Detection and Prediction of Epileptic Seizure Using Different Machine Learning Classifiers

by

Subha Sumaiya B Jamal

18101636

Nuna Tasnim

18101512

Md. Ashikur Rahman

18101074

Afnan Ahmed Crystal

21141046

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

Spring 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/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.

Students' Full Names & Signatures:

Subha Sumaiya Jamal

Subha Sumaiya B Jamal

18101636

Md. Ashikur Rahman

Md. Ashikur Rahman

18101074

Nuna Tasnim

Nuna Tasnim

18101512

Afnan Ahmed Crystal

Afnan Ahmed Crystal

21141046

Approval

The thesis/project titled "Detection and Prediction of Epileptic Seizure Using Different Machine Learning Classifiers" submitted by

1. Subha Sumaiya B Jamal (18101636)

2. Nuna Tasnim (18101512)

3. Md. Ashikur Rahman (18101074)

4. Afnan Ahmed Crystal (21141046)

Of Spring, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on May 29, 2022.

Supervisor:
(Member)

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

Co-Supervisor:
(Member)

Dewan Ziaul Karim
Lecturer
Department of Computer Science and Engineering
BRAC University

Head of Department:

(Chair)

<div style="text-align: center">

_____

Sadia Hamid Kazi, PhD

Chairperson and Associate Professor

Department of Computer Science and Engineering

BRAC University

</div>

Abstract

Analyzing neural signals produced by neurons in the brain, epilepsy can be diagnosed. An electroencephalogram (EEG) measures brain electrical activity, and studying EEG data in order to detect epileptic seizures in their early phases is an important aspect of epilepsy research. Despite optimal medication management, around one-third of epileptic patients continue to experience seizures. As a result, detecting epileptic seizures has become increasingly important in the field of research in recent years. It has been observed that machine learning has a revolutionary effect on classifying EEG data, seizure detection, and identifying sensible patterns without performance deterioration.This study provides a comprehensive summary of works on automated epileptic seizure recognition utilizing a variety of machine learning techniques, including SVC, Logistic Regression, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting, and Multilayer Perceptron (MLP). A dataset provided by the UCI Machine Learning Repository is used to train the model. We considered the F1 score to be our most important performance metric since it handles unbalanced data sets effectively by comparing both precision and recall. Our research found that the Random Forest Classifier achieved a higher F1 score of 97.8261% with a precision of 96.746% and an F1 score of 96.4402% compared to other classifiers when five groups of people were considered. Later, we implemented PCA and clustering to determine if we could improve the Random Forest Classifier's performance. After performing PCA for dimension reduction and K-means for clustering, we compare the F1 scores and cannot find any significant difference, which implies that our data set does not need any further clustering. Hence, the findings suggest that classification performance remains the same after implementing dimension reduction and clustering.

Keywords: EEG; Epilepsy; Seizure; Feature extraction; Classification; PCA

Acknowledgement

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thoughts behind the Prediction Model

Epilepsy is a very common neurological illness, affecting around 70 million people worldwide. 85% of those infected live in underdeveloped nations, and 2.4 million new cases of the disease are diagnosed every year. At least half of all epilepsy cases begin in youth or adolescence, but the disease can strike anyone at any age. For a variety of reasons, including the fact that people with the disease are two or three times more likely to die early than those who do not, epilepsy research has always been important in biomedical research. The difference between epilepsy and epileptic seizures is that the latter is caused by brain abnormalities that indirectly impact a patient's health. These occur suddenly and decrease the life expectancy rate of an individual [30].

Analyzing and classifying EEG data in order to detect seizures in their early phases is an important element of epilepsy research. If a seizure is recognized early enough, neurostimulation can be used to prevent it from progressing and spreading to other areas of the brain. As a result, finding an effective method for automatic seizure detection is necessary [24].

1.2   Motivation

Our goal is to improve and contribute to the field of detecting epileptic seizures, one of the most common brain illnesses. We read and examined various research articles on this subject and decided to take advantage of our opportunity to contribute to this field of research with our own approaches. People with neurological disorders, such as epilepsy, have a difficult time coping with anomalous behavior, activities, and sensations. It is critical to recognize epileptic seizures in order to aid patients in epilepsy diagnosis so that they can receive first aid and be certain of their medical conditions. In this situation, an electroencephalogram (EEG) signal can be used to reveal structural and functional information about the brain, which can aid in identifying malfunctions and abnormalities. There are various varieties of epileptic seizures, each with its own set of symptoms. If it is discovered sooner, first aid can be administered to reduce the severity of the situation [24]. Many study papers show that signal measurement is a very effective approach to identify epilepsy, or to put it another way, it is the only way to point out the disease. We can successfully detect whether a brain has been impacted by an epileptic seizure or not using EEG data analysis and certain other approaches such as signal feature extraction and classification.

1.3   Thesis Overview

An epileptic seizure is a period of symptoms induced by abnormal or excessive neuronal activity in the brain, characterized by uncontrolled movements, shaking, and electrical motions in all or part of the body, as well as varied levels of consciousness or merely a short momentary loss of awareness [45]. These episodes usually last less than a minute or two, but they take some time to return to normal. It's possible that you'll lose bladder control as a result of this seizure [45]. There are several actions that can cause an epileptic seizure to occur. However, in the majority of cases, it is caused by unnoticed or unknown factors, or even minor sleep deprivation.

2

When it comes to epilepsy, there are two types: provoked and unprovoked. In their lifetime, up to 10% of people will suffer at least one epileptic episode [45]. According to research, about 3.5 out of every 10,000 people experience triggered seizures each year, while 4.2 out of every 10,000 experience unprovoked seizures. After a single seizure, the likelihood of another is approximately 50%. Also, about 1% of the population suffers from epilepsy at any one time, and 4% suffer from it at some point in their lives [21] [17]. Furthermore, more or less 80% of people suffering from epilepsy live in underdeveloped nations. For safety reasons, many places even mandate residents to cease driving until they haven't had a seizure in a particular amount of time. To assist someone who is experiencing an epileptic seizure, the first step is to determine whether or not the person is experiencing an epileptic seizure. Electrical signals produced by the human brain (also known as EEG signals) can be separated into several frequency bands (e.g., gamma, alpha, so on). In the real world, a guy suffering from an epileptic seizure exhibits signals that are distinct from those seen in the average brain (see Figure 1.1). Through viable signal evaluation, it can be determined whether the signals are non-seizure or seizure impacted. EEG refers to electrophysiological examination of the brain's electrical activity. This electrical movement alluded to signs of brain function. Typically, the electrodes are put on the scalp. The voltage fluidity caused by ionic current inside the brain's neurons is estimated using EEG [9]. EEG is a term used in clinical contexts to describe the recording of the brain's unrestricted electrical flexibility for an indefinite period of time using electrodes placed on the scalp's surface. Either the event's potential outcomes or the intriguing content of EEG are usually the emphasis of specific segments. In this work, EEG motions in the frequency range are used to analyze the type of neural motion known as "brain waves." Any changes in those brain waves or signals indicate a problem. As a result, signal analysis is required to diagnose epilepsy since it identifies a person's brain signal that is comparable to the epilepsy.

Figure 1.1 shows the EEG signal of a healthy brain and a seizure affected brain. Many recordings of the EEG signals that occur in the brain of an epileptic patient
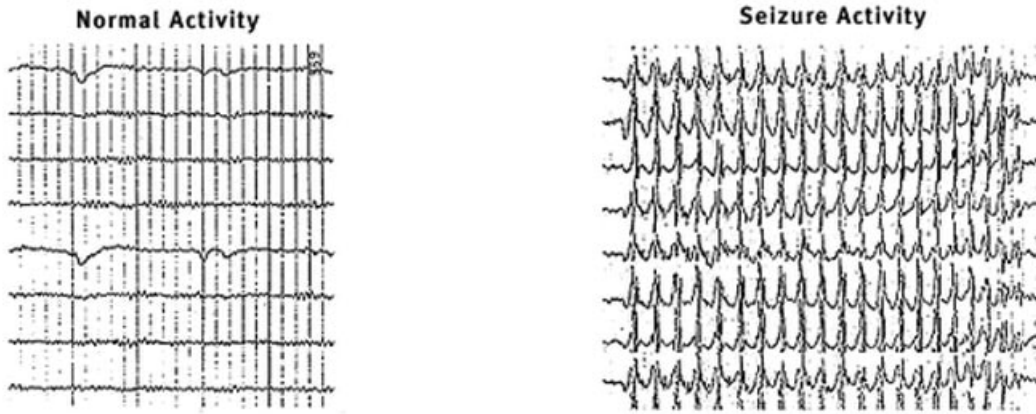
Figure 1.1: Normal activity vs seizure activity of the brain [20]

have been made over the years. A pattern emerges from all of the EEG recordings of these impulses. Anyone with this pattern of signal in their EEG is thought to be having an epileptic seizure. Some applications, followed by an EEG measurement, are required to discover anyone who exhibits the pattern. Feature extraction and classification are performed after the signal has been captured. The entire procedure is comparable to a BCI (brain computer interface) system. Although the Brain Computer Interface system has a specific goal in mind, and the field of BCI research and development has always centered on neural prosthetic applications that aim to restore damaged or handicapped movement [6], the system is similar to the identification of damaged brain activity movements. A BCI system is recognized for tracking certain patterns in the human brain and relaying this information to people while they take action. The task at hand is to translate these patterns into a haphazard order. Following that, various EEG signal processing methods are applied to the commands in order to characterize the patterns. Later on, the efficient methods for identifying diverse patterns of human brain signals and classifying them through the use of a classifier will be discussed. As a result, the focus of this thesis is mostly on the feature extraction and feature classification positions. In all of the applications, several researchers have already discovered a variety of ways. We'll concentrate on what produces the best results for us, or simply which ones we're best suited to deal with. This study employed data from a Bonn University dataset [22] that included seizure-affected signals as well as healthy signals with multiple

subsets. Many time-frequency decomposition approaches, such as Discrete Wavelet Transformation (DWT) [16], short time Fourier transformation (STFT) [35], empirical mode decomposition (EMD) [22], power spectral density (PSD) [35], Pseudo Wigner distribution (PWD) [11], and others, perform well for extracting features. Many traditional machine learning techniques are used for EEG signal classification, such as SVC, Logistic Regression, Random forest, MLP classifier, K-Means, ANN, KNN ,Naive Bayes, among other functions to smoothly identify data. We used 80 percent of the data set is for fitting models, and 20 percent is for model evaluation.We used SVC, Logistic Regression, Random forest, MLP classifier, K-Means, Decision tree, Gradient Boosting for classification. Moreover, classification performance of different machine learning algorithms is compared to find out the best suitable model for our dataset.

1.4 Thesis Orientation

The thesis is organized in the following manner: We gave an overview and introduction to epilectic seizures in Chapter 1. We've also talked about our research objectives and why we're doing it. In Chapter 2 is the literature review that focuses on relevant work and existing methods that are based on our topic. The background knowledge essential to our task is thoroughly examined in Chapter 3. In Chapter 4, we showed how to extract the dataset and its description, as well as how to preprocess it to suit the models. In Chapter 5, we demonstrated research methodology as well as the workflow of our proposed model. We also discussed the machine learning models that we employed in our thesis. We examined and then analyzed the data in Chapter 6. Finally, in Chapter 7, we explored the challenges we faced in compiling the models throughout the thesis, as well as our future intentions.

Chapter 2

Literature Review

Given the severity and perilous health risks of epilepsy, research on finding methods to curb it continues to be a hot topic in the biomedical field. However, there is still a long way to go in completely understanding and preventing/curing this medical phenomenon. Given the spontaneous and untriggered nature of epileptic attacks, any temporary symptoms, such as minute aberrations in movement, might go unnoticed, leading to a major epileptic seizure. Therefore, assessment of the seizures must be continuous, which poses multiple engineering challenges. In this section, we are going to review some similar research conducted in the field.

The capacity to employ EEG signal analysis to identify a variety of structural brain alterations associated with various illness states, as well as the effectiveness in computer-based detection of a variety of brain diseases, could be extremely useful for diagnosis and therapy [10].A human brain can be in a variety of states depending on the messages generated. The spectrum variations of these signals have been used to classify them into five distinct groups: delta, theta, alpha, beta, and gamma [34]. Changes in these signals can indicate a variety of illnesses, including epileptic seizures. A model signal of the brain with epilepsy is created from samples of signals acquired from patients experiencing epileptic seizures. Because epilepsy is a common neurological complication, programmatic classification of epileptic EEG events has become a focus of study in the past few years. Various research has been conducted on the signal analysis of epileptic seizures, in which time-frequency

distribution has proven to be quite effective in signal decomposition [35].

Researchers have utilized a variety of decomposition approaches, with some of them being particularly popular. For example, Krishnaveni et al. in [37], Abdulhamit et al. in [16], and Hasan et al. in [19] used the discrete wavelet transform (DWT) as the decomposition method in their research publications. They claimed that utilizing DWT as a pre-processing method improved performance over not using any at all. In this scenario, DWT is one of the most commonly utilized decomposition strategies. The signal's breakdown directs a collection of coefficients known as wavelet coefficients. A linear combination of wavelet capacity, weighted by those wavelet coefficients can be used to the signal in this case. The importance of registering the proper number of coefficients for obtaining a correct signal structure reproduction is critical. The wavelet's energy is confined in DWT to a certain time interval [16]. Wavelet Packet Decomposition (WPD) [31], power spectral density (PSD) [35], short time Fourier transform (STFT) [20], empirical mode decomposition (EMD) [32], and others are regularly used approaches. WPD differs from the previously described DWT approach in a few ways. The WPD creates twice as many distinct sets of coefficients for n levels of decomposition. The WPD generates twice as many recognizable coefficient arrangements for n dimensions of decomposition. Because it is confined to wavelet bases that expand by an intensity of two towards the low frequencies, the typical wavelet transform may not offer the greatest result in terms of compression [31]. PSD is a signal that exists throughout an unmatched time period, or a time period that is sufficiently expansive that it may have existed across an infinite time interval [35].

STFT is a Fourier transform that determines the sinusoidal frequency and stage material of neighboring portions of a signal as it varies over time [36]. EMD relies on the delivery of smooth envelopes characterized by local grouping maxima and minima, as well as the subtraction of the mean of these envelopes from the underlying succession. To supply the upper and lower envelopes, this necessitates the identification of all local extrema, which are then linked by cubic spline lines.

All of the methods discussed above can be combined into a single system. Working with several time-frequency distribution approaches, on the other hand, has yielded better experimental results. Alexandros et al. demonstrated feature extraction using the STFT and PSD time-frequency distributions [20]. Although the number changes based on the number of classifications, this study found that combining these methods improved their accuracy over utilizing them separately. In most situations, great outcomes have been achieved by applying a single method. It is up to the researchers to determine how they will use them in accordance with their techniques. However, combining various strategies can sometimes, if not always, produce better results. The signals must be categorized after they have been preprocessed and their features extracted.

One of the variety of approaches for classifying signals is the machine learning technique. There are several methodologies to approach machine learning. SVMs (support vector machines) and ANNs are two of the most widely used among them (artificial neural networks). Over the last two decades, ANNs have been routinely used to identify EKG and EEG signals. For epileptic seizure identification, a number of alternative ANN-based techniques have been reported in the literature [13]. Epilepsy and non-epileptic EEG are used to create models for neural network design, which can be used to categorize EEG signals. Preparatory data sets are used to gather features for model construction. The characteristics were selected for the objective of recording differences between epileptic and normal EEGs. Extracting features is critical to the success of ANN classification [19]. The SVM classifier is a general classifier that has shown outstanding results in identifying EEG signals based on data transfer capacity factors [32]. Linear SVM, in particular, is a method that is used in a variety of applications. It creates plots of several separated classes using a hyperplane to isolate them. Depending on the information, the hyperplane partition might be regarded as 2D or 3D. Apart from these two, Rajendra et al [31] use a Gaussian mixture model classifier, in which each class's probability density capacity is predicted as a combination of multidimensional Gaussian distributions.

A Gaussian mixture model anticipates and treats each information focus as if it were made up of a small proportion of Gaussian distributions with unknown variables. This classifier can be regarded as a clustering summarization. Indeed, our classifier outperforms regression trees and KNN (K-nearest neighbor) [31].

In [57], a number of classic machine learning methods are used to look at the EEG dataset including logistic regression as a base learner/classifier. The chance that an instance with attribute values $x_1, x_2, ...x_n$ is a logistic regression can be defined as follows:

$$p = 1 + \frac{1}{1 + e^{-(a_0 + a_1 x_1 + a_2 x_2 + ... + a_n x_n)}} \qquad (2.1)$$

Here, in equation (2.1) a_i i = 0, 1, …, n are all constants.

From the labeled training data, we can figure out what the constants are in the equation, and the values of constants are obtained using the maximum likelihood estimation technique. Logistic regression prediction is simple, and If the parameters are accurate, the prediction will be mechanical. Stacking is implemented using logistic regression as a meta classifier. By fitting stages one by one, AdaBoost creates an additive logistic regression model. Training data is randomly segmented in a random forest classifier, and the decision tree is built with each sample data. In random forest, only a restricted feature is evaluated when splitting a non-leaf node. It was revealed that the deviation of a decision tree classifier is greater than that of a random forest classifier and greater than that of an extra tree. When the random forest classifier was used, it constructed a series of decision trees from a randomized piece of the training dataset and merged the scores from different decision trees to get the resulting class of the test object.

1_A: X $\rightarrow$ {0, 1}. By multiplying constant Y_{jm} with b_{jm}, the loss function is minimized. XGBoost is a distributed gradient boosting toolkit that optimizes and outperforms GBM architecture.

Chapter 3

Background Analysis

Some prior knowledge is required before we dive into our work. As a result, we found resources to gather crucial knowledge about the human brain, the topic we're working on, and the methodologies that can be used before deciding on our own tactics. The following is a discussion of the most important aspects of our gathered knowledge.

3.1　Brain Anatomy

The brain is housed in the skull. It is in close proximity to all of the sense organs. All of the functions of the human body are governed by the human brain, which acts as a control center. The brain is also in charge of managing our knowledge of a situation, our speech, the operation of our limbs, and many of our body's organs [50]. The Brain Stem, Cerebrum,and Cerebellum are the three primary components of the brain. The descriptions of various other components are given below, in addition to these three.

