

# A Review on Seizure Detection Systems with Emphasis on Multi-domain Feature Extraction and Classification using Machine Learning

*Dattaprasad Torse*

Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Jnana Ganga, Khanapur Road, Udyambag, Belagavi, Karnataka 590008, Tel.: +91 831 240 5500, Belagavi, India  
datorse@git.edu

*Veena Desai*

Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Jnana Ganga, Khanapur Road, Udyambag, Belagavi, Karnataka 590008, Tel.: +91 831 240 5500, Belagavi, India  
veenadesai@git.edu

*Rajashri Khanai*

Department of Electronics and Communication Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Udyambag, Angol Main Road, Belgaum, Karnataka 590008, India  
Tel/: +91 831 244 0322, Belagavi, India  
rajashri.khanai@gmail.com

## **Abstract**

At present, manual observation of the electroencephalogram (EEG) signals is the prime method for diagnosis of epileptic seizure disorders. The method is a time consuming and error prone as it involves errors due to fatigue in continuous monitoring of nonlinear and nonstationary EEG signals. Out of approximate 1% of the world's epilepsy patients more than 25% cannot be treated correctly due to erroneous diagnosis. The automated seizure detection system can prove efficient by making the process reliable and faster. This paper reviews multi-domain feature extraction and machine learning classification techniques used in automated seizure detection systems. To analyse subtle variations in EEG, signal decomposition algorithms have been used in time, frequency, joint time-frequency, and nonlinear domain. The statistical and entropy parameters are the key features to discern normal from the seizure EEG signals. Machine learning plays a critical role in extracting meaningful information out of the extracted features. The paper also evaluates the performance of Multilayer Perceptron Neural Network, naïve Bayes, Least Square Support Vector Machine, k nearest neighbour, and random forest classifiers using sensitivity, specificity and accuracy metrics. A seizure detection technique is developed by decomposing the EEG signals by means of Tunable-Q Wavelet Transform (TQWT). To quantify the complexity of the individual multivariate sub-bands of the biomedical signals TQWT proves effective with varied values of Q factor suitable for analyzing signals with oscillatory and non-oscillatory nature. The highest accuracy of 97.3% is obtained using random forest classifier for the combination of spectral, Shannon and Kraskov entropy features. The paper compares the performance of feature extraction and classification techniques for the implemented system. The comparison explores possibility of hardware implementation of real time seizure detection scheme.

**Keywords:** Seizure Detection, Tunable-Q Wavelet Transform, Shannon entropy, Kraskov entropy, Least Square Support Vector Machine, Random Forest

## 1. Introduction

Epilepsy is a widespread brain disorder which affects a variety of mental and physical actions. When more than two episodes of seizures occur in a lifespan of a person then they are categorized as a seizure patient. Epileptic seizures are provoked by group of nerve cells which affect a person's normal behavior. This sudden brain signal change is life intimidating in few cases as it can cause physical injury to the affected person. In the form of partial and generalized seizures the abnormal brain activity poses a very important health concern to the patient. Partial seizures start with a specific area of brain and usually called the epileptic foci. Partial seizures may or may not affect conciseness of a person. Generalized seizures involve seizure signals originating from most part of the brain and cause loss of mental alertness and muscle spasms. The process of 'epileptogenesis' is highly unpredictable and the risk involved in the form of injury is very high (D. Buck, 1997). The seizure disorder occurs due to several causes such as birth asphyxia, stroke, traumatic brain injury or brain infections. The seizure disorders are not preventable or in some cases not completely curable but with the help of anticonvulsant drugs the life threatening seizures can be controlled in majority of the cases (Englander J., 2014).

The episode of epileptic seizures occurs as the brain's controlled neonatal firing circuit malfunctions and causes excessive electrical discharge by a group of nerve cells in the brain cortex. This processing is sudden and unpredictable. Depending upon the side of cortex, out of four sides namely frontal, parietal, occipital and temporal which originates the abnormal signals, the abnormalities in the motor control results in tonic-clonic movements of muscles and joints. The discharge of electrical energy in a normal brain cells is controlled and produces variations that are in normal magnitude ranges. However an abrupt and large transient rush of energy by the brain cells results in epileptic seizures. An epileptic seizure can show variation in properties of brain waves which can result in a short term muscle movement to severe convulsions. These variations mainly depend on the area of the brain from which the energy is generated, the level of electrical energy discharge and the total area over which this energy is extended in the event of abnormal activity (Acharya U., 2013).

The working of brain and its properties that cause epileptic activities are still a mystery. When a person experiences epileptic activity the possible observable signs are sudden movement of the body parts, loss of concentration, muscle involuntary movement, disturbance in visual and auditory senses and mood disorder. There can be several changes in a person suffering from mild to severe epileptic attack which are beyond the range of normal observations. When the seizures are seen in children who have limited knowledge about the situation that they experience it become difficult to notice the seizure onset. This pre-seizure behavior changes in children are linked to the behavioral disorder. Hence, children with epileptic disorder need continuous monitoring and thus the epilepsy observation is a continuous process. In order to make the process fully automated with indication of seizure occurrence many signal processing algorithms need to be considered with detailed analysis. In order to detect epilepsy using automated Computer Assisted Diagnostic (CAD) techniques using EEG signals understanding the physiological aspects of the seizure signal class is essential (Sanei, Saeid).

In this work, a Tunable-Q Wavelet Transform (TQWT) (IW Selesnick, 2011) based seizure detection system is proposed which uses spectral and entropy based features to test performance of five classification algorithms. Figure 1 shows the proposed block diagram of TQWT sub-band's spectral and entropy feature based seizure classification system. As shown in figure, the features for two TQWT sub bands namely, sub-band 1 and sub-band 16 are taken to consideration for the feature extraction from normal and seizure EEG signals. The oscillatory information contained in the signal is reflected in the TQWT sub bands with low frequency content represented in the first sub band and the last sub band representing the high frequency oscillation. The EEG signal decomposition technique quantifies the sub band spectral and entropy features for low and high frequencies and this can be a widespread method to detect seizures from other EEG recording with appropriate choice of the Q-parameter. The efficacy of features extraction and classification

algorithms is tested using computational complexity and the effectiveness of the proposed technique for hardware implementation. The performance of the designed algorithm is tested on the real time data recorded from epileptic and normal patients from a local hospital. The efficacy of the tested method motivates to build an automated seizure detection system to assist neurologists to diagnose the epilepsy and related disorders with improved accuracy.

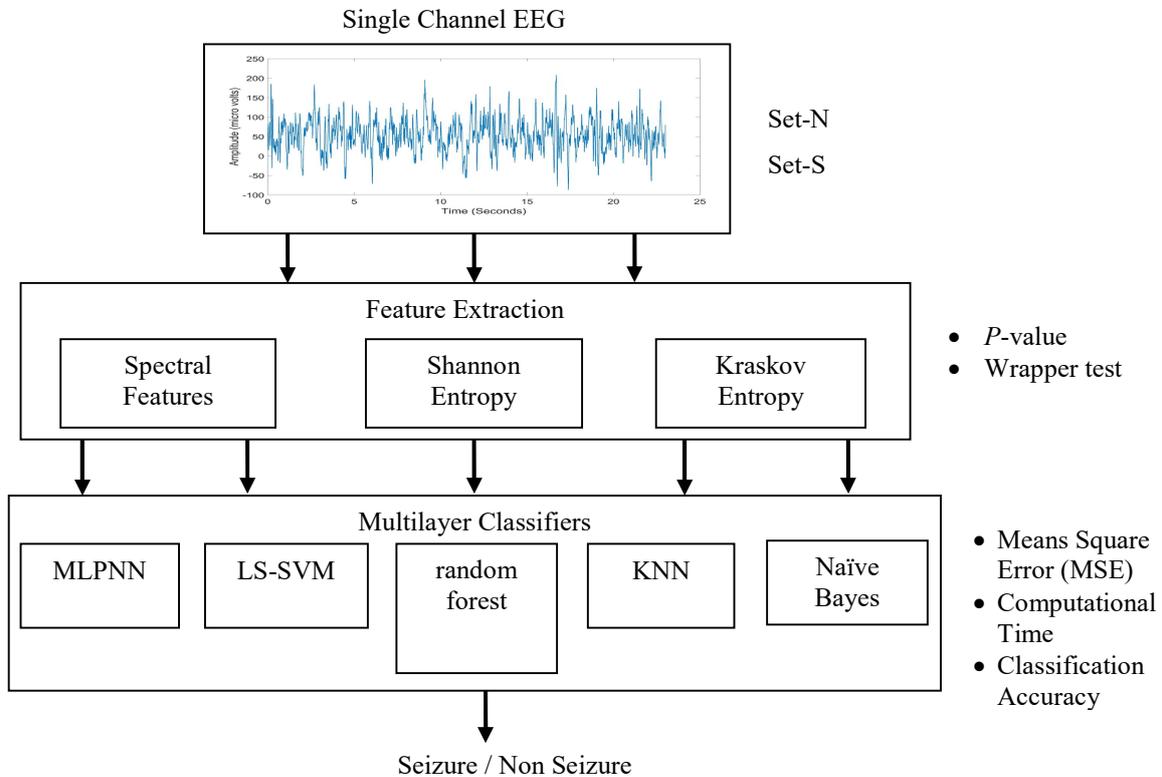


Figure 1. Flow diagram of the structure representing a step-by-step idea of the proposed method

The remaining part of the paper is arranged as follows: In the following section, the different methods are reviewed that are currently used to detect epilepsy. In Section 2, a review on state-of-the-art algorithms of EEG signal feature extraction and classification is presented for seizure detection system. The results obtained using the TQWT feature extraction and classification techniques are illustrated in the section 3 and discussion on comparison between various methods is presented in section 4. Section 5 concludes the paper.

## 2. Literature Review

This paper reviews seizure detection systems that have been developed using the EEG database from University of Bonn, Germany (R. G. Andrzejak, 2001). The decomposition of EEG signals have been implemented time, frequency, joint time-frequency, and nonlinear domain. The statistical and entropy based parameters are used as the key features in machine learning algorithms to separate signals.

