An optimized design of seizure detection system using joint feature extraction of multichannel EEG signals

Dattaprasad Torse\textsuperscript{1,\ast}, Veena Desai\textsuperscript{1}, Rajashri Khanai\textsuperscript{2}

\textsuperscript{1}Department of Electronics and Communication Engineering, KLS Gogte Institute of Technology, Belagavi, Karnataka 590008, India; \textsuperscript{2}Department of Electronics and Communication Engineering, KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Belagavi, Karnataka 590008, India.

Abstract

The detection of seizure onset and events using electroencephalogram (EEG) signals are important tasks in epilepsy research. The literature available on seizure detection has discussed the implementation of advanced signal processing algorithms using tools accessed over the cloud. However, seizure monitoring application needs near sensor processing due to privacy and latency issues. In this paper, a real time seizure detection system has been implemented using an embedded system. The proposed system is based on ensemble empirical mode decomposition (EEMD) and tunable-Q wavelet transform (TQWT) algorithms. The analysis and classification of non-stationary EEG signals require the wavelet transform with high $Q$-factor. However, direct use of TQWT increases the computational complexity of feature extraction from multivariate EEG signals. In this paper, the first step is to process the signal by using EEMD to obtain 8 intrinsic mode functions (IMFs). The Kraskov (KraEn), sample (SampEn), and permutation (PermEn) entropy features of IMFs are extracted and based on optimum values, and 4 IMFs are decomposed using TQWT. Secondly, centered correntropy (CenCorrEn) features of the 1\textsuperscript{st} and 16\textsuperscript{th} sub-band of TQWT have been used as classifier inputs. The performance of multilayer perceptron neural networks (MLPNN), least squares support vector machine (LSSVM), and random forest (RF) classifiers has been tested on the multichannel EEG data recorded from a local hospital. The RF classifier has produced the highest accuracy of 96.2\% in classifying the signals. The proposed scheme has been employed in developing an embedded seizure detection system to assist neurologists in making seizure diagnostic decisions.

Keywords: seizure detection, electroencephalogram, ensemble empirical mode decomposition, intrinsic mode function, tunable-Q wavelet transform

Introduction

Epilepsy is caused by irregular changes in neural activity\cite{1}. The recurrent impulsive seizure from intricate processes is related to numerous neurotransmitters in the cholinergic, glutamatergic and GABAergic system\cite{2}. As the World Health Organization reports\cite{3}, out of 450 million people reported no conflict of interests.

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.
worldwide with neurological or behavioral disorders, 65 million are affected by epilepsy with nearly 7 million cases in India[5]. The medical diagnosis of epileptic disorders is made by medical professionals using EEG signals recorded from the patient's scalp[5]. The two major types of EEG signals are focal and non-focal in generalized epilepsy. The focal EEG signals may be categorized as epileptic or non-epileptic. Epileptic spikes are the interictal indication of a patient with involvement of limited area of the brain. Non-epileptic deformities are characterized by variations in normal signals or by the manifestation of abnormal signals. While these two types often coexist, separating them is conceptually useful and helps in diagnosis of focal epilepsy[6]. The major treatment option for epilepsy is anti-epileptic drugs (AEDs). The best possible AEDs fail to control seizures in practically one-third of epileptic patients and seizures persist for lifetime. The design and development of automated seizure detection systems may have the potential to improve the diagnostic accuracy when compared to manual observations of EEG signals. The signal feature computation and then classification are major steps in computerized seizure detection[7]. EEG signals are a powerful tool in different clinical applications for diagnosis of brain disorders. The computational complexity of signal processing algorithms is a prime factor in real time seizure detection systems[8].

Recently, several automated EEG classification algorithms based on empirical mode decomposition (EMD) and tunable-Q wavelet transform (TQWT) are proposed on single-channel univariate signals for 2-class and 3-class problems. Hassan AR et al have proposed an automated epilepsy diagnosis scheme based on TQWT and bootstrap aggregating, which reveals applicability of spectral features in the TQWT domain in combination with bagging[9]. Based on statistical moment based features extracted using EMD, sleep scoring on single channel EEG classification using adaptive boosting and decision trees has been presented by Hassan AR et al[10].

