this paper.

Chapter 3

Background Study

In this chapter, we conduct an extensive review of the context data linked to our research. In section 3.1, we study depression as the central aspect of our work. Section 3.2 gives us the idea about basic features and structure if the brain. Functions of human brain is given in section 3.3. The next section 3.4 is about brainwaves and their types, like Alpha, Beta, Gamma, Delta, Theta and section 3.5 is a brief description about EEG. Section 3.6 explains the different algorithms that has been used in relation to the EEG Analysis of the brain, such as LR, SVM and NB. Description of 10 fold cross-validation is given in the following section.

3.1 Depression and its classification

Depression is a prevalent illness worldwide and over 264 million people are affected by it [76]. It affects an average one in 15 adults in any given year (6.7 per cent) and one in six people (16.6 percent) at some point in their lives will experience depression [14]. Research shows that about one-third women suffer from depression at least once on their entire lives. A loved one's death, loss of a job or termination of a relationship are traumatic experiences a person may endure which is natural to develop feelings of sorrow or grief.These can be often described as depression as grieving process might have same features of depression but being sad and having depression is not similar. Depression typically makes people feel unworthy and thus despise themselves.
Major depression, persistent depressive disorder, bipolar disorder, and seasonal affective disorder are the four most common forms of depression according to Harvard Medical School. Other than these four types, there are two more types which are unique to women. Each of the types will be discussed briefly in the following section.

3.1.1 Major depression

Major Depressive Disorder (MDD), more commonly known as clinical depression, is the most severe form of depression. MDD is a psychiatric condition that involves affective and mood disturbances, neurovegetative functions (such as appetite and sleep disorders), cognition (such as excessive remorse and feelings of worthlessness), and psychomotor dysfunction (such as distress or mental impairment). People with MDD might have an all consuming dark mood and have suicidal tendency in some cases [102].

3.1.2 Persistent depressive disorder

It is formerly known as dysthymia [5]. It is less severe than any other type of depression but the duration can be long-lasting such as two years or more. According to Diagnostic and Statistical Manual of Mental Disorders, two prior diagnosis merge the persistent depressive disorder: chronic major depressive episode and dysthymic disorder. Deep inside many people with this disorder feel low most of the time nonetheless they are the able to work normally [81].

3.1.3 Bipolar disorder

Bipolar disorder (also known as manic-depressive syndrome or bipolar depression) is a psychiatric illness. Bipolar disorder contributes to erratic mood shifts, changes in motivation as well as changes in the ability to perform day to day tasks and activity levels. People with bipolar disorder may experience significant mood swings like having high-energy manic episodes to low-energy depressive episodes [5],[102].

3.1.4 Seasonal affective disorder (SAD)

A seasonal depression that occurs annually, usually beginning in autumn, worsening in winter and ending in spring. People having this disorder also known as winter blues can experience lethargy, depression, and the feeling of never going to be warm or see the sun again. Women tend to develop this disorder than men [81].

While women are at increased risk of general depression, they experience two more different forms of depression because of the reproductive hormones.

3.1.5 Perinatal depression

Being pregnant is never easy. Suffering from morning sickness, mood swings, gaining weight etc. are common for pregnant women. But perinatal depression is much more serious than these common illness. It can occur during pregnancy or after giving birth. Extreme sadness or anxiety lead to this depression and can have devastating effects on the women,their families and even on the infants [5].

3.1.6 Premenstrual Dysphoric Disorder (PMDD)

It is a severe form of premenstrual syndrome, or PMS. It is a serious debilitating disorder with symptoms such as lethargy, wrath, depressed mood, depression, suicidal thoughts and bloating [81].
We have chosen to work on MDD patients as it is the most fatal form of depression out there.

3.2 Basic Structure of the Human Brain

The human brain is one of our body's most diverse and complex organs, an essential part of our nervous system. It is situated within the head, protected by the skull and near all other neural organs. The brain is the center of all intelligence,

interpreting information and knowledge from the outside world through the senses of sight, smell, hearing, taste and touch. Furthermore, controlling body movements and regulating all of the body's functions. Also, it is responsible for the control of behavior essentially containing and representing the mind and soul. Intelligence, imagination, creativity, thoughts, emotions, empathy, memory and breathing are just a few of the plethora of aspects that the brain controls. In other words, it is the essence of all the attributes that characterize and define humanity. The Brain consists of billions of neurons that interact in trillions of links, in order to control most of the functions of our body. The brain and spinal cord make up the central nervous system (CNS). Spinal nerves branching out of the spinal cord and cranial nerves branching out of the brain form the peripheral nervous system (PNS) [106]. All the components of the brain function together but each part has its own special properties. The brain may be divided into three fundamental units: Forebrain, Midbrain, Hindbrain [28].

3.2.1 Forebrain

The Forebrain is the largest part of the brain and mainly consists of the Cerebrum and it is linked to higher functions of the brain such as the interpretation of touch and learning. The cerebrum is made up of right and left hemispheres. The two cerebral hemispheres interact with each other, following the separation, through a dense tract of nerve fibers called the corpus callosum. Even though, the two hemispheres appear to be identical to each other, they are different and they control opposite parts of the body. They do not share all the same functions, in fact, the left hemisphere controls speech, writing, cognition, calculations while the right hemisphere controls imagination, creativity, artistic or musical skills [106].
The gray surface consists of nerve cells. White nerve fibers are found below the surface and they transfer signals between nerve cells. Each half of the cerebrum can once again be split into four key parts called lobes as we can see in the Figure 3.1 below: Frontal lobe, Parietal lobe, Occipital lobe and Temporal lobe [28].

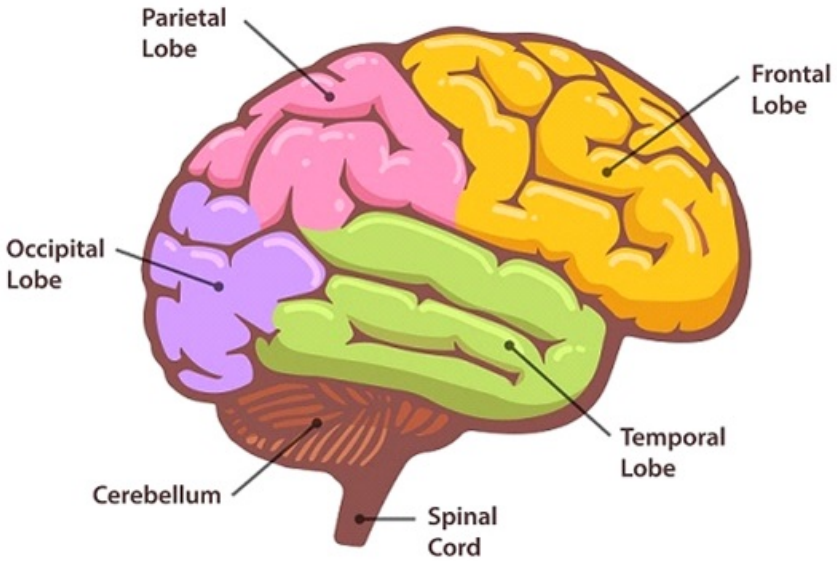


Figure 3.1: The forebrain[105]

Frontal Lobe

The frontal lobe lies just beneath our forehead. It is involved in making our brain to think, making facial expressions, solving problems, arranging, organizing, paying attention, intelligence, preparing and planning, speaking, spontaneity, controlling inhibitions, recalling and managing emotions. The motor area, residing on the rear end of each frontal lobe controls voluntary movement. Thoughts come out as words or speech due to Broca's area which is located on the left frontal lobe [98].

Parietal Lobe

It is located to the rear of the frontal lobe and happens to control many of the complex behaviors exhibited by humans which includes the senses such as vision, hearing, touch, temperature, pain and spatial orientation. As a further matter, it also plays a role in out spatial and sensory perception and processing, ability to build, manipulation of objects, positioning and movement of the body, self-awareness, understanding of numbers, language comprehension. The parietal lobe performs a vital function of integrating the various sensory information from different parts of the body [28].

Occipital Lobe

It is located at the back of the brain. It receives information from the eye and is concerned with our visual processing [98]. It interprets color, light, visual recognition, spatial awareness, perception of body language through vision such as movements, gestures, expressions and postures [28].

Temporal Lobe

The temporal lobe resides on the side of the brain near our ears. It is responsible for receiving information from the ears and it deals with processing such auditory stimuli, to understand the spoken language, linguistic development, verbal memory, visual memory and general information [28].

3.2.2 Midbrain

The Midbrain is located just below the cerebrum and is found in the center of the cranial cavity. It happens to be the smallest part of the brain. This consists of tectum, tegmentum, cerebral aqueduct, peduncles of the cerebrum and several nuclei and fasciculi. The primary purpose of this is to serve as a relay station for the visual and auditory systems in humans.

3.2.3 Hindbrain

The Hindbrain includes the upper part of the spinal cord, the brain stem, and the cerebellum. It is associated with controlling vital functions of the body such as breathing and digestion [98].

3.2.4 Brain Stem

The brain stem is found below the cerebrum. It functions as an interfacing system that connects the cerebrum and cerebellum to the spinal cord. It carries out automatic tasks of the body, for example, cardiac rhythm, temperature of the body, breathing, coughing and sneezing, digestion [85]. It consists of three parts: Midbrain, Pons and Medulla Oblongata [28].

3.2.5 Cerebellum

It is located beneath the cerebrum, posterior to the brain stem. It also has two hemispheres like the cerebrum and takes the form of a wrinkled ball tissue due to its highly folded surface. Its primary function is to coordinate motor control of the body like muscle movements, maintaining posture and balance.

3.3 Brain Functions

The brain along with the remainder of the nervous system comprises of many different types of cells, however, neurons are the main functioning entity. Billions of interconnected neurons make up the brain and these very neurons relay information by way of electrochemical pulses as Figure 3.2 shows. All stimuli, gestures, perceptions, experiences, memories, thoughts and emotions are the product of messages flowing through these neurons. Neurons come in many shapes and forms, nevertheless, they all consist of a cell body, dendrites and an axon. Synapses are tiny gaps over which neurons transmit their energy [28]. Dendrites function as antennae gathering messages from other nerve cells and passing them on to the cell body. Whether the messages should be relayed is determined by the cell bodies. The signals then may pass from the cell body to an axon from which it may pass into another neuron, a muscle cell, or cells in another target organ. Alongside neurons, glial cells also exist and they provide the neurons with nourishment, protection and structural support. They ensure synapses are transmitted properly and also help in repairing in the nerves [106].

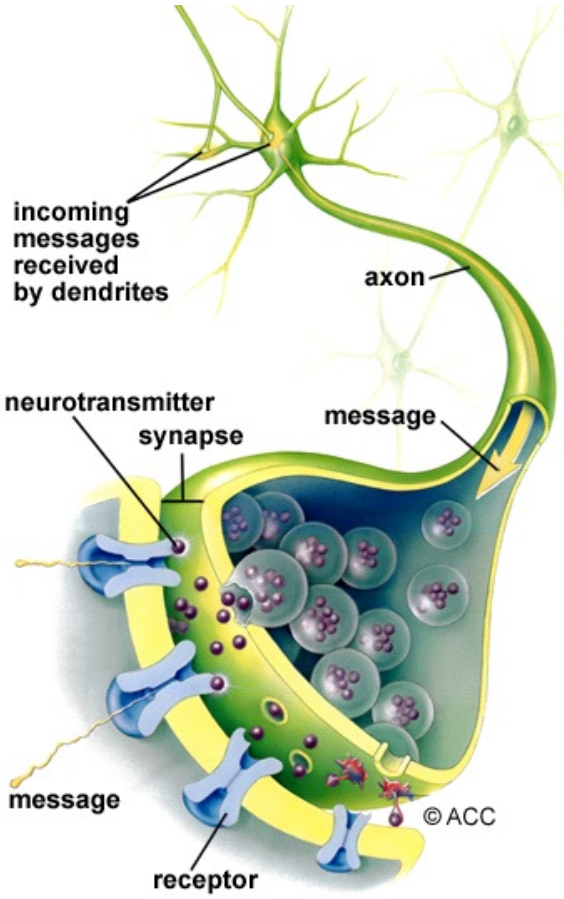


Figure 3.2: Functions of neurons in Brain [106]

3.4 Brain Waves

Our brain is an electrochemical organ, consisting of billions of neurons where in turn each neuron is connected to a further of thousand other neurons. The communication that takes place between the massive network of neurons within our brain is the foundation of all our feelings, thoughts, emotions and behaviors. Masses of neurons communicating with each other generate electrical pulses. Brainwaves are the result of these synchronized electrical activity. Such electrical activity can be detected from outside the brain by using sensors (electrodes) on the scalp. This is what is known as electroencephalography, or EEG. This allows us to take readings of the aforementioned electrical activity and amplify it for further analysis and visualization. The speeds of brainwaves are measured in Hertz (cycles per second).
Different brainwaves reflect upon the different states of mind. Some of these brain waves are more easily detectable than others. The brain comprises of different regions, which are specialized to perform certain tasks and processes. Therefore, certain brain waves correspond to certain parts of the brain or certain networks communicating with each other within the brain [85]. The amplitudes and frequencies are used to identify different brainwave patterns. These brainwaves can subsequently be classified according to their frequency or activity levels. Brainwaves are divided into five types or bandwidths: Gamma, Beta, Alpha, Theta, Delta. Depending on what we are feeling or doing at any particular moment, our brainwaves can change. We

may feel tired, lazy, sluggish or dreamy when brainwaves that are slower dominate. When we feel wired, or hyper-alerted that means higher frequencies are dominating [73]. Slow activity means lower frequency and high amplitude. Fast activity means higher frequency and smaller amplitudes [85].

3.4.1 Gamma Waves

Gamma brainwaves are the fastest of brain waves (high frequency). Simultaneously processing information from different areas of the brain is related to Gamma waves. Gamma waves swiftly and quietly pass on information. It is the most subtle of the brainwaves. Figure 3.3 shows a sample of gamma waves. States of altruism, universal love and higher virtues, the gamma waves are active. In addition to this, people who have practiced meditation for over a long period of time, gamma waves were found to be much stronger and more frequent. The mind has to be in a state of quietness to access the gamma waves [73].

Figure 3.3: Gamma Waves [85]

3.4.2 Beta Waves

Our normal awakened state of consciousness is governed by beta waves, as a result it is the most common pattern which is observable most of the time. These occur when our attention is focused on cognitive tasks relating to the outside world. Beta waves are usually generated in the frontal lobes; thus, it is present when someone is alert, actively thinking, self-awareness, judgement, making choices, problem solving. Beta brainwaves are further divided into three bands; Lo-Beta (Beta1, 12-15Hz), Beta (Beta2, 15-22Hz) and Hi-Beta (Beta3, 22-32Hz). Figure 3.4 represents beta waves. There exists a downside to these Beta waves. Operating the brain at such continuous high-frequency processing is not very effective because it requires a great deal of energy. Not only does it take a huge amount of energy from us, it can also lower our feelings of emotions and creativity. Higher concentration of beta waves means higher levels of anxiety or stressful thinking along with a lack of relaxation. Additionally, it can lead to an increase in adrenaline levels and cause elevations in excitement [67].

Figure 3.4: Beta Waves [85]

3.4.3 Alpha Waves

Alpha brainwaves were found first because they are among the easiest to detect. They inhabit a low frequency range because they are generated when the brain is in a state of calm of calm and thoughtful relaxation. It is the state of resting. They can be detected when the eyes are closed and also along with other activities such as yoga, meditation, being creative or the moment just before on falls asleep. Alpha waves aids and promotes mental coordination and sync, calmness, alertness, the integration of the mind and body, learning and knowledge.
Some studies have also stated that alpha rhythm has two subsets: Lower Alpha (8-10 Hz) and Upper Alpha (10-12 Hz). A sample of alpha waves is shown in the Figure 3.5.
In the right hemisphere, alpha appears to be the strongest and a lack of alpha in the right hemisphere correlates with negative behaviors such as social withdrawal. People with depression have too much frontal alpha [96].

Figure 3.5: Alpha Waves [85]

3.4.4 Theta Waves

The source of these waves is probably from the frontal regions of the brain. Theta waves occur when dreaming in sleep but also predominate in deep meditation or daydreaming. Research shows that Theta waves are associated with memory, creativity and psychological well-being. We are disconnected from the outside physical world and more focused on our inner selves while in Theta. It is a place where we see dreams and vibrant images, where we see our fears and nightmares. Figure 3.6 shows a sample of theta waves.