### 2.1. Datasets

The dataset used in this work is mainly consisting of five subsets, out of which two cases, namely, normal (set N), and seizure (set S) were considered for the system developed in this work. The EEG signals of subsets N was recorded in normal signal intervals for 5 patients (subjects) in the epileptogenic region. The subset-S is composed of EEG signals with seizure action recorded from all the electrodes showing seizure activity. The bio potentials of subsets N and S were recorded intracranially. The 128-channel amplification system is used to record the signals with an average

common. To record the signal, depth electrodes were inserted symmetrically into the hippocampal formations. The basal and lateral regions of the neocortex were used to insert strip electrodes. The EEG was recorded during epileptic seizures is termed as ictal activity. Figure. 2–3 show sample recordings of datasets N and S respectively. Each dataset contains 100 single channel recording with a 4097 samples and a sampling rate of 173.6 Hz.

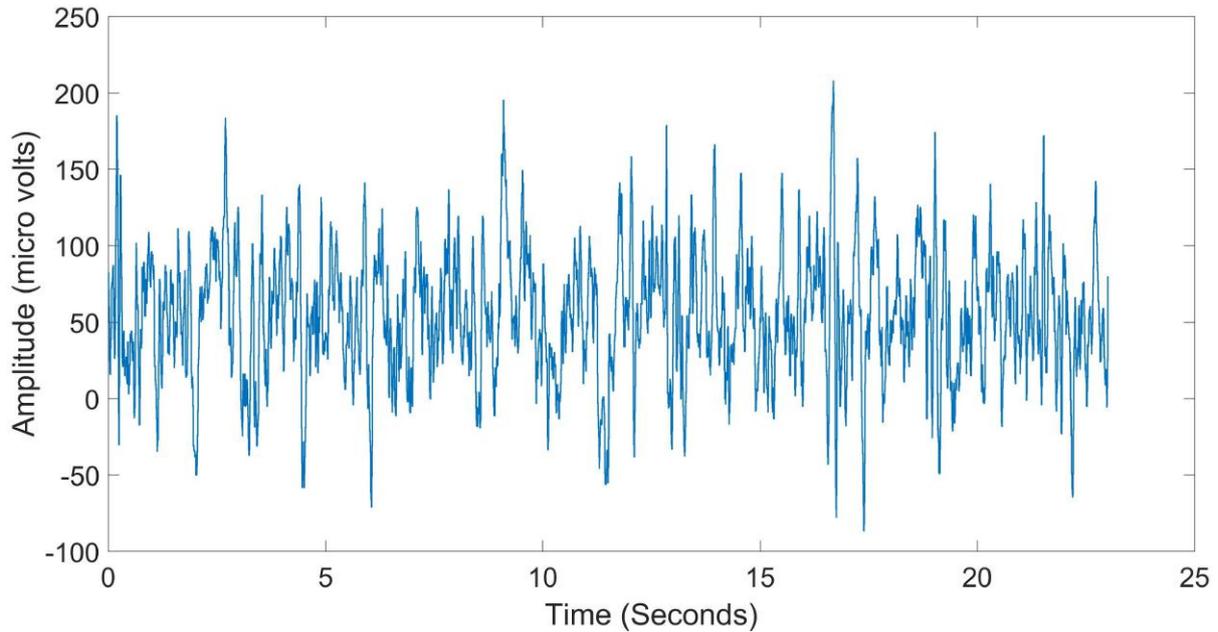


Figure 2. Normal EEG signal sample

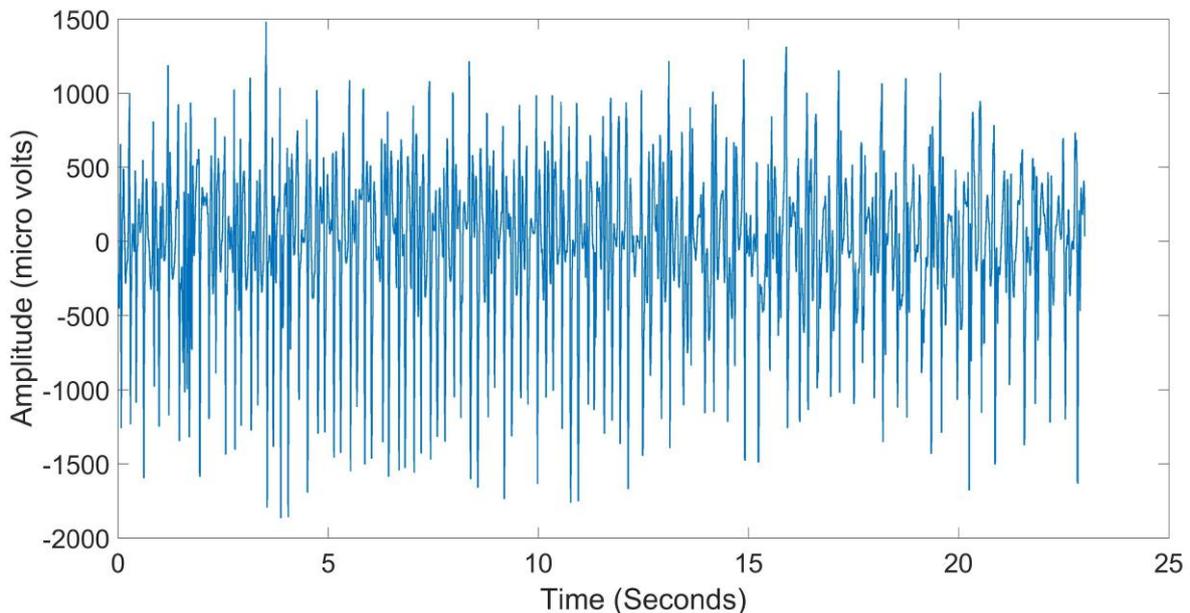


Figure 3. Epileptic EEG signal sample

Another clinical data was collected from patients receiving routine EEG examinations at Dr. Mohire’s Neurology Research Centre, Kolhapur, India. In total, routine EEG data from 30 participants (15 men and 15 women) was included whose ages varied from 12 to 25 years. There are normal and seizure activity in the recorded EEG data signal. The dataset includes signals of patients who suffered headache but diagnosis did not reveal any kind of epilepsy. The written permission was obtained from the patients and the neurologist to approve the research work. To find the latent

of the technique in real-time, training and test samples were divided into 70% and 30% respectively.

## **2.2. Pre-processing**

The process of 10-20 EEG recording system can suffer from contaminations and thus generate ambiguous signal affected by various noise sources. Although the vigilant design of the recording scheme and apt recording of signal procedures will be able to minimize the noise generated, a number of electrophysiological signals, mainly electrooculogram (EOG), needs careful elimination using signal processing techniques. The main cause of electrical spike generation during eye blink is an electrical dipole in the human eye which is a result of a positive cornea and negative retina. The Independent Component Analysis (ICA) was independently used to filter artifacts from EEG signals by (Maarten Mennes et al., 2010, Manousos A. Klados et al., 2011, Nadia Mammone, 2012). The combination of Discrete Wavelet Transform (DWT) and ICA is studied by (Mingai Li, 2012, Nadia Mammone et al., 2014). Another technique of adaptive noise cancellation using DWT was tested as a online predictive tool for ocular artifacts (Hong Peng, 2013). A method (Qinglin Zhao, 2014) demonstrates the elimination of ocular artifact by utilizing the Adaptive Predictive Filter (APF) techniques to improve true EEG by finding EEG eye movement artifact signal amplitudes. An automated online filtering is developed by using a combination of wavelet decomposition, ICA, and thresholding. The adaptive filtering has been compared with the ICA and PCA based methods in one of the studies to find computationally efficient method (2016, Dattaprasad A. Torse). The use of joint time-frequency domain approaches to process the EEG signals and preserve the useful information is explored by means of Empirical Mode Decomposition (EMD) (Gang Wang, 2016). The EMD method outperform the wavelet based methods as in EMD the signal is sifted for the predefined levels. The section of preprocessing stage depends on the application, dataset type used and complexity of the algorithm in terms of computation time.

## **2.3. Feature Extraction**

In many research papers, use of University of Bonn, Germany, database is made to categorize normal and epileptic EEG signals by means of pattern classifiers such as supervised and unsupervised. In (İnan Güler, 2005, Pari Jahankhani, 2006, Hojjat Adeli, 2007), wavelet based coefficients have been used as features and classifiers were tested for classification of EEG into normal and seizure states. Most of the studies have aimed on implementation using software tools against the development of low cost, computationally efficient hardware for seizure detection. The combination of wavelet and PCA and Independent Component Analysis (ICA) with SVM classifiers was studied to classify EEG signals (Subasi and Gursoy, 2010). The epileptic EEG was analyzed with entropy features using Principal Component Analysis (PCA) enhanced Radial Basis Function (RBF) neural network and Support Vector Machine (SVM) classifier by (Ghosh-Dastidar et al., 2008, Oliver Faust et al., 2010) respectively. The Discrete Wavelet Transform (DWT) has been extensively used to decompose EEG signal into different sub-bands to find detail and approximation coefficients (Tapan Gandhi, 2011, Rami J. Oweis, 2011, Esma Sezer, 2012). The kNN classifier was tested for DWT based variance features (Shufang Li, 2013). The 1-NN classifier resulted in the overall accuracy of 99% as compared to complicated SVM classifier. The DWT-entropy feature set resulted 100% accuracy in another study by (Yantindra Kumar, 2014). Using multiple classifiers the performance of DWT and entropy is assessed in (Oliver Faust, 2015). In another study on DWT (Jiang-Ling Song, 2016), automated detection of epileptic EEGs is presented using a novel fusion features which characterize the similarity between signals and extreme learning machine. The LS-SVM, feed forward multi-layer perceptron neural network, kNN classifiers have been considered to identify epileptic seizure using Wavelet Packet Decomposition (WPD). Five mental tasks were classified using SVM for categorization of epileptic seizures using

WPD and approximate entropy (ApEn) and sample entropy (SampEn) (Yong Zhang et al., 2015). The effect of WPD based log energy entropy is studied by (Raghu, Sriraam, 2015).