A study on use of normal inverse Gaussian (NIG) pdf modeling of TQWT sub-bands in sleep stage classification has been presented by Hassan AR et al[11]. An automated sleep staging method has been developed by Hassan AR et al by using TQWT and different spectral features obtained from TQWT sub-bands. The classification has been performed by random forest (RF) classifier[12]. Hassan AR et al have shown the efficacy of automated EEG signal processing systems using TQWT and EMD[13–16].

As in the above literature, the EEG signals are classified using statistical and entropy features. The decomposition of EEG signal and computation of features is done separately in two steps, using either ensemble EMD (EEMD) or TQWT. The single stage decomposition used earlier shows increased computational complexity of the classifiers due to the inclusion of single level decomposition and features. The main advantage of EEMD is in identifying statistical characteristics of white noise when deciding the final number of IMFs. TQWT has an advantage of operating on small/no oscillatory signals with a low Q-factor, but not on high oscillatory signals with a relatively high Q-factor. TQWT handles the issue by adjusting its Q-factor and hence may be used as an effective algorithm for EEG signal analysis. In this paper, advantages of EEMD and TQWT are combined to extract optimum features in two stages to enhance the classification accuracy in a real time seizure detection system.

The proposed joint EEMD-TQWT is shown in Fig. 1. In the first step, the EEG signals from normal and seizure classes are decomposed using EEMD. Since the maximum oscillatory information about the signals lies in the first few IMFs, four initial IMFs are selected for decomposition using TQWT based on optimum values of Kraskov entropy (KraEn), permutation entropy (PermEn), and sample entropy (SampEn). The sub-bands of the signals are obtained by choosing low-Q values for IMFs with low oscillatory behavior, and high-Q values for more oscillatory IMFs. In the second step, experimental results show that the centered correntropy (CenCorrEn) features computed only for the 1st and 16th sub-bands of the selected IMFs results in the highest classification accuracy. The first sub-band of TQWT represents high-pass sub-band and the last low-pass sub-band. Since we have considered 4 IMFs from the EEMD step, a total of 16 sub-bands are produced by the TQWT step. The features from these sub-bands represent oscillatory behavior of decomposed signals. Moreover, with the higher order of sub-band, a reasonably good resolution can be attained concurrently in high and low frequency region of the input signals. The selection of optimum features from this set is done using the wrapper method of feature selection[17]. Finally, a comparison of the classification performance is presented for LS-SVM and RF classifiers.

Materials and methods

Electroencephalogram (EEG) dataset

This research work used two datasets, namely
public and local. The public dataset (http://www.dtic.upf.edu/~ralph/sc/) contains multichannel intracranial EEG records of 5 patients who were suffering from pharmacoresistant temporal lobe epilepsy. Each EEG signal was sampled with 512 Hz sampling frequency and contained 10 240 samples. The band-pass filter of frequency range between 0.5 to 150 Hz has been used to filter noises [18]. In this dataset, 3 750 files were belonging to each class, focal and non-focal and 80% of the dataset have been used to train the machine learning algorithm and test its performance of local dataset. The local dataset contained the EEG data recorded from 10 subjects. The feature extraction and classification algorithms always performed exceedingly well for the online datasets. However, the performance of the same system considerably deteriorated when tested on the real time data recorded from patients. The reason for the adverse performance is unbalanced data resulting from the recording and sampling process in real time. In this work, all four entropy values were computed for multiple channels (19-channels) related to both EEG datasets for detection of seizure activity. The dataset contained 5 healthy subjects' EEG (3 male and 2 female), few with history of chronic headache and none with symptoms of epileptic seizures, and 5 epileptic subjects (3 male and 2 female), aged between 10 to 35 years. The extracranial recording was done with eyes open and closed using standard 10-20 electrode placement. Out of 5 epileptic subjects, only one child subject of age 12 years was recorded with drug induced sleep EEG. The recording was done for 10-to-15-minutes sessions. The expert neurologist was consulted to read the EEG data manually and decide epileptic seizures identified from the specific channel of the recording system. The system used was Recorders & Medicare Systems (RMS) medical equipment data acquisition 10-20 standard, which enabled recording of EEG from 19 active unipolar channels (FP1, FP2, F7, F3, Fz, F4, F8, T3, C3, CZ, C4, T4, T5, P3, PZ, P4, T6, O1, and O2) with A1 and A2 as a reference connected to total 21 electrodes placed on the subject's left and right head. To maintain uniformity in data samples taken from public and local dataset, selection of 19 channels was done based on comparison with public dataset. An electrode fixed in the middle of distance between FP1 and FP2 on the subject's forehead was the ground. For this pilot research work, time series of approximately 23 seconds (4 000 samples) from the EEG recording was selected for 5 healthy and 5 epileptic patients under the expert observation of the neurologist as shown in Fig. 2. The total number of samples was 176/second and hence the EEG data was sampled at 176 Hz. From the recording of original datasets and manual observation, only artifact-free EEG signals were segmented from large datasets.

