Analysis of Epileptic Seizure Using Wavelet Transform

by

Muhammad Dinie Akmal Bin Mustaffa

15965

Dissertation submitted in partial fulfilment of

The requirements for the

Bachelor of Engineering (Hon)

(Electrical & Electronics)

JAN 2016

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

32610 Tronoh

Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Analysis of Epileptic Seizure Using Wavelet Transform

by

Muhammad Dinie Akmal Bin Mustaffa

15965

A project dissertation submitted to the

Electrical and Electronics Engineering Programme

Universiti Teknologi PETRONAS

In partial fulfillment to the requirement for the

BACHERLOR OF ENGINEERING (Hons)

(ELECTRICAL & ELECTRONICS)

Approved by

____

(DR. NORASHIKIN BT. YAHYA)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

Jan 2016

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons

____

(MUHAMMAD DINIE AKMAL BIN MUSTAFFFA)

ABSTRACT

In this work, wavelet transform (WT) is used to analyze epileptic seizure in recorded EEG signals. Wavelets allow non-stationary EEG signals to be decomposed into elementary forms at different positions and scales. The extracted features from the WT decomposition is then expressed in terms wavelet based and classical based features to be further analyzed. In general, the coefficients of a 1-D wavelet decomposition comprises of approximate and detail coefficient, arranged in a single row. The number of wavelet coefficients depends on the decomposition level with more coefficients at high decomposition level. The features generated from wavelet transform is tested in terms of discriminatory information and the highly informative features will be identified. To select the best features, Fisher Discriminant Ratio (FDR) is implemented and classification error was calculated using Support Vector Machine (SVM). When FDR is applied, amongst all the 23 channels, certain channels will be dominant over the other channels in terms of value and these channels are then be chosen for the reduced feature analysis. Comparisons are made between full feature (23 channels) and reduced feature analysis (8 channels) of wavelet and classical based features and also between the two based features as a whole. Results generated show that features of wavelet based exhibits a lower classification error overall with Mean Absolute Deviation (MAD) generating the lowest error which is as low as 0.0391. This means that it has an accuracy of around 96%. Classical based a higher error overall which makes wavelet based the better and ideal features to be extracted and analyzed. Between reduced and full feature, reduced feature have a lower classification error and full feature.

ACKNOWLEDGMENT

First and foremost, I would like to express my greatest gratitude to the Almighty for granting me a sound mind which great ideas and good judgment comes from and also a healthy body, which allows me to perform my duties as a student to its fullest without much difficulty. With his blessing, I was able to complete the last 4 months of my final year and also my FYP 2 course.

Although my major is in power, I am able to expand my scope into other areas for my FYP 2 thus, gain a little more experience. This is all possible thanks to my supervisor, Dr. Norashikin Yahya for providing a title with a signal and communication based subject. In addition to my major knowledge obtained in lectures, I also acquire some understanding in communication and signals which could be useful in the future. Without her never ending support, the project would have been hard to complete.

Not forgetting, this would not have transpired without the encouragement and assistance from my family members. Their support, regardless in terms of emotional or material are highly appreciated and what keeps me going until the end.

Last but not least, I would like to express a million thanks to all my friends which I have spent my years with in UTP. Together, we had many sweet and bitter experiences and these kind of experiences will make us a more mature. Without their support and guidance, I would have trouble surviving in UTP.

TABLE OF CONTENTS

LIST OF ALL FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Background of Study

Epileptic seizure is a condition where a person will experience a short episode of signs or symptoms as a result of an unusual excessive or synchronous neuronal activity in the brain. The disease related to the brain which causes the epileptic seizure is called epilepsy [1, 2]. Epilepsy is a quite common condition which affects around 0.5% to 1% of the world population [3]. There are variety of symptoms that represents an epileptic seizure such as uncontrolled jerking movement (tonic-clonic seizure) to as sophisticated state as a brief loss of consciousness (absence seizure). In a more detailed statistics, the possibilities of having a seizure when a person is still in their youth is little to none, however 5% to 10% people who are in their 80s will have at least one case of epileptic seizure along their lifetime [4] and chances for the second seizure to occur is between 40% to 50% [5].

Most public are confused between epilepsy and seizure and thought the two are the same. Epilepsy occurs when the brain is not sending normal patterns of electrical signal to the body. The neurons which contains a variety of information may be fired at 500 times a second which is much faster than a normal condition. This will result in an uncontrollable movement of the body as an outcome of the vast information in the neurons which cannot be process by the brain. A person who is experiencing a seizure may not have epilepsy. Epilepsy is a disease which are connected to the brain while a seizure is a state where the person cannot control their body movement which is more related to the motor sensory. A seizure may also appear if a person experience a certain trauma which they have not overcome. Undergoing a seizure once does not mean that a person has an epilepsy but multiple seizures requires a medical checkup to be safe.

Seizures have 2 main categories which are focal and generalized seizure. Focal seizure are separated into simple and complex seizure. Simple seizure happens when a person is still awake but they will undergo an abnormal sensation or have an unusual feelings. They may also hallucinate and hear, smell or feel things that are not real. For complex seizure, it can occur when they are conscious or unconscious which is much more dangerous. Repetitive movements and strange behaviors can be observed such as cutting things with a knife repeatedly. Generalized seizure occurs when both sides of the brain undergo an abnormal neuronal activity. This can lead to many type of minor seizures for instance stiffening of muscles (tonic seizure), repeated jerking movement (clonic seizure), upper body jerking and twitching (myoclonic seizure), normal muscle tone loss (atonic seizure) and a mixture of all the mention symptoms (tonic-clonic seizure).

1.2. Problem Statement

In our present time, more and more people have been diagnosed with diseases related to the brain. As we know, our brain is the most important part of our whole body system. No matter how trivial the illness is, if it is brain related, something must be done right way. Diseases such as seizure is no exception. Analysis of EEG signals play an important role in detection of epileptic seizure from EEG recordings. A good analysis algorithm will improve detection accuracy of epileptic seizure algorithms.

1.3. Objective

The primary goal of this study is:

I. To investigate the discriminatory information and class separability of features extracted from EEG signals using wavelet transform

1.4. Scope of Study

1.4.1. Epileptic Seizure

Epilepsy is one of the most widespread brain disease which has affected many people at random. It is a disorder which causes the human body to have a tendency of repeating epileptic seizure. Rather than a single disease, it can be said that it is the source for other diseases. This condition happened because of random and abundant electrical activities occurred all at once in the brain. As a result, our brain could not interpret which activity to be done thus, the activities are all done at the same time making our body moved in an uncontrollable manner or making a person unconscious [6]. Researches have done a lot of study on epileptic seizure and it can be caused by various factors but for some people the cause cannot be identified. Due to its random nature, people are usually oblivious to this disorder. This thinking need to be changed as it can lead to a serious physical injury. As stated in the background study, seizures are separated into 2 types which are focal and generalized seizures. In this project, no types of seizures are being specified. All types will be detected and analyzed.

1.4.2. Electroencephalogram (EEG)

EEG is one the most vital device for the diagnosis and analysis of epilepsy or other diseases that are related to the brain. The function of it is to record any electrical activity produced by the movement of neurons in the brain along the scalp in other words, it records the cortical electrical activity [7]. The readings were taken from the electrodes which are placed at key points on a person's head at the scalp. Usually, the readings were very small in values around microvolts (μV). EEG waves can be divided into several different frequencies with each representing certain activities of the human brain. The frequencies are as follows:

Table 1.1 Brain Waves

Frequency Bands	Frequency Range (Hz)	Description
Delta (δ)	0-4	Deep sleep condition
Theta (θ)	4-8	Inspiration or meditation
Alpha (α)	8-12	Relaxed consciousness
Beta (β)	13-30	Active brain activity
Gamma (γ)	> 30	Brain illness

Besides frequency, there are other variables which can be used to classify EEG activities such as the voltage level which represents the amplitude, morphology which is the shape of the waveform, synchrony which refers to the random emergence of rhythmic frequency and morphology and lastly periodicity which refers to random distribution pattern.

Not much was known about its other applications however, it has been commonly used in the medical field to treat patients with brain diseases [8]. Doctors and specialists in hospitals need a good detection and analyzing tools for this disease as by far, it is the only tool which was proven effective against brain related illness.

1.4.3. Proposed Techniques

From all these years, there are many techniques which have been proposed to detect and analyze EEG signal. Some of them are Wavelet Transform (WT) which decomposes the signal into smaller wavelet coefficients, Short-Time Fourier Transform (STFT) which describes the frequency and spectral content of a signal, Gabor Transform (GT) which is a part of Fourier Transform (FT) but uses a sliding window called the Gaussian Window, pattern recognition approach which detects certain patterns of the EEG signal to detect the seizure and many more which are not mentioned. Various methods that have been proposed will be further elaborated in the literature review section.

CHAPTER 2

LITERATURE REVIEW

The world we live in now is evolving everyday as we speak. The number of achievements and discoveries in science and engineering are limitless. Up to this day, even though the technologies are already at its peak, research and experiments are still being conducted for the future benefit of human kind. This is all possible due to the past knowledge and research of our ancestors. Without them, the pathway to new technologies being developed nowadays would have probably never existed.