#### 2.4. Tunable Q Wavelet Transform

In the literature, existing time-frequency domain analysis techniques are providing constant Q-factor in accordance with the signal. On the other hand, TQWT can tune the wavelet's Q-factor in accordance with the signal. In TQWT it is apparent that the EEG signal under study can be decomposed into sub-bands with different properties. Thus TQWT is a flexible and fully discrete wavelet transform that is particularly suitable for analyzing oscillatory signals (IW Selesnick, 2011). It achieves flexibility by adjusting its input parameters such as Q-factor ( $Q$ ), rate of over-sampling or redundancy ( $r$ ), and number of levels of decomposition ( $J$ ). The excessive noise that is unwanted is controlled by the parameter ' $r$ ' in order to localize the wavelet temporally. Without affecting the shape of the decomposed signal ' $Q$ ' controls the number of oscillations of the wavelet. TQWT decomposes an input signal  $x(n)$  into  $(J+1)$  sub-bands for ' $J$ ' level of decomposition by employing two channel filter banks in an iterative method. The two channel filter banks are applied to the low-pass sub-band. In every stage,  $x(n)$  is decomposed into  $l^0(n)$  and  $h^0(n)$ . The  $l^0(n)$  is low-pass sub-bands and  $h^0(n)$  high-pass sub-bands. The low and high pass sub-bands have sampling frequency of  $\alpha f_s$  and  $\beta f_s$  respectively. The scaling factors are denoted by  $\alpha$  and  $\beta$  and  $f_s$  is the sampling frequency of  $x(n)$ . The Low-pass filter  $H_0(\omega)$  and  $H_1(\omega)$ , with low-pass scaling  $\alpha$  and high-pass scaling  $\beta$  are applied to produce  $l^0(n)$  and  $h^0(n)$ . However, perfect reconstruction is ensure without redundancy when  $\alpha$  and  $\beta$  have relations given by:  $0 < \alpha < 1$ ,  $0 < \beta \leq \alpha^{-1}$  and  $\alpha + \beta > 1$ . The  $H_0^J(\omega)$  and  $H_1^J(\omega)$  are the equivalent frequency response generated after  $J$ -level for low and high-pass sub-bands respectively and mathematically represented as:

$$H_0^J(\omega) = \begin{cases} \prod_{n=0}^{J-1} H_0(\omega/\alpha^n) & \text{if } |\omega| \leq \alpha^J \pi \\ 0 & \text{if } \alpha^J \pi \leq \omega \leq \pi \end{cases} \quad (1)$$

$$H_1^J(\omega) = \begin{cases} H_1(\omega/\alpha^{J-1}) \prod_{n=0}^{J-2} H_0(\omega/\alpha^n), & \text{if } (1-\beta)\alpha^{J-1}\pi \leq |\omega| \leq \alpha^{J-1}\pi \\ \text{else } \omega \in [-\pi, \pi] \end{cases} \quad (2)$$

where,

$$H_0(\omega) = \theta \left( \frac{\omega + (\beta - 1)}{\alpha + \beta - 1} \right)_0$$

$$H_1(\omega) = \theta \left( \frac{\alpha\pi - \omega}{\alpha + \beta - 1} \right)_0$$

The  $\theta(i)$  is the Daubechies filter's frequency response with two vanishing moments given by:

$$\theta(i) = 0.5(1 + \cos(i))\sqrt{2 - \cos(i)}, \quad |i| < \pi \quad (3)$$

The scaling factors are related to 'r' and 'Q' parameters as follows:

$$r = \frac{\beta}{1 - \alpha} \quad \text{and} \quad Q = \frac{2 - \beta}{\beta} \quad (4)$$

The prime reason to select TQWT over other time-frequency techniques is due to various advantages of TQWT. Firstly, when a signal with little or no oscillatory activities, e.g. EEG, needs analysis demands a wavelet transform having a low Q-factor. On contrary, oscillatory signals demand relatively high Q-factor. However, majority of the wavelet transforms are unable to tune the Q-factors for signals containing varied oscillatory behavior. TQWT solves the difficulty by adjustment in Q-factors. Secondly, TQWT has been applied for the analysis of various biomedical signals (Hasan Ahnaf, 2016, Ram Bilas Pachori, 2016, Abhijit Bhattacharyya, 2017, Shivnarayan Patidar, 2017). The "rational transfer functions" of the filters in TQWT enhance computational efficiency and enables a perfect reconstruction of wavelet transform. These advantages motivate the use of TQWT in decomposing the EEG signal into sub-bands for further processing in the proposed scheme. In this paper, computational complexity of the algorithm is analyzed in order to improve the current Information Transfer rate of current BCI based seizure detection system.

### 2.5. Spectral and Entropy Features

The idea of the paper is to test the classification performance of five classifiers namely, Multilayer Perceptron Neural Network (MLPNN), Least Square Support Vector Machine (LS-SVM), Naïve Bayes (NB), k Nearest Neighbour (kNN), and Random Forest (RF). For this purpose, spectral domain features are used to build primary feature space by combining minimum and maximum value, mean, median and standard deviation of EEG signals. This primary feature subset is combined with entropy features to build robust feature vector. The minimum and maximum values of EEG signals decomposed using TQWT are separately stored to create the first feature vector. The second feature vector is build using mean frequency, median frequency of the sub bands. In the third set only standard deviation of the absolute value of the coefficients are computed for each TQWT sub-bands. The standard deviation is effective frequency domain feature as it represents the average deviation of a random nature signal (Phinyomark A. et al.,2102).

The Mean Frequency (MF) is an average value of frequency computed as the summation of product of the EEG signal power spectrum divided by the total summation of the power spectrum (Phinyomark et al., 2012). The MF is also explained as the central frequency ( $f_c$ ) in (Du & Vuskovic, 2004). The definition of mean frequency (MF) is given by:

$$MF = \frac{\sum_{i=1}^N f_i P_i}{\sum_{i=1}^N P_i} \quad (5)$$

where  $f_j$  is the frequency component of EEG power spectrum at the frequency bin  $i$ ,  $P_i$  is the EEG power spectrum at the frequency bin  $i$ , and  $N$  is the length of frequency bin. In the study of EEG signal in time domain,  $N$  is usually defined as the next power of 2 from the length of EEG data.

Median Frequency (MFR) is a frequency at which the EEG power spectrum is divided into two regions with equal amplitude (Phinyomark A., et al., 2012a). MFR is moreover described as half of the total power. The description of MFR is as follows:

$$\sum_{i=1}^{MFR} P_i = \sum_{i=MFR}^N P_i = \sum_{i=1}^N P_i \quad (6)$$

In the standard deviation the averaging is computed with power and not using amplitude. It is calculated, first by squaring each of the deviations previous to computation of the average. In the final step, the square root is considered to balance for the initial squaring. The standard deviation is given by:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2 \quad (7)$$

where  $\mu$  is the mean.

### 2.5.1 Shannon Entropy

For a given discrete probability distribution, Shannon entropy (ShEn) is defined as a measure of on average how much information is necessary to identify random samples from the given distribution. ShEn gives the average information present in the EEG signal. It uses a non-normalized method for estimation of entropy (Shannon, 1948) based on energy content in the EEG. The non-normalization Shannon entropy is given by: (Shannon, 1948),

$$ShEn = H(x) = -\sum_x P(x) \log p(x) \quad (8)$$

As shown in Figure. 4, the ShEn feature values are negative due to the fact that the difference in the entropy parameters is negative.

### 2.5.2. Kraskov Mutual Information

Normally, the seizure detection is carried out by extracting a set of analytic features extracted from EEG signals. These features need to fulfill similarities for signals of same class and represent variations for different class. The hidden nonlinearity and non-stationarity in EEG signals often results in variations in analyzing these signals for extracting features. The time-frequency signal representations using wavelet transform based features is a suitable method for analyzing nonlinear and non-stationary EEG signals. Recently, the Kraskov entropy based nonlinear features were developed and found applicable in EEG signal analysis (K.A. Veselkov, et al., 2010, A. Kraskov, et al., 2008). For measuring and characterizing nonlinearities of EEG signals many authors have employed the Kraskov entropy. It measures the Shannon entropy or differential statistical entropy of the EEG signals using the kNN sample with some distance measures such as Euclidean or Hamming.

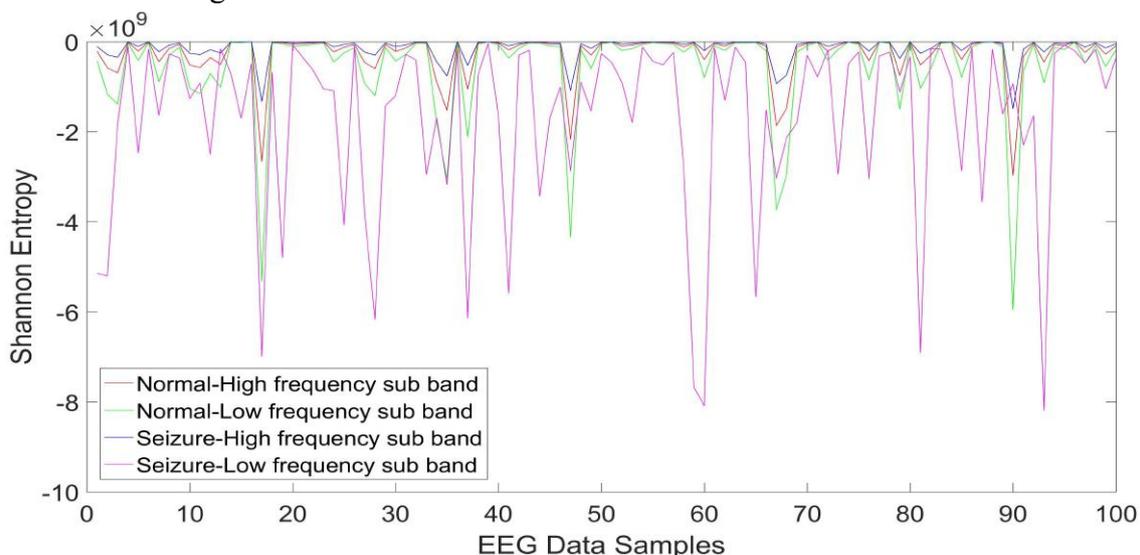


Figure 4. Plot of the Shannon entropy values for normal and seizure signal's high and low frequency sub bands