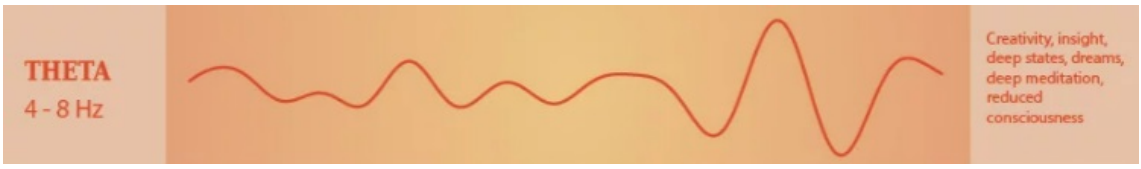


Figure 3.6: Theta Waves [85]

3.4.5 Delta Waves

Delta waves are slow and loud (low frequency and high amplitudes). They first start to appear in stage 3 of the sleep cycle and they overshadow almost all other EEG activity by stage 4. They are also during deep meditation. It is the origin of empathy and it completely disables external awareness. It is imperative to get enough sleep every night because of sleeps restorative properties. This is when healing and rejuvenation are stimulated [67]. The range of delta waves can be from

0.5-4Hz and a sample is given in the Figure 3.7.

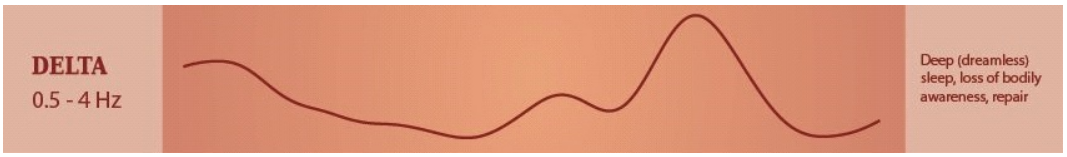


Figure 3.7: Delta Waves [85]

3.5 Electroencephalography (EEG)

The brain is made up of billions of neurons and these neurons are again connected to other neurons by synapses, thus, creating a dense interconnected network. These networks act as the gateway to how our brains operate; capacity of thinking, dreaming, seeing and sensing, taking in information, connecting information together and creating an experience for us, which we call our reality.

Figure 3.8: Electroencephalography (EEG) [107]

Postsynaptic potential is a subtle electrical impulse that is induced by every synaptic activity. The electrical impulse generated by a single neuron is far too difficult to detect without being directly in contact with the neuron itself. Nevertheless, electric fields that are capable of detection (large enough magnitude to spread through tissue, bone and skull) are generated when thousands of neurons function in unison. This kind of synaptic activity can then be detected from the outside.
Electroencephalography (EEG) is an electrophysiological monitoring technique for measuring the brain's electrical activity. There are both invasive and noninvasive techniques available. More typically, the noninvasive method involves placing electrodes along the scalp at specific locations. Figure 3.8 shows, how we can conduct EEG on a subject by using EEG headgears.
Electrocorticography (ECOG) is an invasive method where invasive electrodes are placed directly on the surface of the brain. For our research, we are using EEG, which is noninvasive.

Each electrode is connected to a single cable in typical systems. Others may use caps embedded with electrodes. Majority of the clinical and research applications use the International 10-20 system to specify the names and locations of the electrodes. This allows the parameters to be kept consistent throughout various experiments. 19 electrodes are used in most clinical applications along with two others for ground and system reference. We have also followed this practice and have used 19 channels. When a clinical or research-related application needs enhanced spatial resolution in a certain brain area, extra electrodes may be fitted to the standard setup. Up to 256 electrodes may reside within a high-density array cap [67].

3.6 General Supervised Algorithms

In the sense of artificial intelligence (AI) and machine learning, supervised algorithm is a type of algorithm in which both the input and desired output data are given. The basis of supervised algorithm is to match inputs with their appropriate outputs correspondingly.Based on its previous training, it can even decide the label for the new unknown inputs[1]. It can be further divided into classification and regression. The classification problem arises when the independent variables are of a continuous type while the dependent variable is of the categorical type. An example of such a classification can be placing a tumor in categories of malignant or benign. Likewise, regression tries to predict the relationship between the dependent and independent variables like a student's grade based on the hour studied. We have chosen supervised algorithm as we have both inputs and outputs [88].

3.6.1 Logistic Regression Classification

Logistic Regression is the appropriate form of solving classification problems in cases where dependent variable is dichotomous , as in binary. The logistic function is used in Logistic Regression. The logistic function (another name for it is the sigmoid function) is a curve that takes the shape of an S. What this function does is that it maps any real number between 0 and 1 [88]. The logistic function is:

$$F(z) = E(\frac{Y}{x}) = \frac{1}{1 + e^{-z}} \tag{3.1a}$$

The logistic function has an appearance like the Figure 3.9:

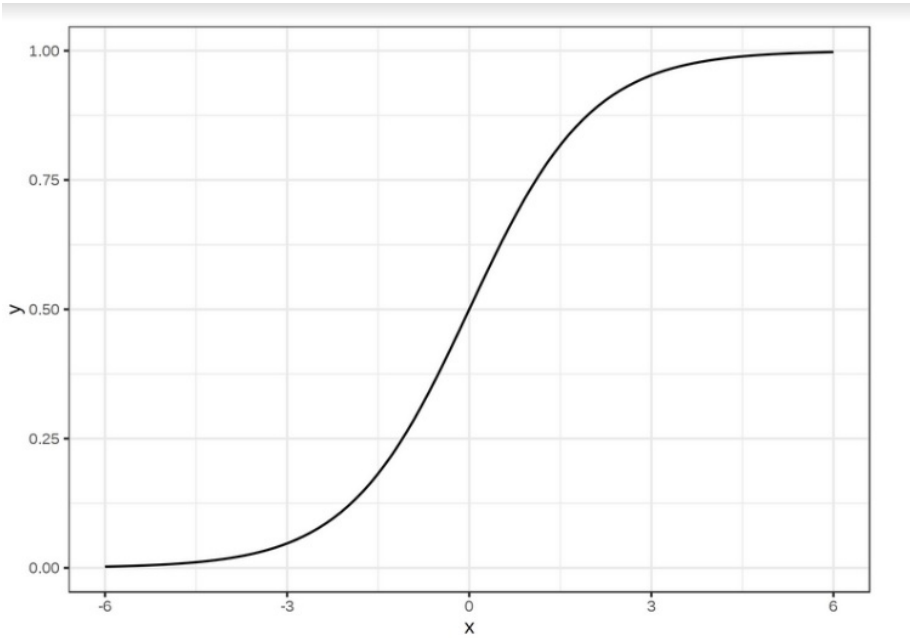


Figure 3.9: Logistic Function Graph[100]

The equation 3.1a shows an equation of logistic function where y is the class labels which can be assigned a value of one of the two classes. The x represents the various features.
We model the relationship between the outcome and the features with a linear equation as done in the linear regression model [100].

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \tag{3.2a}$$

Equation 3.2a shows that z is a linear combination of α plus β_1 multiplied with X_1, plus 2 multiplied with X_2, and plus k multiplied with X_k, where the X_k are the independent variables and α and β_i are constant terms representing unknown parameters.
For classification, as mentioned earlier, we require probabilities between 0 and 1. For this, the right-hand side of the linear equation is wrapped into the logistic function shown in equation 3.3a, thereby forcing the output to give values that are only between 0 and 1.

$$F(z) = E(\frac{Y}{x}) = \frac{1}{1 + e^{-(a+\sum \beta_i X_i)}} \tag{3.3a}$$

3.6.2 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm. It can be used for both regression and classification problems. For this paper, we have used SVM for classification. To carry out the SVM algorithm, every data piece is plotted as a point in an n-dimensional space (n is equal to the number of features) as shown is Figure 3.10. In addition, the value of a particular coordinate takes the value of a feature in our data [104]. The goal of the SVM algorithm is finding a hyperplane that can distinctly partition the classes [88].

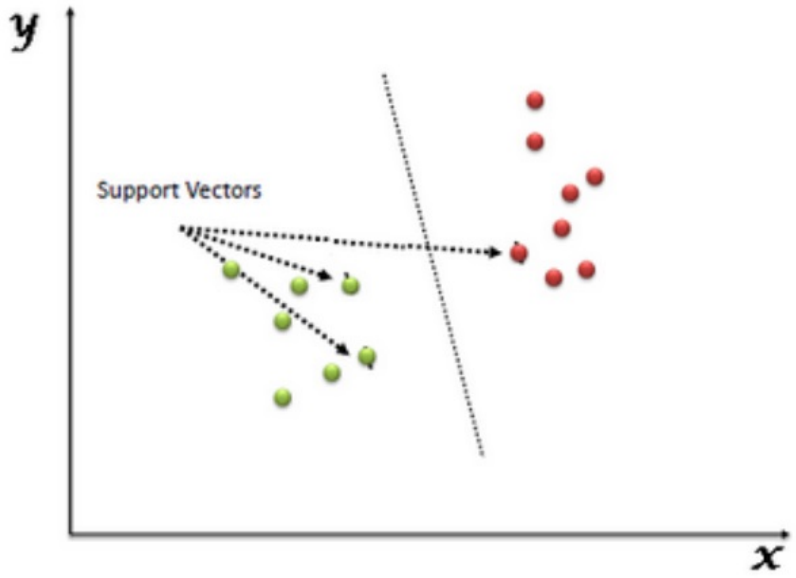


Figure 3.10: Support Vector Machine[87]

The hyperplane is known as the decision boundary. The hyperplane helps us in differentiating between the data points, the two sides of the hyperplane represent different classes and data points that belong to either side can be attributed to those classes. The number of features determine the dimension of the hyperplane. For example, if the number of features is 2 and 3, the hyperplane is a straight line and two-dimensional plane respectively. There are several possible hyperplanes that can be plotted for separation between the two classes of data points but our goal is to find such a plane that has the maximum margin. Maximum margin means the distance between the data points of both classes must be the maximum. This is of utmost importance because by maximizing the margin distance, we are creating reinforcement for when newer data points are to be classified so that they can be done with more reliance.
Data points that are closer to the hyperplane or decision boundary are called Support Vectors. Both the position and orientation of the hyperplane are affected by these Support Vectors. These points are what allows us to build our SVM model. We can maximize the margin of the classifier using such Support Vectors. Even deleting a point will have an effect on the hyperplane.
Grid search technique is used to optimize the accuracy of a classification model. But it provides a large computational problem. Upon GridseacrCV, we found that an optimum value for the parameter "c" was 1000 and not the default value to 100. Inputting the value of c to be 1000 increased the accuracy of the classification model but provided a computation backlog.

3.6.3 Naive Bayes Classifier

Naive Bayes classifier is a type of probabilistic machine learning model based on the Bayes Theorem, used in classification tasks. For each sample of data it generates conditional posterior probabilities. Due to its simplicity and linear run-time, the Naive Bayes classifier is identified as a common learning algorithm for data mining

applications [17]. It not only handles a large number of variables and large data sets but also handles both discrete and continuous variables of attributes [36]. The naive Bayes classifier refers to learning tasks where a combination of attribute values defines each instance x and where the target function f(x) will assume any value from the same finite set V [10]. It produces two probabilities in case of binomial classification. The following equation 3.4a represents NB:

$$R = \frac{P(i|X)}{P(j|X)} = \frac{P(i)P(X|i)}{P(j)P(X|j)} = \frac{P(i)\prod P(X|i)}{P(j)\prod P(X|j)} \tag{3.4a}$$

Here P indicates probability, both x and j are labels. X includes independent variables. The greater probability suggests that the class label is more likely to be a real class label when considering these two probabilities.

3.7 10 fold cross-validation

Cross validation is a machine learning technique used to determine the performance of a classification model. This is done by at first dividing the original dataset into a training dataset used to train the model, and then a test set to evaluate it [63].
In a k-fold cross validation, the original dataset is sub-divided into K equal sized subsamples. From the subsamples, a single subsample is selected as the testing data used to test the model and make the necessary predictions. The other K-1 subsamples are to be used as the training dataset. The process is then repeated k times. These produces a total of K number of accuracies. The mean or standard deviation of those accuracies can be used to test the evaluate the performance of the model.
In our work, we have selected the value of k to be ten, hence the name of 10-fold cross-validation. The dataset is divided into 10 subsamples, one of each is the testing data while the remaining nine are to be the training data. The process is then repeated a total of ten times.
10-fold CV is a popular method to evaluate classifier performance because it is a relatively simple method to understand with a moderately easy code implementation. This method also results in a less biased estimate of the model.

Chapter 4

Proposed Model

In this chapter, we discuss about data description and workflow of our proposed model.

4.1 Dataset

Our initial plan was to use the EEG headset available in Brac university and collect the data ourselves from hospitals. Unfortunately, due to various complications provided by each hospital in terms of acquiring permissions and various bureaucratic problems it was difficult to collect data. Hence, we scoured the internet for a dataset relevant for our research purpose and found a reliable dataset. The dataset we found was obtained from two professors Dr. Malik and Dr. Wajid Mumtaz who conducted the experiment on 68 participants [68].
The dataset contains 181 separate EDF files within four categories. The categories are:

- H stands for Healthy controls
- MDD stands for Major Depressive Disorder
- EO stands for Eyes open
- EC stands for Eyes closed

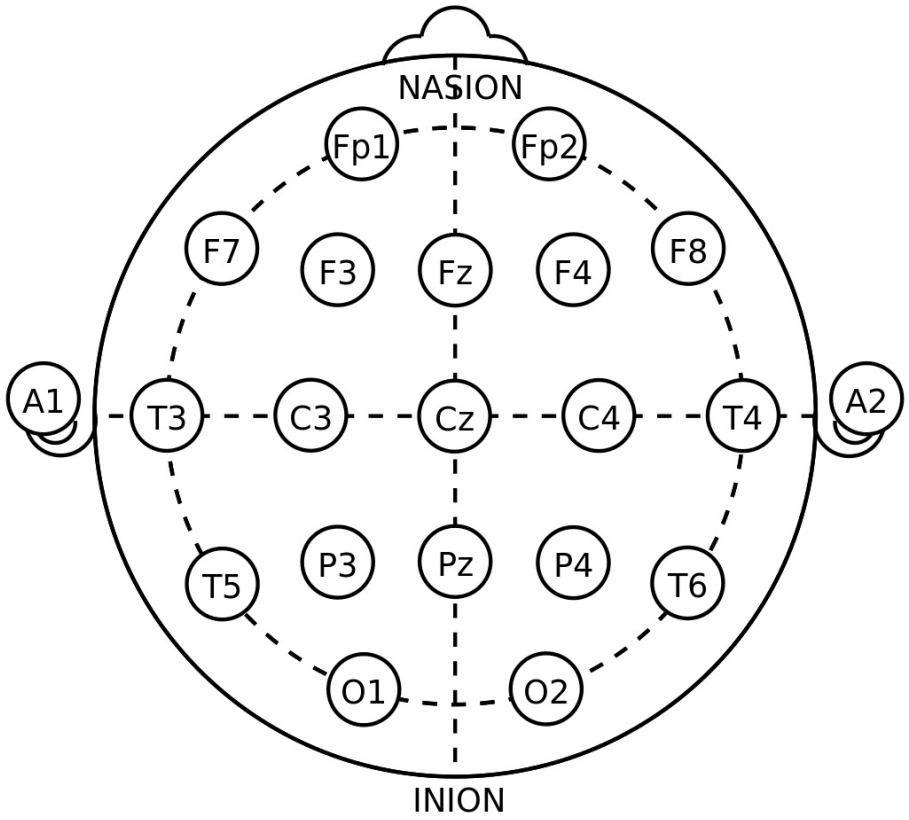


Figure 4.1: The International 10-20 system on the participant's scalp [67]

In this research, as stated above, the dataset we used consists of two groups: MDD patients and Healthy controls. The MDD group is made of a total of 34 patients among which 17 are males and 17 are females. The Healthy control group is made up of 30 participants among which 21 are males and 9 are females. 19 electrodes where placed on the participant's scalp following the International 10-20 system as shown in Figure 4.1 above, to take the EEG readings. The electrodes are Fp1, Fp2, F7, F8, F3, Fz, F4, T3, C3, Cz, C4, T4, A1, T5, P3, Pz, P4, T6, A2, O1, O2.