Methods to analyze EEG signals have also been researched, discovered and improved by many experts such as spike averaging, linear and non-linear correlation, non-linear dynamic methods, wavelet transformations, and Fourier spectral analysis. These methods have been developed in order to ease the work of medical doctors in detecting and examining abnormalities or diseases which are related to the brain. Each of these methods have their own advantages and disadvantages in analyzing EEG signals, some are accurate but slow and others are fast but inaccurate. In this paper, the method that will be focused on is WT.

With all that being said, since the beginning, there are countless methods that have been proposed over time to analyze epileptic seizure until now. One of those would be a method proposed by Quiroga et. al [9] which uses a technique that is similar to Fourier Transform (FT) which is known as the Gabor Transform (GT). The only difference between these two techniques is that the original signal is applied with a sliding window of 1.25s called the Gaussian Window. FT is calculated for the Gaussian windowed signal to obtain the frequency representation. Then, the bandrelative intensity ratio (RIR) for each frequency band that were defined was plotted for the signal. Characterization of the analyzed signal was acquired by converting the frequency representation to time representation. The mean of RIR is then calculated for the pre-ictal and ictal phase where both will be compared to the lower intensity areas (plateaus) observed in the seizure state. The results could be count as successful because a significant reduction in delta band activity were spotted in 70% of the seizure which alpha and delta band were emphasized. Seizure activities are dominant in frequency less than 30 Hz [10, 11].

Another method was proposed by Ahmad et. al [12] where the planning was to predict epileptic seizure before it happens to a person. By using the spike averaging approach, spiking features of the signal are extracted as well as features which can be distinguished from the rest. The features will then be collected and analyzed in a 2D and 3D feature space. Classifying steps for the extracted features will be conducted using Support Vector Machine (SVM). Subspace method was used for extracting features in order to reduce the dimension of the data and represent them properly thus, it helps in extracting only dominant features. The classification is done by multiplying the basis and data vectors. The prediction property of this project will be very useful to the public, but it is has still not been implemented yet.

In a more recent study, Kumar et. al [13] utilized the wavelet entropy (WEN) technique to form feature vectors for classification of epileptic seizure. WEN measures the order or disorder of a signal. In this method, the EEG data was separated into six frequency sub-bands using the 5th level wavelet decomposition by applying discrete wavelet transform (DWT). Decomposed wavelet coefficient (detail and approximate) were passed through the high-pass and low-pass filter to filter any noise or artifacts. WEN values for each of the sub-bands were calculated to form the feature vectors to be analyzed further. Based on the results, the value of mean for the seizure data is higher than the non-seizure data while standard deviation and variance for the seizure data is less than the non-seizure data. This shows that a seizure activity signal has less data dispersion (more orderly) than that of a non-seizure data. The distinguishing features used are successful.

An SVD centered method was proposed by Shahid et. al [14] which employs the singular value based technique to detect seizure states. Singular values were calculated for each matrix data with 18 different channels and a window with a period of 1 second long. The distribution of energy within the matrix are represented by the maximum and minimum energy of the data [15]. The classifier will analyze the singular values to identify seizure activities.

CHAPTER 3

METHODOLOGY

3.1 Gant Chart and Key Milestone

Table 3.1 FYP 1

Activities/Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Selection of Project Topic														
Preliminary Research Work														
Submission of Extended Proposal														
Further Research on Methodology														
Research Completed							◆							
Proposal Defense														
Getting Data from CHB-MIT														
Extracting Features from WT Data														
Represent Results in Scaled Image												◆		
Interim Draft Report Submission														
Submission of Interim Report														

Table 3.2 FYP 2

Activities/Week	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Extracting classic features from WT data														
Represent results in frequency polygon plot														
Representation of results completed														
Progress Report														
Use fisher method to determine which feature to be analyzed														
Classification accuracy measured using SVM														

Milestone

Task

Important

3.2 Project Methodology

3.2.1 Sample Data

To be able to perform an EEG analysis using various analyzing method, sample data would be needed. The data will consist of brainwaves of patients which are obtained using the EEG technique. In this case, the data would be acquired from patients who are having epileptic seizures and also patients who are free from this disease. After analyzing both data, it will then be compared with one another to determine and detect at which part of the data does the brain react differently, and this will give us the characteristics that we need in order to identify an epileptic seizure case.

For this project, the set of data was obtained from the CHB-MIT EEG Scalp database. This set of data consist of many data from 23 different patients with each having 23 separate channels being recorded. The data was quite large as each subject has several EEG data with seizure and non-seizure activity. However, the type of seizure is not specified so it is assumed that all types of seizure are present in the data. An example of a patient's EEG data is shown below.

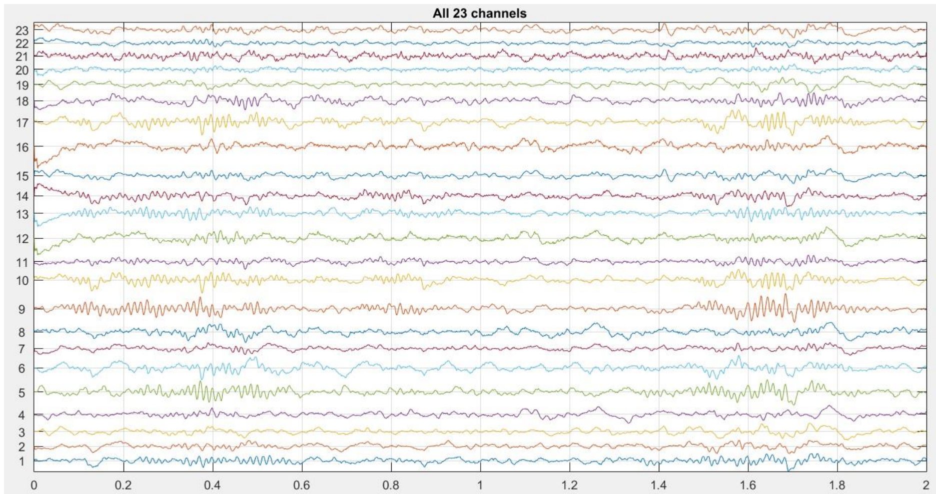

Figure 3.1 All channels of a patient using EEG scalp

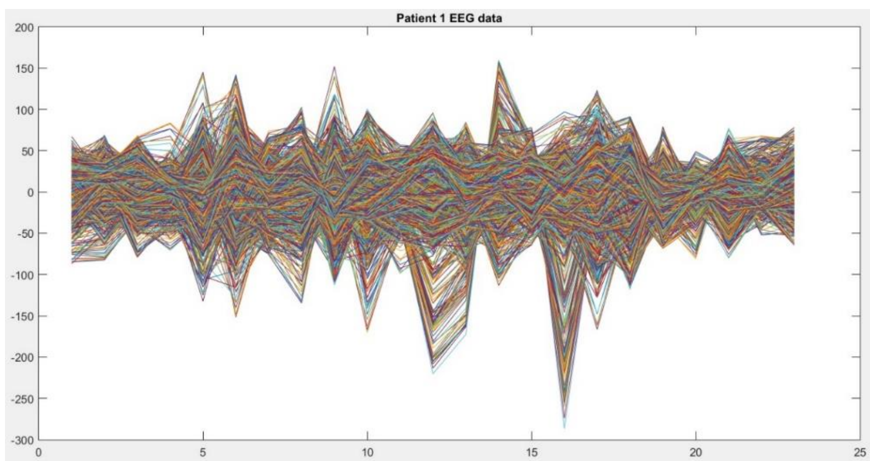

Figure 3.2 All channels are combined to become one waveform

3.2.2 Wavelet Transform (WT)

As per the title, this project will be focusing only on Wavelet Transform (WT) as a main method to analyze the signals. EEG signals have a non-stationary characteristics which means that the signals have a shifting statistical property [16]. Linear analyzing method such as FT which is only used for stationary signals is not a suitable method to be used in the characterization of EEG signal. Being a non-stationary signal, EEG signals are able to be analyzed by any methods which utilize the time-frequency analysis and WT is one of them. Even though WT employs the linear method, this method has a high successful rate in detecting epileptic seizures as it is able to precisely extract the brief and temporary features which are abnormal in the time and frequency domain [10]. WT provides an excellent accuracy on frequency information at low frequency as well as accurate time information at high frequency [17].

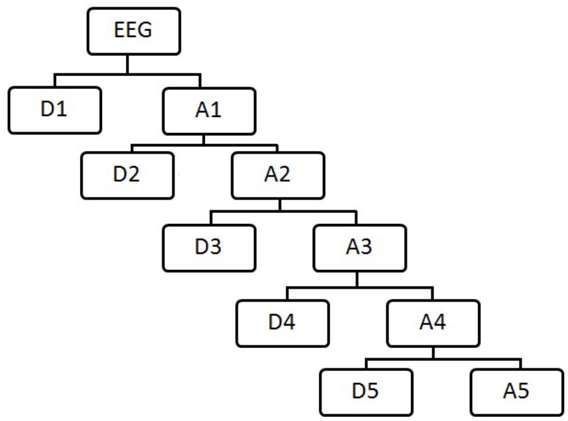

Figure 3.3 Five Wavelet Level Decomposition

A signal is represented by wavelet transform through the combination of many linear functions (wavelet functions) that was obtained from a single function which is called mother wavelet through the process of dilation and translation. The wavelets are categorized by two different labels, one for time and another for frequency [18]. To reconstruct the original signal, wavelet coefficients which are acquired from the decomposition of the signal are used together with the wavelet functions. For a precise reconstruction, the value for each wavelet coefficients need to be as accurate as possible so, a five level decomposition is applied as shown in Figure 3 based on the dominant frequency of the signal to obtain a certain frequency range. EEG signal has a wide range of frequency but not all frequencies give a useful information. Sometimes noise in the form of artifacts may exist inside the signal. It is concluded that EEG signals above 30Hz does not contain much information which are worth to analyze [10].