**Ensemble empirical mode decomposition (EEMD) and tunable-Q wavelet transform (TQWT)**

In time-scale domain, the sparsity of IMFs obtained from EEMD can be best represented by decomposing them using TQWT. The novel technique of application of TQWT to the IMFs to characterize the oscillatory behavior of EEG was used on F and NF signals. The assessment and selection of IMFs were based on the optimum values of Kraskov, sample and permutation entropies.

**Ensemble empirical mode decomposition (EEMD)**

EMD is an adaptive joint time-frequency data
The methods work on the principle of decomposing the EEG signal into a finite set of oscillatory elements called intrinsic mode functions (IMFs). The IMFs are considered as zero-mean, amplitude and frequency modulated (AM-FM) components. To obtain the correct form of IMF, it should satisfy two conditions\(^{(18)}\): (1) In the complete dataset, the total number of extrema values and the number of zero-crossings have to be either equal or vary at most by one; (2) The mean of two envelopes, one defined by connecting local maxima and the other by local minima, at any point should be zero.

In EMD the original signal \(x(t)\) is expressed as

\[
x(t) = \sum_j c_j(t) + r(t) \quad (1)
\]

\[
c_j(t) = \text{Re}\left\{a_j(t) \exp\left[i f_j(t)\right]\right\} = \text{Re}\left\{a_j(t) \exp\left[i \int_{-\infty}^{t} \omega_j(t') dt'\right]\right\} \quad (2)
\]

where \(c_j(t)\) indicates IMFs and \(r(t)\) is the residual non-oscillating signal. The \(a_j(t)\) denotes a time-dependent amplitude, \(\int_{-\infty}^{t} \omega_j(t') dt'\) represents a time-dependent phase.

Although the data adaptive EMD is a powerful method to decompose nonlinear signal, it has the disadvantage of the frequent emergence of mode integration. This indicates a mode mixing problem, where a single IMF either consists of signals of widely dissimilar scales, or a signal of a scale alike existing in unlike IMFs. An improved technique of a noise-assisted data analysis (NADA), known as the EEMD, was suggested to eliminate this drawback\(^{(19)}\).

In EEMD, an ensemble of data sets was generated by adding uniformly distributed white noise with finite amplitude \('a'\) to the original signal \(x(t)\). The EMD analysis was then used to each of the data series of the ensemble. Finally, the IMFs were obtained by averaging the individual components in every realization over the ensemble. In EEMD algorithm various parameters such as ensemble number \(n_e\), number of siftings \(n_s\), amplitude of white noise \(a\), number of IMFs \(n_{\text{IMF}}\) and IMF \(c_m\) were defined based on the time complexity of the signal being decomposed. The algorithm was initialized by setting the IMF value \([c_m(t)]\) equal to zero. The resulting signal decomposed into IMFs and the trial loop was executed for the maximum value of \(n_s\). Finally, the ensemble mean of the computed IMFs from the decompositions was given by\(^{(20)}\)

\[
a_f = a/\sqrt{n_e} \quad (3)
\]

In this work, decomposition of bivariate EEG signal was done by adding optimum values of white noise of amplitude 0.4 and number of ensembles was equal to 10 based on the experimentation performed on both datasets. The IMFs extracted from sample non-focal and focal EEG channel difference signals using EEMD technique were depicted in Fig. 3–4.

