A DWT-based Multiclass Classification of Epileptic Activity in Patients

A Major II Project Report

Submitted in partial fulfillment of the requirements for the award of the degree of

MASTER OF TECHNOLOGY

in

Computer Science and Engineering

by

Abhra Gupta

(Roll No. 2K17/CSE/02)

under the guidance of

Prof. Daya Gupta

DEPT. OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, New Delhi- 110042

JUNE ,2019

CANDIDATE'S DECLARATION

I hereby declare that the Major Project-II work entitled "A DWT-based Multiclass Classification of Epileptic Activity in Patients" which is being submitted by me to the **Department of Computer Science and Engineering** to Delhi Technological University, in partial fulfilment of requirements for the award of the degree of Master Of Technology(Computer Science and Engineering) is a bonafide report of Major Project-II carried out by me. I have not submitted the matter embodied in this dissertation for the award of any other Degree or Diploma.

Place: Delhi (Abhra Gupta)
Date: (Roll no. 2K17/CSE/02)

CERTIFICATE

This is to certify that Project Report entitled "A DWT-based Multiclass Classification of Epileptic Activity in Patients" submitted by Abhra Gupta (roll no. 2K17/CSE/02) in partial fulfillment of the requirement for the award of degree Master of Technology (Computer Science and Engineering) is a record of the original work carried out by him under my supervision. The work has not been previously submitted for the fulfillment of any other Degree or Diploma to this University or elsewhere.

Place: Delhi
Date:

Prof. Daya Gupta
SUPERVISOR

ACKNOWLEDGEMENT

I am highly indebted to **Prof. Daya Gupta** and obliged for giving me the autonomy of working and exploring different avenues regarding thoughts. The sustaining and blooming of the present work is primarily because of her significant direction, proposals, insightful judgement, valuable feedback and an eye for flawlessness. My mentor always answered the myriad of my doubts with smiling graciousness and immense patience, never letting me feel that I am novices by always lending an ear to my views, appreciating and improving them and by giving me a free hand in my project. It's only because of her overwhelming interest and helpful attitude; the present work has attained the stage it has.

I express my deepest and sincere gratitude to **Mrs. Divyashikha Sethia** who personally supervised, supported and encouraged me all through this work. She has been helpful and a constant source of inspiration throughout my project.

Also, I would like to express my deepest gratitude to **Dr. Vibha Sharma**, Head of Department, Neurology at Institute of Human Behaviour and Allied Sciences (IHBAS) for providing me with an opportunity to get real-time exposure to the Electroencephalogram (EEG) research and activities going on in the Neurocomputing Lab at the Institute. The visit at IHBAS provided me with ample knowledge and technical suggestions during my research pursuit.

(Abhra Gupta)
(Roll no. 2K17/CSE/02)

ABSTRACT

The identification of seizure activities in non-stationary electroencephalography (EEG) is a challenging task. The seizure detection by human inspection of EEG signals is prone to errors, inaccurate as well as time-consuming. Several attempts have been made to develop automatic systems so as to assist neurophysiologists in identifying epileptic seizures accurately. The proposed study suggests using Discrete Wavelet Transform to decompose the EEG signals into frequency sub-bands. We choose a certain subset of the frequency sub-bands for feature selection. Following the DWT decomposition, we calculate the Standard Deviation for each sub-band present in the subset. Finally, the standard deviation values of the sub-bands are fed to a Support Vector Machine. The proposed work consists of 5 experiments which are essentially classification problems: 3 of which are Multi-class classification problems and the rest two are Binary Classification problems. In the proposed work, we investigate the three-class classification problems focused on classifying an EEG signal into one of the three classes, which are 1. Healthy patient 2. Seizure-free epochs:inter-ictal stage 3. Epileptic Activity:ictal stage. The dataset used in the proposed work is obtained from the Department of Epileptology of the University of Bonn.The accuracy achieved in one of the Multi-class classification experiment in the proposed work is 98.45% which beats the state of the art accuracy in this three-class problem. Additionally, the proposed method has achieved highest accuracy of 100% in classifying normal EEG signals(eyes closed) and seizure EEG signal and an accuracy of 100% in classifying normal EEG signals(eyes open) and seizure EEG signal which is comparable with the existing state of the art EEG signal classification techniques. Six different classification techniques have been used in each of the five experiments conducted where every classification technique has been used with 8 different Daubechies wavelets db1 to db8. The results obtained from these experiments provide valuable insights establishing that SVM performs the best in most of the experiments with the db4 wavelet among the 8 wavelets achieving the highest accuracy.
Keywords: Epilepsy · Electroencephalogram(EEG) Signals · Signal Processing · Wavelet Decomposition · Classification · Machine Learning

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Epilepsy is a chronic condition caused in the brain where seizures occur multiple times unreliably. It generally occurs where the victim loses consciousness or there are tremendous convulsions over the entire body. This lowers the life expectancy[2] . On a biological basis, the seizures in epileptic patients are caused because they seem to have inability to have control over the level of their cortical neurons. With negative potential shifts in the patients, the seizures start occurring because it causes the over-excitation of the cortical neuronal networks [6]. As a measure to remedy this, research was done on how to suppress this negative shifts. Neurofeedback was conducted on the patients to reinforce them to modify their baseline levels of potential[26]. In most Human Behavioral Science Hospitals, the onset of an epileptic seizure is detected using Electroencephalogram(EEG) which is worn on the scalp. It has multiple channel electrodes to capture the brain waves. The readings on the EEG are the result of the electrical activities of the neurons. Electrodes are pasted with the help of a gel on the scalp of the patient. The standard method to apply the electrodes on the scalp is the 10-20 system. Automatic seizure detection using patient non-specific classifiers has two main advantages in medical science. Firstly, this would mean we can get the seizure data with greater accuracy. Secondly, as a great therapy this would avoid the danger caused by seizures [42]. In a therapeutic sense, a closed system can be built which can send signals to the patient about the onset of a seizure, and thus helpful in performing various kinds of Neurofeedback methods [25]. But, there are problems associated with patient non-specific classifiers. They might display very low accuracy along with a significant delay in detection of seizures.[Wilson et al, 2004] because the seizure activity and non-seizure activity is variable from patient to patient. The designing of patient-specific classifiers for seizure detection is a challenge because of several reasons:

1. There has to be a sharp distinction between detector sensitivity and specificity because epileptic patients have seizure states and non-seizure states overlapping with each other

2. The transition from seizure state to non-seizure state is a non-stationary process, and therefore the detection has to be done within a short critical time

3. Algorithms have to be designed keeping in mind the scarcity of training data

Seizure Detectors are classified into two types:

1. Seizure Onset Detectors

2. Seizure Event Detectors

The purpose of each of these is complimentary to the other. In other words, a seizure onset detector has to recognize a seizure with the shortest possible delay relaxing on the accuracy of the prediction. On the other hand, the purpose of a seizure event detector is to identify a discharge as accurately as possible with a relaxation on the latency of seizure detection.

1. **Seizure Onset Detection**

   In a therapeutic sense, seizure onset detection could be used to trigger the neurostimulators. This has an effect while a seizure is occuring . A well-working seizure onset detection model can help aware the doctor to administer an anti-convolusant. This does not mean doctors can wait for too long to warn the patient. It increases the possibility that the patient will be incapable of any reaction to the seizures.

2. **Seizure Event Detection**

   Seizure event detectors can help doctors for better therapy of epilepsy over time. Due to various reasons, doctors end up prescribing too much or too little. What can be done in order to alleviate this is that a device can be worn on the scalp which can do seizure event detection to provide the statistics(number, frequency, duration) of the seizures. Correlating this information with medication can maximally benefit the patients.

1.2 Motivation

Almost 80% of the patients with Epilepsy live in low earning countries. They have a response to medication 70% of the time. About three-quarters of the people who have epilepsy do not get the treatment they need. In many areas of the world, epileptic patients are still suffering from the discrimination and dogmatism. To talk about the rates of the disease, there are almost 50 million people suffering from epilepsy. On a scale of 1000 people, we have 4-10 persons having epilepsy. Research shows that in the poorer countries the ratio is much higher around 7 to 14 people have epilepsy out of 1000 people. Almost 2.4 million people are identified with Epilepsy every year. In high-income countries, around 0.03-0.05% of the people in the total popluation have epilepsy. In poorer countries it can be twice the numbers. Epilepsy is about 0.6% of the diseases, based on a statistical study that considers years of life cut short because of early death. An Indian study done in 1998 concluded that the expenses invested in epilepsy treatment on a patient was about 88% of the country's economic production value. On a study on Statistics Of Epilepsy done by M. Leuret, of the Bictre, Paris, the following evidences are taken from the paper on this subject. "24 out of 106 cases started getting seizures at an age between 10 and 14. 18 experienced their first seizure between 15 and 19 years. 16 were attacked between 20 and 24 years. Of these 106 cases it was found out that the father and mother had Epilepsy in 6 cases only. Of these 106 patients, 30 were drunkards, 15 addicted. 30 of the 106 patients had attacks regularly once a fortnight; 17 used to have seizures once a month; 13 once a week; 9 every three or four days; 4 almost daily and 2 daily,1 every 2 months, 3 every three months and 24 at very irregular intervals. In 29, they were as many in the day as in the night. In 8 they occurred in the day, in 12 they occur frequently in the day. In 8 cases, seizures occur in the night only; in 3 cases seizures happened in the morning only; and in 1 case seizures happen in the evening only." **In order to minimize the rate of epileptic seizures, the thesis attempts the problem of seizure detection using Machine Learning Algorithm. This will require acquisition of the EEG signals of the patients.**

1.3 Problem Statement

Around 50 million people in the world are estimated to suffer from Epilepsy. The convulsions that occur in these patients are as a result of excess electrical discharges in large number of neurons. In medical institutions , epileptic activity is detected manually by using the technique called Electroencephalography(EEG). **The manual method to detect epileptic activity in patients is to look for a "spike followed by a smooth" complex in the EEG signal**. This process is very time consuming and erroneous. This motivates the need to develop fast, reliable and accurate techniques for seizure detection in epileptic patients.
The goal of seizure detection is to classify from the EEG readings, whether the activity is seizure or non-seizure. Unfortunately, the problem of seizure classification cause many algorithms to form multiple segmentations beyond simple seizure and non-seizure states. So, a supervised framework has to be used instead of unsupervised.

Before automated systems were developed, the EEG signals used to be manually analyzed by neurologists for detection of seizure. Some of the related works for eg. Sharmila et al. [41] and Guo et al.[17] in the field of Epileptic Seizure Detection use Discrete Wavelet Transform for decomposition of the EEG signals. The wavelet decomposition is done upto eight levels. The wavelet coefficients obtained after the decomposition are fed to the Machine Learning Classifiers used in the works respectively with very good results. The proposed work closely resembles the research done by Sharmila et al [41] and achieves better results. The highest accuracy achieved on a Multi-class classification on Healthy, Inter-Ictal and Ictal patients is 98.45% in the proposed work which beats the state of the art accuracy. **The purpose of the thesis is focused on building an autonomous system which achieves the task of epileptic seizure detection using various Machine Learning Algorithms.**
The proposed work does a 8-level Wavelet Decomposition of the EEG signals and then uses a subset of the coefficients obtained unlike the research in Sharmila et al. [41] which uses all the coefficients obtained from Wavelet Decomposition. The proposed work achieves a greater accuracy because of selecting a subset of the coefficients which helps avoid redundancy. The coefficients considered in the proposed work are the Detail Coefficients D_3, D_4 and D_5 and the Approximation Coefficient A_8. **The reason for selecting these four coefficients is because the most important information related to epileptic activities are found to be present in these frequency sub-bands of Gamma and Delta** [19],[36]. The proposed

method, results and comparisons along with the insights into them will be discussed in the coming chapters eventually.
The novelty of the proposed work lies in the feature selection process. The wavelet coefficients used in the proposed work are the Detail Coefficients D_3, D_4 and D_5 and the Approximation Coefficient A_8 unlike the existing works which use all the wavelet coefficients.

1.4 Scope of the work

The scope of the work is to classify the EEG signals into healthy, inter-ictal and ictal classes using Machine Learning Algorithms like SVM, Decision Tree, Random Forest, Nearest Neighbors, Naive Bayes and AdaBoost classifiers. The data obtained from the University of Bonn(available in public domain) have been used in this work. The Discrete Wavelet Transform is applied on the sampled EEG signals and their statistical standard deviation has been calculated for the four wavelet coefficients obtained after Wavelet Decomposition upto 8 levels. This standard deviation measure is fed as a feature vector to various Machine Learning classifiers after which a comparison is made between them in all the classification problems.
The following work deals with five experiments which are the following:

1. Classification between Healthy Patients(with eyes open)(set A) and Patients experiencing seizure activity (set E)

2. Classification between Healthy Patients(with eyes closed)(set B) and Patients experiencing seizure activity (set E)

3. Classification between Healthy Patients(with eyes closed)(set B) and Patients in inter-ictal state whose EEG readings obtained from hippocampal formation of opposite hemisphere of the brain(set C) and Patients experiencing seizure activity (set E)

4. Classification between Healthy Patients(with eyes closed)(set B) and Patients in inter-ictal state whose EEG readings obtained from hippocampal formation which was the epileptogenic area(set D) and Patients experiencing seizure activity (set E)

5. Classification between Healthy Patients(with eyes open)(set A) and Patients in inter-ictal state whose EEG readings obtained from hippocampal formation of which was the epileptogenic area(set D) and Patients experiencing seizure activity (set E)

The proposed work is then compared with the state of the art among which the work closely resembles Sharmila et al. [41] and Guo et al. [17]. This establishes our approach.

1.5 Organization of the Thesis

The organization of the Thesis is as follows: **Chapter 2** discusses the Related Works in the field of Epileptic Seizure Prediction using various Wavelet Decomposition techniques and then feed the features extracted to Machine Learning Algorithms like Neural Networks, SVM etc. The features used are usually obtained by calculating the Approximate Entropy or Maximum value of the wavelet coefficients. **The proposed work computes the standard deviation of the coefficients and then feeds it to the classifiers.**

Electroencephalography in **Chapter 3** with an understanding of Epileptic Seizures followed by a summary on Electroencephalography(EEG) which discusses the device and its electrodes briefly. The chapter ends with some EEG signals.

Chapter 4 discusses the EEG dataset followed by some performance metrics used to evaluate the performance of the seizure detector.

Chapter 5 describes the Wavelet Transform, its comparison with Short-Time Fourier Transform followed by examples on Approximation and a discussion on Multiresolution Analysis.