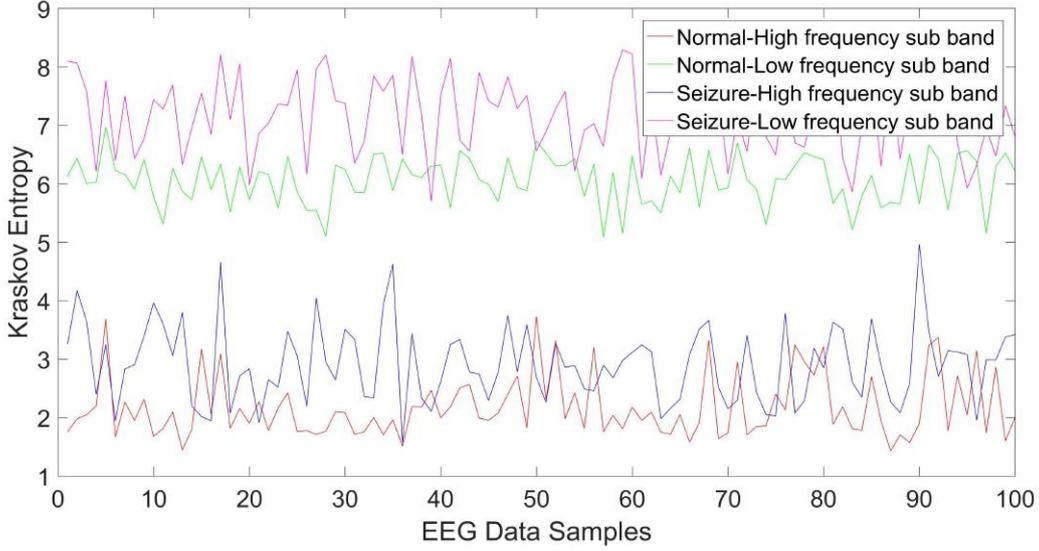


Figure 5. Plot of the Kraskov entropy values for normal and seizure signal's high and low frequency sub bands

In continuous-time domain, the differential statistical entropy of  $d$ -dimensional random variable  $X$  with unknown density function  $f(x)$  is defined as:

$$H(x) = -\int dx \mu(x) \log \mu(x) \quad (9)$$

The density function  $dx$  can be obtained by probability distribution function for the distance between  $x_i$  and the  $k$ -nearest neighbor samples. The above equation results in Kraskov entropy by measuring the  $k$ -NN entropy. It can be expressed as follows (Kraskov, et al., 2008):

$$\hat{H}(X) = -\psi(k) + \psi(N) + \log(C_d) + \frac{d}{N} \sum_{i=0}^N \log \varepsilon(i) \quad (10)$$

where  $(x_1, x_2, x_3, \dots, x_n)$  are  $n$  random samples of  $d$ -dimensional random variable  $X$ ,  $\psi(t)$  symbolizes the digamma function,  $C_d$  represents the volume of the  $d$ -dimensional unit ball that depends on the sample space. The term  $\varepsilon(i)$  is the distance between sample  $x_i$  and its  $k$ NN sample points in  $d$ -dimensional sample space. A more detailed explanation on mathematical aspects and other application is available from (Kraskov, et al., 2008). In Figure. 5, the KraEn features values are plotted for normal and seizure signal's low and high frequency sub bands.

## 2.6. Classification

In machine learning, classification of EEG signals deal with the task of categorizing a set of classes to which a new reading belongs, on the basis of a training set of EEG feature set containing occurrences whose class relationship is identified. Based on the spectral and entropy features of TQWT sub bands, the classification of EEG signals is tested using five classifiers (Bishop, Christopher M., 2006).

### 2.6.1 Multilayer Perceptron Neural Network (MLPNN)

The MLPNNs with the ability to learn and generalize are most commonly used classifiers in EEG analysis and seizure detection. They need smaller training set and work fast with less complexity in implementation (Haykin, 2001). In the MLPNN, every neuron  $j$  in the hidden layer sums its input signals by multiplying the signal  $x_i$  and the strengths of the individual correlation weights  $w_{pq}$ . The output  $y_q$  is then computed as a function of the summation as:

$$y_q = \left( \sum W_{qp} x_p \right) \quad (11)$$

where  $f$  is the activation function. The activation function is to be selected depending on the application as a sigmoid or radial basis function. The addition of differences between the desired and actual values of the output neurons is squared and computed as:

$$Sum = \frac{1}{2} \sum_q (y_{dq} - y_q)^2 \quad (12)$$

where  $y_{dq}$  is the desired value of output neuron  $q$  and  $y_q$  is the actual output of the neuron. Each weight  $w_{qp}$  is tuned to decrease  $Sum$  as fast as possible. Depending on the training algorithm employed the  $w_{qp}$  value is set for the computation (Haykin, 2001). The ANN model development primarily focuses on the training algorithms. A suitable training algorithm can result in the better model for prediction. An optimized training algorithm can also reduce the training time and provide a better accuracy. There are several training algorithms used to train a MLPNN and the most often used is the backpropagation algorithm (Haykin, 2001). The backpropagation algorithm is relatively easy to implement in which a search of an error surface using gradient descent is carried out. However, the error remaining in the local minima for indefinite time is a major challenge in the backpropagation algorithm. The long training sessions is another major concern in the backpropagation algorithm. Therefore, a lot of deviations to advance the convergence of the backpropagation were proposed in the literature.

### 2.6.2 Least Square Support Vector Machine (LS-SVM)

The focus in this paper is on the binary classification and LS-SVMs are least squares of SVM, perfectly suited for the application. In LS-SVM, a set of related supervised learning techniques analyze data, recognize patterns, and hence are used for classification and regression analysis. The method is based on using a quadratic error criterion with equality constraints as an alternative to a convex quadratic programming (QP) problem for classical SVMs. Least squares SVM classifiers, were proposed by (Suykens, 2002). The LS-SVMs are a class of kernel-based learning methods. The SVM is assuring classifier that reduces the error and increases the boundary to classify the data by separating hyper plane. The LSSVM is least squares formulation of SVM and contains the equality constraints. For two-class SVM, decision function is as follows (Suykens, 2002):

$$f(x) = Sign[\omega^T g(x) + b] \quad (13)$$

where  $w$  -  $d$ -dimensional weight vector,  $b$  - bias, and  $g(x)$  is a function that maps  $x$  into the  $d$ -dimensional space. To obtain  $w$  and  $b$  values, the subsequent optimization problem can be created in the subsequent way:

$$J(\omega, b, e) = \frac{1}{2} \omega^T \omega + \frac{y}{2} \sum_{i=0}^N e_i^2 \quad (14)$$

Minimize

subject to equality constraints

$$y_i[\omega^T g(x_i) + b] = 1 - e_i, i = 1, 2, 3, \dots, N \quad (15)$$

where  $x_i$  and  $y_i$  are  $N$  i/o pairs.

### 2.6.3. Naïve Bayes

The Bayesian classifier is a supervised learning technique and a statistical method for classification. It is one of the classification algorithms that apply density estimation to the data to work out diagnostic and predictive problems. This classifier uses Bayes theorem, and naively presupposes that the predictors are conditionally independent for the given class of data. Despite the fact that the hypothesis is usually desecrated in practice, this classifiers yields posterior distributions

those are stout to biased class density estimates, predominantly where the posterior is 0.5 (the decision boundary) (Hastie, T., R., 2008). The naive Bayes assumption is that all the features are conditionally independent given the class label:

$$P(x|y = c) = \prod_{i=1}^D p(x_i|y = c) \quad (16)$$

These classifiers allocate observations to the most probable class. The algorithm works as:

1. An estimation of the densities of predictors inside each class.
2. Modeling subsequent probabilities according to Bayes theorem, i.e., for all  $k = 1, \dots, K$ ,

$$\hat{p}(Y = k | X_1, \dots, X_p) = \frac{\prod_{j=1}^p \Pi(X_j | Y = k)}{\sum_{k=1}^K \prod_{j=1}^p \Pi(X_j | Y = k)} \quad (17)$$

where:

$Y$  is the random variable equivalent to the class index of a sample set,  $X_1, \dots, X_p$  are the random predictors, and  $\pi(Y=k)$  is the former probability for a class index  $k$ .

3. Classification of the sample space by finding the posterior probability for all classes, and then assigning the sample space to the a class resulting in the highest subsequent probability (Hastie, T., R., 2008).

#### 2.6.4. *k* Nearest Neighbour

kNN is a non-parametric supervised learning algorithm (D. T. Larose, 2004) in which use of class labels is made to stores all available cases and classify new data based on a similarity (distance function) measure. For the new sample data to be tested  $k$  number of training data closest to the test sample are estimated and the class that is most familiar amongst that  $k$  nearest neighbors is allocated as the class to the new test data. In this paper, the  $K$  nearest neighbors have been varied from 2 to 6 and achieved the highest accuracy for  $k = 2$ . The distance was calculated by means of Euclidean distance similarity measure. The kNN are an example of instance based supervised learning in which based on the number of nearest neighbour value  $k$  is used to make the classification order to memorize the training set data. The classifier takes the decision about the class label based on the computation done in the previous step. In this work, the combination of stand deviation and Kraskov entropy features for the first and sixteenth sub band of the TQWT decomposed EEG signals were used to form the feature vector for kNN classifier.

#### 2.6.5. *Random Forest*

A significant enhancement in classification accuracy are resulted from developing an ensemble of trees and allowing the vote for the most relevant class. Random forest is an ensemble tool that takes a subset of features and a subset of class variables to construct a decision trees. A merger of such multiple decision tree is achieved to obtain a more precise and steady classification. The bagging techniques is used to develop the ensembles by repeatedly producing the random vectors to administer the growth of every tree in the group (Breiman, 1996). Another techniques used to build the ensemble is to split the selection randomly from  $k$  best splits (Dietterich, 1998). In the optimization of RFs, a new training sets was produced by randomizing the outputs in the original training set (Breiman, 1999). In several papers on “the random subspace”, a technique was presented that does a random selection of a subset of features to use to develop each tree (Ho, 1998). The RF is a practically suited classification algorithm as:

- RFs are non-parametric and can model arbitrarily complex relations between inputs and outputs, without any a aforementioned assumption;
- RF can classify nonlinear and nonstationary EEG data;

- RF are strong to errors in classes and are easily interpretable.

For the given an assembly of classifiers  $h_1(x), h_2(x), \dots, h_K(x)$ , and by the training set drawn at random from the distribution of the random vector  $Y, X$ , describe the margin function as

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \quad (18)$$

where  $I(\cdot)$  is the indicator function. The margin measures the average number of votes at  $X, Y$  for the right class surpasses the average vote for any new class. For the classification to be promising, margin should be more. The generalization error is as follows:

$$PE^* = P_{X,Y} mg(X, Y) < 0 \quad (19)$$

where the subscripts  $X, Y$  indicate that the probability is over the  $X, Y$  space.

