T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Information Technologies

# FOURIER TRANSFORM BASED EPILEPTIC SEIZURE FEATURES CLASSIFICATION USING SCALP ELECTRICAL MEASUREMENTS USING KNN AND SVM

Athar AL-AZZAWI

Master of Science

Supervisor

Asst. Prof. Dr. Abdullahi Abdu Ibrahim

Istanbul, 2021

# FOURIER TRANSFORM BASED EPILEPTIC SEIZURE FEATURES CLASSIFICATION USING SCALP ELECTRICAL MEASUREMENTS USING KNN AND SVM

by

**Athar Hussein Ali**

Information Technologies

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2021

The thesis titled “FOURIER TRANSFORM BASED EPILEPTIC SEIZURE FEATURES CLASSIFICATION USING SCALP ELECTRICAL MEASUREMENTS USING KNN AND SVM” prepared and presented by “Athar AL-AZZAWI” was accepted as a Master of Science Thesis in Information Technologies.

__________________________________

Asst. Prof. Dr. Abdullahi Abdu Ibrahim

Supervisor

Thesis Defense Jury Members:

| | | |
|---|---|---|
| Asst. Prof. Dr. Abdullahi Abdu Ibrahim | School of Engineering and Natural Sciences, Altinbas University | ________________ |
| Asst. Prof. Dr. Sefer Kurnaz | School of Engineering and Natural Sciences, Altinbas University | ________________ |
| Prof. Tareq Abed Mohammed | Faculty of Computer Science and Information Technology Kirkuk University | ________________ |

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Athar Hussein Ali

Signature

# DEDICATION

This thesis is dedicated to:

To God Almighty my creator, the source of all the Grace that I had through all my life and especially on this journey.

To the person who is no longer in this world and no one ever can replace him, the one who made me how I am, father.

To my source strength and my partner in confrontation the life, mother.

To the source of my motivation of challenging and hard-working, who supported me all the way, my sister and my brothers.

To all my relatives and friends who supported me and pray for me.

To my two special persons, (Ruaa Tamim, Shahad Kawaz) whom without them I would not have the opportunity to write this at these moments.

## ACKNOWLEDGEMENTS

To whom, if not for their support and enthusiasm this thesis would not see the sun :

My research supervisor is Dr. Abdullahi Abdu Ibrahim, Department of Computer Engineering, Altinbas University.

In the real world, we keep looking for a true person who would be our friend before anything else. So, when our supervisor became that person that means we won the lottery. In these few lines, I want to spread some happy feelings and express the grateful ones that I have to him for giving me the opportunity to do my research under his supervision. It was a great privilege and honor to work and study under your guidance.

My mentor Assoc. Prof. Dr. Adil Deniz Duru, Faculty of Sport Science, Marmara University.

This is the hardest part I faced ever; due to that everything can be repaid except the knowledge. The word " thanks" is meaningless in front of his spirit, ethic, generosity, and his patience all these months. He had been like an angel who guards me through this study epic, shows me the way, and lets me make mistakes then encourages me to do the right and show me how to do it. By his directive and instruction, my dream saw the light, and my sun shined. So, I would like to thank him very much for the precious time that he spent it teaching me. Where every step in my future road will have his signature on. And I am hoping that he will guide me in the next big step when I fighting to get the Ph.D.

My friend and my Colleague Saif.Al-Jumaili.

My thank would not be enough to express my gratitude for every time he let me sneak and steal his time to polish my prose in the final manuscript; also inspiring me to move on to the next level in my education. On the same side, he was like a cold-blooded killer when it came to reviewing my work! so, thank you for making my work better.

## ABSTRACT

# FOURIER TRANSFORM BASED EPILEPTIC SEIZURE FEATURES CLASSIFICATION USING SCALP ELECTRICAL MEASUREMENTS USING KNN AND SVM

Al-Azzawi, Athar,

M.Sc., Information Technologies, Altınbaş University,

Supervisor: Asst. Prof. Abdullahi Abdu Ibrahim

Date: 08/2021

Pages: 83

The great development that took place in the technology of the interaction between humans and computers led to remarkable and incredible success in many scientific fields. Until this day, researchers' studies are continuing in this field to reach the highest possible accuracy to obtain the results that approximate the accuracy of human work. Electroencephalogram (EEG) is one of the devices that took a wide space and led many studies to amazing results in the field of recording, analyzing, detecting, and classifying brain signals. Where this technology was able to monitor the disorders, which happened in the brain and provide the ability to study the state of health of the brain. In addition, machine learning had Various techniques that were successfully involved in the classification of EEG signals. Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) were our specialties here among the most suitable techniques for classifying EEG data. This thesis aims to build a system to assist clinicians on three levels, first assisting in patients' automatic monitoring which leads to a digital memory easy to memorize. Secondly,

reducing the work time by Supporting the clinicians' decision which assists acceleration in getting the goals with the least number of errors. Finally, assisting in identifying the appropriate medication and medical care that patients may need. Thus, for achieving this purpose, the dataset used here was from Temple University Hospital Seizure Corpus (TUH) [1]. TUH is considered one of the largest open-source databases and is the most extensive [2-4]. Moreover, to ensure the best results preprocessing techniques were implemented to Dataset. There are many techniques for EEG signal processing that feed the classification models with the best data. Fast Fourier Transform (FFT) was studied as one of the types of feature extraction methods for processing EEG signals. Eventually, the results associated with classifying seizure types showed SVM got the best classification accuracy compared with KNN where the accuracy was 99.5 % and 99 %, respectively.

**Keywords:** Electroencephalogram (EEG), Fast Fourier Transform (FFT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM).

# Contents

Pages

# LIST OF TABLES

**Pages**

## LIST OF FIGURES

Pages

## LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| AI | : | Artificial Intelligence |
| ML | : | Machine Learning |
| HCI | : | Human-Computer Interaction |
| BCI | : | Brain Computer Interaction |
| SVM | : | Support Vector Machine |
| KNN | : | K-Nearest Neighbor |
| FFT | : | Fast Fourier Transform |
| MLE | : | Maximum Likelihood Estimation |
| LVQ | : | Learning Vector Quantization |
| AR | : | Auto-Regressive |
| MFCC | : | Mel Frequency Cepstral Coefficients |
| TFD | : | Time Frequency Distributions |
| WT | : | Wavelet Transform |
| EEG | : | Electroencephalogram |
| MRI | : | Magnetic Resonance Imaging |
| CNS | : | The Central Nervous System |
| TUH | : | The Temple University Hospital |
| EDF | : | European Data Format |
| GNSZ | : | Generalized Non-Specific Seizure |
| FNSZ | : | Focal Non-Specific Seizure |
| TCSZ | : | Tonic-Clonic Seizure |
| TP | : | True Positive |
| TN | : | True Negative |
| FP | : | False Positive |
| FN | : | False Negative |

## LIST OF SYMBOLS

| | | |
|---|---|---|
| S(f) | : | Frequency Domain Signal |
| S(t) | : | Time Domain Signal |
| F | : | Frequency |
| T | : | Time |
| N | : | Number of Dimensions |
| pi, qi | : | Data Points |

# 1. INTRODUCTION

Artificial intelligence was in the lead of the fields that shared and led to a revolutionary technological development in Improving the quality of life. AI almost replaced humans in some places and achieved better results with some additional advantages such as the time factor, and the error rate. For instance, in the medical field, the time factor, one of the most critical factors in life, depends on it or it could determine the lifestyle. Therefore, lots of devices, techniques, and methods were developed and connected with AI fields to enhance the life pattern as much as possible. Electroencephalogram (EEG) was one of the devices that associated AI with the life problem and was a source of a lot of successful research and studies.

## 1.1 THE ELECTROENCEPHALOGRAM (EEG)

The magic wand that allows us to unpack brain-behavior which is the most complicated, significant, and mysterious organ ever in the human body is called the electroencephalogram (EEG). EEG is a tool that enables us to detect and measure what happens inside the neurons and record the electrical activity that occurs in the brain [2, 5]. It is one of the devices that made a technological revolution and important development in the scientific aspect in general and the medical aspect in particular [6-11]. Hans Berger the German neurologist in the 1920s was the first one who applied EEG on humans [3]. The EEG device remains at the forefront in recording brain signals, despite some new technologies used for the same purposes such as computed tomography (CT) and magnetic resonance imaging (MRI), etc. Due to its advantages which are lower costs, excellent time resolution where it catches in the same occurs time, and easy to carry and use which enable us to perform the scan anywhere and even if there is no behavioral appearance [12]. Although EEG is considered a non-invasive tool, in some cases it is necessary to record the signals from inside the skull, for example, when patients' signals appear normal or when recorded signals contain a lot of artifacts that damage the disease's diagnosis. Many diseases related to the brain have been successfully diagnosed and have been handled by EEG such as epilepsy, coma, brain death, Alzheimer’s, sleep disorders, and other brain disorders [13-17]. Brain activity that occurs in the brain and between neurons is captured by attaching small pieces of electrodes and recording all brain parts. The duration of the routine EEG recording is 20-30 minutes generally, in some cases no seizures may appear in the first session, so patients may need several sessions or longer duration sessions [13, 18-22]. In addition to the signals of the brain, the EEG reports contain information about the patient's and his medical condition, description of the sessions such as time, duration, locations of disorder, and if the EEG scan was normal or abnormal. Interpretation of the EEG results is performed by certified physicians and may take some time according to the physician's skill and decision. The abnormalities that had been observed through the EEG were ictal, interictal, and in-between sequences seizures [23]. In addition to all EEG Advantages, what improved the power of EEG in the clinical neurophysiology field was its ability to reflect all the activities of the brain [24-28].

## 1.2 BRAIN STRUCTURE

The brain on average is around 1.4 kg and holds one hundred billion nerve cells. The most complicated, significant, and mysterious organ ever in the human body is the brain. Memory, movements, decision, and feeling are controlled by the brain which is the central nervous system (CNS), all these information are transmitted by electrical signals [29]. The brain is categorized according to its functionality as shown in Figure 1.1 which illustrates the brain's Central Nervous System [30]. The first part is called the cerebrum which is also known as the cortex which is the front larger part of the brain, it handles the higher mental tasks such as thinking and sensory processing [31]. The second part is called the cerebellum and is responsible for controlling and balancing body movements. The third part is called the brainstem which is responsible for any function in the unconscious or autonomic process like breathing or heartbeat. The fourth part is the Limbic System, which is the emotional part of the brain and turns on feelings for different situations such as happiness, fairness, excitement, and love. The fifth part is The Occipital Lobe which is responsible for processing all the functions that are related to vision. The sixth part is the Temporal Lobe is a special unique part, where it is dealing with language and emotional feeling, and visual memory. The seventh part is the Parietal Lobe that deals in between outside source information and inner response and integrating them such as coordinating between the eyes and the hand's movement. The last part of the CNS is the Frontal Lobe, which is involved in lots of functions, where you make your decision, controlling movements in the awareness state. The most important thing about this part is containing dopamine which is responsible for lots of things such as making plans or motivation.

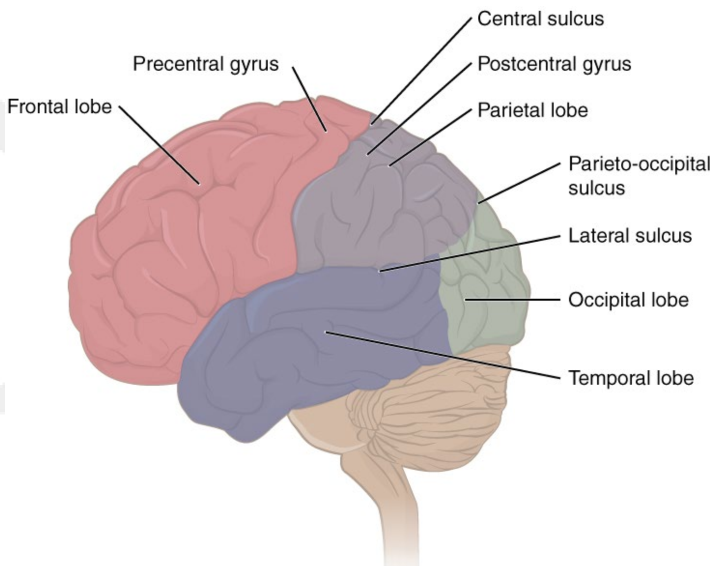



**Figure 1.1**:The Brain Central Nervous System

The brain nervous system consists of a neuron network that collects information and processes it and transmits it through the body. Where the CNS includes Central Nervous System (CNS) and Peripheral Nervous System (PNS) [32]. All the thoughts, emotions, and other mental tasks are completed through information processing and the CNS response. Memory, concentrate, and thoughts are the cerebral cortex's responsibilities which is the out layer of the cerebrum. Besides, the convoluted area of the cerebral cortex surface is the reason for the increment of the neural area. The cortex is subtracted into two symmetrical sides and the central sulcus in the middle.

### 1.3 EEG SCAN

EEG captures and records the electrical changes that occur in the brain through small pieces called electrodes. Two or more electrodes are placed to capture electrical signals, and these electrodes are distributed on the scalp according to an international system as shown in Figure 1.2. In 1994 the American Brain Imaging Association created a standardized system for determining electrode distribution sites [4, 33, 34]. One of the most popular and used systems is the 10-20 system, which later expanded to high-intensity electrode meters which are the 10-10 and 10-5 systems. These electrodes are coded and indicated by the first letter of their location on the scalp, as shown below in Table 1.1.

**Table 1.1**: Electrode Letters Corresponding to The Region

| Letters | The regions on the scalp |
|---|---|
| Fp | Prefrontal Lobe |
| F | Frontal Lobe |
| T | Temporal Lobe |
| C | Central Lobe |
| P | Parietal/Parasagittal Lobe |
| O | Occipital Lobe |
| A | Mastoid Process Lobe (bone near the outer ear) |

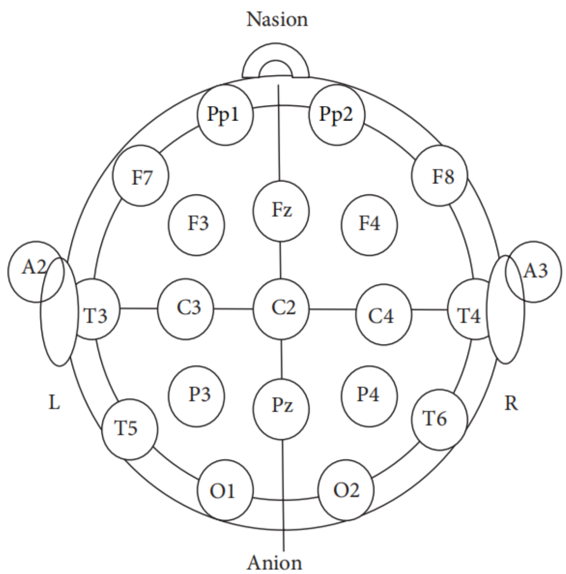



**Figure 1.2**: The International Electrodes Distributed System

These electrodes are distributed on the scalp, which ensures that all areas of the brain are covered, and care is taken to avoid areas that may contain artifacts, thus affecting the electrical signals in different shapes such as head cap as shown in Figure 1.3. The 10/20 system distributes the electrodes as follows: On the right side the even numbers of electrodes are distributed, and on the left side the odd numbers of electrodes are distributed, and the letter Z represents the midline of the scalp Where the EEG measurements usually have signals and a little noise ratio. Due to the size of the signals, they digitized the data that had been recorded and sent to an amplifier. Then after amplifying the data, it is presented as a time series of voltage values. Moreover, a lot of signals are affected or fade due to the neuron's electrical activities and even of the scalp shape [35]. Moreover, the data recorded from a person with epilepsy is completely different when a healthy one.

`

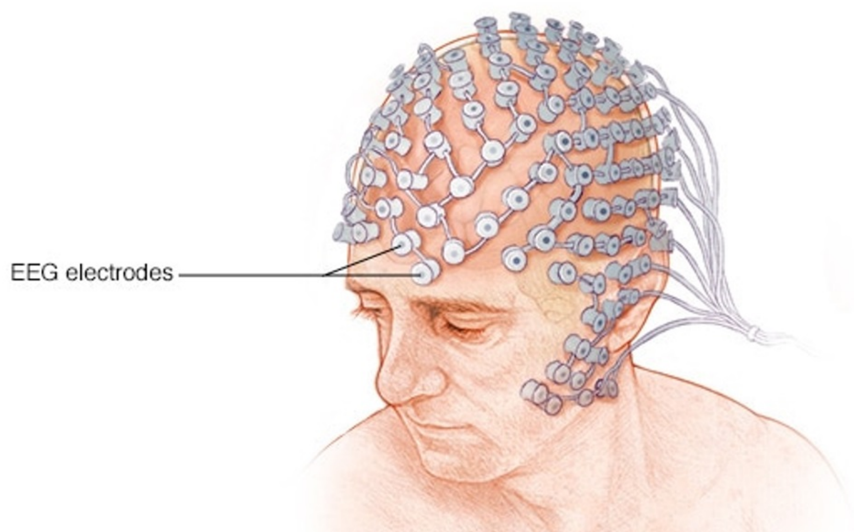



**Figure 1.3**: EEG Head Cap

After the patient undergoes an EEG examination, a specialist neurologist prepares the session report. These reports generally include examination session information, medical history, and notes on any normal or abnormal activity that occurred during the session. EEG examination can be performed for everyone, any age, and it is safe and anyone with a little experience can do it because it does not require difficult steps. The specialist examines multiple states which are sleep, and wakefulness and asks the patients to do some mental activity or physical activity. The signals measured by EEG for these neurons can always be a mixture of several underlying base frequencies. Each signal is decomposed into several frequency bands [33]. These bands differ depending on several factors such as the state of the examination, age, and emotional state.

### 1.3.1 SIGNAL'S FREQUENCY BANDS

According to the EEG, the signals recorded from the scalp are decomposed into several frequency bands [33]. These bands hold the information about the brain state and can tell what we cannot see and what the behavior hides. Frequency bands differ when performing the examination depending on several factors such as age, the person's state when taking the scan. These bands provide information and details for the characteristics that appear in the EEG scan [36]. Hence, the clinicians can make their decision about the record state if it is normal or abnormal and determine the brain disorder. The frequency bands can be summarized as follows:

I. Delta: It occurs during sleep and is most pronounced in the right part of the brain and it is between 0.5-4 Hz in frequency, Figure 1.4 represents the Delta wave.

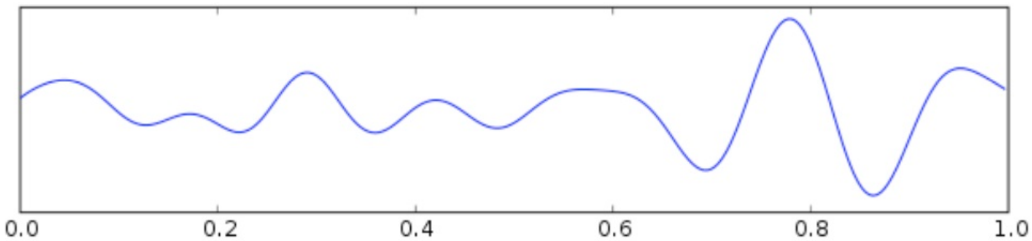



**Figure 1.4**: Delta Wave

II. Theta: They appear with the brain processes underlying mental workload or working memory, they can be recorded from all parts of the cortex, and it is between 4-8 Hz in frequency. Figure 1.5 represents the Theta wave.

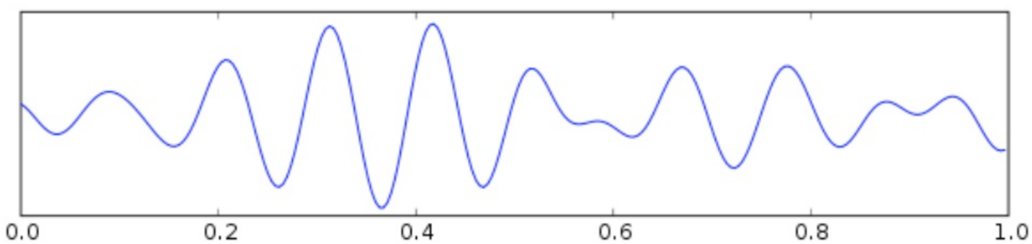



**Figure 1.5**: Theta Wave

III. Alpha: alpha appears and its domain strength while you relax mentally, eyes closed, walking, reading, and practicing yoga and it is between 8 - 13 Hz in frequency. Figure 1.6 represents the Alpha wave.