Chapter 6 describes the Continuous Wavelet Transform followed by a short study on Comparison among the Fourier Transform, Short-time Fourier Transform and Wavelet Transform.

Chapter 7 discusses the Discrete Wavelet Transform, its concept of Subband coding through which the process of Wavelet Decomposition is explained followed by an example of Subband Coding. The chapter concludes with a summary of Daubechies Wavelet, which has been used in the proposed work.

Chapter 8 discusses the Proposed Method and describes briefly the Classifiers used.

Chapter 9 describes the Results obtained from the five experiments in the proposed work followed by important conclusions and insights from the results.

The thesis concludes with Conclusion and Future Work on this topic in **Chapter 10**.

Chapter 2

Related Works

As the seizure event detectors were developed, the research began[16]. Initially, the seizure event detectors were patient non-specific. Naturally, for reasons discussed above the accuracy was poor. Then to improve the performance, researchers came up with patient-specific event detectors. These detectors were better in comparison because seizure and non-seizure states across an individual aren't much distinguishable.

Among the earliest patient non-specific seizure event detector designed was by Gotman(1982)[16]. The Gotman Algorithm was driven by the search of sustained rhythmic activity in the brain with a frequency lying in the range of 3-20 Hz and amplitude thrice the value in the background signal. A seizure event was detected by the algorithm whenever a rhythmic activity was recorded on at least two electrodes persisting for 4 seconds.

So far the Gotman algorithm could successfully detect activities below 20 Hz. But its disadvantage was soon exposed when it had to detect seizure with a mixture of frequencies or those with low amplitude high frequency activities. In a paper[37] it was found out the Gotman algorithm only detects 50% of the test seizures. In that paper, 28 patients with a total of 126 seizures were taken as the data. The algorithm produced 0.5 false detections per hour.

Clearly after this research there was a need to work more on the signal processing part to characterize that a particular signal is rhythmic. One such effort was the Reveal Seizure Detector by Wilson[51]. This algorithm focused on decomposition of EEG signals on a time-frequency domain by taking 2-second EEG Epoch. The Reveal algorithm then uses Artificial Neural Networks whether the features obtained from the electrodes are consistent with the seizures already detected on the patient.

The Wilson paper reported that 76% of the test seizures from the dataset of 672 seizures comprising of 426 persons. Also, it only produced 0.11 false detections per hour. On further improvement, the sensitivity increased to 78%. The Wilson algorithm had better sensitivity but it had poor specificity when studying the abnormal non-seizure activity of patients with scalp EEG.

A patient non-specific seizure event detector was developed[37]. The concept of wavelet decomposition was used for feature extraction to measure the probability of an impending seizure. Whenever the probability exceeded a threshold decided apriori that would mean the algorithm declare an onset of seizure. It could successfully identify 78% of the test seizures with 0.86 false detections per hour. The research identified the failure to identify the seizures was due to focal activity, mixed frequency or short duration; It might be due to intense eye movements and chewing that the false detections have occurred.

It was followed by the first patient-specific seizure onset detection algorithm[33, 34, 35]. It uses nearest-neighbor classifier to classify a feature-vector to be belonging to a seizure class or non-seizure class. The training on the classifier is done from the feature vectors that are already labelled. The features included are seizure's average amplitude, highest frequency and rhythmicity. A seizure is declared if the classifier choses half of the channels on which the seizure turned out to be positive. This method did predict 100% of the test seizures with a delay of 9.35 seconds and the false detections were 0.03 per hour. The non-seizure data on which false detection was studied was made by concatenating epochs of EEG obtained from daily brain waves of the patient at regular intervals. When compared to Gotman's work, this work had much better specificity and sensitivity, but the detection latency was still questionable.

Meier developed a patient non-specific seizure detector which was seizure-specific[24]. The research classified seizure into 6 classes based on the dominant frequency that appears on the onset of the seizure. Then for each type a Support Vector-Machine was trained. It was then seen whether the feature vector extracted from an EEG-epoch was on the verge of being a seizure activity. A single feature vector consisting of all the features was constructed. Its performance was then evaluated. 91 seizures and 1,360 hours of non-seizure EEG of 57 patients was taken for the purpose. It detected 96% of the test seizures. Average lag time was 1.6 seconds and false detections were 0.45 per hour. But as No Free-Lunch would have it, it came with a shortcoming in the very classification of the frequency it had done. If the seizure type was not already included among the 6 categories there was a very high probability that the seizure would not be recognized at all.

In other studies, The SVM has been the most common classifier to distinguish between seizure and non-seizure events. Using the CHB(Children's Hospital of Boston) database and a patient-specific prediction methodology, a research [42] by A.Shoeb did his PhD Thesis on this topic where he used SVM on a dataset with 24 patients. The classification accuracy 96% with a false positive rate of 0.08 false detections per hour. Something similar research was done, five records from the CHB database were taken and study was conducted using the Linear Discriminant Analysis[22]. The classification accuracy achieved was 91.8%.

Acharya et al used entropy for his work on EEG and seven different classifiers were used among which the best performance was by Fuzzy Sugeno Classifier. It achieved a classification accuracy of about 98.1%. The worst performing was the Naïve Bayes Classifier which achieved 88.1% classification accuracy [1].

Nasehi and Pourghassem worked on the same dataset with a Particle Swarm Optimization (a genetic algorithm). The Neural Network gave 98% for sensitivity and false detections were 0.125 per hour[27].

Yuan et al. gave a patient-specific seizure detection system and trained a neural network. The system was trained on 21 seizure records and tested on 65 records. The accuracy achieved were an average of 91.2% for sensitivity and 95% for specificity and the overall accuracy achieved was 94.9% [53].

Patel et al. gave a classification algorithm consuming low power which could classify rhythmic activities as seizures. The FRE Dataset was used. The study compared between multiple classifiers like Linear Discriminant Analysis(LDA), Quadratic Discriminant Analysis, Mahalanobis Distance Classifier and SVM. It did the study on 13 samples. LDA gave the best accuracy 87.7%. Overall the average accuracy was 76.5% [31].

In a recent study by Fergus et al.[11] the previous ideas have been explored further. The fact is that EEG capturing the data and the time taken by the experts to interpret the data is time taking. So, Automated detection of correlation of seizure states across the brain has been explored. The dataset used contained 342 records(171 discharge activities and 171 non-discharge activities). The k-NN Classifier was used and 93% accuracy was achieved with this method.

In a research conducted on neo-natal patients [46], the system was able to report an accuracy of 89% with 1 false detection per hour, 96% with two false detections per hour, 100% with four false detections per hour. The classification system created allows the control of the final decision by choosing the confidence factors according to which the false detections vary.

Most of the recent works are based on classification between healthy patients with eyes closed and patients experiencing seizure activity. The dataset used is described in chapter 4. Research has been done suggesting the use of a hybrid SVM [44]which has been optimized by the use of a Genetic Algorithm(GA) and Particle Swarm Optimization(PSO). Das et al. [9] have used SVM where the Normal Inverse Gaussian (NIG) parameters of the frequency sub-bands have been used as features giving a maximum accuracy of 100%.

Sharma et al. [40]have introduced Analytic Time-Frequency Flexible Wavelet Transform(ATFFWT) for EEG signal decomposition and Fractal Dimensions have been calculated for each sub-band which is fed to a Least Squares-SVM(LS-SVM) to obtain a maximum of 100% accuracy on some datasets.

In other techniques[5], tunable-Q wavelet transform(TQWT) has been used to decompose the signal following which a Quality Factor(Q) based entropy value of the sub-bands obtained have been fed to a Support Vector Machine to get a maximum accuracy of 100%. Other works [29] include using Discrete Wavelet Transform to decompose the EEG signals and calculation of Approximate Entropy(ApEn) of the detail and approximate coefficients obtained at each level of decomposition. Kumar et al [23] used a wavelet-based fuzzy approximate entropy(fApEn) method. The fApEn values of the sub-bands are calculated and the feature vectors are fed for classification. The highest accuracy obtained is 100%. Salem et al. have used a three-stage algorithm [38] consisting of Signal Decomposition followed by Feature Extraction and finally Classification. The extracted features are fed to an Ant-Colony Classifier to finally achieve detection rate of 100%. Ocak et al. [29] have used DWT followed by calculating Approximate Entropy values where a threshold value was chosen. The difference in Approximate Entropy between epileptic and normal EEG helped detect seizures with upto 96% accuracy.

Wani et al.[50] worked on the multi-class classification problem of seizure detection. Given an EEG epoch the task was to classify it into one of the three classes: 1. Healthy, 2. Inter-Ictal and 3. Seizure Activity. The highest accuracy achieved was 95%.

Ullah et al.[49] proposed a P-1D-CNN system with very less learnable parameters providing an accuracy of $99.1 \pm 0.9\%$. Guo et al. [17] used ApEn for EEG analysis to obtain an accuracy of 99.85%. While, Srinivasan et al [43] used Probabilistic Neural Networks by calculating ApEn values of the sub-bands to get 100% accuracy. Nicolau et al [28] used Permutation Entropy(PE) as a feature for SVM classifier. The Table 2 shows the related

Authors	Method Used	Type Of Experiment	Accuracy
Tzallas et al. [48]	Time frequency Features using ANN	A-E ABCD-E	100% 97.73%
Guo et al. [18]	DWT and line length feature using ANN	ABCD-E	97.77 %
Orhan et al. [30]	DWT and clustering using multilayer perceptron ANN	A-E ABCD-E	100% 99.60%
Gandhi et al. [14]	DWT and energy , std and entropy features;using SVM and Probabilistic NN	ABCD-E	95.44 %
Nicolau et al. [28]	Permutation Entropy and SVM	A-E B-E C-E D-E ABCD-E	93.55% 82.88% 88.00% 79.94% 86.10%
Kaya et al. [21]	1D LBP and functional tree 1D LBP and BayesNet	A-E D-E CD-E	99.5% 95.5% 97%
Samiee et al [39]	Rational Discrete Short time fourier transform using multilayer perceptron	A-E B-E C-E D-E ABCD-E	99.80% 99.30% 98.50% 94.90% 98.10%
Peker et al. [32]	DTCWT using complex valued NN	A-E ABCD-E	100% 99.15%
Swami et al [45]	DTCWT and energy, std, Shanon entropy features using GRNN	A-E B-E C-E D-E AB-E CD-E ABCD-E	100% 98.89% 98.72% 93.33% 99.18% 95.15% 95.24%

Authors	Method Used	Type Of Experiment	Accuracy
Sharma et al .[40]	ATFFWT and FD feature using LS-SVM	A-E	100%
		B-E	100%
		C-E	99%
		D-E	98.50%
		AB-E	100%
		CD-E	98.67%
		AB-CD	92.50%
		ABCD-E	99.20%
Tiwari et al .[47]	Key point based LBP and SVM	ABCD-E	99.3 %
Chen [8]	DTCWT and Fourier features with NN classifier	A-E	100%
		ABCD-E	100%
Bajaj and Pachori [4]	Amplitude and frequency modulation bandwidths of IMFs and Least Squares-SVM	ABCD-E	99.5 %
Yuan et al .[52]	ApEn, hurst Exponent, scaling exponents of EEG and Extreme Learning Machine Algorithm	ABCD-E	99.5%
		D-E	96.5%
Bhattacharya et al .[5]	TQWT-based multi-scale K-NN Entropy	A-E	100%
		B-E	100%
		C-E	99.5%
		D-E	98%
Das et al. [9]	SVM using NIG parameters as features in dual-tree complex wavelet transform domain	A-BCDE	99%
		ABCDE	100%
		A-E	100%
		AD-E	100%
		D-E	100%
		C-E	100%
A.Sharmila P.Geethanjali [41]	Naïve Bayes/k-NN Classifiers	A-E	100%
		B-E	99.25%
		C-E	99.62%
		AB-E	99.16%
		AC-E	99.50%
		BC-E	98.25%
		CD-E	98.75%
		ABC-E	98.68%
Proposed Work	Discrete Wavelet Transform, SVM Classifier	A-E	100%
		B-E	100%
		BC-E	98.45%
		AD-E	95%
		BD-E	96.87%

Table 2.1: Results and Features used in Previous and proposed method using the same EEG dataset

works in seizure detection which have used the dataset from University of Bonn, the very same that has been used in the proposed work. Some of the methods use DWT for Wavelet Decomposition, however the proposed work closely resembles the work done by Sharmila et al. [41], achieving better results than state of the art.

2.1 Conclusions

From the Table 2.1, it is evident that the works by Sharmila et al. [41] and Guo et al. [17] give the best results. But these works use all of the wavelet coefficients obtained after Wavelet Decomposition. On the other hand, the proposed work uses a subset of the 9(nine) wavelet coefficients obtained which requires choosing only those coefficients which have minimum redundancy whilst having the most important information present in their window.
Infact, the proposed work beats the state of the art accuracy in the Multi-class classification experiment of classifying an EEG epoch into healthy,inter-ictal and epileptic. The accuracy obtained in the work done by Sharmila et al. [41] is 98.25% while the proposed work achieves a higher accuracy of 98.45%

Chapter 3

Electroencephalography And Epileptic seizures

This chapter reviews the features and summary of epileptic seizures and how the Electroencephalogram uses the electrographic properties of seizures for therapeutic and diagnostic purposes.

3.1 Seizures

To understand the physiology underlying epileptic seizures we have to start looking from neurons which are the cells within our brain. They can generate and send electric signals to each other. Neurons interconnect among each other to form a neural network. Neurons receive messages which can be inhibitive or excitatory in nature. In the former the activity in the neuron ins suppressed while in the latter activity is encouraged [20]. When patients suffer the epileptic seizure, these are recurring periods of hyperactivity of cortical neurons which are caused by negative shifts in potential. This involves a massive number of neurons in one or more neural networks. The seizure states are separated by periods of transition of states which creates an imbalance causing excitation to arise instead of inhibition. This is primarily caused in these patients because they specifically have defects in neural networks organization. The disorders may be due to genetic reasons or from a shock to the Central Nervous System. Epileptic seizures can be divided into many types based on which part of the brain their origin is. Focal seizures originate in a small region of the brain's cortex and it has been clinically proven that they affect that part during seizure activities. To explain this, the best scenario would be to consider the temporal lobe of the brain which processes emotions and short-term memory. So when focal seizures occur in the temporal lobe it causes feelings such as euphoria, fear meaning all the negative emotions of hallucinations etc. When Focal seizures spread to adjacent areas of the brain, it causes convulsions in the whole-body. There are generalized seizures which affect the entire cerebral cortex of the brain. The generalized seizures often result in loss of consciousness. The general signs in the patient are jerking or atonic seizure.