In random forests,  $h_k(X) = h(X, \Theta_k)$ .

### 3. Results and discussions

The automated seizure detection algorithm was developed and tested using two sub sets from Bonn University dataset. The efficacy of the designed algorithms has also been verified using EEG data recorded from a local hospital. The sample normal and epileptic EEG signals of 23.6 seconds duration are shown in Figure 2 and 3 respectively.

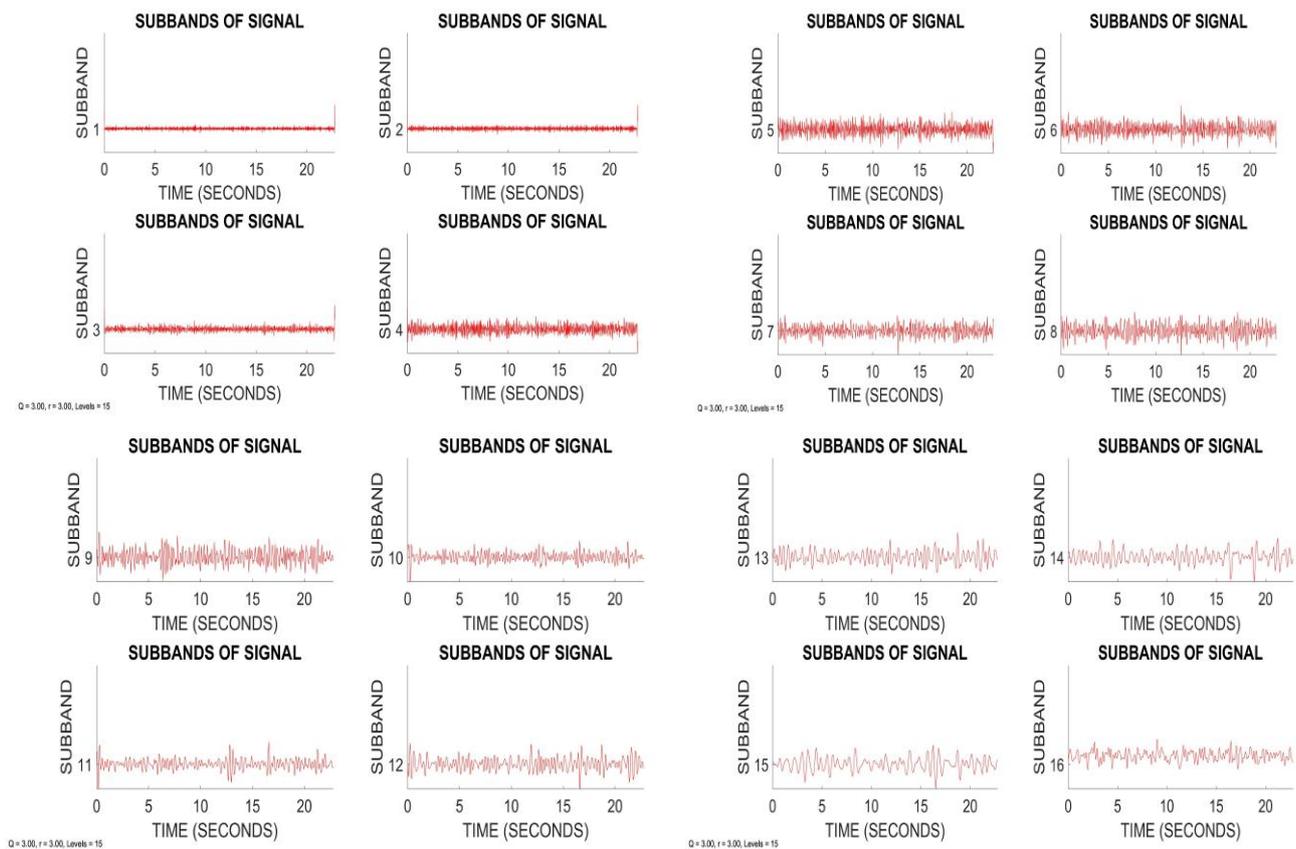


Figure 6. 16 sub bands of sample normal EEG signal decomposed using TQWT

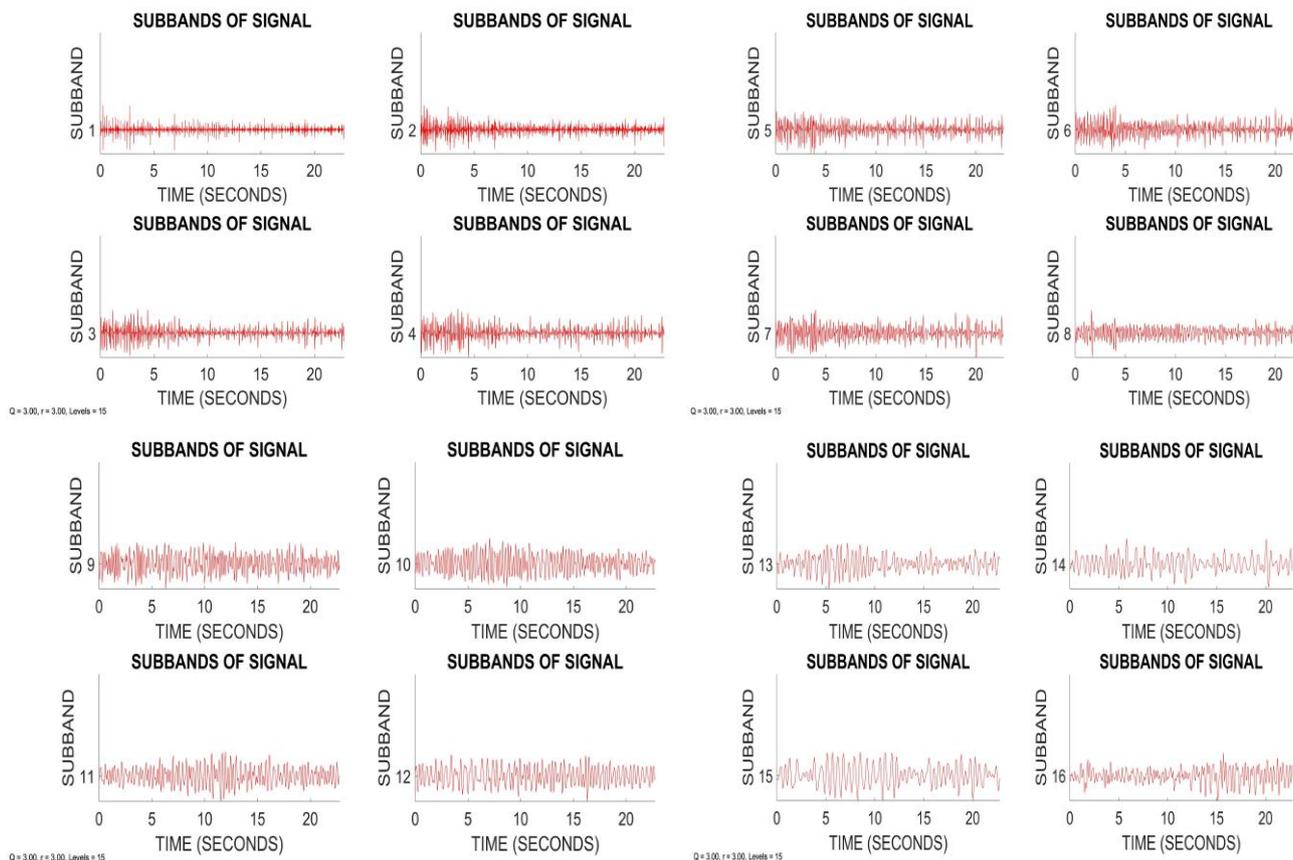


Figure 7. 16 sub bands of sample epileptic EEG signal decomposed using TQWT

The selection of most encouraging value of  $Q$  and  $J$  is a significant step in signal decomposition using TQWT. To select the optimum value of  $Q$  and  $J$ , numbers of experiments were performed by taking only Shannon and Kraskov entropy features as a result of maximum classification accuracy. Further, the values of  $J$  were varied by keeping  $Q = 3$ . It was noted that the highest possible value of  $J$  with  $Q = 1$  is 15. Hence, the value of  $J$  is varied from 3 to 15 and Shannon and Kraskov entropies were computed for each subband. Then, the classification was performed for varied values of  $J$ . The 16 sub bands of normal and ictal EEG data for a sample data are plotted in Figure 6 and Figure 7 respectively.

It can be noticed that at  $J = 15$ , maximum variation in low and high frequency sub bands was obtained which resulted in more relevant entropy parameters and improved classification accuracy. Therefore,  $J = 15$  was selected for additional experiments. After selecting  $J = 15$ , the value of  $Q$  was varied in steps from 1 to 3 and entropies are computed from 16 subbands for each value of  $Q$ .

The plot of frequency response for normal and seizure sample EEG signal is depicted in Figure 8. The increase in  $r$ , while keeping  $Q$  unchanged, has shown the effect of increase in the overlap between adjacent frequency responses. The parameter  $r$  has no effect on the general shape of the wavelet of frequency response as they are governed by  $Q$ . With  $r > 3$ , the number of levels  $J$  need to be increased to cover the same frequency range as a result of the increased overlap. The figures show the frequency responses on log frequency scale with an  $r$  of 3.