**Tunable-Q wavelet transform (TQWT)**

The tunable \(Q\)-factor is a prime quality of TQWT that makes it most suited for transform in analysis of IMFs with varying oscillatory nature. Ideally, the ratio of center frequency of an oscillatory signal to its bandwidth, defined as the \(Q\)-factor of a wavelet transform should be selected based on the oscillatory frequency components in the signals\(^{(14)}\). However, fixed value of \(Q\)-factor poses limitation on decomposition of nonlinear IMFs. The analysis of high oscillatory IMFs should be done using large value of \(Q\) whereas less oscillatory IMFs need smaller values of \(Q\). The TQWT is bound by the \(Q\)-factor and

![Fig. 2 Amplitude plot and frequency maps of normal EEG dataset of subject for FP1 to O2 channels.](image-url)
its redundancy. In decomposing selected IMFs, the TQWT parameters defined are the $Q$-factor ($Q$), total over-sampling ($r$), and the number of decomposition phases termed as ($j$). The number of oscillations of the wavelet is determined by the $Q$ parameter, which has the property of tuning in terms of $\alpha$ and $\beta$.

The parameters $r$ and $Q$-factor for $j$ decomposition levels is defined as

$$r = \frac{\beta}{1 - \alpha}; \quad Q = \frac{\omega_c}{BW} = \frac{2 - \beta}{\beta}$$

The term $BW$ is bandwidth of the frequency response resulting in $j$ sub-bands.

The selection of $Q$-factor and 'r' depend on the number of levels of decomposition to cover the IMF's frequency range. The filter bank parameters $\alpha$ and $\beta$ are selected to achieve the desired values. For a constant value of $Q$, as 'r' increases, overlapping between neighboring frequency responses increase. TQWT filters are easier to implement since the construction of these filters are based on non-rational transfer functions.

Fig. 3 Intrinsic mode functions from difference of sample 2-channel non-focal EEG signal

Fig. 4 Intrinsic mode functions from difference of sample 2-channel focal EEG signal
The multi-stage TQWT decomposition of IMFs can be achieved by repeatable connection of 2-band filter banks to its low pass channel. A ‘j-level’ TQWT is depicted in Fig. 5.

At each level of decomposition, the input signal $x(n)$ is converted into low-pass and high-pass signal components. For the input signal sampling rate $f_s$, the resulting low-pass and high-pass sampling frequencies are $\alpha f_s$ and $\beta f_s$, where $\alpha$ and $\beta$ are the same scaling constants for all levels of decompositions. The low-pass and high-pass sub-bands are computed using their respective frequency responses and scaling factors. The low-pass and high-pass frequency response factors are determined as shown in equations$^{[14]}$

$$F_l(\omega) = \begin{cases} 1, & |\omega| < (1-\beta)\pi \\ 0, & \alpha\pi \leq |\omega| \leq \pi \\ 0, & |\omega| < (1-\beta)\pi \\ 1, & \alpha\pi \leq |\omega| \leq \pi \end{cases}$$

where $\theta(\omega)$ is expressed as$^{[13]}$

$$\theta(\omega) = 1/2 (1 +\cos (\omega)) \sqrt{\cos (\omega)}, \quad |\omega| \leq \pi$$

Finally, the reconstruction of original input signal is accomplished by using the same filter banks used in the decomposition stage.