As a solution, the wavelet coefficients are utilized to the fifth level to get the lowest frequency as epileptic characteristics in the form of interictal spike discharges are most obvious in the low frequency range which is around 0-4Hz [10]. The approximate coefficients would be high scaled but has a low frequency while detailed coefficients are vice versa, making it low scaled but has a high frequency. Table 4 below shows the decomposition of signal's frequency. This method is applicable to any signal with a finite energy [19].

Table 3.3 Example of frequency ranges for Decomposed Signal

Decomposed signal	Frequency Band
Detailed Coefficient, D1	43.4026 - 86.805 Hz
Detail Coefficients, D2	21.7013 - 43.4025 Hz
Detail Coefficients, D3	10.8507 - 21.7012 Hz
Detail Coefficients, D4	5.4254 - 10.8506 Hz
Detail Coefficients, D5	2.7127 - 5.4253 Hz
Approximate Coefficients, A5	0 - 2.7126 Hz

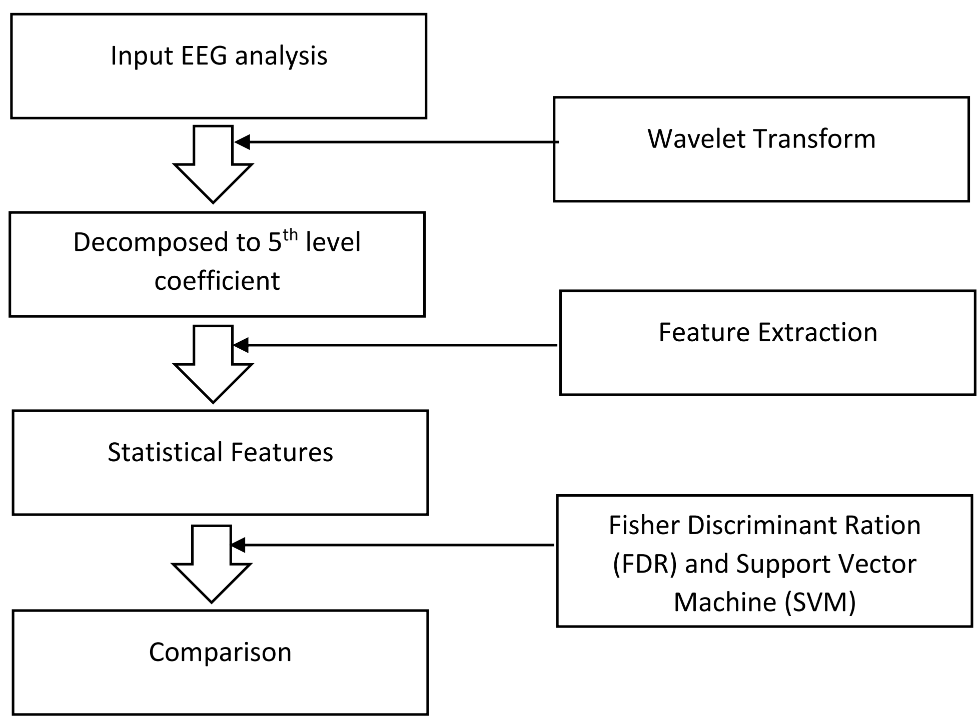

Figure 3.4 Block diagram for WT

Figure 4 above shows the general procedure on conducting a wavelet analysis. First of all, wavelet transform of 5^{th} level coefficient is applied to the EEG signal, separating them into its wavelet coefficients which are divided into approximation and detailed coefficient. Then, it will be divided further into various frequency sub-bands with each having their own frequency range. The sub-bands can be classified as delta (0-4Hz), theta (4-8Hz), alpha (8-12Hz), beta (13-30Hz) and gamma (>30Hz) [20, 21]. After that, certain features was extracted from each sub-bands using certain characteristics or properties such as energy, variance and so on. The purpose of feature extraction was to reduce the original signal to certain features which will make it easier to differentiate from one signal to another [21]. The signal of the sub-bands are then dedicated to FDR to choose which among the 23 channels are more dominant. From there, SVM was also applied to determine the classification error for the chosen (reduced) channels through FDR and for the overall channels. Comparisons will then be made to see if there are any difference between each of the features for reduced and overall channels.

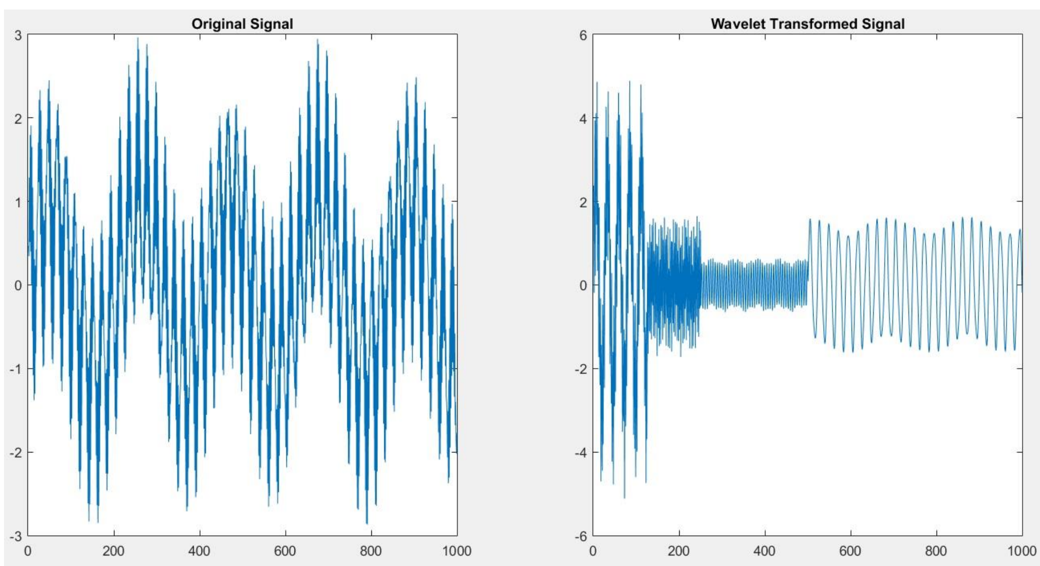

Figure 3.5 Comparison between original and decomposed signal

Figure 13 above shows the effect of WT when it is implemented on a signal. In this case, a 3^{rd} level decomposition of wavelet was used using daubechies (db1) wavelet transform. This results in an approximate coefficient at level 3 and detailed coefficient at level 1, 2 and 3. The figure below demonstrates the steps on getting the wavelet coefficients which is more or less similar to the wavelet decomposition figure which has been stated in the literature review above.

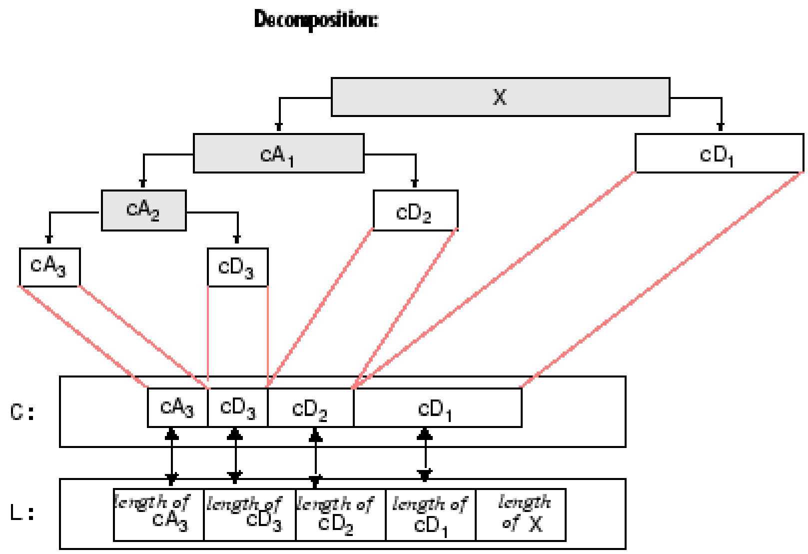

Figure 3.6 Decomposition steps for 3rd level

3.2.3 Feature Extraction

A 5th level (a5) decomposition was applied on the EEG signal using the daubechies (db4) wavelet. After the decomposition phase, the wavelet coefficients (approximate and detail) are classified into different frequency sub-bands depending on their frequency range. Amongst all the coefficient, only those which has an abundant of information on seizure activities will be retained for feature extraction. The 5th level coefficient is utilized as lower coefficient has a lower frequency band.