A flowchart below in Figure 4.2 shows how we can collect data from the participants:-

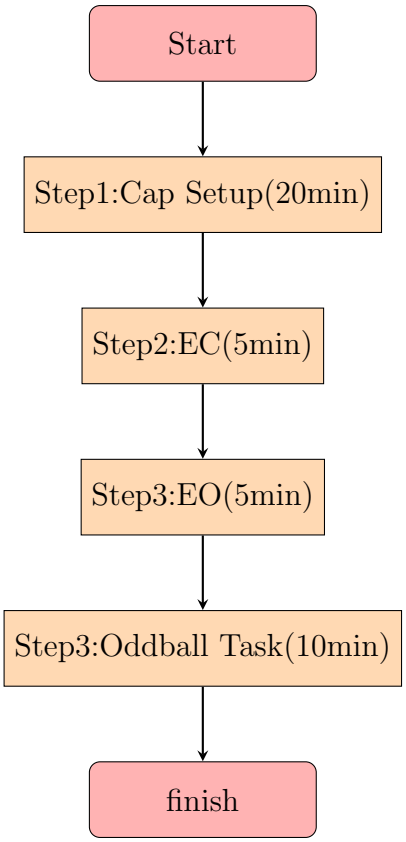


Figure 4.2: EEG Data Collection Method

After setting up the EEG cap was completed on the participants, the participants were subjected to an eyes closed (EC) session. The patients had their eyes closed while sitting in a chair and staying in a state of highest alertness. The EC recordings took five minutes. The next step, the participants were subjected to an eyes open (EO) session in which they focused their eyesight on a fixed point with minimal eye movements and blinks, on a computer screen. The EO recordings also required five minutes. The subsequent step involved a visual three stimulus oddball task; displayed on a computer screen in front of the participants were a random sequence of shapes. There was a total of 3 shapes and only one was shown at a time. The participants had to press certain keys on the keyboard to confirm a "yes" for the target shape and a "no" for the other shapes. These recordings were done for a total of 10 minutes.

4.2 Data Description

We are aiming to create a system that will detect and differentiate between brain signals from healthy and depressed patients, thereby, allowing us to detect depression. This process commences with us obtaining raw EEG data from the participants. Yet, within these raw EEG data reside unnecessary noise which can conceal weaker signals. These can originate from artifacts such as eye blinks, eye movements and even other physiological signals such as heart and muscle activities. Therefore, these noises must be removed and this removal is referred to as pre-processing. Moreover,

there were two electrode channels which showed no signal coverage. This was seen by plotting a significant number of raw EEG signals manually. The two channels, particularly channels 23A-23R" and "24A-24R", as seen from the diagram below, was dropped as a part of pre-processing. The channel names were further mapped to their appropriate names as required to set up a 10-24 electrode montage required to set up the necessary code base. Figure 4.3 below represents a picture of the raw EEG data from our code.

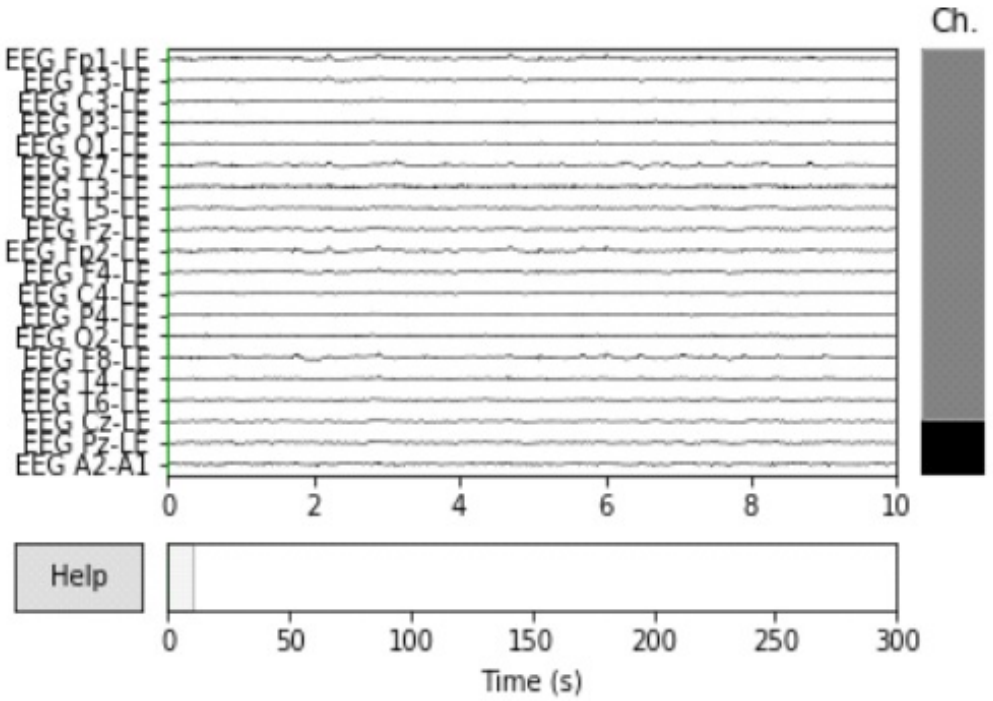


Figure 4.3: Raw EEG data

Afterwards, feature extraction is carried out and an EEG data matrix is formed from the derived features. Features from both eyes closed (EC) and eyes open (EO) conditions are extracted and included in the EEG data matrix. The rows of the matrix represent each participant while the columns portrayed individual features. The features extracted (Absolute Band-power and Alpha Interhemispheric Asymmetry) were specific to the problem domain, i.e Depression Classification, that we are addressing. Following this, z-score standardization was carried out on each column of the EEG data matrix, i.e., it was carried out for each feature. Z-score standardization is required because the EEG data matrix is a culmination of all the data from different participants and will therefore possibly require to be normalized. The EEG data matrix then goes through three classification algorithms: Logistic regression (LR), Support Vector Machine (SVM) and Naïve Bayesian (NB). Finally, 10-fold cross-validation was carried out to validate the classification models. Figure 4.4 shows the flowchart that sums up the entire process of our work.

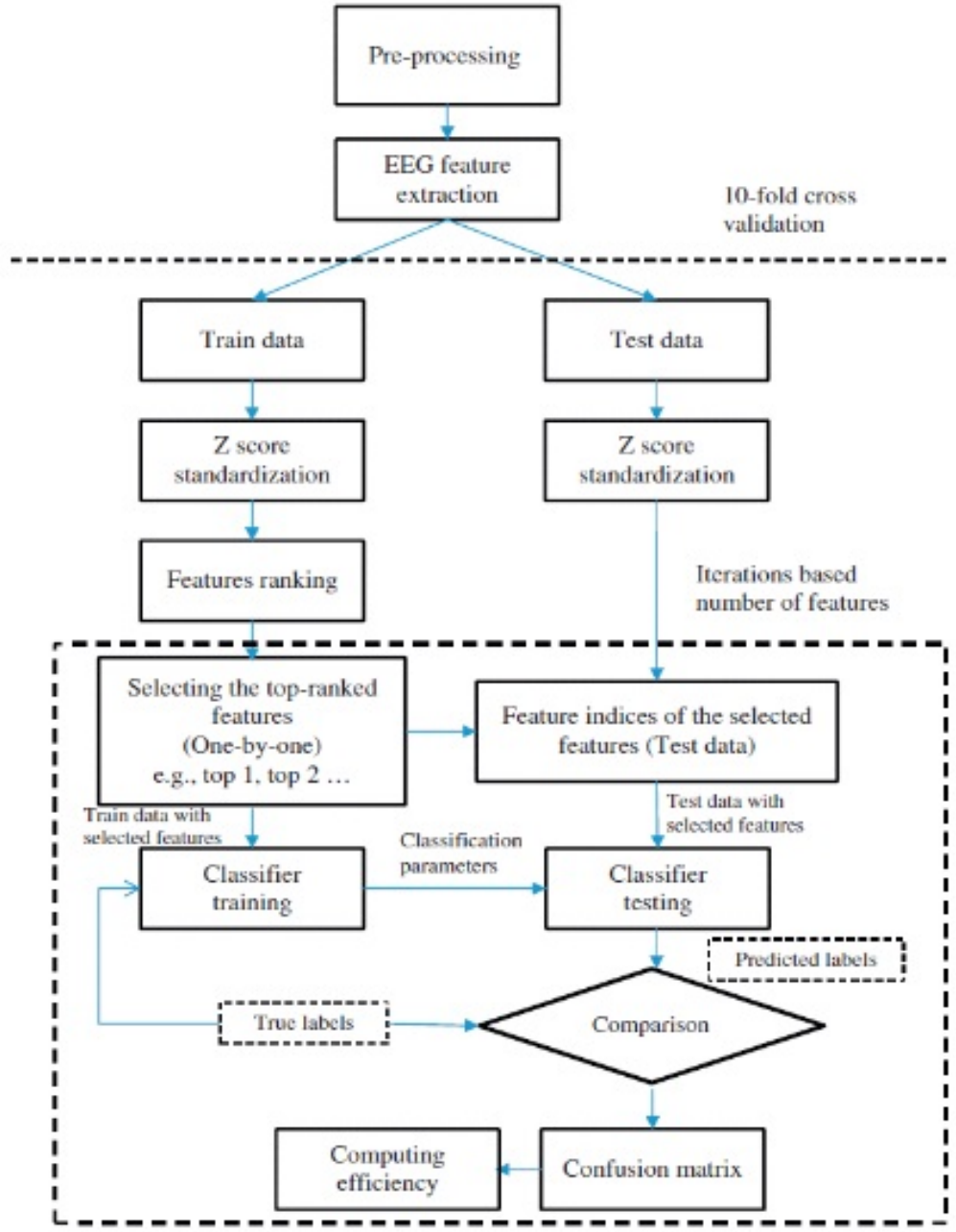


Figure 4.4: The entire processing

4.3 EEG Pre-processing

As mentioned earlier, we collected a data-set [68], that consisted of raw EEG signals that were collected by placing electrodes on the scalp but had some unwanted noise and thereby is not an accurate portrayal of the signals originating from the brain. These artifacts may stem from various sources such as eye movement blinking, heart activities, muscle activities and even from power lines causing the data to be skewed. So, noise removal is needed so that we can attain a clearer version of that data true to the actual neural signal. We want to be able to isolate only the relevant neural

signals from random neural activity occurring during the EEG recordings. However, the removal of certain artifacts from the original signal could hamper or result in the loss of useful information unknowingly. Hence, rather than removing the artifacts entirely, we opted for their correction.
Five minutes of noise removal was executed on both EC and EO data so that we could acquire at least two minutes of clean EEG data of both EC and EO for each participant. Clean EEG samples of two minutes in length provided a clear reflection of the neural activity that underlies it [94].
In reading EEG signals, one should recognize artifacts that distort the signal and cause computational error. Artifacts are EEG signals that do not come from the brain. Some of these artifacts Some devices can elicit genuine epileptic abnormalities or seizures. It is important for us to be aware of the logical topographic field of distribution for true EEG abnormalities when differentiating brain waves and artifacts [94].
Due to their spatial region on the scalp being near the eyes themselves, the Fp1 and Fp2 frontal sensors, are the regions where artifacts relating to eye blinks usually show up. From the raw signal plot of the EEG data sample shown in Figure 4.5, it could clearly be identified visually that eye blinks result in a region of high amplitude in the EEG signals while also acting as a producer of distortion in the nearby signals, consequently distorting the entire EEG segment [94].

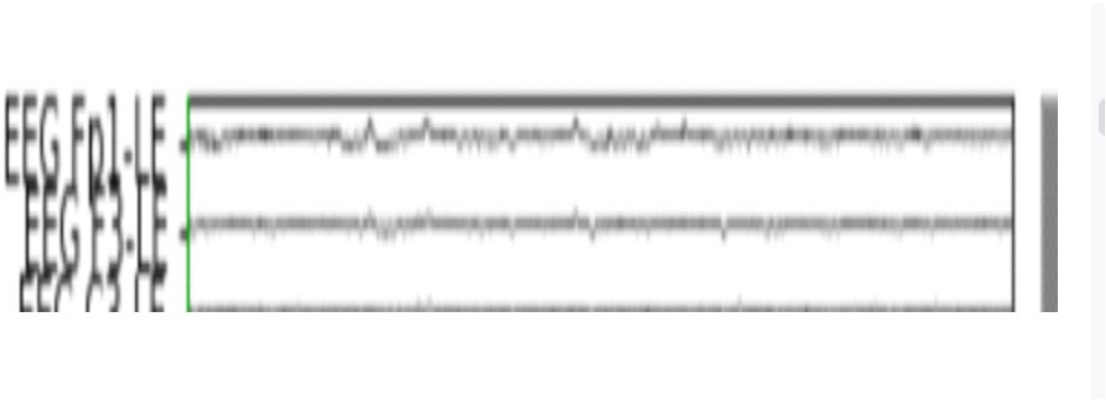


Figure 4.5: EEG Data distortion for eye blinks

In this research we used Python and mainly the Python MNE library for dealing the EEG dataset. The MNE library for Python is an open source library for investigating, visualizing and analyzing human neurophysiological data of Electroencephalogram (EEG), Magnetoencephalogram (MEG), Stereo-electroencephalography (sEEG), Electrocorticography (ECoG) and many more [109]. Processes such as these measure the brain's electrical activity. Understanding and analyzing such signals from the brain is a tremendous task that require skills in physics, signal process, statistics, and numerical methods. MNE solves these problems by offering state of the art algorithms spanning a plethora of methods as for example, data pre-processing, source localization, statistical analysis, and estimation of functional connectivity between distributed brain regions in one whole package [45].
The EEG data that we have used is in the file format EDF that stands for European Data Format. The function *mne.io.read_raw_edf*() is a core modular function built in with the MNE python package. With this function, we were able to read and load EDF files onto the Jupyter Notebook IDE. All of the files were loaded onto a

python list data structure using this function. A list in python can hold a wide variety of data types and has built in functions such as "append" giving us the ability to sequentially insert one EDF after the other running through a for loop, thereby iterating through every single EDF File in the dataset. Based on observation and as stated previously the EEG channels "EEG-23A-23R", "EEG-234-24R" and "STI 014" showed no signals and thereby had no contribution to our purpose. Thereby these channels had to be manually dropped to ensure correct further data processing.
As the above mentioned three channels were dropped, we had to find a replacement for the data coming from these three channels. This is when we applied the concept of interpolation. Interpolation is particularly important when a channel is emitting particularly weak or no EEG signals. Interpolation is the approach where rebuilding these poor or non-existent EEG signals according to the distance it pertains from its neighboring electrode in the same or similar head space which also is indicative of the electrode with the smallest distance from the defected EEG channel. In our code, we used the interpolation_bad core function of the MNE python package which takes in an array containing the defected channels labelled bad_channels. The labelling is to be done manually. The interpolated data is now used as the core raw EEG signal from the previously mentioned defected electrode (which is "EEG-23A-23R", "EEG-234-24R" and "STI 014")
Afterwards, a specific montage had to be set. Montages contain sensor position in three-dimensions. It is used to set up the physical position of the sensors in the head. The *set_montage*() function is used to make the codebase understand of the physical location of the sensors based on the EEG electrode names. In doing so the channel names had to also be transformed to names set according to a specific montage. The montage we used was the "standard_1020" electrode system. Applying this makes bad component detection and further signal processing mechanisms easier. This montage contains information regarding sensor configuration, channel positions and head digitization.

Independent component analysis (ICA):
As mentioned previously in this research, BESA software was not used for noise reduction since it is outdated and a lot of people are now unfamiliar with it.However we used a better technique:- ICA and PCA.
ICA is a method for assessing autonomous source signals from a lot of accounts in which the source signals were combined in unknown proportions. It is a method for extracting a independent source from a mixture of recordings. ICA is an as of late created strategy in which the objective is to locate a direct portrayal of non-Gaussian information with the goal that the segments are statistically autonomous, or as free as could be allowed [11]. It deforms power within the range of 5-20Hz [16]. Another method is the Principal components analysis (PCA) which needs no distributional presumptions and, all things considered, is especially a versatile exploratory strategy which can be utilized on numerical information of different sorts [66]. It successfully reduces the artifacts related to eyes with the least amount of spectral distortion. However, only performing PCA is not effective and sufficient for the deformation of the spectrums and reports suggest to go for further examination by ICA method [16]. Hence, we chose to perform ICA.