The brain anatomy is depicted in Figure 3.1, with different sections of the brain playing distinct roles.
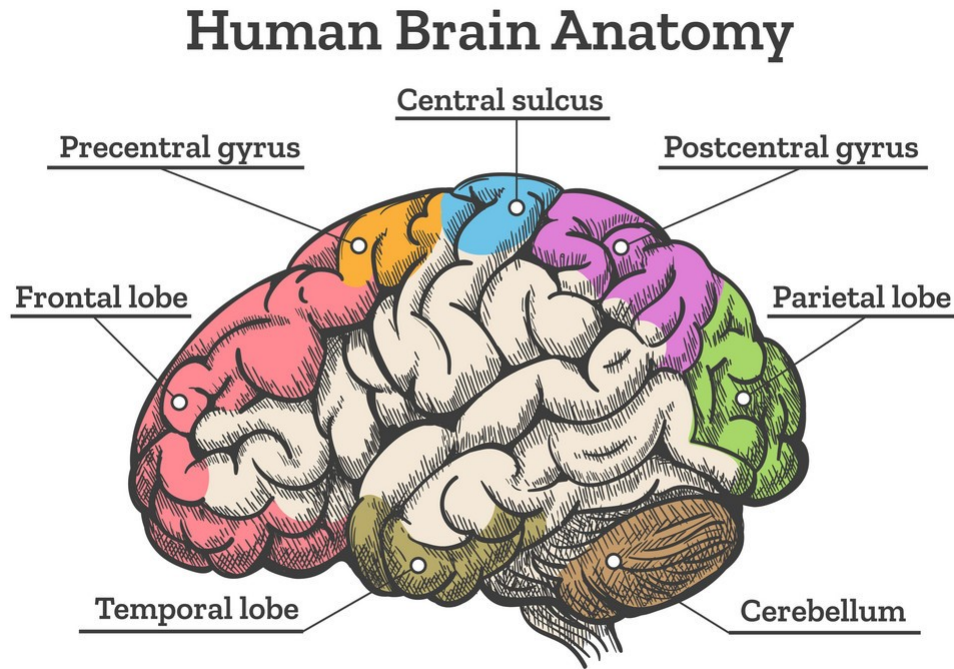
Figure 3.1: Anatomy of the Human Brain [23]

3.1.1   Cerebral Cortex

Humans are distinguished from other animals by their cerebral cortex. Each of the four lobes of the human brain corresponds to one of the four portions of the cerebral cortex [50].

3.1.2   Cerebellum

The cerebellum is a brain area that regulates fine motor actions. It is the sole portion of the nervous system that runs straight through the middle. The cerebellum is primarily in charge of movement execution, timing, and multi-limb coordinates. It is also known as the ” small brain” [38].

3.1.3   Brain Lobes

The cerebral cortex is divided into four lobes, each with its own purpose. The frontal lobe, the parietal lobe, the temporal lobe, and the occipital lobe are all components of the brain.

Frontal Lobe

The closest section of the brain to the front is the frontal lobe. It is capable of thinking, motor abilities, higher-order cognition, and expressive language, among other things. The frontal lobe gets signals from the other lobes and controls body motions using them [50].

Parietal Lobe

In the exact middle of the brain is where you'll find the parietal lobe, which is involved in the processing of tactile input such as tension, contact, and discomfort [50].

Temporal Lobe

It is a portion of the brain located at the brain's base. This lobe contains the auditory cortex. The forebrain cortex is responsible for sound and language interpretation [50].

Occipital Lobe

A portion of the brain is located in the back. This lobe has multiple functions associated with the comprehension of visual input and data. The primary visual cortex is housed within the occipital lobe [50].

3.1.4 Brain Stem

The brainstem is also involved in the regulation of cardiac and respiratory function. The brainstem is in charge of facilitating communication between the brain's motor and sensory systems and the rest of the body.

The midbrain, medulla, and pons are the three sections of the brain stem, as shown in Figure 3.2.

Midbrain
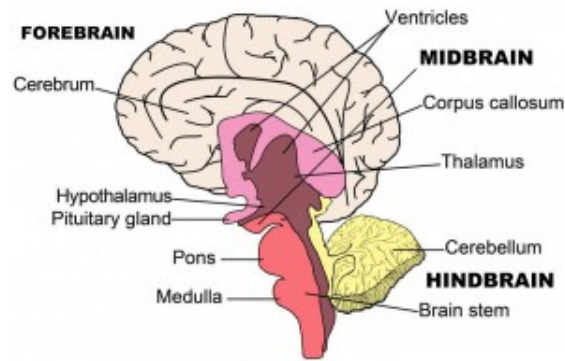
Figure 3.2: Brain Stem Structure [41]

The midbrain is a section of the brain that is relatively tiny. It functions as a transfer point for both audible and visual information [41].

Medulla

The medulla is found in the lowest part of the brain stem, in a section called the spinal cord. The medulla regulates numerous vital functions, including heartbeat, breathing, and blood pressure [41].

Pons

By way of the pons, the medulla is linked to the cerebellum. The pons is responsible for various autonomic activities, including activating breathing and coordinating sleep cycles, among other things [41].

3.1.5   Thalamus

The thalamus is a structure that sits atop the brain stem. The thalamus processes and transmits sensory information and bodily movements. As a transfer point, the thalamus receives and transmits sensory information to the cerebral cortex [50].

3.1.6   Hypothalamus

In the base of the brain, the hypothalamus is near the pituitary gland. It's a cluster of nuclei.It is in charge of the pituitary gland. As a result, the hypothalamus is involved in a variety of bodily activities [50].

3.1.7 Limbic System

The limbic system is divided into four sections.Among these are the amygdala, hippocampus, limbic cortex, and septal region. Through these four regions, the limbic system communicates with the hypothalamus, thalamus, and cerebral cortex.The limbic system is known as the "control center" for emotional responses and learning [41].

3.1.8 Basal Ganglia

A huge cluster of nuclei covers the thalamus. The basal ganglia are composed of several nuclei, which is essential for movement regulation. The red nucleus and ventral striatum are both components of the midbrain, and the basal ganglia is associated with both structures [41].

3.2 EEG Waves

An EEG is a record of the brain's surface electrical stimulation. This activity measurement appears as waveforms of variable frequency, shape, and amplitude on the EEG machine's display. The cerebral electrical activity is reflected in the recorded waveforms.

Figure 3.3 shows the main frequencies of the human EEG wave.

3.2.1 Delta

Delta brainwaves have a frequency dimension of 4 Hz or less and are quite noisy [9]. It is very common as an overwhelming emotion in newborns up to one year of age. It may occur centrally, or it may occur across the brain with diffuse damage, metabolic encephalopathy, hydrocephalus, or significant midline lesions [2].

Figure 3.3: Human brain waves [47]

3.2.2 Theta

The theta frequency range of 4 to 8 Hz corresponds to moderate activity [9]. It is entirely natural in children under the age of 13 and when they are sleeping, but it is considered strange in adults who are awake.. It's commonly thought to be an indication of central subcortical lesions. It's also where we keep our apprehensions, haunted histories, and horrible dreams.

3.2.3 Alpha

Alpha has a frequency that varies between 8 and 13 Hz [35],is most often found in the back parts of both sides of the skull, with the prevailing side having a higher adequacy. It usually appears when closing one's eyes and unwinding, then fades by opening one's eyes or being alarmed by any device [9].

3.2.4 Beta

Beta is a fast action wave with a frequency of 13 Hz and higher [9]. It is most visible frontally and can be seen on both sides in symmetrical appropriation. Beta brain waves, which may be absent or diminished in areas of cortical damage, regulate our regular waking state of consciousness whenever contemplation is coordinated towards subjective errands.

3.2.5 Gamma

In comparison to other types of neural signals, gamma waves are the most rapid and possess the highest frequency. They are also linked to the concurrent processing of information received from a number of different brain regions. Gamma brainwaves are the least noticeable of the brainwave frequencies, passing info fast and quietly [7].

3.3 Epileptic Seizure

An epileptic seizure is a momentary set of signs or symptoms caused by unusually fast or well-coordinated nerve cell activity in the brain.The external manifestations might range from uncontrollable jerking movements to a temporary loss of consciousness. There are many different types of epileptic seizures. A single epileptic seizure can happen to anyone at any time in their lives. This is not the same as epilepsy, which is a tendency to have seizures that originate in the brain [14].

Figure 3.4 shows Brain impulses during a seizure, partial seizure and generalized seizure EEG.

3.3.1 Generalized Epileptic Seizures

The peculiar electrical aggravation affects the entire brain in generalized seizures, and the person begins to lose consciousness. The interval when the individual is
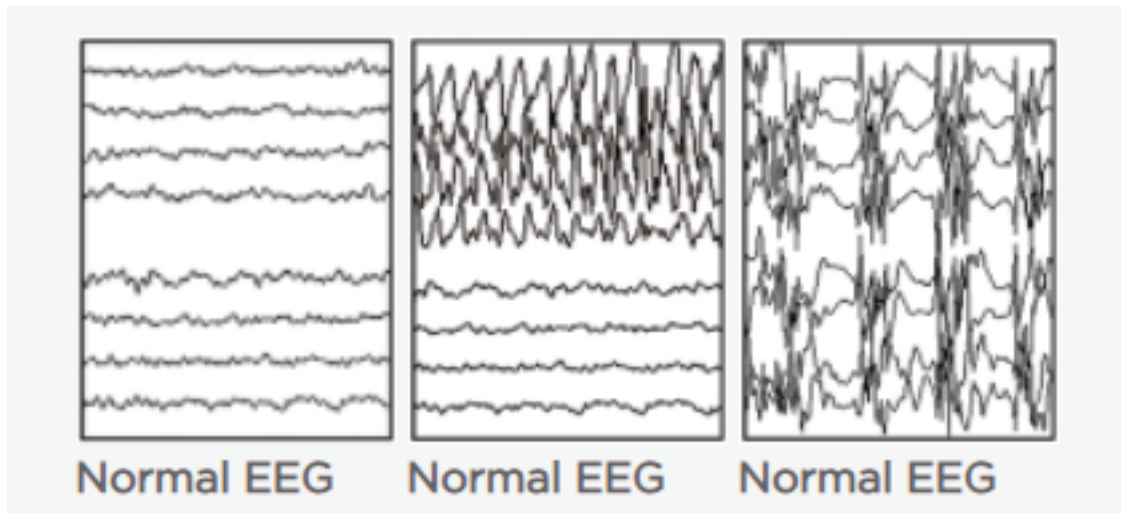
Figure 3.4: Brain impulses [48]

ignorant can be extremely brief and easily overlooked [39].

Seeing a child or anyone else suffering from a seizure can be frightening. Because the child is clueless to what is going on during generalized seizures, they are unaware of what is going on.

They may have unusual side effects before a generalized seizure, which will alert them to the start of a seizure. If this occurs, get them to a safe area, even if it's on the floor.This note or quality is most likely the commencement of a seizure in only one participant in the brain before it expands across the full brain.

Tonic Seizure

Tonic seizures are characterized by increased muscular tone or solidity. Tonic seizures can be a steady progression or a massive yank, depending on how quickly they start and how long they last. Due to muscle stiffening, a child may be pushed forward or backward, and may even fall. Due to the fact that their chest muscles have also solidified, they may become blue and appear to stop breathing. Tonic seizures last 10 to 15 seconds on average, although they can last up to a minute. They frequently occur while a child is sleeping or shortly after waking up. For a little period after the seizure, the tyke is misidentified.

Arms, legs, and trunk muscles became stiff [65]. These usually last less than 20

seconds and occur while people are sleeping. In any event, if you're standing at the time, you risk losing your balance and falling. These are becoming more common in people who have a kind of epilepsy known as Lennox-Gastaut condition, but they can also occur in people with other types of epilepsy.

Clonic Seizure

Clonus refers to the repetitive quick stiffening and relaxation of a muscle [14]. To put it another way, it's jerking. Restraining or moving the arms or legs will not stop the movements.

A person may lose control of physiological functions and begin twitching in various portions of the body during a clonic seizure. He or she may have a brief period of unconsciousness, followed by confusion [5].

Clonic seizures usually begin in childhood. They frequently resemble a case of large bilateral epileptic myoclonus, however the motor aspects are often asymmetrical. Clonic seizures are uncommon and mostly affect newborns. Clonic motions are most commonly noticed as part of a tonic-clonic seizure [65].

Tonic-Clonic Seizures

A grand mal seizure is an unsettling influence on the asymmetrical functioning of the brain. This is due to the incorrect transmission of electrical signals through the brain. This will eventually result in signals being transmitted to muscles, nerves, and glands [34]. The proliferation of these signals in the brain can cause you to lose consciousness and experience extreme muscle constriction.

Tonic-clonic seizures are named for the two distinct stages they go through. People lose awareness, and they may fall down during the tonic phase of the seizure. Quick muscle constrictions, often known as convulsions, characterize the clonic stage [47].

Absence Seizures

Petit mal seizures are also known as absence seizures. They are short, usually lasting less than 15 seconds, and they have unfathomable side consequences [47].

Regardless, even for such a short period of time, loss of awareness can make these dangerous.

These are more common in children than in adults, and they might occur often. When a person who has had an absence seizure becomes oblivious for a short period of time. Their eyelids may squint or vacillate, or they may stop what they're doing, glance up, and gaze. They are unresponsive to what is going on around them. If they are strolling, they may continue to stroll, but they will be unaware of what they are doing [3].

Atonic Seizures

Atonic seizures are a type of seizure in which muscular quality is suddenly lost. Akinetic seizures, drop ambushes, and drop seizures are other names for these seizures. When in doubt, the individual remains conscious and does not collapse to the ground [1].

Myoclonic Seizures

Myoclonic seizures are short, stun-like muscle branches or groups of muscles. "Myo" means muscle, and "clonus" means a muscle that is rapidly compressing and unwinding, snapping or jerking. They usually don't last more than a second or two [47]. Myoclonic seizures typically involve both sides of the body simultaneously, and the individual may lose their balance. They can be found in a variety of epileptic syndromes, each with its own set of symptoms:

- Juvenile myoclonic epilepsy: The neck, shoulders, and upper arms are the most commonly affected areas. Seizures frequently occur shortly after a patient is awakened in many individuals. They usually begin during adolescence or, on rare occasions, early adulthood in people with a typical knowledge base.

- Lennox-Gastaut syndrome: A once-in-a-lifetime condition characterized by a wide range of seizures. It begins in childhood. The neck, shoulders, upper arms, and frequently the face are all affected by myoclonic seizures. They

may be quite substantial and difficult to handle.

- Progressive myoclonic epilepsy: This group includes unusual syndromes that combine myoclonic and tonic-clonic seizures. Treatment usually does not last long since the patient eventually breaks down.

Atypical-Absence Seizure

In contrast to conventional absence seizures, atypical absence seizures are one-of-a-kind, startling, or unusual. They're a type of generalized first seizure, meaning they start on both sides of the brain. Atypical absences are epileptic seizures that typically affect children with severe learning and neuralgic deficits caused by epileptic encephalopathies, the most common of which is Lennox-Gastaut syndrome. They differ from conventional absences in that they have a gentle beginning and conclusion, have low cognitive impedance, and are typically linked with considerable tone changes [18] [8]. The onset of an atypical absence seizure is less abrupt, and it is counterbalanced by loss of awareness, as opposed to conventional absence seizures. With the patient performing an action gradually or with slip-ups, the loss of awareness may be insignificant [47].

Infantile Spasms

Early-life spasms and seizures characterize infantile spasms, a devastating epileptic encephalopathy. It is also called the West Syndrome, and it can result in severe epilepsy that can't be treated [47].

Infantile spasms typically affect children under the age of two [33]. When compared to ordinary offspring of comparable ages, children with infantile spasms and hypsarrhythmia EEGs displayed considerable differences from the standard in soundness and dreadful power. During rest, delta, theta, alpha, and beta coherence increased, notably at long distances between cathode separations, whereas cognizance decreased in the theta and beta ranges, particularly in the frontal region. A few children can achieve seizure control and a typical early recognition with proper therapy

and a careful demonstration assessment [15].

3.4   Reasons of Epilepsy

Children account for around 30% of all epilepsy case instances. Epilepsy has no recognized cause in approximately half of individuals who suffer from it. A number of factors can be blamed for the disease in the other half, including:

1. Genetic influence: It's possible that these occurrences have a hereditary component. In these circumstances, there is most certainly a hereditary component.

2. Prenatal Injury: It has been shown that some kids are immune to brain injury before birth, which can be connected to a variety of factors, including oxygen shortage or a mother's sickness. The infant may develop epilepsy as a result of these circumstances.

3. Brain condition: Epilepsy can also be caused by brain tumors or other types of brain injury or head trauma.

Besides these factors, there are some specific factors which may increase the risk of epilepsy, such as brain injury, dementia, lack of sleep, anxiety, stress, et cetera [40].

Treatment of epilepsy

Treatment can help most people with epilepsy minimize or eliminate their episodes. Some of the therapies are as follows:

Anti-epileptic drugs are administered during surgery to remove a small portion of the patient's brain, which is thought to be the source of the seizures.They are called anti-epileptic drugs (AEDs), a method of seizure control that involves implanting a small electrical device within the body. A certain sort of food (the ketogenic diet) can help avoid seizures [40].

Chapter 4

Dataset Analysis

In this section we talked about the dataset we used for our thesis purpose ,from where we collected it ,and then how we processed it into comprehensible format to feed it into different machine learning classifiers for finding the best model.

4.1   Data Collection

Our work employed a preprocessed version of a commonly used epileptic seizure detection dataset from the UCI Machine Learning Repository [46].

4.2   Data Description

The dataset consists of 500 patients and contains 4097 EEG readings for each subject over the course of 23.5 seconds. After that, the 4097 data points were split up into 23 unique chunks for each patient, and each of those pieces was turned into a single row in the dataset. Each row has 178 readings, which are then transformed into columns; to put it another way, one second's worth of EEG readings is 178 columns. There are a total of 11,500 rows and 179 columns, with the very last column indicating the patient's status, which indicates whether or not the patient is experiencing a seizure at the moment. The response variable, y, can be found in column 179, and the independent variables are X1,.., and X178.

y includes a response variable, which can be written as y in 1, 2, 3, 4, or 5:

5	When the EEG was being recorded, eyes were open
4	When the EEG signal was being recorded, eyes were closed
3	Found the position of the tumor within the brain, as well as the recorded EEG activity from the healthy part of the brain
2	The EEG is recorded at the tumor's location
1	Recordings of seizure activity

Table 4.1: Description of Response Variable (Y)

Table 4.1 illustrates a brief description of the response variable in our dataset, where there are a total of 5 classes.

	Unnamed	X1	X2	X3	X4	X5	X6	X7	X8	X9	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	y
0	X21.V1.791	135	190	229	223	192	125	55	-9	-33	...	-17	-15	-31	-77	-103	-127	-116	-83	-51	4
1	X15.V1.924	386	382	356	331	320	315	307	272	244	...	164	150	146	152	157	156	154	143	129	1
2	X8.V1.1	-32	-39	-47	-37	-32	-36	-57	-73	-85	...	57	64	48	19	-12	-30	-35	-35	-36	5
3	X16.V1.60	-105	-101	-96	-92	-89	-95	-102	-100	-87	...	-82	-81	-80	-77	-85	-77	-72	-69	-65	5
4	X20.V1.54	-9	-65	-98	-102	-78	-48	-16	0	-21	...	4	2	-12	-32	-41	-65	-83	-89	-73	5

5 rows × 180 columns

Figure 4.1: Epileptic Seizure Dataset

It is shown in Figure 4.1 that all subjects from classes 2, 3, 4, and 5 did not experience an epileptic seizure. Epileptic seizures are only a concern for class 1 individuals.