The features extracted were energy, coefficient of variation and two statistical features which are interquartile range and mean absolute deviation. During a seizure, a signal will display a strong rhythmic characteristic which means the same waveform will be repeated and at the same time exhibiting most of the energy in limited ranged scales while coefficient of variation evaluates the rhythmicity of a signal by its amplitude [11] so that is why these two factors are chosen.

These features are being extracted in two different types of data in order to compare their accuracy in detecting seizure. The main method was to extract the features from the normal EEG data without any alteration. However, the validity of the result will not be highly regarded as there are no multiple solutions to be compared to. In conjunction to this, another technique has been applied which is to extract the features from the delta frequency band of the data. According to [10], spikes during seizure or better known as interictal spikes are most obvious in the low frequency range which would make delta band the ideal frequency to be analyzed on as it has a frequency range of 0-4Hz. Delta band was obtained by filtering the data with a low pass Butterworth filter at the 6th level which indicates the delta frequency band. It is also filtered with a cutoff frequency so that only useful frequency range with usable data will be acquired.

The energy was calculated using the equation:

$$E(l) = \sum_{i=1}^{N} {x_n}^2$$

Where x_n represents the values of signal, N is the number of samples and l is the decomposed level. Coefficient of variation is given as:

$$Cov = \frac{\mu^2}{\sigma^2}$$

Where μ represents the mean and σ is the standard deviation. The statistical dispersion is calculated using interquartile range and it is given by:

$$IQR = Q3 - Q1$$

Q3 is the 'middle' value of the second half set while Q1 is the 'middle' value of the second half set of the data. Lastly, mean absolute deviation is the mean of the whole data which it was obtained by calculating the mean of the original data and subtract it with the original data. The absolute value was taken from each subtraction and the mean of the subtracted set of data is the mean absolute deviation.

Another set of features were also extracted besides the four above for comparison purposes. However, these features does not require for the original data to be decomposed using wavelet transform. Instead, it is extracted directly from the data as it is. The four features extracted are the mean, variance, power and power delta.

Mean was find for the intention of knowing the average of the total values in the data set which was calculated using:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Where X_i is the data value at instance i and N is the total number of data values.

Variance was calculated to determine the dispersion within the data set in other words, how far is the value of each data from the mean. Variance is given as:

$$\sigma^2 = \frac{\sum_{I=1}^{N} (X_i - \bar{X})^2}{N}$$

Where the mean, $\bar{X}$ is being subtracted from each data value and then squared. It is then divided by the total number of data. Power is to find the average energy of data is given by:

$$power = \frac{\sum_{i=1}^{N} (X_i)^2}{N}$$

For the last feature, power delta, the formula is the same as power which is stated above except, the difference lies on the data used. The delta frequency band was extracted from the data and it was used to calculate the power.

In addition to the previous methodology, some improvements have been made. Instead of just relying on the uncertain images of the spectrogram to differentiate between seizure and non-seizure conditions, a new method has been adopted which is the frequency polygon plot. Similar to spectrogram, it uses the four features extracted from the signal. This method is very identical to histogram except, there are no bars used to represent the data. The difference being that the data which were extracted are represented in lines instead of bars.

This method is implemented mainly because of its beneficial traits. The main purpose of it is to make analyzation easier by understanding the shape of distributions and it also helps in comparing sets of data. Data are more easily compared with a shape which is distributed in a simple manner, in this case, a single line. Comparison between data can easily be made by analyzing the distribution.

Frequency polygon plot was constructed in the same manner as histogram. In the beginning, the total number of bars and the class interval which will be used to represent each of the data is calculated using the Sturge's rule formula so that the bars are equally divided. The value is rounded off to the nearest integer. Sturge's rule formula is as follows:

$$k = 1 + log_2(n)$$

The width of each bins were calculated by dividing the product of subtraction between the upper side and lower side of each bin by two. This will result in a position in the middle of each bin where they are represented by dots. The dots are then connected together to form a line in which it becomes the frequency polygon plot.

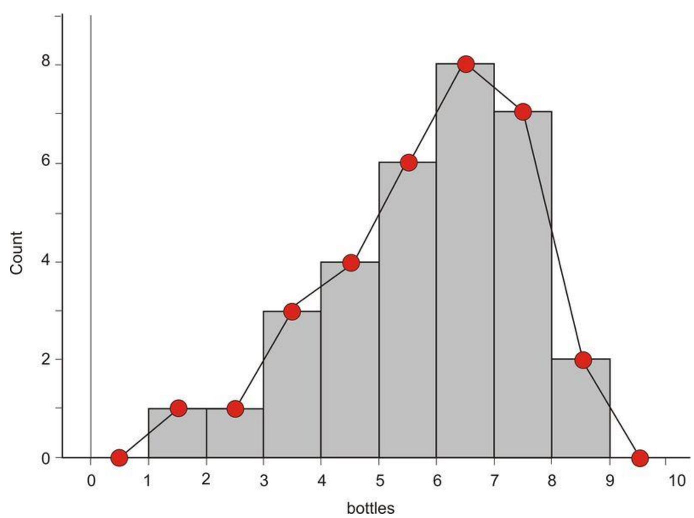

Figure 3.7 Example on plotting a frequency polygon plot

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Feature Extraction using WT and Classical Features

The EEG signals were recorded using 23 channels for most. In some cases, 24 or 26 were used. The signals were sampled at 256 per second with a 16-bit resolution. Each patient has an average amount of 6 files and each of the files has a different period of seizure and non-seizure activities. However, in this project, only a single patient is being analyzed which is patient 1 and it has a total of 7 files. Further details on the CHB-MIT data can be referred to Shuaib & Guttag [22].

Before applying the WT, the signals are preprocessed using a band pass filter at specific frequency in order to remove unwanted frequency and artifacts. After pre-processing, the EEG signals are divided into L sec-long intervals. Since the sampling frequency for the EEG recording is 265 Hz, one second in the recording is represented by 256 points. During the epileptic seizure detection stage, a sliding window of interval length 256 will be used.

5th level decomposition using daubechies (db4) wavelet [23] is applied to the EEG signals. The decomposition will form 2 types of wavelet coefficient which are approximate (a) and detail (d). Approximate coefficient (a5) will have a low frequency ranging between 0-4 Hz while detailed coefficient which comprised of 5th, 4th and 3rd (d5, d4 and d3) will have a high frequency value varying between 4-8 Hz, 8-16 Hz and 16-32 Hz respectively. These frequencies will be kept for the purpose of feature extracting. Since the seizure activities are superior in frequency range less than 30 Hz [10, 11], the frequencies collected have a limited range with a minimum and maximum frequency of 0 and 32 Hz. Hence, this frequency range can be used to classify between a seizure and non-seizure activities. The plotted figure of the combined extracted features (Energy, COV, IQR and MAD) which were mentioned above in the methodology section are shown in Figure 4.1.

In this experiment, WT is applied to the EEG signals and the four features of wavelet (Energy, COV, IQR and MAD) as well as the four features of classical (mean, variance, power and power delta) are calculated. The four features are displayed in two different formats, one as a combined features as shown in Figure 4.1 and the other as a frequency polygon plot shown by Figures 4.2 to 4.7.

All the results displayed below are conducted in phases. In the beginning phase, the extracted data are combined and represented in terms of color diagram. The figures below are the plot for combined features and frequency polygon of the extracted features for three different variables of the data which are wavelet transformed, delta band and classical features.

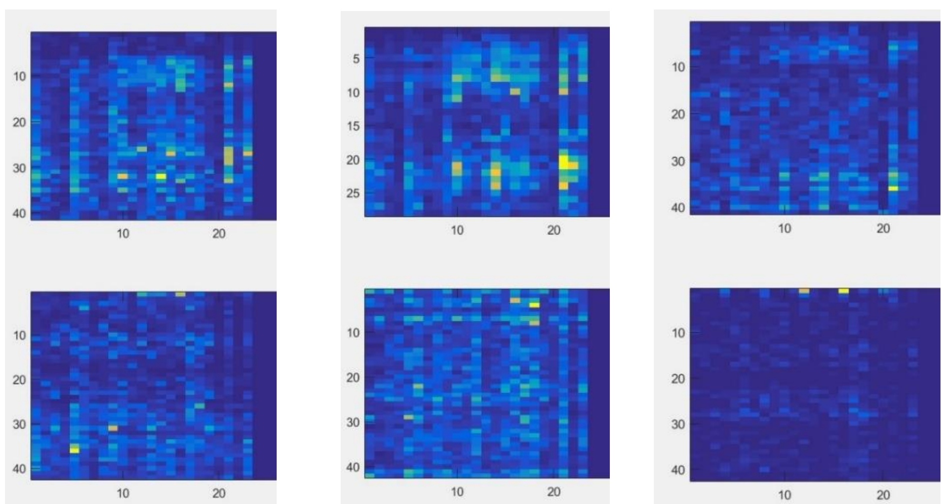

Figure 4.1 2-D plot of combined features (energy, coefficient of variation, interquartile, deviation) in matrix form of seizure (top row) and non-seizure (bottom row), EEG signals extracted using WT from 3 different files of Patient #1

From figure 4.1 above, the image plot of the combined features indicate differences between seizure and non-seizure EEG signal. In a glance, the seizure signals have more pixel variation than the non-seizure signals. In particular, the seizure image exhibits more frequent variety of colors such as yellow, orange and green while the non-seizure image does not show as many colors.