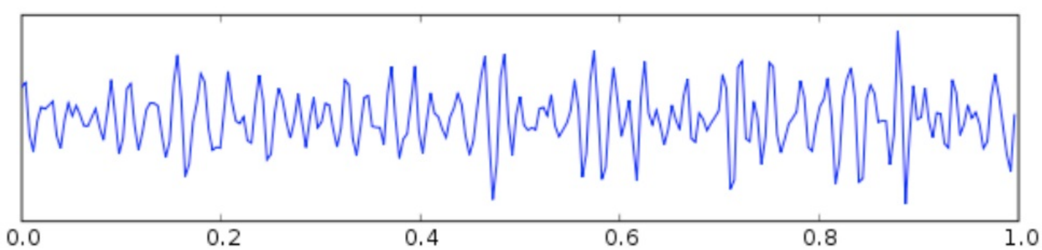



**Figure 1.6:** Alpha Wave

IV. Beta: It occurs due to acts of thinking or anxiety and concentration, and this hesitation appears in both the front and back areas and it is between 13 - 30 Hz in frequency. Figure 1.7 represents the Beta wave.

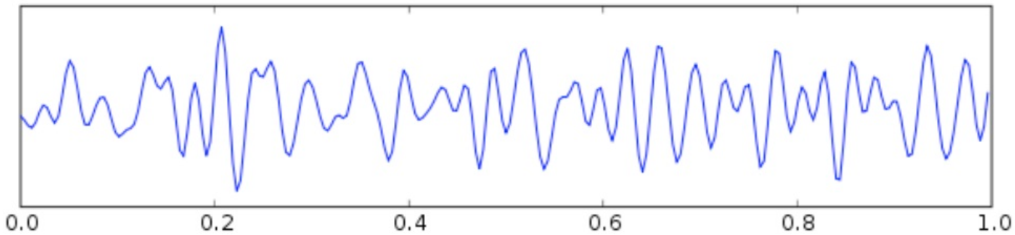



**Figure 1.7**: Beta Wave

V. Gamma: It appears in the somatosensory cortex. The fastest brain waves are related to the decision-making of the brain, and it is larger than 30 Hz in frequency. Figure 1.8 represents the Gamma wave.

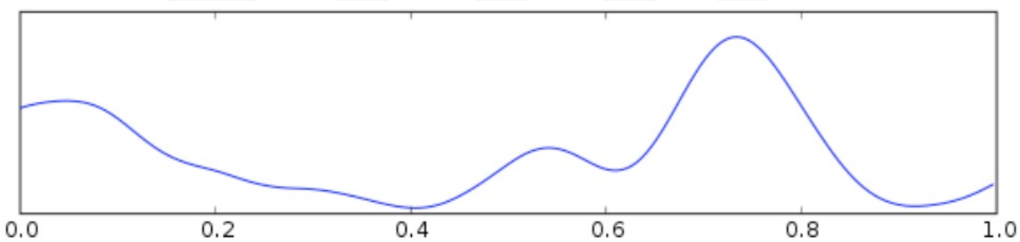



**Figure 1.8**: Gamma wave

## 1.4 BRAIN DISORDERS

The brain is subjected to several disturbances some of which reasons are known and some of which are unknown where that leads to brain abnormal activity. Some of the known causes, for instance, are due to a genetic condition, accident, or certain diseases affecting the brain. Among the diseases that affect the brain and cause an imbalance in its electrical activity are epilepsy, tumors, brain death, sleep disorders, and other diseases that affect the brain [2-4, 13, 34].

## 1.5 EPILEPSY

Epilepsy is one of the ancient diseases that was known for a long time that affects the brain, as it disrupts the electrical activity of the central nervous system [37, 38]. The first physician who realized that Epilepsy was a disease that affects the brain and tried to treat it was The Greek physician Hippocrates, where it was called back then in the Greek language Epilepsies. This disorder affects the behavior and causes an abnormal action and is accompanied in some cases by loss of consciousness. Epilepsy can affect all races of all ages, and it has no specific time, as it can be occurring at different times of life. Epilepsy does not depend on specific circumstances or events, as it can be hereditary or due to an accident or due to one of the brain diseases that cause brain damage, infectious diseases. For instance, meningitis or prenatal injury, and developmental disorders such as autism and neurofibromatosis. Epilepsy affects around 1% of the people in the world, which is around 65 million people [39]. The places that had seen the Highest infection rate, were the poor countries that do not provide medical services.

Epileptic seizures differ in symptoms, duration, and location of occurrence from one person to another, as well it can differ from time to time for the same person. Clinicians divide epileptic seizures into two main types according to where the seizures occur in the brain: one is called Focal (partial) seizures and the second is called Generalized Non-Specific seizures [37, 38, 40].

The brain electrophysiological is changing during a seizure and for each one comes next [41, 42]. Before the seizure which is the pre-ictal phase the communication of neurons becomes weak, then the infected neurons are surrounded by a circle [43].

### 1.5.1 FOCAL SEIZURES

These seizures are also called partial seizures due to their occurrence in one part of the brain and they are segregated into two types the first seizures without loss of consciousness and the second seizures with impaired cognition [44, 45].

I. Focal seizures without loss of consciousness: these seizures named partial simple seizures, they may lead to Involuntary twitching of one part of the body or spontaneous sensory symptoms but without loss of consciousness [37, 38, 45].
II. Focal seizures with impaired awareness: these Seizures with impaired cognition, called Complex partial seizures: These seizures are accompanied by loss of consciousness or a behavior change. During a seizure, you cannot communicate or

respond to the reality surrounding you, and you may make some movements repeatedly, such as chewing, moving hands, or swallowing [37, 38, 45].

### 1.5.2 GENERALIZED NON-SPECIFIC SEIZURES

These seizures occur all over the brain and it subdivided into six kinds of subcategories according to occurs time, symptoms, age, and seizure long interval which are described and explained as following [37, 38, 46] :

I. Absence seizure: Earlier called minor epileptic and it happens in children, especially children with Lennox-Gastaut Syndrome (LGS). It comes accompanied by eyes blinking, smacking the lips, or a brief loss of consciousness. It may be difficult to know when the seizure times end, and the time for the next seizure begins. The reason is due to the level of alertness and activity that may not improve between epileptic eruptions that occur during long periods while awake and often during sleep. Figure 1.9 represents an example of an Absence seizure taken from an EEG device.

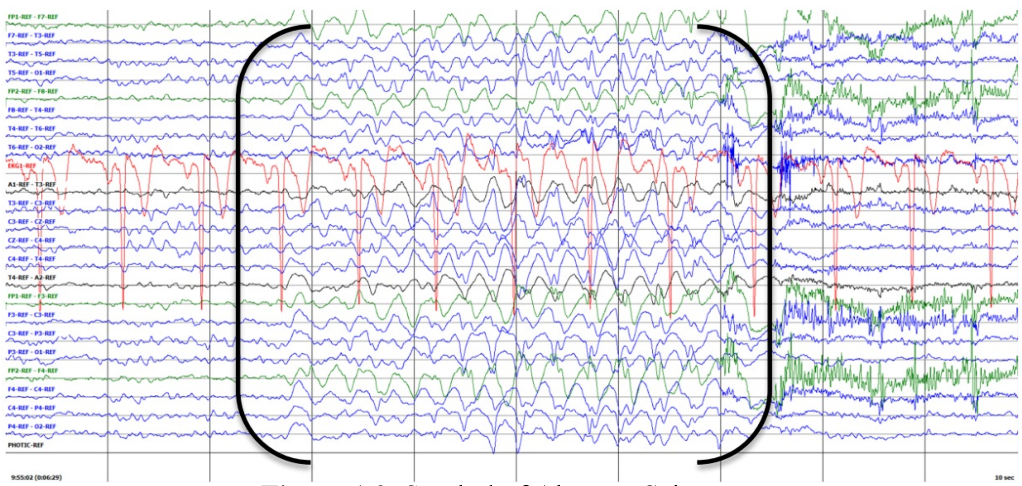


**Figure 1.9**: Symbol of Absence Seizure.

II. Tonic seizure: These seizures result in body stiffness and last from a few seconds to a minute and usually occur during sleep, but if it occurs during the waking period, it will cause the person to fall.

III. Atonic seizure: Epileptic seizures, falling fits, also called "drop attacks," are seizures that lead to a sudden loss of control of the body's muscles, which leads to falls and injury to the head or face. These seizures are brief, usually lasting a few seconds. Figure 1.10 represents an example of an atonic seizure taken from an EEG device.

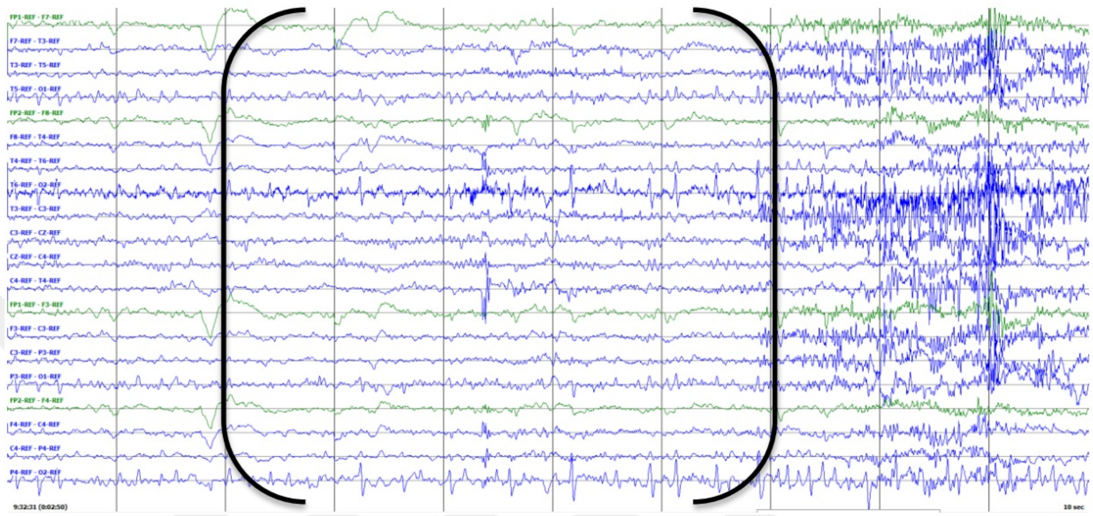


**Figure 1.10**: Symbol of Atonic Seizure.

IV. Clonic seizures: These seizures appear to affect the face, neck, and arms, and they are characterized by repetitive rhythmic and jerky muscles. Figure 13 represents its symbol. Also, it seems that it is usually this seizure that comes along with Tonic-Clonic seizures as shown in Figure 1.11.

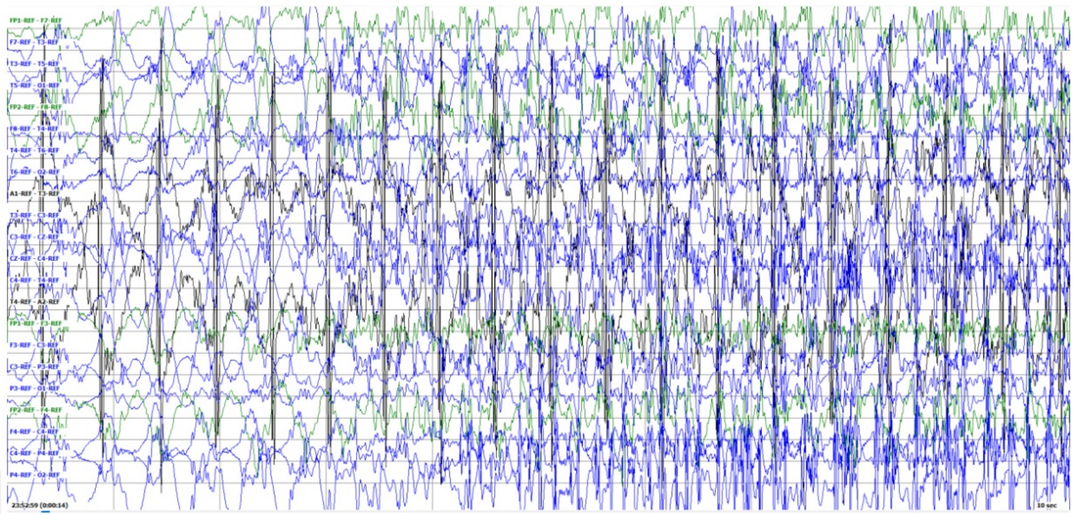


**Figure 1.11**: Clonic seizures combining with Atonic Seizure.

V. Myoclonic seizures: These seizures which are shown in Figure 1.12, cause sudden spasms in the legs and arms and are for short periods.

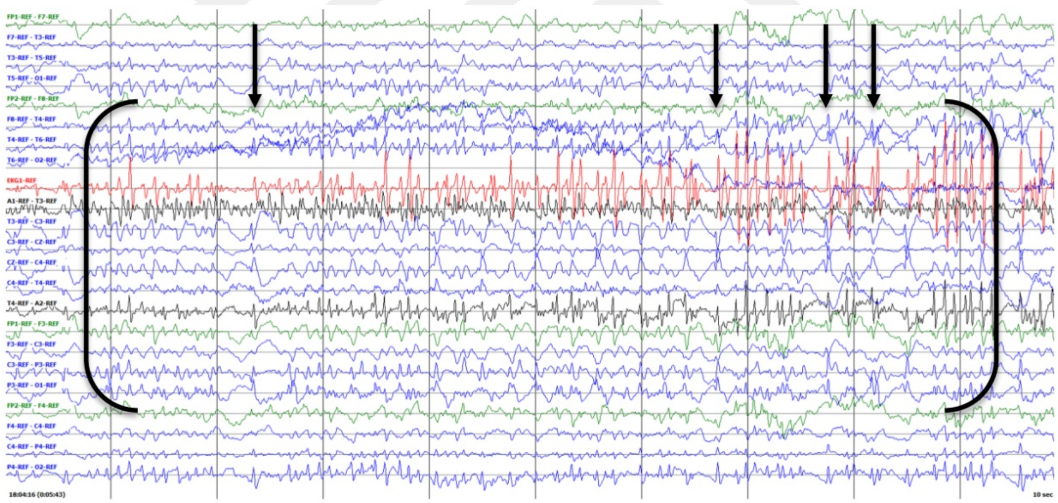


**Figure 1.12**: Myoclonic Seizures

VI. Tonic-Clonic seizures: as shown in Figure 1.13, these were formerly known as major epileptic seizures. They cause a sudden loss of consciousness and often some problem with controlling the bladder or tongue.

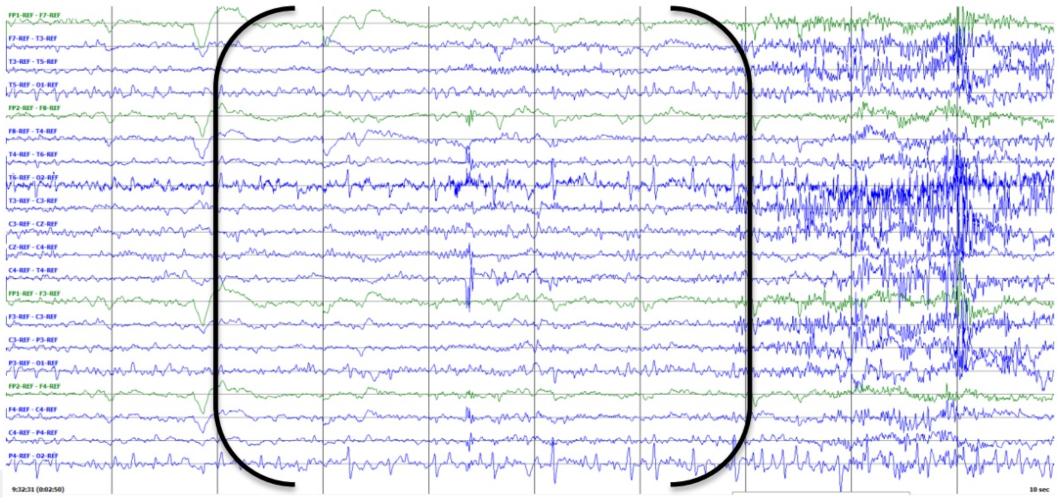


**Figure 1.13**: Tonic-Clonic Seizure

There are places where seizures occur more frequently among children, and likewise, there are places where seizures occur more frequently in adults. In children, seizures are more frequent in the temporal and frontal lobes, but in adults, seizures recur in the medial or middle part of the temporal lobe. It cannot be known how many seizures a patient will suffer from in general, where some can suffer from hundreds of times per day and some who are lucky once every few years [47]. Some of the patients may need Surgical treatment options and these groups represent 8%, and some may not need any kind of treatment and these groups represent 33% Where 25% which is the rest it cannot be controlled [48].

## 1.6 DATABASE RESOURCE

There are a lot of sources in this field, which contain many databases that have been released to help scientists and researchers to make progress in this area. Many open resources are available to serve the science in this field such as Temple University Hospital (TUH), Bonn University [49] and the Children's Hospital Boston (CHB-MIT), and Freiburg University [50], etc.

Temple University Hospital Seizure Corpus (TUH) is considered one of the most extensive open-source databases [2, 4]. where it includes a lot of information such as time of seizure occurrence, seizure types, the seizure position in the brain, seizure strength, and time interval between each seizure. Therefore, it will be much easier for Experienced researchers in analyzing data to deal with that kind of dataset. In addition, the EEG report data for the sessions is written by a board-certified neurologist. These reports contain the patient's information, age, gender, and medical history and include two areas, impression, and clinical correlation, which are observed by the neurologist's findings [34]. The TUH EEG database generally segregates into sessions. EEG data signals collected by the neurologist are stored in a file with six formats which are[51, 52]:

I. edf: represent European Data Format (EDF), which contains EEG sampled data.
II. txt: represent the text file, which contains the session and patient information.
III. tse: term-based comments by using multiple classes.
IV. tse_bi: term-based comments using two classes.
V. lbl: event-based annotations using multiple classes.
VI. lbl_bi: event-based annotations two classes.

Where the main objective of the above file categories is to describe the comments basics and to save the files that hold the report information. In addition, The TUH EEG deals with the EEG data using twenty-seven labels and they release it with the data reports. The EDF file includes the important session information for the patients and is divided over 24 fields that display them. The section below explores a brief description of the types of epileptic seizures the TUH includes in their database [37, 38].

A- TUH Database advantages:

I. Contain an accurate characterization of clinical conditions.
II. The report is written by a specialist neurologist.
III. Providing a big variety in terms of patients and seizures.

B- TUH Database disadvantages:

A. Despite the great specificities for some architectures, the sensitivities still had no good rate.

## 1.7 EEG SIGNAL PROCESSING

EEG Signal Processing Consist of several phases as shown in Figure 1.14. At First, the raw EEG signal is typically going through the filtering process and artifact elimination before features are extracted. Later, it turns into two phases which are the training phase and the testing phase.

Where the classifier learns the data features in the training phase, and it is predicting which class the data features belong to in the testing phase.

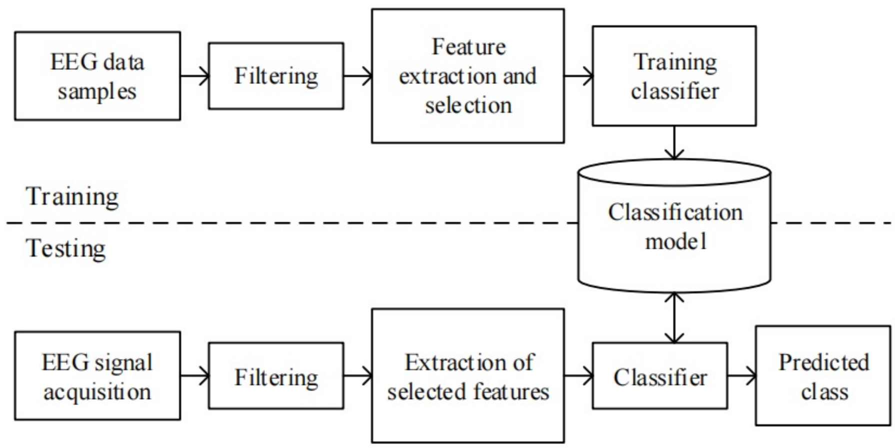



**Figure 1.14**: EEG Signal Processing Phases

### 1.7.1 FEATURE EXTRACTION

Feature extraction is a method used to make input data more relevant to the task at hand in whatever domain such as frequency domain, time domain, or time-frequency domain [53, 54]. Feature extraction is considered one of the various data pre-processing techniques. For extract features, the regular feature extraction methods include two main subdivisions: syntactic descriptions and statistical characteristics. These techniques reduce the artifacts of the data, thus improving and maximizing the accuracy and efficiency of the classification algorithms [55-57]. Indeed, there are a lot of patterns of feature extraction that have been used to normalize data. Each one has a unique mechanism for extracting the feature from the signal such as Entropy, approximate entropy, sample entropy, fuzzy entropy, wavelet transform, time-frequency distributions, and Fast Fourier Transform [6]. Moreover, each technique has an advantage in improving the data properties such as visualization, accuracy, and overfitting risk. There are a lot of artifact resources that affect the EEG signals such as eye movement, heart activity, and even from the EEG itself [58-60].

