Classification of Focal and Non-focal Epileptic EEG signals in EMD-Wavelet Domain

A thesis submitted to the

Department of Electrical and Electronic Engineering (EEE)

of

Bangladesh University of Engineering and Technology (BUET)

In partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

by

Anindya Bijoy Das

(Roll No.: 1014062224 P)

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING (EEE)

BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY (BUET)

December 2017

The thesis titled "**Classification of Focal and Non-focal Epileptic EEG signals in EMD-Wavelet Domain**" submitted by Anindya Bijoy Das, Roll No.: 1014062224 P, Session: October 2014, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING on December 31, 2017.

Board of Examiners

31.12.17

1. **Dr. Mohammed Imamul Hassan Bhuiyan** — Chairman (Supervisor)
   Professor
   Department of EEE, BUET, Dhaka.

2. **Dr. Quazi Deen Mohd Khosru** — Member (Ex-Officio)
   Professor and Head
   Department of EEE, BUET, Dhaka.

3. **Dr. Md. Kamrul Hassan** — Member
   Professor
   Department of EEE, BUET, Dhaka.

4. **Dr. Khawza Iftekhar Uddin Ahmed** — Member (External)
   Professor
   Department of EEE
   United International University
   UIU Bhaban, House No 80, Road No 8/A,
   Satmasjid Road, Dhanmondi, Dhaka-1209

Declaration

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Signature of the candidate

Anindya Bijoy Das

Anindya Bijoy Das

(Roll No.: 1014062224 P)

Dedication

to my Family

Acknowledgements

I would like to convey my heart-felt gratitude and thanks to my thesis supervisor Dr. Mohammed Imamul Hassan Bhuiyan, Professor of the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka for the continuous guidance and support I have received from him during this work. He was there with his kind help and direction anytime I needed them. His inspiration, motivation and friendly attitude helped make my work under him a great experience. Desire and ambition to emulate his knowledge and expertise in the field of research have also acted as a motivation to carry through my work. It was a great pleasure to be a part of his research group and work under him.

I would also like to express my appreciation to the members of my thesis committee, Dr. Quazi Deen Mohd Khosru, Dr. Md. Kamrul Hassan and Dr. Khawza Iftekhar Uddin Ahmed, for their valuable feedback and suggestion on my work.

In addition, my heart-felt thank goes to all the faculty members and my course teachers for their continuous help and guidance throughout the course of my M.Sc. program. I also thank my family members, without whose support and encouragement, it would be impossible to carry on with my study smoothly.

Last but not the least, I want to express my gratitude to the Almighty for all the favors, material and non-material, He has provided me with.

Abstract

In this thesis, a comprehensive analysis of focal and non-focal electroencephalography is carried out in the empirical mode decomposition (EMD) and discrete wavelet transform (DWT) domains. First, the analysis is carried out in the EMD domain and its variants, for example, in ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) domains. A number of spectral entropy-based features such as the Shannon entropy, log-energy entropy and Renyi entropy are calculated in EMD, EEMD and CEEMDAN domains. In lieu of using the direct signals from the EEG channels, the differences between two adjacent EEG channels are used due to its robustness to noise and interference. The EEG signals are obtained from a publicly available electroencephalography database that consists of 7500 signal pairs which contain over 80 hours of electroencephalogram data collected from five epilepsy patients. Then, the ability of the entropy-based features in separating the focal and non-focal EEG signals is explored utilizing the one-way ANOVA analysis and the box-whisker plots. After that, well-known classifiers like support vector machine (SVM) and k-nearest neighbor (KNN) have been utilized to classify focal and non-focal EEG signals.

Next, similar analysis is carried out in discrete wavelet transform domains and the efficacy in discriminating the focal and non-focal EEG signals is investigated. It is observed that the entropy-based features perform better in DWT domain to classify the EEG signals than in EMD domain. In this regard, it is interesting to investigate the capability of the same features to discriminate the EEG data in the combined EMD-DWT domain. It is shown that in the log-energy entropy, when calculated in the combined EMD-DWT domain, gives a better discrimination of these signals as compared to that of the other entropy measures as well as to that obtained in EMD or DWT domain, and utilizing a KNN classifier, it provides 89.4% accuracy (with 90.7% sensitivity), which is higher than that of the state-of-the-art methods. Overall, the proposed classification method reports a significant improvement in terms of sensitivity, specificity and accuracy in comparison to the existing techniques. Besides, for being computationally fast, the proposed method has the potential for identifying the epileptogenic zones, which is an important step prior to resective surgery usually performed on patients with low responsiveness to anti-epileptic medications. The analysis may encourage the researchers to develop improved algorithms to classify these signals.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

About 1% people of the world is affected with epileptic seizure which is a common neurological disorder of the brain characterized by recurrent seizures [1]. When epilepsy attacks a limited area of the brain, then it is called as the focal or partial epilepsy. Among various types of focal epilepsy, simple partial seizures affects a small part of one of the temporal lobes, and it is often a precursor to a larger seizure named as complex partial seizure. Generalized seizures include tonic-clonic (grand mal), absence (petitmal), myoclonic, clonic, tonic, and atonic seizures [2]. Although anti-epileptic drugs are available, there is a problem with the 25% of the epilepsy patients who do not respond well to these drugs [3]. In such cases, the treatment option is resective surgery where a section of the brain, identified as the epileptogenic focus (or the onset of early seizure), is removed. However, such surgeries are not risk free especially in the case of multi-focal epilepsy and may affect other eloquent areas responsible for language, primary motor and vision. Thus, efficient methods for the identification of those epileptogenic area of the brain is an important step prior to surgery.

Epilepsy is a common health problem in Bangladesh. It is estimated that there are at least 1.5 to 2.0 million epilepsy patients in Bangladesh. 30-40% of patients are still treated by traditional healer. The most common cause of non-compliance is cost of drug. 50-60% patients remain symptoms free with 4 common drugs. Epidemiological study was conducted at Epilepsy Clinic, Neurology foundation Hospital, Dhaka, Bangladesh, and total of 2200 patients were included. Men are more often affected than female and rural populations are affected more than the urban populations. The common ages of epileptic patients in Bangladesh are between 16 to 31 years. The etiology varies with age. Birth trauma, birth asphyxia, central nervous system infections are common in neonate and infancy whereas head trauma, brain tumor, stroke, infections are common causes in middle aged and elderly. Appropriate antiepileptic drugs are sometime unavailable in Bangladesh. The BSMMU study showed 23% of patients found it difficult to continue treatment due to financial problem.

1.1 Epilepsy and EEG Signals

Epilepsy is a neurological disorder which is characterized by a predisposition to generate seizures, a transient occurrence that arises due to abnormal and excessive or hyper-synchronous activity in the brain. It is one of the most common diseases of the human brain, with a prevalence of more than 3 million patients in Europe alone. Epilepsy is characterized by sudden changes in brain dynamics that lead to abnormal synchronization of extended brain networks, which is called seizure. These seizures are characterized by transient impairments of sensation, thinking and motor control and intermittent abnormal ring of neurons in the brain. Brain activity in the ictal state (during a seizure) differs significantly from the activity in the normal state with respect to frequency and pattern of neuronal ring. Temporal lobe epilepsy (TLE) is probably the most common focal epilepsy in humans. TLE consists of simple partial seizures without loss of awareness and complex partial seizures (i.e. with loss of awareness).

Electro-encephalogram (EEG) contains a set of electric potential differences developed as a result of volume currents spreading from active neural tissue throughout the conductive media of the brain. These measurements can be obtained either using sensors on the scalp or by placing special intra-cranial electrodes. About 50 million people worldwide are suffering from epilepsy and 85% of those live in developing countries. Each year 2.4 million new cases are estimated to occur globally. In most patients, seizures are infrequent, occupying much less than 0.1% of the time. The unpredictable nature of seizures may lead to unexpected injury. It has considerable economic implications in terms of health care needs, premature mortality, and losses in productivity. Patients are, however, suffering from restrictions in several domains, e.g., physically due to the risk of trauma, socially due

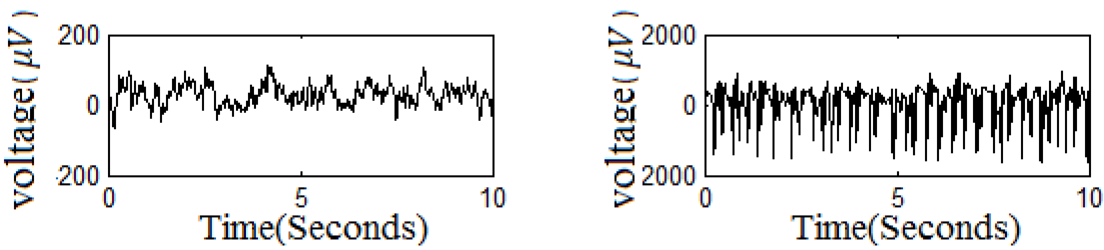

Fig. 1.1. EEG Signals from a normal person (left) and a patient (right)

to driving and occupational restrictions, psychologically due to a feeling of helplessness and sometimes, it may lead to death of the matured persons. The EEG records can easily display these electrical discharges as a rapid change in potential differences, and thus have long been used for the diagnosis of epilepsy. 10 seconds EEG data from a healthy person and a patient are shown in Fig. 1.1.

1.2 Different Kinds of Epileptic Seizures

Doctors have described more than 30 different types of seizures. They classify seizures by how much of the brain is affected. Seizures are divided into two major categories- focal seizures and generalized seizures. However, there are many different types of seizures in each of these two categories and there are some unclassified which do not fit into the previous categories.

Focal seizures, also known as local or partial seizures, are caused by abnormal electrical activity in a specific, smaller part of the brain. The part of the brain causing the seizure is called the seizure focus. Focal seizures are divided into simple and complex seizures. Some focal seizures evolve into generalized ones and are called secondarily generalized seizures.

During the simple focal seizures, you remain conscious although some people can't speak or move until the seizure is over. Uncontrolled movements, such as jerking or stiffening, can occur throughout your body. You also may experience emotions such as fear or rage or even joy; or odd sensations, such as ringing sounds or strange smells. During complex focal seizures, you are not fully conscious and may appear to be in a dreamlike state. Typically, they start with a blank stare. You may involuntarily chew, walk, fidget, or perform other repetitive movements or simple actions, but actions are typically unorganized or confused. These seizures typically last between 30 seconds and a minute. The secondarily generalized seizures seizures begin as a focal seizure and develop into generalized ones as the electrical abnormality spreads throughout the brain. When the seizure begins, you may be fully conscious but then lose consciousness and experience convulsions as it develops.

Generalized seizures are a result of abnormal neuronal activity on both sides of the brain. These seizures may cause loss of consciousness, falls, or massive muscle spasms. There are mainly six types of generalized seizures. Absence seizures, also called petit mal seizures, are characterized by staring and subtle body movement. These seizures can cause a brief loss of awareness. Tonic seizures cause stiffening of your muscles. These seizures usually affect muscles in your back, arms and legs and may cause you to fall to the ground. Clonic seizures are associated with rhythmic, jerking muscle movements. These seizures usually affect the neck, face and arms. Myoclonic seizures usually appear as sudden brief jerks or twitches of your arms and legs. Atonic seizures, also known as drop seizures, cause a loss of muscle control, which may cause you to suddenly collapse or fall down. Tonic-clonic seizures, also called grand mal seizures, are characterized by a loss of consciousness, body stiffening and shaking, and sometimes loss of bladder control or biting your tongue.

1.3 Literature Review

In [4], EEG signals obtained (during seizure) from the epileptogenic area are named as focal EEG signals, whereas those originating from other areas as non-focal signals. Several methods are available in the literature for identifying the epileptogenic focus by classifying the focal and non-focal EEG signals [5]-[10]. Epileptic seizures are often detected from time or frequency or time-frequency domain analysis of EEG signals [11]-[32]. Two steps of these detection processes include appropriate feature extraction and classification. Appropriate features may include entropy, higher order statistics, etc. whereas the classification may be carried out using well-known classifiers like support vector machine (SVM) or k-nearest neighbor (KNN).

Among various types of features which have been shown to be quite effective to differentiate the EEG signals, higher order moments, like variance, skewness and kurtosis have been used for discriminating EEG signals in [11], [30]; whereas Lyapunov exponents and correlation dimension are used in [12], [16]. In [15], the EEG signals and their sub-bands are modeled by a probability density function, and then, the model parameters are utilized to separate various types of EEG signals. Various types of entropy measures like Shannon entropy or Renyi entropy, are used in [16], [24] to classify the EEG signals. The entropy-based features are considered due to their successful application in detecting epileptic activities [7], [16]. The motivation behind utilizing the spectral entropy measures like Shannon entropy comes from its previous success to provide the measurement of the variability related to different sub-bands to classify the EEG signals [5]. The log-energy entropy is also considered to characterize the complexity related to the EEG sub-bands for classifying the EEG data successfully [24]. Besides, Renyi entropy provides the measurement of the randomness of the wavelet coefficients of the EEG signals [5]. In [25], the correlation among the sub-bands of the EEG signals has also been introduced to differentiate the EEG signals.