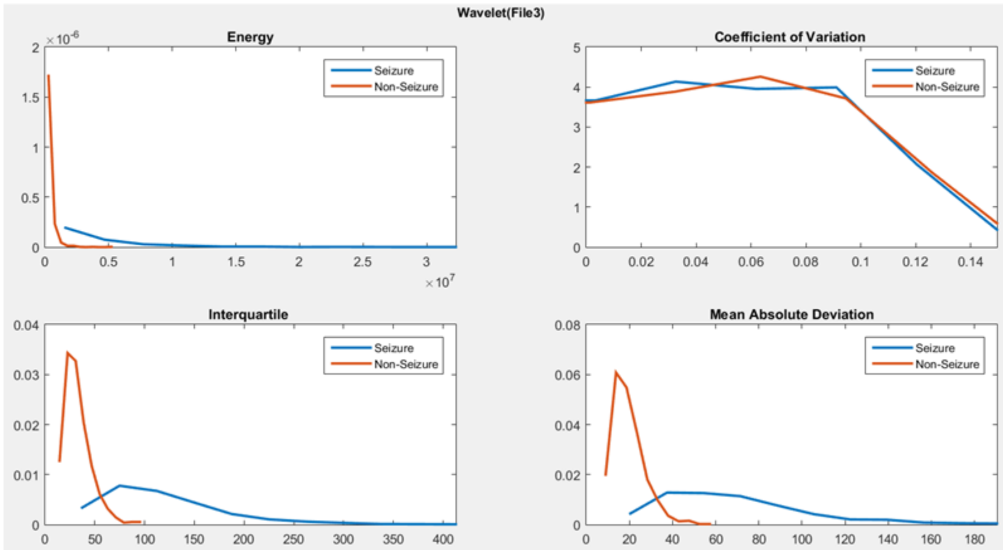

Figure 4.2 Polygon plot of energy, COV, IQR and MAD for WT data (a) File 3

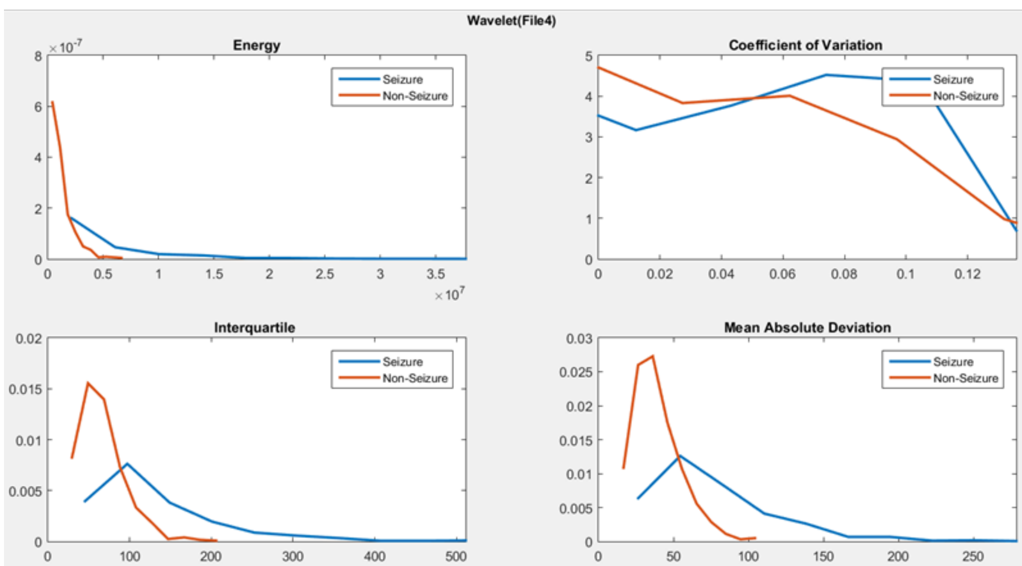

Figure 4.3 (b) File 4

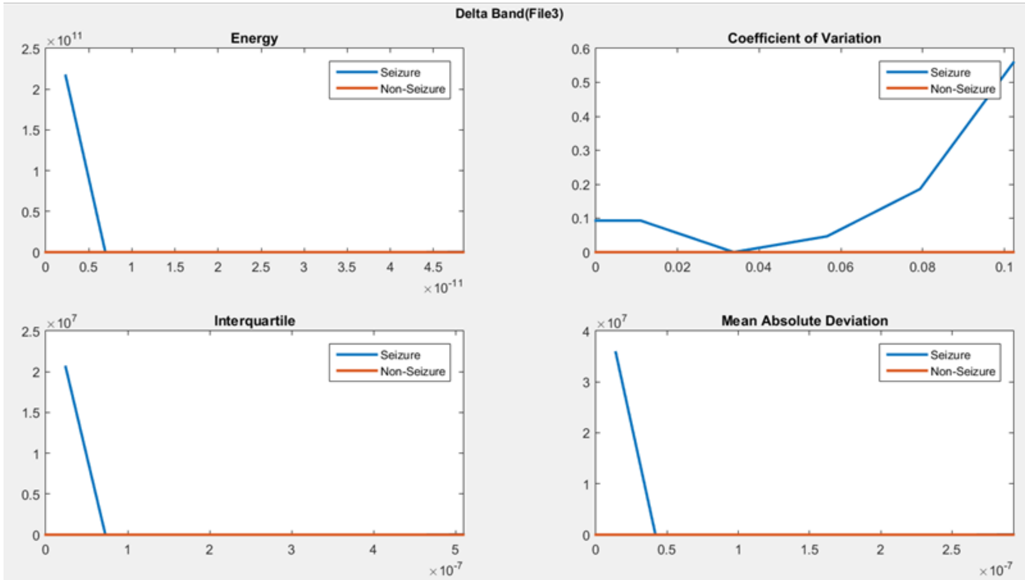

Figure 4.4 Polygon plot for delta band data (c) File 3

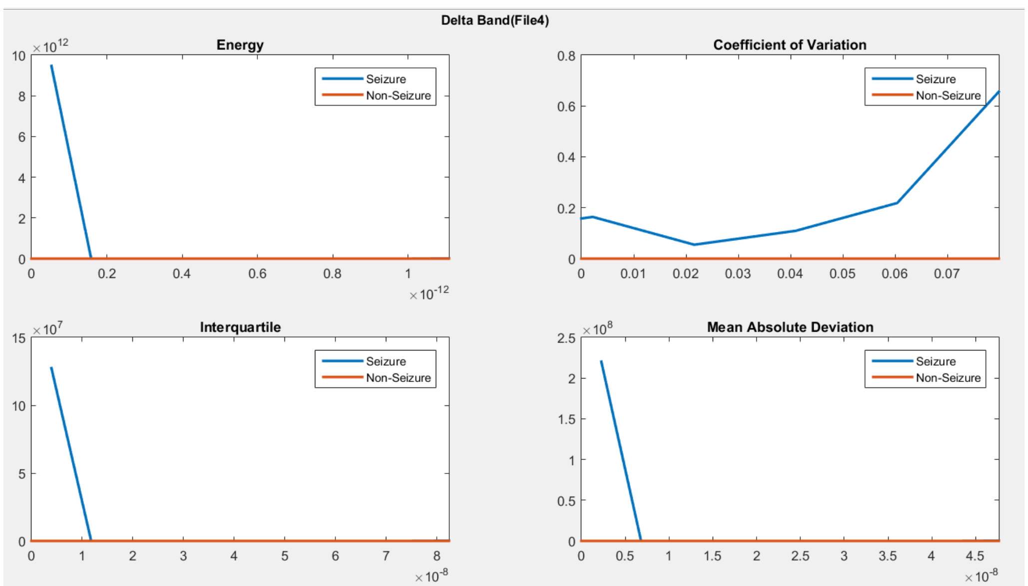

Figure 4.5 (d) File 3

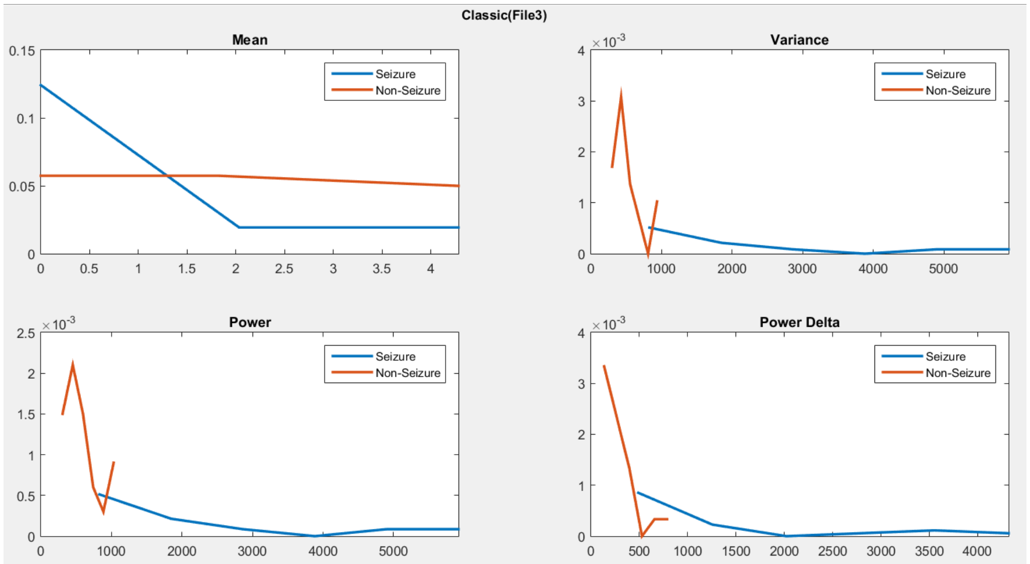

Figure 4.6 Polygon plot of mean, variance, power and power delta for classical based (e) File 3