ICA is an automated, dependable, instantaneous and a pragmatic tool that decreases the lengthy and time consuming manual thorough selection of ICs of eliminating artifacts. Though, it has a small channel of 25 electrodes but it gives surprisingly amazing results [61].It helps to remove neurological activity from the artifacts [90].Using ICA technique we do not need to subtract original EEG from the corrupted EEG, as it does not depend on the direct artifact signals. Positively, it is not confined to a specific number or range so it adds up to its advantages.
We can even intensify the recognition ability of the automatic artifact signals by combining them with another classification tool which is the linear discriminant analysis [38].
Using ICA, we can separate different components such as eye blinks or heartbeats from the brain signal because these artifacts have characteristic shapes and are easy to identify. This method does not assume orthogonal or gaussian behavior of the individual signals, which other techniques still assume despite them being unreasonable to assume. For this very reason, it is deemed the best. Nonetheless, an assumption that the signals are statistically independent is made. It is possible for us to isolated separate signals from each other using ICA if the signals are statistically independent and non-gaussian. The gist of the process is shown in Figure 4.6.

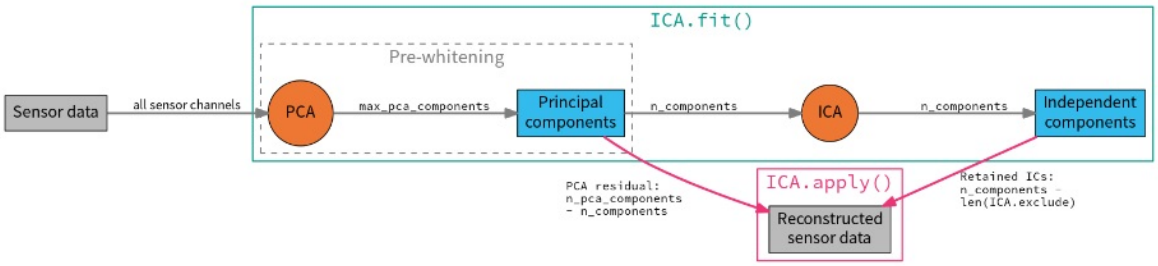


Figure 4.6: ICA process [80]

ICA looks for a linear transformation to find a mutually independent representation of the original signals. Assuming an n-dimensional demonstrated signal by x= (x1, x2, . . . , xn)T which is a linear mixture of the independent components, say x= As where s= (s1, s2, ... , sl)T. Here A is an nl linear matrix. Suppose y= (y1, y2, ... , yl)T denotes an l- dimensional signal which is the calculated vector of independent components. The following linear relationship described in the equation 4.1a represents the ICA problem:

$$y = Wx \tag{4.1a}$$

In equation 4.1a, W represents an unknown l n matrix by which we can obtain independent component.
A common example for the implementation of ICA is the “blind source separation” which is as follows: with 3 musical instruments playing in the same room, and 3 microphones recording the performance (each picking up all 3 instruments, but at varying levels), can you somehow “unmix” the signals recorded by the 3 microphones so that you end up with a separate “recording” isolating the sound of each instrument?

This analogy is quite similar to our problem domain as we want to extract raw EEG brain signals from a mixture of signals resulting from signals due to eye blinks, heart-beats, muscular activities, etc. As long as the sources are statistically independent , it is possible to separate sources using ICA and then rebuild the signal without the sources that is to be removed.
In our codebase , ICA was implemented using the MNE-Python package. ICA in python implements three algorithms, which are Fastica, Picard and Infomax. In our code base we have implemented the Fastica ICA algorithm due to its robust nature and fast computation power.
At first we imported the ICA module from the "mne.preprocessing" core library. Then we created an ICA object from the library we had imported earlier. This ICA object takes two things as its parameter, namely "n_components" and "random state". Here the parameter n_components represents the number of principal components that has been passed to the ICA algorithm during fitting. This was done to increase the speed of computation by reducing the number of components to be selected. The principal components selected were 15. Another parameter "random_state" of value 97 is an integer used as a seed for Random State. As ICA fitting is not deterministic so we specified the random seed to check and validate the data when running the code multiple times.
This ICA object created is not fitted with the raw EEG data using the $ica.fit()$ method. This method was individually applied to all the signal samples via a simple for loop and appended onto a python list data structure.
As we know that one of the most prominent and inevitable artifacts during EEG recording is eye blink because it causes an adjustment in the electric fields that encompass the eyes, which influences the electric field over the frontal scalp produced by neural potentials. Therefore, the primary artifact to be removed would be the artifacts caused due to blink movements of the eye.
The detection the bad ocular artifacts require the usage of the "ica.find_bad_eog" method which takes in a single data sample and a EOG channel. This presented two problems.
The first challenge we had was that the data we had did not have an EOG channel. Thereby in situations like this, it is advised to simulate a channel to be an EOG channel. For us, the Fp-1 channel was selected as a simulated EOG channel and was sent in as a second parameter for the "ica.find_bad_eog" method.
The second challenge was a huge computation problem. Due to the large data samples, it was difficult to manually select ICA components for all the data samples. Thereby for this, we implemented the selection of ICA components using template matching. This is done using the "corrmap" function of the MNE.preprocessing library.
We first take the fitted data from the very first sample, i.e, the first item from the ICA data signal list and find the bad components from just that first sample. These components are stored as "eog_inds" and "eog_scores" as returned data from the "find_bads_eog" method that was previously mentioned.
Using *mne.preprocessing.corrmap*() it is possible to use the bad components, like Figure 4.7 shows,from the first data sample as a template for choosing which IC bad components to remove from other subjects' data. We set a threshold value of 0.6 as a default standard and ran *corrmap*() with the third parameter labelled as "blink".

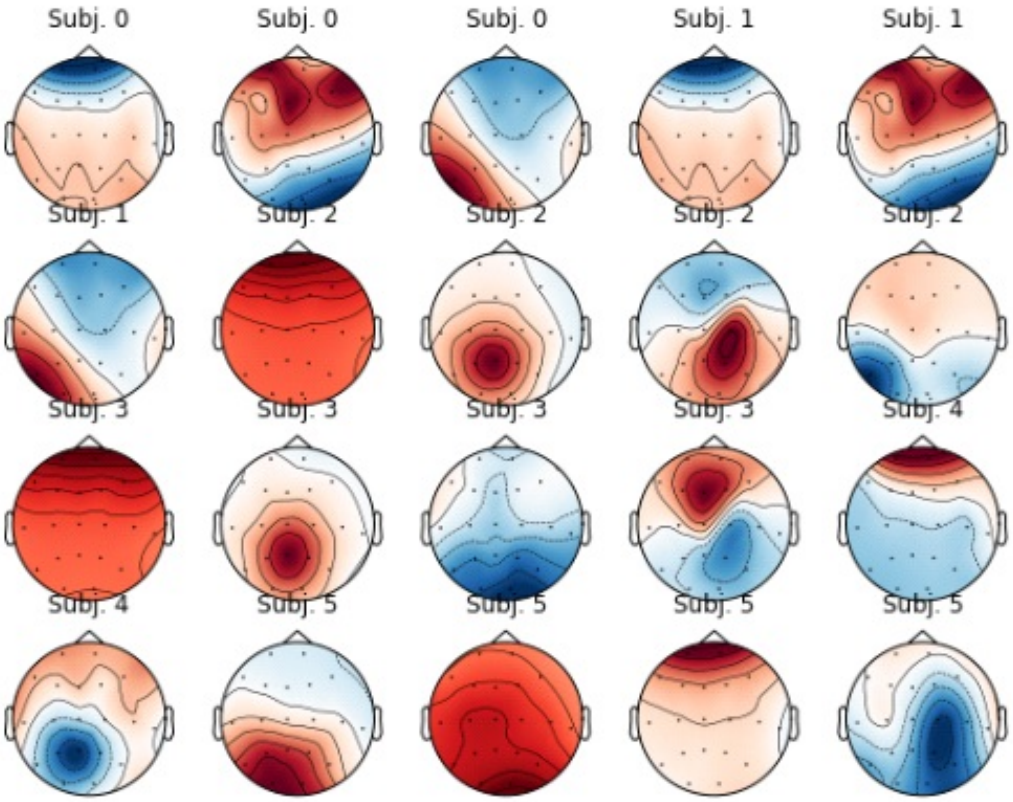


Figure 4.7: Bad Components

This allowed us to extract only the bad components as they are uniquely labelled. These components are included in an array called ica.exclude for each data sample. Note that the each data sample excluding the first was subjected to the same procedure via a simple for loop implementation. Each data sample was subjected to the method *ica.apply*(). This ica.apply method has two functionalities. It at first looks for the bad components in the ica_exclude array and removes it from the signal. It then reconstructs the signal without the bad components, hence giving a noise and artifact free signal as we can see in Figure 4.8.

```
    ica.exclude = [ica.labels_ for ica in HDD_icas][c]['blink']
    ica.apply(x)
    c = c+1
```

```
Transforming to ICA space (15 components)
Zeroing out 3 ICA components
Transforming to ICA space (15 components)
Zeroing out 3 ICA components
Transforming to ICA space (15 components)
Zeroing out 4 ICA components
Transforming to ICA space (15 components)
Zeroing out 4 ICA components
Transforming to ICA space (15 components)
Zeroing out 2 ICA components
Transforming to ICA space (15 components)
Zeroing out 4 ICA components
Transforming to ICA space (15 components)
Zeroing out 3 ICA components
Transforming to ICA space (15 components)
Zeroing out 2 ICA components
Transforming to ICA space (15 components)
Zeroing out 3 ICA components
Transforming to ICA space (15 components)
Zeroing out 2 ICA components
Transforming to ICA space (15 components)
Zeroing out 2 ICA components
Transforming to ICA space (15 components)
Zeroing out 4 ICA components
Transforming to ICA space (15 components)
Zeroing out 2 ICA components
Transforming to ICA space (15 components)
Zeroing out 1 ICA components
```

Figure 4.8: Reconstruction of signals

4.4 Feature Extraction

4.4.1 Absolute Power

After noise removal, the clean EEG signals that were extracted underwent feature extraction.
The features we extracted were absolute power of each frequency band and EEG alpha interhemispheric asymmetry.
We chose to extract the absolute power of each frequency band as upon research, we found out that EEG signals from the frontal and temporal regions have been associated with cognitive deficits and functional impairments which are common characteristics of depression [95]. The reduced left frontal activity (increased alpha power or amplitude) has been associated with depression, according to a recent review [50]. The spectral power of the various EEG bands shows different abnormalities for MDD patients. As a result, studying and analyzing the EEG spectral powers of both MDD and healthy participants can allow us to classify depression, as we can notice from Figure 4.9. Hence, EEG spectral power is one of the features of this study. One study has shown that EEG frontal asymmetry is a marker

for depression [84]. Also, a decrease in alpha waves has been linked to depression[50]. Consequently, EEG alpha interhemispheric asymmetry could be regarded as a biomarker for automatic depression diagnosis. Other bands namely the theta band has also been associated with depression in addition to the alpha band.

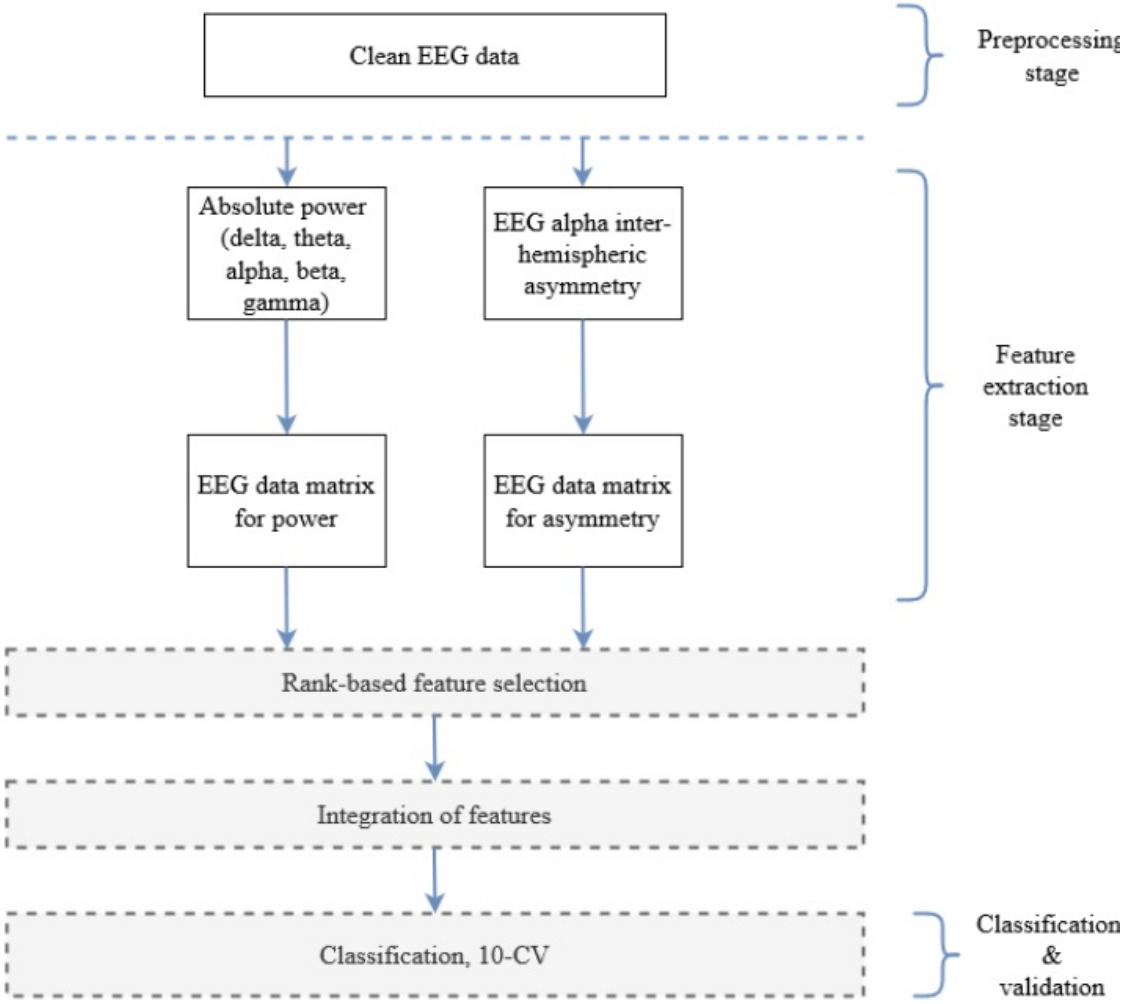


Figure 4.9: Feature Extraction

Computing Absolute Power for Different EEG Bands

At first the EEG signals must be filtered and separated into individual frequency bands; gamma (30-70 Hz), beta (12-30 Hz), alpha (8-12 Hz), theta (4-8 Hz) and delta (0.5-4 Hz). Next the power spectral density (PSD) is calculated and finally the individual powers of the bands like Figure 4.10 illustrates.