However, among the various techniques to classify focal and non-focal EEG signals, methods employing entropy-based features obtained from the EEG signals in discrete wavelet transform (DWT) and empirical mode decomposition (EMD) domain are shown to be most promising [5], [7]. It should be noted that the effectiveness of wavelet-based classification depends on the appropriate choice of the basis function [23]. On the other hand, the EMD is data driven and does not require a prior basis function [13], [28]. As noted in earlier works [12], the characteristics of EEG signals are spread over different frequency bands, well described by DWT sub-bands. In [14], it is shown that epileptic seizures can be identified effectively from the Hilbert magnitude spectrum of the intrinsic mode functions (IMF) of EEG signals. The DWT is also an orthogonal transform as the Hilbert Transform while providing the time-frequency representations of a signal.

Furthermore, it has also been shown in earlier works [33], [34] that non-stationary signals such as ECG can be effectively analyzed by processing the DWT sub-bands obtained from the IMFs in the EMD domain. Thus, it would be interesting to study whether the focal and non-focal EEG signals are classified better in the combined EMD-DWT domain.

1.4 Objectives of the Thesis

The objectives of this thesis are:

- (i) To develop a statistical method to classify focal and non-focal EEG signals.
- (ii) To investigate the capability of different spectral entropy features to differentiate focal and non-focal EEG signals.
- (iii) To study the EEG signals in various time-frequency domains, like EMD, EEMD and CEEMDAN domains.
- (iv) To explore the EEG signals in a hybrid domain, like in EMD-DWT domain.

(v) To investigate one-way analysis of variance (ANOVA) and receiver operating characteristic (ROC) curves for different entropy-based features

(vi) To study the performance of well-known classifiers like KNN and SVM to classify the EEG signals utilizing the entropy-based features.

1.5 Thesis Overview

This thesis is divided into four chapters.

Chapter 1 provides general introduction followed by necessary background, literature review and the objectives of the work.

Chapter 2 presents comprehensive analysis of EEG signals in EMD domain and in its variants, for example, EEMD and CEEMDAN domains. Various entropy measures, such as Shannon entropy, log-energy entropy and Renyi entropy, are measured, and due to their ability to characterize the complexity and measure the variability of the EEG signals, they are utilized to classify the EEG signals using different well-known classifiers like support vector machine (SVM) or k-nearest neighbor (KNN). The performance of KNN and SVM classifiers are studied using the well-known criteria such as sensitivity, specificity and accuracy.

The EEG signals are first analyzed in DWT domain in chapter 3. Next, a hybrid method using entropy-based features obtained from EEG signals in the EMD-DWT domain is proposed for classifying focal and non-focal EEG signals. The ability of the previously mentioned features to discriminate focal and non-focal EEG signals is investigated employing one-way analysis of variance (ANOVA) and receiver operating characteristic (ROC) curves.

Chapter 4 contains the concluding remarks along with suggestions for future work on the topic.

Chapter 2

Analysis in EMD Domain and its variants

In this chapter, a comprehensive analysis for the discrimination of the focal and non-focal electroencephalography (EEG) signals is carried out in the empirical mode decomposition (EMD) domain, and its variants, like ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) domains. A number of spectral entropy-based features such as the Shannon entropy, log-energy entropy and Renyi entropy are calculated in EMD, EEMD and CEEMDAN domain of the EEG signals. In lieu of using the signals from the EEG channels, the differences between two adjacent EEG channels are used due to its robustness to noise and interference. The ability of the entropy-based features in separating the focal and non-focal EEG signals is explored utilizing the one-way ANOVA analysis and the box-whisker plots. The results reveal that among the entropy measures computed in the EEMD and CEEMDAN domains, the quadratic Renyi entropy and log-energy entropy measures are most promising in discriminating the focal and non-focal EEG signals. Finally, k-nearest neighbor (KNN) and support vector machine (SVM) are employed to classify the focal and non-focal EEG signals utilizing the extracted features.

2.1 EEG Dataset

The EEG signals are obtained from a publicly available EEG database [35], [4]. For the convenience of the readers a brief description of the database is provided in this Section. The reasons behind using this database are: (i) availability in public domain, (ii) a considerable volume, which is necessary to provide the statistical significance of the discriminating features, and (iii) use in the literature of EEG (See references [5], [8], [10]).

The database consists of two types EEG signals- Focal and Non-focal EEG. Each class contains 3750 pairs of simultaneously recorded signals, denoted as x and y. Each of these 3750 EEG records contains data of 20 seconds duration, with 10240 samples, since the sampling frequency is 512 Hz. For focal recordings, signal x is collected randomly from any of the five patients from all those channels that detects the first ictal EEG signals and signal y corresponds to the one of the neighboring focal channels. All other channels included in the recordings are classified as non-focal EEG channels where the recordings of two adjacent channels are also named as signals x and y. Before inclusion in the database, EEG signals containing prominent measurement artifacts, are discarded.

Moreover, recordings of seizure activity and three hours after the last seizure are excluded. In this thesis, first, 50 sets of signals for focal and non-focal EEG, named Data_F_50 and Data_N_50 respectively, are utilized for the investigation of the appropriate features. All the sets (7500 sets in total) are used for the classification.

Figure 2.1 shows 20 seconds sample EEG from focal and non-focal EEG signals. First column shows signal x, signal y and signal x-y for the focal signals and the second shows the same respectively for the non-focal EEG signals. The motivation to analyze the signal x-y, the difference between two adjacent channels, comes from its robustness to noise and interference reported in various algorithms related to seizure detection [21].

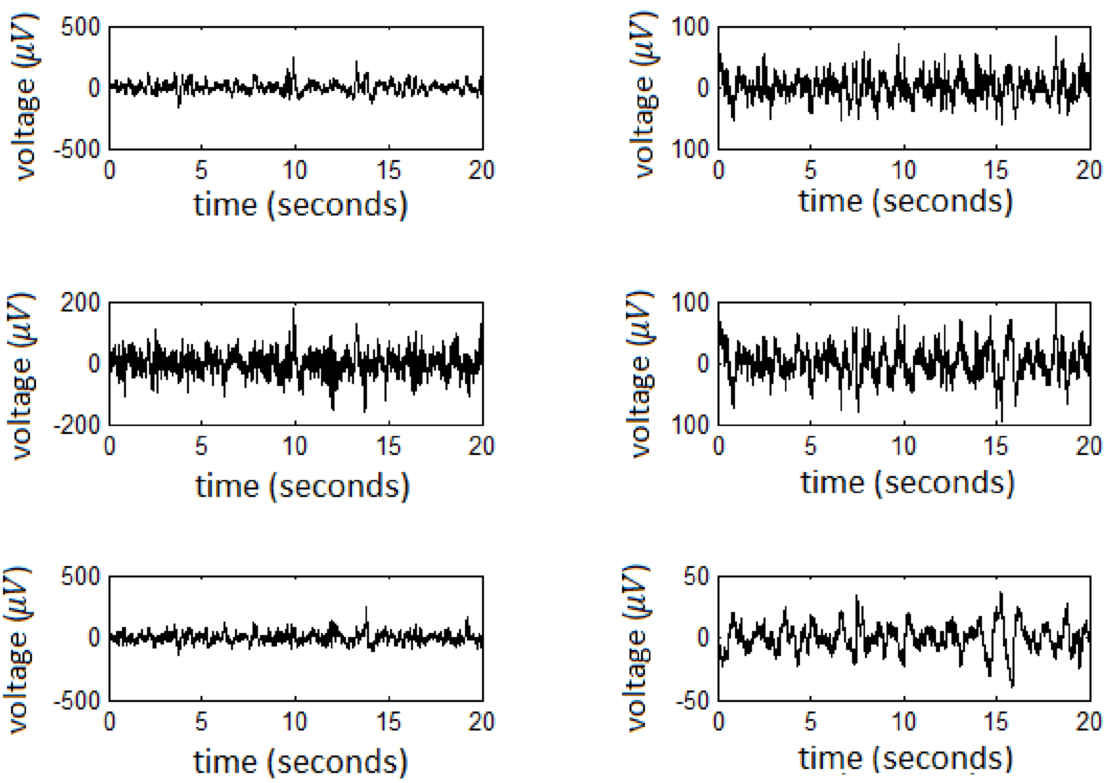

Fig. 2.1 Raw focal(Left) and non-focal(Right) EEG signals with duration of consecutive 20 seconds for signals x (first row), y (second row) and x-y (third row).

The frequency range, that is interesting for the EEG signal analysis, spans over 0 to 60 Hz. The frequencies beyond 60 Hz may come from noise [36]. On the other hand, as the sampling frequency is 512 Hz, the highest frequency component of an EEG segment of the database is 256 Hz. The frequencies beyond 60 Hz are removed by using a 6th order Butterworth filter.

2.2 Classification Methods

In the EEG literature, there are, in general, two steps of any type of classification algorithms. First of all, the appropriate features are extracted from the EEG signals, and then, those features are utilized in suitable classifiers to classify the signals in clinically relevant cases.

2.2.1 Feature Extraction

Entropy is a quantity that measures the randomness in a signal and characterizes the disorder of a chaotic system. The non-linear behavior of entropy to measure the complexity of a signal maximizes the capability to describe and differentiate the EEG signals which are also non-stationary by nature.

Among various types of entropy, spectral entropy parameters (e.g. Shannon entropy, log-energy entropy and Renyi entropy) have been utilized in the literature to measure the spectral complexity of a time series data [7], [16], [37]-[42]. For this reason, the entropy features are expected to have a significant capability to discriminate EEG signals into clinically relevant cases. Shannon entropy quantifies the potential reduction of uncertainty if the outcome of the probabilistic process is known. For the non-linear signals like EEG signals, Shannon entropy measures the average information contained in the probability distribution function (pdf). On the other hand, Renyi entropy can be used to derive the continuous family of mutual information measures and can be applied not only in these kinds of statistical signals but also in economics, ecology or quantum information too. Besides, the capability of log-energy entropy to characterize the non-linear dynamics of EEG signals helps to describe the electrophysiological behavior of epileptogenic regions successfully. These spectral parameters measure the power spectral density (PSD) which represent the power distribution according to the frequencies.

If the power level of the i^{th} frequency component is p_i, for a given data-set, with length N, the corresponding Shannon-entropy (SE), log-energy entropy (LE) and Renyi entropy (RE) are expressed as

$$SE = -\sum_{i=1}^{N} {p_i}^2 \times log({p_i}^2) \tag{2.1}$$

$$LE = \sum_{i=1}^{N} log({p_i}^2) \tag{2.2}$$

$$RE = \frac{1}{1-\alpha} log \ \{ \sum_{i=1}^{N} (p_i)^{\alpha} \} \tag{2.3}$$

The renyi entropy (RE) is of order α, where $\alpha > 0$ and $\alpha \neq 1$. If α is equal to 2, then the measurement equally emphasizes the sub-gaussian and the super-gaussian components [42]. This feature has been successfully utilized in the EEG literature for classifying the EEG data [5], [7], [42]. However, if $\alpha = 2$, the measurement is called as quadratic Renyi entropy, that is given by-

$$RE = -log \ \{ \sum_{i=1}^{N} (p_i)^2 \} \tag{2.4}$$

2.2.2 Classifiers

In this thesis, the entropy-based features are used to classify the EEG signals employing two different well-known classifiers, namely support vector machine (SVM) and k-nearest neighbor (KNN).

Support Vector Machine (SVM), proposed by Vapnik [43], is a widely used classifier to separate various types of EEG signals [5], [13]-[15]. It is a non-linear binary classifier which maps the input features onto the higher dimensional hyper-plane. The best hyper-plane is used to separate all data points of one class from other to classify the signals. The SVM has the added advantage of reduced over-fitting. To train the classifier, a proper kernel function for a certain problem is dependent on the specific data. In this thesis, the performance of the SVM classifiers is measured using both the radial basis function and polynomial kernel function, as well as employing the least square method for training.

K-nearest neighbor (KNN) is a supervised non-parametric classifier that does not need a prior assumption about the statistics of the training samples. Any new testing sample is classified by measuring the distance from the nearest training case where K number of training points closest to the test sample are calculated and the most common class among these k-nearest neighbors is selected as the class. The number k decides how many neighbors influence the classification, for example, when k = 1, it becomes simply the nearest neighbor algorithm. The KNN classifier has mainly two parameters to tune; those are the distance parameter and K. In this thesis, the performance is observed by varying K and varying the distance parameter as Euclidean, cosine, correlation and city-block. In various cases of various distance parameters, different values of K provide the optimum accuracy. Moreover, in the literature, this classifier has been shown to be quite promising to classify various types of EEG signals [44] - [46].

For both classifiers, SVM and KNN, randomly selected 20% segments are used for training, and the other 80% segments have been used for testing. Due to the non-stationary nature of EEG signals, each twenty second sample data has been segmented into ten non-overlapping parts. Thus, each segment contains two seconds data and corresponds to 1024 samples. The performance criteria of the proposed classification method in classifying different types of EEG data are-

$$Sensitivity = \frac{TP}{TP+FN} \times 100\ \% \tag{2.5}$$

$$Specificity = \frac{TN}{FP+TN} \times 100\ \% \tag{2.6}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\ \% \tag{2.7}$$

where TP, TN, FP and FN stand for true positive, true negative, false positive and false negative events, respectively [15].