**Entropy based feature estimation**

Normally, classification of seizure and non-seizure EEG signals is achieved using a set of diagnostic features extracted from the signal components. The optimum feature sets with intra-class identical characteristics and inter-class fluctuations are well suited for automated systems. However, proper selection of feature sets and their discrimination ability depends on signal decomposition stage. The entropy features, developed by Shannon are a measure of spread of the data. The entropy is used to quantify unpredictability within the signal and to measure complexity of nonlinear signals. Acharya, U. Rajendra, et al. have proposed a review on applications of entropies in automated diagnosis of epilepsy$^{[21]}$.

**Krauskov entropy (KraEn)**

The entropy measure is useful in analyzing nonstationary and nonlinear EEG signals. The Shannon’s entropy for a data source $S$ is defined using

$$H_k(S) = -\int f(S) \log [f(s)] ds$$

where $f(S)$ is probability density function (pdf). The differential statistical entropy of nonlinear signals known as the Shannon entropy is measured using KraEn. In this work, KraEn is computed for the EEG signals using the “k-nearest neighbor” sample. By default the Euclidean distance has been used for estimating $H_k(S)$ from EEG signals of ‘n’ samples of signal ‘S’. Equation 8 is considered as an average of log $\mu(S_i)$. For unbiased estimator equation 8 is defined as

$$\hat{H}_k(S) = -\frac{1}{N} \sum_{i=1}^{N} [\log \mu(S_i)]$$

To obtain the estimate (log $\mu(S_i)$), we consider the probability distribution $P_i(e)$ for the distance between $S_i$ and its $k^{th}$ nearest neighbor. For Euclidian norm, the final entropy is given by

$$\hat{H}_k(S) = -\Psi(k) + \psi(N) + \log C_d + \frac{d}{N} \sum_{i=1}^{N} \log \epsilon(i)$$

where $d$ is the length of source $S$, $C_d$ is volume of the $d$, $\psi(N)-\psi(k)$ is the expected value of the data source, is twice the distance from $S_i$ to its $k^{th}$ neighbor$^{[22]}$.

**Permutation entropy (PermEn)**

By combining the concept of entropy and symbolic dynamics, Bandt and Pompe introduced a new concept of PermEn as a measure of complexity of nonlinear signals by$^{[23]}$

$$H_p(S) = -\sum_{i=1}^{n!} p'_i \log(p'_i)$$

where $p'_i$ indicates the relative frequency permutations.

The PermEn per symbol is as given by

$$h_p(S) = -\frac{1}{(n-1)} \sum_{i=1}^{n!} p'_i \log(p'_i)$$

where $0 \leq h_p(S) \leq 1$.

The PermEn is computed with permutation order ‘n’ using a method which considers the first occurrence order$^{[23]}$.

**Sample entropy (SampEn)**

SampEn devised by Richman and Moorman$^{[24]}$, is
defined as measure of the unpredictability of the \( i.e. \) how much a given data point depends on the values of \( k \) of previous data points, averaged over the entire signal range. It is also a measure of data regularity like approximate entropy. SampEn is mostly independent of record length. The detection irregularity in the EEG data is associated to larger values of SampEn.

The SampEn is computed by finding points that are matching within the tolerance \( \varepsilon' \) for all match points, whereas the number of template matches are stored in counters \( A(dim) \) and \( B(dim) \) for all lengths \( dim \) up to \( \varepsilon' \). SampEn is defined by

\[
SampEn(dim;r;N) = \ln \left[ \frac{A(dim)}{B(dim-1)} \right] \quad (13)
\]

where \( B(0) = N \), the length of the input series and \( dim \) is the embedding dimension and \( r \) is the tolerance value. The false nearest neighbor has been used in this work to extract SampEn. With the appropriate \( k \) value the false neighbor value reaches zero. In this work, \( k=2 \) was appropriate value for SampEn computation of IMFs\(^{(24)}\).