4.3   Dataset Preprocessing

Data preprocessing is a process that puts unprocessed data into a format that is easy to understand so that our dataset can be better understood. Initially, we eliminated the first column from the dataset because it was irrelevant to our machine learning algorithm. Since our main goal is to determine whether or not a patient is having a seizure using 178 EEG signals per second, we converted the multiclass dataset to a binary classification problem by replacing the y column values greater than 1 with 0 and leaving the 1 values unchanged. In this case, 1 denotes a patient who is experiencing a seizure and 0 indicates that the patient is not having a seizure.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	...	X170	X171	X172	X173	X174	X175	X176	X177	X178	y
0	135	190	229	223	192	125	55	-9	-33	-38	...	-17	-15	-31	-77	-103	-127	-116	-83	-51	0
1	386	382	356	331	320	315	307	272	244	232	...	164	150	146	152	157	156	154	143	129	1
2	-32	-39	-47	-37	-32	-36	-57	-73	-85	-94	...	57	64	48	19	-12	-30	-35	-35	-36	0
3	-105	-101	-96	-92	-89	-95	-102	-100	-87	-79	...	-82	-81	-80	-77	-85	-77	-72	-69	-65	0
4	-9	-65	-98	-102	-78	-48	-16	0	-21	-59	...	4	2	-12	-32	-41	-65	-83	-89	-73	0

5 rows × 179 columns

Figure 4.2: Epileptic Seizure Dataset after Manipulation

In Figure 4.2, the first 178 columns are called independent parameters. The last column is the response parameters, and it says y = 0 or 1, where y = 0 means no seizure activity was found and 1 means seizure activity was found.

Class	Number of cases
0	9200
1	2300

Table 4.2: Total values present in dataset after binary classification

Table 4.2 represents the number of cases for the classes being used, with a total of 2300 cases for seizure class (1) and 9200 cases for non- seizure class (0).

Then, we verify whether any null values exist in our dataset.

X1	0
X2	0
X3	0
X4	0
X5	0
...	...
X175	0
X176	0
X177	0
X178	0
y	0

Length: 179, dtype: int64

Table 4.3: Result after checking null values in dataset

Table 4.3 shows that our dataset has no missing values and that we are dealing with the complete dataset.

To find out whether or not feature engineering is required for our dataset, for which we determine the log column. Here, the number of columns that are sufficiently skewed to require transformation is "0", indicating that all columns are sufficiently skewed and no additional feature engineering is required for our dataset. Therefore, we may avoid transforming feature columns.

4.3.1   Splitting Training and Test set

The Stratified Shuffle Split algorithm is utilized in order to split our dataset into training and testing sets. We chose Stratified Shuffle Split as our binary classifier as it consists of 0 and 1, with a total of 9200 0s and 2300 1s. However, in this instance, implementing a random split could result in a bias toward 9200. To overcome this issue, we used Stratified Shuffle Split, which takes 80 percent of data from both class 0 and class 1, then combines them to create a combined 80 percent trained dataset that is optimal and unbiased because it is a blend of both classes.

4.3.2   Dimension Reduction

We performed dimension reduction to determine if the performance of our best classifier improved after dimension reduction. As our dataset is so huge, dimension reduction may be necessary. It is really difficult to manipulate this massive dataset. We used principal component analysis (PCA) to cut down on the number of dimensions as much as possible, hence removing ineffective columns.

PCA

PCA reduces the variance of large data sets by turning a huge proportion of variables into a small number of variables, which contains the dataset's most important information. The goal of analysis is to reduce the data set's dimensional complexity. The data's dimensionality is reduced by creating basis vectors. A linear combination of basis vectors can be used to reproduce any sample from a data collection using basis vectors. In essence, PCA strives to lessen the amount of parameters in our collection of data while retaining quite enough information as feasible [43].

Step 1: Standardization

This step normalizes the range of continuous initial input variables so that they all have a substantial impact on the analysis [42]. If the starting variable values vary markedly, the factors with a wide variety will outnumber those with lower ranges (for instance, a parameter ranging from 0 to 100 will outnumber a parameter ranging from 0 to 1), leading to an unfair outcome [61]. Following standardization, all variables will be scaled to a certain level.

Step 2: Covariance Matrix Computation.

$$Z = \begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix} \qquad (4.1)$$

The variances of each initial variable are found on the major diagonal of Equation 4.1, which is a covariance matrix. The principal diagonal in Z runs from top left to bottom right because the covariance of a parameter with itself equals its deviation (Cov (a, a) = Var (a)). The values of the covariance matrix are symmetrical with regard to the major diagonal because the covariance is (Cov (a, b) = Cov (b, a). This covariance matrix summarizes the connections between two or more variables.

Step 3: Using the eigenvectors and eigenvalues of the covariance matrix to find principal components

Now, PCA tries to fit quite enough information as it can into the first component, followed by the second component. Because there are as many variables in the data as there are principal components, the first principal component explains the most variation in the data. T The second principal component is found in the very same manner as the first, but it must not be related to the first and account for following the largest variance. The above technique is repeated until the number of variables equals the total number of primary components [202]. The eigenvectors of the covariance matrix are the directions of the axis that have the most variance. These are the principal components.

Step 4:Creation Of Feature Vector

By figuring out the eigenvectors and putting them in order by decreasing eigenvalue, we can find the order of importance of the main parts. Then, we decide if we want to keep all of these parts or get rid of the ones with poor eigenvalues. With the ones we keep, we make a matrix of vectors termed "feature vectors" [61].

Step 5: Rearranging data along the axes of the main components

The data is rearranged from the primary axis to the axes indicated by the principal components through the feature vector. In the last step, the transpose of the initial data set is multiplied by the transpose of the feature vector to get this result.

$$FinalDataSet = FeatureVector^T * StandardizedOriginalDataset^T \qquad (4.2)$$

We rearranged the data along with the primary component axes using equation 4.2, then we used it to reduce dimension.

To reduce the dimensions of our dataset, first we figured out if we needed to make any changes to our dataset.
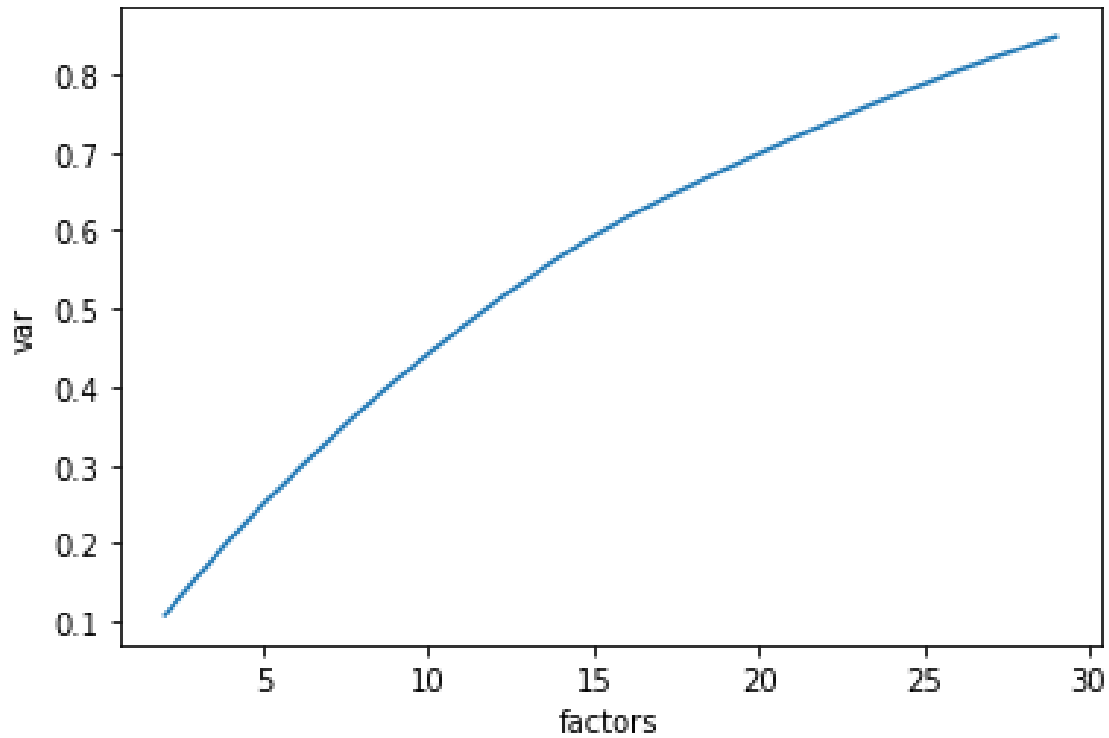

Figure 4.3: Variance vs. Number of Components in PCA

In Figure 4.3, there is no elbow present in the PCA graph, and it is gradually increasing. It indicates that we don't have to make any changes to our dataset. If any elbow exists, we have to make changes in our dataset before proceeding.
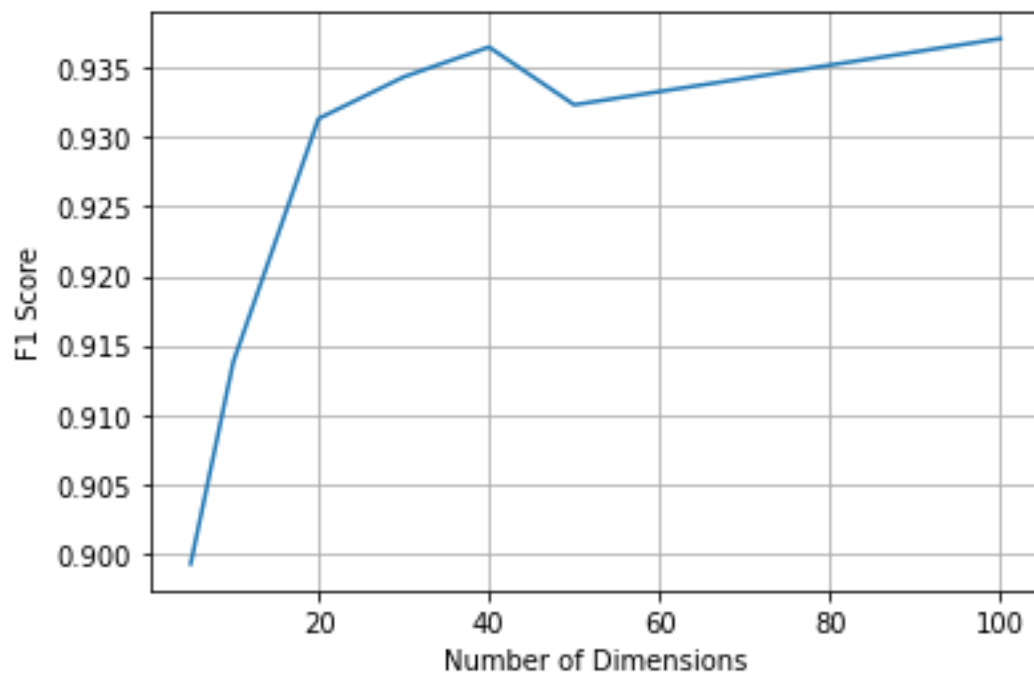
Figure 4.4: F1-score vs Number of Dimensions

With only 40 components, our graph 4.4 achieved a good score. This means that instead of 179 features, we can use just 40 components. So, we tried to figure out if we could achieve high accuracy with 40 components instead of using 178 features.

Chapter 5

Research Methodology

In this section, we will describe our model workflow, the machine learning classifiers we applied, and the testing, validation, and evaluation methods used in our study on epileptic seizure detection and prediction thoroughly.
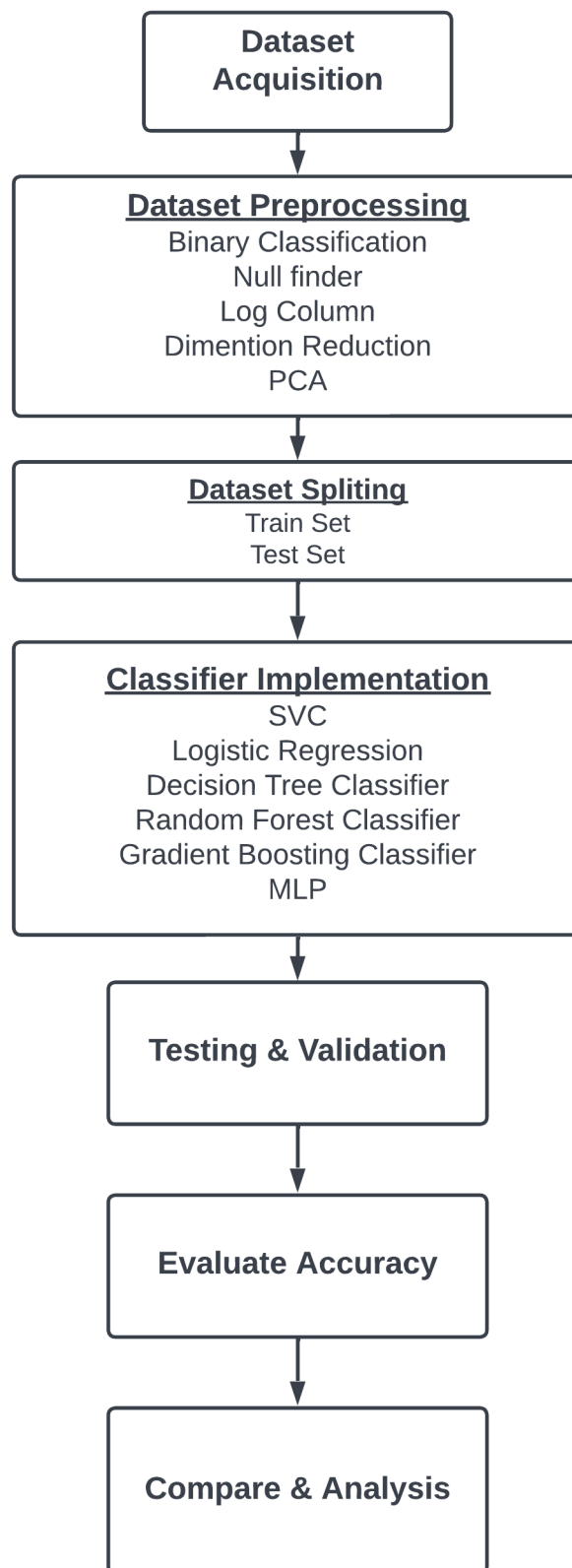
5.1 Model Workflow

Figure 5.1: Complete Flowchart of Model Workflow

In Figure 5.1, the flowchart for our thesis work is shown, which includes data pre-processing, data splitting, classifier implementation, testing & validation, evaluation of performance, and analysis.

5.2   Classifiers Used

SVC

The Support Vector Classifier, or SVC, is the most appropriate machine learning method for our case. SVC does not allow multiclass categorization by default. It enables binary classification and data item separation into two categories. A SVC's goal is to fit the data given to it, providing a "best fit" hyperplane that divides or categorizes it. After obtaining the hyperplane, we may input some characteristics to the classifier in order to get the "predicted" class. This makes this algorithm particularly ideal for our needs, although it may be used in a variety of circumstances.

It's a one-vs-one C-support vector classification that deals with multiclass support [25].

In our dataset, we applied SVC as it let us choose from a variety of non-linear classifiers. kernel rbf, c=0.025.

In SVC, the Radial Basis Function is a regularly used kernel.

$$K(x, x) = e^{\frac{-||x - x'||^2}{2\sigma^2}} \tag{5.1}$$

In 5.1 $||x - x||^2$ is the squared Euclidean distance. There are two parameters (C and γ) in an SVC classifier using the RBF kernel. C is applicable to all SVM kernels, and removes the training set's categorization for simplicity on the surface of a decision. A low C flattens the decision surface, whereas a high C detects all training data properly. The gamma value represents the potency of a particular training example. The closer the other samples are to the gamma, the more likely they are to be

affected [26].

There is the training vector, $x_i \, \epsilon \, \mathbb{R}^p$, i =1,.., n, in two classes, and another vector $y \, \epsilon \, (1, -1)^n$

Our objective is to figure out $w \, \epsilon \, \mathbb{R}^p$ and $b \, \epsilon \, \mathbb{R}$ so that the detection is correct for the most cases. The primal problem solved by SVC is given in equation 5.1.

$$\min_{w,b,\zeta} \tfrac{1}{2} w^T w + C \sum_{i=1}^n \zeta_i$$
$$\text{subject to } y_i \left(w^T \phi \left(x_i \right) + b \right) \geq 1 - \zeta_i \qquad (5.2)$$
$$\zeta_i \geq 0, i = 1, \ldots, n$$

We're attempting to maximize the margin in 5.2 (by reducing) $||w||^2 = w^T w$, while incurring a penalty when a sample is misclassified or falls under the margin.

$$\min_\alpha \tfrac{1}{2} \alpha^T Q \alpha - e^T \alpha$$
$$\text{subject to } y^T \alpha = 0 \qquad (5.3)$$
$$0 \leq \alpha_i \leq C, \; i = 1, \ldots, n$$

In 5.3, e is the vector of all ones, and Q is an n by n positive semidefinite matrix.

$Q_{ij} = y_i y_j K(x_i, x_j)$, where $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ is the kernel. The terms in 5.3, α i are called the dual coefficients, and they are upper-bounded by C.

As demonstrated in Figure 5.2, the RBF kernel can handle scenarios in which the relationship between class labels and characteristics is nonlinear because it non-linearly inserts samples into a higher-dimensional domain. In addition, the linear kernel with a penalty parameter C" functions identically to the RBF kernel with a few specialized settings (C, γ). Moreover, for some special parameters, the sigmoid function performs like RBF [12].

The second argument is that the number of hyperparameters impacts the model selection complexity. There are fewer parameters in the RBF kernel than in the polynomial kernel. For this reason, the RBF kernel is not numerically complex.

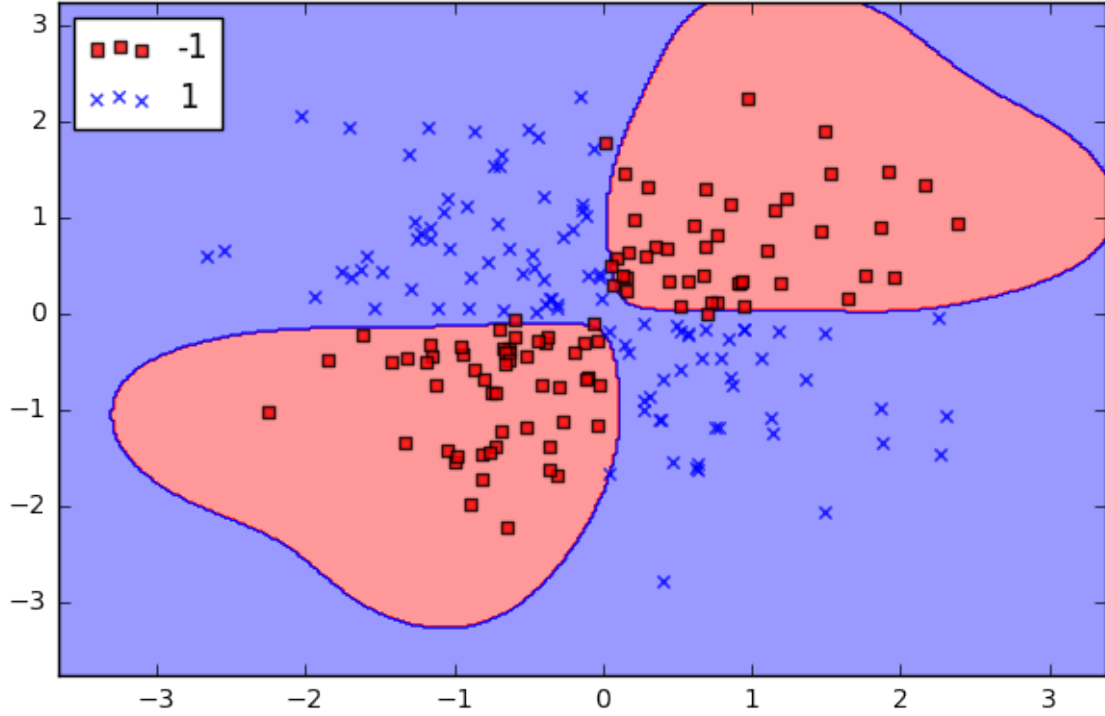And accordingly, it is efficient in terms of computational complexity [4].