2.3 Analysis of Focal and Non-focal EEG Band-limited Signals

In this thesis, the entropy-based features are first explored for the bandlimited EEG segments. For 50 EEG records of 20s duration each, the values of the entropies are calculated for each of the signals x, y and x-y. To investigate whether these values can discriminate the focal EEG from the non-focal ones or not, one-way ANOVA is carried out. The small p-values obtained from ANOVA indicate that differences between column means are significant. Box plots are also used to see if the entropy values can differentiate focal and non-focal EEG signals. From the box plots in Figure 2.2, it is obvious that the values of focal and non-focal EEGs are mostly overlapping for these band limited signals. Besides, the p-values obtained from ANOVA, shown in Table 2.1, are, in general, not small enough to establish the evidence of capability of the features to discriminate the signals. Furthermore, in the ROC curves in fig. 2.3, in general, none of these three entropy parameters indicates a significant level of discrimination capability. Thus, the overall scenario underscores the necessity to decompose the EEG signals in sub-band levels to investigate in the decomposed versions whether they may be separable in sub-band levels.

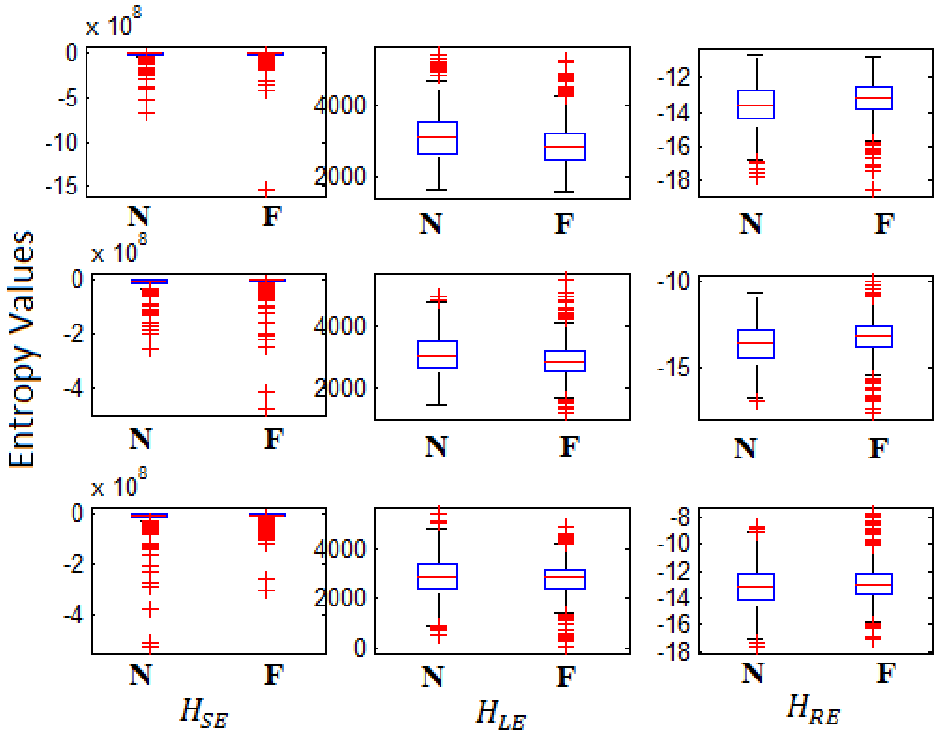

Fig. 2.2 Box Plots for Band-limited EEG signals corresponding to Shannon entropy (first column), Log-energy entropy (second column) and Quadratic Renyi Entropy (third column) for signals x (first row), y (second row) and x-y (third row).

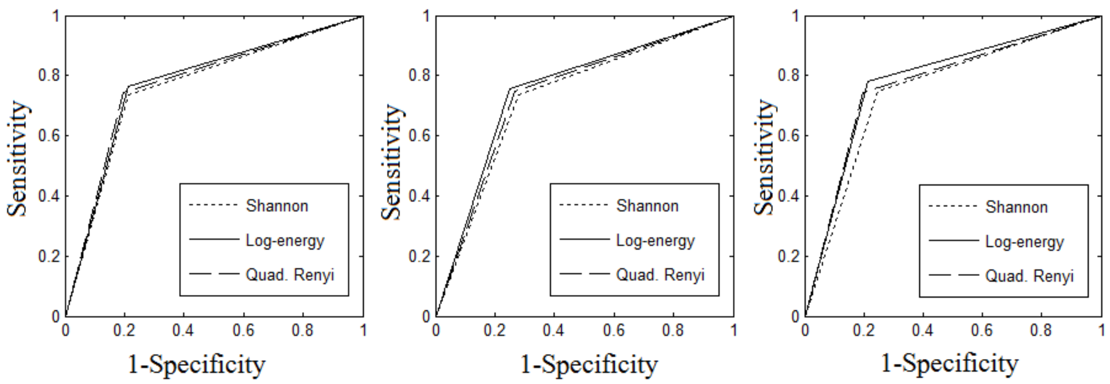

Fig. 2.3 ROC curve for entropy values for band-limited signals x, y and x-y, respectively

Table 2.1 p-values for various features for the band-limited signals

Signals	Shannon Entropy	Log-energy Entropy	Renyi Entropy
Signal x	0.1015	2.4546e-04	0.0446
Signal y	0.0721	5.1087e-03	5.8915e-03
Signal x-y	0.0165	0.0220	0.0160

2.4 Analysis in EMD Domain and in Its Variants

The Empirical Mode Decomposition(EMD) provides amplitude and frequency modulated oscillatory patterns known as intrinsic mode functions (IMF), which are derived using the basis obtained from the signal subjected to EMD. As discussed before, a major problem with EMD is mode-mixing; recently, variants of EMD such as EEMD and CEEMDAN have been proposed to eradicate this problem. In the following, a brief review of these variants is provided.

2.4.1 Empirical Mode Decomposition

The empirical mode decomposition (EMD) provides amplitude and frequency modulated oscillatory patterns, known as intrinsic mode functions (IMF), of a nonlinear and non-stationary signal [7], [13]. For an N -point data X $(x_1, x_2, x_3, \ldots \ldots \ldots, x_N)$, IMFs are obtained as-

- First step is to set the input as m_0, such that $m_0 = X$, and then, $m_{0ld} = m_0$
- After the identification of the local maxima and minima of m_{0ld}, cubic spline interpolation is used to obtain the envelopes of those.
- After that, the mean values(m) of e_{min} and e_{max} are calculated using and subsequently the value is subtracted from m_{old} as-

$$m_{new} = m_{old} - \frac{e_{max}+e_{min}}{2} \tag{2.8}$$

- Finally, m_{old} is set equal to m_{new}
- Stop the decomposition if

$$\frac{\sum|m_{new}-m_{old}|^2}{\sum m_{old}{}^2} < \alpha \tag{2.9}$$

where $0.2 \leq \alpha \leq 0.3$. Otherwise, the second and third step will be repeated to get the IMFs [16].

Figure 2.4 shows the band-limited signals and first five IMFs for a sample focal (left column) and a sample non-focal (right column) EEG signals comprising 4s recordings of signal x for each.

Tables 2.3 and 2.4, respectively, show the p-values obtained from the various IMFs of signals x, y and x-y. It is seen that log-energy entropy provides, in general, smaller p-values than the quadratic Renyi entropy or Shannon entropy. It indicates the superior capability of the log-energy entropy to separate the focal EEG data from the non-focal ones than the other entropy parameters for various IMFs. It should be noted that the p-values for log-energy entropy parameters for IMFs 1 and 2, are significantly smaller than others for signals x, y and x-y.

Table 2.2 p-values for various features for different IMFs in EMD Domain

Levels	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	x	y	x	y	x	y
IMF 1	0.203	0.183	3.0e-15	2.7e-15	0.011	0.001
IMF 2	0.512	0.162	9.8e-19	3.6e-20	3.1e-12	5.5e-13
IMF 3	0.192	0.910	9.5e-17	7.5e-07	0.1963	0.041
IMF 4	0.269	0.483	2.4e-07	1.2e-09	5.8e-11	1.6e-13
IMF 5	0.312	0.245	3.8e-09	4.3e-08	0.0038	0.0062

Besides, the box plots in fig 2.5 shows a better level of discrimination for the log-energy entropy values (second column) than Shannon entropy (first column) and Renyi Entropy (third column) for IMF 1. Furthermore, in fig. 2.6, the ROC curves are shown when the three entropy parameters obtained from IMF 1 are utilized. It is seen that, in general, the area under the ROC curves for the IMFs has been increased from the case of band-limited signals in fig. 2.5 which indicates significant discrimination level in the EMD domain than in the band-limited signals. Finally, in table 2.5, classification performance is shown for the features in EMD domain.

Table 2.3 p-values for various features for the first five IMFs for signal x-y

Entropy	Shannon	Log-energy	Renyi
IMF 1	0.1035	1.13e-13	0.0035
IMF 2	0.7470	3.94e-14	1.30e-05
IMF 3	0.8246	4.45e-06	0.0601
IMF 4	0.1542	0.0056	5.01e-04
IMF 5	0.2316	0.0073	0.0751

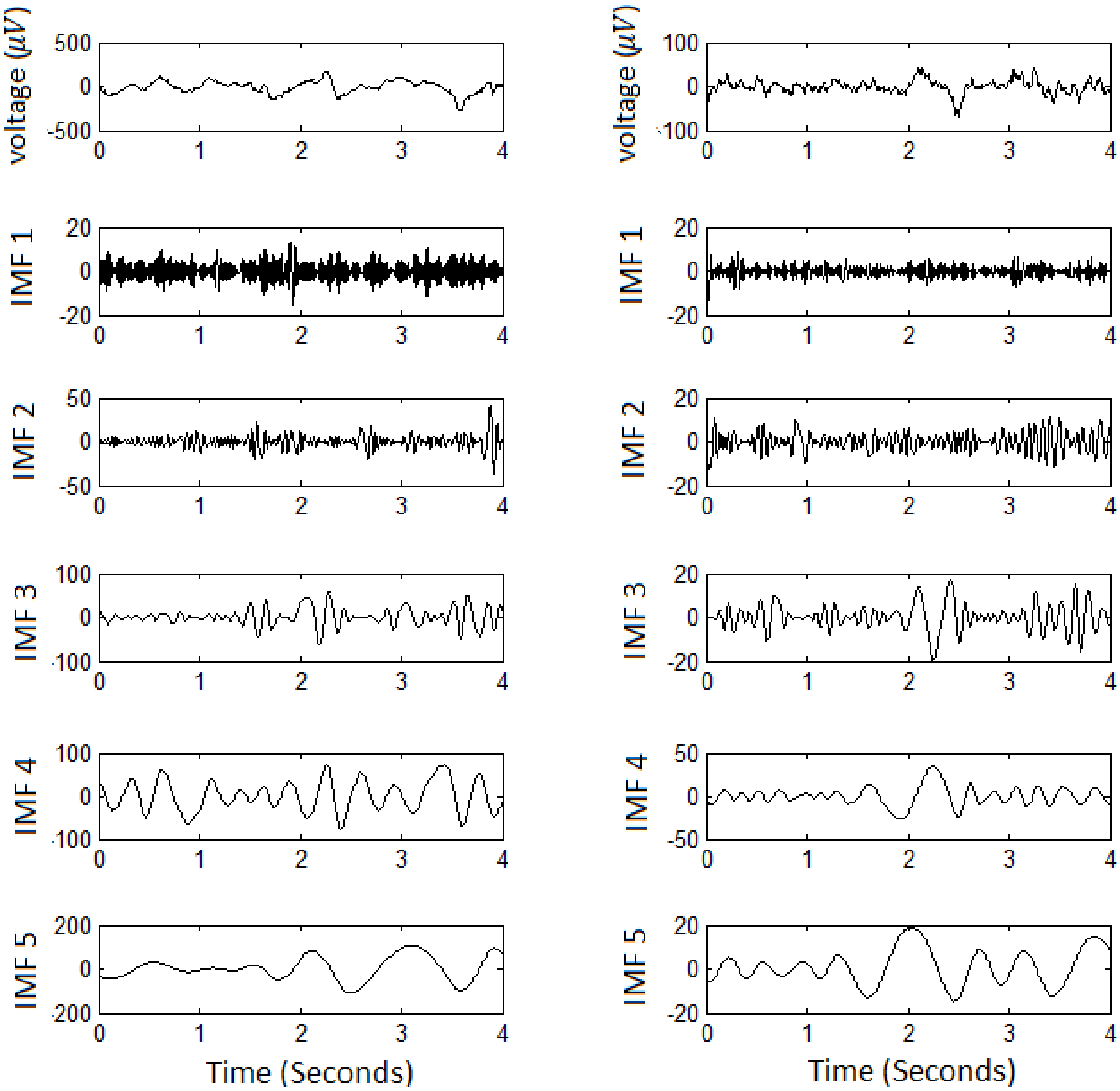

Fig. 2.4 First five IMFs extracted from EMD for focal(left) and non-focal(right) EEG signals for signal x.