A seizure that starts as a focal seizure and then gradually starts affecting not only the temporal lobe but also the entire cerebral cortex and becomes a generalized seizure is called a secondarily generalized seizure.

3.2 Electroencephalography(EEG)

EEG is the measure of the ionic current that flows from the neurons in our brain and the electrical activity is recorded on the EEG. The clinical EEG equipment are used for patients suffering from ADHD, Alzheimer disease, Epilepsy. Usually the device kept is called the scalp EEG. The electrodes are used to measure the electrical potentials generated by the neurons. The EEG device which has the electrodes arranged symmetrically on the scalp which is used to get a temporal and spatial summary of the synchronous excitation of numerous neurons within the brain can be seen in Figure 3.1. Odd numbered electrodes are place on the left side of the scalp and even

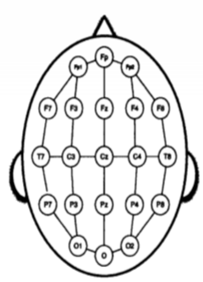

Figure 3.1: EEG Electrodes placed symmetrically on the scalp

numbered on the right. An EEG Signal is taken by measuring the potential difference measured between two electrodes. Consider, the channel FP2-F8. It is formed by taking the potential difference between the potential at FP2 and that in F8. Each EEG channel records the ionic electrical activities of a region of the brain such as the channel FP2-F8 records the activity of the frontal lobe of the right hemisphere.

In case of a seizure in the focal region of the brain, only a few EEG channels that lie in the proximity of the site of origin of the seizure are active. But when a generalized seizure occurs all the EEG channels become active. The neurons lying closest to the surface of the scalp are most responsible for the scalp EEG. Also, there are brain fluids which inhibit the amplitude of higher frequency brain waves.

So, the problem that arises with this is that the small, deep region within the brain involved in a seizure cannot be detected by using scalp EEG.

EEG describes the ionic activities in the brain based on the portion of the brain whether its frontal, posterior, lateral and bilateral lobe of the brain.

An EEG wave has a delta component if the highest frequency is less than 4Hz, a theta component if it's 4-8 Hz, an alpha component when it is 8-12 Hz, a beta component if the dominant frequency is from 12-30 Hz or a brain wave whose gamma component when its greater than 30 Hz. Scalp EEG depends a lot upon the dominant frequency of the brain waves and on the spatial features. It is however different for a person when awake and during sleep. EEG readings can be foiled by sweating, chewing, rapid eye-movements. The figure 3.2 shows 10 seconds of awake EEG followed by 37 seconds eye-blinking causing deflection in the EEG channels FP1-F7, FP1-F3, FP2-F4,FP2-F8.

The figure Fig 3.3 shows sample EEG readings while patient is asleep. The activity between 12-14 seconds is called sleep spindle. While the figure 3.4 shows 10 seconds of awake EEG followed by chewing causing physiological artifact in scalp EEG.

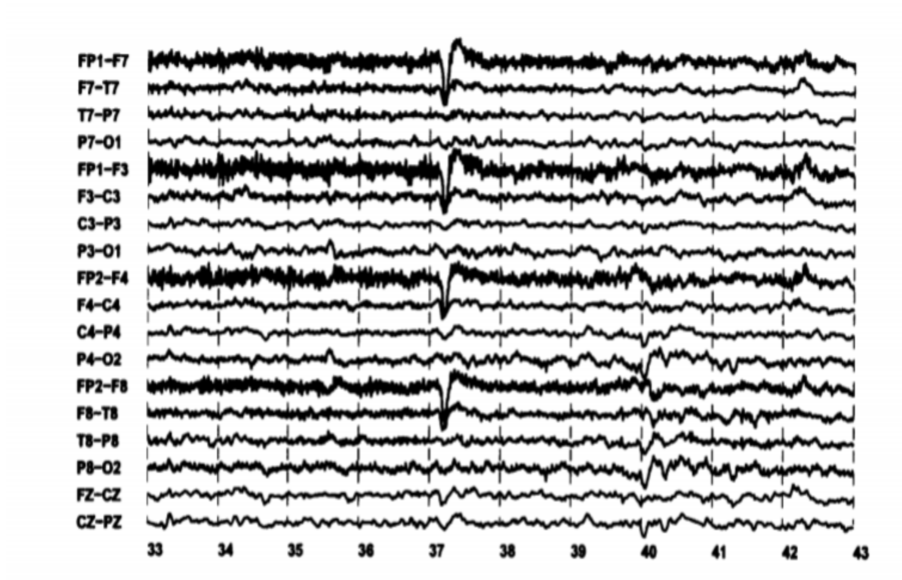

Figure 3.2: 10 seconds of awake EEG followed by 37 seconds eye-blinking

3.2.1 Seizures within the scalp Electroencephalogram

Within the scalp EEG, the occurrence of seizures are exhibited as a spike increase in the spectral energy in the brain. That happens because there are spectral components within the patients which start increasing or decreasing(depending upon the patient). This has been noticed it occurs with appearance or disappearance of frequency components within the range 0-25Hz [15]. Along with this variability it can also be said that there is variability in the EEG channels where the spectral energy varies across patients.

Let us look at an example. In Figure 3.5 Consider patient A whose seizure begins at 1723seconds and consists of flattening of the waves. Then for a few seconds the amplitude of this rhythm increases as its frequency decreases. The figure 3.6 is depicting the seizure of patient B whose seizure begins at 6313 seconds. The rhythmic activities can be clearly seen on the channels F7-T7, T7-P7. It is now known that the rhythmic activities are visible on a scalp EEG while the seizure is occurring, however the spatial and spectral energies are variable across patients. The figure 3.7 shows an excess spike in the EEG of the epileptic patient. It can be easily observed from 2884-2892 seconds that there are spikes. Although these spikes are present on the EEG of the patient A these are not a result by the physical activities associated

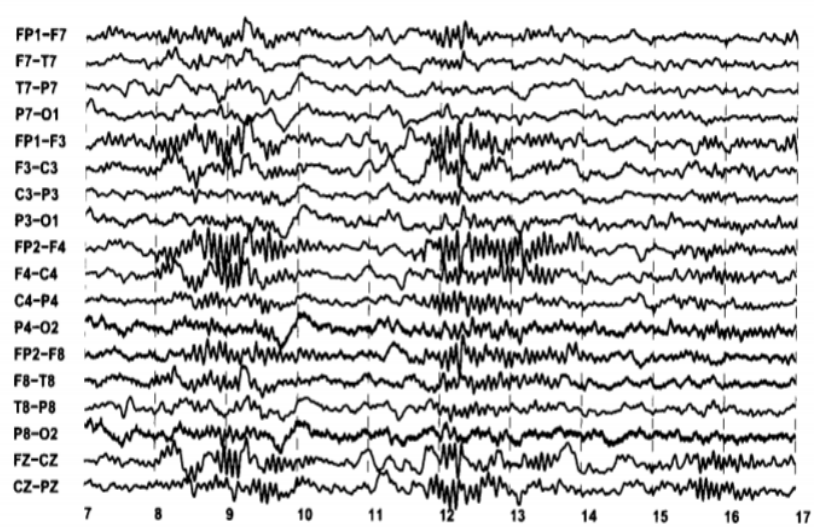

Figure 3.3: Demonstration of 10 seconds of sleep EEG. Sleep Spindles occurring at 12-14 seconds

with Patient A's attack in Fig 3.5. Whereas in the next figure, the abnormal discharge is a result with physical activity is visible in the Figure 3.7.
There are many examples like this to illustrate that the spectral and spatial features during seizure and non-seizure states are variable across patients. So, clearly this is the primary reason why patient non-specific seizure detectors have poor sensitivity and specificity [16]. Experts in the study of EEG have observed that if a patient has not suffered any brain disorder it is highly likely that the seizures recorded months apart would exhibit very similar spectral and spatial features. Seizure Attacks are recurring periods of neural network malfunctioning. The symptoms are variable across patients depending on whether it is a generalized seizure or focal seizure. The scalp and the intracranial EEG measure the ionic activity of neurons are used to detect seizures. The scalp EEG has poor spatial resolution but high spatial coverage. On the other hand, the intracranial EEG has the opposite features.

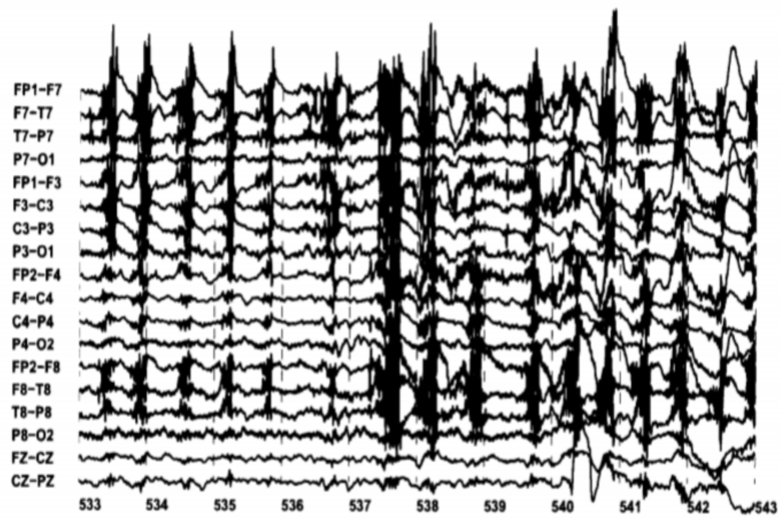

Figure 3.4: Demonstration of 10 seconds of awake EEG followed by chewing

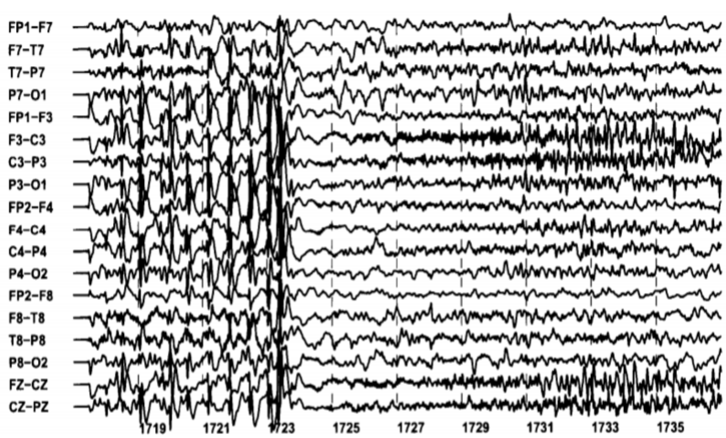

Figure 3.5: Example of a seizure as recorded on a scalp EEG of a patient.

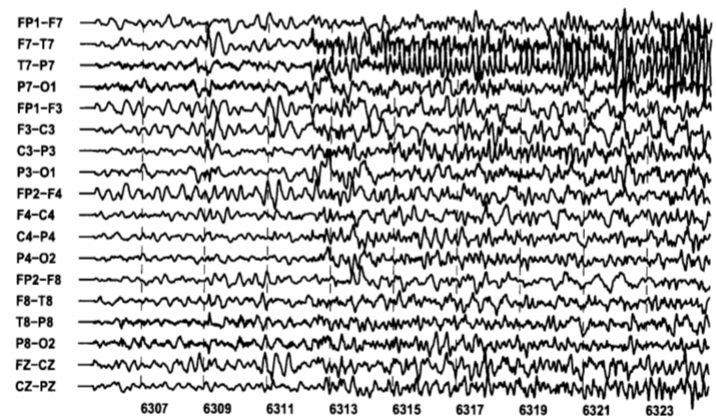

Figure 3.6: Example of a seizure of another patient as seen on the scalp EEG.

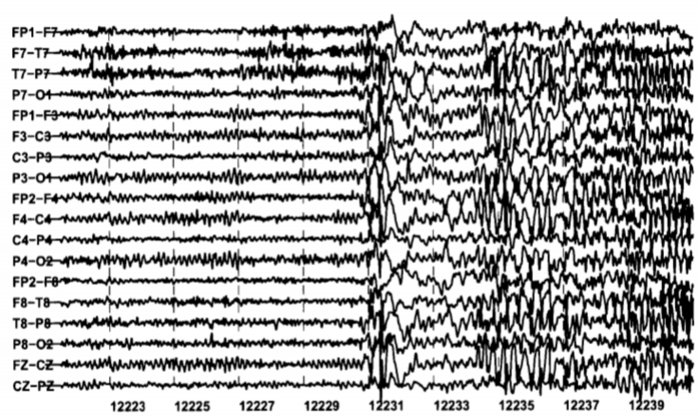

Figure 3.7: Seizure recorded within the scalp EEG where the onset of the seizure is accompanied by a spike

Chapter 4

EEG Data and Testing

In this chapter, the scalp EEG Dataset will be discussed along with the testing. We find the latency, sensitivity and specificity of the patient-specific seizure onset detector.

4.1 EEG Dataset

The dataset [3] used in our work is obtained from the Department of Epileptology of the University of Bonn. Each data set consists of 100 single channel EEG epochs. It consists of EEG recordings having a sampling rate of 173.61 Hz for the duration of 23.6 seconds. It has 5 sets of data consisting of 100 EEG recording for each set. Set A and B are for healthy patients with the eyes open and close respectively. Set C and D are for the Interictal case of epilepsy. The EEG signals in in sets C were recorded from a hippocampal formation of an opposite hemisphere of the brain while the EEG signals in set D were recorded from the hippocampal formation which was the epileptogenic area. Set E is for Ictal or the patient having a seizure attack.

4.2 Performance Metrics

There are three metrics used to characterize the performance of our seizure onset detector:

1. **ELECTROGRAPHIC SEIZURE ONSET DETECTION LATENCY** $EO_{LATENCY}$
   On the EEG, the detection of a seizure is usually indicated by the

changes in potential across scalp EEG. However the clinical definition of a seizure is that the physical symptoms start occurring. $EO_{latency}$ is defined as the time elapsed between the onset of the seizure on EEG and electrographic onset. It is usually $>= 0$.

2. **SENSITIVITY**
   We have used the term called 'sensitivity' a lot in the previous chapters. Sensitivity is defined as the percentage of test seizures detected by the seizure onset detector.

3. **FALSE ALARMS PER HOUR** It is defined as the number of times, for one hour, that the seizure onset detector reports a false positive, meaning that it detects n epileptic seizure whereas it is actually not.