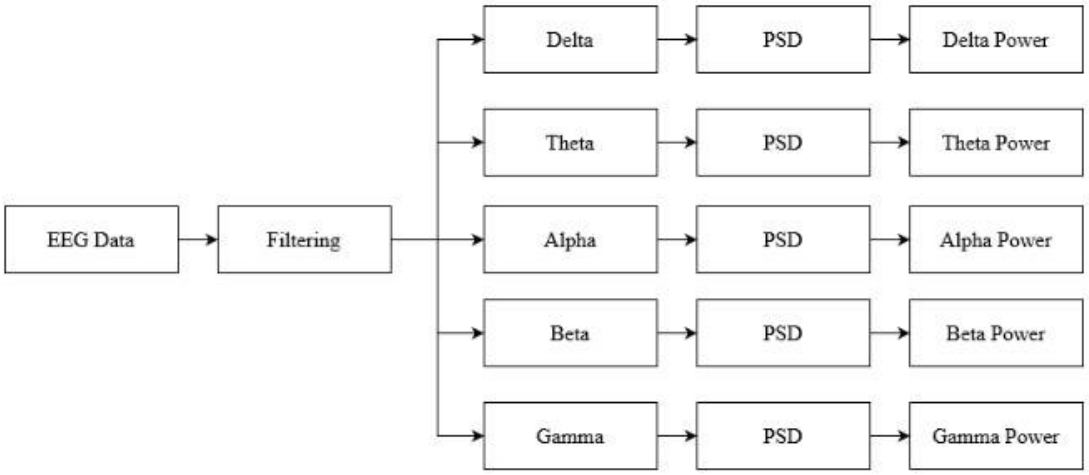


Figure 4.10: How EEG signals are filtered

The absolute power calculation includes 19 EEG channels. The power values are averaged according to the number of channels corresponding to each brain lobe such as frontal (Fp1, Fp2, F3, F4, F7, F8, Fpz), temporal (T3, T4, T5, T6), parietal (P3, P4, P7, P8), occipital (O1, O2), and central (C3, C4). The average power of individual bands was calculated using the Welch periodogram method [7] and the Hamming window function described in 4.2a [6].
Data is broken down into a number of weighted sinusoids, this is the purpose of spectral analysis. Due to the decomposition of the data, we are now free to examine the frequency content of the phenomenon we are currently studying. This phenomenon can either be distributed across a wide range of frequencies or even found within a narrow frequency band. Two main areas of spectral analysis are the Fourier transformation and Power spectral density (PSD). Deterministic data does not contain any random noise, disturbances or effects. In our case, the data does indeed consist of random noises and effects. For this very reason we must compute a PSD. Averaging or smoothing is required for us to expose the underlying phenomenon [7]. Power spectral density (PSD) of a signal shows the distribution of the signal's power over a frequency spectrum [13]. While there are many ways to calculate this, we have used the Welch's periodogram method.
A complex object, for example a time series, can be split into the sum of simple objects. These simpler objects can be further studied to find out which is unimportant and subsequently discarded, while using the remainder of what is left to make an approximation of the original object. This technique leads us to a periodogram of a time series. Time series data sets can be described using periodograms, which makes it an excellent tool to use [1]. The relevance of various frequencies in time series data is measured in a periodogram to determine underlying periodic signals [83].
Estimating power spectra using Welch's method (also known as the periodogram averaging method and weighted overlapped segment averaging (WOSA) method) is conducted by segmenting the signal into sequential fragments, thus creating periodograms for each of the fragments and then averaging them [7]. The parameters are; K is the number of segments, M is the length of each segment, S is the number of new points in each segment. The number of points in common to two adjacent

segments is M — S. Two adjacent segments are said to overlap by M — S points. When S = M, the segments do not overlap. When S = 0.5M, the segments contain 50% overlap. In our study, we divided the EEG signal into 8 segments with 50% overlap. It also contains a window function. Hann or Hanning, Hamming, Blackman, Blackman-Harris, and Kaiser-Bessel are among common windows that can be used [3]. We have used the Hamming window function.
PSD calculations involve Fourier transformation and when we carry out a Fourier Transformation on data, the way periodic and non-periodic data is affected by a window is different [22]. A window is not needed for periodic data, conversely, non-periodic data requires a window. Non-periodic spectral leakage will be diminished but not completely eradicated. EEG signals are non-periodic and that is exactly why we have utilized a window. A common usage of the Hamming window is when dealing with operation noise and vibration measurements. It takes the shape of a cosine wave having 1 cycle. It is always positive, a result created from having a 1 added [103]. The starting and ending value is 0 with a center value of 1. It aids with minimizing the spectral leakage due to its smooth change from 0 and 1 values. Now comes the question for an optimal window size. The most commonly used approach is to apply a window size that is sufficient to take in at least two cycles of the lowest frequency available in the signal of interest. Since the signals we are using are brain signals and the lowest possible frequency of interest is 0.5 Hz which is of the delta frequency band, the window size for us is $2/0.5 = 4$ seconds.

$$\hat{S}_{xx}(k) = \frac{1}{N}\left|\sum x(n)e^{-2\pi jkn/N}\right|^2 \tag{4.2a}$$

To perform these calculations, we have utilized the YASA library for Python. YASA (Yet Another Spindle Algorithm) is a Python 3 toolbox that can allow us to detect sleep micro-structure events from EEG recordings [99]. We use the function called *bandpower*() from the YASA library *yasa.bandpower*() to calculate the average band power of all the individual bands at once. The function takes in an array containing 2D EEG data. The length of the sliding window is defined, which is required for the Welch Power-Spectral-Density calculation. The bands are defined with the lower and upper frequencies along with the band name. The window used, in this case, the Hamming window, is also to be specified. This function then calculates and returns to us a pandas Dataframe comprising of the average power of each band. The rows represent different channels and the columns represent different bands. Since we are interested in the five frequency bands namely Delta, Theta, Alpha, Beta and Gamma, we specifically labelled the columns in an increasing order of frequency.
This method of calculating and extracting individual frequency bands are done for each of the data samples via implementation of a simple for loop. A pandas dataframe was returned for each of the data samples. To build an EEG data matrix, we first had to create an empty dataframe and then consecutively concatenate it with each of the dataframes that were returned from individual data samples.
The dataframe also consisted of redundant constant numeric columns representing the frequency response and non-numeric columns representing relative which had to be dropped from the dataframe.
This procedure was done for both MDD patient dataset and healthy patient dataset.

4.4.2 EEG Alpha-Interhemispheric Asymmetry

Next came the feature extraction of the alpha interhemispheric asymmetry mode. If we see anatomically or even functionally, the two hemispheres of the human brain are not identical. In the most usual cases the frontal lobe of the brain is broader and hence extends a bit more towards the right hemisphere compared to the left hemisphere [27]. Reports also say how the left frontal region is more active when a person shows positive emotions or is in a peaceful state of mind and how on the other hand when a person portrays negative emotions the right frontal region is more responsive [4]. Depression is also one kind of negative emotion and there are links of MDD with the EEG alpha interhemispheric asymmetry. Therefore, we can use interhemispheric difference patterns as a classification tool and with the aid of it we can distinguish healthy and MDD patients. Participants with MDD are more likely to have greater asymmetry and it is more likely to outshine the asymmetry of the healthy patients. The method tells us about the difference of spectral power in alpha bands between the left and right brain coordinates. The left and right power hemisphere can be represented by the following equations 4.3a and 4.4a respectively:

$$W\prime_{Lmn} = \frac{\sum_{f=f_1}^{f_2} S_{Lmn}}{\sum_{f=0.5Hz}^{30Hz} S_{Lmn}} \tag{4.3a}$$

$$W\prime_{Rmn} = \frac{\sum_{f=f_1}^{f_2} S_{Rmn}}{\sum_{f=0.5Hz}^{30Hz} S_{Rmn}} \tag{4.4a}$$

where f_1 represents the lower frequency limit and f_2 the upper frequency limit of the EEG frequency bands we selected respectively. The left and right hemispheric spectral power densities are denoted by S_{Lmn} and S_{Rmn} respectively. Finally, W_{Lmn} and W_{Rmn} show the left and right signal power individually.
Now we figure out the alpha interhemispheric asymmetry for each patient both healthy and the MDD victims. We divide the electrodes based the left and right sides and place a name for the left and right hemisphere (O1 and O2). The odd electrodes are Fp1, F7, F3, T3, T5, C3, P3 and the even electrodes are Fp2, F8, F4, T4, T6, C4 and P4.Hence we have to find out the difference in the absolute alpha power band between signals emitted from the electrodes between the left and right hemisphere.
We discard the electrodes that are at the center region as indicated by the Figure 4.1, as these electrodes namely electrode "Fz" , "Cz" and "Pz", do not categorize as part of either the left nor the right hemisphere. We also discard the electrode "A2-A1" as they are in the outer most region of the head and contribute the least to our feature of interest, i.e. , the Alpha-Interhemispheric Asymmetry.
Finally, the following equation 4.5a provides the mathematical description of the interhemispheric asymmetry.

$$A_{mn}(f_1, f_2) = \frac{W\prime_{Lmn} - W\prime_{Rmn}}{W\prime_{Lmn} + W\prime_{Rmn}} \times 100 \tag{4.5a}$$

In the equation, W_{Lmn} and W_{Rmn} are the left and right signal power, respectively.

The A_{mn} (f1, f2) is the EEG alpha interhemispheric asymmetry. We computed the EEG alpha interhemispheric asymmetry with each channel pair for brain lobes such as the frontal (Fp1, Fp2, F3, F4,F7, F8), temporal (T3, T4, T5, T6), parietal (P3, P4, P7, P8), occipital (O1,O2), and central (C3, C4) individually. Ultimately, we found out the ratios of the differences for the channels, and found out the possible combinations. For instance for Fp1 the ratios were Fp1-Fp2; Fp1-F4; Fp1-F8; Fp1-T4; Fp1-T6; Fp1-P4; Fp1-P8;Fp1-O2; and Fp1-C4. The same combination were applied for the other remaining seven electrodes of the left hemisphere. Basically In the order given above, Fp1 would be replaced by F3 then C3 then P3 then O1 then F7 then T3 and finally T5.
Each algebraic division of the permutations given above returned a dataframe. And as with the previous EEG Data Matrix of absolute band power, this too involved an initial empty dataframe that is to be concatenated with the resultant dataframes to create a large data matrix big enough for our classification model.

4.5 Z-Score Standardization

Our research work makes use of Z-score standardization which makes sure any possible outliers in both of our data matrices gets eliminated. There are various methods for standardizing variables. According to research a standardized variable as found with the usual 'z-score' formula has been converted to have a zero mean and unity variance [97]. A z-score, when expressed in standard deviations units, represents the location of a raw score in terms of their distance from the mean. It helps to measure the likelihood of a score that occurs within a typical standard distribution and allows to compare two scores which could have different thresholds and standard deviations from different samples. Each feature is subjected to standardization separately in z-score. Z-score is calculated by the subtraction of each element value by its mean and then dividing by its corresponding standard deviation. The standardization of our data was performed by the following formulas described in equation 4.6a and 4.7a respectively.:

$$Z = \frac{X_i - \bar{X}}{S} \tag{4.6a}$$

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}, \qquad S = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}} \tag{4.7a}$$

Here X_i represents an element of each feature with "i" as an index. For each feature, the mean and the standard deviation S are calculated respectively. The total number of elements in a feature is represented by N.
The value of z-score shows exactly how many standard deviations we are from the mean. If the z-score is equal to 0 then it is on the mean. On the other hand, a positive z-score indicates that the standard deviation is above the mean whereas negative z-score shows exactly the opposite i.e standard deviation below the mean. In our code base, we have implemented the usage of the "z-score" function. This is an extremely powerful built-in function from the scipy.stats library, a core modular library for python data processing. This function makes the task of standardization

devilishly simple. It required only one line of code execution. It works by taking the dataframe in consideration and calling the *.apply*() method on it and passing on the previously imported Z-score callback function. This one line of code execution effectively returns a dataframe with all the values standardized following the principles of the formula given above.
This procedure is done for both of the data matrices of Absolute power and Alpha-Interhemispheric Asymmetry resulting in two EEG data matrices with no outliers. So far the EEG data matrices for the healthy and MDD patients for both of the features were separate. After the z-score standardization was done, The two EEG data matrices for the Absolute Power feature and the two data matrices for the Alpha-Interhemispheric Asymmetry were combined with each other. As we are performing binary classification, our target variable "Depression" was added to each of the two newly formed data matrices. The value of 0 was input for the Healthy patients and a value for 1 was the input for patients with MDD.

4.6 Feature Selection

Feature Selection is the process of selecting a specific number of input variables from a set of input variables. This is done as a mode of efficiency when developing a predictive classification model. Reducing features like these and giving them as input reduces the computation cost of a predictive model. It also aids in increasing the efficiency of the model, as discussed later on.
Our objective here was to implement a feature-based feature selection. This method involves deducing the relationship of the input variables with the target variable (i.e., the feature we are trying to predict) and selecting those input variables that have a high correlation with the target variable. We also determine the relationship between input variables and take it into consideration. After selecting the input variables that have a high correlation with the target variable, we select the input that are not highly correlated with each other. This is done as input variables with high correlation will induce similar effects on our predictive model and thereby, taking both of them would burden our model with unnecessary computation cost.
Correlation is a statistical analysis that measures the relationship or association along with the direction of the relationship or association too [92]. This analysis is done between a pair of variables; hence, it is a bi-variate analysis. The correlation coefficient can have a value from -1 to 1. They can never be lower than -1 or greater than +1. A value of -1 means that the two variables happen to be perfectly negatively linearly related. On the other hand, a value of 1 means that the two variables are perfectly positively linearly related. A correlation of 0 indicates that there is no linear relation between the two variables. So, the value gives us the degree of association and the sign indicates the direction of the relation between the two variables [93].
The first challenge faced here was to determine which statistical feature selection method we are to implement. Upon research we found that there are essentially two types of feature selection method as stated below.

1.Wrapper feature selection method: In here, we create many models with different subsets of input features and select those features that result in the best performing model according to a performance metric. These methods are uncon-

cerned with the variable types, although they can be computationally expensive.

2.Filter-feature selection method: This method uses statistical techniques to evaluate relationship between input variables and between input variables and the target variable.
Based on our problem domain, filter-feature selection method was chosen due to its effective and robust computational nature and its wide use in predictive model testing.

Next came the question of which statistical measure to use based on the fiter-feature selection method. There are a wide range of statistical measures that we can choose from such as Pearson's, Spearman, Anova, Chi-squared and so on based on the data type of our input and target variable. To determine this, we had to determine the data –type of the input and our target variable. Since we are doing a binary classification problem, both of our input and target variable were numerical. Based on these data-types, Pearson's co-efficient seemed like the best choice for our problem domain. Figure 4.11 further helps us in illustrating our point.

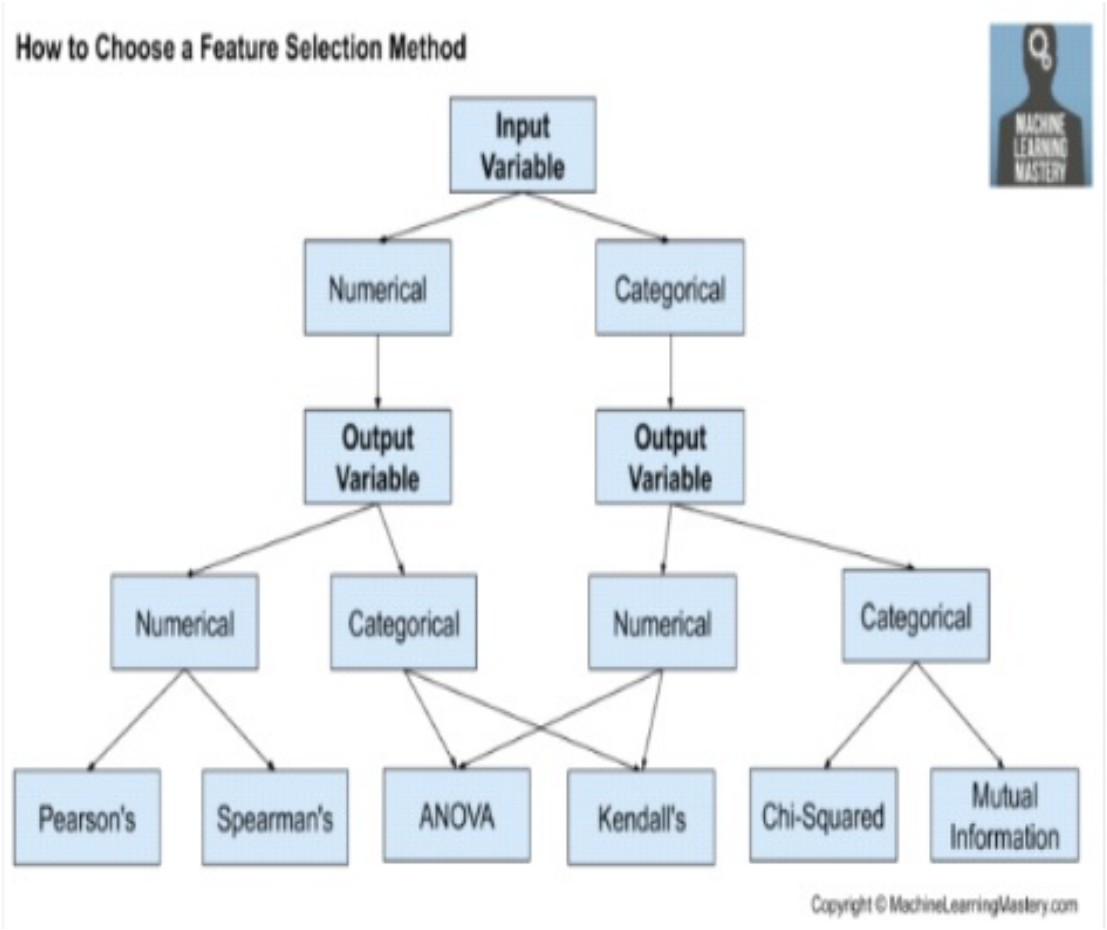


Figure 4.11: How to choose feature selection [63]

There a few assumptions to be made when using the Pearson correlation. Both of the variables need to be normally distributed. There must not be any significant outliers. It is possible to get skewed results if there exist significant outliers as the Pearson's correlation is very sensitive to these outliers. There should be a linear

relationship between the variables and they should also be continuous. A corresponding observation must be made of the dependent variable for each observation of the independent variable; for instance, 50 observations of the variable "X", there should also be 50 observations of the variable "Y". The data should also be Homoscedascity, a scatter plot of the observations will show that there is an equal amount of points on both sides of the line of best fit [93].
Using the Pearson correlation co-efficient we have plotted a Heatmap. The heatmap was imported from the seaborn library, a core module shown in Figure 4.12.