## 1.8 CLASSIFICATION

Classification is a common task in machine learning to analyze and information uptake data. Classification creates models called classifiers that perform the process of analyzing data and assign labels to them. Classification's real goal is to split data which in turn gives us a bigger ability to understand and evaluate the data. The signal classification phase can be done in multiple ways, such as linear analysis, nonlinear analysis, neural networks, fuzzy techniques, clustering, and adaptive algorithms.

Classification falls within the supervised learning family where machine learning contains many classification algorithms that have been proposed and developed by researchers to deal with the quantity and quality of input data with the advancement and development of science and life.

However, it is not possible to define a specific algorithm to solve a specific problem. Where the accuracy is determining the classifier performance for the data that is evaluated.

## 1.9 SIGNIFICANCE OF THE STUDY

During the past era, machine learning took the lead in many health fields, included the development of medical research, handling of patient data, the process of disease discovery, and giving a hand in defining treatment. In addition, reducing time and giving accurate results made machine learning an integral part of medical applications and research. The manual adjustment process is still present, but its accuracy is up to 50% compared to the automatic operation [39]. Artificial intelligence technologies have served science amazingly and are still developing day after day to help humans in various fields through scientific applications. In this study, we focused on the application of machine learning, due to the remarkable success it received in scientific, medical, and even commercial applications. In this study, we deal with one of the most famous and prominent brain disorders that are still being researched and analyzed, which is epilepsy. Thus, the primary goal of this work was to present techniques that could classify different types of epileptic seizures with the lowest error rate, i.e., with the highest possible accuracy.

## 1.10 THESIS OVERVIEW

**Chapter 1** starts with a definition of EEG devices and their relationships with a lot of diseases including epilepsy. This chapter also includes a brief for brain structure and brain signal types according to their location. As well as a basic introduction of the classification techniques used in this thesis which are SVM and KNN classifiers. Finally, the significance of the study and the outline.

**Chapter two** introduces a preview of the brain-computer interface and its techniques that serve science and help improve the fields including medical, engineering, and other fields. Moreover, we explored some studies that had similarities with our work and goals. Where They used the same data set or the same pre-processing methods, or the same systems technologies are applied.

**Chapter 3** aims to extract features from the EEG signals that were acquired Clinically. Presented here is an effective way to analyze and extract the parameters for the classification of the EEG signals for achieving our goal. we explore the methodology that we applied by the two classification techniques for supervised machine learning algorithms. Where K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are considered among the highest and most efficient algorithms used in classification problems.

**In Chapter 4**, we discuss the results we got from applying our classifier combined with our feature extraction method on our data and compare our performance with others. Finally, Chapter Five provides the conclusion of our work.

**Finally, chapter 5** provides the conclusion achieved from the work presented in the previous chapters.

# 2. LITERATURE REVIEW

## 2.1 HUMAN-COMPUTER INTERACTION (HCI)

Human-computer interaction (HCI) is considered one of the most important modern fields and the most active and developed in recent times, due to its intervention in many different human fields [61-65]. HCI aims to build and produce a design that provides services between people and computers to modernize and improve the nature of life in the best possible way [66]. Human-computer interface applications have gained rapid and massive popularity due to their association with human actions, reactions, and feelings [67, 68]. Brain-computer interface (BCI), a subfield of HCI, is of great importance as it relates to the most important and mysterious part of the human body, the brain. BCI is a direct bridge that connects the brain with the computer to evaluate, understand and repair cognitive and sensory-motor functions of the body. This section will summarize multiple areas of the EEG.

Moreover, this section discusses some studies that included our field by exploring their database source, preprocessing techniques, and classification algorithms they used. Where every approach has its merit and demerit that we should be aware of it to achieve as much as possible the best results. Hence, the classification accuracy usually depends on the type of data and the methods used to extract the features where the main goal of these techniques is to decrease the error rate as much as possible.

## 2.2 NORMAL EEG

The human brain is a nonlinear complicated system therefore, EEG normal or abnormal reports cannot be reliable from the normative databases. Where these databases use linear analysis, and it may lead to serious errors. A normal EEG for an adult in a waking state consists of stable electrical activity and a low amplitude rate. Although, the EEG state differs from one person to another according to age, gender, and mental state. Where in children it has the same frequency range, but the amplitude is greater and its spread in more spots.

Noteworthy, brain determinism is translated by the ability of the brain to invent or the ability to learn. Brain signals collected by EEG have three familiars' qualities, which are Non-stationarity, Non-linear, and Noisy [14, 69, 70]. In addition, the information that EEG acquires from the brain is evaluated by the frequency waves. Therefore, according to that diagnosis, it will be more appropriate to use nonlinear techniques with brain signals although the linear techniques proved their efficiency in that field. [2, 71, 72]. Clinicians interpret the EEG signals in two phases, first, they analyze the EEG signal background then identify them [73]. The features that observe in the EEG signals it is what defines the normal and abnormal decision about the seizure. The analyzing phase for the signal background included many steps, where it should observe frequency, voltage, wave shape, regulation, position, responsiveness, and interhemispheric coherence [69, 74, 75]. Hence, there is no normal EEG, just comparing the scan at different times.

## 2.3 SEQUENTIAL DATA

EEG signals are the result of the brain physiological process such as speech, but it reveals itself in special circumstances. For example, Figure 2.1 illustrates the spectrogram for an EEG seizure in one channel.

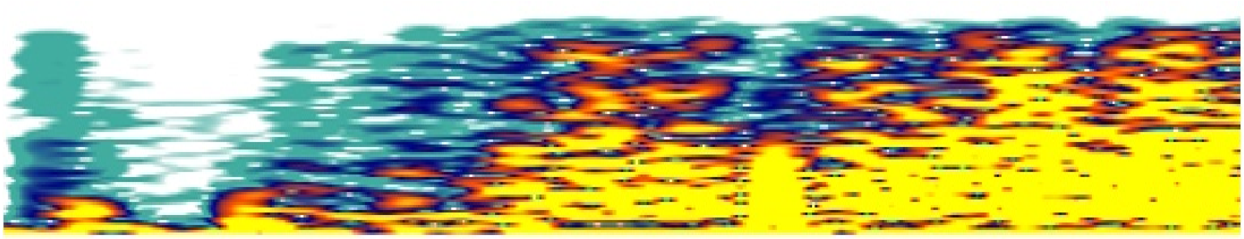


**Figure 2.1**: The Spectrogram of The Signal

According to that, Bishop [76], pointed out that the approaches of machine learning that deal with the independent data would not use the sequential nature of the data. Moreover, the frequency spectrum of the EEG refers to HMMs as promising methods to classify abnormal signals.

## 2.4 EPILEPTIC SEIZURES

Clinical neurology is one of the areas that deal with the diagnosis of epileptic patients, sleep disorders, and other brain disorders. Sudden electrical activity in the brain cells is one of the main reasons for epileptic seizures [77]. The electrical disorders cannot be observed by eyes that is why many researchers have studied epileptic seizures to discover and identify the changes during the seizures. Therefore, researchers studied the mechanisms of the brain and tried to develop algorithms that fit with epilepsy. Due to becoming the most prevalent Nervous disease worldwide and becoming one of the important era issues [78]. In turn, numerous studies deal with EEG since it took a wide and direct place in Clinical neurology. Epileptic seizures classification had opened the door to be more interesting to get involved with because it may lead to predicting seizures [79]. Researchers in their studies used several types of techniques trying to achieve the highest accuracy of their work.

One of these studies J.C. Rajapakse et al. [80] they discussed the intensity of the seizures on the brain signals by using several different ICA applications. They illustrated the ability of epileptic seizure detection and the ability to use it in clinical research. Also, Niederhoffer et al. [81] where they used their research's raw data information to develop the algorithms used in detecting the onset of epileptic seizures. Subasi et al. [4], used three features which were Auto-Regressive, Fast Fourier transform, and Maximum Likelihood Estimation (MLE). Although, these features present information about the frequency and the energy. However, these features cannot provide the seizure's start time which led to an inability to detect and analyze the data. Some researchers' studies deal with the epilepsy topic as two phases which are normal and ictal, while others deal as three phases that are normal, interictal, and ictal.

Seizure detection is the next step after confirming that the seizure that happened is because of epilepsy as well trying may lead to predicting the start [82]. Figure 2.2 shows the three epilepsy stages.

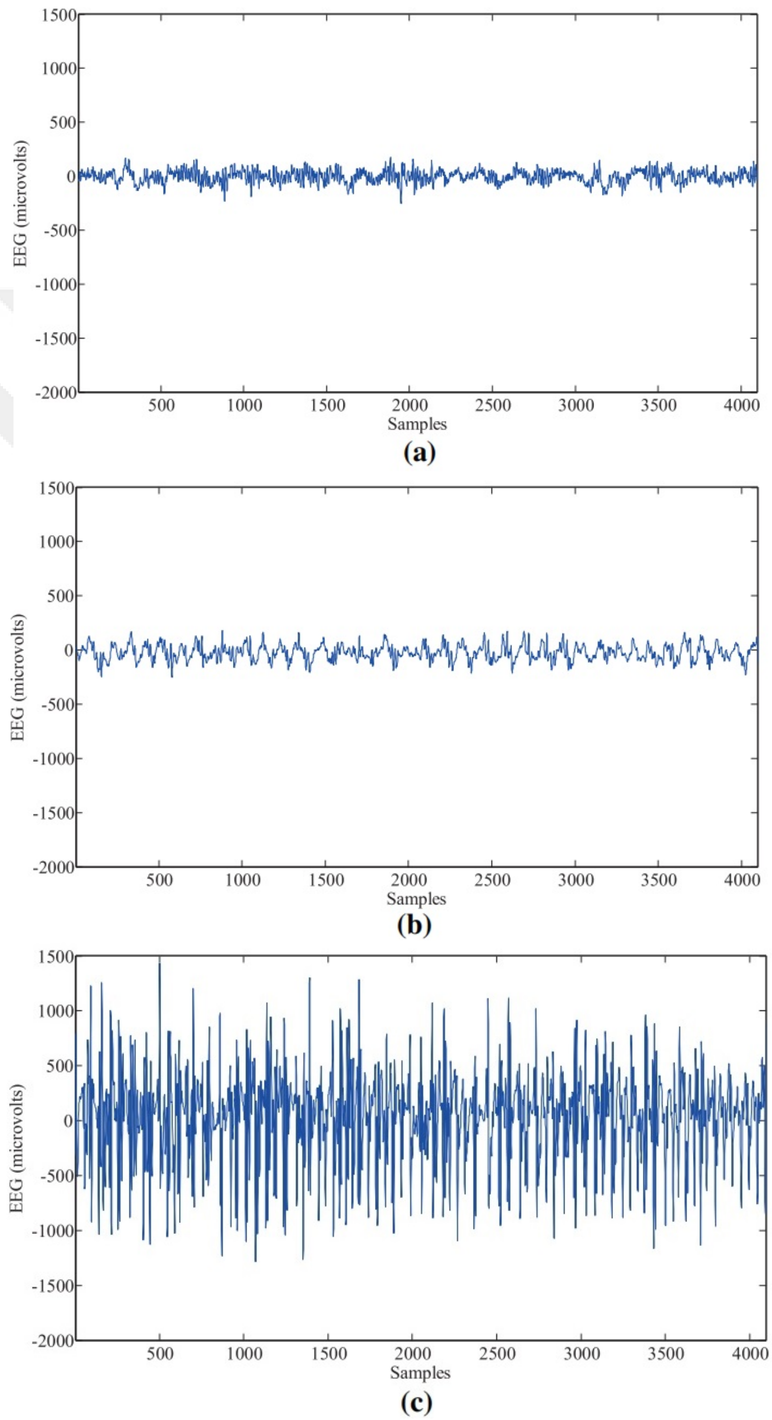



**Figure 2.2**: EEG signals Phases (a) Normal, (b) Interictal and (c) Ictal.

## 2.5 ARTIFACTS

The brain generates signals for all mental states, normal or abnormal (illness), which are recorded by the EEG. In addition to brain signals, the EEG records other signals from different regions non-cerebral that are useless called artifacts [83]. These artifacts lead to inaccurate decisions and are more difficult in the diagnosis because it is confused with the neural signals that distinguish existing defects [84]. Extracting and analyzing EEG signals is not easy as long as there will always be artifacts. Artifacts are caused by various reasons such as involuntary patient movement, electrode problems, sweating, or some electromagnetic source [59]. The dataset we used here which is the TUH contains some artifacts that are considered the biggest source of error in machine learning as described below.

I. Muscle Artifact: These artifacts are related to the patient's muscle movement and are characterized by the signal high frequency as shown in Figure 2.3. It is associated with tonic-Clonic seizures and cannot recognize their pattern easily.

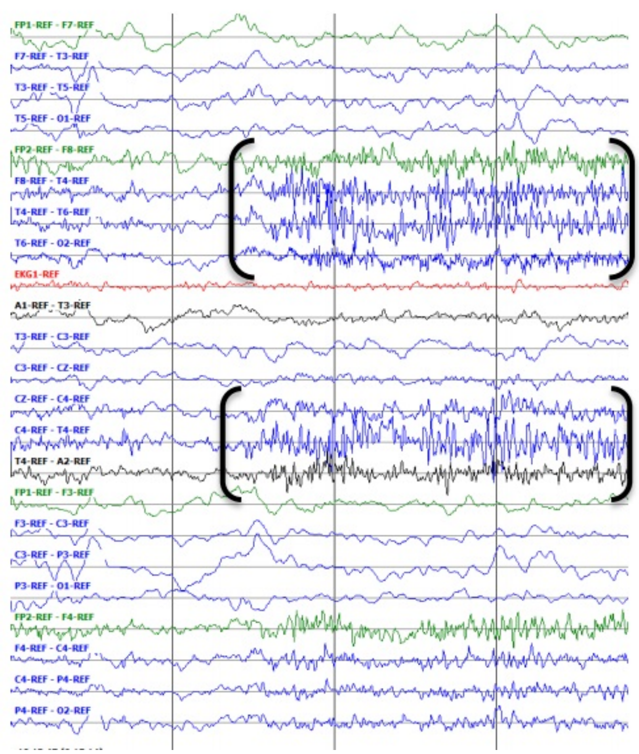



**Figure 2.3**: A part of Muscle Artifact

II. Shivering Artifact: This is a subcategory of the muscle artifact that is very rare to happen and appears mostly in the beta band frequency as shown in Figure 2.4 below.

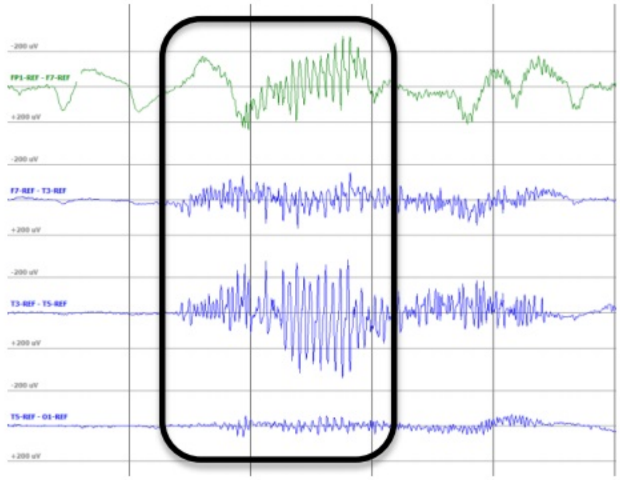


**Figure 2.4**: Part of Shivering Artifact

III. Chewing Artifact: This is a subcategory of the activity of the jaw muscles and mainly appears on the temporal channels. This artifact has a high-frequency rate with small periods between the movement of the jaw as shown in Figure 2.5.

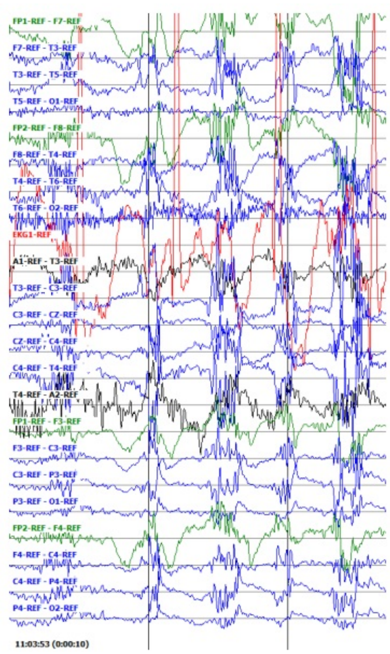


**Figure 2.5**: Chewing Artifact

IV. Eye Blink artifact: this artifact happens because of the frontal electrodes that are near the eyes and appears as a single high amplitude sharp wave then comes a slow wave oppositely. Eye artifacts may be associated with neurological dysfunction as shown in Figure 2.6.

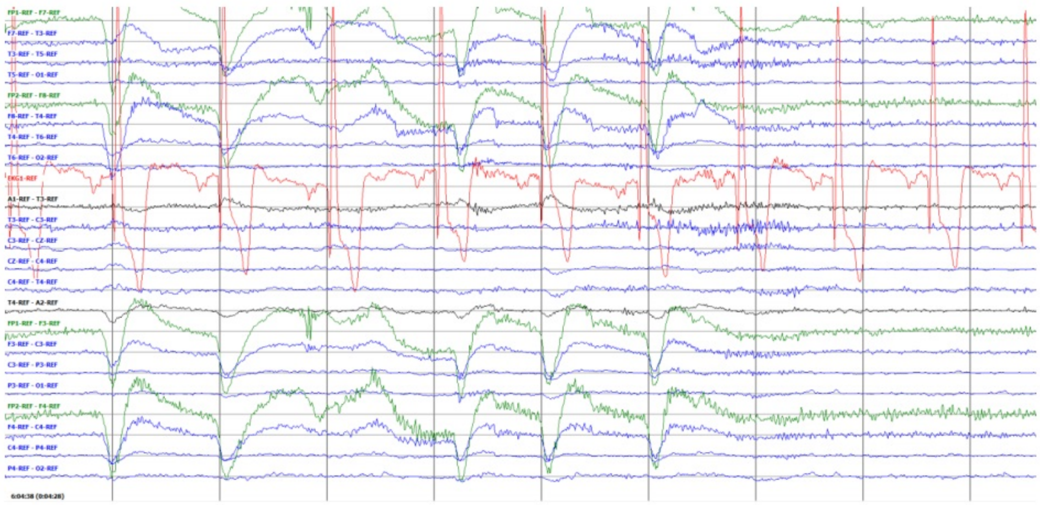


**Figure 2.6**: Eye Artifact may be associated with Neurological Dysfunction**.**

V. Electrode Artifact: these artifacts happen because of various electrode problems such as some bubbles in the gel that used to stick the electrode on the scalp or some fault in the electrode. Where usually the electrodes are exposed to damage and that affects the seizure extraction hardly. Figure 2.7 shows two-electrode pops.