4.3 Performance Metric Measurement

The performance metric measurement is used to evaluate the performance of the seizure detector. Let us assume that N_{NS} denotes the number of non-seizure samples and N_S denote the number of seizure samples.

4.3.1 Measuring $EO_{latency}$ and S

The detector is trained using "leave-one-out" cross-validation technique. In this method, the detector is trained on all the non-seizure records of the patient and all but one seizure records. The seizure onset detector is then given the task of detecting the seizure record which was left out. This process is repeated N_s times and each time one of the N_s seizure records is considered as the test data. Let $S_m \in \{ 0,1 \}$ be a binary variable where 0 means no seizure and 1 means a seizure, let $EO_{latency,m}$ denote the latency with which the detector records the Electrographic Seizure Onset Detection Latency. Let $FA_{s,m}$ be the number of false alarms given out while testing the m^{th} seizure record. Let K denote the total number of detected seizures. The following equations relate the quantities with the performance of the detector:

$$S = \frac{1}{N_s} \sum S_m \tag{4.1}$$

$$EO_{latency} = \frac{1}{K} \sum (S_m * EO_{latency,m}) \tag{4.2}$$

4.3.2 Measuring the False Alarms FA

The training is done on all the seizure records and all but one non-seizure records. It is then made to test the left-out non-seizure record. Now this process is repeated N_{ns} times where in each of the time one non-seizure record is left out of the training data and then taken as the testing data.

The equation used to calculate the number of False Detections of the detector:

$$FA = (\frac{1}{N_s + N_{ns}}) * (\sum FA_{ns,n} + \sum FA_{s,m}) \tag{4.3}$$

Chapter 5

Wavelet Transform

5.1 Overview

5.1.1 Introduction

The Fourier Transform is useful for converting time domain signal to frequency domain signals. But the standard Fourier Transform is only localized in frequency, therefore it becomes difficult to tell at what particular time a particular frequency exists. The Short- Time Fourier Transform(STFT) gives information on both the time and frequency with a problem that since the window used is small, the resolution in frequency is limited.

The problem with STFT lies in the **Heisenberg's Uncertainty Principle**. The problem is simply that the exact time-frequency information of a signal is not possible.

In Fourier Transform, it is known exactly which frequencies exist and the value of the signal at every time instance is also known. Thus there are no frequency resolution and time resolution problem. The window length is shorter in STFT, which creates the frequency resolution problem. The problem that arises is that there is no more information about what frequencies exist, rather only a band of frequencies that exist is known. To summarize the scenario, using a narrow window improves the time resolution of the signal at the cost of frequency resolution. The problem to the dilemma of the choice of the window is application dependent. If the frequency components are well separated, it is a good choice to look for good time resolution in this case.

It seems that the solution to the above problem is Wavelet Transform. It uses small wavelets on which scaling and shifting is applied which allow to analyze the signal in various scales and in various locations respectively.

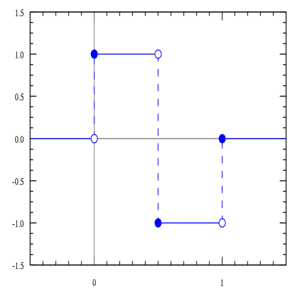

Figure 5.1: The Haar Wavelet

5.1.2 History

The first publication on wavelet transform dates back to 1909, where a mathematician named Alfréd Haar proposed the Haar Wavelet. The term wavelet was invented in the year 1984. For a long time, the Haar Wavelet remained the only known orthogonal wavelet until in 1985, Yves Meyer proposed the second orthogonal wavelet called the Meyer Wavelet.

In 1988, Mallat and Meyer brought the concept of Multiresolution.

In 1989, the Fast Wavlet Transform was introduced by Mallat which promised several applications in the signal processing domain.

5.2 Approximation Theory and Multiresolution Analysis

The Heisenberg's uncertainty principle gives rise to time resolution and frequency resolution problems. The alternative approach to analyze any signal is **Multiresolution Analysis(MRA)**. MRA does not analyze every frequency component equally.

MRA is designed in a way such that for high frequencies, it gives good time resolution and for lower frequencies, it gives good frequency resolution. This makes sense when we realize that the signals that exist in applications have higher frequency components which exist only for a short duration and the lower frequency components exist for longer durations.

5.2.1 A Simple Approximation Example

Consider a periodic function $x(t)$ with period T

$$x(t) = 1 - \frac{2|t|}{T}, \quad |t| < T. \tag{5.1}$$

We can decompose a periodic signal with period T into higher order harmonics,

$$x(t) = \sum_{k=-\infty}^{\infty} a_k exp(j\frac{2\pi kt}{T}) \tag{5.2}$$

The following analysis equation describes the orthogonal property of exponential equations having complex parts,

$$a_k = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)exp(-j\frac{2\pi kt}{T})dt \tag{5.3}$$

The Fourier Series coefficients are

$$a_k = \frac{\sin^2(\pi k/2)}{2(\pi k/2)^2} \tag{5.4}$$

The physical significance of calculating the Fourier Series coefficient is that it indicates the amplitude and the phase of the higher order harmonics, indexed by k. The higher the value of k, the higher frequency it approximates.

Intuitively, as k approaches infinity, the reconstructed signal resembles the original.

5.2.2 Abstract Idea in the Approximation Example

A signal can be decomposed into linear combination of the basis signals given by,

$$f(t) = \sum_k a_k \phi_k(t) \tag{5.5}$$

where a_k are expansion coefficients and the $\phi_k(t)$ are expansion functions. If we choose the basis function correctly, there exists its dual function also. It is referred by $\phi_k^D(t)$. $\phi_k(t)$ and $\phi_k^D(t)$ are orthonormal. The inner product is given by

$$< \phi_k(t), \phi_k^D(t) >= \int \phi_k(t)\phi_k^D(t)dt = \delta_{ij} \tag{5.6}$$

The expansion coefficients are given by:

$$\begin{aligned} < f(t), \phi_k^D(t) > &= \int f(t)(\phi_k^D)^*(t)dt \\ &= \int (\sum_{k'} a_{k'}\phi_{k'}(t))(\phi_k^D)^*(t)dt \\ &= \sum_{k'} a_{k'}\delta_{k'k} \\ &= a_k \end{aligned}$$

So, the expansion coefficients can be defined by the following equation:

$$a_k =< f(t), \phi_k^D(t) >= \int f(t)(\phi_k^D)^*(t)dt \tag{5.7}$$

It is very important that a good choice of basis function and its dual is made. Usually, some of the expansion coefficients of are critical values and some decay to zero. This property helps in data compression while maintaining the resemblance to the original signal also.

5.3 Example about Multiresolution

5.3.1 Approximate discrete-time signals using delta function

Consider a discrete-time signal

$$x[n] = (\frac{1}{2})^{|n|} \tag{5.8}$$

Now, the next task is to find the expansion coefficients of $x[n]$. To do that, a basis function and its dual have to be chosen. Thereafter, we check if the orthonormal property holds true:

$$< \phi_k(t), \phi_k^D(t) >=< \delta[n-i], \delta[n-j] >= \sum_{n=-\infty}^{\infty} \delta[n-i]\delta[n-j] = \delta_{ij} \tag{5.9}$$

From equation 5.7, the expansion coefficients come out to be

$$a_k = < x[n], \delta[n-k] > = \sum_{n=-\infty}^{\infty} (\frac{1}{2})^{|n|} \delta[n-k] = (\frac{1}{2})^{|k|} \tag{5.10}$$

The above example demonstrates the translations of the delta function $\delta[n]$. $\delta[n-k]$ would mean that the impulse is at $n = k$. the reconstruction using the expansion coefficients is straight forward. For instance, if we wish to know the signal when n $\in$ [-1,0], the coefficients a_{-1}, a_0 are used for reconstruction of the original signal in that range.

5.3.2 Reconstruction Using Scaling

Consider a continuous function $\phi(t)$ such that it is defined by,

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

The single delta function is a rectangular function of width 1. Its scaled

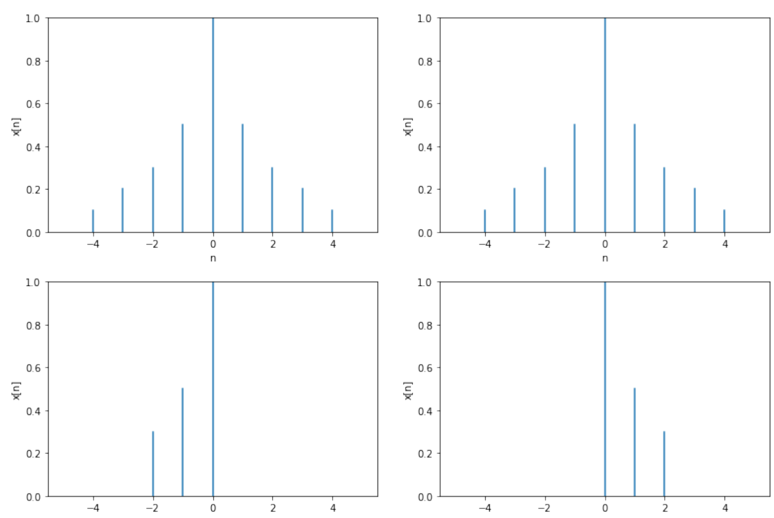

Figure 5.2: The result of approximation using delta functions. (a) Original signal $x[n]$ (b) Obtained coefficients a_k (c) Reconstructed signal $x_1[n] = \sum_{k=-2}^{0} a_k \delta[n-k]$ (d) Reconstructed signal $x_2[n] = \sum_{k=0}^{2} a_k \delta[n-k]$

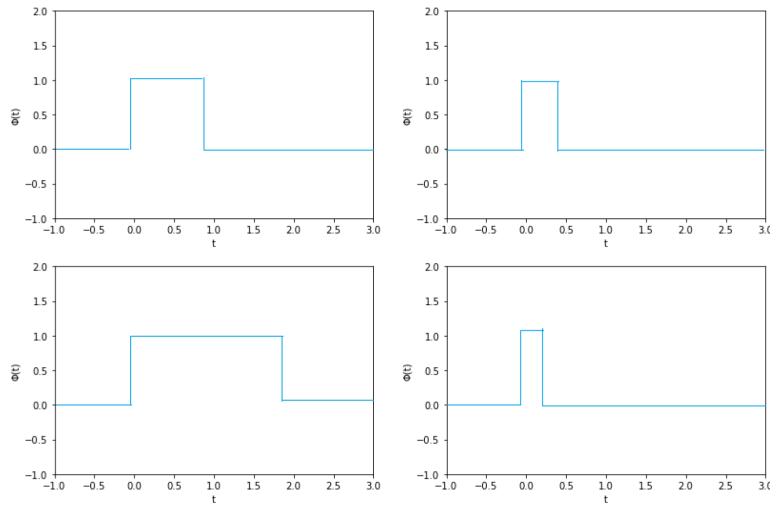

Figure 5.3: Plot of the scaled version of the basis. (a)$\phi_n(t) = \phi_0(t)$ (b) $\phi_n(t) = \phi_1(t)$ (c) $\phi_n(t) = \phi_{-1}(t)$ (d) $\phi_n(t) = \phi_{-2}(t)$

version is given by

$$\phi_s(t) = \phi(st) \tag{5.11}$$

For brevity purposes, let $s = 2^n$ where n is an integer. Unfortunately, the scaled versions of the function $\phi_s(t)$ are not orthogonal to each other. To obtain orthonormal basis functions from $\phi_n(t)$, we apply

$$\phi'_0(t) = \phi_0(t) = \phi(t)$$

$$\begin{aligned}
\phi'_1(t) &= \phi_1(t) - \frac{< \phi_1(t), \phi_0(t) >}{< \phi_0(t), \phi_0(t) >}\phi_0(t) \\
&= \begin{cases} 1/2 & \text{if } 0 \leq t < 1/2 \\ -1/2 & \text{if } 1/2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases} \\
&= \psi(t)/2
\end{aligned}$$

This process is continuously applied to extend the basis.It is mathematically found out that $\phi(t)$ has more concentration at lower frequencies while $\psi(t)$ has more concentration at high frequencies. $\phi(t)$ is called the Scaling Function which gives the approximation coefficients while $\psi(t)$ is called the Wavelet Function which gives the detail coefficients.

5.3.3 Multiresolution Analysis

In the section 5.3.1, an orthonormal function was found corresponding to its translated version. Various Frequency resolutions can be achieved by varying the scaling versions of the original signal.

With the above two properties, the basis functions can be constructed using the functions $\phi(t)$ and $\psi(t)$. These are defined by:

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \tag{5.12}$$

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{5.13}$$

where j is the scaling parameter and k is the translation parameter. The subspace swept by the Scaling Function and the Wavelet Function:

$$V_j = Span\{\phi_{j,k}(t)\} \tag{5.14}$$

$$W_j = Span\{\psi_{j,k}(t)\} \tag{5.15}$$

Some Observations worth mentioning here as follows:

1. The scaling function does not overlap with the translated version of the scaling function. They are orthogonal.
2. It can be observed from equation 5.3 that $\phi_{-1}(t) = \phi_0(t) + \phi_0(t-1)$. This means the scaling functions at lower scales are enveloped within the scaling functions at higher scales.
3. This helps in representation of functions with arbitrary precision.