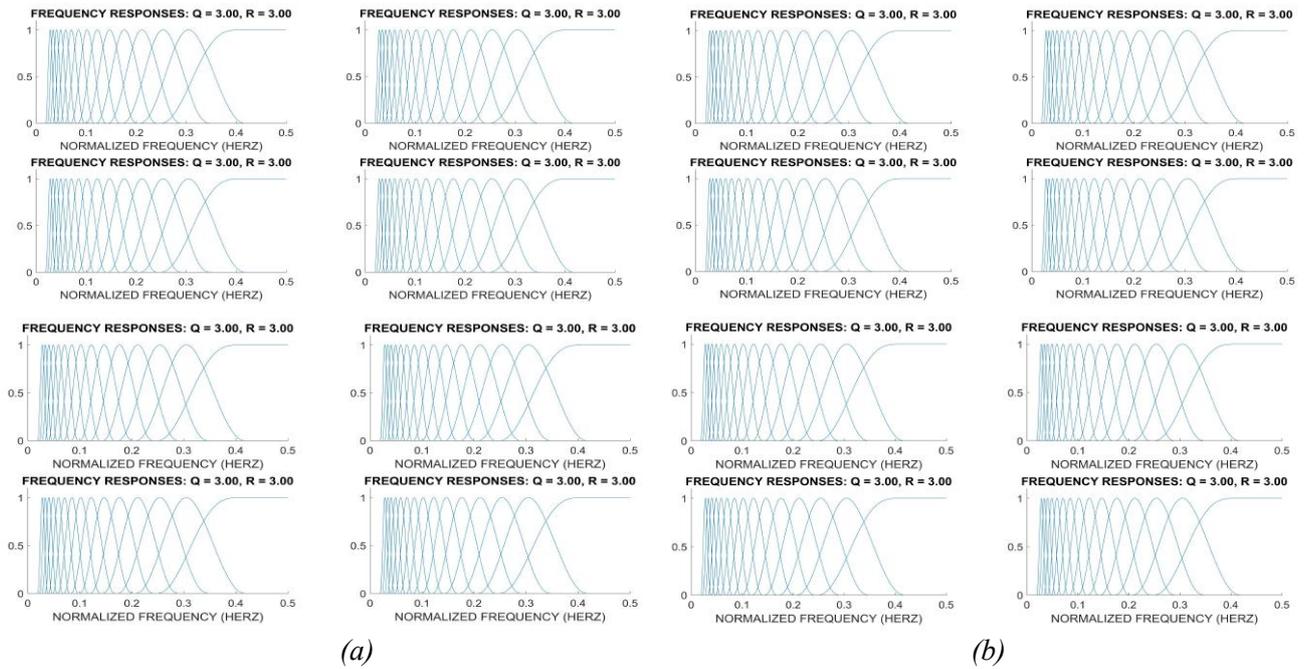


Figure 8. Frequency response of TQWT sub bands for (a)normal, (b)seizure EEG signal sample

The Figure 9 shows wavelets for 3-15 sub bands for  $r = 3$ . The figures display the wavelets and the frequency responses when Q is set to 3.0.

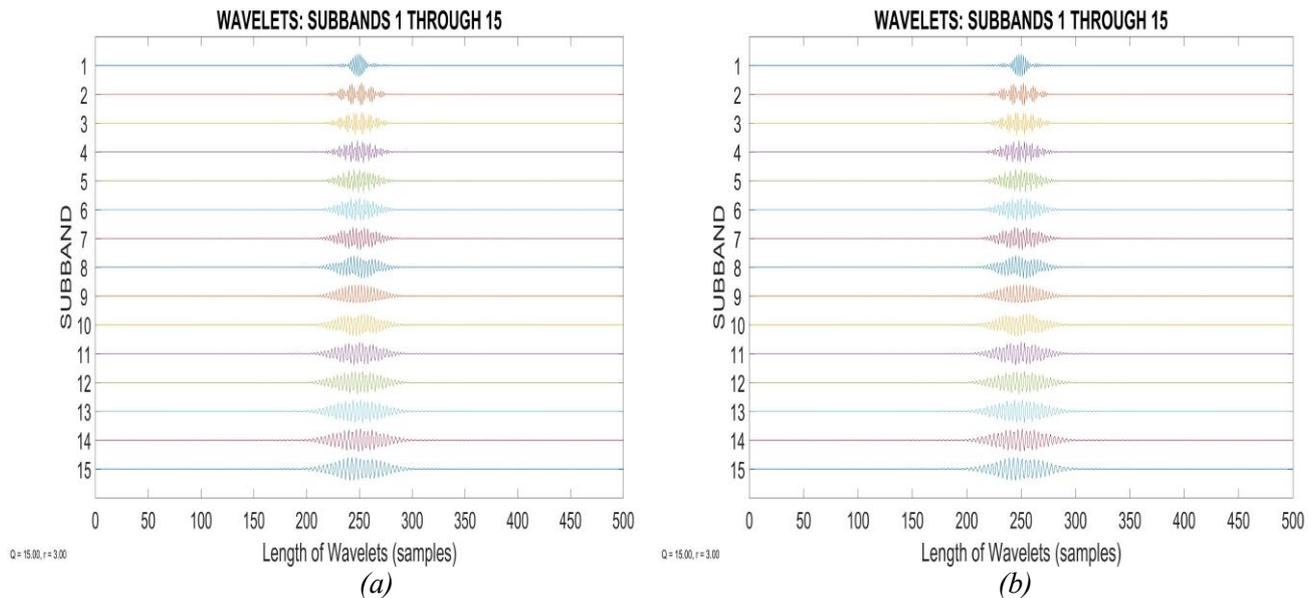


Figure 9. Wavelets for 3-15 sub bands for (a) normal and (b) epileptic EEG signal sample

It is to be noted that the frequency responses are more narrow for  $Q = 3$ , compared to the cases where Q was set to 1.0. With increasing Q from 1.0 to 3.0, more stages are needed to span the same frequency range because each frequency response is narrower. Here, we made use of 16 stages divided into 2 parts for the purpose of display.

The figures show a sample normal and seizure EEG signal decomposed using the TQWT, displays the subbands, and the distribution of energy across subbands. The seizure sample signal has a more oscillatory behavior than the normal signal used in the demonstration. It is interesting to note that out of total 16 stages specified the high-pass subbands (subband 1-8) have negligible energy as compared to the low-pass sub band (sub band-16). Because the first eight subbands have

essential zero energy, it was decided not to compute entropy features for these sub bands for normal signals.

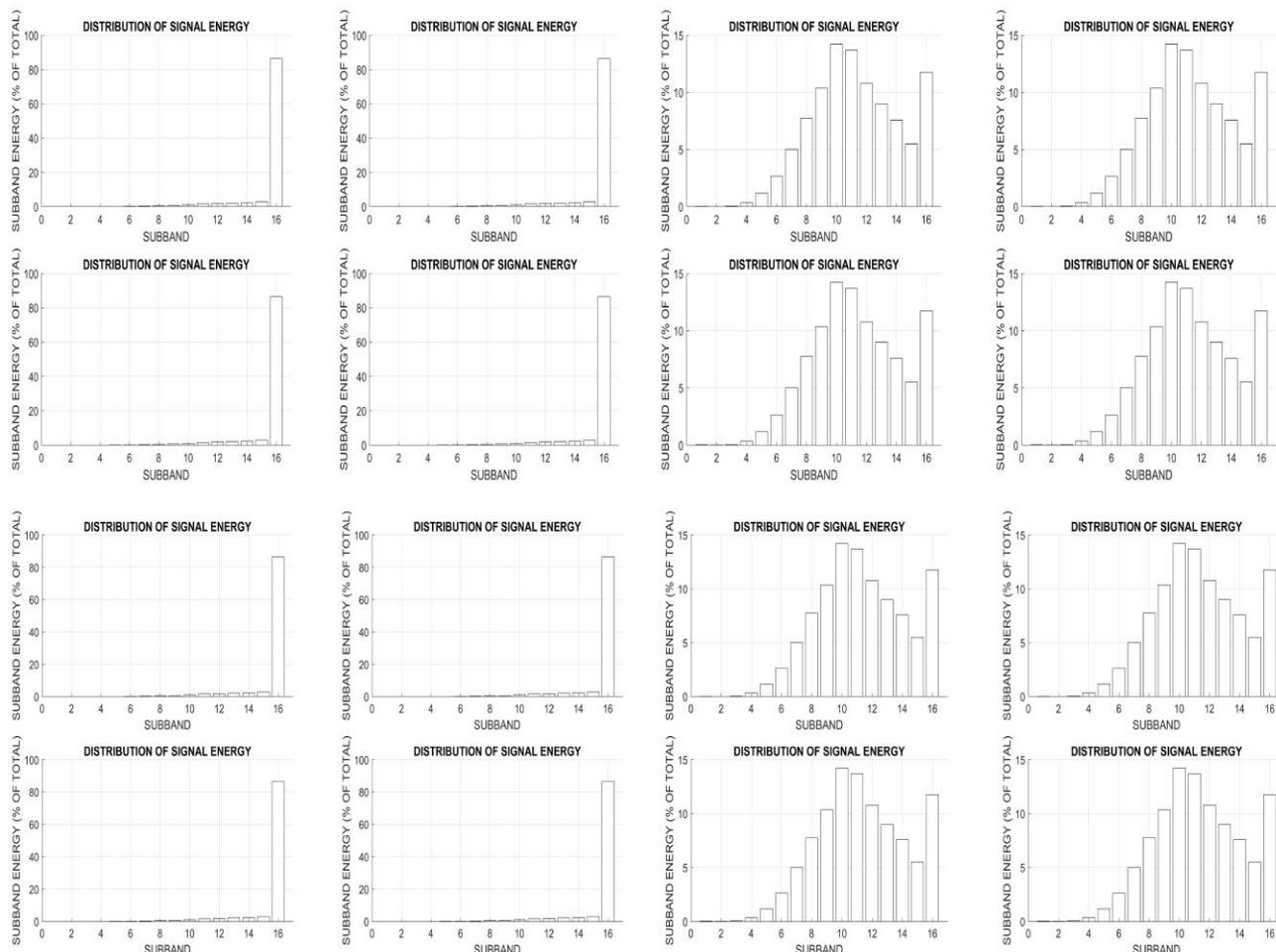


Figure 10. Energy distribution for sub bands(1-4 and 13-16) for (a) normal, (b) Seizure EEG signal sample

The subscript in Subband-1 denotes the first level of subband. The subbands from Subband-1 to Subband-16 are in decreasing order of frequencies. First 15 subbands were reconstructed from the detail coefficients, and 16<sup>th</sup> subband was reconstructed from the approximate coefficients. In this way, the values of  $Q$  and  $J$  were selected for further experiments. The different features were computed from all subbands and experiments were performed to find the best combination of features using various ranking methods and five different classifiers. Further, with  $Q = 3$  and  $J = 15$  different nonlinear features are computed from each decomposed subband. All the computed features are combined to form a feature set of size  $140 \times 4$ .