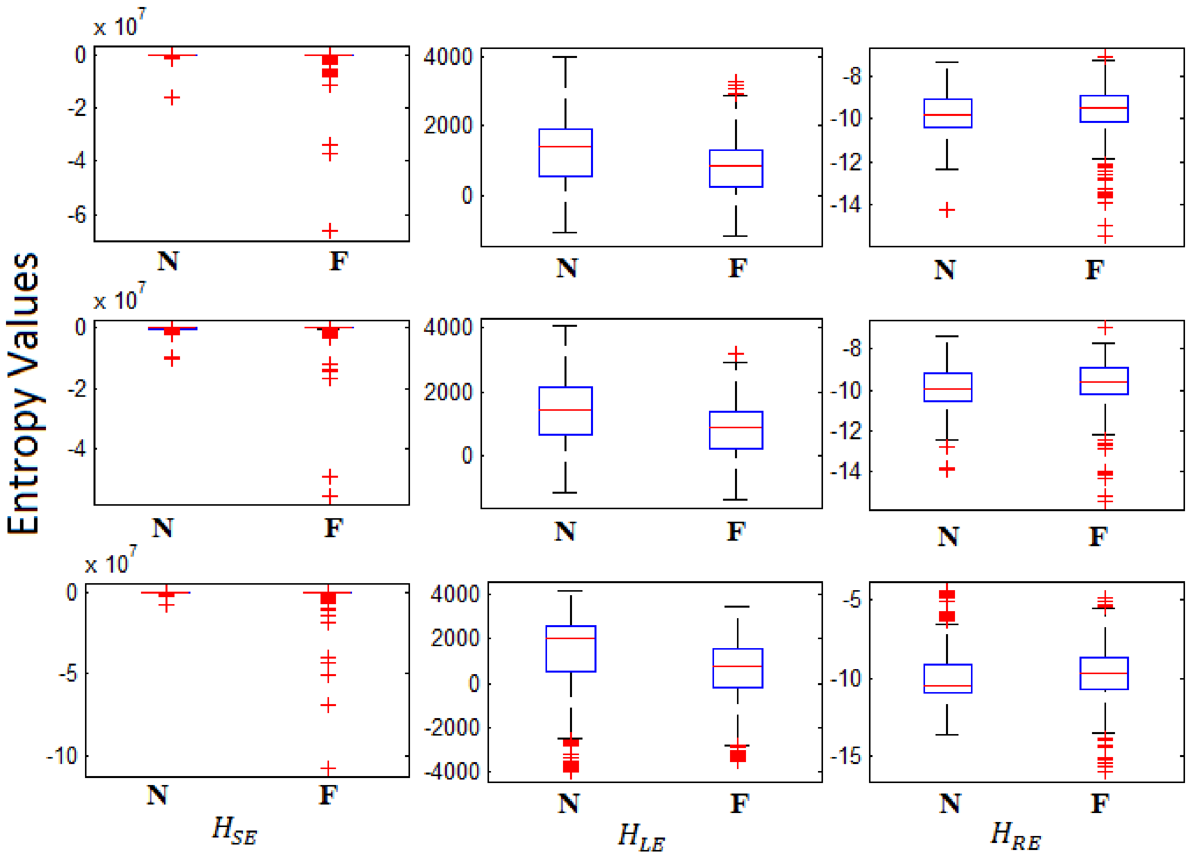

Fig. 2.5 Box plots for IMF 1 for signals x, y and x-y in first, second and third row respectively for Shannon entropy, log-energy entropy and quadratic Renyi entropy in first, second and third column respectively

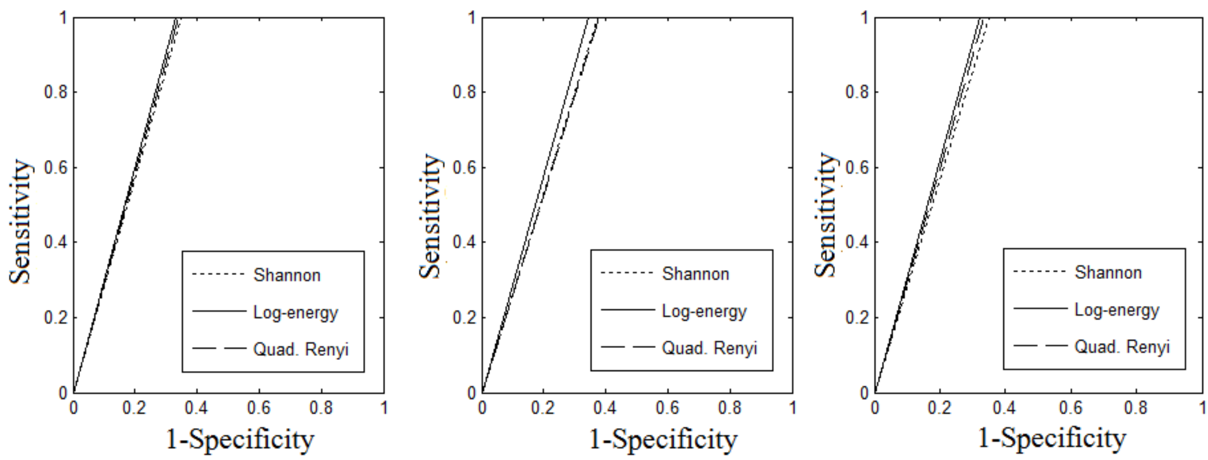

Fig. 2.6 ROC curve for the entropy values obtained from the rst IMF of the EEG signals x, y and x- y respectively.

Table 2.4 Classification Accuracy in EMD Domain

IMFs	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	KNN	SVM	KNN	SVM	KNN	SVM
1	76.2 %	75.1 %	79.1 %	78.3 %	77.4 %	78.1 %
2	73.8%	72.7 %	76.5%	75.7 %	75.8%	74 %
3	71.1 %	70 %	74.3 %	72.6 %	74.4 %	72.6 %
4	69.2 %	68.3 %	74.4 %	73.3 %	73.9 %	71.1 %
5	69.8 %	69.6 %	72.4 %	72.9 %	70.2 %	69.8 %
All	81 %	79.8 %	85.7 %	84.6 %	82.3 %	81.1 %

2.4.2 Ensemble Empirical Mode Decomposition (EEMD)

The EEMD [47] provides the EMD of ensemble average of Gaussian noise-assisted data. A low-level Gaussian noise is mixed with the original signal which is then decomposed by EMD [10]. The steps of EEMD are as follows:

- First, from the actual signal X, X_i is calculated as $X_i = X + w_i$ where w_i indicates the realizations of white Gaussian noise (while i = 1,2,3,…,R) with zero mean and unit variance, and R is the number of realizations of the noise.
- Next, each of the signals of X_i is subjected to EMD, thus giving the IMFs, denoted as IMF_i^k, where k = 1, 2, 3, …, L represents the different modes.
- Finally, the mode functions in EEMD domain, IMF^k is calculated as-

$$IMF^k = \frac{1}{L}\sum_{i=1}^{R} IMF_i^k \tag{2.10}$$

In Fig. 2.7, the first three mode functions of the focal and non-focal EEG signals in EEMD domain have been shown for the signal x-y in both cases. Table 2.5 shows the p-values obtained from the various IMFs of signal x-y, and table 2.6 shows the classification performance.

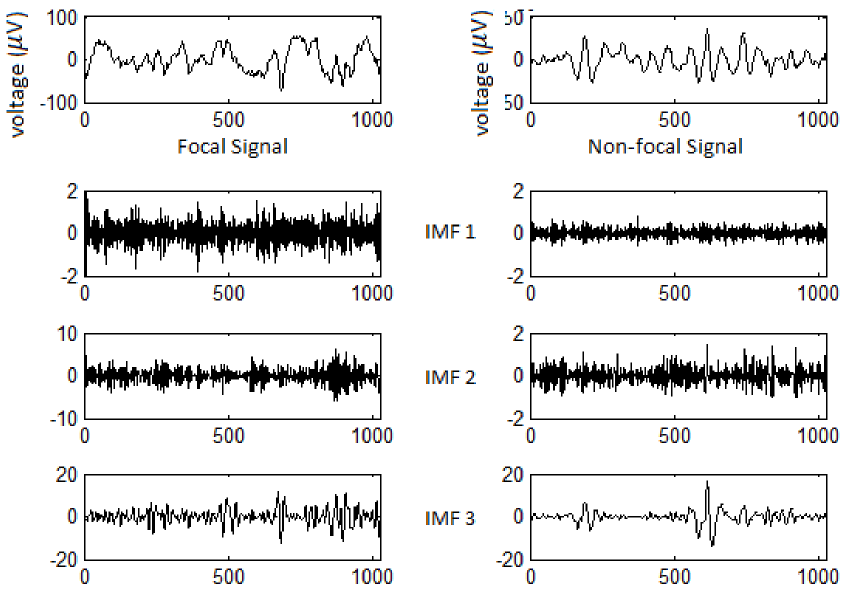

Fig. 2.7 First three IMFs extracted from EEMD for focal (left) and non-focal (right) EEG signals for signal x-y

Table 2.5 p-values for various features for the EEMD Domain for signal x-y

IMFs	Shannon Entropy	Log-energy Entropy	Renyi Entropy
1	0.1013	0.0742	2.12e-06
2	0.0044	4.84e-04	3.26e-08
3	0.0457	0.0085	1.53e-07
4	0.0776	0.0035	3.75e-06
5	0.0136	0.2831	2.71e-04

Table 2.6 Classification Accuracy in EEMD Domain

IMFs	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	KNN	SVM	KNN	SVM	KNN	SVM
1	76.8 %	74.3 %	77.1 %	77.3 %	79.4 %	79.1 %
2	73.2%	73.7 %	78.5 %	75.8 %	77.9 %	78.9 %
3	72.3 %	71.4 %	75.6 %	74.6 %	76.2 %	77.2 %
4	70.9 %	69.1 %	72.7 %	71.1 %	77.1 %	76.1 %
5	69.2 %	69.8 %	73.4 %	72.7 %	75.2 %	74.8 %
All	82.8 %	81.7 %	85 %	83.4 %	81.9 %	81.2 %

2.4.3 Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)

Although EEMD successfully minimizes the mode-mixing problem, due to the use of different realizations, it can give different number of modes and residual noise. The CEEMDAN [49] gives an exact reconstruction of the data while computationally less expensive as compared to EEMD [13]. In CEEMDAN, an operator M_j is defined that produces the j^{th} mode of EMD. Unlike EEMD, the added white Gaussian noise has a certain standard deviation denoted as σ_0. Thus, the steps of the CEEMDAN algorithm is given as below-

- First, from the actual signal X, X_i is calculated as $X_i = X + \sigma_0 w_i$ and transformed in EMD to obtain the first mode functions.
- Next, the first mode IMF, named as IMF^1 and the first residue r_1 in CEEMDAN are calculated as-

$$IMF^1 = \frac{1}{L}\sum_{i=1}^{R} IMF_i^1 \tag{2.11}$$

$$r_1 = x - IMF^1 \tag{2.12}$$

- If σ_1 is the standard deviation of the white Gaussian noise at this stage, then IMF^2 is determined as

$$IMF^2 = \frac{1}{L}\sum_{i=1}^{R} M_1[r_1 + \sigma_1 M_1(w_i)] \tag{2.13}$$

- Subsequently, the k^{th} residue and the $(k+1)^{th}$ mode functions are obtained as

$$r_k = r_{k-1} - IMF^k \tag{2.14}$$

$$IMF^{k+1} = \frac{1}{L}\sum_{i=1}^{R} M_1[r_k + \sigma_k M_k(w_i)] \tag{2.15}$$

- The previous step is repeated until the residue becomes a monotonic function such that further extraction of an IMF is impossible. If k is the total number of modes and r_k is the final residue, the input X can be reconstructed from all the IMFs as

$$X = \sum_{k=1}^{L} IMF^k + r_k \tag{2.16}$$

In Fig. 2.8, the first three mode functions of the focal and non-focal EEG signals in CEEMDAN domain have been shown for the signal x-y in both cases. Table 2.8 shows the p-values obtained from the various IMFs of signal x-y, and table 2.9 shows the classification performance. Fig. 2.9 shows the box-plots for EEMD and CEEMDAN domains.