Figure 5.2: SVC Parameters When Using RBF Kernel [44]

Logistic Regression

The categorical dependent variable is predicted using logistic regression, applying a set of independent factors. However, we used this algorithm to solve classification problems in our model. The Logistic Regression problem's outcome can only be between 0 and 1. When the probabilities between two classes must be calculated, logistic regression can be utilized. For example, if a seizure exists or not, 0 or 1, true or false, and so on. In logistic regression (LR), the probability, P1, of a binary result occurrence is connected to a set of independent variables in the form

$$
\begin{aligned}
\text{logit}\,(P_1) = \ln\left(\frac{P_1}{1 - P_1}\right) &= \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n \\
&= \beta_0 + \sum_{i=1}^{n} \beta_i x_i
\end{aligned}
\tag{5.4}
$$

in equation 5.4 where 0 is the intercept and 1, 2, ..., n are the coefficients for the explanatory variable x1, x2,..., xn. A binary variable has only two possible values,

such as yes/no, on/off, survive/die, or 1/0, and is used to describe the presence or absence of a particular event (e.g., epileptic seizure/not).

Instead of fitting a regression line, we fit a "S" shaped logistic function (shown in Figure 5.3), that predicts two maximum values (0 or 1), but instead of giving precise values, it produces probabilistic values that are between 0 and 1. When a decision criterion is included, logistic regression transforms into a classification procedure [59].
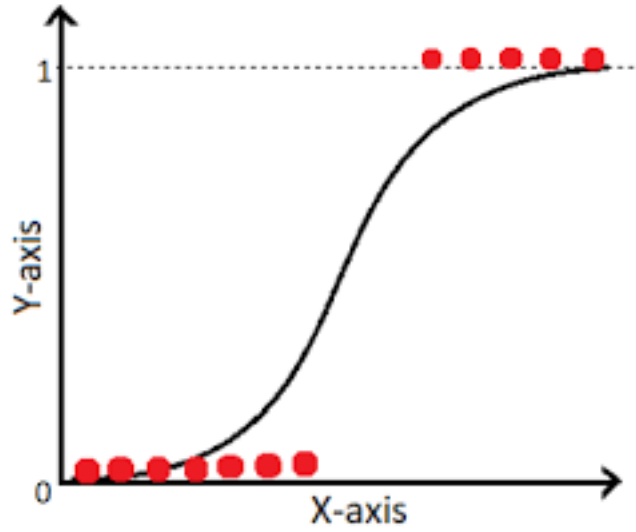

Figure 5.3: Logistic Regression sigmoid curve [53]

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

$$\mathcal{L}(\hat{y}, y) = -(y \log \hat{y} + (1 - y) \log(1 - \hat{y}))$$

$$z = x_1 w_1 + x_2 w_2 + b$$

$$\hat{y} = \sigma(z)$$
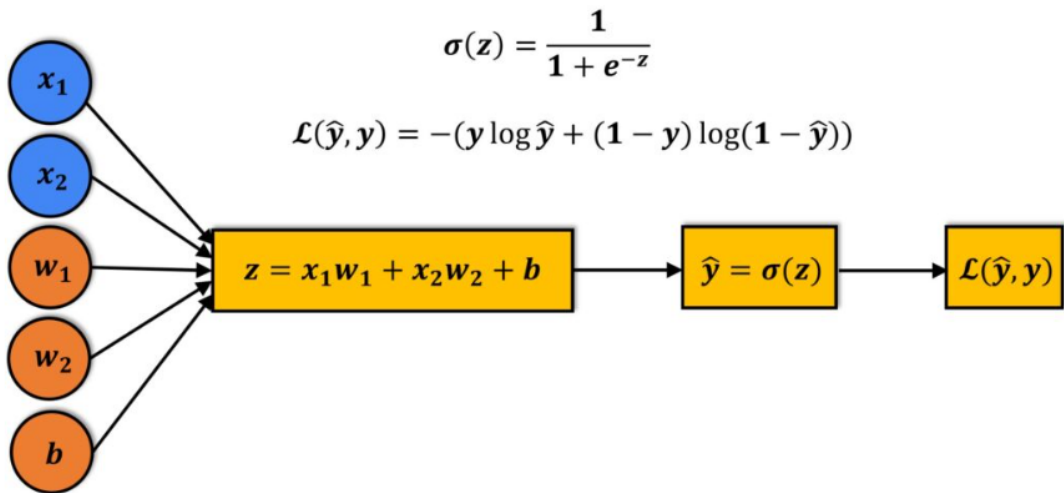
$$\mathcal{L}(\hat{y}, y)$$

Figure 5.4: Working procedure of the Logistic Regression Model [59]

The Logistic Regression Model is illustrated in Figure 5.4. To predict the outcome,

a sigmoidal activation function is given a linear equation. To evaluate the model's performance, we calculate the loss. The most common loss function is mean squared error. The mean squared error is not the best option because the output of logistic regression is a probability value between 0 and 1.

Decision Tree

Decision Tree is the most effective even among highly prominent classifiers, which have outstanding performance in several areas. This is as a result of its capacity to construct a number of rational explanatory logic rules with excellent prediction accuracy.

In this tree-structured classifier, input data comprises dataset characteristics and variables, branches indicate decision rules, and each leaf node delivers the output. A decision tree has two nodes: the decision node and the leaf node.Unlike Leaf nodes, which are the results of decisions and have no further branches, Decision nodes can be used to make any kind of decision. The dataset is stored in the root node. While implementing a decision tree, we must pick the best characteristic for the root node and subnodes. Attribute Selection Measure (ASM) is a method that has been devised. We can simply determine the best characteristic for the tree's nodes using this measurement. The two common ASM techniques are information gain and gini index. We divide the node and build the decision tree based on the value of information gain. The Attribute Selection Measure (ASM) finds the best attribute in the dataset. By dividing the root into subsets, it constructs the node of the decision tree with the best attribute. Using subsets of the dataset, it creates further decision trees in a recursive manner. The process is repeated until no more nodes can be classified, at which point the final node is identified as a leaf node, resulting in the final output. A pure sub-split indicates that in our decision tree, the output is either "yes" or "no". The Entropy formula is shown in equation 5.5

$$\mathbf{E}(\mathbf{S}) = -\mathbf{p}_{(+)} \log \mathbf{p}_{(+)} - \mathbf{p}_{(-)} \log \mathbf{p}_{(-)} \tag{5.5}$$

Here in 5.5, p+ is the probability of positive class,

p+ is the probability of positive class,

p− is the probability of negative class,

S is the subset of the training example

Our goal here is to minimize the uncertainty in our dataset. Here, "Information gain" is a new measure for this purpose, which reveals to us how much the parent entropy has decreased after being separated by a feature.

$$\text{Information Gain} = E(Y) - E(Y|X) \tag{5.6}$$

Here in 5.6, E(Y) represents entropy for Y, and E(Y | X) represents conditional entropy for X given Y. It reduces disorder in our target variables by deciding how data is divided by a decision tree.

Decision trees are useful for finding patterns in large amounts of data that are difficult to detect manually. The patterns found can be used to make decisions by offering insight into outcome(s) of future circumstances.

Random Forest

Random forest generates decision trees from numerous samples by using the majority vote for classification and the average for regression. This approach is used in regression and classification to handle data sets containing continuous and categorical variables. However, it produces higher classification accuracy. In our paper, we used it for classification to get our final output, which is based on majority voting [62].

To understand how a random forest algorithm works, we must first understand the decision trees, which consist of three types of nodes: root node, decision node, and leaf node. From the root node, the set begins to divide. Decision nodes are the nodes that result from dividing a root node, while leaf nodes are nodes that cannot

be divided further [55].

We created a decision tree, which uses a tree-like flowchart to represent the predictions that result from a series of feature-based splits. A Random Forest is a bagging method that makes predictions using a subset of the original dataset, which helps to overcome overfitting. Random forest creates many decision trees, each with a different collection of observations, rather than a single decision tree [55] [27].

The formula for calculating the feature in Scikit-learn is in equation 5.7:

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)} \tag{5.7}$$

Where,

ni_j = the importance of node j

w_j = weighted number of samples reaching node j

C_j = the impurity value of node j

$w_{left(j)}$ = child node from left split on node j

$C_{right(j)}$ = child node from right split on node j

$$fi_i = \frac{\sum_{j: \text{ node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{ all nodes}} ni_k} \tag{5.8}$$

Where,

fi_i = the importance of feature i

ni_j = the importance of node j

With the help of equation 5.8, the value of each feature is computed.

$$\text{norm } fi_i = \frac{fi_i}{\sum_{j \in all} \text{ features } fi_j} \tag{5.9}$$

They may be normalized to a number between 0 and 1 by dividing by the total of all feature importance values, as shown in equation 5.9.

The average of all the trees is the most essential attribute during the Random Forest stage. The total number of trees is divided by the sum of the importance ratings of each attribute on each tree.

$$RFfi_i = \frac{\sum_{j\in \text{ all trees}} \text{norm } i_{ij}}{T} \tag{5.10}$$

Where,

$RFfi_i$ = the importance of feature i calculated from all trees in the Random Forest model

$normfi_{ij}$ = the normalized feature importance for i in tree j

T = total number of trees

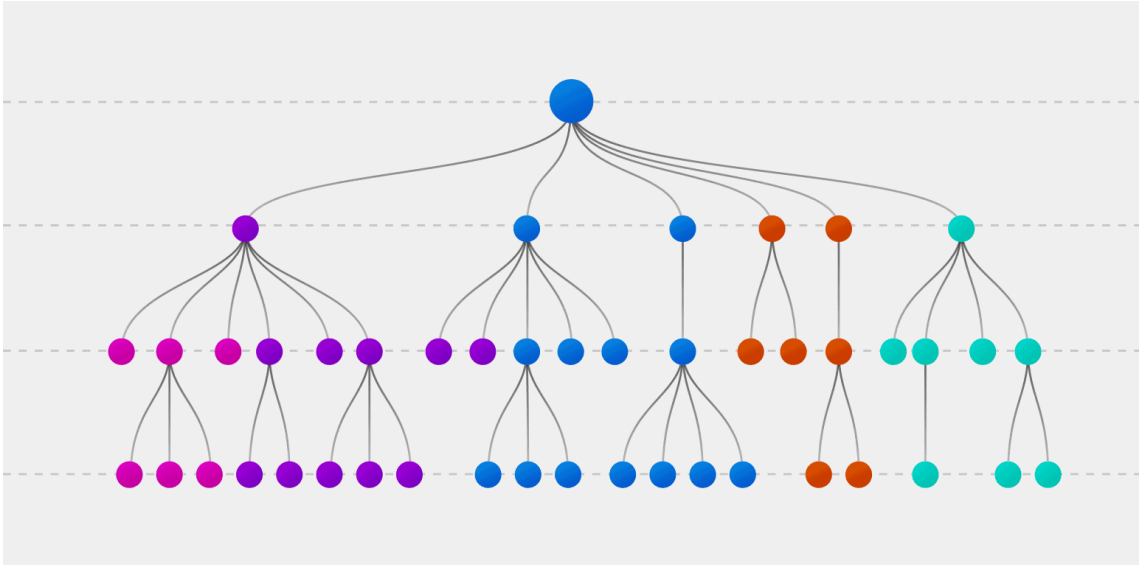The output of equation 5.10 [55] is the final feature.

Figure 5.5: Decision tree [63]

Figure 5.5 shows how final output is obtained through individual decision trees in random forest classification.

It is an ensemble approach. Bootstrapping is the process of repeatedly sampling different rows and characteristics from training data to build each decision tree model, as represented in Figure 5.5. Here, aggregation is the process of aggregating

all of the results from each decision tree to obtain a final result using majority votes.
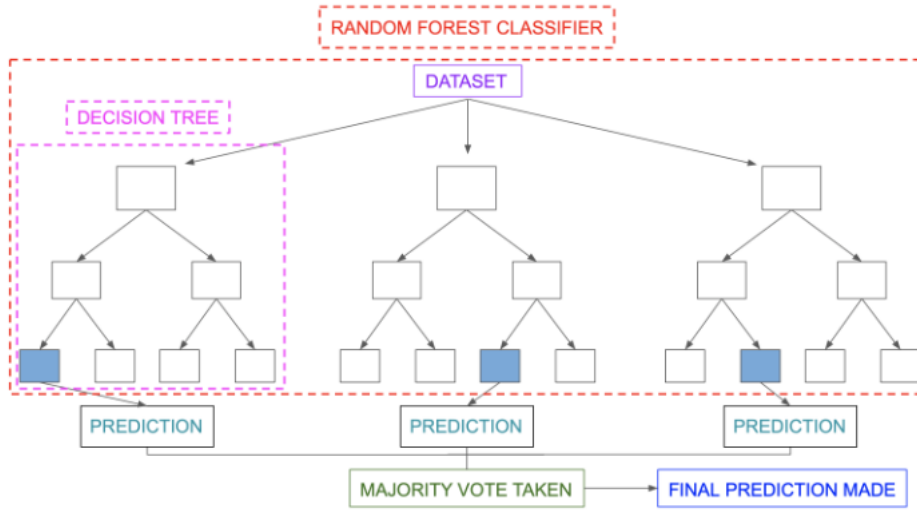

Figure 5.6: Classification in Random Forest [63]

Figure 5.6 is the map of random forest classification with row and feature sampling and parallel connections between all decision trees.

Gradient Boosting

One of the boosting algorithms used to lower a model's bias error is gradient boosting. We chose this algorithm as a classifier because our dataset comprises categorical variables and the cost function is log loss. The Gradient Boosting Classifier implements a loss function. This classifier can use a user-defined loss function, as well as a variety of standardized loss functions. The loss function, however, must be differentiable. In classification algorithms, logarithmic loss is commonly utilized. In addition, we used logarithmic loss in our analysis [54].

The steps outlined below are the steps to generating a gradient boosting classifier.

First of all, the model has to be fitted. Afterwards, the parameters and hyperparameters of the model should be tuned. We make some predictions and examine the outcomes and interpret the results. It needs some active decision-making to turn the hyperparameters of our model. We tried a number of inputs and hyperparameters to improve the accuracy of our model. For better results, we used three

41

modifications to our gradient boosting algorithm —tree constraints, weighted updates, and random sampling. These include tree depth and tree numbers employed in the ensemble. Weighted updates are used to restrict how much each tree contributes to the ensemble, such as a learning rate. Fitting trees on random subsets of features and samples is an example of random sampling. Gradient descent is used to improve the training loss of gradient boosting models. The gradient of the training loss is to change the target variables for each successive tree.

The loss function for the classification problem is given below in equation 5.11:

$$L = \frac{1}{n} \sum_{i=0}^{n} (y_i - \gamma_i)^2 \tag{5.11}$$

Equation 5.11 is our loss function, where y_i is the observed value and Γ is the predicted value.

Now we must determine the gamma value that will result in the smallest loss function [207]. We set the derivative of it to 0 in relation to log(odds). As a function, our initial prediction will be the minimal value of the loss function. The next step in the Gradient boosting regressor is to generate the pseudo residuals by multiplying the loss function's derivative by -1. We proceed as before, but this time the loss function is different, and we deal with the probability of a result.Then we have our first decision tree and find the final outputs of the leaves.

Learning rate is one of the most important parameters for gradient boosting decision trees. The learning rate refers to how quickly the model picks up new information. The overall model is altered by each new tree. The learning rate measures the strength of the change.Primarily, we are concerned about the classifier's accuracy on the validation set, although it shows that a learning rate of 0.05 provides the best validation set performance as well as decent training set performance. We can now analyze the classifier by estimating its accuracy and generating a confusion matrix . By generating a new classifier and setting in the best learning rate found, an output

of the tuned classifier is found.

Multilayer Perceptron (MLP)

MLP learns functions by using a dataset with a predetermined number of input and output dimensions. Based on the model's collection of features, it can construct a nonlinear function predictor for classification.
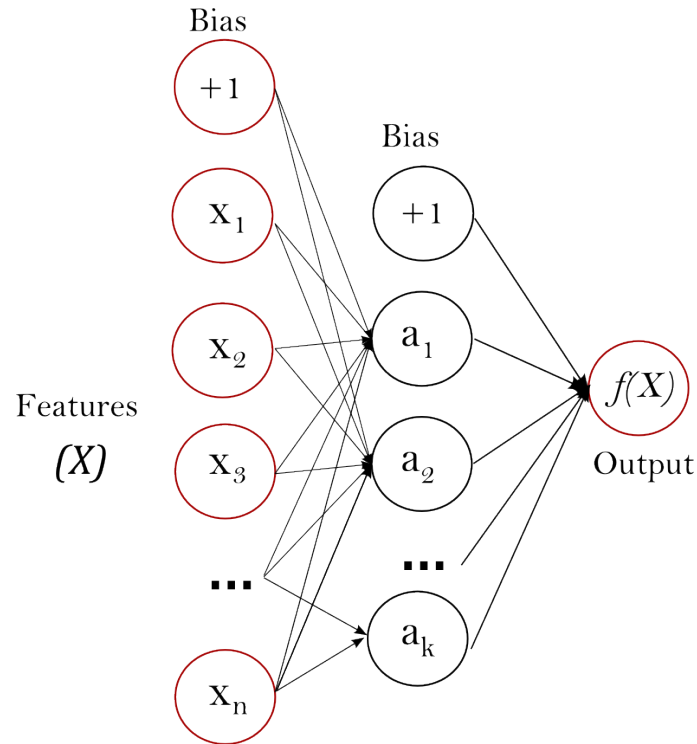

Figure 5.7: One hidden layer MLP [28]

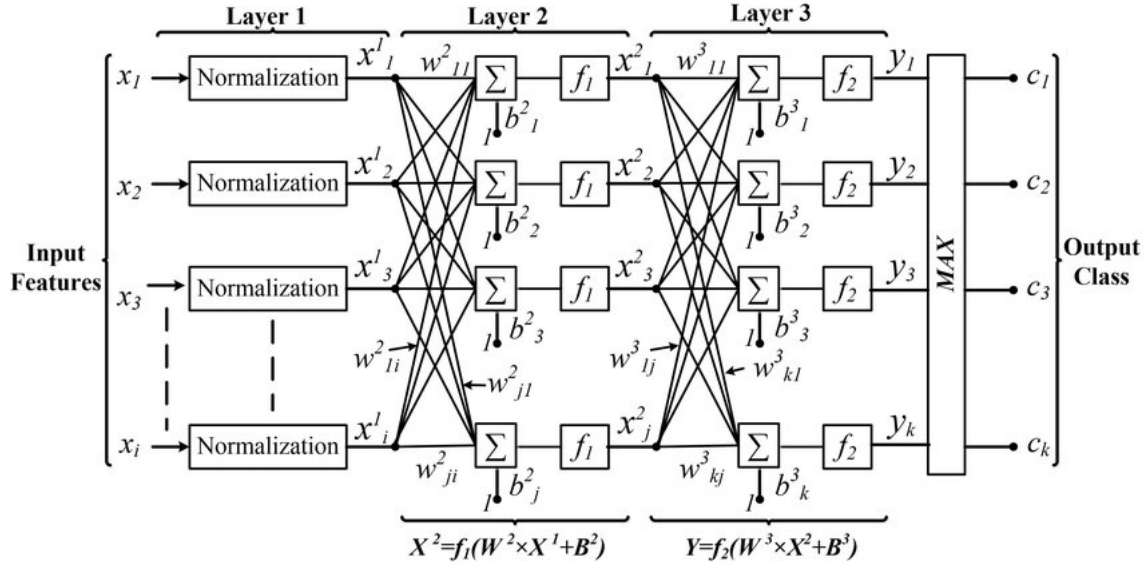Figure 5.7 depicts a scalar output MLP with only one hidden layer [56].