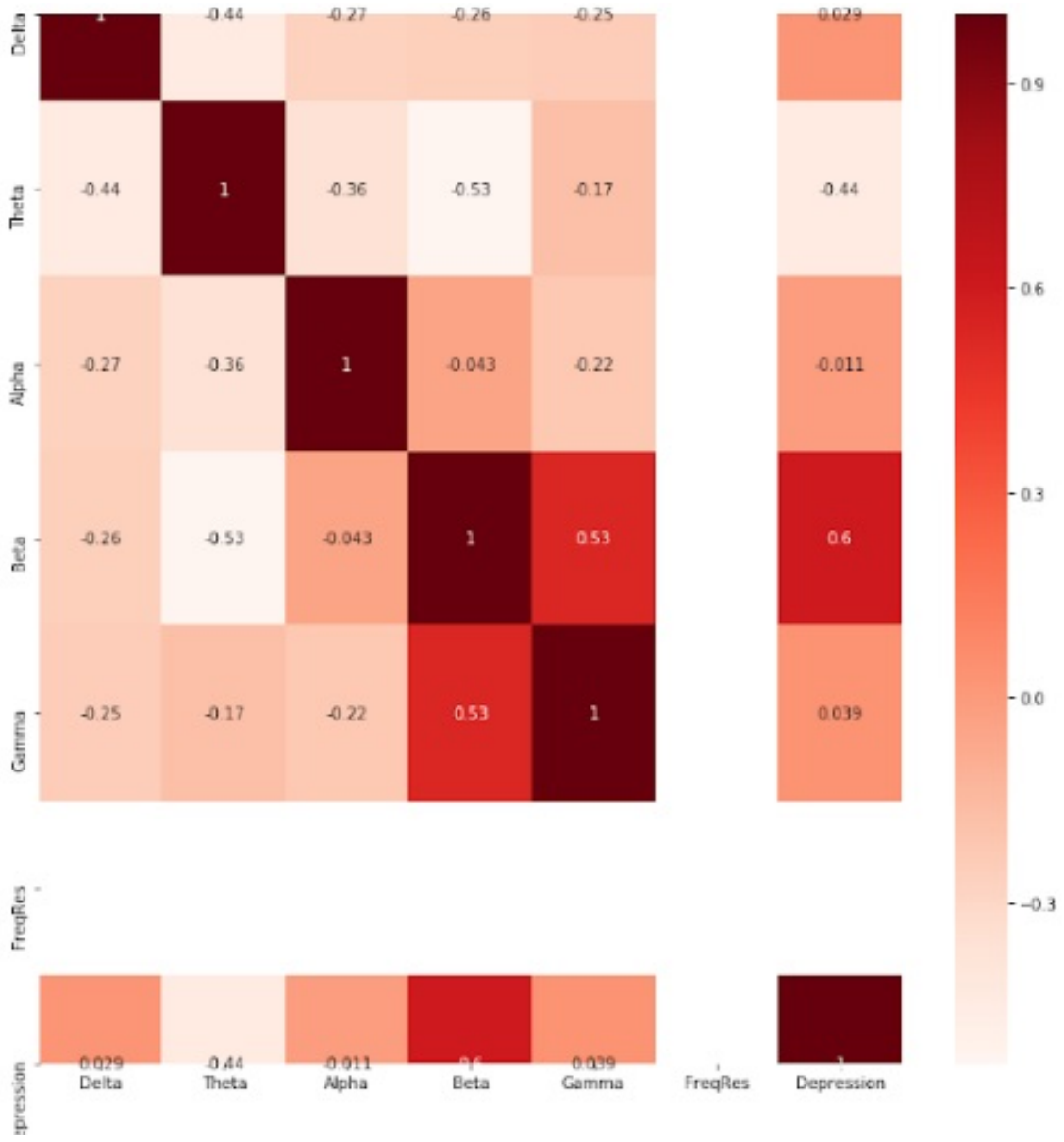


Figure 4.12: Heatmap

This heat map shows the correlation of the variables with the target variable, i.e. the depression column. The intensity of color depicts the gravity of correlation. The darker the color, the higher the correlation and vice-versa. As expected, the depression column has the highest correlation of 1 with itself and is represented on the diagonal line. Each box here also represents a numerical value depicting the

correlational value.
From this heatmap we can see in Figure 4.13, that Beta and Theta waves have some correlation with depression. By setting the threshold value of 0.4 or greater, we have found

Figure 4.13: Heatmap results

Furthermore, we had checked the correlation between the two input variables selected. There is strong negative correlation between Beta and Theta waves. Therefore we were able to choose either one or both of the feature values and test out the accuracy of the classification model based on these features alone.

4.7 Classification Model

A classification model has the functionality to determine an outcome from a set of observed values, i.e., input variables. The outcomes are to be labelled and can be applied to the dataset of concern. In our study, as stated earlier, we had added a column of depression that represented binary values of 0 and 1. 0 was indicative of no depression and 1 was indicative of depression.
There are mainly two approaches to implementing classification algorithms in the field of machine learning. These two are supervised and unsupervised learning.
In a supervised model, a training dataset is applied to the classification algorithm in order to build and train the model. This stage prepares the model to make predictions based on a target variable. This is done using a testing dataset. The testing dataset is then applied to the trained model and predict and classify the target variable initially provided. This type of learning falls under the category of classification.
In our research work, we have implemented the supervised learning method. The training and test data set was derived from the EEG data Matrices of the two features that we have build so far. The dataset was divided into a 70%:30% ratio. 70% of the data belonged to the training dataset while 30% of the data belonged to the testing dataset. This was done using the function $test_train_split()$ imported from the module "sci-kit-learn".
This kind of splitting of the dataset was done for all three of the classification

algorithms, namely Logistic Regression (LR), Support Vector Machine (SVM) and Naïve Bayesian (NB).

4.7.1 Logistic Regression Classification (LR)

The LR classifier gave as likelihood value of l(x) where if l(X) was greater than the threshold of 0.5, the subject was categorized as an MDD patient. But if it was less than the threshold, then the subject was classified as a healthy patient.
Let us say that we have a sample of n independent observations (x_i,y_i), where i = 1, 2, ..., n. Here x_i is the value of the independent variable for the i-th term and y_i is the value of a dichotomous outcome variable. After feature reduction, we have seen that there is a correlation between Beta and Theta waves with depression and because of this, we have combined the absolute power of Beta and Theta waves, which is represented by X (the independent variable as mentioned earlier). The values of depression (either 0 or 1) from the individual EEG channels of each patient represents Y (the dichotomous outcome variable). The number of individual EEG channels of each patient represents our sample size of n. We have divided this sample into a train-test split with the train:test ratio being 70:30 of the entire sample. A random_state value of 101 was given for maximum accuracy.
The logistic model was then trained using the training data as stated earlier and by fitting the model onto the training data using the function fit() imported from the "sci-kit" library. After the model was trained, we used it to predict outcomes on the test data. This was accomplished by using the function predict() on the test data.

4.7.2 Support Vector Machine (SVM)

We have two classes, MDD and healthy patients. Therefore, the hyperplane segregating the two classes is linear. That is why, for this study we have used the SVM classifier with a linear kernel. Using a linear kernel SVM over a non-linear kernel has some advantages, one of them is that it reduces the risk of overfitting in small datasets. It also improves upon the classification performance by remarkably reducing the complexity of the model [68]. As we have mentioned earlier, in logistic regression, the output of the linear function is wrapped using the sigmoid or logistic function, forcing it to take a value within the range of 0 and 1. If the value is greater than a threshold value of 0.5, the subject is assigned a value of 1, otherwise it is assigned a value of 1. In SVM, it is similar in a manner of speaking. If the output is greater than 1, we are able to identify the subject with one class and if the output is -1, we are able to identify it with the other class. The threshold values in SVM are therefore 1 and -1. This range of -1 to 1 acts as the margin [87].
We have seen that there is a correlation between Beta and Theta waves with depression after feature reduction. Because of this, we have combined the absolute power of Beta and Theta waves, which is represented by X. The values of depression (either 0 or 1) from the individual EEG channels of each patient represents Y (the categorical value). We have divided this sample into a train-test split with the train:test ratio being 70:30 of the entire sample. A random_state value of 101 was given for maximum accuracy.
As per the discussion above, the SVM model uses a linear model. The SVM model was then trained using the training data by fitting the model onto the training data

using the function fit(). After the model was trained, we used it to predict outcomes on the test data. This was accomplished by using the function predict() on the test data.

4.7.3 Naive-Bayes Classification (NB)

In our research work we have implemented the use of Gaussian Naïve Bayes algorithm. This can be imported from the "scikit-learn" library. A classifier object instantiated from the GaussianNB is created. This created classifier object is fitted with the training data to train and make the model ready. After the training is done, the predict method is applied to the classifier object with the testing data as parameter.

4.8 10 fold cross-validation

We at first calculated the 10 fold CV for each of the classification algorithms we had used, namely, Logistic regression, Sample vector machine and Naive-Bayes.
This was done using the built-in-function "cross_val_score" imported from the "scikit-learn" library. The function takes in the training data as the parameter along with the model built and the number of folds (i.e. 10) the method is to be cross-validated. The function then returns a two-dimensional array of ten values, each representing the accuracy of the model trained using 9 of the 10 training data samples and tested with the remaining one data sample. An mean and standard deviation of these two returned accuracy values was also calculated. This procedure was repeated for all of the classifier models.
The array of results of the three classifier was then used to draw a box-plot diagram for visual comparison of the effectiveness of the algorithms.

Chapter 5

Results and Discussion

The main purpose of this chapter is to describe results and their interpretation based on the features derived from EEG signals and machine learning algorithms.

Confusion Matrix

A confusion matrix is a table that can help us explain the output of a classification model on a collection of test data for which the true values are known. [94]. This is a table that includes four different combinations of predicted and real values as shown in Figure 5.1. It is extremely useful for measuring Recall, Precision, Specificity, Accuracy and most importantly AUC-ROC Curve.

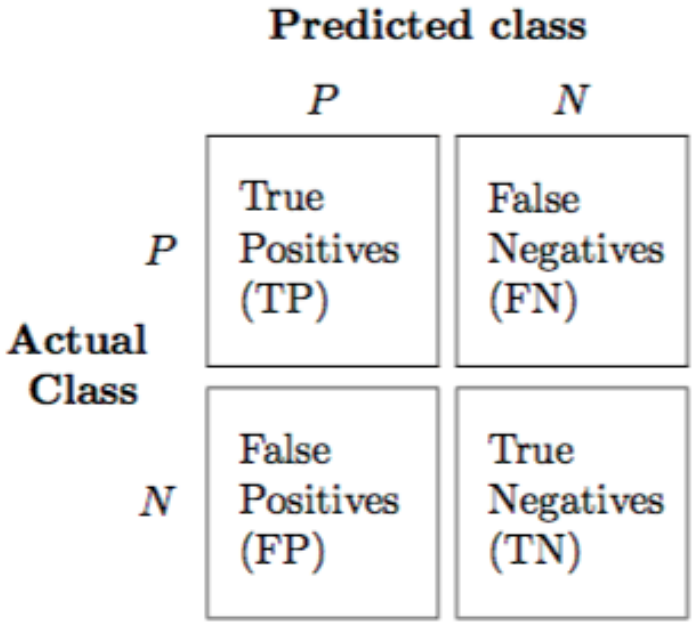


Figure 5.1: Confusion Matrix

The recall (also known as sensitivity) of a classification model is defined as the percentage of true cases (TP) that are correctly classified.

$$Sensitivity = \frac{TP}{TP + FN} \tag{5.1a}$$

The specificity of a classification model is defined as the percentage of true non-cases (TN) correctly classified as non-cases.

$$Specificity = \frac{TP}{FP + TN} \tag{5.2a}$$

The accuracy of a classification model is defined as the percentage of correctly classified cases and non-cases among all the example points.

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{5.3a}$$

The precision of a classifier is defined as the fraction of predicted positives events that are actually positive.

$$Precision = \frac{TP}{TP + FP} \tag{5.4a}$$

The f1 score is the harmonic mean of recall and precision, with a higher score as a better model.

$$F - score = \frac{2TP}{2TP + FP + FN} \tag{5.5a}$$

The equations 5.1a, 5.2a, 5.3a, 5.4a and 5.5a shows Sensitivity, Specificity, Accuracy, Precision ad f-score respectively where TP is the number of true positive, TN is the number of true negative, FP is the number of false positive and FN is the number of false negative.

5.1 Analysis of Supervised Algorithms

In this section, the results of our analysed EEG data using supervised algorithms and generated confusion matrix of each algorithm has been discussed.

5.1.1 Logistic Regression

Table 5.1 shows what we observed:

Table 5.1: Result obtained from LR

Sr.	Precision	Recall	F1-score	Support
0	0.77	0.84	0.80	513
1	0.84	0.77	0.81	573
accuracy			0.81	1086
macro avg	0.81	0.81	0.81	1086
weighted avg	0.81	0.81	0.81	1086

This is what we have obtained from the array shown in Figure 5.2 and the confusion matrix:

```
array([[431,  82],
       [129, 444]], dtype=int64)
```

Figure 5.2: Array obtained

- TP = 431
- FP = 129
- FN = 82
- TN = 444

Table 5.2 shows performance of extracted EEG features based on Logistic Regression (LR) classification while discriminating MDD patients and healthy controls.

Table 5.2: Recall, Specificity, Precision and F1-score of LR

Sr.	Recall	Specificity	Precision	F1-score
Healthy patients	84%	75%	77%	80%
MDD patients	77%	87%	84%	81%

- The accuracy we got from this classifier is 81%.

Receiver Operating Characteristics (ROC)
Performance measurement is an important task when it comes to a classification problem. Predicting the possibility of an event belonging to each class can be more versatile in a classification problem than predicting classes directly. This versatility refers to the way in which probabilities can be evaluated through different standards that cause the model operator to trade-off issues in the model's errors, such as the number of false positives compared to the number of false negatives [63]. We used ROC curve on our Logistic Regression Model to measure the effectiveness of the output quality.
ROC stands for Receiver Operating Characteristic. Usually, ROC curves show true positive rate (TPR) also referred to as sensitivity on the Y axis, and false positive rate (FPR) also referred to as inverted specificity on the X axis. Sometimes, the ROC is also known as relative operating characteristic curve, since it shows an evaluation between TPR and FPR when the measures are altered. This indicates the plot's top left corner is the "ideal" point - a false positive zero score and a true positive score of one. Additionally, the 'steepness' of ROC curves is significant because it is optimal to maximize the true positive rate while decreasing the false positive rate [110]. We plotted a ROC curve for the LR model using roc_curve()

scikit-learn function. This function is used to get the probability array for all the points generated by predict_proba() and the array of actual labels (y).Therefore, a ROC curve can be termed as a susceptibility of fall-out as it helps us figure out how a model can differentiate the classes. It then determines the threshold values that range from 0 to 1. Based on the predicted and actual values we built the confusion matrix and calculated TPR and FPR values. Finally, it returns the threshold array containing the TPR and FPR values for each threshold value. The ROC curve we got is based on plotting the TPR and FPR arrays. The red dotted line is our ROC curve of our model. As we know an ideal classifier stays as far away from that line as possible. To define this divergence, we calculated the area under the ROC curve AUC. AUC stands for Area under the curve which measures the accuracy of the test performed by the LR model. The AUC value lies between 0.5 to 1 where 0.5 is defined as a poor classifier and 1 is an excellent classifier [54].If the classifier is 1 it means it has a good sense of separability and similarly the one close to 0 means the opposite. The higher the value of the AUC, the better the it is as the classifier is close to an ideal classifier.