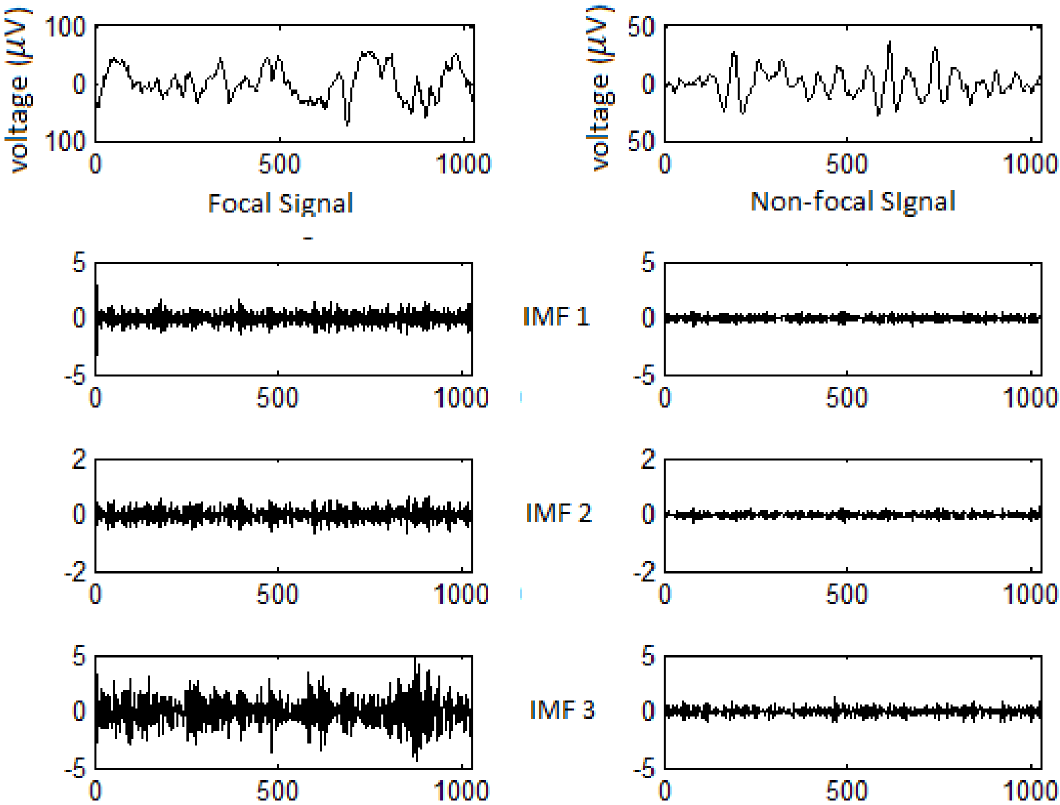

Fig. 2.8 First three IMFs extracted from CEEMDAN for focal(left) and non-focal(right) EEG signals for signal x-y.

Table 2.7 p-values for various features for the CEEMDAN Domain for signal x-y

IMFs	Shannon Entropy	Log-energy Entropy	Renyi Entropy
1	0.1084	0.0959	2.21e-06
2	0.0393	0.2990	8.81e-07
3	0.0306	0.3668	4.35e-04
4	0.0072	0.1546	8.43e-06
5	0.0321	1.02e-08	1.04e-10

Table 2.8 Classification Accuracy in CEEMDAN Domain

IMFs	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	KNN	SVM	KNN	SVM	KNN	SVM
1	76.5 %	76.3 %	77.9 %	78.3 %	78.9 %	79.3 %
2	74.2 %	75.7 %	79.2 %	78.8 %	77.5 %	78.3 %
3	72.8 %	73.9 %	78.1 %	76.6 %	76.8 %	77.8 %
4	71.2 %	72.5 %	74.8 %	72.1 %	73.2 %	75.4 %
5	71.9 %	71.8 %	73.9 %	73.7 %	71.2 %	72.5 %
All	84.1 %	82.9 %	85.4 %	84.7 %	82.3 %	81 %

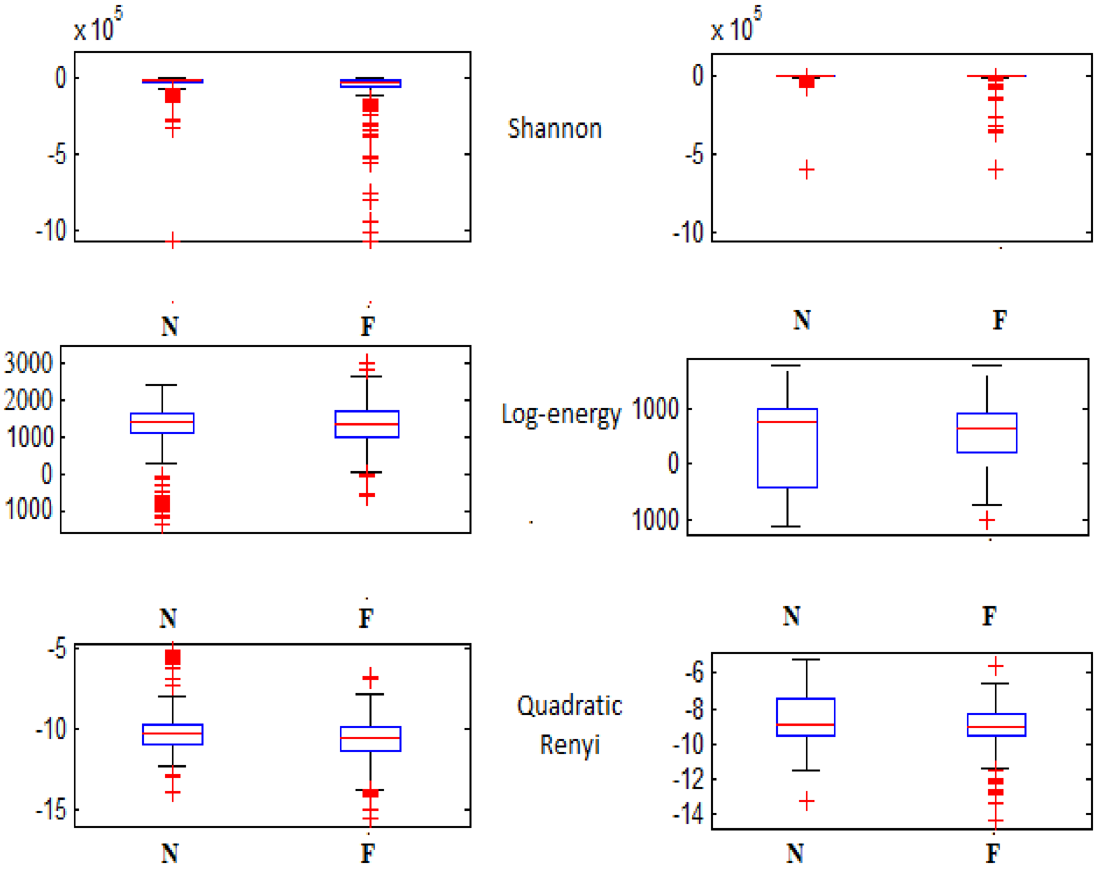

Fig. 2.9 Sample box plots for IMF 3 for EEG signal x-y in for Shannon entropy, log-energy entropy and quadratic Renyi entropy in EEMD domain(left) and CEEMDAN domain(right)

2.5 Discussions

The following observations can be made based on the analysis in section 2.4:

- The p-values obtained from EMD domain and its variants, like in EEMD and CEEMDAN domains, are small enough to indicate the discrimination capability of the entropy-based features.
- The p-values for quadratic Renyi entropy and log-energy entropy are significantly smaller than the those for Shannon entropy which indicates their superior capability to distinguish between the focal and non-focal EEG signals.
- The box plots further confirm that Renyi entropy and log-energy entropy perform a better discrimination than Shannon entropy.
- Moreover, the box plots also confirm that these features can separate EEG signals more for signal x-y than for signal x or signal y.
- Between the two classifiers, KNN and SVM, in almost all the cases, KNN provided better performance than SVM.
- From the classification performance in tables 2.5, 2.7 and 2.9, the maximum accuracy has been achieved up to 85.7 % from log-energy entropy in EMD domain for KNN classifier.

Chapter 3

Analysis in EMD-DWT Domain

In this chapter, EEG signals are first analyzed in discrete wavelet transform (DWT) domain. Next, being motivated by the success of discriminating various types of non-stationary signals, like ECG, where the frequency bands are extracted from the IMFs [14], [33]-[34], EEG signals are next analyzed in EMD-DWT domain. The hybrid method using entropy-based features obtained from EEG signals in the EMD-DWT domain is proposed for classifying focal and non-focal EEG signals. Entropy-based features such as Shannon entropy, log-energy entropy and Renyi entropy are calculated in the EMD, DWT and EMD-DWT domains. In this chapter, the ability of these features to discriminate focal and non-focal EEG signals is investigated employing one-way analysis of variance (ANOVA) and receiver operating characteristic (ROC) curves. It is noted that the best scenario in differentiating focal and non-focal EEG signals is obtained while using the combined EMD-DWT domain and employing log-energy entropy values.

3.1 Analysis of EEG signals in Discrete Wavelet Transform (DWT) Domain

Discrete wavelet transform (DWT) has emerged as one of the most important tools in non-stationary signal analysis because of its capability to extract the time and frequency localization from the time domain signals. Various studies are available in the literature exploiting the superior time-frequency localization of DWT for identifying epilepsy activities [5], [23], [30]. The DWT decomposes EEG signals onto basis functions by expanding, contracting and shifting a single prototype function (the mother wavelet) providing a coarse approximation ($C_{j,k}$) of the signal and detail information ($D_{j,k}$) as given by-

$$C_{j,k} =< f(t), \varphi_{j,k}(t) > = \int_R f(t) \times 2^{-j/2} \overline{\varphi(2^{-j}\, t - k)}\, dt \tag{3.1}$$

$$D_{j,k} =< f(t), \psi_{j,k}(t) > = \int_R f(t) \times 2^{-j/2} \overline{\psi(2^{-j}\, t - k)}\, dt \tag{3.2}$$

where $\psi(t)$ is the mother wavelet, $\phi(t)$ is the basic scaling, j is the scale index, and k is the translation parameter. The inverse discrete wavelet transform is given by-

$$f(t) = \sum_k C_{j,k} \times 2^{-j/2} \varphi(2^{-j} t - k) + \sum_k D_{j,k} \times 2^{-j/2} \psi(2^{-j} t - k) \tag{3.3}$$

Contrary to the band-limited signals, the DWT decomposition captures the dynamics of EEG signals better [12]. Certain features that are not so prominent in band-limited signal can be emphasized in the time frequency domain sub-bands [15], [23]. Hence-fore, it is more appropriate to analyze the EEG signals at DWT sub-band levels. For the purpose of analysis, the band-limited EEG signals are subjected to a 4 level DWT decomposition and the obtained sub-bands are shown in table 3.1. In this thesis, Daubechies-4 (db4) is chosen to be the mother wavelet for performing the DWT decomposition. The smoothing feature of the db4 wavelet has made it to be more appropriate to detect continuous changes of the EEG signals [23]. However, if the mentioned five sub-bands are reconstructed utilizing the inverse DWT, it would approximately correspond to the five physiological EEG sub-bands [25] which is listed in table 3.1. In fig. 3.1, two seconds band-limited (BL) EEG signals from focal and non-focal ones are shown in the first row in the first and second column, respectively.

Table 3.1 DWT Sub-bands

Sub-band Name	Frequency Range	Physiological EEG Sub-bands
a4	0-4 Hz	Delta band
d4	4-8 Hz	Theta band
d3	8-15 Hz	Alpha band
d2	15-30 Hz	Beta band
d1	30-60 Hz	Gamma band

Now, for of the sub-bands, for all the two seconds segments Shannon entropy, log-energy entropy and quadratic Renyi entropy have been estimated, and then, the one-way ANOVA has been conducted. The corresponding p-values for signal x and signal y have been shown in Table 3.2. Furthermore, for the signal x-y, the p-values have been shown in Table 3.3.

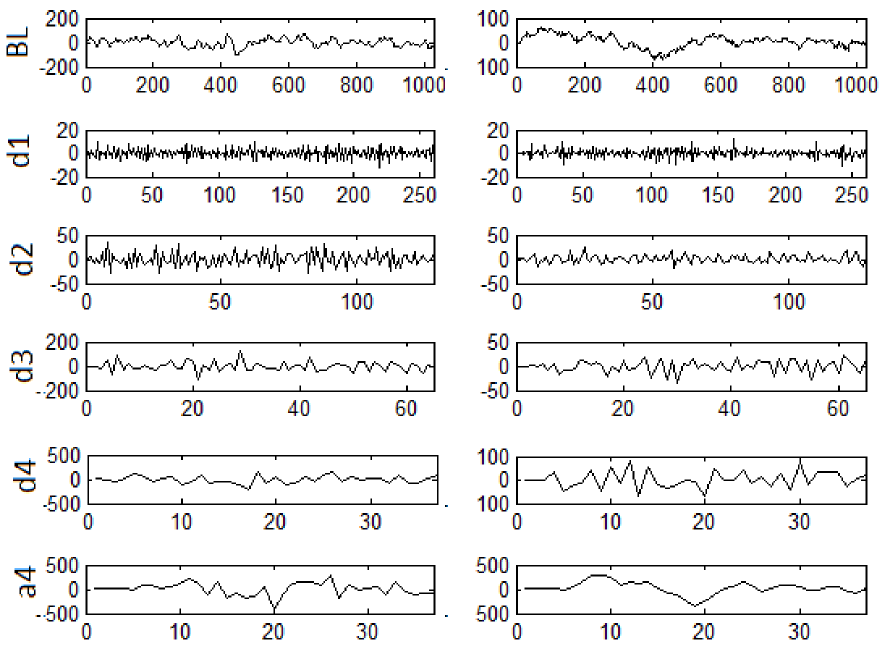

Fig. 3.1 Five DWT sub-bands extracted in DWT domain for focal (left) and non-focal (right) EEG signals for signal x.