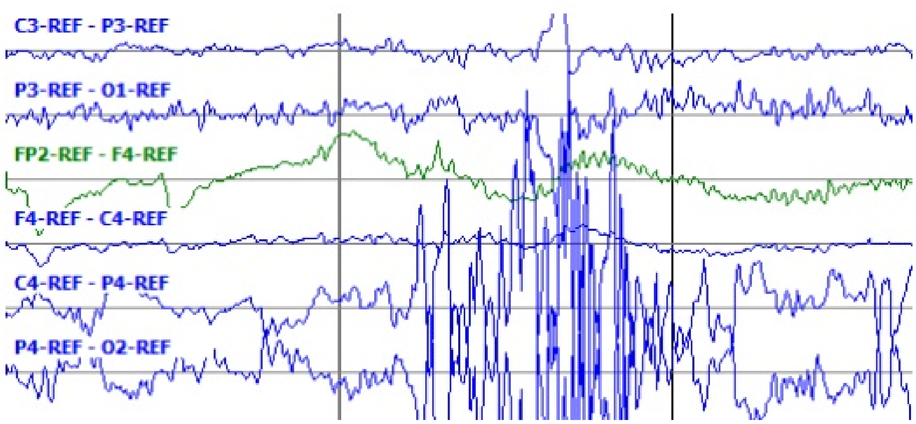



**Figure 2.7**: Two Electrodes Artifact

Thus, it will make our job more difficult to distinguish between brain waves and noise [2]. Various research considered removing the artifacts as an essential matter in preprocessing technique for EEG raw data [83, 85, 86]. Hence, with the progress and development of science in this area, EEG signals have been analyzed using many ways. There are four main methods that researchers use to analyze EEG data (time, frequency, time-frequency domain, and nonlinear methods).

## 2.6 FEATURES EXTRACTION AND SELECTION

Feature selection considers an effective method to analyze the signals [87, 88]. It is a way to minimize the dimensions by choosing a subset of features from the original inputs. The researchers studied the optimal choosing ways that provide the best feature selected, where it categorized into two models: the wrapper model [89], and the filter model [90, 91]. These models determine the efficiency of the feature subset by learning algorithms. After all, the features are evaluated and can use just the relevant one [92]. One of the methods that were used in the literature was in the EEG feature field was Fourier transform [93]. Also, Lei Yu et al. [91], was one of these studies that discussed the optimal way to select features. They came with a new concept of predominant correlation and suggested a fast filter way to determine the features and redundancy between them without pairwise correlation analysis.

## 2.7 EPILEPTIC SEIZURE DETECTION

Detection and classification of seizures became a very popular area recently, where it may be the way of paving to seizure prediction [79]. The researchers studied the neuron and the transporters component to analyze, understand, and detect the main reason behind the seizures. Many research mentions that spike-wave and sharp wave is the most common wave that can be detected [94-96]. Although, during epileptic seizures, these waves are barely appearing due to the patients sleeping most of the time. Therefore, and due to the multiformity of seizure shape, many methods had been used to detect it by the EEG signals automatically [97, 98]. Seizures may not be due to epilepsy, therefore besides neurological tests, the diagnosis required some other tests such as blood and cerebrospinal fluid. Also, may use the CT to make sure there are no abnormalities in the brain. Hojjat Adeli et al. [99], used wavelet transforms to analyze the data of epileptic seizures. Where they are used with patients who suffer from the absence of seizures Wavelet transform to analyze and describe epileptiform discharges. Niina P¨aivinen et al. [100], studied seizure detection by mathematical methods, where a group of features had been calculated through small time windows. Three features used are time domain, frequency domain, and nonlinear. They showed when making a combination of features will present the best accuracy for linear or nonlinear data. In addition, Osorio et al. [77], studied the probability for seizure detection through the intensity of the seizures, and later they used the direct stimulation of the brain by developing a plan and manipulating the real-time [101].

Various studies show that the seizure detection technique has various requirements due to the different dataset's characteristics and circumstances. One of the techniques to detect seizures is Pattern recognition where these techniques focus on enhancing the accuracy of the classification with the help of various feature extraction and classification methods [102]. Florian Mormann et al. [103], studied the statistical moment of the EEG amplitudes to detect the pre-ictal stage and the features were Hjorth parameters. For predicting seizure onset Wim van Drongelen et al. [104], used the power and P.E. Mc Sharry et al. [105], used signal variance which are both a type of linear features. Semih Altunay et al. [106], observed the spikes and sharp waves in the EEG recording data using a linear prediction filter, and according to the results, it is considered an effective way to detect seizures.

Z. Roshan Zamir [107], used another linear feature to detect seizures which is the relative fluctuation index and the Linear least squares as preprocessing models.

## 2.8 EPILEPTIC SEIZURES CLASSIFICATION

Earlier, EEG waveforms were described using descriptive ways. Hence, through the development in this field, the EEG signals had been analyzed in different ways. Pradhan et al. [108], analyzed the EEG signals where they trained Neural networks using two layers which are called Learning Vector Quantization (LVQ). The first layer was learning to classify the inputs and the second layer was transforming the result from the first layer to the goal identified by the user. Neurologists have been assisted by several different proposals to detect and classify epileptic seizures. Many classification methods are used generally but in the literature Support Vector Machines [109, 110], and Artificial Neural Networks were the most used [111]. Petrosian et al. [112], applied two techniques which are recurrent neural networks (RNN) and signal wavelet decomposition trying to predict the onset of epileptic seizures.

Forrest Sheng Bao et al. [113], used Probabilistic Neural Network (PNN) to classify epileptic patients. While Ling Guo et al. [82] used more than one feature, they used KNN as a classifier and wavelet transform, line length as a feature [114]. Also, they classify the EEGs with three-layer MLPNN by calculating the ApEn feature and by using wavelet energy-based features. A proposal has been made by Iscan et al. [115], to use time and frequency features with an SVM classifier. While Lima et al. [116] achieved the highest accuracy in classification by using SVM with wavelet transform, Wang et al. [117], used the K-Nearest neighbor ( KNN) classifier and wavelet packet entropy, and Orhan et al. [118], used DWT and Artificial neural networks. Antti Saastamoinen et al. [119], detect epileptic seizures by using a radial basis function (RBF). Varun Bajaj [120], used an SVM classifier with the EMD method to classify the EEG signals. Also, Ashkan Yazdani et al. [121], used the Dempster-Shafer theory of KNN in classifying EEG signals. The used data were five classes according to different mental tasks they are signing with. They used wavelet decomposition and autoregressive (AR) to extract the EEG signal features. Their result was higher than the classical voting KNN classifier and the distance weighted KNN classifier. In addition, KQ Weinberger et al. [122], and Belongie et al. [123], showed that when KNN was combined with different techniques the system achieved amazing results. KNN classifies the goal data by computing the distances between her and its neighbors, while some other studies such as Chopra et al. [124], measure the similarity which leads to an increase in the performance of the classification system. Thao Nguyen Thieu et al. [125], suggested a method to detect epileptic seizures. Where they used data from 8 patients who have temporal and extratemporal lobe epilepsy at frequency sampling of 500 Hz and 16 channels. Then the data had been segmentation into several segments according to the record time. For extracting features, they used FFT, and for the detection techniques, they used an artificial neural network (ANN), Naïve Bayes, and Logistic

Regression. The most effective results were ANN+FFT where they achieved a rate 89.8% of accuracy. Here we preview some articles that share the data, features, and classification techniques. Subhrajit Roy et al. [39], classified the seizure types and aimed to distinguish between normal and abnormal seizures. They used the TUH EEG dataset, and a combination of the FFT algorithm as a feature extraction method with several classifiers' techniques SGD, KNN, XGBoost, and CNN for the classification process. They applied two pre-processing methods on the EEG signals and these methods were described as efficient in analyzing the data [126, 127]. Both methods applied the feature extraction which is called Fast Fourier Transform (FFT) on every second of the data channel with specific seconds of overlap. After that, in the first method in the range 11 fmax Hz they took log10() of the magnitudes of frequencies. Thus, the training sample data dimension became (N,47) and N represents the EEG channel number. In the second method, they normalized some data that clipped which are from 1 to fmax Hz through frequency buckets. Then they calculated the real eigenvalues. KNN, SGD, and XGBoost used HyperOpt as a method to choose the best hyperparameters. The famous Resnet50 method was chosen to be used with CNN. They selected KNN and SGD classifiers to apply, due to considering them as the fastest classifiers to explore the design space effectively [128, 129]. KNN achieved outstanding performance compared to SGD with 90% accuracy while SGD was 56%.

Also, Saputroet et al. [130], aimed to classify the seizure phases which were called pre-ictal, ictal, inter-ictal, and postictal, analyzing differences in characteristics in each phase. They used 120 signals from Temple University Hospital Seizure Corpus that were recorded with a sampling rate of 250 Hz and 256 Hz, and the 10-20 System of Electrode Placement. Then they applied a combination of three feature extraction where MFCC uses several steps to extract the features and then generate 14 coefficient features. Also, Hjorth Descriptor is working in the time domain and with three parameters are Activity, mobility, and complexity. ICA is known to split the blind source and split the mixed-signal under specific terms. In the classification technique, they used the Support Vector Machine (SVM) as the classifier. They classified Generalized Non-Specific Seizure (GNSZ), Focal Non-Specific Seizure (FNSZ), and Tonic-Clonic Seizure (TCSZ) and a normal EEG signal. They used the linear SVM kernels, quadratic, cubic, fine, medium, and coarse Gaussian. In their research, they did five scenarios: first, they used Hjorth's Descriptor with his three parameters. Where the second, they used MFCC which produced 14 features. In the third one, they used a combination of MFCC and Hjorth Descriptor which led to extract 17 features. While the fourth one, they used a combination of ICA and Hjorth Descriptor and produced four features. And the last scenario was by using the combination of MFCC, Hjorth Descriptor, and ICA and produced 18 characteristics. The highest result they achieved was the combination of MFCC+ Hjorth, the parameters that evaluated and enabled the system to detect seizure type were sensitivity, specificity, and accuracy average and the result was 90.25%, 97.83%, and 91.4% respectively.

In the same scenario, K. Polat et al. [131], used the dataset of the UCI machine learning repository  They used three features which were (time domain, frequency domain, and chaotic ), then they used three methods to normalize namely minimum-maximum, MAD and z-score. And for classification techniques, they used Support Vector Machine (SVM) with the three models of kernels which are (Linear, Cubic, and Medium Gaussian) SVM. Their aim was the detection and classification of epileptic seizures from EEG signals where they classify five healthy cases, an epileptic seizure, eyes closed, eyes open, and tumor region. After applied, the classification techniques with and without the normalization method the best result that they got was when they used Cubic SVM as a classification technique with MAD normalization methods which got an accuracy of 82.50%. Also, Manish Sharma et al. [38], was aiming to help clinicians to design and build an automatic system for disease detection and provide them with clinical data. Thus, they created as possible an accurate system that improved the results got by using the manual seizures classification work. They used Temple University Hospital Seizure Corpus (TUSZ) dataset and three feature extraction methods FuEn, LSN, and FrD features. Moreover, they used for the feature’s classification the 10-fold cross-validation. When the original signals are decomposed into 6 sub-bands the features are extracted. thus, the classifier takes these features as input and divides them into 10 parts, nine for training and one for validation. where this procedure is repeated for all validation parts and the system performance will be considered as the ten-folds average. As a classification technique, they used the Support Vector Machine (SVM) for classifying normal and abnormal seizures. The combination of FuEn, LSN, and FrD features with SVM got the best accuracy of 79.34%.

# 3. METHODOLOGY

## 3.1 INTRODUCTION

The brain generates electrical signals to represent the body in each state even if it was a normal or abnormal state. EEG is considered the gold tool in neurology due to its ability to reflect brain activity even if the signals were small and weak. Thus, for getting the most benefit from EEG characteristics we should choose the right data, preprocessing techniques, and analyzing methods that serve our goal.

In this section, we discussed the signal collecting method by EEG, the channel's distribution on the scalp, the system used to distribute the channels, the data size for each class we use, and the frequency rate for the data recorded. Then we explained the data preparation strategies for the collected data to convert the data from the original domain to a suitable domain that can fit the classification used technique. Where it is very common that the original brain signal contains an unwanted signal, which causes a misdiagnosis, they are called artifacts. Pre-processing techniques implemented on the data are increasing the data quality by making the data free as much as possible from the artifacts.

Later, we discussed the features extraction methods we used that present us with the most suitable data that fit with our model and can achieve our goal with the highest accuracy rate. Finally, the last phase of our work was using machine learning algorithms through two simple and powerful classifiers which are SVM and KNN. Figure 3.1 below illustrates the diagram of the EEG classification techniques approach for this work.

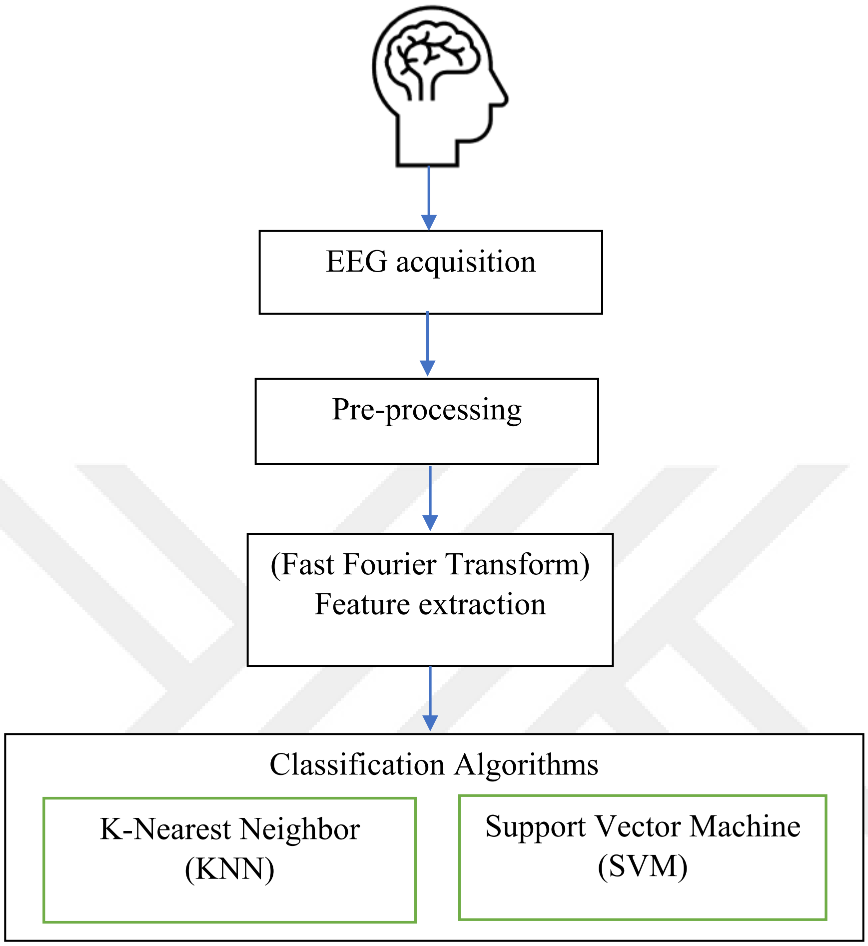



**Figure 3.1**: EEG Classification Approach

## 3.2 SIGNAL RECORDING (ACQUISITION)

EEG is a non-invasive tool where the brain signals are picked up and recorded through the metal pieces attached to the scalp (electrodes). Generally, EEG signals are recorded through several electrodes about 8 to 64 which enable the data to become kind of high-dimensional. These metals are put in specific places and do not touch the scalp but there is a gel or paste in between. The electrodes play a main role in the EEG capturing process [132]. Recording an accurate signal needs to beware of some instructions such as the scalp clean and the contact impedance. Where the contact impedance between the scalp and the electrode should be from 1kΩ to 10kΩ because otherwise, it will affect the signals collected. In this study, we identified three types of seizures namely complex partial seizures (CPSZ), generalized non-specific seizures (GNSZ), and tonic seizures (TNSZ). The datasets were 864 signals consisting of 149 CPSZ, 660 GNSZ, and 55 TCSZ with 16 channels and 250 HZ samples per second. The selected scalp EEG signals were recorded according to the International 10/20 system and EEG channels were distributed on the scalp and can be divided into two main groups which are odd can be on the left side and the even can be on the right side as shown in Table 3.1.

**Table 3.1:** The Channels Labels in a 10/20 System as Represented in TUEG

| Channel Label | Position on Scalp | Channel Label | Position on Scalp |
|---|---|---|---|
| Fp1 | left forehead | Fp2 | right forehead |
| T5 | left outer posterior | T6 | right outer posterior |
| F7 | left outer frontal | F8 | right outer frontal |
| F3 | left inner frontal | F4 | right inner frontal |
| P3 | left inner posterior | P4 | right inner posterior |
| C3 | left inner medial | C4 | right inner medial |
| T3 | left outer medial | T4 | right outer medial |
| O1 | left back of the head | O2 | right back of the head |

### 3.3 SIGNAL PRE-PROCESSING

Representing the EEG data is one of the many classification problems and an important step in the BCI system. In general EEG signals have two important information which is time and location. The features used may discard the signal patterns that occur through the electrodes or through time. Hence, signal preprocessing is an important step in the data mining process to reduce noise in the EEG signal. Also, pre-processing improves data efficiency and thus guarantees a good result. Rapid development in the size of databases, diversity, and their fetching from different sources affects the quality of the data and thus reduces the quality of the results.

There are various data pre-processing techniques; One of these is to reduce information (feature reduction). Where this technique reduces massive data features to a small number, by producing new value features from the initial ones [128]. These processes led to an increase in the processing of the learning model Automation, thus improving the accuracy and efficiency of the algorithm. One of the pre-processing advantages is that you can combine more than one technology to work together to achieve maximum efficiency and accuracy. Every technique has an advantage for reducing risks such as overfitting. Using feature extraction extra advantages appear like (accuracy improvements, overfitting risk reduction, speed up in training, improved data visualization, and increase in explain the ability of our model). Indeed, there are a lot of patterns of feature extraction such as Entropy, approximate entropy, sample entropy, fuzzy entropy, wavelet transform, time-frequency distributions, and Fast Fourier Transform (FFT) [6], where each one has a unique mechanism for extract the feature from the signal.

### 3.4 FAST FOURIER TRANSFORM (FFT)

EEG success in the analysis is attached to the quality of the information they extracted from the signals. There are many methods to extract features according to the data type if it was in the time domain or frequency domain. Despite EEG signals being considered as non-linear data, some scientists deal with it as linear data. Fast Fourier Transform is the feature extraction method used in this study to deal with our data. FFT was suggested by the scientists C. Cooley and John Tookey in 1965 [133]. It is considered one of the most popular methods compared to other techniques, this algorithm aims to speed up finding a discrete Fourier transform. FFT transforms the EEG signal from the time domain to the frequency domain and the output of this process is called a spectrum [134, 135]. This method uses mathematical tools for EEG data analysis, the mechanism of action and the idea are based on the sequencing division. Where four of the frequencies of the bands are included in the major characteristic waveforms of the EEG spectrum [4]. Also, most research and studies show that features extracted in the frequency domain are the best to identify mental tasks based on EEG signals [136]. Figure 3.2 shows the mathematical equation for the Fourier transforms.

$$s(f) = \int_{-\infty}^{\infty} s(t)\, e^{-j2\pi f t} dt$$

**Figure 3.2**: Fast Fourier Transforms Equation

| | |
|---|---|
| Where: | |
| s(f) | = frequency domain signal |
| s(t) | = time domain signal |
| $e^{-j2\pi f t}$ | = Constant |
| f | = frequency |
| t | = time |

The data obtained from FFT are collected from multiple areas of the scalp, therefore the data acquired multiple features. Due to these features set varying degrees of size, scope, and units we had to use normalization. Below we FFT explored advantages and disadvantages.

A- Advantages:

I. Work perfectly with a stationary signal.
II. More compatible with the narrowband signal, such as sine wave.
III. In most real-time applications it has an enhanced speed.
IV. For the analysis method, use the frequency domain.
V. Fits with both types of signals Narrowband and stationary signals.

B- Disadvantages:

I. Suffer from noise and it does not keep a data record with a short duration.
II. Has not the ability to discover the localized spikes and complex seizures in EEG signals for epileptic seizures.
III. Vulnerability in analyzing nonstationary signals.
IV. Not appropriate for short signals and have not a good spectral estimation.

## 3.5 NORMALIZATION

The most feature scaling techniques used in machine learning are normalization and standardization. Normalization is considered an important step in data preparation. Due to using FFT, the signals must normalize which enables us to bring all the data to the same range to be ready for classification. There are many transformation methods used for data normalization such as Z-score, T-score, MinMax, Logistic, LogNormal, and TanH. The most common ways to normalize data in linear transformation include z-score or t-score, this is usually called standardization. Feature scaling, have values between 0 and 1 which enable us to compare data more easily and the results become inputs for the next step which is classification and that will improve and enhance our accuracy result. Moreover, normalization for data applying in one method at a time, and some algorithms have their data normalization or scaling. Figure 3.3 below shows two types of EEG signals: normal EEG signals and normalized EEG signals.