The typical range of spectral and entropy features obtained from sub band 1-16 are shown in Table 1 and 2 respectively. The  $p$ -value statistical test (T. Dahiru, 2008) was applied to examine the discrimination ability of various features. Apart from spectral features, all other features were found to be significant with less  $p$ -values ( $p < 0.05$ ) indicating their suitability for good discrimination of normal and seizure EEG signals. Further, the features were ranked using Receiver Operating Characteristics (ROC) (Zweig Mark H., 1993).

The classification stage followed the steps where decomposition of TQWT and computation of spectral and entropy features was achieved for both normal and seizure types of EEG signals. After feature computation the next task was to divide features into training and test datasets for the performing classification of signal into two classes. To improve the classification performance and

to provide faster and more cost-effective classifier the selection of the optimal set of both the feature sets is important. The  $p$ -test alone was not suited to describe the efficacy of the features employed in the classification. Hence, the method called wrapper-based feature selection, suggested in (Kohavi Ron, 1997), was employed for feature selection. The performance of five classifiers have been tested. The performance parameters of classification methods, such as specificity (Spec), sensitivity (Sens) and accuracy (Acc), were found by the 10-fold cross-validation process (Van Stralen Karlijn J., et al., 2009). The 10-fold cross-validation process has been used in this work to get unbiased performance of the classifier. Along with the above classifiers the wrapper-based feature selection algorithms have been carried out by the MATLAB machine learning and statistics toolbox (MATLAB, 2016) software on Intel ® Core™ i5-7200 U CPU @ 2.5 GHz with 8 GB RAM.

The parameters of LSSVM and RBF kernel like cost and gamma were set to 1 for the selected number of features. The performance evaluation of the proposed Shannon and Kraskov entropy features has been presented for EEG classification. The distinguishing ability of the five classifiers have been tested using the two classes (normal, and seizure). The Shannon and Kraskov entropies of decomposed EEG signals have been computed and used as a separate feature set as compared to the spectral features set.

Table 1. Spectral Features of the extracted TQWT sub bands ( $Q = 3, R = 3, J = 16$ ).

Scale No.	Mean		Standard Deviation	
	Normal	Seizure	Normal	Seizure
1	9.834927	8329.427	11.56066	71.58755
2	58.77755	11110.04	23.84113	65.25688
3	47.77503	8932.082	20.76151	85.51468
4	10.30816	446.1276	20.08941	114.6693
5	130.0165	1539.636	428.3669	99.01905
6	6.069761	243.3806	24.44858	106.087
7	34.06279	5926.254	16.71747	95.69294
8	33.06874	1857.898	11.7173	76.64777
9	522.3913	2032.91	35.768	71.35347
10	72.81399	28592.28	11.83377	251.9867
11	16.94179	7297.517	6.223891	298.7179
12	13.56822	3044.189	23.99828	116.3269
13	7.512193	7857.243	6.081293	132.1923
14	7.613255	426.0846	13.88737	287.1622
15	6.310997	552.7713	113.0568	282.8694
16	48.03316	86.0241	13.83567	79.17186

In Table 1, the spectral features, namely, mean and standard deviation are projected. It signifies the computed value of SD by the projected technique for normal categories of EEG signals are less significant than expected, because of the absence of abnormalities present in signals.

Table 2. Entropies of the extracted TQWT sub bands ( $Q = 3, R = 3, J = 16$ ).

Scale No.	Shannon Entropy		Kraskov Entropy	
	Normal	Seizure	Normal	Seizure
1	-7832.583117	-2120182.678	1.754068	3.25435
2	-19938.80744	-5836813.985	1.977948	4.173407
3	-21701.92126	-6969011.533	2.047737	3.642304
4	-31714.94771	-153011.0566	2.197772	2.397555
5	-1111243.781	-2075609.65	3.687368	3.24926
6	-133957.6349	-102607.3996	1.663588	1.940963
7	-33054.87133	-4459470.757	2.26557	2.829904

8	-11657.37693	-1497161.665	1.949023	2.905165
9	-39067.18007	-622563.9997	2.312583	3.393844
10	-10312.37088	-5192562.954	1.677248	3.961785
11	-11278.44848	-5773207.133	1.815736	3.598474
12	-26697.77175	-3530204.881	2.095272	3.058711
13	-4380.662994	-5064415.974	1.443809	3.797015
14	-8794.364987	-76324.63628	1.803797	2.19443
15	-310926.8264	-106834.1615	3.172345	2.015378
16	-18999.05462	-15945.06889	2.065524	1.944525

In Table 2, the Shannon and Kraskov entropy values have been presented for normal and seizure signals. The use of MLPNN classifier depends on the number of iterations and the learning rate used for a specific transfer function employed in the design. In this work, the highest classification of normal and seizure signal using MLPNN reported was 92.5% for the Shannon entropy features. In the classification task, the entropy features outperformed the spectral feature. It was inferred from the MLPNN study that the tan-sigmoid and pure linear transfer functions were best possible for the application with backpropagation algorithm for the learning purpose. Two different datasets were tested using MLPNN for the learning rate of varying between 0.1 to 0.4 and the classification tasks were assessed using sensitivity, specificity and accuracy. The simulation results showed that the classification accuracy indirectly varies with the mutual range of entropy features,  $p$ -value, features selected using the wrapper test and Mean Square Error (MSE).

Though, reasonably good classification accuracies were achieved from the five classifiers, the classification accuracy for ictal and non-ictal (S-N) EEG signals using KNN and naïve Bayes classifiers are 86.1% and 84.6%, which are significantly low.

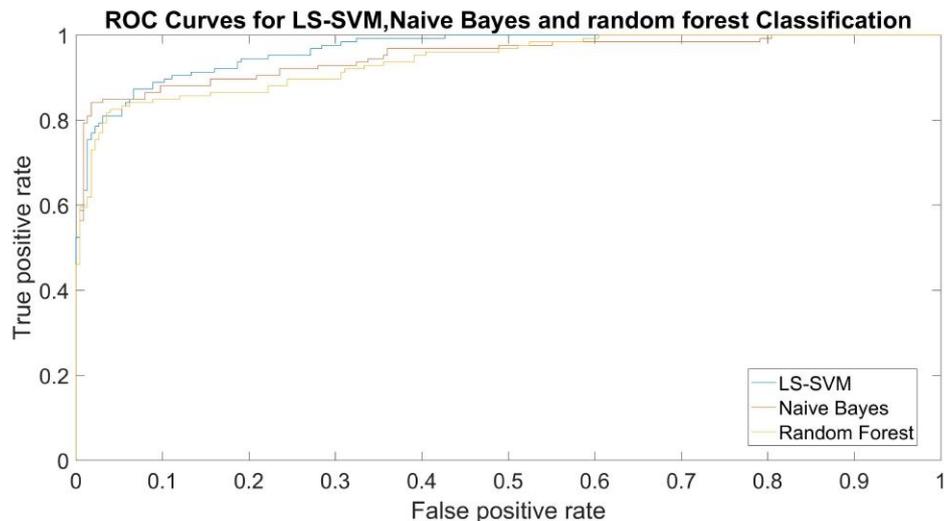


Figure 11. Receiver Operating Characteristics (ROC) graphs for LS-SVM, naïve Bayes and random forest classifiers

As compared to conventional KNN and naïve bayes classifiers, the performance of the proposed random forest classifier with feature selected using wrapper test is very promising. In Table 3, a comparison of classifier performances are presented using proposed features. In this technique, the parameter ( $Q$ ) is varied to obtain enhanced discrimination between two classes. Many decomposition levels of TQWT are set to 8 and 16, respectively, to set the significance of the number of sub-bands in the entropy based classification. The ROC curves for LS-SVM, naïve Bayes and KNN classifiers is shown in Figure 11.

In all the classifiers, the accuracy increases by taking into account  $J = 16$ . For  $Q = 3$ ,  $R = 3$  and  $J = 16$ , major enhancement in classification accuracies have been achieved for random forest

classifier. Many classification tasks, like normal-seizure and seizure-interictal have shown promising results for entropy feature. The maximum accuracy of 97.3% was attained for naïve bayes and kNN classification tasks. The maximum classification accuracy of 97.3% was obtained using random forest classifier. For the normal-seizure classification, the highest classification accuracy obtained was 98% with the Kraskov entropy with decomposition using TQWT for the values specifies as  $Q = 3$ ,  $R = 3$  and  $J = 16$ . The performance of classifier through the combination of Shannon and Kraskov entropy features is also noteworthy.

Table 3. Classification performance of MLPNN, LS-SVM, naïve Bayes, KNN, and random forest classifier

Classifier	Sens(%)	Spec(%)	Acc(%)
MLPNN	91.5	90.5	92.5
LS-SVM	94.5	93.2	95.6
naïve Bayes	85.2	85.3	84.4
KNN	87.3	84.3	86.1
random forest	94.5	95.6	97.3

Table 3 reveals the classification accuracies for the majority of classifiers presented in this paper. The Table 3 presents comparison of the proposed entropy features based classifiers and their classification performance on the EEG database acquired from Dr. D. M. Mohire’s Neurology Research Centre, Kolhapur, INDIA. In most of the cases, the combined entropy features have shown equivalent performance with the optimum algorithms described in the literature. The proposed system may prove useful in the detection of seizures and aid the neurologists to take accurate diagnostic decision pertaining to the epileptic seizure disorders.

#### 4. Conclusion

Manual monitoring of EEG to diagnose epilepsy is a very challenging task with cumbersome work of observing long recordings and decision making through experience. The automated seizure detection is a promising tool for neurologists in making epilepsy diagnosis. In this paper, an automated method based on varied values of  $Q$  is developed that decomposes EEG signal, computes Shannon and Kraskov entropies and detects seizure signal using random forest. The method achieved classification accuracy of 97.3% and sensitivity and selectivity of 94.5% and 95.6% respectively. A literature survey is presented on the current studies that are related to single channel seizure detection. A comparison table showing different seizure detection methods show that most techniques use joint time-frequency domain signal decomposition and entropy features. Several research papers have explored multi-domain features to build robust feature space. It is clear also that the TQWT is a promising trend for seizure detection and prediction that needs further investigation. The separation of normal and seizure event based on the extracted features is achieved by employing state-of-the-art machine learning classifiers. Majority of the studies focus on improving classification accuracy using remotely accessed resources. However, there is increasing demand to implement the algorithms on local embedded system to reduce computational complexity. The main goal behind this review is to explore the field of EEG signal analysis in real time and accept the same to detect epileptic disorders using EEG recordings.