Table 3.2 p-values for various features for the DWT sub-bands

Sub-bands	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	x	y	x	y	x	y
d1	0.016	0.017	2.3e-10	2.6e-12	4.1e-12	3.3e-14
d2	0.921	0.596	2.7e-08	2.7e-06	3.5e-04	3.8e-05
d3	0.220	0.805	2.2e-14	8.6e-12	8.4e-10	2.0e-10
d4	0.159	0.149	2.3e-09	7.2e-08	0.0144	0.0313
a4	0.102	0.847	2.1e-05	6.0e-07	0.0373	0.0077

Table 3.3 p-values for various features for the DWT sub-bands of signal x-y

Sub-bands	Shannon	Log-energy	Renyi
d1	0.0036	2.02e-04	2.38e-05
d2	0.6225	7.61e-07	0.0019
d3	0.4200	4.58e-12	6.62e-06
d4	0.2188	4.77e-15	2.40e-04
a4	0.9099	5.58e-19	6.34e-06

From Tables 3.2 and 3.3, it is shown that except the sub-band d1(30-60 Hz), log-energy entropy provides smaller p-values than the quadratic Renyi entropy or Shannon entropy. It indicates the superior capability of the log-energy entropy to separate the focal EEG data from the non-focal ones than the other entropy parameters for those sub-bands. The box plots in fig. 3.2 shows the discrimination for d4 sub-band for Shannon entropy, Log-energy entropy and Renyi entropy in the first, second and third column, respectively, which confirms the better discrimination ability of the log-energy entropy parameters. Furthermore, in fig. 3.3, the ROC curves are shown when the three entropy parameters obtained from the five sub-bands are utilized and the area under the ROC curves is also increased by a significant level than the case of band-limited signals in chapter 2. Finally, table 3.4 shows the classification performance of the entropy-based features in DWT domain for various sub-bands.

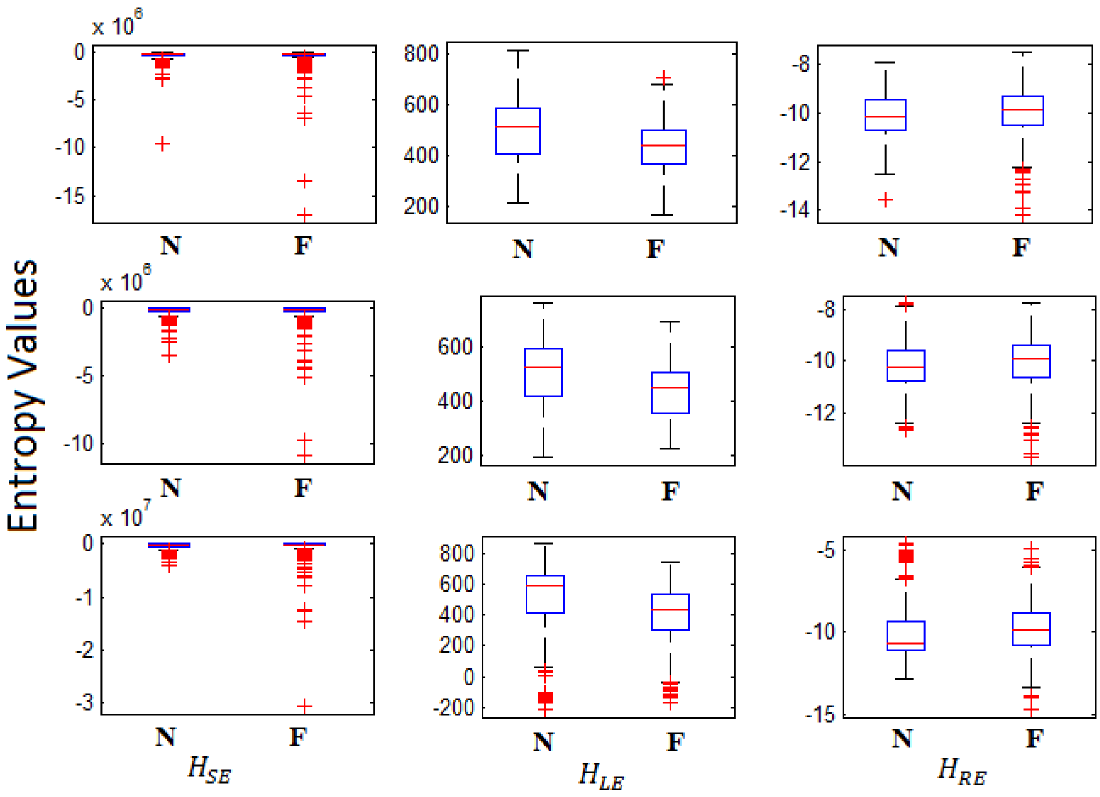

Fig. 3.2 Box plot for d4 sub-band for Shannon, log-energy and Renyi entropy, respectively.

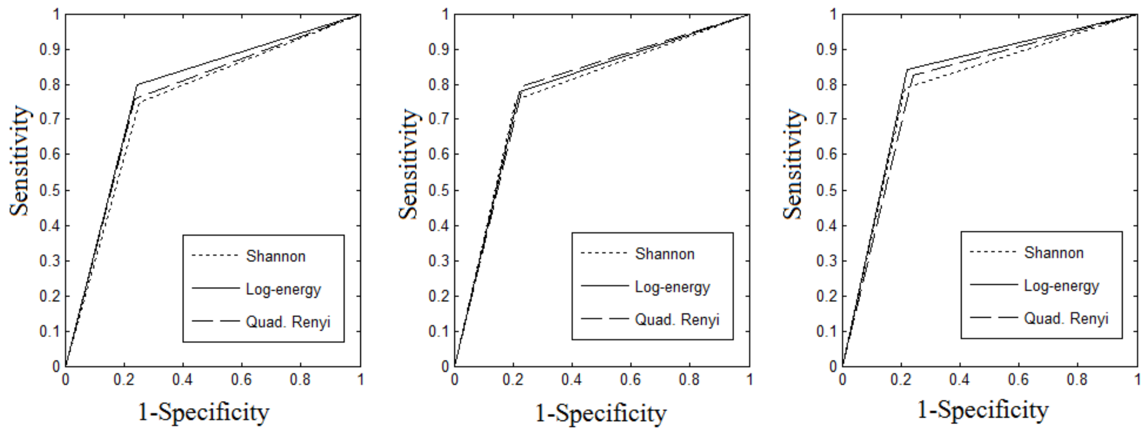

Fig. 3.3 ROC curve for Shannon, log-energy and Renyi entropy when all the sub-bands in DWT domain are utilized.

Table 3.4 Classification Results in DWT Domain for signal x-y

Sub-bands	Shannon Entropy		Log-energy Entropy		Renyi Entropy	
	KNN	SVM	KNN	SVM	KNN	SVM
d1	81.5 %	80.3 %	77.4 %	78.4 %	78.7 %	79.1 %
d2	76.2 %	77.1 %	78.2 %	78.5 %	75.4 %	76.3 %
d3	78.3 %	78.9 %	81.1 %	80.5 %	78.8 %	77.8 %
d4	79.2 %	80 %	81.8 %	81 %	76.2 %	75.9 %
a4	73.9 %	72.8 %	82.3 %	81.5 %	77.9 %	76.5 %
All	84.6 %	83.8 %.	87.2 %	86.1 %	82.8 %	83.6 %

3.2 Analysis of EEG signals in EMD-DWT domain

From the analysis in section 3.1 and chapter 2, it is quite obvious that the log-energy entropy parameter can discriminate the EEG data quite well in comparison to the other features in both of the EMD and DWT domains. In this regard, it would be interesting to investigate the capability of the features to classify the EEG data in the combined EMD-DWT domain. The motivation comes from the success of discriminating various types of non-stationary signals, like ECG, where the frequency bands are extracted from the IMFs [14], [33]-[34]. For that reason, in this thesis, first the EEG signals are decomposed into the EMD domain, and then, each of the IMF is decomposed in the sub-bands through the DWT. Next, the log-energy entropy values are calculated for each of the sub-bands for all the IMFs and the

ANOVA is conducted. The corresponding p-values are shown in Tables 3.5, 3.6 and 3.7. Furthermore, the sample box-plots for all the sub-bands for IMF 1 are shown in Figure 3.4.

Table 3.5 p-values for signal x in EMD-DWT domain

Sub-bands	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
d1	1.1e-06	5.0e-05	0.0440	1.9e-11	0.2389
d2	3.3e-07	9.6e-07	6.4e-08	0.1904	0.7866
d3	2.3e-09	1.3e-06	1.3e-09	0.4321	0.4238
d4	5.6e-13	1.4e-10	2.0e-04	0.5194	0.0761
a4	1.2e-16	1.9e-18	0.9857	0.1497	0.4573

It is seen from the tables that the p-values for Log-energy entropy parameters are quite small for signals x, y and x y. It should be noted that the p-values for IMFs 1 and 2 are, in general, significantly smaller than those for the IMFs 3,4 or 5. It should be noted that, on average, the values for signal x-y are smaller than those for the others.

Table 3.6 p-values for signal y in EMD-DWT domain

Sub-bands	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
d1	4.3e-08	8.9e-13	2.6e-06	3.3e-11	0.1292
d2	3.3e-07	2.8e-12	1.7e-09	4.9e-05	0.1683
d3	1.0e-09	1.2e-12	1.9e-05	0.0016	0.8419
d4	3.6e-13	1.7e-12	0.4033	0.0017	0.2196
a4	1.6e-16	1.1e-11	0.0078	4.5e-05	0.4695

Table 3.7 p-values for signal x-y in EMD-DWT domain

Sub-bands	IMF 1	IMF 2	IMF 3	IMF 4	IMF 5
d1	3.9e-20	1.7e-25	0.008	2.2e-05	0.118
d2	5.0e-19	4.4e-27	6.9e-08	0.829	0.875
d3	1.7e-24	3.7e-29	3.1e-09	0.271	0.454
d4	4.1e-24	1.1e-19	2.9e-06	0.163	0.199
a4	1.7e-19	1.8e-13	0.006	0.809	0.785

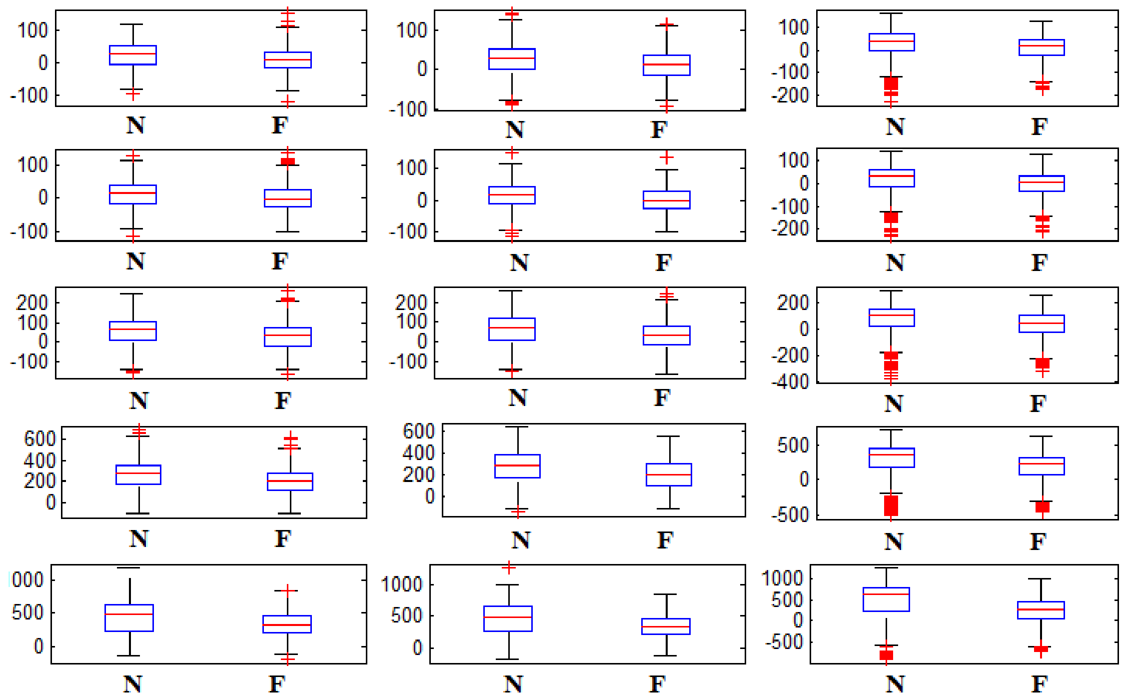

Fig. 3.4 Box-plots for signals x, y and x-y for log-energy entropy for all the DWT sub-bands for IMF 1.