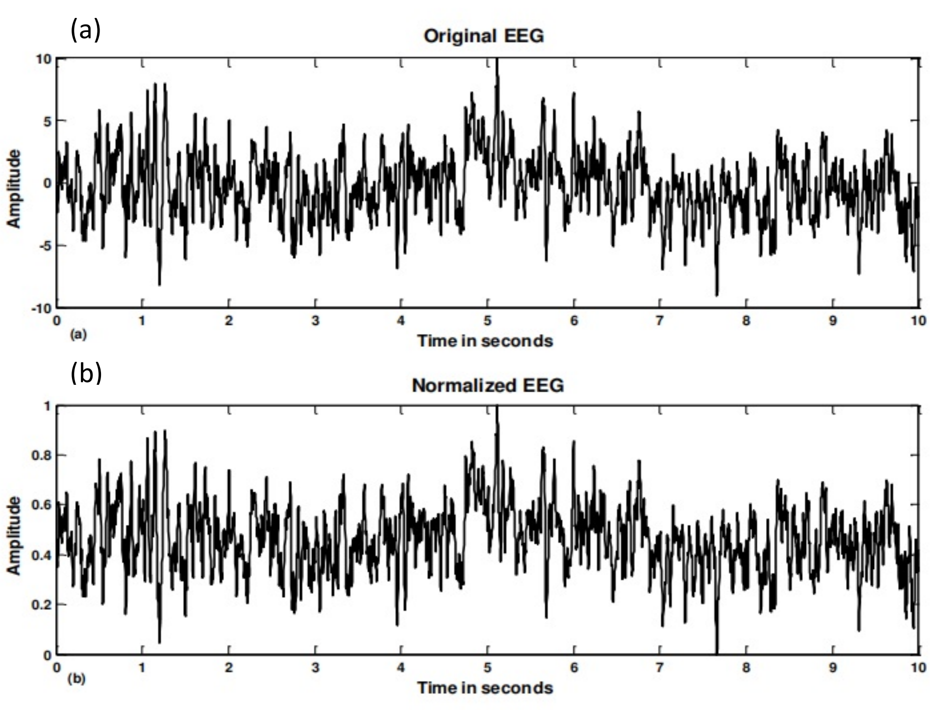



**Figure 3.3**: (a) Normal EEG Signals (b) Normalized EEG Signals

## 3.6 SUPPORT VECTOR MACHINE (SVM)

SVMs are a very sophisticated idea, but it still has a simple implementation. Moreover, it is considered a tool that every civilized person should have in his tool bag. SVM is one of the most important algorithms used in solving various problems which fall under machine learning such as regression, clustering, handwriting recognition, and classification technique [137, 138]. SVM considers one of the linear classification fields and Falls within the supervised machine learning algorithms and is based on what is known as the statistical learning concept [139]. SVM is one of the fastest and most efficient algorithms used with big data and one of the most important prediction methods. The Scientists who developed the SVMs were Vladimir Vapnik and his colleagues in 1963 at AT&T Bell Laboratories [139]. The main idea of SVM is all about decision boundaries, how to choose the best line that separates between the points. where the best line is the farthest line from the points as possible as shown in Figure 3.4.

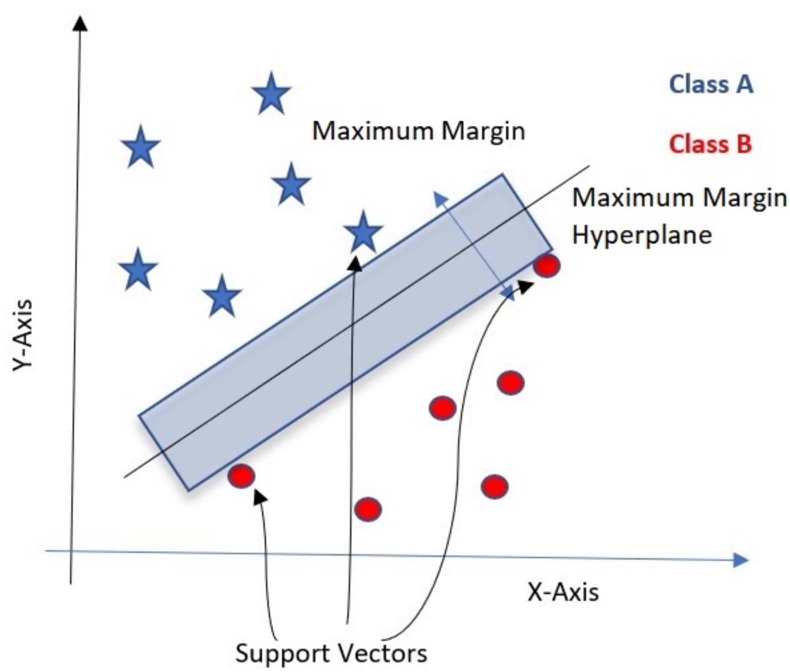



**Figure 3.4**: Support Vector Machine (SVM) Classifier

SVM starts with data in relatively low dimensions and transforms the data into a higher dimension and at the end finds a line that splits the higher dimensional data into two categories.

Where that line is called support vector classifier due to the observation on the edge and within the soft margin are called support vectors. Where the margin is the shorter distance between the observation and the threshold. Moreover, whenever the margin is bigger The generalization error of the classifier will decrease [140].

In addition to linear classification, SVM works perfectly with data that is non-linear which is a high-dimensional space through a special trick called kernel [141]. The kernel mapped the data into higher dimensional space which is supposedly making the separation process much easier. kernel function only calculates the relationships between every pair of points as if they are in the higher dimensions, they do not do the transformation. Moreover, the kernel calculation reduces the amount of computation required for the support vector machine. Three requirements are to control the effectiveness of SVM leading to reducing error and overfitting. These basic requirements are Soft Margin parameter C, Kernel selection, and Kernel parameters.

In concluding, we explored here the advantages and disadvantages of the SVM algorithm :

A- Advantages :

I. Works perfectly for both linear and non-linear data.

II. Use kernel trick for complex data.

III. Soft margin constant C can handle the Outliers well.

IV. Using a convex optimization function.

V. Hinge loss provides higher accuracy

VI. Work perfectly with small train data

B- Disadvantages:

I. Hinge loss leads to sparsity.

II. For getting an accurate accuracy.

III. Hyperparameters and kernels should be tuned.

IV. Larger data requires a longer training time.

## 3.7 K-NEAREST NEIGHBOR (KNN)

KNN classifier is one of the supervised machine learning algorithms that is considered as one of the topmost, simplest, and efficient algorithms used in classification and regression problems [142]. In 1951 Evelyn Fix and Joseph Hodges were the first men who developed this technique [19], then later Thomas Cover expanded it [20]. KNN is a versatile algorithm used in image recognition applications, handwriting detection, and finance issues, etc. KNN is simple to understand as it is based on the feature-similarity approach. KNN classifier assigns the new data to the specific class to which the data belong and depending on the K which is considered the nearest neighbors even without retraining the classifier. The nearest neighbor points for the new data are found by checking the distance between points as shown in Figure 3.5.

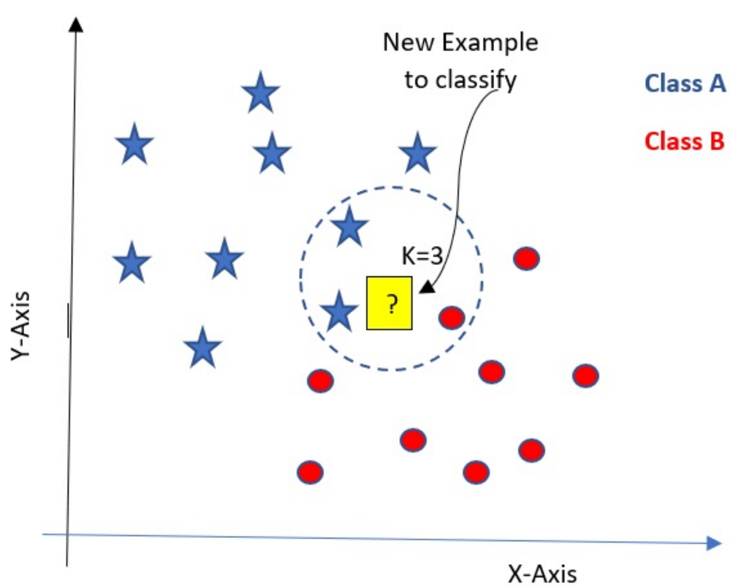



**Figure 3.5**: K-Nearest Neighbor (KNN) Classifier

These distances are measured by several methods such as Euclidean, Hamming, Manhattan, and Minkowski distance metrics. Moreover, the distance measurement gives the ability to improve the data set accuracy through the normalizing of the training data [140]. In addition, the accuracy of data can my improved when using KNN, especially when using specialized algorithms like Neighborhood components analysis or Large Margin Nearest Neighbor. One of the problems that can be encountered in KNN is that k's closest neighbors are approximately the same distance that is, KNN does not include in his calculations the existence of different dimensions. This case is handled by specifying Different degrees of importance to different closest neighbors [143]. Therefore, we will need a suitable way to measure the distances and fit our data such as Euclidean Distance. In concluding, we explored here the advantages and disadvantages of the SVM algorithm.

A- Advantages :

I. Simple and easy to understand.

II. Few hyperparameters to tune.

B- Disadvantages:

I. Choose appropriate K.

II. Big Cost calculation with large data size.

III. Choose the right scale for features.

## 3.8 DISTANCE METRICS IN MACHINE LEARNING

Machine learning has some basic parts in the fields of supervised learning and unsupervised learning. One of these parts is the Distance metrics which is used usually to determine the similarity between data points. In general, Distance metrics can enhance our machine learning model whatever if it was a classification or regression problem. The distance is measured for the points that have the similarity between them in features then determined the similarity. Generally, in Machine Learning there are four ways to calculate Distance Metrics which is Euclidean Distance, Manhattan Distance, Minkowski Distance, and Hamming Distance. Moreover, Distance Measures are chosen according to the data types the model is used.

### 3.8.1 EUCLIDEAN DISTANCE

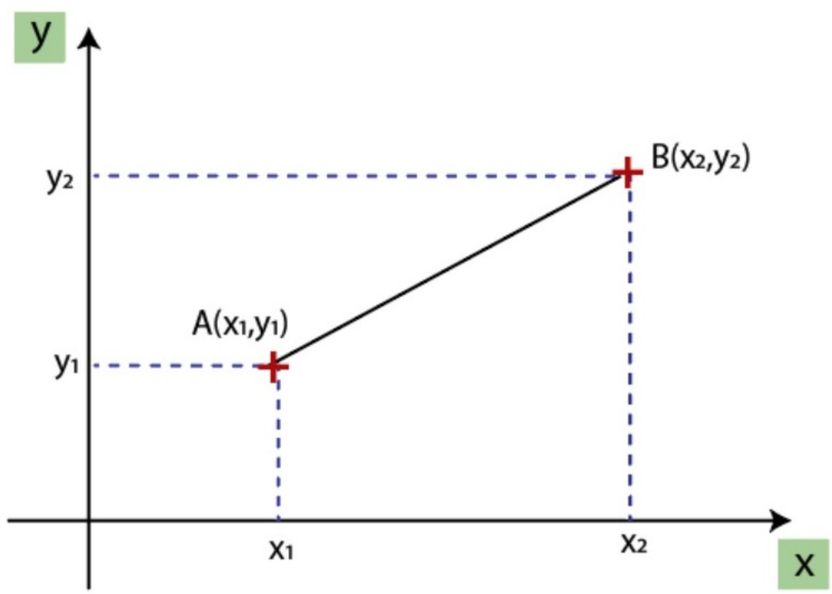
Euclidean Distance can be represented in simple words which is the closest distance between pairs of samples. It is the most common way to measure the distance scale in machine learning. Euclidean Distance has no problem applying with EEG data types and it is more frequent to use with the KNN classifier [144-148]. The distance between the pair is usually measured in an n-dimensional feature space according to the following equation in Figure 3.6 and Figure 3.7 represent the Euclidean distance graph.

$$D_e = \left( \sum_{i=1}^{n} (p_i - q_i)^2 \right)^{1/2}$$

Where,

n = number of dimensions

pi, qi = data points



# 4. EXPERIMENTAL RESULTS

## 4.1 INTRODUCTION

This study is intended to analyze EEG signals for epileptic seizures and classify them using Classification Linear models. Features were extracted by one of the common feature extraction methods used with EEG signal which is Fast Fourier Transform. Then finally, the features extracted had been classified using two classifiers SVM and KNN.

## 4.2 SOFTWARE ENVIRONMENT

There are different types of software that are used to implement classification for epilepsy such as Python, Weka, Octave, and MATLAB. In this study we implemented MATLAB with a setup run on a computer with 16 GB RAM and 2.30 GHz of an Intel(R) processor, running Windows version 10. MATLAB (R2021a) software implements the classification techniques due to the classification algorithms built-in with the Classification Linear toolbox Which is a type of model that is trained for binary classification. This trained model uses techniques that minimize the time needed in high dimensional data, which leads to reducing the function objective by fitclinear. Supervised machine learning provides the ability to detect data, choose features, select validation systems, train the models, and evaluate the results. Many supervised machine learning models allow us to get the best classification model result such as Logistic Regression, Decision Trees, Naive Bayes, Support Vector Machines, and Neural Network Classification.

## 4.3 DATASET

The data used in this work was from the largest open source for EEG data which is (TUH Groups) [149]. Where they separated EEG information and put them in multiple file formats. One of these files was the (file. EDF) where (EDF) refers to European Data Format that contains the data sample that we used. The datasets used here provide different types of seizures with their described characteristics. Choosing accurately measured input data is considered the first and an important step in the classification technique and to find an accurate model that satisfies our requirements. Next, the results we obtained with our dataset were as an outcome to five parameters for making sure of diagnosis validation. These parameters were Sensitivity, Specificity, Confusion Matrix, Classification Accuracy, Receiving Operating Characteristic (ROC) curve.

## 4.4 CONFUSION MATRIX

A confusion matrix is the next important step after preprocessing data due to it checking the performance of our data and its effectiveness. A way to evaluate your model and to explore what it is capable of. Although, it is considered as a simple concept for comparing the actual target with the predictive one but is powerful at the same time. Thus, provide us with a general view of our model performance as well as the types of error it has. Moreover, the Confusion Matrix is defined as a matrix of NxN, where N refers to the classes that want to be classified. In this study, a confusion matrix has been applied, it is considered a great tool that helps in classifying the seizures [150]. Figure 4.1 below explores the mathematical calculation for achieving that purpose through classification accuracy, sensitivity, specificity, and ROC curve [151, 152].

Predicted Values

| Actual Values | Positive | Negative | |
|---|---|---|---|
| **True** | True Positive (TP) | False Negative (FN)<br>**Type 2 Error** | **Sensitivity**<br>$\frac{TP}{(TP + FN)}$ |
| **False** | False Positive (FP)<br>**Type 1 Error** | True Negative (TN) | **Specificity**<br>$\frac{TN}{(TN + FP)}$ |
| | **Percision**<br>$\frac{TP}{(TP + FP)}$ | **Negative Predictaive Value**<br>$\frac{TN}{(TN + FN)}$ | **Accuracy**<br>$\frac{TP + TN}{(TP + TN + FP + FN)}$ |

**Figure 4.1**: Confusion Matrix Table Values

The matrix is a table with four different values for the predicted and actual data values. As shown in the figure above where the actual values are (True-False) while the predicate values are (Positive - Negative). Where these options will be handled when calculating the confusion matrix, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Hence, when the expected value is the same as the real value and it is positive it is called True Positive. As well, True Negative is when the two values are the same and they are negative. In the third and fourth options, both predictions are false, and they classify as type 1 error and type 2 error. When the actual value is negative and the model predicts are positive it is called False Positive, and When the actual value is positive, and the model predicts are negative it is called False Negative.

## 4.5 RECEIVER OPERATING CHARACTERISTIC CURVES (ROC)

ROC is a very common way used with binary classifiers in machine learning. Where ROC is a graphical plot that provides us with the information of our system performance. The system performance is considered as good performance when the sensitivity rises fast, whereas the system specificity does not increase even once until sensitivity gets a high speed. ROC illustrates all the system states, the actual values, and the predictable values where sensitivity is on the Y-axis and specificity is on the X-axis. The area under the ROC curve is called AUC which provides a measure for the general performance of the classifiers. When the AUC area is larger the classification accuracy is better. AUC takes values from (0 - 1) so, when AUC is =1 to the test data that means an optimal separating classifier has been found for our data. Also, when the AUC = 0.5 that means the classifier has no separating power.

## 4.6 K-FOLD CROSS-VALIDATION

K-Fold Cross Validation or it is called rotation estimation is considered an important tool in the data toolkit. When building a machine learning model this method will enable us to have a better use for our data. The idea behind this method is to divide the data either into two packages training and testing or into three which are training, validation, and testing [153]. The training data job is to train the machine learning model, while the validation and test data job is to verify the unseen data results. The validation dataset is almost the most common reason for using K Fold cross-validation according to its job to seek the best Parameters. There are various divisions in the cross-validation method where these parts are called Folds such as 90% -10%, 80% -20%, and can be 70% -30%. The Folds numbers are represented as K, and the unseen results data will repeat K times. The advantage of this method is that we can define the test data size and how many trials we need [154]. K-Fold Cross Validation provides the ability to use all the data for both subset training and testing. In addition, using this method can get more Metrics that provide a better and more clear idea about the data or the algorithms. Figure 4.2 shows the equation of the average error for all k data points.

$$E = \frac{1}{K} \sum_{1}^{K} E_i$$

**Figure 4.2**: The Average Error Of K-Fold Cross Validation

Where: E= MES of fold

K= number of folds

## 4.7 RESULTS

In this study, we implemented two types of classification techniques to classify epileptic seizures. Two algorithms that were performed that considered computationally fastest classifiers are KNN and SVM. The frequency data used in this study was 250 Hz with a 1-second interval with 5-fold cross-validation to reduce the overfitting risk. We classified the three seizures type according to two scenarios. In the first scenario, we combined FFT+SVM, sensitivity and specificity results were summarized in Table 4.1. As shown in the table below the last class which is TNSZ had the same value for the Specificity and Sensitivity.

**Table 4.1:** Average Sensitivity and Specificity for SVM Technique

| Classifier Model | Feature Type | Specificity / Sensitivity | CPSZ | GNSZ | TNSZ |
|---|---|---|---|---|---|
| SVM | FFT | Specificity (%) | 99 | 98 | 100 |
| | | Sensitivity (%) | 98 | 99 | 100 |

The third parameter, the accuracy of SVM model which is the first combination summarized in the table 4.2.

**Table 4.2**: Average Accuracy of The SVM Classifier

| Researchers | Objectives of study | Database Source | Classifiers Techniques | Features Methods | Accuracy (%) |
|---|---|---|---|---|---|
| Present work | Classify Epileptic seizures from EEG signals | TUH | SVM | FFT | 99.5 |

For evaluating the classifiers, we use the Confusion Matrix and the ROC as shown in Figure 4.3 and Figure 4.4.