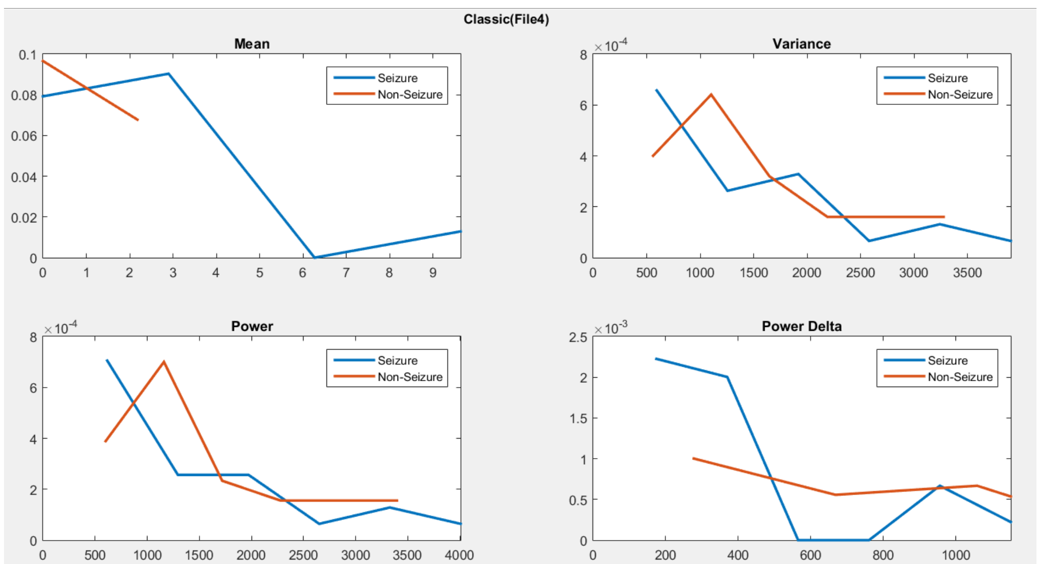

Figure 4.7 (f) File 4

The polygon plots in Figures 4.2 to 4.7 clearly show the different signatures for wavelet and delta based features which are energy, COV, IQR and MAD as well as classical based features which are mean, variance, power and power delta between seizure and non-seizure signals of different files (File 3 and 4). The results for the other files are not shown as they produce similar results. The difference in signatures between the two signals can be used to detect seizure in EEG signal.

For figures 4.2 and 4.3 of the WT data, the distinctiveness between seizure and non-seizure activity can be seen clearly form the energy, IQR and MAD features but not COV. Overlapping between seizure and non-seizure curves are minimum in the three features so the differences can be observed clearly while the same cannot be said for COV. Both curves have almost the same pattern which results in a complication during comparison. To summarize, the features which produces reliable results are all except COV as it shows the least variation between seizure and non-seizure.

From figures 4.4 to 4.5 above, we can conclude that the delta frequency band data extracted features results can only be depended on one feature which is the COV. The produced results of seizure and non-seizure curves are quite dependable as the pattern are completely different so the differences are able to be differentiated clearly. For the other three features, there are no changes in both curves patterns for most cases so it is likely unreliable.

Figures 4.6 to 4.7 shows the seizure and non-seizure curves for four classic features which are mean, variance, power and power delta. From the results, it seems there are no certain patterns for both curves. The curves appear to have a randomized shape which is not very reliable to differentiate between seizure and non-seizure activity as observed from the 3 files above. Plotting graph gives vague results for these features. However, other methods may be utilized to make use of the data in a better way.

4.2 Fisher Discriminant Ratio (FDR) Test on WT and Classical Features

Before applying SVM to calculate the classification error for the EEG signals, FDR is implemented beforehand to save time and increase accuracy in analysis of EEG signal. FDR is performed on both wavelet and classical based features to determine amongst the 23 channels, which will be used for the reduction feature analysis. Figures below show the ranking for each channel for both wavelet and classical based features for all files of patient 1.

Figures 4.8 to 4.11 below shows the FDR results for wavelet based.

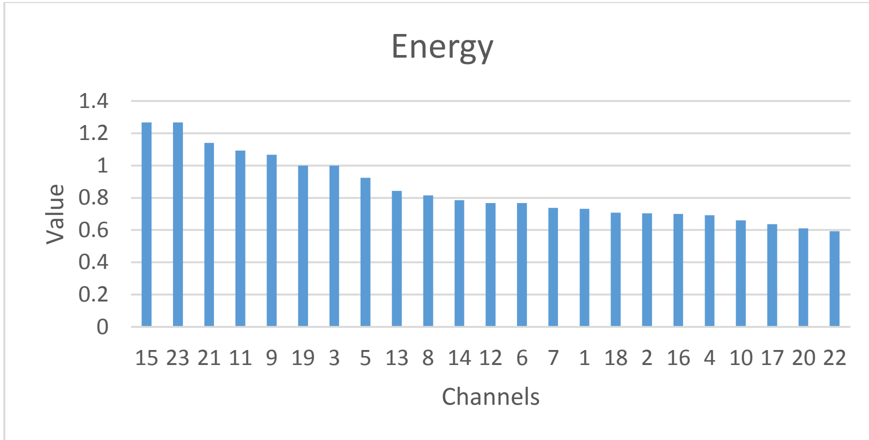

Figure 4.8 Energy

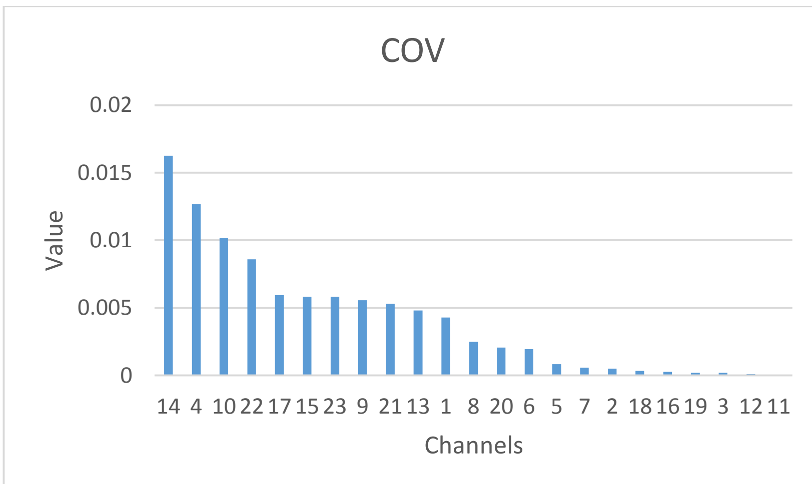

Figure 4.9 COV

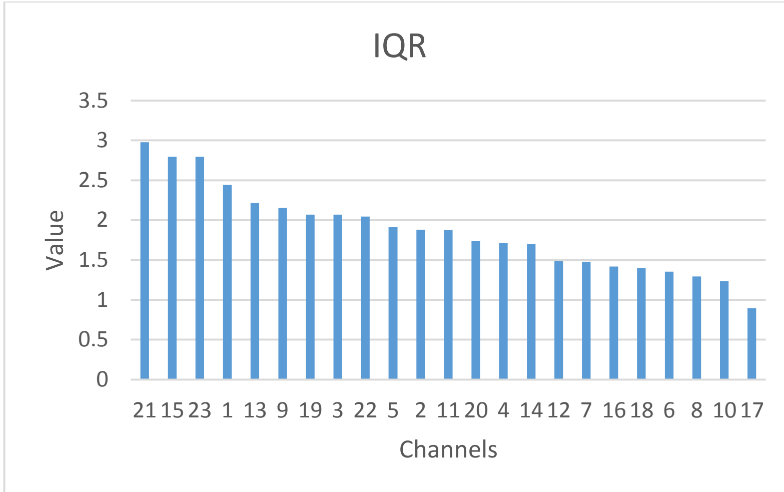

Figure 4.10 IQR

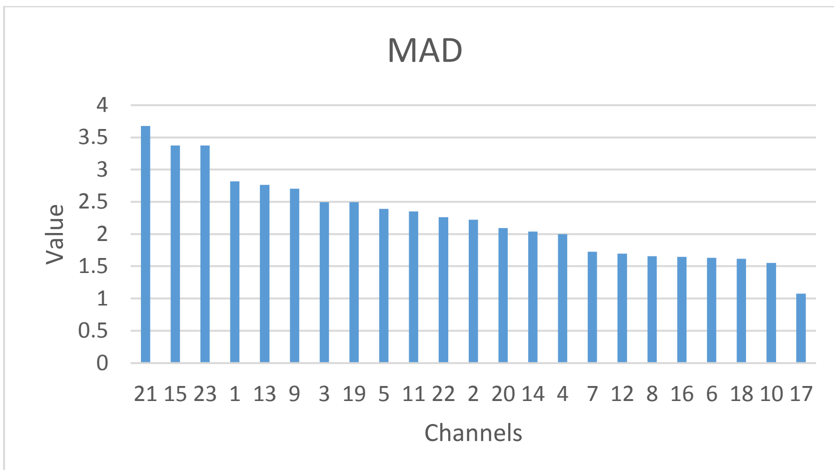

Figure 4.11 MAD

Figures 4.12 to 4.15 below shows the FDR results for classical based.