Therefore, our example of the Haar Scaling Function is given by

$$\phi(t) = \phi_{0,0}(t) = \frac{1}{\sqrt{2}}\phi_{1,0}(t) + \frac{1}{\sqrt{2}}\phi_{1,1}(t). \tag{5.16}$$

The scaling function $\phi_{j,k}$ can be represented in terms of its scaled version and translated version

$$\phi(t) = \frac{1}{\sqrt{2}}(\sqrt{2}\phi(2t)) + \frac{1}{\sqrt{2}}(\sqrt{2}\phi(2t-1)) \tag{5.17}$$

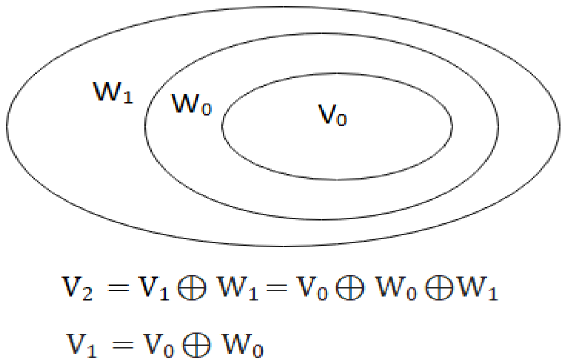

Figure 5.4: The relationship between scaling and wavelet function spaces

The generic form of equation 5.17 is the refinement equation or the dilation equation is given by,

$$\phi(t) = \sum_n h_\phi[n]\sqrt{2}\phi(2t-n) \tag{5.18}$$

The physical significance of equation 5.18 lies in the fact that it relates the scaling function with higher frequency i.e. $\phi(2t)$. Thus, the scaling function $\phi(t)$ can be obtained by applying a low pass filter $h_\phi[n]$.
Similarly, there is a relationship between wavelet functions given by

$$\psi(t) = \sum_n h_\psi[n]\sqrt{2}\psi(2t-n) \tag{5.19}$$

For Haar Wavelets, $h_\phi[n] = \{1/\sqrt{2}, -1/\sqrt{2}\}$ and $h_\psi[n] = \{1/\sqrt{2}, -1/\sqrt{2}\}$. These two filters are related by,

$$h_\psi[n] = (-1)^n h_\phi[1-n] \tag{5.20}$$

V_0 is the approximation when scaling factor $= 0$. To obtain higher order of approximation, the union of the subspaces swept by varying levels of wavelet functions. With the infinite union of these wavelet sets, any set can be represented with arbitrary precision. The total set formed by infinite union is given by

$$L^2(\mathbf{R}) = V_0 \oplus W_0 \oplus W_1 ... \tag{5.21}$$

This shows that any function lying in this set can be decomposed using the Scaling and Wavelet Functions. The advantage of using Wavelet Transform is that unlike in STFT which has constant resolution at all times and frequencies, Wavelet Transform gives good time resolution (and poor frequency resolution) at high frequencies and good frequency resolution (and poor time resolution) at low frequencies. This is beneficial because in case of seizure activity, the EEG signals form spike and slow complex which are characterized by high frequencies. Thus, it is extremely beneficial to get a good time resolution during these discharges. A close look ath Figure 5.5 suggests that

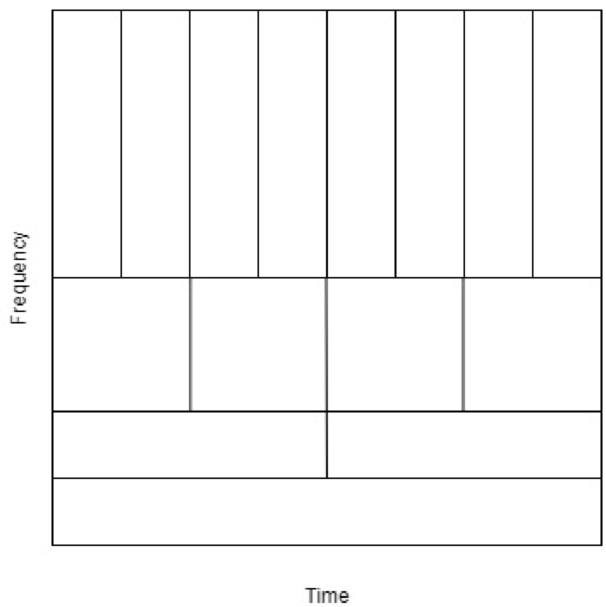

Figure 5.5: The time-frequency resolution in Wavelet Transform

although the heights and widths of the individual boxes change, the overall area is still constant. At low values of frequency, the boxes are shorter in height, thus meaning that at low frequencies the frequency resolution is good but time resolution is poor. At higher frequencies, the boxes become shorter, meaning that the time resolution becomes better and the frequency resolution gets poorer.
In case of STFT, the time resolution and frequency resolution is decided by the length of the window, so for the entire analysis the time and frequency resolutions are constant, thus the time-frequency planes consists of **squares**.

Chapter 6

The Continuous Wavelet Transform

6.1 Introduction

The continuous wavelet transform is an alternate approach to the Short Time Fourier Transform to get around the Resolution Problem. Here, the original signal is multiplied with a function (as in STFT where a small length window is used). There are mainly two differences in both of these approaches:

1. The Fourier transforms of the signals are not computed, so only a single peak will be seen.
2. The width of the window is varied in the wavelet transform.

6.2 Definition

A wavelet function $\psi(t) \in L^2(R)$ is defined as a function which is limited in time, having values in a certain range and zeros elsewhere with zero mean. Therefore,

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{6.1}$$

$$||\psi(t)||^2 = \int_{-\infty}^{\infty} \psi(t)\psi^*(t)dt = 1 \tag{6.2}$$

Having been discussed the translation and the scaling properties, the basis function can be derived from the mother wavelet function,

$$\psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})\Bigg|_{u\in R, s\in R^+} \tag{6.3}$$

where u is the shifting parameter and s is the scaling parameter. Note that $s \in R_+$ because negative scaling is ignored in the wavelet transform. The continuous wavelet transform is given by

$$\begin{aligned} Wf(s,u) &=< f(t), \phi_{s,u} > \\ &= \int_{-\infty}^{\infty} f(t)\psi^*_{s,u}(t)dt \\ &= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})dt \end{aligned} \tag{6.4}$$

It can be observed from the above equation that we can do both the frequency resolution(parameter s) and the time resolution at the same time(parameter u) now.
The inverse wavelet transform is

$$f(t) = \frac{1}{C_\psi}\int_0^{\infty}\int_{-\infty}^{\infty} Wf(s,u)\frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})du\frac{ds}{s^2} \tag{6.5}$$

where C_ψ is

$$C_\psi = \int_0^{\infty}\frac{|\Psi(\omega)|^2}{\omega}d\omega < \infty \tag{6.6}$$

$\Psi(\omega)$ is the Fourier Transform of $\psi(t)$

6.3 Wavelet Transform: An Example

In the previous section, the mother wavelet function was introduced with its definition and properties. To demonstrate an example on Wavelet Transform, the Mexican Hat Wavelet has been taken here:

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}}(\frac{t^2}{\sigma^2} - 1)exp(-\frac{t^2}{\sigma^2}) \tag{6.7}$$

The name "Mexican-hat" comes from the fact that this function resembles a Mexican hat when plotted against time. It is derived from the second derivative of the Gaussian function, $exp(-t^2/(2\sigma^2))$. The wavelet decays fast to

zero because the Gaussian Function drops to zero fast.
The Fourier Transform of this mother wavelet is calculated as

$$\Psi(\omega) = \frac{-\sqrt{8}\omega^{\frac{5}{2}}\pi^{\frac{1}{4}}}{\sqrt{3}}\omega^2 exp(-\frac{\sigma^2\omega^2}{2}) \tag{6.8}$$

The illustration is provided in the figure 6.1. the Ricker Wavelet $r(t)$ along

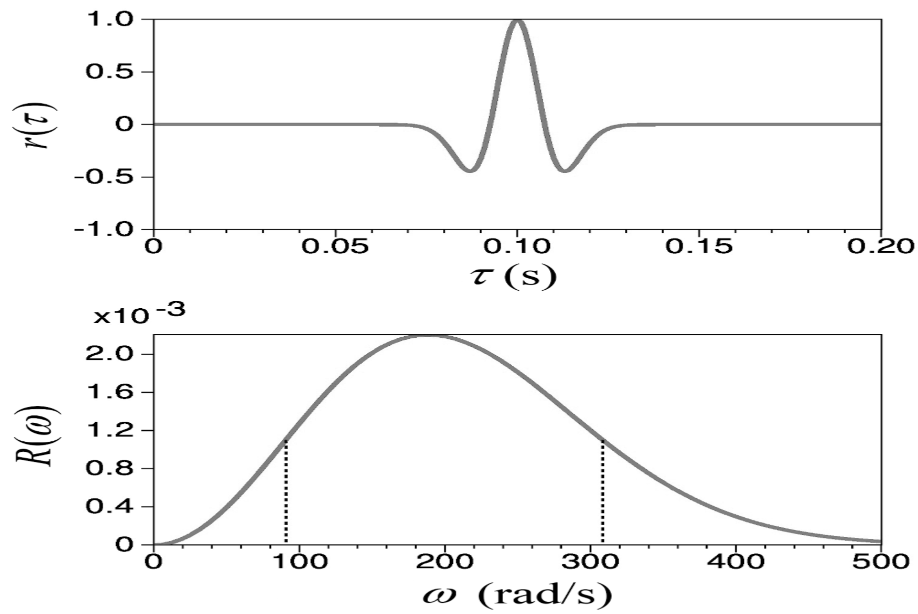

Figure 6.1: Mexican Hat Wavelet and its Fourier Transform

with its Fourier transform $R(\omega)$ where the highest angular frequency is $60\pi rads^{-1}$.The two vertical dashed lines at $\omega = 100$ and $\omega = 300$ denote half of the peak frequency.

6.4 Comparison among the Fourier Transform, Short-time Fourier Transform(STFT) and Wavelet Transform

6.4.1 Forward Transform

Fourier Transform

$$F(\omega) = \int_{-\infty}^{\infty} f(t)exp(-j\omega t)dt \tag{6.9}$$

As already discussed, a signal varying in time is translated to frequency domain with the help of Fourier Transform. However, the disadvantage is that Frequency Resolution is not possible here.

STFT

$$Sf(u, \xi) = \int_{-\infty}^{\infty} f(t)w(t-u)exp(-j\xi t)dt \tag{6.10}$$

The problem with Frequency Resolution in Fourier Transform is addressed here with the use of a small window $w(t-u)$ which only takes a small portion of the original signal and then its Fourier Transform is calculated. The problem that arises here is that the low frequency components are not detected on the spectrum.

Wavelet Transform

$$Wf(s, u) = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})dt \tag{6.11}$$

The previous problem is resolved in the Wavelet Transform. Here, both the Frequency Resolution and Time Resolution is achieved. As the scaling and shifting is done on the mother wavelet, very low frequency components are located at large s while very high frequency components are located at small s.

6.4.2 Inverse Transform

Fourier Transform

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)exp(j\omega t)dt \tag{6.12}$$

STFT

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Sf(u, \xi) w(t-u) exp(j\xi t) d\xi du \tag{6.13}$$

Wavelet Transform

$$f(t) = \frac{1}{C_\psi} \int_{0}^{\infty} \int_{-\infty}^{\infty} Wf(s, u) \frac{1}{\sqrt{s}} \psi(\frac{t-u}{s}) du \frac{ds}{s^2} \tag{6.14}$$

$$C_\psi = \int_{0}^{\infty} \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty \tag{6.15}$$

6.4.3 Time-Frequency Tiling

The Heisenberg Uncertainty Principle in Quantum Physics, states that it is not possible to state both the position and the momentum of a particle at the same time. This is the comparison of the Fourier transform, STFT and the Wavelet Transform:

Fourier Transform

The time resolution is very poor. Frequency resolution is very precise if the signal is integrated over the whole time axis.

STFT

A sliding window is considered on the original signal instead of the original signal. The Frequency resolution depends on the size of the window. The window is uniformly placed, so there is no possibility of zooming in on a particular frequency.

Wavelet Transform

The wavelet transform out of these three techniques strikes a balance between time resolution and frequency resolution and using the parameter s, the higher frequency ranges require higher values of s. Similarly, for lower frequency components, lower value of s is needed. This helps in a better time-frequency analysis.

Chapter 7

Discrete Wavelet Transform

7.1 Why is DWT needed

The discretized continuous wavelet transform obtained by sampling the CWT is not a true discrete transform. The information provided by the wavelet series is very redundant when it comes to reconstructing the signal. The redundancy present takes up a considerable amount of time. While, DWT gives information for analysis and synthesis of the signals with lesser redundancy and lesser computational time. It is easier to implement than CWT.

7.2 Subband Coding

The basic concept is similar to the Continuous Wavelet Transform(CWT). As explained before, the CWT is basically a measure of the correlation/similarity between a wavelet chosen and the original signal with the frequency used as a measure of similarity. The CWT is computed by scaling the window function, shifting the window, multiplying it by the signal and integrating it over the entire time.For discrete CWT, different filters having different cutoff are used to decompose the signal. The signal is then passed through high pass filters and low pass filters to analyze the high frequencies and low frequencies respectively.

The resolution of a signal means the amount of information available in the signal. The resolution can be varied by filters and the scale(frequency) of the signal can be changed by upsampling and downsampling. Upsampling refers to addition of new samples to the signal. Downsampling refers to removal of existing samples from the signal.

The wavelet coefficients in DWT are obtained by sampling the CWT on a dyadic grid, i.e. $s_0 = 2$ and $t_0 = 1$ yielding $s = 2^j$ and $t = k * 2^j$. Let's denote the discrete signal as $x[n]$.

The DWT starts by passing the signal through a halfband low-pass filter, which mathematically means that the convolution of the signal with the impulse of the filter is carried out,

$$x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[n-k] \tag{7.1}$$

Having done this, the frequencies which are more than half the highest frequency in the signal are removed. Now, half of the samples can be removed because the highest frequency is now $p/2$radians, considering it was p radians originally. The signal is now downsampled, so that the signal has half the number of points. Filtering the signal has no effect on the scale. Filtering effects the resolution of the signal. It needs to be mentioned here that the half number of samples are redundant after filtering of the signal. So, half the samples are discarded by downsmapling to remove redundancy.

The mathematical expression of the above process is given by,

$$y[n] = \sum_{k=-\infty}^{\infty} h[k] \cdot x[2n-k] \tag{7.2}$$

The DWT decomposes the signal into various frequency bands with varying resolutions (unlike in STFT) having detail coefficients and approximation coefficients. The highpass filter $g[n]$ and low-pass filter $h[n]$ present achieve the task of decomposing the signal into frequency bands. This signal can now be downsampled by a factor of 2 to eliminate the redundancy. The mathematical equations are as follows:

$$y_{high}[k] = \sum_{n} x[n] \cdot g[2k-n] \tag{7.3}$$

$$y_{low}[k] = \sum_{n} x[n] \cdot h[2k-n] \tag{7.4}$$

Here, $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the highpass and low-pass filters respectively after downsampling.

Having decomposed this signal, a lot of insights need to be explicitly mentioned. The number of samples have been halved, which effectively increases the frequency resolution by a factor of 2. This is called Subband Coding which can be used for successive decomposition. For every level, filtering and down-sampling is done.