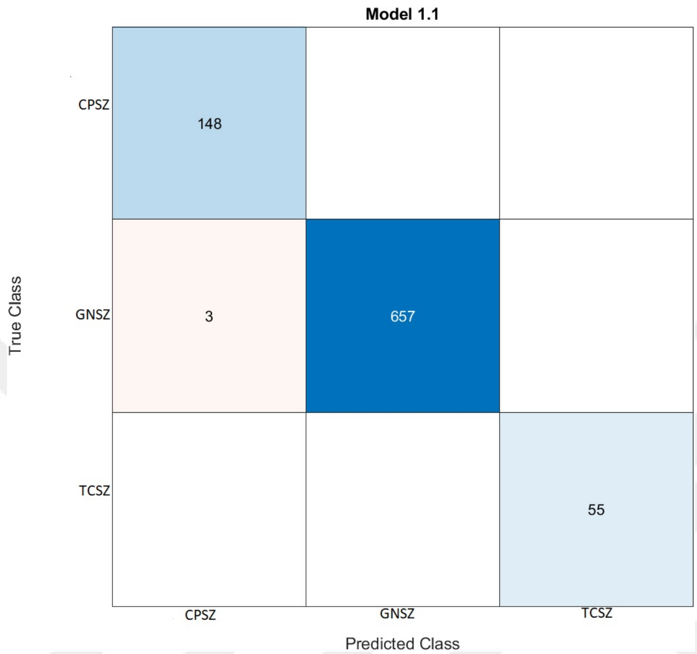



**Figure 4.3**: SVM Confusion Matrix

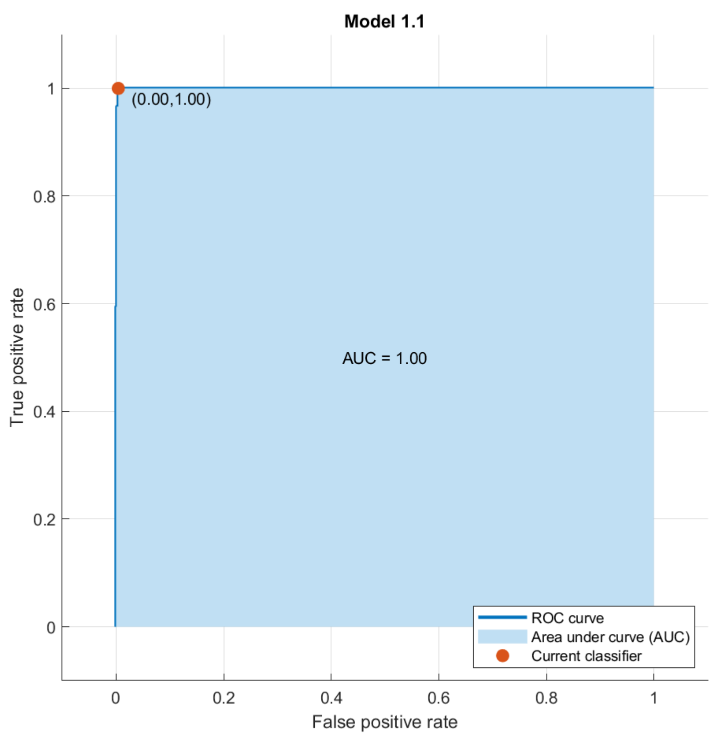



**Figure 4.4**: SVM ROC Graph

Then, in the second scenario, we combined FFT+KNN, sensitivity and specificity results were summarized in Table 4.3. As shown in the table below the last class which is TNSZ had the same value for the Specificity and Sensitivity.

**Table 4.3**: Average Sensitivity and Specificity for KNN Technique

| Classifier Model | Feature Type | Specificity / Sensitivity | CPSZ | GNSZ | TNSZ |
|---|---|---|---|---|---|
| KNN | FFT | Specificity (%) | 99 | 97 | 100 |
| | | Sensitivity (%) | 96 | 99 | 100 |

The third parameter, the accuracy of SVM model which is the first combination summarized in the table 4.4.

**Table 4.4:** Average Accuracy of KNN Model

| Researchers | Objectives of study | Database Source | Classifiers Techniques | Features Methods | Accuracy (%) |
|---|---|---|---|---|---|
| Present work | Classify Epileptic seizures from EEG signals | TUH | KNN | FFT | 99 |

For evaluating the classifiers, we use the Confusion Matrix and the ROC as shown in Figure 4.5 and Figure 4.6.

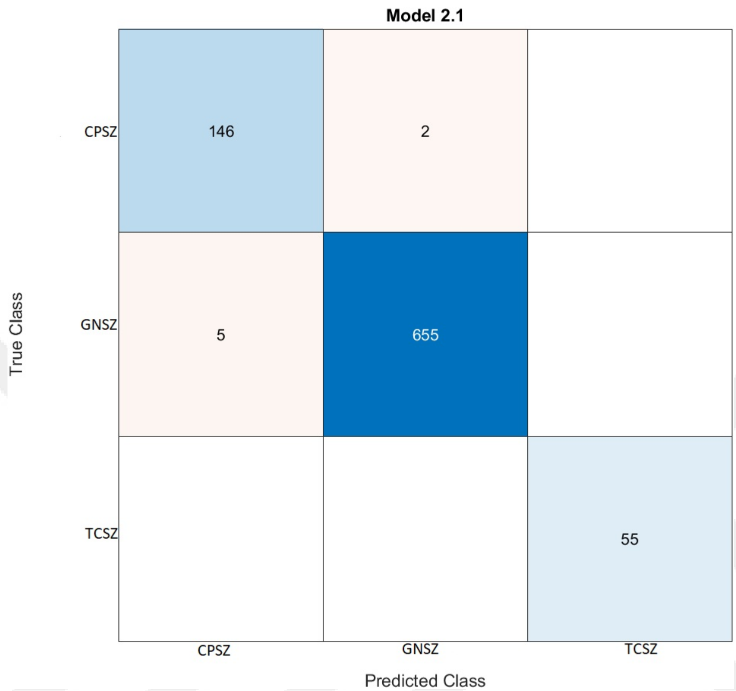



**Figure 4.5**: KNN Confusion Matrix

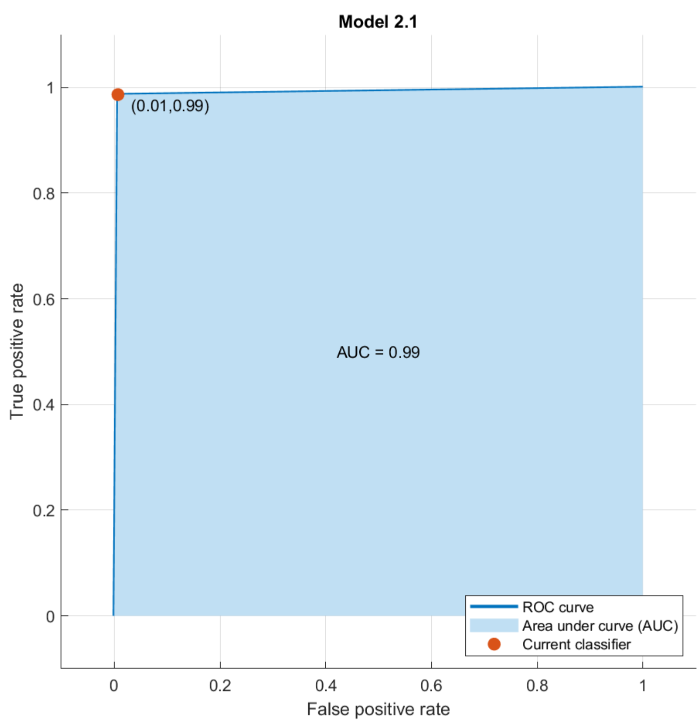



**Figure 4.6**: SVM ROC Graph

Finally, according to the scenarios above the average accuracy results achieved for both classification systems were shown in Table 4.5. SVM gets a higher result accuracy compared with the accuracy of the KNN.

**Table 4.5**: Average Accuracy of The Classification

| Researchers | Objectives of study | Database Source | Classifiers Techniques | Features Methods | Accuracy (%) |
|---|---|---|---|---|---|
| Present work | Classify Epileptic seizures from EEG signals | TUH | SVM | FFT | **99.5** |
| | | | KNN | | 99 |

From the above result, we conclude that using the FFT feature with KNN AND SVM along with the THU data seems to be one of the best ways as we far know to classify the EEG data. Moreover, using KNN and SVM classifiers is an easier and more understandable comparison with neural networks and other classifiers, which are often used for classification. Therefore, our methods will raise the accuracy ceiling of the classification process for epileptic seizure patients in the future.

# 5. CONCLUSION AND FUTURE WORK

## 5.1 CONCLUSION

Epilepsy is one of the diseases of the ages that was and still is a critical case Facing neuroscientists, but the good side, it could be treatable with medication [155, 156]. Due to it been related to the brain the feeding of science in this field will remain open in front of researchers. The researchers will keep trying to understand the way that the brain works and apply their techniques to improve the way of life for sick and healthy people. Seizures are the most obvious sign of a neurological disorder in the brain. There is a massive chance of treating and curing seizures if it is discovered early, but long seizures lead to permanent damage.

As a result, there was a need to have a system that enhances the neuroscientists' decisions which eventually improves the life quality patients have. Moreover, EEG has become an important measurement tool to record and evaluate the brain state. The abilities and characteristics of EEG helped in a significant way in diagnosing several diseases and helping to define treatments for neurological diseases. These systems assisted neurologists by enhancing their evaluation and reducing the time for their decision-making.

In this work, we presented an overview of the classification techniques through related supervised machine learning methods. Two classifier techniques were applied that provided us with the ability to build a digital diary that memorizes patient information in different situations before, after, and during seizures[157]. This work included feature examining which reveals its advantage in EEG signal representation by using Fast Fourier Transform (FFT) which converts the original EEG signals from the time domain to the frequency domain [134]. These features give important information about our signals which gives us the ability to identify normal and abnormal seizures for trying to control the situation. Which will help clinicians to detect and classify seizures to determine the suitable drug and medical care that the patients need. This thesis through two different classifiers for the classification of epilepsy seizures determined as much as possible the highest accuracy. As well as sensitivity and specificity have been calculated through the

Confusion matrix. Here, the TUH EEG Seizure Corpus database was used, which contains several different seizures, and CPSZ, GNSZ, and TNSZ seizures were selected. We used the TUH EEG Seizure Corpus (TUH-EEG) database, which contains several different seizures, where we selected CPSZ, GNSZ, and TNSZ. Although, there is no optimal feature and algorithm to find the best effective solutions these kinds of the system make some limits to the feature and algorithm which should be used to get the best data accuracy. Classification performance has been enhanced by applying the non-linear analysis method to the EEG signals. In addition, the combination of features and classifier improved the classification which led us to a great result in separating the epileptic seizures. SVM and KNN classifiers here improved the performance of the classification system and led to a more efficient analysis of the signals and their characteristics. Therefore, we chose three different epileptic seizures to be classified after using computationally efficient FFT as a feature extraction method.

In this study, the accuracy rate for the classification of epileptic seizures was 99.5% for SVM and 99% for KNN. The achieved results demonstrated the great teamwork effort provided by features and classifiers together. Where finally, selecting the features and models with the right dataset can ensure high accuracy results for our work.

## 5.2 FUTURE WORK

The work presented here we believe will give a good push to the classification techniques in the EEG signal field. We tested the models' abilities as far as possible and in different states to reach the best results. Moreover, with the great results achieved, in the future it can do some of the proposed work below.

I. Testing another dataset even maybe with another disorder.

II. Augmented classifier performance through several parameters such as having more data, dealing with missing data, being involved with the engineering of features, selecting features.

III. Using the same database but with a different frequency rate such as (250,256,400).

IV. Using different types of epochs for more than 1 sec such as (3 sec,5 sec).

V. Convert data to an image by applying the topography method (drawing EEG map for the electrode on the scalp ), then design a deep learning model and apply typography image to extract feature which will be used as input to the linear classifier.

VI. Increase the number of classes by using different types of seizures.

## REFERENCES


1. Obeid, I. and J. Picone, *The temple university hospital EEG data corpus.* Frontiers in neuroscience, 2016. **10**: p. 196.
2. Sanei, S. and J.A. Chambers, *EEG signal processing*. 2013: John Wiley & Sons.
3. Tudor, M., L. Tudor, and K.I. Tudor, *Hans Berger (1873-1941)--the history of electroencephalography.* Acta medica Croatica: casopis Hravatske akademije medicinskih znanosti, 2005. **59**(4): p. 307-313.
4. Subasi, A. and E. Ercelebi, *Classification of EEG signals using neural network and logistic regression.* Computer methods and programs in biomedicine, 2005. **78**(2): p. 87-99.
5. Kannathal, N., et al., *Entropies for detection of epilepsy in EEG.* Computer methods and programs in biomedicine, 2005. **80**(3): p. 187-194.
6. Al-Fahoum, A.S. and A.A. Al-Fraihat, *Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains.* International Scholarly Research Notices, 2014. **2014**.
7. Xing-Yu, W., et al., *Brain control: human-computer integration control based on brain-computer interface approach.* Acta Automatica Sinica, 2013. **39**(3): p. 208-221.
8. Al-Hudhud, G., *Affective command-based control system integrating brain signals in commands control systems.* Computers in Human Behavior, 2014. **30**: p. 535-541.
9. Blasco, J.S., et al., *Visual evoked potential-based brain–machine interface applications to assist disabled people.* Expert Systems with Applications, 2012. **39**(9): p. 7908-7918.
10. Cecotti, H., *Spelling with non-invasive Brain–Computer Interfaces–Current and future trends.* Journal of Physiology-Paris, 2011. **105**(1-3): p. 106-114.
11. Kotchetkov, I.S., et al., *Brain-computer interfaces: military, neurosurgical, and ethical perspective.* Neurosurgical focus, 2010. **28**(5): p. E25.

12. Smith, S., *EEG in the diagnosis, classification, and management of patients with epilepsy.* Journal of Neurology, Neurosurgery & Psychiatry, 2005. **76**(suppl 2): p. ii2-ii7.
13. Duru, A.D., *Determination of increased mental workload condition from EEG by the use of classification techniques.* International Journal of Advances in Engineering and Pure Sciences, 2019. **31**(1): p. 47-52.
14. Übeyli, E.D., *Statistics over features: EEG signals analysis.* Computers in Biology and Medicine, 2009. **39**(8): p. 733-741.
15. Agarwal, R., et al., *Automatic EEG analysis during long-term monitoring in the ICU.* Electroencephalography and clinical Neurophysiology, 1998. **107**(1): p. 44-58.
16. Meyer-Lindenberg, A., *The evolution of complexity in human brain development: an EEG study.* Electroencephalography and clinical neurophysiology, 1996. **99**(5): p. 405-411.
17. Tsolaki, A., et al., *Electroencephalogram and Alzheimer's disease: clinical and research approaches.* International journal of Alzheimer's disease, 2014. **2014**.
18. Browne, T.R. and G.L. Holmes, *Handbook of epilepsy*. 2008: Jones & Bartlett Learning.
19. Acharya, U.R., et al., *Automated EEG analysis of epilepsy: a review.* Knowledge-Based Systems, 2013. **45**: p. 147-165.
20. Medithe, J.W.C. and U.R. Nelakuditi. *Study of normal and abnormal EEG*. in *2016 3rd International conference on advanced computing and communication systems (ICACCS)*. 2016. IEEE.
21. Crawford, B., et al., *Factors influencing progress in establishing community-based marine protected areas in Indonesia.* Coastal Management, 2006. **34**(1): p. 39-64.
22. Begg, R., J. Kamruzzaman, and R. Sarker, *Neural networks in healthcare: Potential and challenges: Potential and challenges*. 2006: Igi Global.
23. Joshi, V., R.B. Pachori, and A. Vijesh, *Classification of ictal and seizure-free EEG signals using fractional linear prediction.* Biomedical Signal Processing and Control, 2014. **9**: p. 1-5.

24. Cook, I.A., et al., *Assessing the accuracy of topographic EEG mapping for determining local brain function.* Electroencephalography and clinical neurophysiology, 1998. **107**(6): p. 408-414.
25. Gotman, J., *Automatic seizure detection: improvements and evaluation.* Electroencephalography and clinical Neurophysiology, 1990. **76**(4): p. 317-324.
26. Jokeit, H. and M. Schacher, *Neuropsychological aspects of type of epilepsy and etiological factors in adults.* Epilepsy & Behavior, 2004. **5**: p. 14-20.
27. Hoffmann, E., *Brain training against stress: Theory, methods and results from an outcome study.* Stress Report, 2005. **4**(2): p. 1-24.
28. Hosseini, S.A., M. Akbarzadeh-T, and M.B. Naghibi-Sistani, *Qualitative and quantitative evaluation of EEG signals in epileptic seizure recognition.* International Journal of Intelligent Systems and Applications, 2013. **5**(6): p. 41.
29. Carlson, N.R., *Foundations of physiological psychology*. 2005: Pearson Education New Zealand.
30. Nykopp, T., *Statistical modelling issues for the adaptive brain interface.* Helsinki: Helsinki University of Technology, 2001.
31. Hendry, S., et al., *Neuropeptide-containing neurons of the cerebral cortex are also GABAergic.* Proceedings of the National Academy of Sciences, 1984. **81**(20): p. 6526-6530.
32. Merlet, I., et al., *Simplified projection of EEG dipole sources onto human brain anatomy.* Neurophysiologie Clinique/Clinical Neurophysiology, 1999. **29**(1): p. 39-52.
33. Jasper, H.H., et al., *Reticular formation of the brain.* Academic Medicine, 1958. **33**(11): p. xviii.
34. Subasi, A., *EEG signal classification using wavelet feature extraction and a mixture of expert model.* Expert Systems with Applications, 2007. **32**(4): p. 1084-1093.
35. Nunez, P.L. and R. Srinivasan, *Electric fields of the brain: the neurophysics of EEG*. 2006: Oxford University Press, USA.

36. Ebersole, J., A. Husain, and D. Nordli Jr, *Current Practice of Clinical Electroencephalography. 4th Edn Philadelphia*. 2014, PA: Wolters Kluwer.[Google Scholar].
37. Vickers, N.J., *Animal communication: when i'm calling you, will you answer too?* Current biology, 2017. **27**(14): p. R713-R715.
38. Sharma, M., S. Patel, and U.R. Acharya, *Automated detection of abnormal EEG signals using localized wavelet filter banks.* Pattern Recognition Letters, 2020. **133**: p. 188-194.
39. Roy, S., et al., *Seizure type classification using EEG signals and machine learning: Setting a benchmark.* arXiv: 1902.01012, 2019.
40. Angeles, D., *Proposal for revised clinical and electroencephalographic classification of epileptic seizures.* Epilepsia, 1981. **22**(4): p. 489-501.
41. Iasemidis, L.D., et al. *Transition to epileptic seizures: Optimization*. in *Discrete Mathematical Problems with Medical Applications*. 1999.
42. Iasemidis, L.D., et al., *Quadratic binary programming and dynamical system approach to determine the predictability of epileptic seizures.* Journal of combinatorial optimization, 2001. **5**(1): p. 9-26.
43. Petitmengin, C., M. Baulac, and V. Navarro, *Seizure anticipation: are neurophenomenological approaches able to detect preictal symptoms?* Epilepsy & Behavior, 2006. **9**(2): p. 298-306.
44. Sharma, M., et al., *An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks.* Knowledge-Based Systems, 2017. **118**: p. 217-227.
45. Beloglazov, A., J. Abawajy, and R. Buyya, *Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing.* Future generation computer systems, 2012. **28**(5): p. 755-768.
46. Sharma, M., S. Shah, and P. Achuth, *A novel approach for epilepsy detection using time–frequency localized bi-orthogonal wavelet filter.* Journal of Mechanics in Medicine and Biology, 2019. **19**(01): p. 1940007.

47. Hall, J.E. and M.E. Hall, *Guyton and Hall textbook of medical physiology e-Book*. 2020: Elsevier Health Sciences.
48. Iasemidis, L.D., *Epileptic seizure prediction and control.* IEEE Transactions on Biomedical Engineering, 2003. **50**(5): p. 549-558.
49. Andrzejak, R.G., et al., *Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state.* Physical Review E, 2001. **64**(6): p. 061907.
50. Zhou, M., et al., *Epileptic seizure detection based on EEG signals and CNN.* Frontiers in neuroinformatics, 2018. **12**: p. 95.
51. Kemp, B. and J. Olivan, *European data format 'plus'(EDF+), an EDF alike standard format for the exchange of physiological data.* Clinical neurophysiology, 2003. **114**(9): p. 1755-1761.
52. Kemp, B., et al., *A simple format for exchange of digitized polygraphic recordings.* Electroencephalography and clinical neurophysiology, 1992. **82**(5): p. 391-393.
53. Subha, D.P., et al., *EEG signal analysis: a survey.* Journal of medical systems, 2010. **34**(2): p. 195-212.
54. Cvetkovic, D., E.D. Übeyli, and I. Cosic, *Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study.* Digital signal processing, 2008. **18**(5): p. 861-874.
55. Kordylewski, H., D. Graupe, and K. Liu, *A novel large-memory neural network as an aid in medical diagnosis applications.* IEEE Transactions on Information Technology in Biomedicine, 2001. **5**(3): p. 202-209.
56. Güler, I. and E.D. Übeyli, *Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients.* Journal of neuroscience methods, 2005. **148**(2): p. 113-121.
57. Übeyli, E.D., *Wavelet/mixture of experts network structure for EEG signals classification.* Expert systems with applications, 2008. **34**(3): p. 1954-1962.
58. Nasir, A.N.K., et al., *InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019*. Vol. 632. 2020: Springer Nature.

59. Jung, T.-P., et al., *Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects.* Clinical Neurophysiology, 2000. **111**(10): p. 1745-1758.