Figure 5.8: Mathematical model of MLP classifier with i-j-k topology [51]

There are three levels in the hypothesized MLP design: input, output, and a hidden layer. Layer 1 is the input layer, as described in Figure 5.8 and it conducts normalization on all I number of features. MLPs typically include one or more hidden layers, but in this study, we just employed one hidden layer to minimize classification time and hardware resource consumption [51].

MLP trains with two arrays. Training samples are stored as floating point feature vectors in Array X, while performance measures (class labels) for the training data are stored in Array Y (n samples, n classes) [56].

The controller parameters are described by a collection of neurons in the input layer. In the hidden layer, each neuron performs a weighted linear sum on the data from the previous layer, followed by a non-linear activation function such as the hyperbolic tan function. The output layer receives the values from the final concealed layer. Finally, it turns them into output values in the output layer [49][60].

In terms of concept; if we put the MLP calculations into a numerical equation, we simply need to choose a notation to refer to different regions of the network. We'll refer to the activations of the input units as xj and the activations of the output unit as y, as in the linear case. The units in the lth hidden layer will be represented

by h_i^l . Because our network is fully linked, each unit receives interactions from the preceding layer's units [52].

This means that each unit has its own bias, and every pair of units in two subsequent layers has a weight. As a result, the calculations of the network can be represented as 5.12:

$$
\begin{aligned}
h_i^{(1)} &= \phi^{(1)} \left(\sum_j w_{ij}^{(1)} x_j + b_i^{(1)} \right) \\
h_i^{(2)} &= \phi^{(2)} \left(\sum_j w_{ij}^{(2)} h_j^{(1)} + b_i^{(2)} \right) \\
y_i &= \phi^{(3)} \left(\sum_j w_{ij}^{(3)} h_j^{(2)} + b_i^{(3)} \right)
\end{aligned}
\tag{5.12}
$$

Distinct layers may have different activation functions in 5.12. We separate Φ^1 and Φ^2. We put the computations in vectorized form because all of these summations and indices might be difficult. Because each layer has numerous units, we use an activation vector h^l to represent activation of all of them. We use a weight matrix W^1 to express each layer's weights since every pair of units in two subsequent levels has a weight [52]. A bias vector b^1 is also present in each layer. As a result, the computations above are represented in vectorized form as in 5.13.

$$
\begin{aligned}
h_i^{(1)} &= \varphi^{(1)} \left(W^{(1)} x + b^{(1)} \right) \\
h^{(2)} &= \varphi^{(2)} \left(W^{(2)} h^{(1)} + b^{(2)} \right) \\
Y &= \varphi^{(3)} \left(W^{(3)} h^{(2)} + b^{(3)} \right)
\end{aligned}
\tag{5.13}
$$

When an activation function is applied to a vector, we imply that it is implemented to each entry independently. All of the hidden units for each layer for all of the training set is maintained as a matrix H^l.

The following are the computations:

$$H^{(1)} = \varphi^{(1)} \left(XW^{(1)T} + 1 \, b^{(1)T} \right)$$

$$H^{(2)} = \varphi^{(2)} \left(H^{(1)}W^{(2)T} + 1 \, b^{(2)T} \right) \tag{5.14}$$

$$Y = \varphi^{(3)} \left(H^{(2)}W^{(3)T} + 1 \, b^{(3)T} \right)$$

We directly translate the equation 5.14 into NumPy code and the code computes the prediction on the dataset. Since we used Sklearn, the hidden layer size in our code is 100 by default. And the activation function used in our thesis is ReLu that outputs the input directly if it is positive and zero otherwise. And the default solver is 'adam' which works well on our dataset [29] [52].

K-Means

K-Means Clustering is used in machine learning and data science to divide unlabeled datasets into groups, with each dataset belonging to only one group with similar qualities. The number of groups that must be made during the process is set by K; for example, if K=2, two pairs will be made.

The algorithm starts with an unlabeled dataset, divides it into k groups, and keeps doing this until no better groups can be found, resulting in each cluster comprising data points that share some characteristics but are isolated from the others. In this algorithm, k should always have the same value [64].
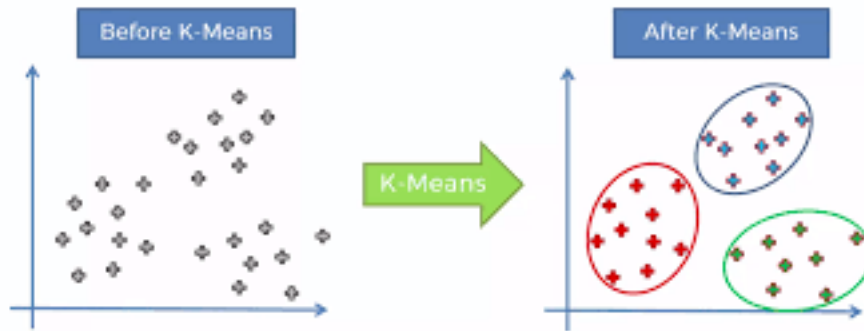
Clustering is shown in Figure 5.9.

Figure 5.9: Clustering of K-means [64]

Variables', or points', center of gravity is computed in order to find new centroids from clusters. This process is repeated by selecting a new centroid as the closest cluster needs to be found. After that, each datapoint is reassigned to the new centroid. This procedure is utilized to identify a median line. Once the model is complete, the assumed centroids can be removed, leaving two final clusters.

Choosing the value of "K number of clusters" in K-means Clustering

The Elbow technique was applied in our thesis work to determine the k value. The WCSS value concept is utilized in this strategy. The following formula can be used to determine the WCSS value for three clusters:

$$\text{WCSS} = \sum_{\text{P}_{i} \text{ in Cluster 1}} \text{distance } (\text{P}_{i}\text{C}_1)^2 + \sum_{\text{P}_i \text{ in Cluster 2}} \text{distance } (\text{P}_{i}\text{C}_2)^2 + \sum_{\text{P}_i \text{ in cLuster3}} \text{distance } (\text{P}_{i}\text{C}_3)^2$$

(5.15)

Each data point and its centroid in a cluster are separated by a distance of P_i in cluster distance $(PiC1)^2$ in equation 5.15, and the same holds for the other two variables.

The elbow approach uses the following procedures to find the optimal cluster value:

- On a dataset, it performs k-means clustering with different K values (ranges from 1-10)

- WCSS is calculated for each K value

- A curve is drawn between the estimated WCSS values and the number of clusters K

- A steep bend or a plot point that resembles an arm is thought to be the best K value

We find the elbow graph for our dataset to figure out the k value(s) of k mean clustering.If there is no visible elbow present we will use the number of y categories
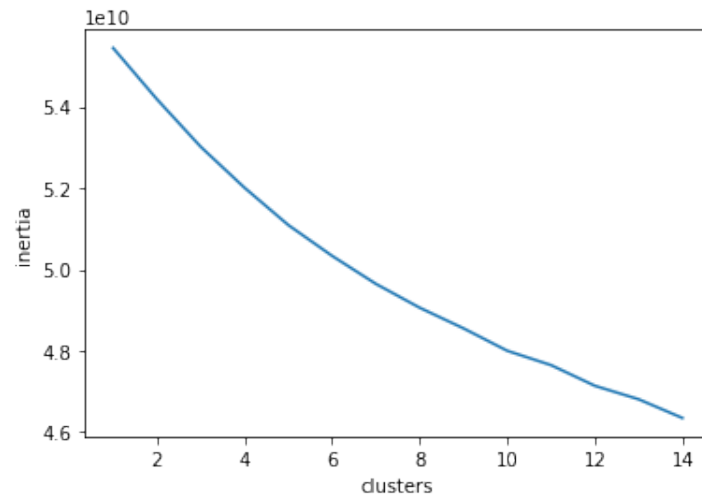
for k value.

Figure 5.10: The Elbow Curve for our used dataset

As in Figure 5.10, there's no obvious elbow here. We do, however, have domain knowledge: these features correspond to 5 y-categories, so there should be 5 clusters. So, we used K=5.

5.3   Evaluation Method

5.3.1   Performance metrics

We used confusion matrices, which have four values: true positive, true negative, false negative, and false positive, to analyze the performance of a binary classification task like seizure prediction.

Figure 5.11: Basic 2x2 Confusion Matrix [58]

As illustrated in Figure 5.11, the four possible classification outcomes are True-Positive (TP), True-Negative (TN), False-Positive (FP), and False-Negative (FN).

True Positives (TP): These are the examples in which we predicted yes ,they have the condition and our prediction was accurate.

True Negatives (TN): As predicted, they do not have the disease.

False Positives (FP): We thought they had the condition, but they do not.

False Negatives (FN): We predicted they would not have the disease, but they actually have.

Sensitivity (or recall) represents the proportion of real positive cases that our model correctly predicted. Here is how we can determine recall (Equation 5.16):

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{5.16}$$

Precision indicates the proportion of instances in which the predictions were accurate and the outcome was positive. Here's how to figure out Precision (Equation 5.17):

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{5.17}$$

F1 score indicates whether our model is accurate or not. It achieves its optimum value when precision equals recall. Here's how to figure out the F1 score (5.18):

$$\text{F1} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5.18}$$

5.3.2 Graphical Analysis

We used the Receiver Operating Characteristics (ROC) curve to graphically represent our findings. A ROC curve is a graph that demonstrates the performance of a classification model over all categorized levels. This curve depicts two parameters: the rate of true positives and the rate of false positives.

True Positive Rate, is the same thing as "recall" (Equation 5.19):

$$\text{TPR} = \frac{\text{TP}}{\text{TP+FN}} \tag{5.19}$$

Here's how to describe the False Positive Rate (Equation 5.20):

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \tag{5.20}$$

Chapter 6

Experimental Results

We evaluated model accuracy, F1-score, recall, and precision, as well as the confusion matrix, while analyzing the results. This section displays the results of classifying the epileptic seizure dataset using multiple classifiers.

6.1 Result Analysis

6.1.1 Confusion Matrix Analysis

It is used to evaluate a classifier's performance in detail. Multiple categorical outputs are produced by classification models. The total error in our model is calculated by most error measures, but we can't detect specific instances of errors in our model. Confusion matrices come in very handy here. It creates a table with all of a classifier's predicted and actual values. The performance of our classification models is compared in detail using confusion matrices for two binary classes, 0 and 1, where 0 means no seizure and 1 means seizure.
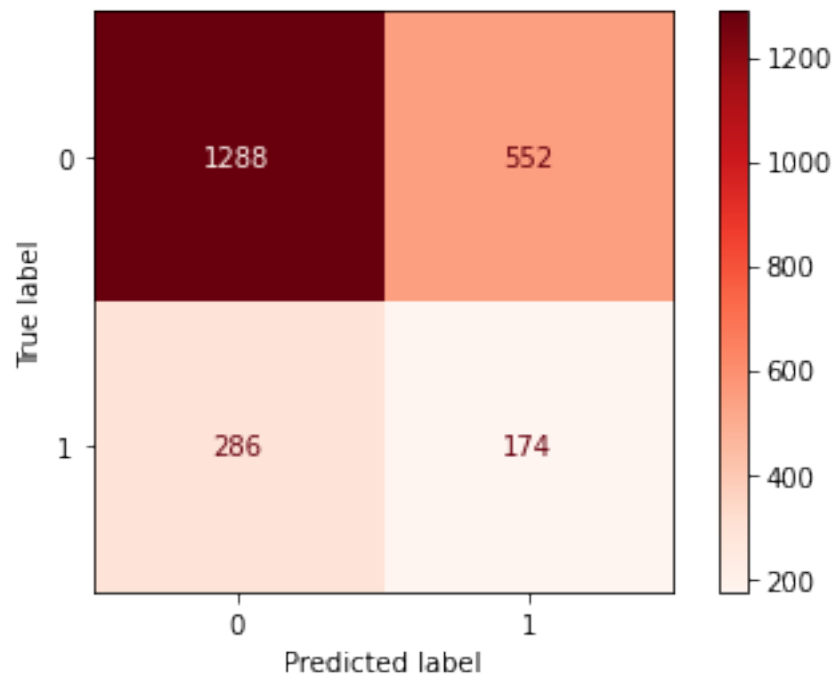
Logistic Regression

Figure 6.1: Confusion Matrix for Logistic Regression

Figure 6.1 depicts the confusion matrix of Logistic Regression. Consequently, the True Positive (TP) value is 174, indicating that seizure classes correctly predicted as seizure are 174, whereas the True Negative (TN) value is 1288, implying that healthy people correctly predicted as healthy are 1288, with some minor errors totaling 838, including sick people predicted as healthy and healthy people predicted as sick.
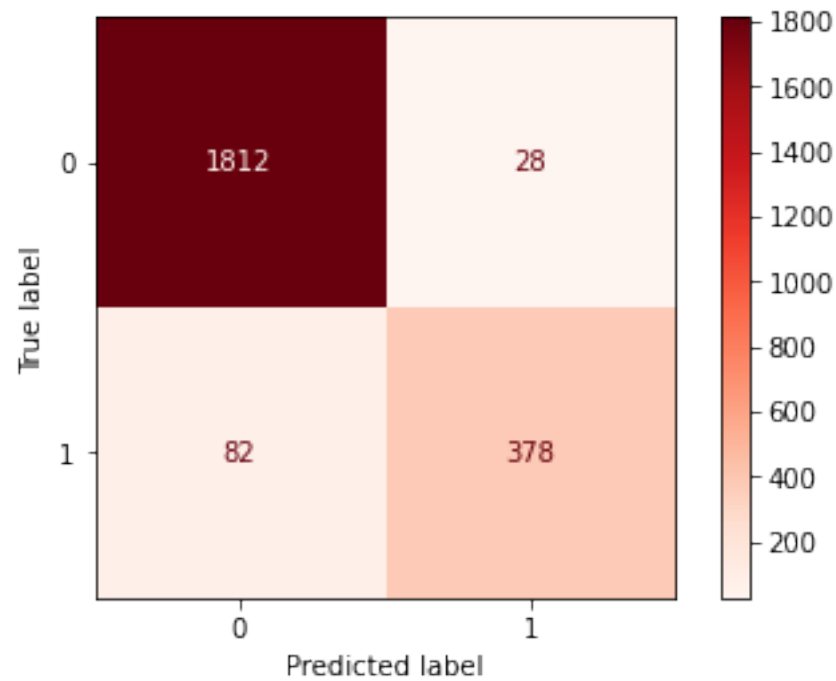
SVC

Figure 6.2: Confusion Matrix for SVC

The SVC confusion matrix is shown in Figure 6.2. Here, the True Positive (TP) value is 378, indicating that seizure classes were correctly predicted as seizure, while the True Negative (TN) value is 1812, indicating that healthy people were correctly predicted as healthy, with minor errors totaling 110, including sick people predicted as healthy and healthy people predicted as sick.
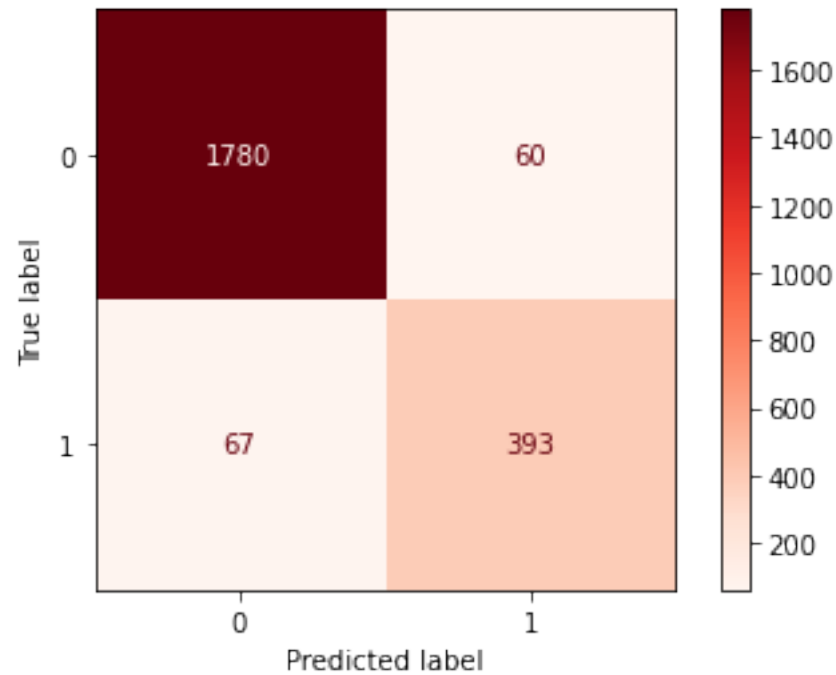
Decision Tree Classifier