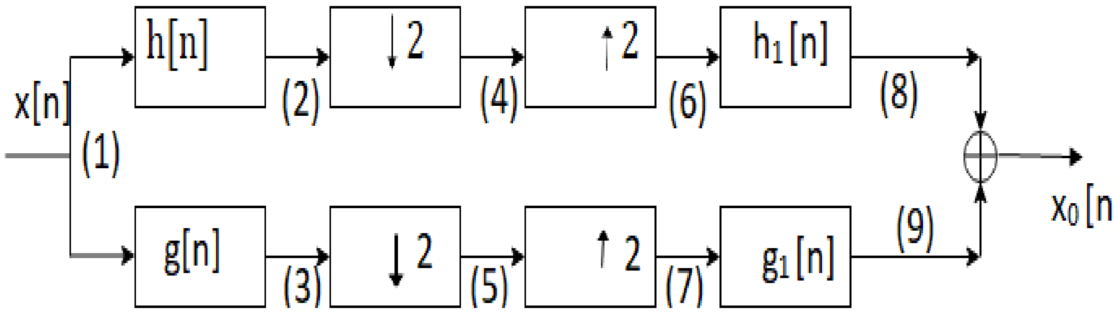

Figure 7.1: The schematic diagram to realize discrete wavelet transform. Here the filter names are changed.

7.3 Example of Subband Coding

Let us take an example to demonstrate Subband Coding. Consider that the original signal $x[n]$ has 512 discrete points to begin with. Let the spectral components range from 0 to p rad/s. On the first decomposition, the output from the highpass filter has p/2 to p rad/s frequencies. These 256 points form the first level of Wavelet Decomposition. The output of the low-pass filter, with the frequencies ranging from is passed through further decomposition. The output from the second high pass filter has frequencies ranging from p/4 to p/2 rad/s which are the second level of DWT coefficients. This window has half the time resolution but twice the frequency resolution than the first level because we have achieved a frequency resolution which is 4 times that of original signal. This is continued until there are two samples left. **In the proposed work, wavelet decomposition done was upto 8 levels. The DWT of the original EEG signal is obtained on combing all the coefficients .** DWT strikes an amazing balance between the time resolution and frequency resolution. As discussed, Time resolution is not possible in Fourier Transform. In DWT, if the information needed lies in the high frequency window,time localization will be precise but if the required information lies at very low frequencies, the time localization is erroneous due to fewer samples. **Most EEG signals are of this type. The epileptic activities are characterized by spike and slow complexes which have high frequency. So, with DWT we get a good time resolution at higher frequencies and good frequency resolution at lower frequencies.** The frequency bands that do not hold much prominence will have very less amplitude. They can be discarded with any major loss of information, allowing data reduction.

7.4 Mathematical Analysis

The high-pass filters and low-pass filters are related by

$$g[L-1-n] = (-1)^n \cdot h[n] \tag{7.5}$$

where g[n] is the high-pass filter and h[n] is the low-pass filter and L is the filter length(number of points). The filtering and down- sampling operations are mathematically given by,

$$y_{high}[k] = \sum_n x[n] \cdot g[-n+2k] \tag{7.6}$$

$$y_{low}[k] = \sum_n x[n] \cdot g[-n+2k] \tag{7.7}$$

$$\tag{7.8}$$

The reconstruction of the signal follows a reverse order. The signals at every level are upsampled by the same factor, followed by synthesis filters (high-pass and low-pass). The reconstruction formula is,

$$x[n] = \sum_{k=-\infty}^{\infty} (y_{high}[k] \cdot g[-n+2k]) + (y_{low}[k] \cdot h[-n+2k]) \tag{7.9}$$

It should be stated that perfect reconstruction is not possible if the filters are not ideal halfband, but under certain conditions some wavelets provide perfect reconstruction. **These are Daubechies' wavelets which have been used in this work.**

7.5 Example of Decomposition by DWT

Consider a signal with 256 samples. Let the sampling frequency be 10 MHz. Let us obtain its DWT coefficients. The highest frequency must be 5 MHz. The first level consists of a low pass filter h[n] and a high pass filter g[n], the outputs of these are then downsampled by 2. The high pass filters contains the frequency range [2.5,5] MHz range. There are 128 such samples. The low pass filter output has frequencies in the range of [0,2.5] MHz are sent for successive DWT decomposition.

At the second level, the high-pass filter has 64 samples and the output of the low-pass filter is passed through a high-pass filter and low-pass filter in the third level.The output of the third high pass filter has 32 samples.

This continues till we are left with 1 DWT coefficient. It should be noted that lesser number of samples are used at lower frequencies, so the time resolution decreases but the frequency resolution increases because the frequency interval decreases. The first few coefficients do not carry much information due to very less time resolution.

7.6 Daubechies Wavelet

The wavelet functions discussed here are based on the lectures by I.Daubechies [10]. Ingrid Daubechies, the brightest name in the world of wavelet research, found something called compactly supported orthonormal wavelets - hence making discrete wavelet analysis practically possible.

Daubechies designed this wavelet for a known vanishing moment with a minimum size discrete filter. The conclusion is that if we want a wavelet function with p vanishing moments, the minimum filter size to be used turns out to be $2p$.

The Daubechies family wavelets are denoted by "db_N ", where N is the order of wavelet, and db denotes the "surname" of the wavelet family. The db1 wavelet, as mentioned earlier, denotes Haar wavelet.

The nine members of the Daubechies Wavlet Family are shown in the following figure.

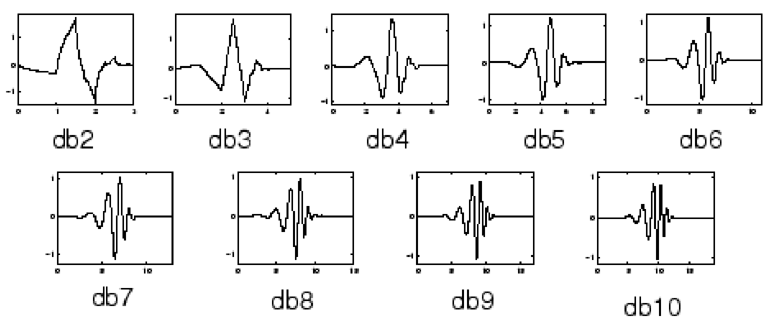

Figure 7.2: Wavelet Functions of the nine members of the Daubechies Wavelet Family

Chapter 8

Proposed Method

The proposed work deals with five experiments, three among which are Multi-class classification problems and the rest of the two are Binary Classifications. The pre-ictal state which the proposed work classifies the EEG signal, is a very useful class for medical purposes. **If the start of the pre-ictal state is predicted correctly, the medication can be provided at the earliest so as to avoid the next seizure attack.** It is important to mention that the proposed work does not propose or recommend any particular method of EEG data acquisition as this is out of the scope of this proposed work. The EEG data in this work is acquired from the University Of Bonn, available in the public domain. The **first stage** of our proposed

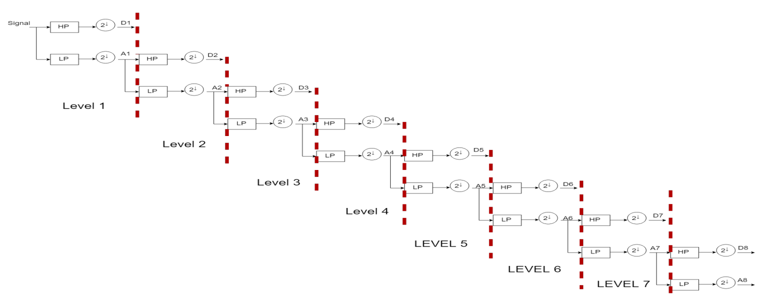

Figure 8.1: Wavelet Decomposition upto 8 levels

method consists of gathering the sampled EEG signals and then perform pre-processing on them.

Having done the pre-processing, the **second stage** is creating 3 matrices corresponding to the three classes used in Multi-Class Classification and the two classes in cases of Binary Classification.

The **third stage** of our proposed method is mainly analysis of the EEG signal of the patient using Discrete Wavelet Transform(DWT). DWT is a convenient technique for this purpose because using DWT, we can decompose the original signal into its frequency components to obtain EEG sub-bands in various ranges of frequency. It is appropriate to mention here that the brain waves existing at different frequency ranges are most active during certain human activities.The main advantage behind using wavelets is that these are localized in both time and frequency whereas on the other hand the Fourier Transform is localized only in frequency. The wavelet transform strikes a balance between the frequency resolution and time resolution of the signal. In the proposed work, Wavelet Decomposition upto the eight level is applied to the dataset considered over which the three-class classification is to be done. The structure of the level 8 wavelet decomposition is shown below. The Proposed Work has been depicted in the form of a flowchart in Fig 8.2. The frequency range for all the coefficients are as follows:

1. D_1: 86.8 - 173.6 Hz
2. D_2: 43.4 - 86.8 Hz
3. D_3: 21.7 - 43.4 Hz
4. D_4: 10.85 - 21.7 Hz
5. D_5: 5.42 - 10.85 Hz
6. D_6: 2.71 - 5.42 Hz
7. D_7: 1.35 - 2.7 Hz
8. D_8: 0.67 - 1.35 Hz
9. A_8: 0 - 0.67 Hz

According to Haddad et al. [19], there is a high correlation between Delta and Gamma sub-bands with Temporal Seizures. Their research reveals that gamma spikes have more probability to originate on the same nodes after delta high voltages have been recorded and there is a high correlation in these events. According to L. Ren et al. [36], it was found out that the Gamma waves often precede the interictal epileptiform spike discharges(IED) in certain brain areas. Their research work shows a strong correlation between the Seizure Onset Zone and gamma-IEDs. **So, the proposed work included the coefficients D_3 and A_8.**

Only few data exist where focal beta activity is manifested as an electroencephalographic seizure pattern. This pattern has only been observed in patients with intractable seizure disorders [7]. The reason for including the coeffcients D_4 and D_5 is because the EEG signals of the healthy patients were recorded with their eyes open and with their eyes closed. It is likely that the alpha waves and the beta waves were the most active in these patients the entire time, considering that alpha waves are most dominant during wakeful alertness, calmness and mental coordination while the beta waves are most active during waking state of consciousness when the patient is paying attention towards cognitive tasks and the outside world. Consid-

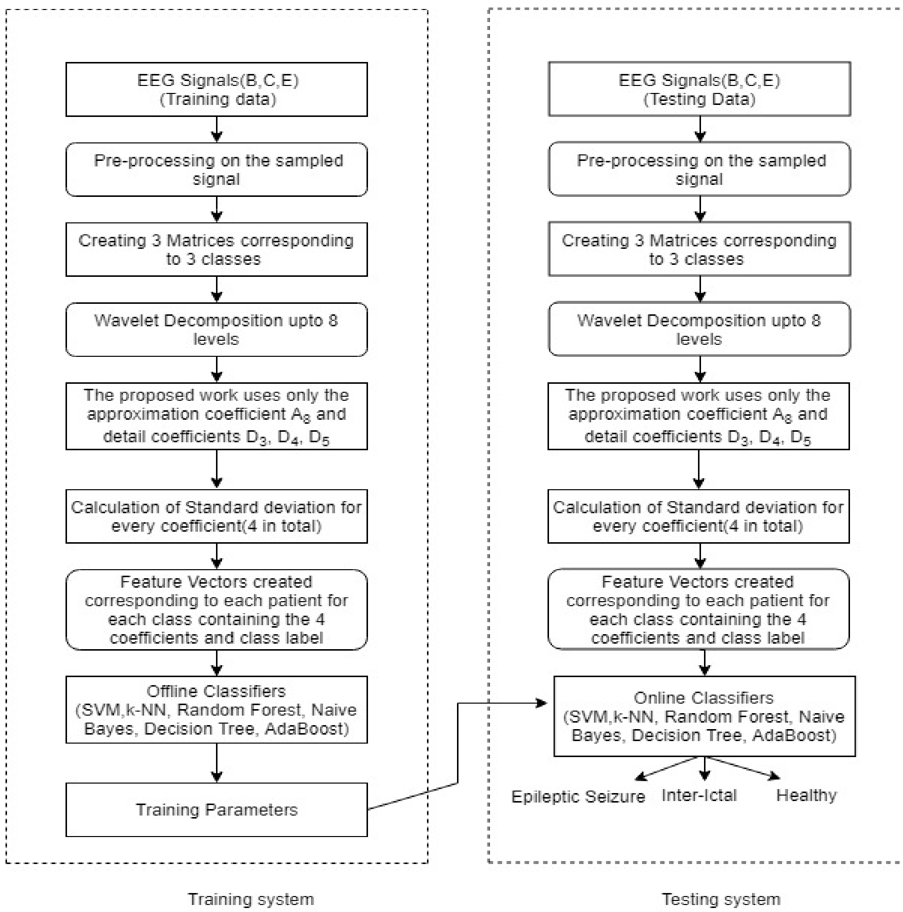

Figure 8.2: A flowchart of the proposed classification framework.

ered coefficients for all healthy, interictal, ictal periods of seizure have been

provided in the following figures 8.3, 8.4 and 8.5.
The coefficients A_8, D_3, D_4, and D_5 have been considered for the proposed method in the **fourth stage**. Having selected the four wavelet coefficients for every feature vector where a feature vector represents a patient of a class, the Standard deviation of the considered coefficients has been computed for all the EEG epochs of a patient in this stage. Thus, we had 4097 discrete EEG readings for every patient to begin with after which Wavelet Decomposition was done upto 8 levels creating 9 coefficients. 4 coefficients where selected out of them whose standard deviation was calculated. The feature matrix consists of five features as the standard deviation of Approximation Coefficient A_8 and Detail Coefficients D_3 , D_4 and D_5 along with the class label for every feature vector.

In the **fifth stage**, a train-test split is made for every class and the training samples are fed to the classifiers. The proposed work uses 60% training data and 40% testing data.
The classifiers are imported and the classification is done in the **sixth stage**. This work consists of five types of experiments which are mainly classification problems, namely:

1. Binary-class classification of healthy patients, with eyes open and patients experiencing seizures

2. Binary-class classification of healthy patients, with eyes closed and patients experiencing seizures

3. Multi-class classification of healthy patients, with eyes open and patients in inter-ictal state(whose EEG readings were recorded from hippocampal formation which was the epileptogenic area) and patients experiencing seizures

4. Multi-class classification of healthy patients, with eyes closed and patients in inter-ictal state(whose EEG readings were recorded from hippocampal formation of an opposite hemisphere of the brain) and patients experiencing seizures

5. Multi-class classification of healthy patients, with eyes closed and patients in inter-ictal state(whose EEG readings were recorded from hippocampal formation in the epileptogenic area of the brain) and patients experiencing seizures

This work compared the result obtained using 8 different Daubechies wavelets namely db1, db2, db3, db4, db5, db6, db7, db8. The results obtained after using each of these wavelets are tabulated in the next chapter and the insights obtained from the results are mentioned thereby.