60. Kousarrizi, M.R.N., et al. *Feature extraction and classification of EEG signals using Wavelet transform, SVM and artificial neural networks for brain computer interfaces*. in *2009 International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing*. 2009. IEEE.

61. Ramirez-Hernandez, A.C., et al. *Detection and interpretation of human walking gestures for human-robot interaction*. in *2009 Eighth Mexican International Conference on Artificial Intelligence*. 2009. IEEE.

62. Zander, T.O., et al., *Enhancing human-computer interaction with input from active and passive brain-computer interfaces*, in *Brain-computer interfaces*. 2010, Springer. p. 181-199.

63. Lin, C.-T., et al., *EEG-based drowsiness estimation for safety driving using independent component analysis.* IEEE Transactions on Circuits and Systems I: Regular Papers, 2005. **52**(12): p. 2726-2738.

64. Blankertz, B., et al., *The Berlin brain–computer interface: non-medical uses of BCI technology.* Frontiers in neuroscience, 2010. **4**: p. 198.

65. Lécuyer, A., et al., *Brain-computer interfaces, virtual reality, and videogames.* Computer, 2008. **41**(10): p. 66-72.

66. Zhang, P. and D.F. Galletta, *Human-computer interaction and management information systems: Foundations*. Vol. 5. 2006: ME Sharpe.

67. Lin, Y.-P., et al. *EEG-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine*. in *2009 IEEE international conference on acoustics, speech and signal processing*. 2009. IEEE.

68. Takeda, K., et al., *In-vehicle corpus and signal processing for driver behavior*. 2009: Springer Science & Business Media.

69. Polat, K. and S. Güneş, *Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform.* Applied Mathematics and Computation, 2007. **187**(2): p. 1017-1026.

70. Abdulhay, E., et al., *Automated diagnosis of epilepsy from EEG signals using ensemble learning approach.* Pattern Recognition Letters, 2020. **139**: p. 174-181.
71. Klonowski, W., *Everything you wanted to ask about EEG but were afraid to get the right answer.* Nonlinear biomedical physics, 2009. **3**(1): p. 1-5.
72. Klonowski, W., E. Olejarczyk, and R. Stepien, *'Epileptic seizures' in economic organism.* Physica A: Statistical Mechanics and its Applications, 2004. **342**(3-4): p. 701-707.
73. Finnigan, S. and M.J. van Putten, *EEG in ischaemic stroke: quantitative EEG can uniquely inform (sub-) acute prognoses and clinical management.* Clinical neurophysiology, 2013. **124**(1): p. 10-19.
74. Marcuse, L.V., et al., *Nonconvulsive status epilepticus in patients with brain tumors.* Seizure, 2014. **23**(7): p. 542-547.
75. Adjouadi, M., et al., *Detection of interictal spikes and artifactual data through orthogonal transformations.* Journal of clinical neurophysiology, 2005. **22**(1): p. 53-64.
76. Bishop, C., *Pattern recognition and machine learning (2nd edn., p. 738)*. 2011, New York: Springer.
77. Osorio, I., M.G. Frei, and S.B. Wilkinson, *Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset.* Epilepsia, 1998. **39**(6): p. 615-627.
78. Hill, D.R. and H.N. Stickell, *Brandon/Hill selected list of print books and journals for the small medical library.* Bulletin of the Medical Library Association, 2001. **89**(2): p. 131.
79. Niedermeyer, E., *Electroencephalography.* Basic principles, clinical applications, and related fields, 2005.
80. Rajapakse, J.C. and A. Cichocki. *Independent component analysis and beyond in brain imaging: EEG, MEG, fMRI, and PET*. in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.* 2002. IEEE.

81. Tetzlaff, R., C. Niederhofer, and P. Fischer. *Feature extraction in epilepsy using a cellular neural network based device-first results*. in *Proceedings of the 2003 International Symposium on Circuits and Systems, 2003. ISCAS'03.* 2003. IEEE.
82. Guo, L., et al., *Classification of EEG signals using relative wavelet energy and artificial neural networks*, in *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation*. 2009. p. 177-184.
83. O'Regan, S., S. Faul, and W. Marnane, *Automatic detection of EEG artefacts arising from head movements using EEG and gyroscope signals.* Medical engineering & physics, 2013. **35**(7): p. 867-874.
84. Lins, O.G., et al., *Ocular artifacts in EEG and event-related potentials I: Scalp topography.* Brain topography, 1993. **6**(1): p. 51-63.
85. He, P., G. Wilson, and C. Russell, *Removal of ocular artifacts from electro-encephalogram by adaptive filtering.* Medical and biological engineering and computing, 2004. **42**(3): p. 407-412.
86. Flink, R., et al., *Guidelines for the use of EEG methodology in the diagnosis of epilepsy: International league against epilepsy: Commission report commission on European Affairs: Subcommission on European Guidelines.* Acta Neurologica Scandinavica, 2002. **106**(1): p. 1-7.
87. Khalid, S., T. Khalil, and S. Nasreen. *A survey of feature selection and feature extraction techniques in machine learning*. in *2014 science and information conference*. 2014. IEEE.
88. Dy, J.G. and C.E. Brodley, *Feature selection for unsupervised learning.* Journal of machine learning research, 2004. **5**(Aug): p. 845-889.
89. Kohavi, R. and G.H. John, *Wrappers for feature subset selection.* Artificial intelligence, 1997. **97**(1-2): p. 273-324.
90. Liu, H. and R. Setiono. *A probabilistic approach to feature selection-a filter solution*. in *ICML*. 1996. Citeseer.
91. Liu, H., H. Motoda, and L. Yu. *Feature selection with selective sampling*. in *In Proceedings of the Nineteenth International Conference on Machine Learning*. 2002. Citeseer.

92. Kumar, V. and S. Minz, *Feature selection: a literature review.* SmartCR, 2014. **4**(3): p. 211-229.

93. Boashash, B., M. Mesbah, and P. Colitz, *Time-frequency detection of EEG abnormalities.* Time-frequency signal analysis and processing: A comprehensive reference, 2003: p. 663-670.

94. Arakawa, K., et al., *Separation of a nonstationary component from the EEG by a nonlinear digital filter.* IEEE transactions on biomedical engineering, 1986(7): p. 724-726.

95. Bishop, C.M., *Neural networks for pattern recognition*. 1995: Oxford university press.

96. Jung, T.-P., et al., *Extended ICA removes artifacts from electroencephalographic recordings.* Advances in neural information processing systems, 1998: p. 894-900.

97. Hyvärinen, A. and E. Oja, *Independent component analysis: algorithms and applications.* Neural networks, 2000. **13**(4-5): p. 411-430.

98. WU, X., H. FENG, and H. ZHOU, *Independent Component Analysis and Its Application for Preprocessing EEG.* Beijing Biomedical Engineering, 2001: p. 01.

99. Adeli, H., Z. Zhou, and N. Dadmehr, *Analysis of EEG records in an epileptic patient using wavelet transform.* Journal of neuroscience methods, 2003. **123**(1): p. 69-87.

100. Päivinen, N., et al., *Epileptic seizure detection: A nonlinear viewpoint.* Computer methods and programs in biomedicine, 2005. **79**(2): p. 151-159.

101. Osorio, I., et al., *An introduction to contingent (closed-loop) brain electrical stimulation for seizure blockage, to ultra-short-term clinical trials, and to multidimensional statistical analysis of therapeutic efficacy.* Journal of Clinical Neurophysiology, 2001. **18**(6): p. 533-544.

102. Geethanjali, P. and K. Ray, *A low-cost real-time research platform for EMG pattern recognition-based prosthetic hand.* IEEE/ASME Transactions on mechatronics, 2014. **20**(4): p. 1948-1955.

103. Mormann, F., et al., *On the predictability of epileptic seizures.* Clinical neurophysiology, 2005. **116**(3): p. 569-587.

104. van Drongelen, W., et al., *Seizure anticipation in pediatric epilepsy: use of Kolmogorov entropy.* Pediatric neurology, 2003. **29**(3): p. 207-213.

105. McSharry, P.E., L.A. Smith, and L. Tarassenko, *Comparison of predictability of epileptic seizures by a linear and a nonlinear method.* IEEE transactions on Biomedical Engineering, 2003. **50**(5): p. 628-633.

106. Altunay, S., Z. Telatar, and O. Erogul, *Epileptic EEG detection using the linear prediction error energy.* Expert Systems with Applications, 2010. **37**(8): p. 5661-5665.

107. Zamir, Z.R., *Detection of epileptic seizure in EEG signals using linear least squares preprocessing.* Computer methods and programs in biomedicine, 2016. **133**: p. 95-109.

108. Pradhan, N., P. Sadasivan, and G. Arunodaya, *Detection of seizure activity in EEG by an artificial neural network: A preliminary study.* Computers and Biomedical Research, 1996. **29**(4): p. 303-313.

109. Nicolaou, N. and J. Georgiou, *Detection of epileptic electroencephalogram based on permutation entropy and support vector machines.* Expert Systems with Applications, 2012. **39**(1): p. 202-209.

110. Zandi, A.S., et al. *A novel wavelet-based index to detect epileptic seizures using scalp EEG signals*. in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2008. IEEE.

111. Tzallas, A.T., M.G. Tsipouras, and D.I. Fotiadis. *The use of time-frequency distributions for epileptic seizure detection in EEG recordings*. in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2007. IEEE.

112. Petrosian, A., et al., *Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG.* Neurocomputing, 2000. **30**(1-4): p. 201-218.

113. Bao, F.S., D.Y.-C. Lie, and Y. Zhang. *A new approach to automated epileptic diagnosis using EEG and probabilistic neural network*. in *2008 20th IEEE International Conference on Tools with Artificial Intelligence*. 2008. IEEE.

114. Guo, L., et al., *Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks.* Journal of neuroscience methods, 2010. **191**(1): p. 101-109.
115. Iscan, Z., Z. Dokur, and T. Demiralp, *Classification of electroencephalogram signals with combined time and frequency features.* Expert Systems with Applications, 2011. **38**(8): p. 10499-10505.
116. Lima, C.A., A.L. Coelho, and M. Eisencraft, *Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study.* Computers in Biology and Medicine, 2010. **40**(8): p. 705-714.
117. Wang, D., D. Miao, and C. Xie, *Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection.* Expert Systems with Applications, 2011. **38**(11): p. 14314-14320.
118. Orhan, U., M. Hekim, and M. Ozer, *EEG signals classification using the K-means clustering and a multilayer perceptron neural network model.* Expert Systems with Applications, 2011. **38**(10): p. 13475-13481.
119. Saastamoinen, A., et al., *Waveform detection with rbf network–application to automated eeg analysis.* Neurocomputing, 1998. **20**(1-3): p. 1-13.
120. Bajaj, V. and R.B. Pachori, *Classification of seizure and nonseizure EEG signals using empirical mode decomposition.* IEEE Transactions on Information Technology in Biomedicine, 2011. **16**(6): p. 1135-1142.
121. Yazdani, A., T. Ebrahimi, and U. Hoffmann. *Classification of EEG signals using Dempster Shafer theory and a k-nearest neighbor classifier*. in *2009 4th International IEEE/EMBS Conference on Neural Engineering*. 2009. IEEE.
122. Weinberger, K.Q. and L.K. Saul, *Distance metric learning for large margin nearest neighbor classification.* Journal of machine learning research, 2009. **10**(2).
123. Belongie, S., J. Malik, and J. Puzicha, *Shape matching and object recognition using shape contexts.* IEEE transactions on pattern analysis and machine intelligence, 2002. **24**(4): p. 509-522.
124. Chopra, S., R. Hadsell, and Y. LeCun. *Learning a similarity metric discriminatively, with application to face verification*. in *2005 IEEE Computer*

*Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005. IEEE.

125. Thieu, T.N. and H.-J. Yang. *Diagnosis of epilepsy in patients based on the classification of EEG signals using fast Fourier transform*. in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. 2015. Springer.

126. Paul, Y., *Various epileptic seizure detection techniques using biomedical signals: a review.* Brain informatics, 2018. **5**(2): p. 1-19.

127. Schindler, K., et al., *Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG.* Brain, 2007. **130**(1): p. 65-77.

128. Gordan, M., et al. *Computational complexity reduction of the support vector machine classifiers for image analysis tasks through the use of the discrete cosine transform*. in *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*. 2006. IEEE.

129. Litt, B., G. Worrell, and Z.G. Ives, *The international epilepsy electrophysiology portal*.

130. Saputro, I.R.D., et al. *Seizure type classification on EEG signal using support vector machine*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.

131. Polat, K. and M. Nour, *Epileptic Seizure Detection Based on New Hybrid Models with Electroencephalogram Signals.* IRBM, 2020. **41**(6): p. 331-353.

132. Dunseath, W.R. and E.F. Kelly, *Multichannel PC-based data-acquisition system for high-resolution EEG.* IEEE Transactions on Biomedical Engineering, 1995. **42**(12): p. 1212-1217.

133. Anscombe, F., *Quiet contributor: The civic career and times of John W. Tukey.* Statistical Science, 2003: p. 287-310.

134. Kamen, E.W. and B.S. Heck, *Fundamentals of signals and systems using the web and matlab*. 2006: Prentice-Hall, Inc.

135. Kamen, E.W. and B.S. Heck, *Fundamentals of Signals and Systems using the Web and MATLAB*. 2000: Prentice Hall.

136. Akrami, A., et al. *EEG-based mental task classification: Linear and nonlinear classification of movement imagery*. in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. 2006. IEEE.
137. Joachims, T. *Text categorization with support vector machines: Learning with many relevant features*. in *European conference on machine learning*. 1998. Springer.
138. DeCoste, D. and B. Schölkopf, *Training invariant support vector machines.* Machine learning, 2002. **46**(1): p. 161-190.
139. Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.
140. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. 2009: Springer Science & Business Media.
141. Boser, B.E., I.M. Guyon, and V.N. Vapnik. *A training algorithm for optimal margin classifiers*. in *Proceedings of the fifth annual workshop on Computational learning theory*. 1992.
142. Parvin, H., H. Alizadeh, and B. Minaei-Bidgoli. *MKNN: Modified k-nearest neighbor*. in *Proceedings of the world congress on engineering and computer science*. 2008. Citeseer.
143. Coomans, D. and D.L. Massart, *Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules.* Analytica Chimica Acta, 1982. **136**: p. 15-27.
144. Kevric, J. and A. Subasi, *Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system.* Biomedical Signal Processing and Control, 2017. **31**: p. 398-406.
145. Kayikcioglu, T. and O. Aydemir, *A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data.* Pattern Recognition Letters, 2010. **31**(11): p. 1207-1215.
146. Fan, J., et al. *A Step towards EEG-based brain computer interface for autism intervention*. in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2015. IEEE.

147. Sabancı, K. and M. Köklü, *The classification of eye state by using kNN and MLP classification models according to the EEG signals.* 2015.

148. Aldea, R., M. Fira, and A. Lazăr. *Classifications of motor imagery tasks using k-nearest neighbors*. in *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*. 2014. IEEE.

149. Golmohammadi, M., et al. *The TUH EEG seizure corpus*. in *Proceedings of the American Clinical Neurophysiology Society Annual Meeting*. 2017.

150. Han, J., M. Kamber, and J. Pei, *Data mining concepts and techniques third edition.* The Morgan Kaufmann Series in Data Management Systems, 2011. **5**(4): p. 83-124.

151. Virmani, J., et al., *Characterization of primary and secondary malignant liver lesions from B-mode ultrasound.* Journal of digital imaging, 2013. **26**(6): p. 1058-1070.

152. Fawcett, T., *An introduction to ROC analysis.* Pattern recognition letters, 2006. **27**(8): p. 861-874.

153. Firman Tempola, M.M. and A. Khairan, *Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation.* Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 2018. **5**(5).

154. Kohavi, R., *Glossary of terms.* Special issue on applications of machine learning and the knowledge discovery process, 1998. **30**(271): p. 127-132.

155. Patterson, V., *Managing epilepsy by telemedicine in resource-poor settings.* Frontiers in public health, 2019. **7**: p. 321.

156. Vaughan, K.A., et al., *An estimation of global volume of surgically treatable epilepsy based on a systematic review and meta-analysis of epilepsy.* Journal of neurosurgery, 2018. **130**(4): p. 1127-1141.

157. Kiral-Kornek, I., et al., *Epileptic seizure prediction using big data and deep learning: toward a mobile system.* EBioMedicine, 2018. **27**: p. 103-111.

**APPENDIX A**

# Fourier Transform based Epileptic seizure features classification using scalp electrical measurements using KNN and SVM

Athar H.Ali[1,a)],Saif Al-jumaili[2,a,b)] ,Abdullahi Abdu Ibrahim[3,c)], Adil Deniz Duru[4,d)]

[1] *Information technologies department, Altinbas University, Istanbul, Turkey*
[2] *Electrical and computer engineering, Altinbas University, Istanbul, Turkey*
[3] *Department of Computer engineering, Altinbas University, Istanbul, Turkey*
[4] *Neuroscience and Psychology Research in Sports Lab, Faculty of Sport Science, Marmara University, Istanbul, Turkey.*
[a)] Athar.alazzawi, Saif.aljuamili@ogr.altinbas.edu.tr
[b)] lolo89.adm,Saifabdalrhman@gmail.com
[c)] abdullahi.ibrahim@altinbas.edu.tr
[d)] deniz.duru@marmara.edu.tr

**Abstract.** Epilepsy classification techniques are one of the areas that are still under searching till now as long as there is no specific method for detection seizures. The brain consists of more than 100 billion nerves that generate electrical activity. These activities are recorded using an Electroencephalogram (EEG) by electrodes attached to the scalp. EEG is considered a big footstep in the medical and technical field where it allows the detection of brain disorders. However, this paper aimed to classify three types of epilepsy seizures with the lowest error rate. As well, assisting clinicians to collect patient information without needing daily observation as a digital notebook, which is easier to memorize. Therefore, we applied two classification techniques namely Support Vector Machine (SVM) and k-Nearest Neighbors (KNN), which rely on the features extracted from the data by the Fast Fourier Transform (FFT) method. The results show SVM obtained the highest accuracy value compared to KNN, accurate scores were 99.5% and 99%, respectively.

**Keywords:** Electroencephalogram (EEG), Fast Fourier Transform (FFT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM).

## INTRODUCTION

The magic wand that allows us to unpack brain behavior which is the most complicated, significant, and mysterious organ ever in the human body is called the Electroencephalogram (EEG). EEG is a tool that provides the ability to detect what happens inside the neurons, recording the brain's electrical activity up to 20–30 min generally from multiple channels, and determine where and when it occurs without affecting the validity of the data [1, 2]. The EEG advantages included lightweight which gives the ability to do a scan anywhere and has a lower cost than other devices [3]. Besides that, EEG is a non-invasive tool, although in some cases, it is necessary to observe the electrical activity from the cortex or deep structures of the brain. Hans Berger was the first German neurologist who applied EEG in the 1920s to humans and had been recorded through surgical operation for a boy by the neurosurgeon Nikolai Guleke [4]. EEG created a fundamental paradigm shift in neuropsychological milestones and was recently considered the head leader of studying brain activity. The brain activity is measured by placing small metal pieces attached to the scalp called electrodes.

Electrode systems locations on the scalp were defined by the American Encephalographic Society in 1994 [5, 6]. The most common systems for electrode placements are referred to as the 10-20 system, which is an international system that expanded later to 10-10 and 10-5. These systems differ in the number of electrodes and the way they are distributed on the scalp. The strength of signals or weakness is distinguished when an EEG scan is performed and according to the patient's situation which includes wakefulness, drowsiness, sleep, and certain activities. The signals can be recorded from any part of the brain and have a diversity of frequencies called the power of frequency bands [6]. These power bands contain information about the mental brain state. Table 1 below shows the location and condition for frequency bands' appearance.