Figure 6.3: Confusion Matrix for Decision Tree Classifier

Figure 6.3 illustrates the Decision Tree Classifier's confusion matrix. The True Positive (TP) value is 393, indicating that seizure classes were correctly predicted as seizure, and the True Negative (TN) value is 1780, indicating that healthy people were correctly predicted as healthy, with 127 minor errors, including sick people predicted as healthy and healthy people predicted as sick.
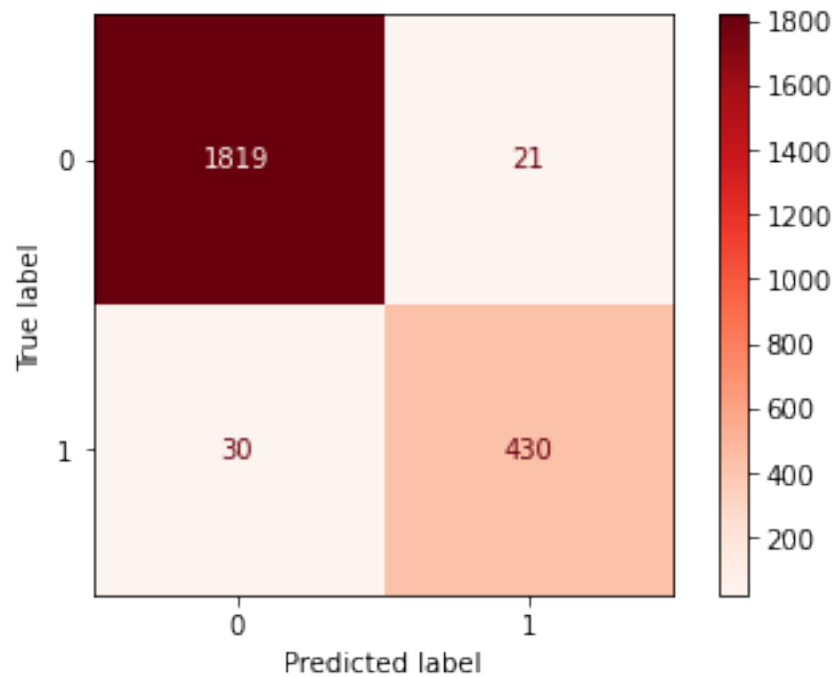
Random Forest Classifier

Figure 6.4: Confusion matrix for Random Forest Classifier

The confusion matrix of the Random Forest Classifier is shown in Figure 6.4. With 51 minor errors, including sick people predicted as healthy and healthy people predicted as sick, its True Positive (TP) value is 430, indicating seizure classes correctly predicted as seizure, and the True Negative (TN) value is 1819, indicating non seizure classes correctly predicted as healthy.
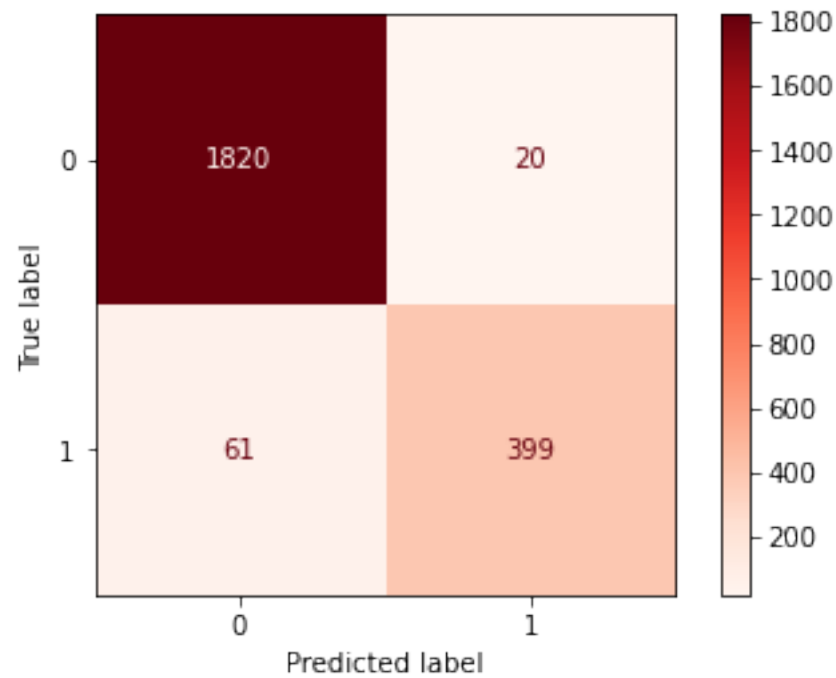
Gradient Boosting Classifier

Figure 6.5: Confusion matrix for Gradient Boosting Classifier

Figure 6.5 demonstrates the Gradient Boosting Classifier's confusion matrix. With 81 minor errors, including sick people predicted as healthy and healthy people predicted as sick, its True Positive (TP) score is 399, indicating seizure classes accurately predicted as seizure, and the True Negative (TN) value is 1820, indicating non seizure classes correctly predicted as non seizure.
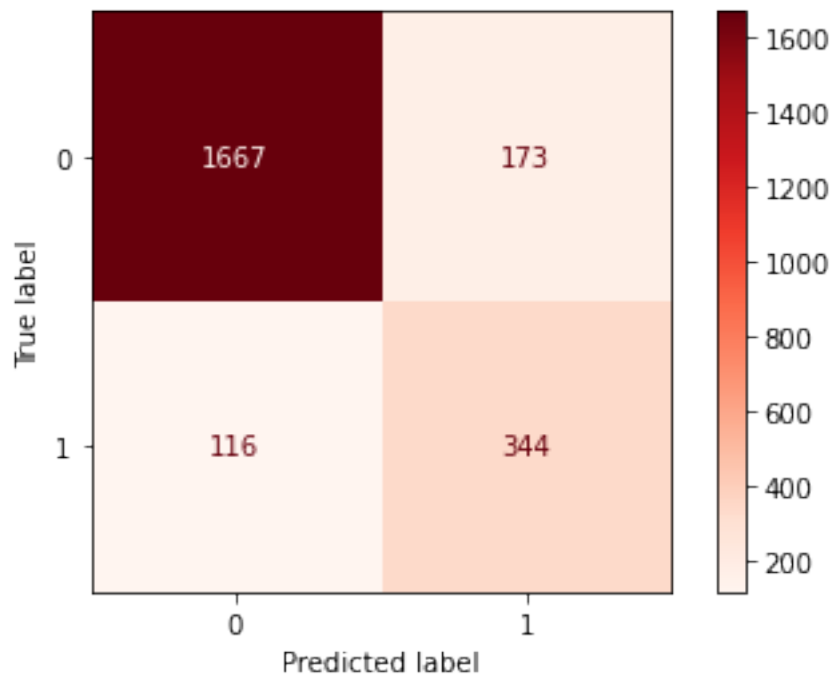
Multi Layer Perceptron (MLP)

Figure 6.6: Confusion Matrix for MLP Classifier

The confusion matrix of the MLP classifier is shown in Figure 6.6. Its True Positive (TP) score is 344, indicating seizure classes accurately predicted as seizure, and the True Negative (TN) value is 1667, indicating non seizure classes correctly predicted as non-seizure, with 289 errors, including sick people was predicted as fine and fine people was predicted as sick. . Following an analysis of the confusion matrix, it is evident that the Random Forest Classifier fits best our dataset as it provides us with the least amount of error (51) compared to other used classifiers while predicting seizures.

6.1.2 Graphical Analysis

We showed the Receiver Operating Characteristic curve, or ROC curve, to graphically represent our binary classifier system's diagnostic ability. The AUC, which is a measure of a classifier's ability to differentiate between classes and is used to summarize the ROC curve, is also included. The model's AUC is calculated and presented in the lower right corner or upper left corner of the ROC plot. AUC might

be anywhere from 0 to 1. An AUC of 0.0 indicates a model that makes 100% wrong predictions, whereas an AUC of 1.0 indicates a model that consistently gets the predictions correct. A model with an AUC of 0.5 is equivalent to one that randomly assigns classes.
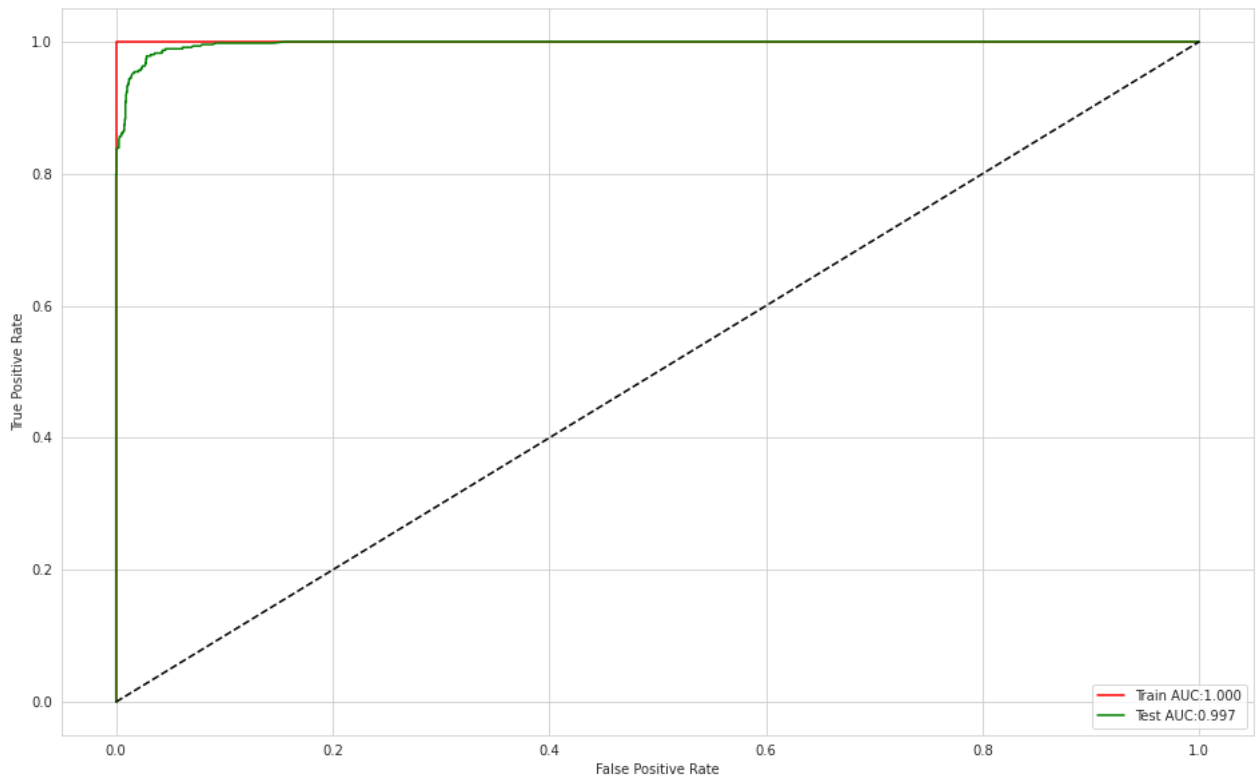
Random Forest Classifier

Figure 6.7: ROC curve of Random Forest Classifier

AUC of 1.00 for training data and 0.997 for testing are shown in Figure **??** for the Random Forest Classifier. Clearly, our data can be identified by the model, as shown by the AUC for both datasets being near to 1.
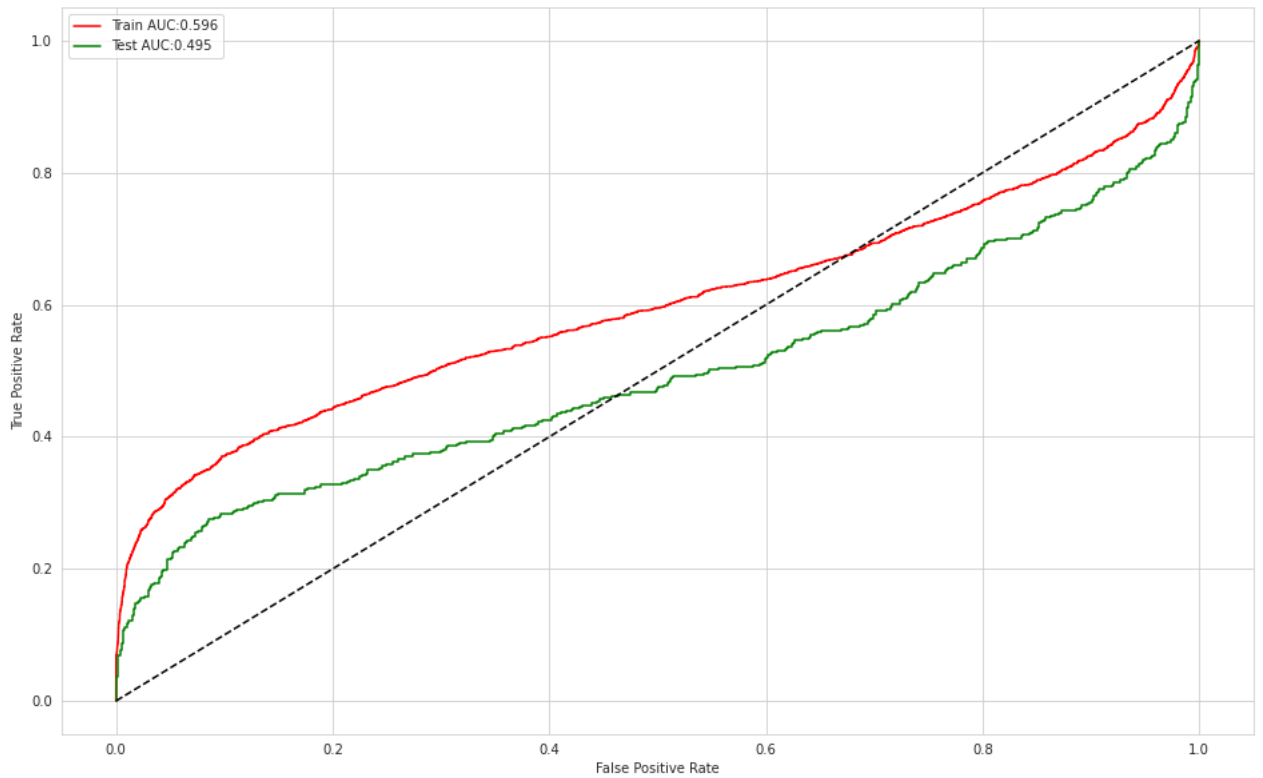
Logistic Regression

Figure 6.8: ROC curve of Logistic Regression

Figure 6.8 illustrates the AUC of this Logistic Regression model for the training set is 0.596 and for the test set is 0.495. Since AUC for both sets of data is near to 0.5, it is evident that the model is incapable of classifying our data.
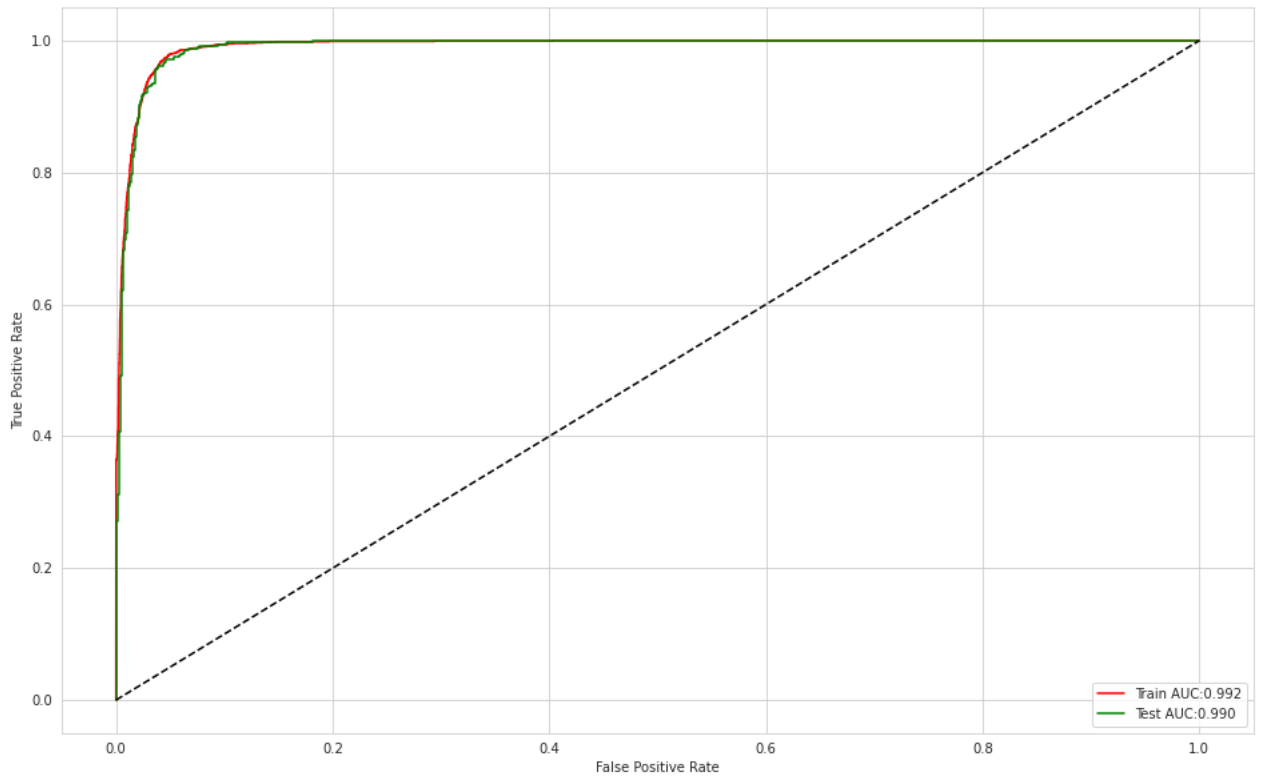
Support Vector Classifier (SVC)

Figure 6.9: ROC curve of SVC

Figure 6.9 illustrates the ROC curve of the Support Vector Classifier, which has an AUC of 0.992 for the training set and 0.9902 for the test set. Given that the AUC for both datasets is near to 1, it is evident that our data can be identified by this model.