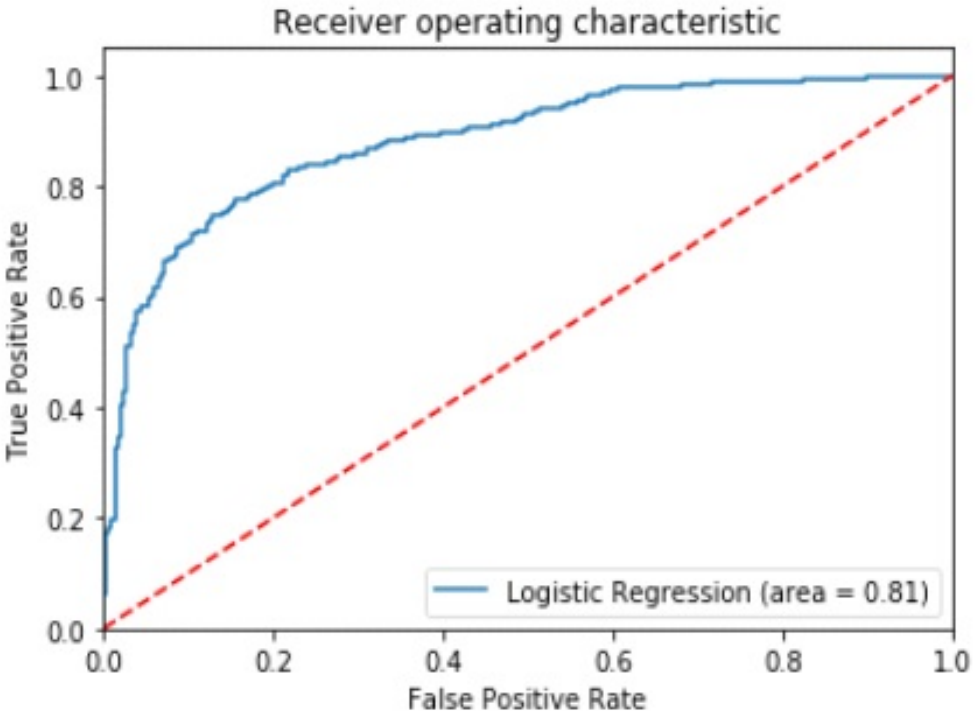


Figure 5.3: ROC curve obtained

The AUC value we got for our ROC curve is 0.81 as shown in Figure 5.3. It means the test performed by our model has a high enough accuracy for it to be categorized as an optimum classifier.

5.1.2 Support Vector Machine

The results we observed are shown in table 5.3:

Table 5.3: Result obtained from SVM

Sr.	Precision	Recall	F1-score	Support
0	0.76	0.87	0.81	513
1	0.86	0.75	0.80	573
accuracy			0.81	1086
macro avg	0.81	0.81	0.81	1086
weighted avg	0.81	0.81	0.81	1086

The array as shown in Figure 5.4 and the confusion matrix we obtained are given below:

```
[[446  67]
 [144 429]]
```

Figure 5.4: Array obtained

- TP = 446
- FN = 67
- FP = 144
- TN = 429

SVM is a machine learning model and like all machine learning models, it is defined as a mathematical model consisting of several parameters which it needs to learn from the data itself. There exists another type of parameter, known as hyper parameters. These cannot be directly learned. Based on trial and error as well as some intuition, these parameters are set by humans. They can affect the performance of a model such as its complexity. Additionally, models may have several hyper parameters. We require to find the right combination of the hyper parameters to obtain the best performance, accuracy and precision from the model. Simply put, one would need to try all the combinations of the hyper parameters and find which combination works best. However, as we are using the Scikit-learn library for our SVM model, we were able to utilize the library's built in function called GridSearchCV [82]. The hyperparameter that we are concerned with is "C". We have obtained the optimum value of "C" using GridSearchCV and it holds a value of 1000. Once again we have trained the model using this value of "C" that we have obtained. After the model was trained, we used it to predict outcomes on the test data. The results we observed now are showed in the table 5.4:

Table 5.4: Result obtained from SVM(2)

Sr.	Precision	Recall	F1-score	Support
0	0.79	0.85	0.82	513
1	0.85	0.80	0.83	573
accuracy			0.82	1086
macro avg	0.82	0.82	0.82	1086
weighted avg	0.82	0.82	0.82	1086

Now the array in Figure 5.5 and the confusion matrix we obtained:

```
[[435  78]
 [115 458]]
```

Figure 5.5: Array obtained

- TP = 435
- FN = 78
- FP = 115
- TN = 458

Table 5.5 shows performance of extracted EEG features based on Support Vector Machine (SVM) classification while discriminating MDD patients and healthy controls.

Table 5.5: Recall, Specificity, Precision and F1-score of SVM

Sr.	Recall	Specificity	Precision	F1-score
Healthy patients	85%	75%	79%	82%
MDD patients	80%	89%	85%	83%

- The accuracy we got from this classifier is 82%.

5.1.3 Naïve-Bayes

Classification report: Table 5.6 shows the results we obtained.

Table 5.6: Result obtained from NB

Sr.	Precision	Recall	F1-score	Support
0	0.80	0.80	0.80	513
1	0.82	0.82	0.82	573
accuracy			0.81	1086
macro avg	0.81	0.81	0.81	1086
weighted avg	0.81	0.81	0.81	1086

Now the array in the Figure 5.6 and the confusion matrix we obtained:

```
array([[411, 102],
       [105, 468]], dtype=int64)
```

Figure 5.6: Array obtained

- TP = 411
- FN = 102
- FP = 105
- TN = 468

Table 5.7 shows performance of extracted EEG features based on Naıve-Bayes (NB) classification while discriminating MDD patients and healthy controls.

Table 5.7: Recall, Specificity, Precision and F1-score of NB

Sr.	Recall	Specificity	Precision	F1-score
Healthy patients	80%	71%	80%	80%
MDD patients	82%	91%	82%	82%

- The accuracy we got from this classifier is 81%.

5.2 10 fold cross-validation

The array of results obtained from the Logistic regression model was

$$[0.77952756, 0.7992126, 0.77952756, 0.81496063, 0.81102362,$$
$$0.78740157, 0.76771654, 0.80555556, 0.8015873, 0.81746032]$$

The mean and standard deviation for logistic regression model are:

- Mean = 0.7963973253343332
- STD = 0.016104653155216208

The array of results obtained from the SVM model was

$$[0.78740157, 0.79133858, 0.78740157, 0.81102362, 0.81102362, \\ 0.80314961, 0.76377953, 0.82142857, 0.78968254, 0.82539683]$$

The mean and standard deviation for SVM model are:

- Mean = 0.7991626046744156
- STD = 0.016104653155216208

The array of results obtained from the Naïve-Bayes model was

$$[0.81102362, 0.81102362, 0.77952756, 0.79527559, 0.80314961, \\ 0.77952756, 0.77952756, 0.79365079, 0.8015873, 0.84126984]$$

The mean and standard deviation for NB model are:

- Mean = 0.7995563054618172
- STD = 0.018101674405661353

A general view of the results as well as the mean and standard deviation show that the Naïve-Bayes model can be a good classifier for our depression classification task. A better and more effective way of comparison is to use a box-plot diagram to visually see and deduce the performance of the classification algorithms and decide which one is the best among the three.

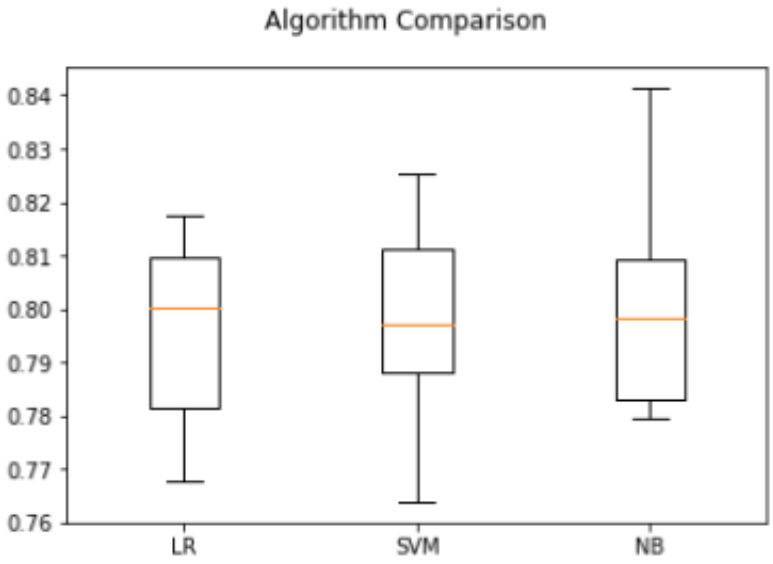


Figure 5.7: Algorithm comparison

From the box-plot diagram as shown in Figure 5.7, it is evident that the Naive-bayes classifier has the highest accuracy among the other three.

Chapter 6

Conclusion and Future Work

This chapter concludes by restating the problem, challenges faced in this research and improvements and further work to be done in this field.

6.1 Conclusion

All in all, depression is a disturbing disease and can prove to be mortal if not nursed in time. Therefore, it is crucial and necessary to cure this plight while it is in the initial stages or else it might aggravate to a worst ending. As already elaborated in the paper people measure depression solely based on doctor's skills and practices over the issue. Unfortunately, there are times when a doctor might fail to address the severity of the disease. Hence, the non-partisan method that we developed to detect the MDD patients using their brain signals will be really accessible in the long run. In contrast to all the previous complex methods of EEG signals, using sequence learning and deep representation, the method we used in this paper is more convenient at the same time accurate enough. Candidates will not feel uncomfortable while we place the electrodes on their scalp since it is not for a long time period plus they can just use it while they are just in resting state.

6.2 Limitations & Future work

We faced few challenges during this research. The primary challenge we had faced was working with the EEG data. We needed to find ways to process the EEG data efficiently with a suitable noise reduction technique. Most of the work done using EEG signals are done using software's like EEGLAB or MATLAB. Our implementation done in python provided many coding challenges to be overcame.
Another challenge was as our research involved the immense study and the gathering of domain knowledge of the EEG signals and its acquisition, how depression was related with EEG signals and what features are important to our required domain. Coming from a CSE background, gathering medical knowledge and working with it proved to be quite challenging.

Our initial plan was to collect the EEG data on our own. Collecting the EEG signal proved to be extremely difficult for us and thus we had to work with EEG signals collected from the internet. The limitations here include the availability of efficient

hardware, i.e., a headgear and moreover the training required to handle such sensitive equipment. As we were intended to collect EEG signal from depressed patients, we needed access to hospitals to allow us to gather the required signals from patients suffering from MDD admitted to the hospital. It was not possible to gain the permission from these hospitals as patient data as well as interaction with patients with MDD is considered a highly sensitive topic in our country.

In the future we hope to overcome limitations like this and work on using our own collected data. We plan to be proficient at the experimental setup required to accurately collect the EEG signal, which includes familiarity with headgear and its usage, understanding the correct stimulus to give to a patient and so on. We plan on working with various other classification algorithms such as KNN and CNN and compare classifications with these to the classifications we have already done.
In the long run having a genetic indicator will be really needed for the enhancement of sure related to MDD. The project we worked on, being a valid bio-marker for MDD, holds the prospective to be handed to the psychiatrists and counsellors. They can use it along with their knowledge of the symptoms of MDD and find a more effective treatment for the victims. Optimistically, the dread and panic behind depression can be eradicated.

Bibliography

[1] M. S. Bartlett, "Periodogram analysis and continuous spectra", *Biometrika*, vol. 37, no. 1/2, pp. 1–16, 1950.

[2] L. S. Radloff, "A self-report depression scale for research in the general population", *Applied psychol Measurements*, vol. 1, pp. 385–401, 1977.

[3] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform", *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[4] W. J. Ray and H. W. Cole, "Eeg alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes", *Science*, vol. 228, no. 4700, pp. 750–752, 1985.

[5] I. Elkin, M. T. Shea, J. T. Watkins, S. D. Imber, S. M. Sotsky, J. F. Collins, D. R. Glass, P. A. Pilkonis, W. R. Leber, J. P. Docherty, *et al.*, "National institute of mental health treatment of depression collaborative research program: General effectiveness of treatments", *Archives of general psychiatry*, vol. 46, no. 11, pp. 971–982, 1989.

[6] A. Oppenheim, R. Schafer, and J. Buck, "Discrete-time signal processing (vol. 2) prentice-hall englewood cliffs", 1989.

[7] O. Solomon Jr, "Psd computations using welch's method", *STIN*, vol. 92, p. 23 584, 1991.

[8] J.-L. Nandrino, L. Pezard, J. Martinerie, F. El Massioui, B. Renault, R. Jouvent, J.-F. Allilaire, and D. Widlöcher, "Decrease of complexity in eeg as a symptom of depression.", *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, 1994.

[9] C. J. Murray and A. D. Lopez, *Alternative visions of the future: Projecting mortality and disability, 1990–2020*, 1996.

[10] T. M. Mitchell *et al.*, *Machine learning*, 1997.

[11] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications", *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[12] R. C. Arnau, M. W. Meagher, M. P. Norris, and R. Bramson, "Psychometric evaluation of the beck depression inventory-ii with primary care medical patients.", *Health Psychology*, vol. 20, no. 2, p. 112, 2001.

[13] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics", *IEEE Transactions on speech and audio processing*, vol. 9, no. 5, pp. 504–512, 2001.

[14] C. L. Keyes, "The mental health continuum: From languishing to flourishing in life", *Journal of health and social behavior*, pp. 207–222, 2002.

[15] K. Rost, J. L. Smith, and M. Dickinson, "The effect of improving primary care depression management on employee absenteeism and productivity a randomized trial", *Medical care*, vol. 42, no. 12, p. 1202, 2004.

[16] G. L. Wallstrom, R. E. Kass, A. Miller, J. F. Cohn, and N. A. Fox, "Automatic correction of ocular artifacts in the eeg: A comparison of regression-based and component-based methods", *International journal of psychophysiology*, vol. 53, no. 2, pp. 105–119, 2004.

[17] M. Hall, "A decision tree-based attribute weighting filter for naive bayes", in *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Springer, 2006, pp. 59–70.

[18] A. Halfin, "Depression: The benefits of early and appropriate treatment.", *The American journal of managed care*, vol. 13, no. 4 Suppl, S92–7, 2007.

[19] J.-S. Lee, B.-H. Yang, J.-H. Lee, J.-H. Choi, I.-G. Choi, and S.-B. Kim, "Detrended fluctuation analysis of resting eeg in depressed outpatients and healthy controls", *Clinical Neurophysiology*, vol. 118, no. 11, pp. 2489–2496, 2007.

[20] W. H. Organization *et al.*, "Who-aims report on mental health system in bangladesh", 2007.

[21] A. C. Deslandes, H. De Moraes, F. A. Pompeu, P. Ribeiro, M. Cagy, C. Capitão, H. Alves, R. A. Piedade, and J. Laks, "Electroencephalographic frontal asymmetry and depressive symptoms in the elderly", *Biological psychology*, vol. 79, no. 3, pp. 317–322, 2008.

[22] L. Illing, *Fourier analysis*, 2008.

[23] R. Bluhm, P. Williamson, R. Lanius, J. Théberge, M. Densmore, R. Bartha, R. Neufeld, and E. Osuch, "Resting state default-mode network connectivity in early depression using a seed region-of-interest analysis: Decreased connectivity with caudate nucleus", *Psychiatry and clinical neurosciences*, vol. 63, no. 6, pp. 754–761, 2009.

[24] K. Gausia, C. Fisher, M. Ali, and J. Oosthuizen, "Antenatal depression and suicidal ideation among rural bangladeshi women: A community-based study", *Archives of women's mental health*, vol. 12, pp. 351–8, Jun. 2009. DOI: 10.1007/s00737-009-0080-7.

[25] ——, "Antenatal depression and suicidal ideation among rural bangladeshi women: A community-based study", *Archives of women's mental health*, vol. 12, no. 5, p. 351, 2009.

[26] V. A. Grin-Yatsenko, I. Baas, V. A. Ponomarev, and J. D. Kropotov, "Eeg power spectra at early stages of depressive disorders", *Journal of Clinical Neurophysiology*, vol. 26, no. 6, pp. 401–406, 2009.

[27] B. Kolb and I. Q. Whishaw, *Fundamentals of human neuropsychology*. Macmillan, 2009.