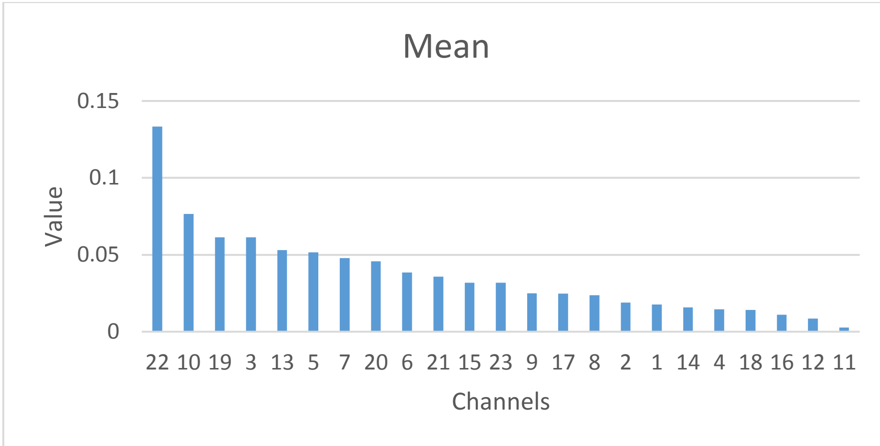

Figure 4.12 Mean

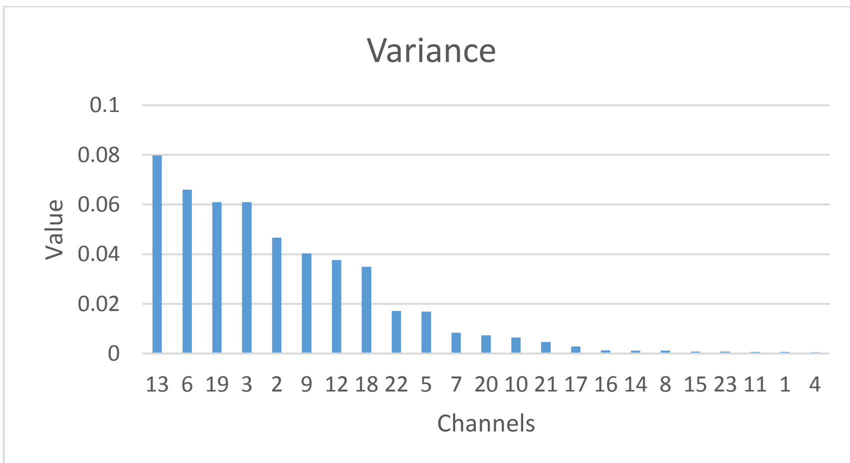

Figure 4.13 Variance

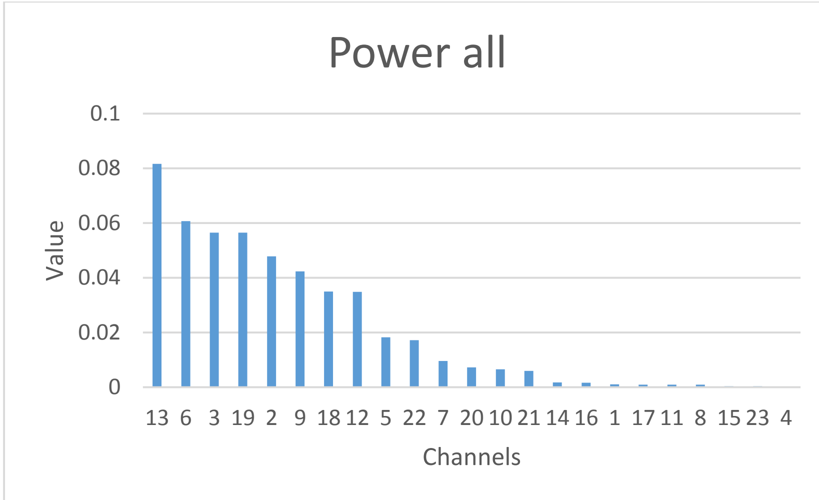

Figure 4.14 Power all

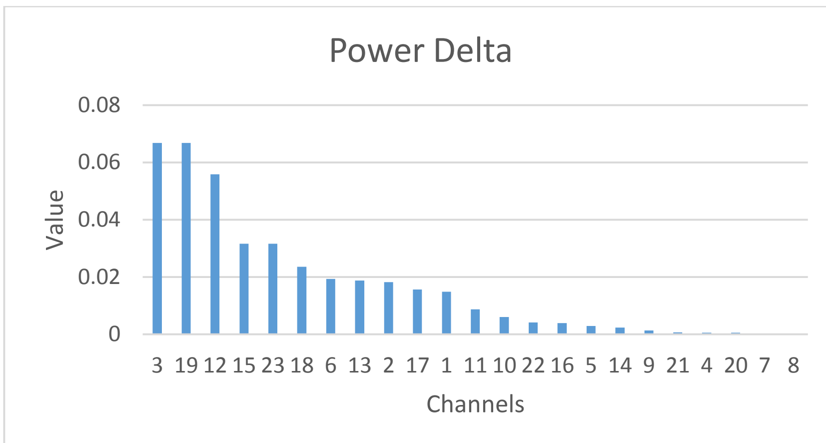

Figure 4.15 Power delta

Based on the FDR results above, it can be observed that for each feature, there are certain channels that has a dominant value over the other. Channel at the most left is the channel which has the highest ranking. In this work, 8 most dominant channels which is exhibited by channels to the left will be chosen to be analyzed in reduction feature analysis instead of using all 23 channels to make analysis faster and increasing the accuracy.

4.3 Classification Error of EEG Signals for WT and Classical Features

To achieve more dependable results, FDR and SVM were implemented to the EEG data. These are the final results obtained for this project. Figures below show the comparison that was made between wavelet and classical based features for reduced and overall channels for all files of patient 1.