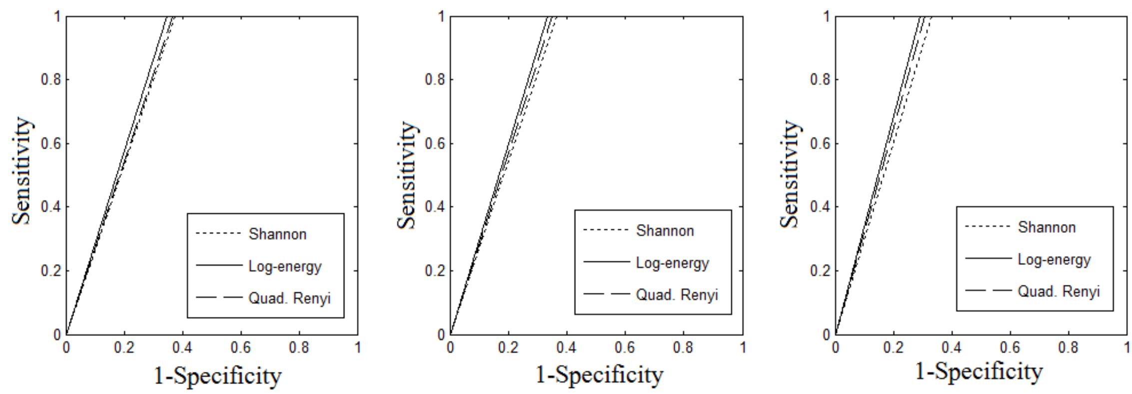

Fig. 3.5 ROC curves for the entropy values obtained from the sub-bands of IMF 1 of the EEG signals x, y and x y respectively

Furthermore, the ROC curves are also shown for all the three entropy features for the sub-bands for IMF 1 which is shown in fig 3.5. It is clearly seen that the log-energy entropy provides a larger area under the curves in comparison to Shannon and Renyi entropy. Besides, it also provides larger area under the curves than the curves for EMD or DWT domain which indicates a better discrimination capability of the log-energy entropy in the EMD-DWT domain than in EMD or in DWT domain.

In summary, after all these analysis, it is clear that the log-energy entropy has a significant better discrimination capability than Shannon entropy or Renyi entropy. Apart from the analysis with the bandlimited signals, it is shown to be quite promising to decompose the EEG signal into EMD and DWT domain providing some IMFs or sub-bands. It is shown that the discrimination is more obvious for the difference of the signals (signal x-y) when the IMFs extracted from EMD domain are decomposed in various sub-bands level through the DWT decomposition. The p-values are smaller for the sub-bands extracted from IMFs 1 and 2 and the separation in the box-plots is also significant. It should be noted that in [27], IMFs 1 and 2 have been considered to be noise, but in this thesis, the noise was already removed using the butter-worth filter. Besides, in [28], IMFs 1, 2 and 3 have been utilized for EEG signal classification while in [29], feature from IMF 2 provide up to 100% accuracy for seizure detection.

Thus, the previous success of IMFs 1 and 2 to classify EEG signals also works as the motivation along with the previous analysis of this thesis. In the next portion of this thesis, the log-energy entropy values for signal x-y in the EMD-DWT domain, would be utilized as the discriminating features for various types of classifiers.

3.3 Results and Discussions

In Table 3.8, the classification performance is shown for SVM classifiers where the parameters need to be tuned for RBF kernel (sigma) and polynomial kernel (order) to obtain a good accuracy. Observing the performance from the RBF Kernel, polynomial Kernel and Least Square Method based SVM classifiers, it is obvious that the accuracy obtained using the signal x-y provides, on average, better performance in comparison to those obtained from signal x or signal y alone. However, in all the cases, RBF-Kernel provides better performance than the others. Next, in Table 3.9, in KNN classifiers, the signal x-y discriminates the focal and non-focal EEG signals better than the signal x or y. Among various types of KNN classifiers, the Cityblock distance KNN provides the highest accuracy 89.4% with 90.7% sensitivity and 88.1% specificity. In Table 3.10 and 3.11, the results of the proposed method are compared with the several recent algorithms available in EEG literature.

From Table 3.10, it is seen that the proposed method provides 91% accuracy which indicates providing the best performance in comparison to the recent algorithms described in [5], [7], [8] and [10]. It is also seen that the log-energy entropy-based features obtained

Table 3.8 Classification Results for SVM classifiers

Signals	SVM Methods	Performance		
		Sensitivity	Specificity	Accuracy
x	Kernel Function RBF	82.6 ($\pm$2.3)	80.2 ($\pm$3.5)	81.4 ($\pm$2.8)
	Kernel Function Polynomial	81.7 ($\pm$2.0)	77.6 ($\pm$3.3)	79.6 ($\pm$2.7)
	Least Square	81.9 ($\pm$3.1)	78.5 ($\pm$3.3)	80.2 ($\pm$3.2)
y	Kernel Function RBF	80.7 ($\pm$1.3)	79.8 ($\pm$2.8)	80.3 ($\pm$1.7)
	Kernel Function Polynomial	79.7 ($\pm$4.3)	75.1 ($\pm$2.7)	77.4 ($\pm$3.3)
	Least Square	78.4 ($\pm$4.9)	76.2 ($\pm$2.8)	77.3 ($\pm$3.7)
x-y	Kernel Function RBF	88.7 ($\pm$3.1)	85.7 ($\pm$4.3)	87.2 ($\pm$3.5)
	Kernel Function Polynomial	88.1 ($\pm$3.0)	83.5 ($\pm$2.7)	85.8 ($\pm$2.8)
	Least Square	84.1 ($\pm$2.8)	83.1 ($\pm$1.3)	83.6 ($\pm$1.9)

Table 3.9 Classification Results for KNN classifiers

Signals	KNN Methods	Performance		
		Sensitivity	Specificity	Accuracy
x	Cityblock	84.8 ($\pm$3.7)	82.0 ($\pm$3.3)	83.4 ($\pm$3.6)
	Euclidean	84.1 ($\pm$2.3)	83.1 ($\pm$2.8)	83.6 ($\pm$2.6)
	Cosine	82.6 ($\pm$4.9)	79.5 ($\pm$5.3)	81.1 ($\pm$4.7)
	Correlation	82.3 ($\pm$3.2)	81.2 ($\pm$4.3)	81.7 ($\pm$4.0)
y	Cityblock	83.6 ($\pm$4.1)	79.1 ($\pm$3.1)	81.4 ($\pm$3.6)
	Euclidean	81.1 ($\pm$4.3)	77.6 ($\pm$4.2)	79.3 ($\pm$4.2)
	Cosine	80.8 ($\pm$3.5)	77.6 ($\pm$3.2)	79.2 ($\pm$3.3)
	Correlation	79.5 ($\pm$3.8)	76.1 ($\pm$4.3)	77.8 ($\pm$4.1)
x-y	Cityblock	90.7 ($\pm$1.9)	88.1 ($\pm$2.2)	89.4 ($\pm$2.1)
	Euclidean	88.3 ($\pm$2.8)	87.9 ($\pm$3.3)	88.1 ($\pm$2.9)
	Cosine	88.2 ($\pm$3.7)	85.2 ($\pm$5.1)	86.7 ($\pm$4.3)
	Correlation	87.1 ($\pm$3.1)	85.9 ($\pm$3.3)	86.5 ($\pm$3.5)

Table 3.10 Comparison of the performance from various methods

Methods	Classifier-Method	Features	Accuracy (%)
Guohun [8], 2013	SVM RBF Kernel	Delay Permutation Entropy	84
Sharma [5], (2015)	SVM-LS Method	DWT, Entropy Measures	84
Sharma [10], (2014)	SVM-LS Method	EMD, ASE, AVIF	85
Rajeev [7], (2015)	SVM-LS Method	EMD, Entropy Measures	87
Proposed Method in EMD Domain	KNN City-block Distance	EMD, Log-energy Entropy	88
Proposed Method in DWT Domain	KNN City-block Distance	DWT, Log-energy Entropy	89
Proposed Method in EMD-DWT Domain	KNN City-block Distance	EMD-DWT, Log-energy Entropy	91

in the first five IMFs in EMD domain give better accuracy than the methods in [5], [8] and [10]. However, when the same features are extracted from the five sub-bands of the DWT sub-bands, the accuracy improves and thus, provides better performance than the method described in [7]. When, the features are obtained from the five DWT sub-bands of the first two IMFs in EMD domain, the KNN classifier provides 91% accuracy. It should be noted that the methods in [5], [7] and [10] provides the classification utilizing only 50 signals from focal and non-focal EEG each without any segmenting into 2s recordings. However in this thesis, the classification performance has also been provided, in table 3.11, utilizing the whole database containing 3750 signals from focal and no-focal EEG each, which leads the performance to be significant in the statistical point of view. Besides, a smaller part of the database (20%) has been used as the training part whereas a larger portion (80%) has been utilized for testing, which makes the proposed method to be more generalized in comparison to the previous methods in [5] and [7]. However, while using the same data selection method as in [5] and [7], but having the segmentation of 2s records, the proposed

Table 3.11 Performance from our method using the whole database

Methods	Classifier-Method	Features	Accuracy (%)
Proposed Method in EMD Domain	KNN City-block Distance	EMD, Log-energy Entropy	85.7
Proposed Method in DWT Domain	KNN City-block Distance	DWT, Log-energy Entropy	87.2
Proposed Method in EMD-DWT Domain	KNN City-block Distance	EMD-DWT, Log-energy Entropy	89.4

method provides 90.5% accuracy with 91.3% sensitivity and 89.7% specificity if the segmentation is applied. The proposed method is also computationally quite fast. It takes on an average of 0.05 to 0.07 seconds to extract the features from a 2s data segment of focal or non-focal EEG records in MATLAB [49] using 2GB RAM in core-2-duo processor. The method could be faster if MATLAB could be run with multiple cores in parallel fashion. In fig. 3.6, the detailed flow chart of the algorithm is shown Some initial results of this analysis are described in [50], whereas the detailed results are published in [51].

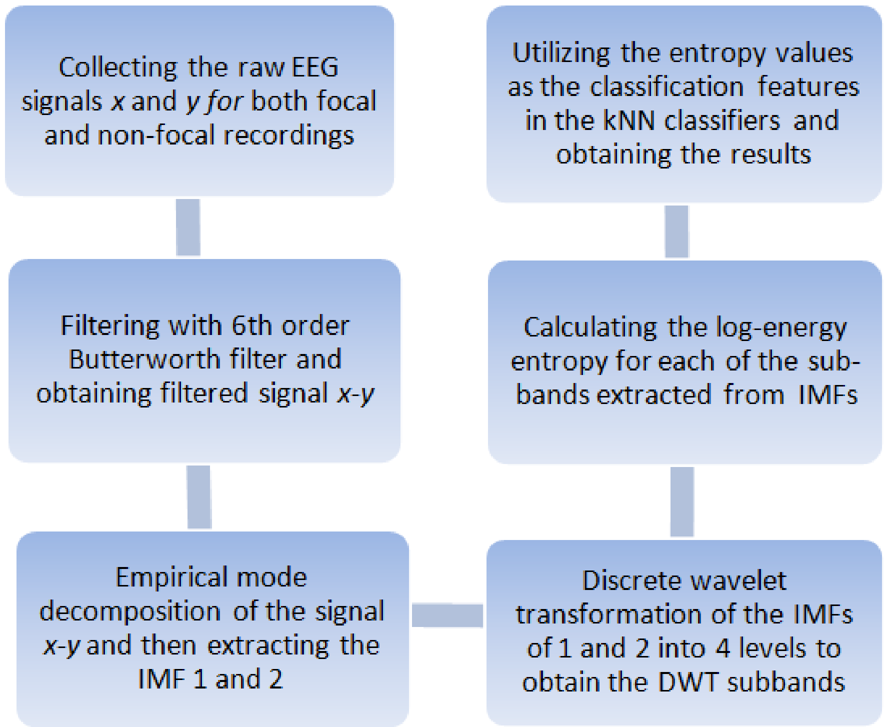

Fig. 3.6 The flow chart of the proposed method to classify the EEG signals

Chapter 4

Conclusion and Future Works

4.1 Conclusion

In this thesis, a statistical analysis of focal and non-focal EEG is carried out in the empirical mode decomposition (EMD) and discrete wavelet transform (DWT) domains. First, the EEG signals are analyzed in the EMD domain and its variants, EEMD and CEEMDAN domains. Several spectral entropy-based features such as the Shannon entropy, log-energy entropy and Renyi entropy are calculated in EMD, EEMD and CEEMDAN domains. Instead of using the direct signals from the EEG channels, the differences between two adjacent EEG channels are used due to its robustness to noise and interference. Next, similar analysis is carried out in discrete wavelet transform domains and the efficacy in discriminating the focal and non-focal EEG signals is investigated. The ability of the entropy-based features in separating the focal and non-focal EEG signals is explored utilizing the one-way ANOVA analysis and the box-whisker plots. After that, being motivated by the previous analysis, the signals are decomposed into a hybrid domain, where the DWT sub-bands are extracted from the IMFs of EMD domain, and the spectral entropy-based features are calculated in the EMD-DWT domain. Finally, well-known classifiers like support vector machine (SVM) and k-nearest neighbor (KNN) have been utilized to classify focal and non-focal EEG signals using appropriate features. It is observed that the entropy-based features perform better in DWT domain to classify the EEG signals than in EMD domain. In this regard, EEG signals have been investigated to observe the capability of the same features to discriminate the EEG data in the combined EMD-DWT domain. In short, the major observations of this thesis are:

(i) The spectral entropy-based features like Shannon entropy, log-energy entropy and Renyi entropy has the capability to discriminate between focal and non-focal EEG signals.

(ii) The features can differentiate the EEG signals more in EMD or DWT domain than in the band-limited case.

(iii) The discrimination capability of these features are quite similar in EMD domain or in its variants like EEMD or CEEMDAN domains, but in EMD domain, it takes less time to extract the features.