**Centered correntropy (CenCorrEn)**

In a specific window controlled by the kernel dimension, correntropy straightly specifies the probability density of closeness of two random variables\(^{(25)}\). Cross-correlation is the similarity between two different signals in time domain and its entropy known as correntropy. It is a similarity measure of multichannel signals resulting in nonlinearity of a feature space. The correntropy is defined by

\[
C(r,t) = E[f(x_r), f(x_t)] \quad (14)
\]

where \( \phi \) maps nonlinearity between samples from the input to the feature space. Correntropy uses the ‘kernel trick’ that gives the inner product of nonlinear combinations called as Mercer kernel and is defined by

\[
K(x_r, x_t) = [f(x_r), f(x_t)] \quad (15)
\]

This kernel-trick method results in relevant feature space without the precise knowledge of nonlinear mapping and hence reduces the expected value of the kernel of a particular choice. The most commonly used Mercer kernel is in the form of Gaussian kernel denoted by

\[
K(x_r, x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{|x_r-x_t|^2}{2\sigma^2}} \quad (16)
\]

where \( \sigma \) is the size of the kernel.

The Fourier transform of correntropy has been introduced as the general power spectral density and defined as correntropy spectral density (CSD)\(^{(25)}\). The CSD is defined by

\[
P_v(\omega) = \sum_{m=-\infty}^{\infty} V[m] e^{-jm\omega} \quad (17)
\]

**Classification techniques**

In machine learning and statistics, classification deals with identifying a set of class to which a data sample belongs in the sample database. In this work, EEG signal is classified into seizure and non-seizure class. An illustration of supervised learning has considered where training data set of correctly classified samples is available for assessing performance of three classifiers.

**Multi layer perceptron neural networks (MLPNN) classifier**

Multilayer feed-forward neural networks (MFNN)\(^{(26)}\) designed for the problem uses multiple neurons interconnected in basic format of input layer, multiple hidden layers, and an output layer. The basic functionality of multilayer neural networks is the same as that of a traditional one-hidden-layer neural network\(^{(26)}\). In the proposed study, four features were used to train the MLPNN and its performance was evaluated. The training and testing the MLPNN algorithm begins with configuring the neural network’s various parameters such as hidden neurons, transfer functions and training algorithms with initialization of the network weights bias. The structure with 10 hidden-neurons was decided with "tansig" function as input-hidden transfer function. An optimal design value was chosen for hidden output transfer function. Four different supervised training functions were tested to decide the optimum performance. The performance function selected were MSE, the number of epochs were multivariate recordings (750), performance goal (0.001), number of hidden neurons (10), maximum training time, and maximum validation failures (6). The network training and testing was split to a standard division of training (70%) and testing (30%) in these cases. The evaluation was done based on simulation results, the computation of average MSE and an average Classification Accuracy (CAcc).

**Least squares support vector machine (LS-SVM) classifier**

Support vector machines (SVM) is dominant in nonlinear classification problems, function, and density estimation in kernel based schemes. The SVM originated from statistical learning theory and
structural risk minimization where quadratic programming is used to solve convex optimization problems. The standard SVMs are re-formulated to obtain the least squares support vector machines (LS-SVM). The LS-SVM has also been used successfully in the classification of Bonn EEG datasets[27]. In this work, the classification performance of LS-SVM was assessed due to its extensive application in machine learning domain.

For a given training set \( \{x_n, y_n\}_{n=1}^N \) and input pattern \( x_n \in \mathbb{R}^p \) and class patterns \( y_n \in \{-1, +1\} \), the classifier output is given by[28]

\[
y(x) = \text{sign}[W^T \varphi(x) + b] \tag{18}
\]

with \( \varphi(.) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) mapping to large dimensional feature which can be up to infinite dimension. The terms \( W \) and \( b \) are weight vector and bias respectively. The term \( \varphi(x) \) can map the \( x \) into \( d \)-dimensional weight vector.