**TABLE 1**. EEG Frequency Bands

| Frequency Bands | EEG Frequency ranges | Description |
|---|---|---|
| Delta | 0.5-4 Hz | It occurs during sleep and is most pronounced in the right part of the brain |
| Theta | 4-8 Hz | They appear with the brain processes that need the mental capability or working memory, they can be recorded from all parts of the cortex |
| Alpha | 8-13 Hz | It appears while you relax mentally, eyes closed, walking, reading, and practicing yoga |
| Beta | 13-30 Hz | It occurs due to acts of thinking or anxiety and concentration, and this hesitation appears in both the front and back areas. |
| Gamma | >30 Hz | It appears in the somatosensory cortex and is related to brain decision-making. |

EEG is used to diagnose various diseases such as Epilepsy, Stroke, Alzheimer's, drug intoxication, tumors, and sleep disorder (Narcolepsy), and trauma [7]. Epilepsy is one of the ancient diseases that was known for a long time and has effects on the brain, as it disrupts the electrical activity of the central nervous system and is accompanied in some cases by loss of consciousness [8, 9]. The first physician who realized that Epilepsy was a disease that affects the brain and tried to treat him was The Greek physician Hippocrates. Epilepsy disease targets, males, females, children, and adults of all races, but on the bright side, it could be treatable with medication [10]. Epilepsy affects around 1% of the people in the world, which is around 65 million people [11]. The disorders occur in the brain parts as an electric wave, irregular behavior appears in a person's routine. In the medical field, clinicians classify seizures based on when and where the abnormal brain activity began and denote them as Focal (partial) seizures when it happens in one part of the brain and Generalized seizures when it takes over all the brain [12]. This study is highlighted as follows, Section 1, Introduction, Section 2, a review of the literature related to our work. Section 3, the database description, Section 4, includes the methodology, which is preprocessing, classification, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN). Section 5, discussing our results and comparing them with other similar studies, and the Conclusion takes place in Section 6.

## RELATED WORK

Since epilepsy became the most prevalent Nervous disease worldwide and became one of the important era issues, numerous studies discussed the classification of epileptic seizures using various types of techniques. One of these studies, Subhrajit Roy et al. [11], classified the seizure types as normal and abnormal seizures. They used the TUH dataset and FFT algorithm as feature extraction and K-Nearest Neighbors (KNN) as a classifier. In addition, classifiers' techniques included XGBoost, Convolutional Neural Networks (CNN), Stochastic Gradient Descent (SGD). They applied KNN and SGD classifiers and achieved outstanding performance with 90% accuracy for KNN while SGD was 56%. Also, IRD Saputroet et al. [13], classified the seizure stages into pre-ictal, ictal, inter-ictal, and postictal and described their characteristics. Where the seizures were Tonic-Clonic (TCSZ), Focal Non-Specific (FNSZ), Generalized Non-Specific (GNSZ), and normal EEG signal. The authors used the TUH dataset and SVM classifier with three different features to extract the signal properties, Independent Component Analysis (ICA), Mel Frequency Cepstral Coefficients (MFCC), and Hjorth Descriptor. They used different SVM kernels which are separated into two categories linear and nonlinear. The nonlinear kernel included quadratic and cubic, fine, medium, and coarse gaussian [14]. The best result obtained was from the combination of MFCC+Hjorth Descriptor, they used different evaluation performances to check classifiers which were sensitivity, specificity, and accuracy the results were 90.25%, 97.83%, and 91.4% respectively. K. Polat et al. [15], classified different types of epileptic seizures by using the UCI machine learning repository which is considered one of the best 100 cited databases in the computer science field. They used an SVM classifier with three models of kernels and three normalization techniques were minimum-maximum, z-score, and MAD. The kernels were Cubic SVM, Medium Gaussian SVM, and Linear SVM. The best result was by using Cubic SVM with MAD which got an accuracy of 82.50 %. Also, Manish Sharma et al. [9], the authors classify seizure types using Temple University Hospital dataset and three feature extraction methods: logarithmic squared norm (LSN), Fuzzy Entropy (FuEn), and fractal dimension (FrD), where the classification technique was used is SVM. The best result they achieved was accuracy of 79.34%.

## DATASET USED IN THIS STUDY

Temple University Hospital Seizure Corpus (TUH) is one of the most extensive open-source databases [1, 16]. Where TUH included extensive information about seizures, such as seizure type, occurrence time, and duration. Moreover, a report written by a board-certified neurologist contains the patient's information, age, gender, medical history and includes two areas, impression, and clinical correlation, which contain the clinician's findings [5]. Table 2 below summarizes the categories of epileptic seizures [8, 9].

**TABLE 2**. Categories of The Main Seizures Classes

| | |
|---|---|
| **Focal (partial) seizures**: which can be separated into two sub-categories [17] | |
| Focal seizures without loss of consciousness [9] | These seizures named partial simple seizures may lead to Involuntary twitching of one part of the body but without loss of consciousness |
| Focal seizures with impaired awareness [9] | It is called Complex partial seizures, which are accompanied by loss of consciousness or a behavior change. During a seizure, you cannot communicate or respond to the reality surrounding you, and you may make some movements repeatedly, such as chewing, moving hands, or swallowing |
| **Generalized Non-Specific**: which can be segregated into six sub-categories [9, 18] | |
| Absence seizures | Earlier known as minor epileptic seizures most often occur in children. maybe accompanied by body movements and brief loss of consciousness. |
| Tonic seizures | Seizures result in body stiffness and last from a few seconds to a minute and usually occur during sleep, in the waking |

| | |
|---|---|
| | period, it causes the person to fall. |
| Atonic seizures | seizures lead to a sudden loss of control of the body's muscles, lead to falls and injury to the head or face and usually last a few seconds. |
| Clonic seizures | These kinds of seizures aimed at the upper part of the body like the face, neck, and the arms with repetitive rhythmic for the spasms muscle |
| Myoclonic seizures | These seizures cause sudden spasms in the legs and arms and are for short periods. |
| Tonic-clonic seizures | cause a sudden loss of consciousness and sometimes loss of bladder control or tongue biting. These were formerly known as major epileptic seizures. |

The selected EEG signals were recorded according to the International 10/20 system electrode at a sampling frequency rate of 250 Hz. The recordings consist of 16 unipolar channels EEG distributed on the scalp and can be divided into two main groups which are odd can be on the left side and the even can be on the right side as shown in Table 3.

**TABLE 3.** The Channels Labels in a 10/20 System as Represented in TUEG

| Channel Label | Position on Scalp | Channel Label | Position on Scalp |
|---|---|---|---|
| Fp1 | left forehead | Fp2 | right forehead |
| T5 | left outer posterior | T6 | right outer posterior |
| F7 | left outer frontal | F8 | right outer frontal |
| F3 | left inner frontal | F4 | right inner frontal |
| P3 | left inner posterior | P4 | right inner posterior |
| C3 | left inner medial | C4 | right inner medial |
| T3 | left outer medial | T4 | right outer medial |
| O1 | left back of the head | O2 | right back of the head |

## METHODOLOGY

Signal preprocessing is an important step in the data mining process to reduce noise in the EEG signal and improve data efficiency which guarantees a good result. Due to the rapid development in the size of databases in the real world and their diversity that affects the quality of the data and thus reduces the quality of the results. There are various data pre-processing techniques; such as reducing information (feature reduction), which reduce massive data features to a small number, by producing new value features from the initial ones [19]. Indeed, a lot of patterns for feature extraction have been proposed such as Entropy, approximate entropy, sample entropy, fuzzy entropy, wavelet transform, time-frequency distributions, and Fast Fourier Transform (FFT) [20]. Each technique has a different advantage that can improve accuracy, speed, and data visualization. In this study, we used FFT as a feature extraction method which was developed by the scientists C. Cooley and John Tookey in 1965 [21]. It is considered one of the most popular methods compared to other techniques, this algorithm aims to speed up finding a discrete Fourier transform. FFT technique works perfectly with signal processing where it converts the signal from the time domain that is recorded through the scalp into the frequency domain which is the spectrum [22]. This method uses mathematical tools for EEG data analysis, the mechanism of action and the idea are based on the sequencing division. Also, the frequency domain showed spectacular work in feature extraction that enabled the detection of mental brain states according to most researches and studies [23]. Since the data obtained from FFT were collected from multiple areas of the scalp, the data acquired multiple features spanning varying degrees of size, scope, and units. The two of the most used feature scaling techniques in machine learning are normalization and standardization. Generally, when data needs standardization some methods can be used such as t-score and z-score. Therefore, in this study, we used Euclidean normalization and used it as input for both classifications' techniques KNN and SVM.

## CLASSIFICATION

Classification is a technique that allows identifying the data classes through classifiers that assign labels to our data. EEG signals classification is what bonded the brain disorder and the brain-computer interface [24]. In this work, MATLAB R2021a software had been used by employing two classifiers SVM and KNN.

## K-NEAREST NEIGHBOR (KNN)

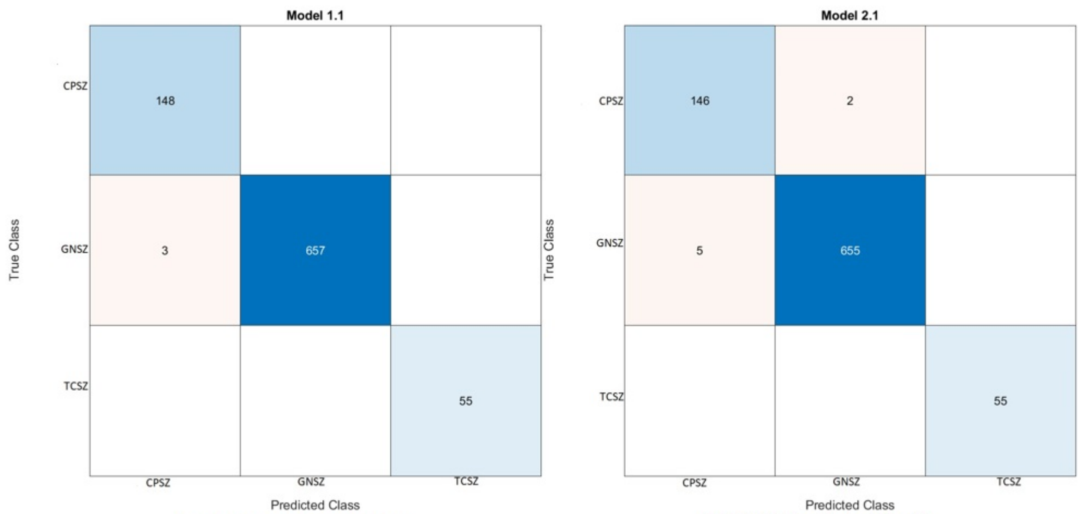
KNN is regarded as one of the supervised machine learning algorithms, also one of the topmost and efficient algorithms used in classification and regression problems. Evelyn Fix and Joseph Hodges were the first who developed KNN in 1951 and later Thomas Cover was who expanded it [25]. KNN is used in image recognition applications and handwriting detection, and so on. KNN is simple to understand as it is based on the feature-similarity approach. The KNN method is working on assigning new unclassified data to the closest neighbors K, even without retraining the classifier. In KNN, K is the nearest neighbor point between the data we computed. Many methods were found to identify the closest similarity neighbor, which is dependent on distance calculation such as Hamming, Manhattan, Euclidean, and Minkowski distance metrics. Where Euclidean distance is a common way to use the distance scale, as long as the Euclidean distance has no problem when it is applied with these types of data and with the KNN classifier [26]. Where KNN gives the ability to deal with various classes and that is what is so important in BCI.

## SUPPORT VECTOR MACHINE (SVM)

SVM is one of the most popular supervised algorithms which belong to the linear classification family. SVM is the golden solution to the problems which are related to clustering, regression, and prediction. SVM algorithm represents the data in n-dimensional space as a point, where these points represent the features. The data and the value of each feature are plotted in the form of specific coordinates. The main idea behind SVM is to start with data in relatively low dimensions and transform the data into a higher dimension. As well separated the data that is in a higher dimension into categories by support vector classifier. SVM's fantastic reputation was gained through dealing with the non-linear data, by using the kernel technique which ensures a better generalization to the data [27]. The kernel function lays on the relationships between the points, where it does not do the transformation of the points but deals with them as if they are in the higher dimensions. In addition, one of the advantages of SVM used is reducing the experimental classification error and increasing the geometric margin.

## RESULTS

The results below are illustrated and measured by the two main evaluation criteria which are specificity and sensitivity as shown in Figure 1(a) the SVM confusion matrix and Figure 1(b) the KNN confusion matrix. while Table 4 summarizes the results. After applied FFT features extraction, the result was used as input for both SVM and KNN classifiers. The obtained results show that SVM outperforms the KNN with (99.5%) compared to (99%) for accuracy.




(a) SVM Confusion Matrix (a) KNN Confusion Matrix

**FIGURE 1**. (a) SVM Confusion Matrix implemented by FFT, (b) KNN Confusion Matrix implemented by FFT

**TABLE 4.** Average Sensitivity and Specificity

| Classifier Model | Feature Type | Specificity / Sensitivity | CPSZ | GNSZ | TNSZ |
|---|---|---|---|---|---|
| SVM | FFT | Specificity (%) | 99 | 98 | 100 |
| | | Sensitivity (%) | 98 | 99 | 100 |
| KNN | | Specificity (%) | 99 | 97 | 100 |
| | | Sensitivity (%) | 96 | 99 | 100 |

Bear in mind CPSZ seizures come with some movements repeatedly, such as chewing, moving hands, or swallowing, and are accompanied by loss of consciousness. While GNSZ deals with the focal nonspecific seizures and covers the skull with larger channels. TNSZ is a seizure that is characterized by muscle stiffening and associated with Tonic-Clonic seizures.

**TABLE 5.** Comparison of The Classification Accuracies, where UCI-ML refers to UCI Machine Learning repository, MG-SVM refers to Medium Gaussian SVM, CES refers to Classify epileptic seizure, and HD refers to Hjorth Descriptor.

| Researcher | Study Objective | Database Source | Classifiers Techniques | Features Methods | Accuracy (%) |
|---|---|---|---|---|---|
| **Present work** | **CES** | **TUH** | **SVM**<br>**KNN** | **FFT** | **99.5**<br>**99** |
| Saputroet et al. [13] | CES | TUH | SVM | (MFCC)+ HD | 91.4 |
| Subhrajit et al. [11] | CES | TUH | KNN<br>XGBoost | FFT | 90<br>56 |
| K.Polat et al. [15] | Detect and CES | UCI-ML | Cubic SVM<br>MG-SVM<br>Linear SVM | FFT | 82.50<br>81.40<br>76.70 |
| Manish et al. [9] | CES | TUH | SVM | FuEn,LSN, and FrD | 79.34 |

Bear in mind that XGBoost is a library that implements algorithms in the machine learning field with high performance and in a fast accurate way even for billions of data. Likewise, Cubic SVM is one of the SVM kernels that deal with nonlinear data, and it is distinguished with the least computation time. Else, Medium Gaussian is also a kernel of SVM classifiers that deal with nonlinear data and its values depending on the distance either from the origin point or some other point. While linear SVM deals with linear 2-D data and is the simplest way for searching for the largest straight-line, it can be separated by the margin. Finally, Table 5 shows that our results were superior compared to other similar works, our results were written in bold. Where the nearest result to our work was 91%, the authors used the same dataset and same linear classifier which is SVM. Where the nearest result to our work was 91%, the authors used the same dataset and same linear classifier which is SVM. Thus, We conclude that linear classification techniques with these dataset types perform perfectly.

## CONCLUSION

This study provided high accuracy results using the two classification techniques SVM and KNN. Besides, the methodology used was dependent on one of the most enormous databases available, the TUH dataset. The recordings were made by using 16 electrodes (16 channels) placed on the scalp with a sampling rate of 250 Hz. The total raw data we used to classification that we applied by FFT feature extraction was (864) for each class were CPSZ (149), GNSZ (660), and TNSZ (55), all these raw data was applied to the 5-fold that will reduce the basis for classification. The output was normalized by Euclidean distance which enabled us to bring all the data to the same range to be ready for classification. We achieved an accuracy rate of 99.5% and 99% for SVM and KNN respectively, which is higher results compared with the previous work aforementioned in section two. Finally, the fast computational characteristics of KNN and SVM and the ability to deal with various classes was the reason to serve and evaluate the performance of our approach which is what is so important in BCI.

## REFERENCES


1. Sanei, S. and J.A. Chambers, *EEG signal processing*. 2013: John Wiley & Sons.
2. Duru, A.D., *Determination of increased mental workload condition from EEG by the use of classification techniques.* International Journal of Advances in Engineering and Pure Sciences, 2019. **31**(1): p. 47-52.
3. Crawford, B., et al., *Factors influencing progress in establishing community-based marine protected areas in Indonesia.* Coastal Management, 2006. **34**(1): p. 39-64.
4. Tudor, M., L. Tudor, and K.I. Tudor, *Hans Berger (1873-1941)--the history of electroencephalography.* Acta medica Croatica: casopis Hravatske akademije medicinskih znanosti, 2005. **59**(4): p. 307-313.
5. Subasi, A., *EEG signal classification using wavelet feature extraction and a mixture of expert model.* Expert Systems with Applications, 2007. **32**(4): p. 1084-1093.
6. Jasper, H.H., et al., *Reticular formation of the brain.* Academic Medicine, 1958. **33**(11): p. xviii.
7. Tsolaki, A., et al., *Electroencephalogram and Alzheimer’s disease: clinical and research approaches.* International journal of Alzheimer’s disease, 2014. **2014**.
8. Vickers, N.J., *Animal communication: when i’m calling you, will you answer too?* Current biology, 2017. **27**(14): p. R713-R715.
9. Sharma, M., S. Patel, and U.R. Acharya, *Automated detection of abnormal EEG signals using localized wavelet filter banks.* Pattern Recognition Letters, 2020. **133**: p. 188-194.
10. Patterson, V., *Managing epilepsy by telemedicine in resource-poor settings.* Frontiers in public health, 2019. **7**: p. 321.
11. Roy, S., et al., *Seizure type classification using EEG signals and machine learning: Setting a benchmark.* arXiv: 1902.01012, 2019.

12. Agarwal, R., et al., *Automatic EEG analysis during long-term monitoring in the ICU.* Electroencephalography and clinical Neurophysiology, 1998. **107**(1): p. 44-58.
13. Saputro, I.R.D., et al. *Seizure type classification on EEG signal using support vector machine*. in *Journal of Physics: Conference Series*. 2019. IOP Publishing.
14. Wijayanto, I., A. Rizal, and S. Hadiyoso. *Multilevel Wavelet Packet Entropy and Support Vector Machine for Epileptic EEG Classification*. in *2018 4th International Conference on Science and Technology (ICST)*. 2018. IEEE.
15. Polat, K. and M. Nour, *Epileptic Seizure Detection Based on New Hybrid Models with Electroencephalogram Signals.* IRBM, 2020. **41**(6): p. 331-353.
16. Subasi, A. and E. Ercelebi, *Classification of EEG signals using neural network and logistic regression.* Computer methods and programs in biomedicine, 2005. **78**(2): p. 87-99.
17. Sharma, M., et al., *An automatic detection of focal EEG signals using new class of time–frequency localized orthogonal wavelet filter banks.* Knowledge-Based Systems, 2017. **118**: p. 217-227.
18. Sharma, M., S. Shah, and P. Achuth, *A novel approach for epilepsy detection using time–frequency localized bi-orthogonal wavelet filter.* Journal of Mechanics in Medicine and Biology, 2019. **19**(01): p. 1940007.
19. Gordan, M., et al. *Computational complexity reduction of the support vector machine classifiers for image analysis tasks through the use of the discrete cosine transform*. in *2006 IEEE International Conference on Automation, Quality and Testing, Robotics*. 2006. IEEE.
20. Al-Fahoum, A.S. and A.A. Al-Fraihat, *Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains.* International Scholarly Research Notices, 2014. **2014**.
21. Anscombe, F., *Quiet contributor: The civic career and times of John W. Tukey.* Statistical Science, 2003: p. 287-310.
22. Kamen, E.W. and B.S. Heck, *Fundamentals of signals and systems using the web and matlab*. 2006: Prentice-Hall, Inc.
23. Akrami, A., et al. *EEG-based mental task classification: Linear and nonlinear classification of movement imagery*. in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. 2006. IEEE.
24. Shuren, Q. and J. Zhong. *Extraction of feature information in EEG signal by virtual EEG instrument with the functions of time-Frequency analysis*. in *2009 Proceedings of 6th International Symposium on Image and Signal Processing and Analysis*. 2009. IEEE.
25. Medithe, J.W.C. and U.R. Nelakuditi. *Study of normal and abnormal EEG*. in *2016 3rd International conference on advanced computing and communication systems (ICACCS)*. 2016. IEEE.
26. Kevric, J. and A. Subasi, *Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system.* Biomedical Signal Processing and Control, 2017. **31**: p. 398-406.
27. Dietrich, R., M. Opper, and H. Sompolinsky, *Statistical mechanics of support vector networks.* Physical review letters, 1999. **82**(14): p. 2975.