(iv) The features also perform quite well in the discrete wavelet transform (DWT) domain.

(v) The best performance is achieved when the signals are analyzed in a hybrid domain like EMD-DWT domain.

(vi) Log-energy entropy provides highest accuracy for classifying focal and non-focal EEG signals when incorporated in KNN classifier.

Overall, the proposed classification method reports a significant improvement in terms of sensitivity, specificity and accuracy in comparison to the existing techniques. This analysis may encourage the researchers to develop improved algorithms classify these signals Furthermore, for being computationally fast, the proposed method has the potential for identifying the epileptogenic zones, which is an important step prior to resective surgery usually performed on patients with low responsiveness to anti-epileptic medications.

4.2 Future Works

Although, the proposed method performs quite well in comparison to the available methods for focal and non-focal EEG data classification, but there are still some scopes to extend this work. For example, instead of DWT, wavelet packet or dual-tree complex wavelet transform (DT-CWT) can be explored whether that can perform better or not. Besides, EEMD or CEEMDAN can be explored for a part of the hybrid domain. Furthermore, the method can be explored with a larger database which would provide more statistical significance of this method.

References

[1] Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., and Engel, J., "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)". Epilepsia, 46(4), 470-472, 2005

[2] Eva Verhellen and Paul Boon, Chia-Hsiang Yang and Herming Chiueh,"EEG source localization of the epileptogenic focus in patients with refractory temporal lobe epilepsy, dipole modelling revisited," Acta Neurologica Belgica, vol. 107, pp. 71-77, September, 2007

[3] Sheng-Fu Liang, Yi-Chun Chen, Yu-Lin Wang, Pin-Tzu Chen, ChiaHsiang Yang and Herming Chiueh,"A hierarchical approach for online temporal lobe seizure detection in long-term intracranial EEG recordings," Journal of Neural Engineering, May, 2013.

[4] Ralph G. Andrzejak, Kaspar Schindler and Christian Rummel, "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients", Physical Review E, 86, 046206 ,2012.

[5] Rajeev Sharma, Ram Bilas Pachori and U. R Acharya, "An Integrated Index for the Identification of Focal Electroencephalogram Signals Using Discrete Wavelet Transform and Entropy Measures", Entropy, vol. 8, no. 17, pp. 5218-5240, 2015.

[6] R. Yadav, M. N. S. Swamy and R. Agarwal "Rapid Identifi cation of Epileptogenic Sites in the Intra-cranial EEG", proc. of 33rd Annual International Conference of IEEE EMBS, September, 2011.

[7] Rajeev Sharma, Ram Bilas Pachori and U. Rajendra Acharya, "Application of Entropy Measures on Intrinsic Mode Functions for the Automated Identifi cation of Focal Electroencephalogram Signals", Entropy, vol. 2, no. 17, pp. 669-691, 2015.

[8] Guohun Zhu, Yan Li, Peng Paul Wen, Shuaifang Wang, and Min Xi, \Epileptogenic focus detection in intracranial EEG based on delay permutation entropy", proc. of AIP International Conf., vol. 31, 2013.

[9] P.W. Mirowski, Yann LeCun, D. Madhavan and R. Kuzniecky, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG", proc. of IEEE Workshop on Machine Learning for Signal Processing, pp. 244-249, 2008.

[10] Sharma, R.; Pachori, R.B.; Gautam, S., "Empirical mode decomposition-based classification of focal and non-focal EEG signals", in Proceedings of the International Conference on Medical Biometrics, pp. 135-140, Shenzhen, China, 2014.

[11] S M Shafiul Alam and M. I. H. Bhuiyan, "Detection of Seizure and Epilepsy using Higher order Statistics in the EMD Domain," IEEE Transactions on Information Technology in BioMedicine, vol. 17, issue. 2, pp. 312-318, March, 2013.

[12] Hojjat Adeli, Samanwoy Ghosh-Dastidar and Nahid Dadmehr , "A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Sub-Bands to Detect Seizure and Epilepsy,"IEEE Transactions on Biomedical Engineering, vol:54 , Issue: 2 , pp: 205-211, Feb, 2007.

[13] V. Bajaj, R.B. Pachori, "EEG Signal Classification Using Empirical Mode Decomposition and Support Vector Machine", Proc. of International Conference on Soft Computing for Problem Solving, December 20-22, 2011.

[14] Kai Fu, Jianfeng Qua, Yi Chai and Tao Zou, "Hilbert marginal spectrum analysis for automatic seizure detection in EEG signals", Elsevier Journal of Biomedical Signal Processing and Control, vol. 18, 2015.

[15] A. B. Das, M. I. H. Bhuiyan and S M S. Alam,"Classification of EEG signals using normal inverse Gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection", Springer Journal onSignal, Image and Video Processing, vol. 10, issue. 2, pp. 259-266, 2014.

[16] S M Shafiul Alam and M. I. H. Bhuiyan, "Detection of Epileptic Seizures using Chaotic and Statistical Features in the EMD Domain," Annual IEEE Conference, India (INDICON), December 2011.

[17] Alexandros T. Tzallas, Markos G. Tsipouras and Dimitrios I. Fotiadis, "Epileptic

Seizure Detection in EEGs Using Time-Frequency Analysis," IEEE Transactions on Information Technology in Biomedicine, vol. 13, no. 5, September 2009..

[18] R. Yadav, M.N.S.Swamy, and R. Agarwal, "Model-Based Seizure Detection for Intracranial EEG Recordings," IEEE Transactions On Biomedical Engineering. vol. 59, no. 5, pp. 1419-1428, May, 2012.

[19] Dragoljub Gajic, Zeljko Djurovic, Stefano Di Gennaro and Fredrik Gustafsson, "Classification OF EEG Signals For Detection of Epileptic Seizures Based on Wavelets and Statistical Pattern Recognition," Biomedical Engineering: Applications, Basis and Communications, vol. 26, Issue: 02, April, 2014.

[20] 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. vol. 09, Issue. 38, March, 2015.

[21] Rajeev Yadav , A. K. Shah, J.A. Loeb, M.N.S. Swamy and R. Agarwal, "Morphology-Based Automatic Seizure Detector for Intercerebral EEG Recordings," IEEE Transactions on Biomedical Engineering,vol. 59, no. 7, 2012.

[22] A. S. Zandi, M. Javidan, G.A. Dumont and R. Tafreshi, "Automated Real-Time Epileptic Seizure Detection in Scalp EEG Recordings Using an Algorithm Based on Wavelet Packet Transform," IEEE Transactions On Biomedical Engineering. vol. 57, issue. 5, pp. 1639-1651, June, 2010.

[23] Liu, Y., Zhou, W., Yuan, Q., Chen, S., "Automatic seizure detection using wavelet transformation and SVM in long term intracranial EEG", IEEE Trans. Neural Syst. Rehabil. Eng. vol. 6, no. 20, 2012.

[24] S. Aydin, HMK Saraoglu, S. Kara, "Log energy entropy-based EEG classification with multi-layer neural networks in seizure", Springer Journal of ANN Biomed Engg., vol. 12, no. 37, pp. 2626-2630, December, 2009.

[25] M. I. H. Bhuiyan and A. B. Das, "A Subband Correlation-based Method for the Automatic Detection of Epilepsy and Seizure in the Dual Tree Complex Wavelet Transform Domain", Proc. of IEEE Conference on Biomedical Engineering and

Sciences, pp. 811-816, Malaysia, December, 2014.

[26] Fuad, N., Shah Alam and Mansor, W. ; Lee, K.Y., "Wavelet packet analysis of EEG signals from dyslexic children with writing disability", proc. of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2890 - 2893, July, 2013.

[27] A. Feltane, G.F. Boudreaux Bartels, Y. Boudria and W. Besio, "Seizure detection using empirical mode decomposition and timefrequency energy concentration", proc. in IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2013.

[28] Lorena Orosco, Eric Laciar, Agustina Garcs Correa, Abel Torres and Juan P. Gragna,"An Epileptic Seizures Detection Algorithm based on the Empirical Mode Decomposition of EEG ", proc. of 31st IEEE Conference onEngineering in Medicine and Biology Society (EMBS), Minnesota, USA, September 2-6, 2009.

[29] V. Bajaj and R. B. Pachori, "Classification of Seizure and Non-seizure EEG Signals Using Empirical Mode Decomposition", IEEE Transactions On Information Technology in Biomedicine, vol. 16, no. 6, November, 2012.

[30] A. B. Das, M. J. H. Pantho and M. I. H. Bhuiyan,"Discrimination of Scalp EEG signals in Wavelet Transform Domain and Channel Selection for the Patient-invariant Seizure Detection", Proc. of IEEE Conference on Electrical and Electronic Engineering, Bangladesh, pp. 77-80, November, 2015.

[31] Panda, R., Khobragade, P.S., Jambhule, P.D., Jengthe, S.N., Pal, P.R. and Gandhi, T.K., "Classification of EEG signal using wavelet transform and support vector machine for epileptic seizure prediction", IEEE Systems in Medicine and Biology (ICSMB)), December, 2010

[32] Shufang Li, Weidong Zhou, Qi Yuan, and Yinxia Liu,"Seizure Prediction Using Spike Rate of Intracranial EEG", IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 21, no. 6, September 2013

[33] Md. Ashfanoor Kabir and Celia Shahnaz, "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains", Elsevier Journal on Biomedical

Signal Processing and Control, vol. 07, issue. 05, pp. 481-489, 2012.

[34] Kopsinis, Y. ; and McLaughlin, S., "Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding", IEEE transaction on Signal Processing,, vol. 57, pp. 1351-1362, 2009.

[35] Bern-Barcelona database: http://ntsa.upf.edu/downloads/andrzejakrg-schindler-k-rummel-c-2012-nonrandomness-nonlinear-dependenceand

[36] G. Repovs, "Dealing with Noise in EEG Recording and Data Analysis,"Informatica Medica Slovenica, vol. 15, no. 1, 2010.

[37] Christophe C. Jouny and Gregory K. Bergey, "Characterization of early partial seizure onset: Frequency, complexity and entropy", proc. of Elsevier Journal on Clinical Neuro-physiology, vol. 123, Issue. 4, pp. 658-669, January, 2012.

[38] El-Kishky, A., "Assessing entropy and fractal dimensions as discriminants of seizures in EEG time series", proc. of 11th International Conference on Information Science, Signal Processing and their Applications, (ISSPA), pp. 92-96, July, 2012.

[39] Dan Xiao, Zhengdong Mu and Jianfeng Hu, "Classification of Motor Imagery EEG Signals Based on Energy Entropy", proc. of IEEE International Symposium on Intelligent Ubiquitous Computing and Education, pp. 61-64, 2009.

[40] Lisha Sun ; Guoliang Chang ; Hongrong Tang, "Wavelet Packet Entropy in the Analysis of EEG Signals", proc. of 8th International Conference on Signal Processing, vol. 4, Beijing, 2006.

[41] See, A.R. , Tainan, Taiwan and Chih-Kuo Liang, "A study on sleep EEG Using sample entropy and power spectrum analysis", proc. of Defense Science Research Conference and Expo (DSR), pp. 1-4, August, 2011.

[42] Inuso, G., La Foresta, F. , Mammone, N. and Morabito, F.C., "Brain Activity Investigation by EEG Processing: Wavelet Analysis, Kurtosis and Renyi's Entropy for Artifact Detection", proc. of International Conference onInformation Acquisition, pp. 195-200, 2007.

[43] V. Vapnik, "The Nature of Statistical Learning Theory, New York: Springer-

Verlag", 1995.

[44] Ashkan Yazdani, Touradj Ebrahimi and Ulrich Ho mann, "Classification of EEG Signals Using Dempster Shafer Theory and a K-Nearest Neighbor Classifier ", proc. of IEEE EMBS Conference on Neural Engineering, 2009.

[45] Fatemeh Bahari and Amin Janghorbani, "EEG-based Emotion Recognition Using Recurrence Plot Analysis and K Nearest Neighbor Classifier", proc. of 20th Iranian Conference onBiomedical Engineering (ICBME), December, 2013

[46] E. Parvinnia, M. Sabetia, M. Zolghadri Jahromi and R. Boostani, "Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm", proc. of 20th Iranian Conference onBiomedical Engineering (ICBME), vol. 26, Issue. 1, pp. 1-6, January, 2014.

[47] M. Torres, M. Colominas, G. Schlotthauer, P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise", Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4144-4147, 2011.

[48] A. R. Hassan and M. I. H. Bhuiyan, "Computer-aided sleep staging using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and bootstrap aggregating", in Elsevier Journal on Biomedical Signal Processing and Control, vol. 24, February, 2016.

[49] MATLAB: http://www.mathworks.com/products/matlab/

[50] A. B. Das and M. I. H. Bhuiyan, "Discrimination of Focal and Non-Focal EEG Signals using Entropy-based Features in EEMD and CEEMDAN Domains", proc. of 9th IEEE International Conference on Electrical and Computer Engineering (ICECE), pp. 435-438, Dhaka, Bangladesh, December, 2016.

[51] A. B. Das and M. I. H. Bhuiyan, "Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain", Elsevier Journal On Biomedical Signal Processing and Control, vol. 29, pp. 11-21, 2016.