Thus for separable data, the assumptions are:

\[
\begin{align*}
W^T \varphi(x_n) + b &> +1, \text{if } y_n = +1 \\
W^T \varphi(x_n) + b &< -1, \text{if } y_n = -1
\end{align*} \tag{19}
\]

which is equivalent to

\[
y_n[W^T \varphi(x_n) + b] \geq 1, k = 1, \ldots, N \tag{20}
\]

In the LS-SVM technique, optimal separation of hyper-plane can be determined to maximize the distance from every class to the hyper-plane. The optimization problem can be addressed by the following equation[29]:

\[
\min_{\omega, \xi} \hat{\mathcal{S}}(\omega, \xi) = \frac{1}{2} \omega^T \omega + c \sum_{k=1}^N \xi_k \tag{21}
\]

which is subjected to equality constraints,

\[
\begin{align*}
y_k \left[ W^T \varphi(x_k) + b \right] &> 1 - \xi_k, \quad k = 1, \ldots, N \\
\xi_k &> 0, \quad k = 1, \ldots, N
\end{align*} \tag{22}
\]

Constructing Lagrangian is given by

\[
\mathcal{L}(\omega, b, \xi; \alpha, \nu) = \mathcal{T}(\omega, \xi_k) - \sum_{k=1}^N \alpha_k \left[ y_k \left[ W^T \varphi(x_k) + b \right] - 1 + \xi_k \right] - \sum_{k=1}^N \nu_k \xi_k \tag{23}
\]

with Lagrange multipliers \( \alpha_k \geq 0, \nu_k \geq 0, (k = 1, \ldots, N) \).

After solving equation 24, the LS-SVM classifier is given by

\[
f(x) = \text{sign} \left[ \sum_{k=1}^N \alpha_k y_k K(x_k, y_k) + b \right] \tag{24}
\]

where represents a kernel function. In this paper, radial basis kernel function has been used for the 2-class problem. The kernel function is defined by

\[
K(x_k, y_k) = e^{-\frac{||x_k - y_k||^2}{2\sigma^2}} \tag{25}
\]

Random forest (RF) classifier

RF are a combination of tree predictors and emerged as an efficient tool in classification for the reason that they do not overfit. The accuracy of the RF classifiers improves with injection of right type of randomness, the individual predictor's strength and their correlations. RFs produce results competitive with boosting and adaptive bagging without change in the training set. With the knowledge of the construction of single classification trees, RF develops several trees for classification. To classify a new sample, the values of the sample are assigned to the trees in the forest. Then, every tree gives a classification, which means it actually "votes" for that class. The forest chooses the classification having the most votes as compared to all the trees in the forest[29].

For \( M \) input variables, a number \( m<<M \) is specified so that at each node, \( m \) variables are randomly selected out of the \( M \) and the finest split on these \( m \) is used to divide the node. The value of \( m=10 \) and it is held constant for the period of the forest development.

In this work, 3-class problem (normal, pre-ictal, and ictal) is designed using RF classifier due to its suitability in multiclass classification over SVM. A random vector \( \xi_n \) is used to assign the \( n \)-th tree. The number of features in the training set are \( x_n=100 \). The tree classifier \( C(x_N, \xi_n) \) for 3-class problem was developed using feature samples and the random vector.

Results

The experiment used datasets with difference in signal from adjacent channels. Differencing stabilized the mean of EEG signals by eliminating changes in the level of the signals, and so eliminated seasonality and drift. Firstly, signals from various channels were decomposed into band-limited, oscillatory and symmetric IMFs. In this work, signal pairs of seizure and non-seizure EEG signals were decomposed using EEMD algorithm to obtain 8 IMFs. A value of 0.4 for standard deviation error and 10 for ensemble number were selected for EEMD decomposition. Three entropy parameters namely, KraEn, PermEn, SampEn were computed for the 8 IMFs. Since maximum oscillatory information about the underlying signal is obtained from the first few IMFs[18], selection of the first four IMFs were made for TQWT decomposition. Accuracy, Sensitivity, Specificity and Receiver
Operating Characteristics (ROC) are standard evaluation parameters for machine learning algorithms and have been used to analyze the performance of a classifier. The resulting observations of 3 different maximum entropy feature values derived from 8 IMFs of real time dataset are shown in Table 1.