#### References

- Buck, D., Baker, G.A., Jacoby, A., Smith, D.F. & Chadwick, D.W. (1997). Patients’ experiences of injury as a result of epilepsy. *Epilepsia*. 38 (4), 439–444.
- Englander, Jeffrey, et al. Seizures after traumatic brain injury. *Archives of physical medicine and rehabilitation* 95.6 (2014): 1223.
- Acharya, U. & Rajendra, et al. (2013). Automated EEG analysis of epilepsy: a review. *Knowledge-Based Systems*. 45, 147-165.
- Sanei, Saeid, & Jonathon A. Chambers. (2013). EEG signal processing. John Wiley & Sons,.

- Selesnick, Ivan W. Wavelet transform with tunable Q-factor. (2011). *IEEE transactions on signal processing* 59.8, 3560-3575.
- Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P. & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, *Physical Review*. 64, 8 pages.
- Mennes, Maarten, et al. (2010). Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. *Psychophysiology* 47.6 , 1142-1150.
- Klados, Manousos A., et al. REG-ICA: a hybrid methodology combining blind source separation and regression techniques for the rejection of ocular artifacts. *Biomedical Signal Processing and Control* 6.3 (2011): 291-300.
- Mammone, Nadia, Fabio La Foresta, and Francesco Carlo Morabito. (2012). Automatic artifact rejection from multichannel scalp EEG by wavelet ICA. *IEEE Sensors Journal* 12.3, 533-542.
- Li, Mingai, Yan, C. & Jinfu, Y. (2013). Automatic removal of ocular artifact from EEG with DWT and ICA Method. *Applied Mathematics & Information Sciences* 7.2, 809.
- Mammone, Nadia, & Francesco C. Morabito. (2014). Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy* 16.12 6553-6572.
- Peng, Hong, et al. Removal of ocular artifacts in EEG—An improved approach combining DWT and ANC for portable applications. *IEEE journal of biomedical and health informatics* 17.3 (2013): 600-607.
- Zhao, Qinglin, et al. (2014). Automatic identification and removal of ocular artifacts in EEG—improved adaptive predictor filtering for portable applications. *IEEE transactions on nanobioscience* 13.2, 109-117.
- Torse, D. A. & Veena V. D., (2016). Design of adaptive EEG preprocessing algorithm for neurofeedback system. *Communication and Signal Processing (ICCSP) International Conference on IEEE*. 392-395.
- Wang, Gang, et al. (2016). The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition. *IEEE journal of biomedical and health informatics* 20.5, 1301-1308.
- Güler, Inan, & Elif Derya Übeyli. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of neuroscience methods* 148.2 113-121.
- Jahankhani, Pari, Vassilis Kodogiannis, and Kenneth Revett. (2006). EEG signal classification using wavelet feature extraction and neural networks. *Modern Computing, 2006. JVA'06. IEEE John Vincent Atanasoff International Symposium on. IEEE*.
- Ghosh-Dastidar, Samanwoy, and Hojjat Adeli. (2007). Improved spiking neural networks for EEG classification and epilepsy and seizure detection. *Integrated Computer-Aided Engineering* 14.3, 187-212.
- Subasi, Abdulhamit, and M. Ismail Gursoy. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications* 37.12, 8659-8666.
- Ghosh-Dastidar, Samanwoy, Hojjat Adeli, and Nahid Dadmehr. (2008). Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering* 55.2, 512-518.
- Faust, Oliver, et al. (2010). Automatic identification of epileptic and background EEG signals using frequency domain parameters. *International journal of neural systems* 20.02, 159-176.
- Gandhi, Tapan, Bijay Ketan Panigrahi, and Sneha Anand. (2011), A comparative study of wavelet families for EEG signal classification. *Neurocomputing* 74.17, 3051-3057.
- Oweis, R.J. & Abdulhey, E.W. (2010). Seizure classification in EEG signals utilizing Hilbert-Huang transform. *Biomed. Eng. Online* 10, 38

- Işik, Hakan, and Esmâ Sezer. (2012). Diagnosis of epilepsy from electroencephalography signals using multilayer perceptron and Elman artificial neural networks and wavelet transform. *Journal of medical systems* 36.1, 1-13.
- Omerhodzic, Ibrahim, et al. (2013). Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. *arXiv preprint arXiv:1307.7897*
- Kumar, Yatindra, M. L. Dewal, and R. S. Anand. Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine. *Neurocomputing* 133 (2014): 271-279.
- Faust, Oliver, et al. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure* 26, 56-64.
- Song, Jiang-Ling, Wenfeng Hu, and Rui Zhang. (2016). Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine. *Neurocomputing* 175, 383-391.
- Zhang, Yong, et al. (2015). Comparison of classification methods on EEG signals based on wavelet packet decomposition. *Neural Computing and Applications* 26.5, 1217-1225.
- Raghu, S., N. Sriraam, and G. Pradeep Kumar. (2015). Effect of wavelet packet log energy entropy on electroencephalogram (EEG) signals. *International Journal of Biomedical and Clinical Engineering (IJBCE)* 4.1, 32-43.
- Hassan, Ahnaf Rashik, Siuly Siuly, and Yanchun Zhang. "Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating." *Computer methods and programs in biomedicine* 137 (2016): 247-259.
- Patidar, Shivnarayan, and Ram Bilas Pachori. "Classification of cardiac sound signals using constrained tunable-Q wavelet transform." *Expert Systems with Applications* 41.16 (2014): 7161-7170.
- Bhattacharyya, Abhijit, et al. "Tunable-Q Wavelet Transform Based Multiscale Entropy Measure for Automated Classification of Epileptic EEG Signals." *Applied Sciences* 7.4 (2017): 385.
- Patidar, Shivnarayan, and Trilochan Panigrahi. (2017). Detection of epileptic seizure using Kraskov entropy applied on tunable-Q wavelet transform of EEG signals." *Biomedical Signal Processing and Control* 34, 74-80.
- Kumar, S. Pravin, et al. (2010). Entropies based detection of epileptic seizures with artificial neural network classifiers. *Expert Systems with Applications* 37.4, 3284-3291.
- Phinyomark, Angkoon, et al. (2012). The usefulness of mean and median frequencies in electromyography analysis. *Computational intelligence in electromyography analysis-A perspective on current applications and future challenges*.
- Du, Sijiang, & Marko Vuskovic., (2004). Temporal vs. spectral approach to feature extraction from prehensile EMG signals. *Information Reuse and Integration, 2004. IRI 2004. Proceedings of the 2004 IEEE International Conference*.
- Shannon, Claude E. (1948). A note on the concept of entropy. *Bell System Tech. J* 27.3, 379-423.
- K.A. Veselkov, et al., A(2010). Metabolic entropy approach for measurements of systemic metabolic disruptions in patho-physiological states, *J. Proteome Res.* 9, 3537–3544.
- Kraskov, A. et al., *Estimating Mutual Information, 2008* [Online]. Available: [arxiv.org/pdf/cond-mat/0305641](https://arxiv.org/pdf/cond-mat/0305641).
- Bishop, Christopher M. *Pattern recognition and machine learning*. Springer, 2006.
- Haykin, Simon S. *Neural networks: a comprehensive foundation*. Tsinghua University Press, 2001.
- Suykens, Johan AK, Tony Van Gestel, & Jos De Brabanter. (2002), *Least squares support vector machines*. World Scientific,.
- Hastie, T., R. Tibshirani, and J. Friedman. (2008). *The Elements of Statistical Learning, Second Edition*. NY: Springer,.
- D. T. Larose, (2004a). *Discovering Knowledge in Data: An introduction to data mining*, New Jersey, USA: Wiley Interscienceb.
- Breiman, Leo. (1996). Bagging predictors. *Machine learning* 24.2, 123-140.

- Dietterich, Thomas G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 10.7, 1895-1923.
- Breiman, Leo. (1999). Random forests. UC Berkeley TR567.
- Ho, Tin Kam. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence* 20.8 , 832-844.
- Dahiru, Tukur. (2008). P-value, a true test of statistical significance? A cautionary note. *Annals of Ibadan postgraduate medicine* 6.1, 21-26.
- Zweig, Mark H., & Gregory Campbell. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry* 39.4, 561-577.
- Kohavi, Ron, & George H. John. (1997). Wrappers for feature subset selection. *Artificial intelligence* 97.1-2, 273-324.
- Van Stralen, Karlijn J., et al. Diagnostic methods I: sensitivity, specificity, and other measures of accuracy. *Kidney international* 75.12 (2009): 1257-1263.
- MATLAB and machine learning and Statistics Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, United States.



**Dattaprasad Torse** (b. April 12, 1979) received his B.E. in Electronics and Telecommunication Engineering (2001), M.E. in Digital Electronics (2005), pursuing PhD in Electronics and Communication from Visvesvaraya Technological University of Belagavi, Karnataka, India. Now, he is Assistant Professor of Electronics and Communication Department of KLS Gogte Institute of Technology, Belagavi, India. His current research interests include different aspects of EEG Signal Processing, EEG analysis and Machine Learning. He has (co-) more than 10 papers, more than 10 conferences participation. He is a member of IEEE and life member of Indian Society for Technical Education (ISTE).



**Veena Desai** (b. August 17, 1969) received her B.E. in Electronics and Communication Engineering (2001), M.Tech in Computer Networking (2005) and PhD (2012) in Electronics and Communication from Visvesvaraya Technological University of Belagavi, Karnataka, India. Now, she is Professor of Electronics and Communication Department of KLS Gogte Institute of Technology, Belagavi, India. Her current research interests include different aspects of Cryptography and network security and machine learning. She has authored more than 30 papers, more than 10 conferences participation. She is a member of IEEE and life member of Indian Society for Technical Education (ISTE).



**Rajashri Khanai** (b. October 15, 1979) received her B.E. in Electronics and Communication Engineering (2000), M.Tech in Digital Communication and Networking (2007) and PhD (2015) in Electronics and Communication from Visvesvaraya Technological University of Belagavi, Karnataka, India. Now, she is Professor of Electronics and Communication Department of KLE's Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, India. Her current research interests include different aspects of error correction coding for wireless communication, biomedical signal processing and machine learning. She has authored more than 20 papers, more than 10 conferences participation. She is a member of IEEE.