[28] M. McCrea, G. L. Iverson, T. W. McAllister, T. A. Hammeke, M. R. Powell, W. B. Barr, and J. P. Kelly, "An integrated review of recovery after mild traumatic brain injury (mtbi): Implications for clinical management", *The Clinical Neuropsychologist*, vol. 23, no. 8, pp. 1368–1390, 2009.

[29] G. I. Papakostas and M. Fava, "Does the probability of receiving placebo influence clinical trial outcome? a meta-regression of double-blind, randomized clinical trials in mdd", *European Neuropsychopharmacology*, vol. 19, no. 1, pp. 34–40, 2009.

[30] C. Chew and G. Eysenbach, "Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak", *PloS one*, vol. 5, no. 11, 2010.

[31] P. Cuijpers, A. van Straten, J. Schuurmans, P. van Oppen, S. D. Hollon, and G. Andersson, "Psychotherapy for chronic major depression and dysthymia: A meta-analysis", *Clinical psychology review*, vol. 30, no. 1, pp. 51–62, 2010.

[32] V. A. Grin-Yatsenko, I. Baas, V. A. Ponomarev, and J. D. Kropotov, "Independent component approach to the analysis of eeg recordings at early stages of depressive disorders", *Clinical Neurophysiology*, vol. 121, no. 3, pp. 281–289, 2010.

[33] G. E. Simon and R. H. Perlis, "Personalized medicine for depression: Can we match patients with treatments?", *American Journal of Psychiatry*, vol. 167, no. 12, pp. 1445–1455, 2010.

[34] M. Chawla, "Pca and ica processing methods for removal of artifacts and noise in electrocardiograms: A survey and comparison", *Applied Soft Computing*, vol. 11, no. 2, pp. 2216–2226, 2011.

[35] J.-P. Lépine and M. Briley, "The increasing burden of depression", *Neuropsychiatric disease and treatment*, vol. 7, no. Suppl 1, p. 3, 2011.

[36] D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, "A 'non-parametric'version of the naive bayes classifier", *Knowledge-Based Systems*, vol. 24, no. 6, pp. 775–784, 2011.

[37] F.-E. Vederine, M. Wessa, M. Leboyer, and J. Houenou, "A meta-analysis of whole-brain diffusion tensor imaging studies in bipolar disorder", *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 35, no. 8, pp. 1820–1826, 2011.

[38] I. Winkler, S. Haufe, and M. Tangermann, "Automatic classification of artifactual ica-components for artifact removal in eeg signals", *Behavioral and Brain Functions*, vol. 7, no. 1, p. 30, 2011.

[39] N. Zamperetti, R. Bellomo, P. Piccinni, and C. Ronco, "Reflections on transplantation waiting lists", *The Lancet*, vol. 378, no. 9791, pp. 632–635, 2011.

[40] J. Zhang, J. Wang, Q. Wu, W. Kuang, X. Huang, Y. He, and Q. Gong, "Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder", *Biological psychiatry*, vol. 70, no. 4, pp. 334–342, 2011.

[41] R. Chunara, J. R. Andrews, and J. S. Brownstein, "Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak", *The American journal of tropical medicine and hygiene*, vol. 86, no. 1, pp. 39–45, 2012.

[42] R. S. Hock, F. Or, K. Kolappa, M. D. Burkey, P. J. Surkan, and W. W. Eaton, "A new resolution for global mental health", *Lancet (London, England)*, vol. 379, no. 9824, p. 1367, 2012.

[43] S. D. Puthankattil and P. K. Joseph, "Classification of eeg signals in normal and depression conditions by ann using rwe and signal entropy", *Journal of Mechanics in Medicine and biology*, vol. 12, no. 04, p. 1 240 019, 2012.

[44] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[45] A. Gramfort, M. Luessi, E. Larson, D. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen, "Meg and eeg data analysis with mne-python", *Frontiers in Neuroscience*, vol. 7, p. 267, 2013, ISSN: 1662-453X. DOI: 10.3389/fnins.2013.00267. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2013.00267.

[46] B. Hosseinifard, M. H. Moradi, and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from eeg signal", *Computer methods and programs in biomedicine*, vol. 109, no. 3, pp. 339–345, 2013.

[47] B. de Kwaasteniet, E. Ruhe, M. Caan, M. Rive, S. Olabarriaga, M. Groefsema, L. Heesink, G. van Wingen, and D. Denys, "Relation between structural and functional connectivity in major depressive disorder", *Biological psychiatry*, vol. 74, no. 1, pp. 40–47, 2013.

[48] W. H. Organization *et al.*, *World health organization sixty-sixth world health assembly. 2013*, 2013.

[49] G. Chandrashekar and F. Sahin, "A survey on feature selection methods", *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[50] A. A. Fingelkurts and A. A. Fingelkurts, *Altered structure of dynamic electroencephalogram oscillatory pattern in major depression*, Dec. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0006322314009950.

[51] M. D. Hossain, H. U. Ahmed, W. A. Chowdhury, L. W. Niessen, and D. S. Alam, "Mental disorders in bangladesh: A systematic review", *BMC psychiatry*, vol. 14, no. 1, p. 216, 2014.

[52] V. M. Prieto, S. Matos, M. Alvarez, F. Cacheda, and J. L. Oliveira, "Twitter: A good place to detect health conditions", *PloS one*, vol. 9, no. 1, 2014.

[53] S. M. Rice, J. Goodall, S. E. Hetrick, A. G. Parker, T. Gilbertson, G. P. Amminger, C. G. Davey, P. D. McGorry, J. Gleeson, and M. Alvarez-Jimenez, "Online and social networking interventions for the treatment of depression in young people: A systematic review", *Journal of medical Internet research*, vol. 16, no. 9, e206, 2014.

[54] N. Wald and J. Bestwick, "Is the area under an roc curve a valid measure of the performance of a screening or diagnostic test?", *Journal of medical screening*, vol. 21, no. 1, pp. 51–56, 2014.

[55] L.-L. Zeng, H. Shen, L. Liu, and D. Hu, "Unsupervised classification of major depression using functional connectivity mri", *Human brain mapping*, vol. 35, no. 4, pp. 1630–1641, 2014.

[56] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. Koh, and A. Adeli, "Computer-aided diagnosis of depression using eeg signals", *European neurology*, vol. 73, no. 5-6, pp. 329–336, 2015.

[57] S. Balani and M. De Choudhury, "Detecting and characterizing mental health related self-disclosure in social media", in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 2015, pp. 1373–1378.

[58] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, "Clpsych 2015 shared task: Depression and ptsd on twitter", in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 31–39.

[59] A. Islam and T. Biswas, "Mental health and the health system in bangladesh: Situation analysis of a neglected domain", *Am J Psychiatry Neurosci*, vol. 3, no. 4, pp. 57–62, 2015.

[60] X. Li, B. Hu, J. Shen, T. Xu, and M. Retcliffe, "Mild depression detection of college students: An eeg-based solution with free viewing tasks", *Journal of medical systems*, vol. 39, no. 12, p. 187, 2015.

[61] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, "Eeg artifact elimination by extraction of ica-component features using image processing algorithms", *Journal of neuroscience methods*, vol. 243, pp. 84–93, 2015.

[62] P. Resnik, W. Armstrong, L. Claudino, and T. Nguyen, "The university of maryland clpsych 2015 shared task system", in *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 2015, pp. 54–60.

[63] J. Brownlee, "Machine learning mastery with python", *Machine Learning Mastery Pty Ltd*, pp. 100–120, 2016.

[64] M. Conway and D. O'Connor, "Social media, big data, and mental health: Current advances and ethical implications", *Current opinion in psychology*, vol. 9, pp. 77–82, 2016.

[65] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media", in *Proceedings of the 2016 CHI conference on human factors in computing systems*, 2016, pp. 2098–2110.

[66] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20 150 202, 2016.

[67] H. Marzbani, H. R. Marateb, and M. Mansourian, "Neurofeedback: A comprehensive review on system design, methodology and clinical applications", *Basic and clinical neuroscience*, vol. 7, no. 2, p. 143, 2016.

[68] W. Mumtaz, *Mdd patients and healthy controls eeg data (new)*, Nov. 2016. [Online]. Available: https://figshare.com/articles/EEG_Data_New/4244171.

[69] R. H. Perlis, "Abandoning personalization to get to precision in the pharmacotherapy of depression", *World Psychiatry*, vol. 15, no. 3, pp. 228–235, 2016.

[70] M. M. Pervin and N. Ferdowshi, "Suicidal ideation in relation to depression, loneliness and hopelessness among university students", *Dhaka University journal of biological sciences*, vol. 25, no. 1, pp. 57–64, 2016.

[71] A. Picardi, I. Lega, L. Tarsitani, M. Caredda, G. Matteucci, M. Zerella, R. Miglio, A. Gigantesco, M. Cerbo, A. Gaddini, *et al.*, "A randomised controlled trial of the effectiveness of a program for early detection and treatment of depression in primary care", *Journal of affective disorders*, vol. 198, pp. 96–101, 2016.

[72] S. Siuly, Y. Li, and Y. Zhang, "Significance of eeg signals in medical and health research", in *EEG Signal Analysis and Classification*, Springer, 2016, pp. 23–41.

[73] M. Tabakcıoğlu, H. Çizmeci, and D. Ayberkin, "Neurosky eeg biosensor using in education", *International Journal of Applied Mathematics Electronics and Computers*, no. Special Issue-1, pp. 76–78, 2016.

[74] M. L. Birnbaum, S. K. Ernala, A. F. Rizvi, M. De Choudhury, and J. M. Kane, "A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals", *Journal of medical Internet research*, vol. 19, no. 8, e289, 2017.

[75] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt, "Detecting depression and mental illness on social media: An integrative review", *Current Opinion in Behavioral Sciences*, vol. 18, pp. 43–49, 2017.

[76] S. I. Hay, A. A. Abajobir, K. H. Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, R. S. Abdulkader, A. M. Abdulle, T. A. Abebo, S. F. Abera, *et al.*, "Global, regional, and national disability-adjusted life-years (dalys) for 333 diseases and injuries and healthy life expectancy (hale) for 195 countries and territories, 1990–2016: A systematic analysis for the global burden of disease study 2016", *The Lancet*, vol. 390, no. 10100, pp. 1260–1344, 2017.

[77] Y. Ophir, C. S. Asterhan, and B. B. Schwarz, "Unfolding the notes from the walls: Adolescents' depression manifestations on facebook", *Computers in Human Behavior*, vol. 72, pp. 96–107, 2017.

[78] R. Strawbridge, A. H. Young, and A. J. Cleare, "Biomarkers for depression: Recent insights, current challenges and future prospects.", *Neuropsychiatric disease and treatment*, 2017.

[79] A. E. Aladağ, S. Muderrisoglu, N. B. Akbas, O. Zahmacioglu, and H. O. Bingol, "Detecting suicidal ideation on forums: Proof-of-concept study", *Journal of medical Internet research*, vol. 20, no. 6, e215, 2018.

[80] L. M. Andersen, "Group analysis in mne-python of evoked responses from a tactile stimulation paradigm: A pipeline for reproducibility at every step of processing, going from individual sensor space representations to an across-group source space representation", *Frontiers in Neuroscience*, vol. 12, p. 6, 2018.

[81] J. Kim, *Christian Ferras and His Struggle with Depression*. California State University, Long Beach, 2018.

[82] S. Putatunda and K. Rama, "A comparative analysis of hyperopt as against other approaches for hyper-parameter optimization of xgboost", in *Proceedings of the 2018 International Conference on Signal Processing and Machine Learning*, 2018, pp. 6–10.

[83] A. M. Somily, "Photometry of recent outbursts of dwarf nova al com", 2018.

[84] C.-T. Wu, D. G. Dillon, H.-C. Hsu, S. Huang, E. Barrick, and Y.-H. Liu, "Depression detection using relative eeg power induced by emotionally positive images and a conformal kernel support vector machine", *Applied Sciences*, vol. 8, no. 8, p. 1244, 2018.

[85] *A deep dive into brainwaves: Brainwave frequencies explained*, Apr. 2019. [Online]. Available: https://choosemuse.com/blog/a-deep-dive-into-brainwaves-brainwave-frequencies-explained-2/.

[86] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, "Early detection of depression: Social network analysis and random forest techniques", *Journal of medical Internet research*, vol. 21, no. 6, e12554, 2019.

[87] H. Canbaz and K. Polat, "Fault detection of cnc machines from vibration signals using machine learning methods", in *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*, Springer, 2019, pp. 365–374.

[88] S. Dankwa and W. Zheng, "Special issue on using machine learning algorithms in the prediction of kyphosis disease: A comparative study", *Applied Sciences*, vol. 9, no. 16, p. 3322, 2019.

[89] gergana007, *Depression: A neglected public health domain in bangladesh*, Oct. 2019. [Online]. Available: https://www.globallyminded.org/home/depression-a-neglected-public-health-domain-in-bangladesh/.

[90] X. Jiang, G.-B. Bian, and Z. Tian, "Removal of artifacts from eeg signals: A review", *Sensors*, vol. 19, no. 5, p. 987, 2019.

[91] D. K, *Top 4 advantages and disadvantages of support vector machine or svm*, Jun. 2019. [Online]. Available: https://medium.com/@dhiraj8899/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107.

[92] H. Kekana and G. Landwehr, "Wind capacity factor calculator", *Journal of Energy in Southern Africa*, vol. 30, no. 2, pp. 118–125, 2019.

[93] J. Magiya, *Pearson coefficient of correlation explained.* Nov. 2019. [Online]. Available: https://towardsdatascience.com/pearson-coefficient-of-correlation-explained-369991d93404.

[94] A. S. Malik and W. Mumtaz, *Electroencephalography-based diagnosis of depression*, May 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128174203000084.

[95] J. J. Newson and T. C. Thiagarajan, "Eeg frequency bands in psychiatric disorders: A review of resting state studies", *Frontiers in human neuroscience*, vol. 12, p. 521, 2019.

[96] Y. Park, W. Jung, S. Kim, H. Jeon, and S.-H. Lee, "Frontal alpha asymmetry correlates with suicidal behavior in major depressive disorder", *Clinical Psychopharmacology and Neuroscience*, vol. 17, no. 3, p. 377, 2019.

[97] S. Thangavelu, S. Akshaya, K. Naetra, K. S. AC, and V. Lasya, "Feature selection in cancer genetics using hybrid soft computing", in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, IEEE, 2019, pp. 734–739.

[98] S. Turnbull and J. Guthrie, "Simplifying the management of complexity: As achieved in nature", *Journal of Behavioural Economics and Social Systems*, vol. 1, no. 1, pp. 51–73, 2019.

[99] E. Urtnasan, Y. Kim, J. Park, H. Choi, K. Lee, *et al.*, "Automatic screening of plms patient based on deep learning model using an electrocardiogram", *Sleep Medicine*, vol. 64, S395, 2019.

[100] C. Molnar, *Interpretable machine learning*, Mar. 2020. [Online]. Available: https://christophm.github.io/interpretable-ml-book/logistic.html#advantages-and-disadvantages.

[101] *Naive bayes algorithm: How it works: Basic models: Advantages*, Mar. 2020. [Online]. Available: https://www.educba.com/naive-bayes-algorithm/.

[102] A. Pietrangelo, *Types of depression: 9 forms of depression and their symptoms*, Mar. 2020. [Online]. Available: https://www.healthline.com/health/types-of-depression.

[103] *hanning sacchetto, llp • arcadia • california • hanningsacchetto.com*. [Online]. Available: https://www.tuugo.us/Companies/hanning-sacchetto-llp7/0310006442941.

[104] F. M. Abujalala, H. R. Abulifa, and A. M. Abushaala, "Training a support vector machine classifier",

[105] *Biology - human brain*. [Online]. Available: https://www.tutorialspoint.com/biology_part2/biology_human_brain.htm.

[106] *Brain anatomy, anatomy of the human brain*. [Online]. Available: https://mayfieldclinic.com/pe-anatbrain.htm.

[107] *Eeg accessories mcscap*. [Online]. Available: https://mks.ru/en/products/mcscap/.

[108] N. Kumar, *Advantages and disadvantages of logistic regression in machine learning*. [Online]. Available: http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages-of.html.

[109] *Mne 0.20.0 documentation*. [Online]. Available: https://mne.tools/stable/index.html.

[110] *Receiver operating characteristic (roc)¶*. [Online]. Available: https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html.