The features extracted which are essentially the statistical standard deviation of the selected coefficients in the previous stage are fed into the Machine Learning classifiers. The classifiers used in the proposed work are SVM, Decision Tree, Random Forest, Naive Bayes, Nearest Neighbors and AdaBoost. The results obtained are mentioned in the next chapter. The results obtained using each classifier in each of the five experiments are mentioned and tabulated in the next chapter.

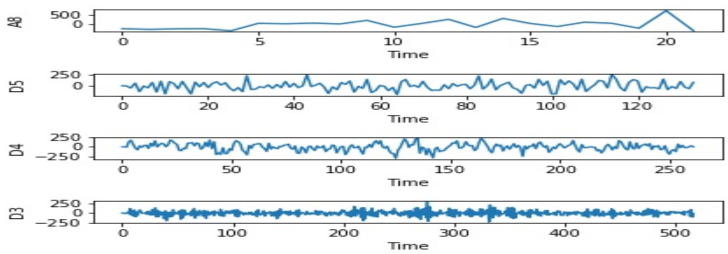

Figure 8.3: Wavelet Decomposition of Healthy EEG signals

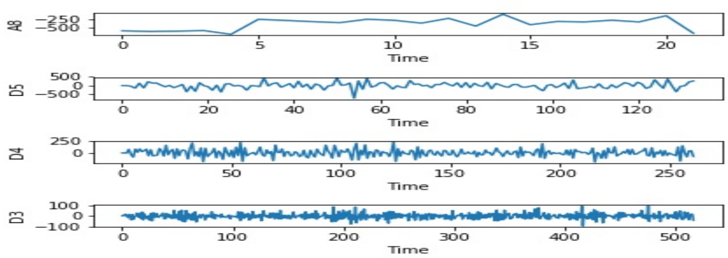

Figure 8.4: Wavelet Decomposition of Inter-ictal EEG signals

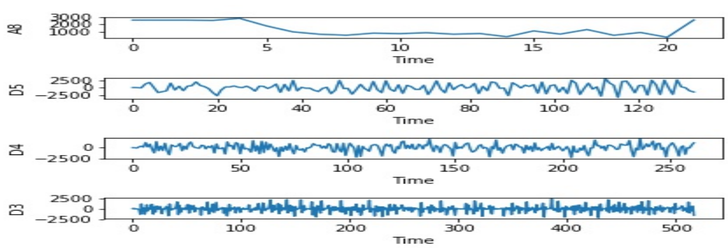

Figure 8.5: Wavelet Decomposition of Ictal EEG signals

8.1 Classifiers Used

The following classifiers were used in this work:

8.1.1 Support Vector Machine

A Support Vector Machine is a powerful classifier which gives the optimal hyperplane. The advantage of using SVM is the optimal hyperplane which maximizes the distance between the nearest datapoints of both the classes. Consider the Figure 8.6 for example. The correct hyper-plane is decided only after maximizing the distances between nearest data point on both the class. The important parameters associated with SVM which have a high impact on model performance are the kernel, gamma, and C. We can choose the kernel to be linear, rbf, poly etc. The gamma parameter is the value or the extent to which the SCM tries to fit the training dataset. The "C" parameter is the penalty parameter of the error term. For higher values of C, the hyperplane has a small margin if it correctly classifies all the training points correctly. Similarly, a larger value of C gives a wide-margin hyperplane even if the hyperplane does a lot of misclassification on the training samples. The Multi-class SVM is implemented using a one-vs-one method.

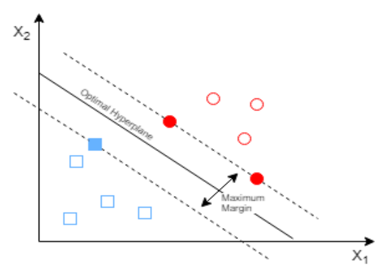

Figure 8.6: Support Vector Machine

The parameter **C** influences the misclassification on training examples. A large value of **C** gives a small-margin hyperplane if that hyperplane looks for correctly classifying all the training points. A small value of **C**, means that we are ready to trade-off classifying every training point correctly in exchange for a broader margin.
C was taken as 0.025 **in this work along with a linear kernel**. **The proposed work consists of 4 features. So, the feature space is 4-dimensional. And thus hyper-planes will be formed to separate two classes in this feature space.**

8.1.2 Decision Tree

Decision Trees are non-parametric classifiers also used for regression. It learns simple decision rules from the variables/features and keeps splitting the feature space. Decision Trees are used for classification and regression. These are tree data structures where each internal node represents a feature and the branch represent a decision. Each attribute is denoted by a node where the most important attribute is the root node. The decision is arrived by starting from the root node and moving downwards until a leaf node is reached. There are mainly two types of Decision Trees:

1. Classification Trees
2. Regression Trees

An example of Decision Tree is given in the figure below. The task is to predict whether a person is fit or unfit and the given information are his age, eating habit and physical activity.

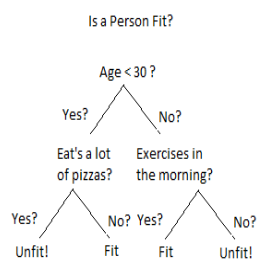

Figure 8.7: Decision Tree

In the proposed work, the Decision Tree is trained using the four attributes which are the standard deviation of the wavelet coefficients D_3, D_4, D_5 and A_8. **The maximum depth of the Decision Tree used in the proposed work was 5.**

8.1.3 Nearest Neighbors

In Nearest Neighbor Classification,classification is achieved with a majority count of the nearest neighbors of each point: the unknown data point is assigned the label which has the most number of data points as the nearest neighbors of the unknown data point.
This can be demonstrated by a simple example. Following is a feature space

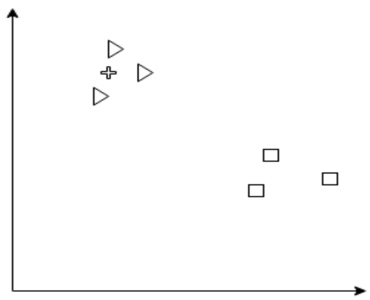

Figure 8.8: Simple Data Spread

containing triangles and squares: The task is to classify the feature vector denoted by "plus". Consider k = 3. Hence, the blue star is now encircled with 3 of its closest neighbors measured by the Euclidean distance. The situation is the following: The three closest points the new feature vectors

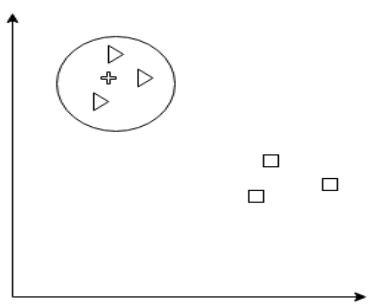

Figure 8.9: Nearest Neighbors in action

are all triangles. Hence, it is classified to be a triangle. The choice of **k** is highly data-dependent: it has been generally observed that a high value of **k** diminishes the noise, but the boundaries formed are less distinct.
k was taken as 3 in the proposed work. **The proposed work consists of 4 features. So, the feature space is 4-dimensional. And hyperspheres will thus be formed while selecting the 3-nearest neighbors.**

8.1.4 Random Forest

The Random Forest Classifier creates a number of decision trees in the training phase. The decision taken by the majority of the trees is considered to be the final classification result. The advantage in using the Random Forest

Classifier in this work is because it avoids overfitting. Random Forest handle the missing values. Also, Random Forests can be used for categorical data too. The Random Forest Pseudocode is as follows:

1. "k" features selected randomly from "m" features
2. Among the "k" features, calculate the node "d"
3. This node is divided into child nodes using the best split approach
4. Steps 1 to 3 are repeated until "l" number of nodes are obtained
5. Build Random Forest by repeating steps 1 to 4 for "n" times to create "n" number of trees

Random Forest pseudocode

1. The testing data is taken and using the rules of each randomly created Decision Trees, every outcome from each tree is stored
2. Each predicted target are taken into votre
3. The highest voted prediction is considered to be the final decision

The maximum depth of the decision trees is taken to be 5. The number of trees are taken as 10. Only one feature was considered for the splitting which was the value obtained after quantization of the sampled signal.

8.1.5 Naive Bayes Classifier

These methods are a set of supervised learning algorithms which are based on Bayes' Theorem.
According to Bayes' Theorem, the *Aposteriori Probability* probability is obtained from *Apriori Probability* and likelihood by the following equation:

$$P(y|x_1,.....,x_n) = \frac{P(y)P(x_1,...,x_n|y)}{P(x_1,...,x_n)} \tag{8.1}$$

The conditional probability is given by,

$$P(x_i|y,x_1,x_2,...,x_{i-1},x_{i+1},...x_n) = P(x_i|y) \tag{8.2}$$

Thus the equation 8.1 can be further simplified:

$$P(y|x_1,.....,x_n) = \frac{P(y)\prod_{i=1}^{n} P(x_i|y)}{P(x_1,....,x_n)} \tag{8.3}$$

Since, the denominator is a constant, the following classification rule can be applied

$$P(y|x_1, ..., x_n) \propto P(y) \textstyle\prod_{i=1}^{n} P(x_i|y)$$

which gives

$$\hat{y} = \underset{y}{\operatorname{argmax}}\, P(y) \prod_{i=1}^{n} P(x_i|y) \tag{8.4}$$

$P(y)$ and $P(x_i|y)$ can be estimated by the Maximum A posteriori(MAP)rule. There are many advantages in using these classifiers. These are faster compared to others and help avoid problems related to the curse of dimensionality.
The disadvantage in this method is its poor estimation.

8.1.6 AdaBoost Classifier

AdaBoost classifier is another ensemble classifier like the Random Forest Classifier. The working of this classifier is characterized by using a classifier on the dataset initially[12]. Then, multiple copies of the classifier are selected at every iteration with various weights assigned to the training instances where a higher weight is assigned to the misclassified item so that it is taken as a training instance in the next iteration of the next classifier.
The most accurate classifiers are assigned more weights, so that they have more impact on the final classification decision.

Chapter 9

Results And Comparison

The proposed work deals with the three-class classification problems where the classes are consisted of Healthy patients(with eyes open and eyes closed), Patients in an inter-ictal stage (whose EEG readings have been recorded from the opposite hemisphere of the brain and epileptogenic zone of the brain) and patients experiencing epileptic activity.
The following work was carried out on a Intel(R) Core(TM)i5-8250U CPU@1.80 GHz with a 64-bit OS, x64-based processor. The language used was Python3 in Anaconda Environment. The packages used were the following:

1. NumPy (ver1.12.1)
2. Pandas (ver0.20.1)
3. Scikit-Learn (ver0.18.1)
4. PyWavelets (ver0.5.2)
5. Matplotlib (ver2.0.2)
6. Time
7. os

9.1 Multi-Class Classification

The following table shows the list of all the classifiers used in the work where the highest classification accuracy achieved was 98.45% (by Support Vector Machine). The accuracy is obtained while using the classifier for Multi-Class Classification on datasets *BC-E*.

Classifier	Accuracy
Support Vector Machine	98.45%
Nearest Neighbors	95.34%
Random Forest	95.34%
Naive Bayes	94.57%
Decision Tree	93.79%
AdaBoost	74.41%

Table 9.1: Comparison between accuracy achieved by different classifiers in Multi-Class Classification of BC-E

Classifier	Accuracy
Support Vector Machine	95.34%
Nearest Neighbors	83%
Random Forest	96.12%
Naive Bayes	86.82%
Decision Tree	95.34%
AdaBoost	94.57%

Table 9.2: Comparison between accuracy achieved by different classifiers in Multi-Class Classification of AD-E

Classifier	Accuracy
Support Vector Machine	96.89%
Nearest Neighbors	94.57%
Random Forest	93.79%
Naive Bayes	92.24%
Decision Tree	92.24%
AdaBoost	76.74%

Table 9.3: Comparison between accuracy achieved by different classifiers in Multi-Class Classification of BD-E

9.2 Two-Class Classification

The following table shows the list of all the classifiers used in the work where the highest classification accuracy achieved was 100% (by Support Vector Machine). The accuracy is obtained while using the classifier for Two-Class Classification on datasets *A-E* and *B-E*.

Cases	Classifier Used	Accuracy
A-E	Support Vector Machine	96.67%
	Nearest Neighbors	100%
	Random Forest	100%
	Naive Bayes	100%
	Decision Tree	96.67%
	AdaBoost	96.67%
B-E	Support Vector Machine	100%
	Nearest Neighbors	96.67%
	Random Forest	96.67%
	Naive Bayes	96.67%
	Decision Tree	100%
	AdaBoost	100%

Table 9.4: Comparison between accuracy achieved by different classifiers in Two-Class Classification

9.3 Comparison of Accuracy with different Daubechies Wavelets

The proposed work deals with the three-class classification problem where EEG epochs are classified into healthy patients with eyes closed, seizure-free intervals and seizure (ictal periods) epochs along with two two-class classification problems where one problem was classification of EEG epochs into healthy patients(with eyes closed) and patients experiencing seizure. The other problem being classification of EEG epochs into healthy patients(with eyes open) and patients experiencing seizure. The following tables show comparisons of accuracy obtained using different wavelets:

Wavelet Used	Accuracy for BC-E	Classifier Used
db1	97.67%	SVM
db2	97.67%	SVM
db3	97.67%	SVM
db4	**98.45%**	**SVM**
db5	96.96%	SVM
db6	96.96%	SVM
db7	96.96%	SVM
db8	96.96%	SVM

Table 9.5: Comparison between accuracy achieved by using different Daubechies wavelets on the dataset BC-E

Wavelet Used	Accuracy for B-E	Classifier Used
db1	100%	SVM
db2	100%	Decision Tree
db3	100%	SVM
db4	**100%**	**SVM**
db5	100%	SVM
db6	100%	Decision Tree
db7	100%	Nearest Neighbors
db8	100%	SVM

Table 9.6: Comparison between highest accuracy achieved by using different Daubachies wavelets in classifiers on set B-E

The description of the datasets *A, B, C, D and E* are summarized in Chapter 4. It is observed that the highest accuracy is obtained using the *db*4 wavelet

Wavelet Used	Accuracy for A-E	Classifier Used
db1	100%	SVM
db2	100%	SVM
db3	100%	SVM
db4	**100%**	**SVM**
db5	100%	SVM
db6	100%	SVM
db7	100%	SVM
db8	100%	Nearest Neighbors

Table 9.7: Comparison between highest accuracy achieved by using different Daubachies wavelets in classifiers on set A-E

Wavelet Used	Accuracy for AD-E	Classifier Used
db1	94.57%	SVM
db2	94.57%	AdaBoost
db3	95.34%	SVM
db4	**96.12%**	**Random Forest**
db5	96.12%	Random Forest
db6	95.34%	Random Forest
db7	94.57%	SVM
db8	94.57%	SVM

Table 9.8: Comparison between highest accuracy achieved by using different Daubachies wavelets in classifiers on set AD-E

Wavelet Used	Accuracy for BD-E	Classifier Used
db1	95.34%	SVM
db2	95.34%	SVM
db3	96.89%	SVM
db4	**96.89%**	**SVM**
db5	94.57%	SVM
db6	95.34%	SVM
db7	96.89%	SVM
db8	96.89%	SVM

Table 9.9: Comparison between highest accuracy achieved by using different Daubachies wavelets in classifiers on set BD-E

which gives 98.45% in the three-class classification problem and 100% in both the binary classification problem. This might be due to the smoothing factor of *db*4 making it suitable to detect changes in the EEG Epoch.