Decision Tree Classifier
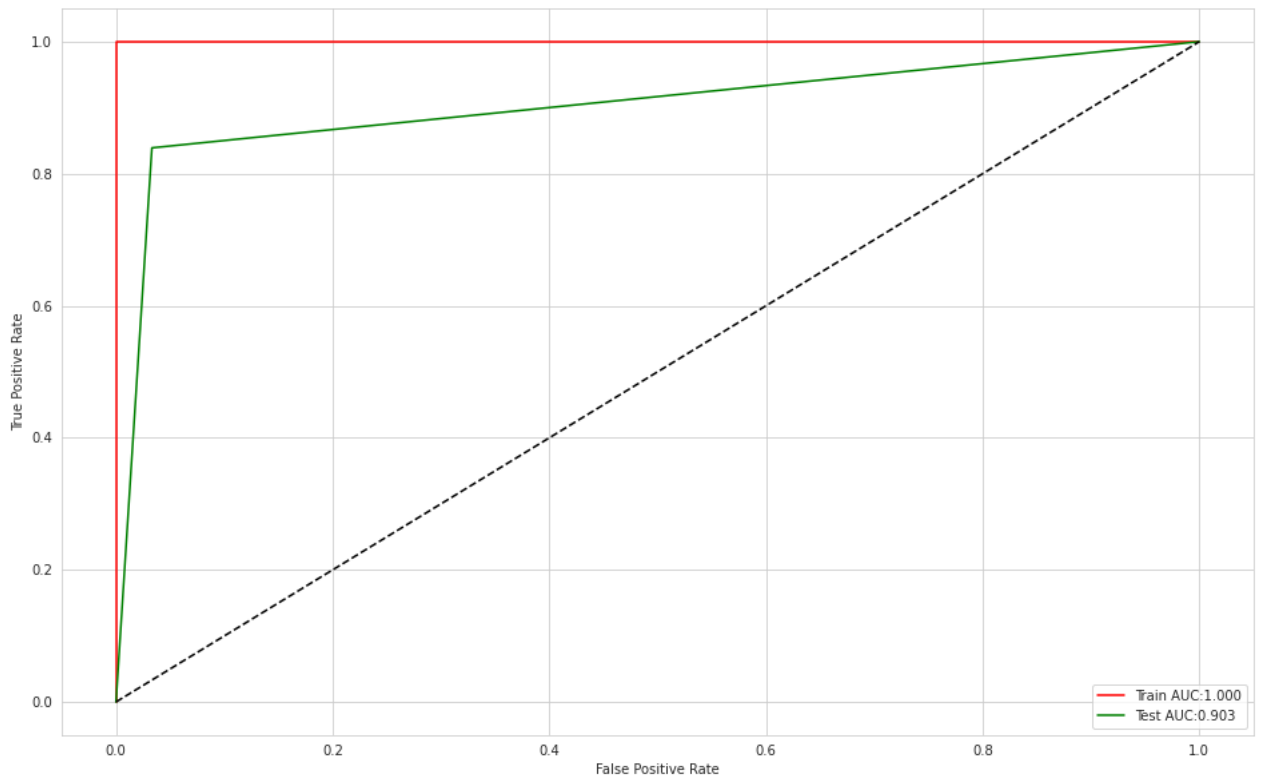
Figure 6.10: ROC curve of the Decision Tree Classifier

Figure 6.10 displays the ROC curve of the Decision Tree Classifier, which has an AUC of 1.0 for the training set and 0.903 for the test set. Given that the AUC for both datasets is close to 1, it is apparent that this model can be used to classify our data.

Gradient Boosting Classifier

Figure 6.11: ROC curve of Gradient Boosting Classifier

With the training data and test data displayed in Figure **??**, the Gradient Boosting Classifier has an AUC of 0.999. Our data can clearly be classified using this model, as evidenced by the AUCs near 1 for both datasets.

Multilayer Perceptron (MLP)
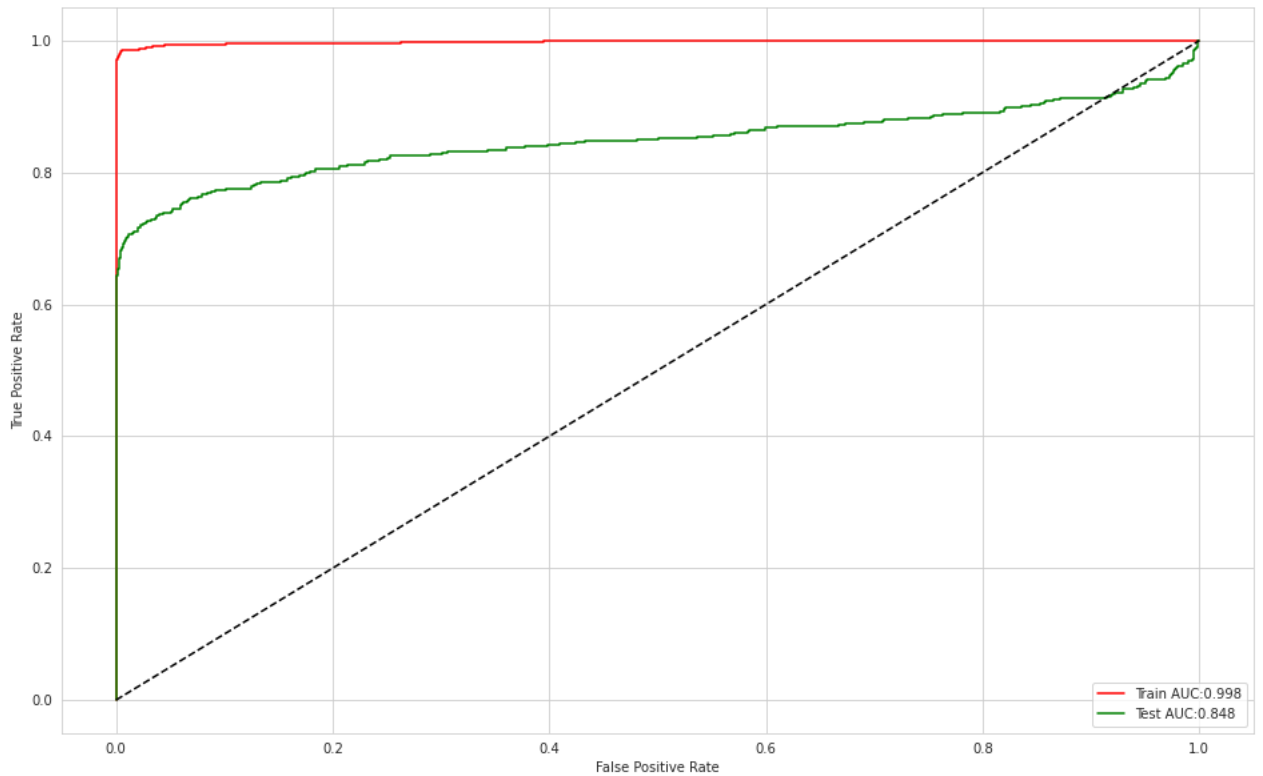
Figure 6.12: ROC curve of Multilayer Perceptron
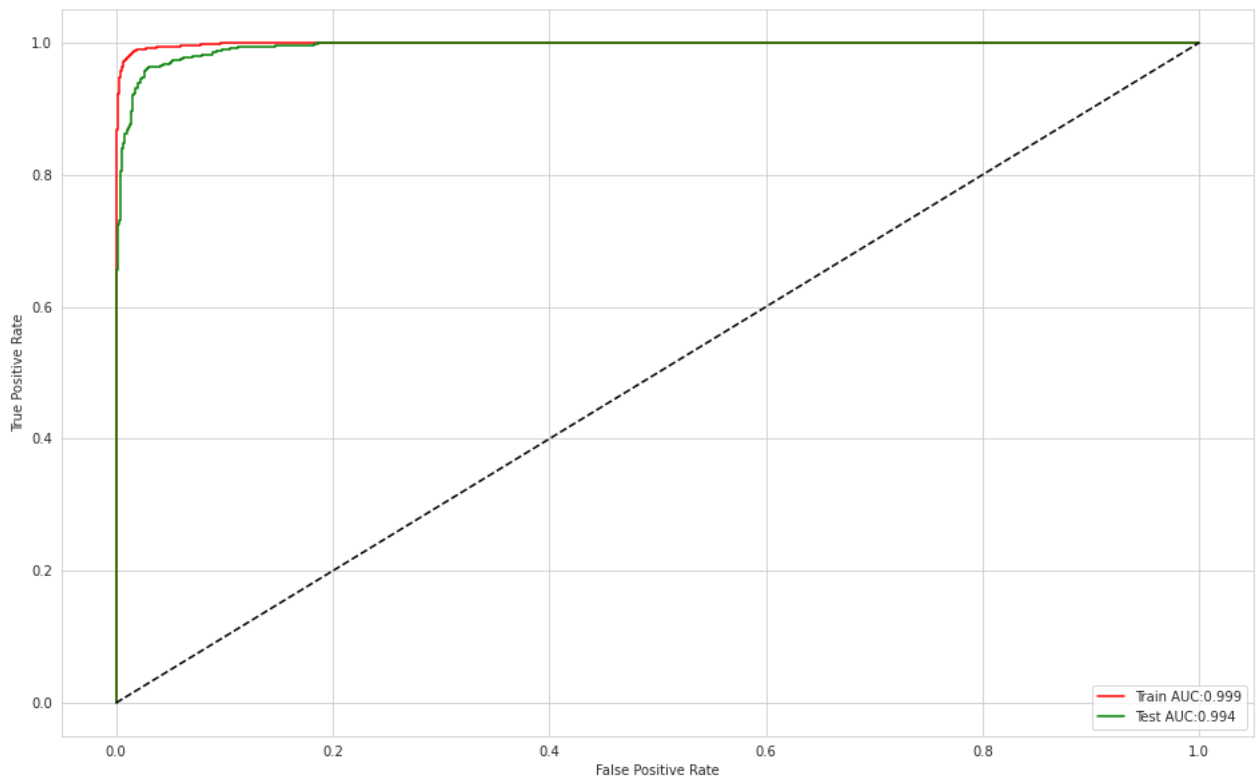
As shown in Figure 6.12, the Multilayer Perceptron Classifier has an AUC of 0.998 for the training data and 0.848 for the test data. Although the AUC for train datasets is close to 1, it is low for the test dataset. This model can still categorize our data, but it will not be efficient compared to other models.

The ROC curves of the six classifiers are shown in the graphs above. AUC is good for all six classifiers except Logistic Regression. Random Forest has the highest train and test AUC, which is 100 and 99 percent, and slightly higher than the Gradient Boosting Classifier's train and test AUC, which is 99.9 and 99.4 percent, respectively. In every experiment, the Random Forest Classifier outperformed all other classifiers. In terms of ROC and confusion matrix, the Random Forest Classification performed impressively.

6.2   Results Comparison

Different performance measures including Accuracy, Recall, Precision, and F1-score were compared using bar charts to illustrate the differences among classifiers.

Accuracy
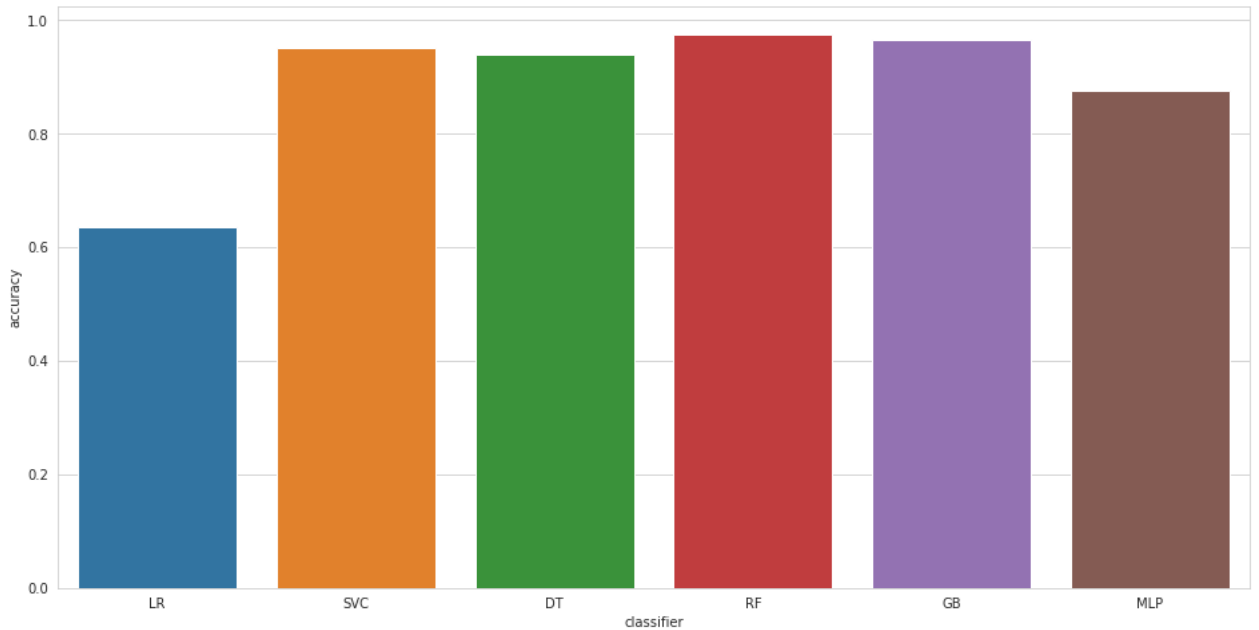
Figure 6.13: Accuracy of Used Classifiers

In Figure 6.13, we presented a bar graph of the accuracy analysis of the classifiers. According to the graph, we can see that the accuracies of the different classifiers don't vary to a large extent. The largest bar is for the random forest classifier, which shows the random forest classifier gives the highest accuracy. The second best classifier in this bar graph is the Gradient Boosting Classifier, which is greater than the SVC and Decision Tree Classifier accordingly. These four classifiers have small differences in accuracy between them. Next comes MLP, which has a much lower accuracy rate compared to the four classifiers mentioned above. And lastly, logistic regression has the lowest accuracy.

Precision

Figure 6.14: Precision Analysis

In Figure 6.14, we presented a bar graph of the precision analysis of the classifiers. According to the graph, we can see that the precision rate of the different classifiers varies. The largest bar is the RF bar, which is the Random Forest Classifier. It indicates that the random forest classifier has the highest precision rate. The second best classifier in this bar graph is gradient boosting, which is greater than SVC or decision tree. Next comes MLP, which has a lower precision rate than the decision tree. According to the bar graph, logistic regression gives the lowest precision rate compared to the other five classifiers.

Recall
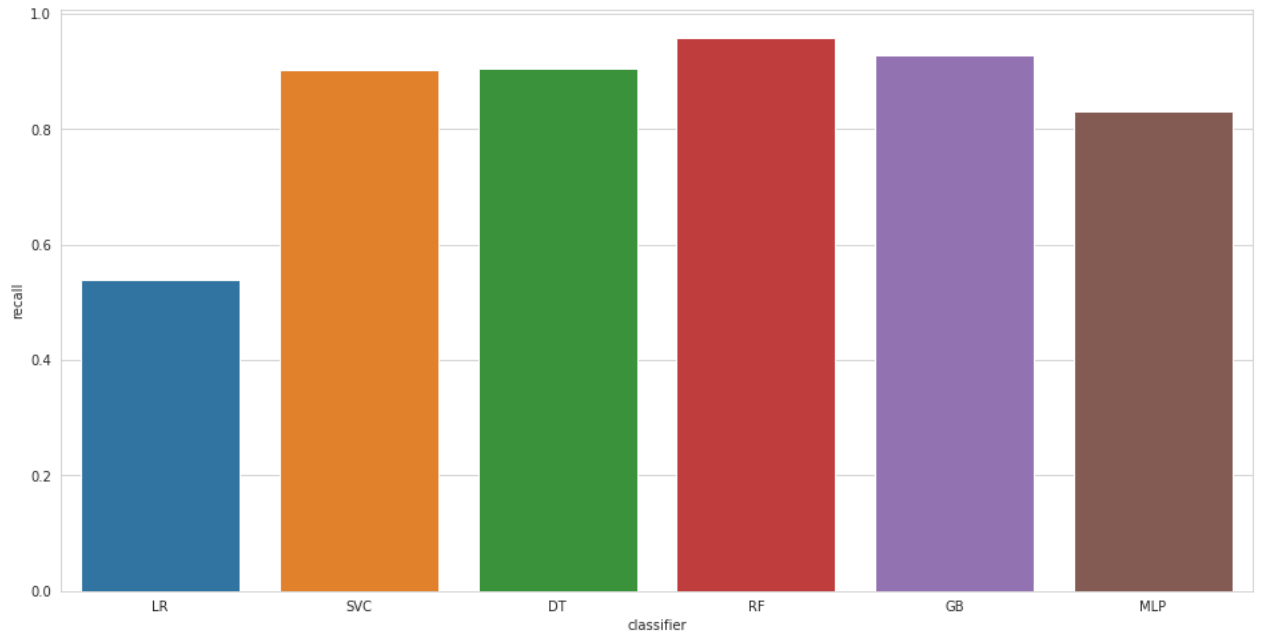
Figure 6.15: Recall Analysis

Figure 6.15 presents a bar graph of the recall analysis of the classifiers. According to the graph, the largest bar is random forest, which illustrates that random forest classifier has the greatest recall rate. The second best classifier in this bar graph is gradient boosting, which is greater than SVC. The decision tree bar is smaller than the SVC, which means it has a lower recall rate than the SVC. The MLP gives a lower precision rate than the decision tree. According to the bar graph, logistic regression gives a much lower recall rate, which is the lowest in the graph.
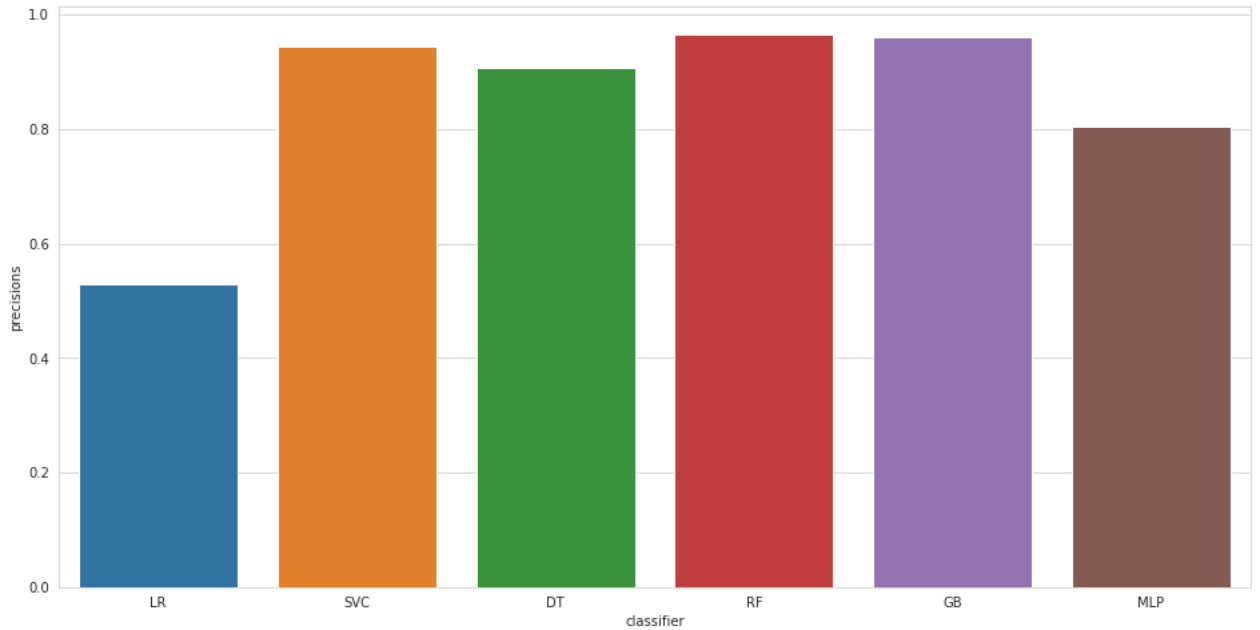
F1 Score
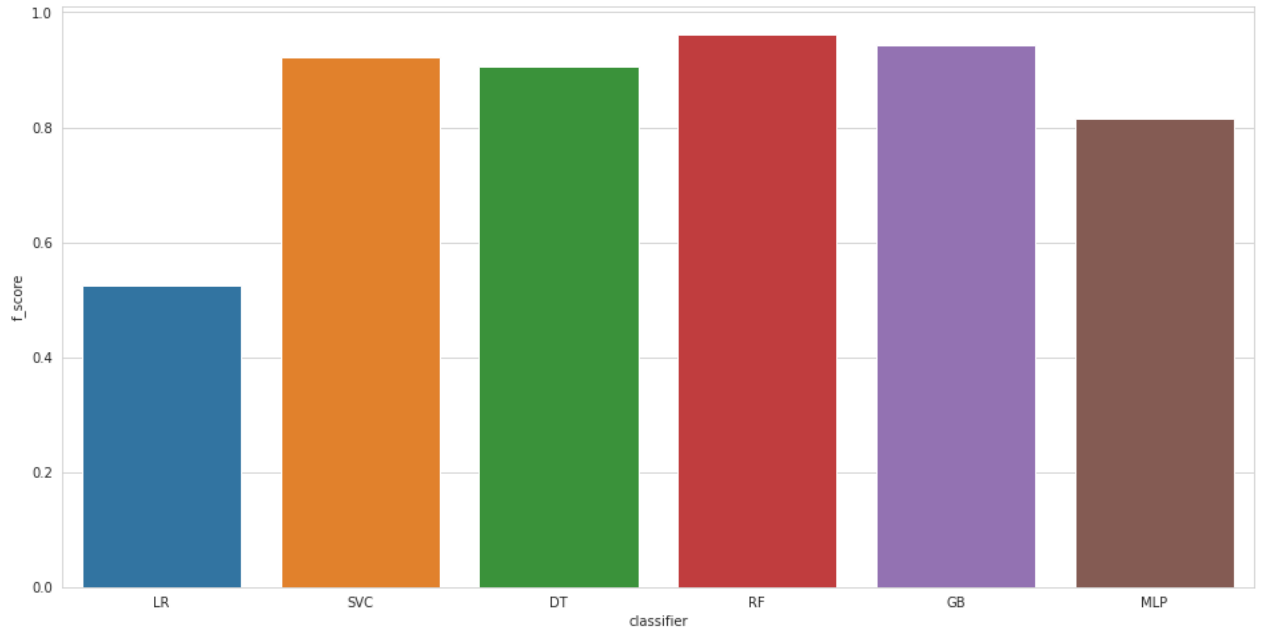
Figure 6.16: F1 Score Analysis

In Figure 6.16, we presented a bar graph of the F1 score analysis of the classifiers. According to the graph, we can see the F1 score of the random forest classifier is at top . The second best classifier in this bar graph is gradient boosting, which is greater than SVC. The decision tree classifier gives a lower f1 score than SVC and a higher f1 score than MLP. The logistic regression has a much lower F1 score than MLP. According to the bar graph, logistic regression gives the lowest precision rate compared to the other classifiers.