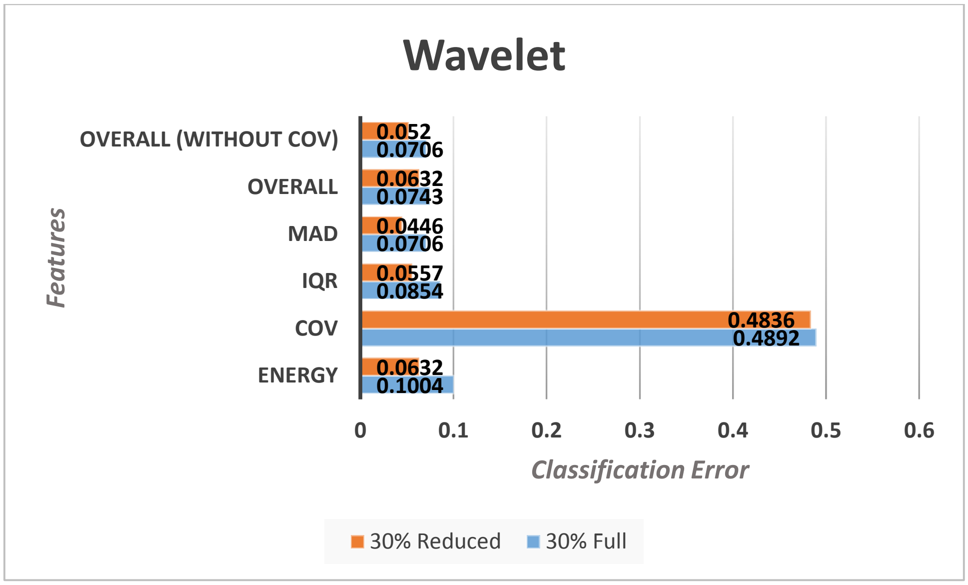

Figure 4.16 Classification error for wavelet based feature

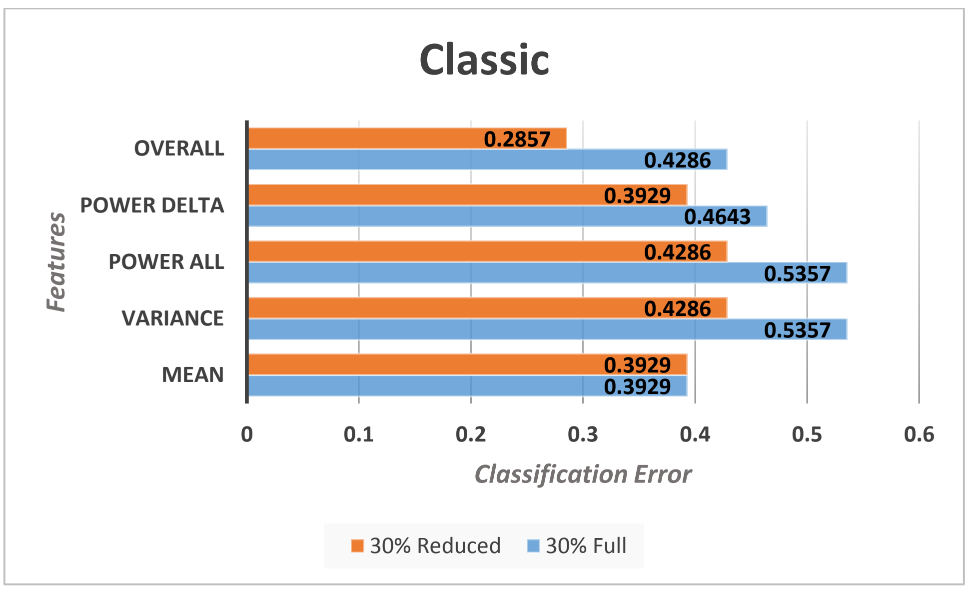

Figure 4.17 Classification error for classical based features

Figures 4.16 to 4.17 above indicate the classification error values for wavelet and classical based features. For each figure, the value of percentage at the bottom represents the holdout value used in SVM. It is the value of the EEG data which is used as a testing sample. In this case, the holdout value is 30%, the other 70% of the data is used for the calculation of the classification error based on the 30% testing sample. Reduced based feature is when FDR is applied to the overall channel. Through this technique, each of the channel will be arrange in terms of its dominance over the others. The top 8 channels will then be chosen for reduced analysis instead of all 23 channels.

Wavelet based features results are represented by figure 4.16. The features extracted for wavelet based were energy, MAD, COV, IQR, overall and overall without COV. As observed from the bar charts, the overall classification error for both reduced and overall channel are relatively low except for COV and the feature which generates the lowest classification error is determined to be MAD for most feature. Since COV has the highest classification error, it will not be chosen as a feature to be extracted and analyzed. Instead, to get maximum accuracy, MAD is the feature to be analyzed. In accordance to the unusual high classification error of COV, another feature has been added for analysis which is the combination of all features excluding COV that is represented by overall (without COV). It is compared with the overall feature including COV and it is proven that without COV, the classification error is lower. Comparing the reduced and overall channel, reduced channel produces a lower classification error than overall channel even though it not so obvious.

As for classical based features, they are represented by figure 4.17. The features extracted were mean, variance, power, power delta and overall. From the bar chart, it can be perceived that as a whole, the features produces a higher classification error than wavelet based. Even then, the trend of reduced feature producing a lower classification error than full feature still exist in classical based. Although, there are a few features which generate a lower classification error such as mean, it is still not as low as wavelet based features. Since it generally produces a higher classification error, classical based features may not be suitable to be extracted to accurately analyze epileptic seizure in EEG data.

CHAPTER 5

CONCLUSION

In the first phase, the results of the extracted features were represented in the frequency polygon plot. Since the features were extracted from different types of data, the results produced for the seizure and non-seizure curves vary from one type to another. For some cases only COV is reliable and other cases proved that all except COV is dependable. Even though there are variation of results, seizure and non-seizure activity can be observed clearly using this plot.

In the second phase, FDR is applied and it is perceived that for wavelet based features channel 21 exhibits high FDR value for 2 features while channel 13 displays high FDR value for 2 features in classical based.

Based on the final results, it can concluded that wavelet based features produces a lower classification error than classical based features which means that in terms of accuracy, wavelet based are better than classical based in analyzing epileptic seizure with MAD generating the lowest classification error of 0.0391 that is equivalent to about 96% if converted to accuracy rate. The same can be said for reduced and full features. Since reduced features produce a lower classification error than full features, it is also better in terms of accuracy.

REFERENCES

[1] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee, *et al.*, "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia,* vol. 46, pp. 470-472, 2005.

[2] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, *et al.*, "ILAE official report: a practical clinical definition of epilepsy," *Epilepsia,* vol. 55, pp. 475-482, 2014.

[3] D. J. Thurman, E. Beghi, C. E. Begley, A. T. Berg, J. R. Buchhalter, D. Ding, *et al.*, "Standards for epidemiologic studies and surveillance of epilepsy," *Epilepsia,* vol. 52, pp. 2-26, 2011.

[4] J. A. Wilden and A. A. Cohen-Gadol, "Evaluation of first nonfebrile seizures," *Am Fam Physician,* vol. 86, pp. 334-40, 2012.

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

[6] Y. Yuan, "Detection of epileptic seizure based on EEG signals," in *Image and Signal Processing (CISP), 2010 3rd International Congress on*, 2010, pp. 4209-4211.

[7] L. Boubchir, S. Al-Maadeed, and A. Bouridane, "On the use of time-frequency features for detecting and classifying epileptic seizure activities in non-stationary EEG signals," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, 2014, pp. 5889-5893.

[8] W. Min and G. Luo, "Medical applications of EEG wave classification," *Chance,* vol. 22, pp. 14-20, 2009.

[9] R. Q. Quiroga, S. Blanco, O. Rosso, H. Garcia, and A. Rabinowicz, "Searching for hidden information with Gabor Transform in generalized tonic-clonic seizures," *Electroencephalography and clinical Neurophysiology,* vol. 103, pp. 434-439, 1997.

[10] E. S. Kollialil, G. K. Gopan, A. Harsha, and L. A. Joseph, "Single feature-based non-convulsive epileptic seizure detection using multi-class SVM," in *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on*, 2013, pp. 1-6.

[11] N. Rafiuddin, Y. Uzzaman Khan, and O. Farooq, "Feature extraction and classification of EEG for automatic seizure detection," in *Multimedia, Signal Processing and Communication Technologies (IMPACT), 2011 International Conference on*, 2011, pp. 184-187.

[12] R. F. Ahmad, A. S. Malik, N. Kamel, and F. Reza, "A proposed frame work for real time epileptic seizure prediction using scalp EEG," in *Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on*, 2013, pp. 284-289.

[13] Y. Kumar, M. Dewal, and R. S. Anand, "Wavelet entropy based EEG analysis for seizure detection," in *Signal Processing, Computing and Control (ISPCC), 2013 IEEE International Conference on*, 2013, pp. 1-6.

[14] A. Shahid, N. Kamel, A. S. Malik, and M. A. Jatoi, "Epileptic Seizure Detection using the singular values of EEG signals," in *Complex Medical Engineering (CME), 2013 ICME International Conference on*, 2013, pp. 652-655.

[15] B. De Moor, J. Staar, and J. Vandewalle, "Oriented energy and oriented signal-to-signal ratio concepts in the analysis of vector sequences and time series, SVD and signal processing: algorithms, applications and architectures," ed: North-Holland Publishing Co., Amsterdam, The Netherlands, 1989.

[16] L. Chen, E. Zhao, D. Wang, Z. Han, S. Zhang, and C. Xu, "Feature extraction of EEG signals from epilepsy patients based on Gabor Transform and EMD Decomposition," in *Natural Computation (ICNC), 2010 Sixth International Conference on*, 2010, pp. 1243-1247.

[17] K. A. H. Kulasuriya and M. U. S. Perera, "Forecasting epileptic seizures using EEG signals, wavelet transform and artificial neural networks," in *IT in Medicine and Education (ITME), 2011 International Symposium on*, 2011, pp. 557-562.

[18] L. Yong and Z. Shengxun, "Apply wavelet transform to analyse EEG signal," in *Engineering in Medicine and Biology Society, 1996. Bridging Disciplines for Biomedicine. Proceedings of the 18th Annual International Conference of the IEEE*, 1996, pp. 1007-1008.

[19] F.-C. Tsui, C.-C. Li, M. Sun, and R. J. Sclabassi, "A comparative study of two biorthogonal wavelet transforms in time series prediction," in *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, 1997, pp. 1791-1796.

[20] A. Kumar and M. H. Kolekar, "Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals," in *Medical Imaging, m-Health and Emerging Communication Systems (MedCom), 2014 International Conference on*, 2014, pp. 412-416.

[21] R. Shantha Selva Kumari and J. Prabin Jose, "Seizure detection in EEG using time frequency analysis and SVM," in *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*, 2011, pp. 626-630.

[22] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 975-982.

[23] Y. Khan and J. Gotman, "Wavelet based automatic seizure detection in intracerebral electroencephalogram," *Clinical Neurophysiology,* vol. 114, pp. 898-908, 2003.