9.4 Conclusions from Results

The results obtained from the various experiments conducted provide many insights to look into. Working with an automated system which accomplishes the task of Multi-class classification instead of a system which does Two-class classification is a better choice because in the former, the inter-ictal activity is detected separately. This can help physicians to administer the vaccines and provide immediate medication to these patients in inter-ictal state before they experience the next epileptic activity. This is not possible to achieve in the Binary-Classification scenario.

The results obtained in Table 9.1 and Table 9.2 show that a higher accuracy is achieved while performing classification using the sets *BC-E*. This gives meaningful insights that set *A* has healthy patients whose EEG were recorded with eyes open and set *B* has healthy patients whose EEG were recorded with eyes closed. The blinking of the eyes from patients in set *A* contribute to physiological artifacts which are not present if set *B* is used. And, thus an accuracy of 98.45% is achieved in the Multi-class classification using set *BC-E* while the highest accuracy achieved using set *AD-E* is 95%.

The results obtained from Table 9.4 shows that both the sets *A-E* and *B-E* are linearly separable considering we achieved 100% accuracy using SVM in the case of *B-E*. SVM achieves 96.67% accuracy in set *A-E* which may be due to outliers. However on using k-NN classifier where k= 3, we achieve 100% accuracy in set *A-E*. This explains that the outliers are very few in population thereby leading the k-NN classifier to the right classification everytime the testing point is assigned the label of the majority count.

It is observed from the tables that the highest accuracy is achieved using the fourth-order Daubechies(db4) wavelet. This can be attributed to their better localization performance in both time domain and frequency domain compared to other Daubechies wavelet. Furthermore, for Signal Processing in EEG, db4 wavelets are better suited due to their smoothing factor and shape[13].

Chapter 10

Conclusion and Future Work

It is a very time consuming and computationally expensive task to detect epileptic seizure activities from lengthy EEG signals. This process is usually done by trained professional. In this work an automated epileptic seizure detection system has been developed based on DWT followed by classification of the EEG epochs into one of the three classes namely healthy, interictal and ictal by an SVM classifier. The EEG recordings obtained from healthy patients with eyes closed and patients having seizures have been classified with an accuracy of 100%. The EEG recordings obtained from healthy patients with eyes open and patients having seizures have been classified with an accuracy of 100%. The EEG recording obtained from the three classes during healthy, interictal and ictal dataset have been classified with an accuracy as high as 98.45 %. It would be beneficial if the proposed method is used in real time for three-class classification. It is expected that with this system, subtle information which are usually hidden in the EEG data will be revealed to the clinicians with the help of which better decision can be taken. The proposed model must be validated using some other database than the one used here. Additionally, if the proposed model is to be used in real-time epilepsy seizure detection the proposed model needs to be tested with long duration EEG records with more patients.

As a part of future work, we plan to deploy the method to test and predict seizure activity in epileptic patients in a clinical environment and apply deep learning methods.

Bibliography

[1] U Rajendra Acharya, Filippo Molinari, S Vinitha Sree, Subhagata Chattopadhyay, Kwan-Hoong Ng, and Jasjit S Suri. Automated diagnosis of epileptic eeg using entropies. *Biomedical Signal Processing and Control*, 7(4):401–408, 2012.

[2] U. Rajendra Acharya, S. Vinitha Sree, G. Swapna, Roshan Joy Martis, and Jasjit S. Suri. Automated eeg analysis of epilepsy: A review. *Know.-Based Syst.*, 45:147–165, June 2013.

[3] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E. Elger. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E*, 64:061907, Nov 2001.

[4] V. Bajaj and R. B. Pachori. Classification of seizure and nonseizure eeg signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine*, 16(6):1135–1142, Nov 2012.

[5] Abhijit Bhattacharyya, Ram Bilas Pachori, Abhay Upadhyay, and U. Rajendra Acharya. Tunable-q wavelet transform based multiscale entropy measure for automated classification of epileptic eeg signals. *Applied Sciences*, 7(4), 2017.

[6] Niels Birbaumer. Clinical-psychological treatment of epileptic seizures: a controlled study. *Perspectives and promises in clinical psychology*, 1991.

[7] Leo H Bonati, Yvonne Naegelin, Heinz-Gregor Wieser, Peter Fuhr, and Stephan Ruegg. Beta activity in status epilepticus. *Epilepsia*, 47(1):207–210, 2006.

[8] Weiting Chen, Jun Zhuang, Wangxin Yu, and Zhizhong Wang. Measuring complexity using fuzzyen, apen, and sampen. *Medical Engineering & Physics*, 31(1):61–68, 2009.

[9] Anindya Bijoy Das, Mohammed Imamul Hassan Bhuiyan, and S. M. Shafiul Alam. Classification of eeg signals using normal inverse gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection. *Signal, Image and Video Processing*, 10(2):259–266, Feb 2016.

[10] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.

[11] Paul Fergus, A Hussain, David Hignett, Dhiya Al-Jumeily, Khaled Abdel-Aziz, and Hani Hamdan. A machine learning system for automated whole-brain seizure detection. *Applied Computing and Informatics*, 12(1):70–89, 2016.

[12] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

[13] Dragoljub Gajic, Zeljko Djurovic, Jovan Gligorijevic, Stefano Di Gennaro, and Ivana Savic-Gajic. Detection of epileptiform activity in eeg signals based on time-frequency and non-linear analysis. *Frontiers in computational neuroscience*, 9:38, 2015.

[14] Tapan Gandhi, Bijay Ketan Panigrahi, and Sneh Anand. A comparative study of wavelet families for eeg signal classification. *Neurocomputing*, 74(17):3051 – 3057, 2011.

[15] J_ Gotman, JR Ives, and P Gloor. Frequency content of eeg and emg at seizure onset: possibility of removal of emg artefact by digital filtering. *Electroencephalography and clinical neurophysiology*, 52(6):626–639, 1981.

[16] Jean Gotman. Automatic recognition of epileptic seizures in the eeg. *Electroencephalography and clinical Neurophysiology*, 54(5):530–540, 1982.

[17] Ling Guo, Daniel Rivero, Julián Dorado, Juan R. Rabuñal, and Alejandro Pazos. Automatic epileptic seizure detection in eegs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods*, 191(1):101 – 109, 2010.

[18] Ling Guo, Daniel Rivero, and Alejandro Pazos. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *Journal of neuroscience methods*, 193(1):156–163, 2010.

[19] Tahar Haddad, Naim Ben-Hamida, Larbi Talbi, Ahmed Lakhssassi, and Sadok Aouini. Temporal epilepsy seizures monitoring and prediction using cross-correlation and chaos theory. *Healthcare technology letters*, 1(1):45–50, 2014.

[20] Eric R Kandel, James H Schwartz, Thomas M Jessell, Department of Biochemistry, Molecular Biophysics Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

[21] Yılmaz Kaya, Murat Uyar, Ramazan Tekin, and Selçuk Yıldırım. 1d-local binary pattern based feature extraction for classification of epileptic eeg signals. *Applied Mathematics and Computation*, 243:209–219, 2014.

[22] Yusuf Uzzaman Khan, Nidal Rafiuddin, and Omar Farooq. Automated seizure detection in scalp eeg using multiple wavelet scales. In *2012 IEEE international conference on signal processing, computing and control*, pages 1–5. IEEE, 2012.

[23] Yatindra Kumar, ML Dewal, and RS Anand. Epileptic seizure detection using dwt based fuzzy approximate entropy and support vector machine. *Neurocomputing*, 133:271–279, 2014.

[24] Ralph Meier, Heike Dittrich, Andreas Schulze-Bonhage, and Ad Aertsen. Detecting epileptic seizures in long-term human eeg: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *Journal of clinical neurophysiology*, 25(3):119–131, 2008.

[25] Mohamed EL Menshawy, Abdelghani Benharref, and Mohamed Serhani. An automatic mobile-health based approach for eeg epileptic seizures detection. *Expert Systems with Applications*, 42(20):7157–7174, 2015.

[26] Renee S Monderer, Daniel M Harrison, and Sheryl R Haut. Neurofeedback and epilepsy. *Epilepsy & Behavior*, 3(3):214–218, 2002.

[27] Saadat Nasehi and Hossein Pourghassem. Patient-specific epileptic seizure onset detection algorithm based on spectral features and ipsonn classifier. In *2013 International Conference on Communication Systems and Network Technologies*, pages 186–190. IEEE, 2013.

[28] Nicoletta Nicolaou and Julius Georgiou. Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Systems with Applications*, 39(1):202–209, 2012.

[29] Hasan Ocak. Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, 36(2):2027–2036, 2009.

[30] Umut Orhan, Mahmut Hekim, and Mahmut Ozer. Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, 38(10):13475 – 13481, 2011.

[31] Kunjan Patel, Chern-Pin Chua, Stephen Fau, and Chris J Bleakley. Low power real-time seizure detection for ambulatory eeg. In *2009 3rd International Conference on Pervasive Computing Technologies for Healthcare*, pages 1–7. IEEE, 2009.

[32] M. Peker, B. Sen, and D. Delen. A novel method for automated diagnosis of epilepsy using complex-valued classifiers. *IEEE Journal of Biomedical and Health Informatics*, 20(1):108–118, Jan 2016.

[33] Hao Qu. Self-adapting algorithms for seizure detection during eeg monitoring. 1997.

[34] Hao Qu and Jean Gotman. Improvement in seizure detection performance by automatic adaptation to the eeg of each patient. *Electroencephalography and clinical Neurophysiology*, 86(2):79–87, 1993.

[35] Hao Qu and Jean Gotman. A seizure warning system for long-term epilepsy monitoring. *Neurology*, 45(12):2250–2254, 1995.

[36] Liankun Ren, Michal T Kucewicz, Jan Cimbalnik, Joseph Y Matsumoto, Benjamin H Brinkmann, Wei Hu, W Richard Marsh, Fredric B Meyer, S Matthew Stead, and Gregory A Worrell. Gamma oscillations precede interictal epileptiform spikes in the seizure onset zone. *Neurology*, 84(6):602–608, 2015.

[37] ME Saab and Jean Gotman. A system to detect the onset of epileptic seizures in scalp eeg. *Clinical Neurophysiology*, 116(2):427–442, 2005.

[38] Osman Salem, Amal Naseem, and Ahmed Mehaoua. Epileptic seizure detection from eeg signal using discrete wavelet transform and ant colony classifier. In *2014 IEEE International Conference on Communications (ICC)*, pages 3529–3534. IEEE, 2014.

[39] K. Samiee, P. Kovács, and M. Gabbouj. Epileptic seizure classification of eeg time-series using rational discrete short-time fourier transform. *IEEE Transactions on Biomedical Engineering*, 62(2):541–552, Feb 2015.

[40] Manish Sharma, Ram Bilas Pachori, and U. Rajendra Acharya. A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension. *Pattern Recognition Letters*, 94:172 – 179, 2017.

[41] A. Sharmila and P. Geethanjali. Dwt based detection of epileptic seizure from eeg signals using naive bayes and k-nn classifiers. *IEEE Access*, 4:7716–7727, 2016.

[42] Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009.

[43] Vairavan Srinivasan, Chikkannan Eswaran, Sriraam, and N. Artificial neural network based epileptic detection using time-domain and frequency-domain features. *Journal of Medical Systems*, 29(6):647–660, 2005.

[44] Abdulhamit Subasi, Jasmin Kevric, and M. Abdullah Canbaz. Epileptic seizure detection using hybrid machine learning methods. *Neural Computing and Applications*, 31(1):317–325, Jan 2019.

[45] Piyush Swami, Tapan K. Gandhi, Bijaya K. Panigrahi, Manjari Tripathi, and Sneh Anand. A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Systems with Applications*, 56:116 – 130, 2016.

[46] A Temko, E Thomas, W Marnane, G Lightbody, and G Boylan. Eeg-based neonatal seizure detection with support vector machines. *Clinical Neurophysiology*, 122(3):464–473, 2011.

[47] A. K. Tiwari, R. B. Pachori, V. Kanhangad, and B. K. Panigrahi. Automated diagnosis of epilepsy using key-point-based local binary pattern of eeg signals. *IEEE Journal of Biomedical and Health Informatics*, 21(4):888–896, July 2017.

[48] Alexandros T Tzallas, Markos G Tsipouras, and Dimitrios I Fotiadis. Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Computational Intelligence and Neuroscience*, 2007, 2007.

[49] Ihsan Ullah, Muhammad Hussain, Hatim Aboalsamh, et al. An automated system for epilepsy detection using eeg brain signals based on deep learning approach. *Expert Systems with Applications*, 107:61–71, 2018.

[50] SM Wani, S Sabut, and SL Nalbalwar. Detection of epileptic seizure using wavelet transform and neural network classifier. In *Computing, Communication and Signal Processing*, pages 739–747. Springer, 2019.

[51] Scott B Wilson, Mark L Scheuer, Ronald G Emerson, and Andrew J Gabor. Seizure detection: evaluation of the reveal algorithm. *Clinical neurophysiology*, 115(10):2280–2291, 2004.

[52] Qi Yuan, Weidong Zhou, Shufang Li, and Dongmei Cai. Epileptic eeg classification based on extreme learning machine and nonlinear features. *Epilepsy research*, 96(1-2):29–38, 2011.

[53] Qi Yuan, Weidong Zhou, Yinxia Liu, and Jiwen Wang. Epileptic seizure detection with linear and nonlinear features. *Epilepsy & Behavior*, 24(4):415–421, 2012.