Classifiers	Accuracy	Recall	Precision	F1 score
Logistic regression	63.5652%	53.9130%	52.8983%	53.9130%
SVC	95.2174%	90.3261%	94.3870%	90.3261%
Random Forest Classifier	97.6087%	95.8967%	96.5848%	95.8967%
Decision Tree Classifier	94.0000%	90.4620%	90.7283%	90.4620%
Gradient Boosting Classifier	96.4783%	92.8261%	95.9919%	92.8261%
Multilayer Perceptron	87.6522%	82.9891%	80.3321%	82.9891%

Table 6.1: Comparison table among Logistic Regression, SVC, Decision Tree, Random Forest, Gradient boosting, MLP

Here, Table 6.1 illustrates the comparison of performance among the six classifiers which we used in our research with respect to different performance metrics such as accuracy, recall, precision, and F1-score. Logistic regression has an accuracy rate of 63.5652%, which is the lowest accuracy in our findings with a recall rate of 53.9130%, a very low precision rate of 52.8983% and an F1 score of 53.9130%. It is because there are fewer observations than features. The MLP classifier performs better than logistic regression. We got an 87.6522.0% accuracy rate and a 80.3321% precision rate. The decision tree classifier obtains 94.0000% accuracy and a 90.7283% precision rate. The SVC classifier gives better results compared to logistic regression, MLP, and decision tree classifiers. It performs with 95.2174% of accuracy and 90.3261% of precision, with an F1 score of 90.3261%.

We achieved the best performance by using the random forest classifier in our findings. It gives 97.6087% accuracy with a high precision of 96.5848%, a recall rate of 95.8967% and an F1 score of 95.8967%. The very high values of the indicators might be due to the embedded feature selection of our model. The second best performance was obtained by the gradient boosting classifier. It gives a 96.4783% accuracy rate and a 95.9919% precision rate, which is also sufficiently good.

We considered the F1 score as our key indicator because it compares precision and recall and works well with unbalanced datasets. Therefore, compared to the other five models, the random forest classifier is the model of preference in this study, with the highest F1-score for our unbalanced dataset.

We compared the performance of our best-fit classifier, Baseline Random Forest, to that of PCA with RF and clustering to see if clustering could increase performance. We used k-means clustering and hierarchical clustering to show the difference.

F1 score of Random Forest with K-means Clustering	0.9989
F1 score of Random Forest with Hierarchical Clustering	0.9978
F1 score of Random Forest with No Clustering	0.9989
F1 score of Basic baseline Random Forest	0.9589

Table 6.2: Comparison of F1 scores of Random Forest with and without clustering

Table 6.2 shows that the F1 score of the random forest increased after performing dimension reduction. But if we compare the F1 scores of three different clusters, clustering makes a slight difference here, but the effect is negligible because the performance is already maxed out. Thus, we can come to the conclusion that since our dataset was preprocessed, it doesn't need any further clustering.

Chapter 7

Conclusion

Reliable and accurate detection is becoming more common, so epilepsy is becoming more necessary nowadays. The accurate detection of seizures from a huge dataset is exceedingly challenging. In this research, we have used a preprocessed data set featuring epileptic seizure detection, which was collected from the UCI Machine Learning Repository [46]. In this study, we thoroughly studied and examined a wide range of machine learning classifiers for seizure detection. This topic has revealed the development of continuous effort, but it also raises a few important research issues that we have faced during the research. We outline important problems in this section that can help to further future research in this field. The crucial choice of choosing appropriate features and machine learning classifiers to reduce computation time due to the large volume and high dimension of our data set was a crucial one. Since our dataset has a large set of features, we needed to reduce the dimension to work with the selected features.

Therefore, we come to the conclusion that the 'Random Forest' classifier comes out to be the most successful in our study. It gives 97.6087% accuracy with a high precision of 96.5848% and an F1 score of 95.8967%. We considered model accuracy, model loss, recall, and precision, as well as the confusion matrix, when evaluating performance. The results suggest that by implementing dimension reduction and clustering, classification performance can be slightly improved. For example, dimension reduction improved the F1 score of the random forest classifier by 4%, which is

99.89%, but clustering couldn't bring any changes to the performance as the dataset was already preprocessed and clustered.

Our future goal for this research is to implement hyperparameter tuning of our random forest model. We hope that it will solve the bias and variance performance of the model. Developing ensemble models may be of benefit. Although our performance is nearly excellent at this moment.

Bibliography

[1] R. J. Racine, "Modification of seizure activity by electrical stimulation: Ii. motor seizure," *Electroencephalography and clinical neurophysiology*, vol. 32, no. 3, pp. 281–294, 1972.

[2] P. Gloor, G. Ball, and N. Schaul, "Brain lesions that produce delta waves in the eeg," *Neurology*, vol. 27, no. 4, pp. 326–326, 1977.

[3] O. C. Cockerell, J. Sander, Y. Hart, S. Shorvon, and A. Johnson, "Remission of epilepsy: Results from the national general practice study of epilepsy," *The Lancet*, vol. 346, no. 8968, pp. 140–144, 1995.

[4] N. Vapnik Vladimir, *The nature of statistical learning theory*, 1995.

[5] T. Büttner, B. Lack, M. Jäger, *et al.*, "Serum levels of neuron-specific enolase and s-100 protein after single tonic-clonic seizures," *Journal of neurology*, vol. 246, no. 6, pp. 459–461, 1999.

[6] S. P. Levine, J. E. Huggins, S. L. BeMent, *et al.*, "A direct brain interface based on event-related potentials," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 180–185, 2000.

[7] E. Başar, C. Başar-Eroglu, S. Karakaş, and M. Schürmann, "Gamma, alpha, delta, and theta oscillations govern cognitive processes," *International journal of psychophysiology*, vol. 39, no. 2-3, pp. 241–248, 2001.

[8] M. A. Cortez, C. McKerlie, and O. C. Snead, "A model of atypical absence seizures: Eeg, pharmacology, and developmental characterization," *Neurology*, vol. 56, no. 3, pp. 341–349, 2001.

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

[10] J. Ashburner, J. G. Csernansk, C. Davatzikos, N. C. Fox, G. B. Frisoni, and P. M. Thompson, "Computer-assisted imaging to assess brain structure in healthy and diseased brains," *The Lancet Neurology*, vol. 2, no. 2, pp. 79–88, 2003.

[11] H. G. Feichtinger and T. Strohmer, *Advances in Gabor analysis.* Springer Science & Business Media, 2003.

[12] H.-T. Lin and C.-J. Lin, "A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods," *submitted to Neural Computation*, vol. 3, no. 1-32, p. 16, 2003.

[13] M. K. Kiymik, A. Subasi, and H. R. Ozcalık, "Neural networks with periodogram and autoregressive spectral analysis methods in detection of epileptic seizure," *Journal of Medical Systems*, vol. 28, no. 6, pp. 511–522, 2004.

[14] M. A. Leone, A. Solari, E. Beghi, *et al.*, "Treatment of the first tonic-clonic seizure does not affect long-term remission of epilepsy," *Neurology*, vol. 67, no. 12, pp. 2227–2229, 2006.

[15] W. D. Shields, "Infantile spasms: Little seizures, big consequences," *Epilepsy Currents*, vol. 6, no. 3, pp. 63–69, 2006.

[16] A. Subasi, "Eeg signal classification using wavelet feature extraction and a mixture of expert model," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1084–1093, 2007.

[17] A. T. Berg, "Risk of recurrence after a first unprovoked seizure," *Epilepsia*, vol. 49, pp. 13–18, 2008.

[18] G. M. Fenichel, *Clinical pediatric neurology: a signs and symptoms approach.* Elsevier Health Sciences, 2009.

[19] H. Ocak, "Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2027–2036, 2009.

[20]   A. T. Tzallas, M. G. Tsipouras, and D. I. Fotiadis, "Epileptic seizure detection in eegs using time–frequency analysis," *IEEE transactions on information technology in biomedicine*, vol. 13, no. 5, pp. 703–710, 2009.

[21]   A. Subasi and M. I. Gursoy, "Eeg signal classification using pca, ica, lda and support vector machines," *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.

[22]   V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure eeg signals using empirical mode decomposition," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 6, pp. 1135–1142, 2011.

[23]   E. J. Donner, "Explaining the unexplained; expecting the unexpected: Where are we with sudden unexpected death in epilepsy? explaining the unexplained; expecting the unexpected," *Epilepsy currents*, vol. 11, no. 2, pp. 45–49, 2011.

[24]   C. C. Jouny, P. J. Franaszczuk, and G. K. Bergey, "Improving early seizure detection," *Epilepsy & Behavior*, vol. 22, S44–S48, 2011.

[25]   F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[26]   ——, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[27]   ——, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28]   ——, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[29]   ——, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[30]   U. R. Acharya, F. Molinari, S. V. Sree, S. Chattopadhyay, K.-H. Ng, and J. S. Suri, "Automated diagnosis of epileptic eeg using entropies," *Biomedical Signal Processing and Control*, vol. 7, no. 4, pp. 401–408, 2012.

[31] U. R. Acharya, S. V. Sree, A. P. C. Alvin, and J. S. Suri, "Use of principal component analysis for automatic classification of epileptic eeg activities in wavelet framework," *Expert Systems with Applications*, vol. 39, no. 10, pp. 9072–9078, 2012.

[32] R. B. Pachori, "Classification of seizure and nonseizure eeg signals using empirical mode decomposition," 2012.

[33] J. W. Wheless, P. A. Gibson, K. L. Rosbeck, *et al.*, "Infantile spasms (west syndrome): Update and resources for pediatricians and providers to share with parents," *BMC pediatrics*, vol. 12, no. 1, pp. 1–9, 2012.

[34] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas, "A new eeg acquisition protocol for biometric identification using eye blinking signals," *International Journal of Intelligent Systems and Applications*, vol. 7, no. 6, p. 48, 2015.

[35] N. K. Ambulkar and S. Sharma, "Detection of epileptic seizure in eeg signals using window width optimized s-transform and artificial neural networks," in *2015 IEEE Bombay section symposium (IBSS)*, IEEE, 2015, pp. 1–6.

[36] B. Boashash, *Time-frequency signal analysis and processing: a comprehensive reference*. Academic press, 2015.

[37] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based eeg processing for computer-aided seizure detection and epilepsy diagnosis," *Seizure*, vol. 26, pp. 56–64, 2015.

[38] R. Llinas and M. N. Negrello, "Cerebellum," *Scholarpedia*, vol. 10, no. 1, p. 4606, 2015.

[39] C. E. Stafstrom and L. Carmant, "Seizures and epilepsy: An overview for neuroscientists," *Cold Spring Harbor perspectives in medicine*, vol. 5, no. 6, a022426, 2015.

[40] D. Triantafyllopoulos, P. Korvesis, I. Mporas, and V. Megalooikonomou, *Real-time management of multimodal streaming data for monitoring of epileptic patients - journal of medical systems*, Dec. 2015. [Online]. Available: https://link.springer.com/article/10.1007/s10916-015-0403-3.

[41]   J. W. Kalat, *Introduction to psychology*. Cengage Learning, 2016.

[42]   L. Li, S. Liu, Y. Peng, and Z. Sun, "Overview of principal component analysis algorithm," *Optik*, vol. 127, no. 9, pp. 3935–3944, 2016.

[43]   D. Vance, S. H. Little, C. Archer, *et al.*, "The oceanic budgets of nickel and zinc isotopes: The importance of sulfidic environments as illustrated by the black sea," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2081, p. 20 150 294, 2016.

[44]   C. Albon, *Svc parameters when using rbf kernel*, Dec. 2017. [Online]. Available: https://chrisalbon.com/code/machine_learning/support_vector_machines/svc_parameters_using_rbf_kernel/.

[45]   K. E. Misulis and E. L. Murray, *Essentials of Hospital Neurology*. Oxford University Press, 2017.

[46]   Q. Wu and E. Fokoue, *Epileptic seizure recognition data set*, May 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Epileptic+Seizure+Recognition.

[47]   R. Karim, S. Nitol, M. Rahman, *et al.*, "Epileptic seizure detection by exploiting eeg signals using different decomposition techniques and machine learning approaches," Ph.D. dissertation, BRAC University, 2018.

[48]   S. T. Snyder, "Time series modeling of clinical electroencephalogram data-an information theory approach," Ph.D. dissertation, Ohio University, 2018.

[49]   N. Sriraam, S. Raghu, K. Tamanna, *et al.*, "Automated epileptic seizures detection using multi-features and multilayer perceptron neural network," *Brain informatics*, vol. 5, no. 2, pp. 1–10, 2018.

[50]   R. Carter, *The human brain book: An illustrated guide to its structure, function, and disorders*. Penguin, 2019.

[51]   N. B. Gaikwad, V. Tiwari, A. Keskar, and N. Shivaprakash, "Efficient fpga implementation of multilayer perceptron for real-time human activity classification," *IEEE Access*, vol. 7, pp. 26 696–26 706, 2019.

[52] R. Grosse, *Multilayer perceptrons*, 2019. [Online]. Available: https://www.cs.toronto.edu/~mren/teach/csc411_19s/lec/lec10_notes1.pdf?fbclid= IwAR0vcjnHrYh4-XioVa2myysCshz55IO9WPBPTeRl3OzF-5jWiIxPxEako3Y.

[53] A. Navlani, *Python logistic regression tutorial with sklearn & scikit*, Dec. 2019. [Online]. Available: https://www.datacamp.com/tutorial/understanding-logistic-regression-python.

[54] D. Nelson, *Gradient boosting classifiers in python with scikit-learn*, Aug. 2019. [Online]. Available: https://stackabuse.com/gradient-boosting-classifiers-in-python-with-scikit-learn/#:%5C%7E:text=Gradient%5C%20boosting%5C%20classifiers%5C%20are%5C%20a,used%5C%20when%5C%20doing%5C%20gradient%5C%20boosting..

[55] S. Ronaghan, *The mathematics of decision trees, random forest and feature importance in scikit-learn and spark*, Nov. 2019. [Online]. Available: https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3.

[56] F. PEIXOTO, *A simple overview of multilayer perceptron (mlp) deep learning*, Dec. 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/12/mlp-multilayer-perceptron-simple-overview/.

[57] R. Sahu, S. R. Dash, L. A. Cacha, R. R. Poznanski, and S. Parida, "Epileptic seizure detection: A comparative study between deep and traditional machine learning techniques," *Journal of integrative neuroscience*, vol. 19, no. 1, pp. 1–9, 2020.

[58] A. Bhandari, *Confusion matrix for machine learning*, Jul. 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/.

[59] H. Bonthu, *An introduction to logistic regression*, Jul. 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/07/an-introduction-to-logistic-regression/#:~:text=Logistic%20Regression%20is%20a%20%E2%80%9CSupervised,used%20for%20Binary%20classification%20problems.

[60] M. Fuchs, *Nn - multi-layer perceptron classifier (mlpclassifier) - michael fuchs python*, Feb. 2021. [Online]. Available: https://michael-fuchs-python.netlify. app/2021/02/03/nn-multi-layer-perceptron-classifier-mlpclassifier/?fbclid= IwAR0vcjnHrYh4-XioVa2myysCshz55IO9WPBPTeRl3OzF-5jWiIxPxEako3Y# mlpclassifier-for-binary-classification.

[61] Z. Jaadi, *A step-by-step explanation of principal component analysis (pca)*, Dec. 2021. [Online]. Available: https://builtin.com/data-science/step-step-explanation-principal-component-analysis?fbclid=IwAR3xKASJ7Z3LaFwtWdNoF2Qhnh3D96

[62] S. E. R, *Random forest: Introduction to random forest algorithm*, Jun. 2021. [Online]. Available: https : / / www . analyticsvidhya . com / blog / 2021 / 06 / understanding-random-forest/.

[63] A. Saini, *Random forest algorithm for absolute beginners in data science*, Nov. 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/10/ an-introduction-to-random-forest-algorithm-for-beginners/.

[64] P. Sharma, *Understanding k-means clustering algorithm*, Nov. 2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/11/understanding-k-means-clustering-in-machine-learningwith-examples/.

[65] M. Smith, *Types of seizures and their symptoms*. [Online]. Available: https: //www.webmd.com/epilepsy/types-of-seizures-their-symptoms.