A total of 2,562 samples of normal and seizure EEG dataset were considered due to computational complexity. However, by employing EEMD, the feature set for 2,562 samples of the normal and seizure EEG signals had been considered and found computationally efficient with an average computation time of 11.63 seconds. The algorithm was tested using MATLAB 9.0.1 tool on Intel ®Core™ i5-7200 U CPU @ 2.5 GHz with 8 GB RAM. The computation time was obtained for the entire 2,562 samples for each of the entropy features and is depicted in Table 2. The table shows the average values of computation time for randomly chosen 100 data samples from both signal classes. The minimum average time for computation of KraEn, PermEn, and SampEn was 0.042 for each IMF. The computation time depicts the potential of the proposed EEG classification algorithm in real time hardware implementation. The CenCorrEn of the 16th sub-band of TQWT is separately computed and computation time of the same has not been considered in this paper.

The KraEn estimated the Shannon entropy of the dataset using k-nearest neighboring sample. The KraEn of IMFs with Euclidean distance produced a valuable feature set for 2-nearest neighbor sample. The PermEn computes the permutation entropy $H$ of the IMFs using permutation order $n$, time lags from $\tau$, and a method to treat equal values. These parameters describe the accuracy of the values in the IMF data series by the number of decimal places. In this work, the value of $n=1, 2, 3$ were considered to experiment and the optimum PermEn feature set was obtained for $n=3$, for which maximum classification accuracy has been achieved. The SampEn computes sample entropy of input signal vector for maximum template length $M$. The default value of $M=5$. Experimental values of $M=1, 2, 3, 4, 5$ have been employed in this work. When the SampEn is estimated for $M=1, ..., M-1$, the feature set size increases with standard error estimates (se), and therefore the number of matches for $m=1, ..., M$ and $m=0, ..., M-1$ are computed separately. In this work the maximum accuracy of RF classifier has been reported for SampEn feature vector computed for $M=1$.

Initially, 100 EEG signals (50 each from normal

<table>
<thead>
<tr>
<th>IMF</th>
<th>Feature parameters</th>
<th>Real time dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KraEn</td>
<td>FP-channel</td>
</tr>
<tr>
<td>IMF1</td>
<td>PermEn</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>2.15</td>
</tr>
<tr>
<td>IMF2</td>
<td>PermEn</td>
<td>2.79</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>2.19</td>
</tr>
<tr>
<td>IMF3</td>
<td>PermEn</td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>2.19</td>
</tr>
<tr>
<td>IMF4</td>
<td>PermEn</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>1.74</td>
</tr>
<tr>
<td>IMF5</td>
<td>PermEn</td>
<td>4.84</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>1.31</td>
</tr>
<tr>
<td>IMF6</td>
<td>PermEn</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>2.18</td>
</tr>
<tr>
<td>IMF7</td>
<td>PermEn</td>
<td>3.88</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>1.11</td>
</tr>
<tr>
<td>IMF8</td>
<td>PermEn</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>SampEn</td>
<td>1.07</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IMF</th>
<th>Computation time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KraEn</td>
<td>0.055</td>
</tr>
<tr>
<td>PermEn</td>
<td>0.043</td>
</tr>
<tr>
<td>SampEn</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Table 1 The maximum value of the entropy features computed for fourteen IMFs of real time datasets sample EEG signals

Table 2 Computation time for entropies of IMFs

and seizure) were selected to compute entropy features. Using 8 IMFs and 3 entropy parameters the size of the feature vector became 42×1 000. In the next step, to minimize the classifier complexity, selection of the IMFs containing the best possible
entropy for further decomposition using TQWT was achieved using Student's $t$-test ($P \leq 0.05$) [31]. The reduced feature set after performing the test was of $12 \times 1000$ size is shown in Table 3.

Further, the pair of IMFs was decomposed using TQWT for various values of $Q$ and $J$. The maximum classification accuracy was obtained for subband$_{16}$ of IMF pairs IMF$_1$, IMF$_3$, and IMF$_4$ for $Q=3$ and $J=16$. The minimum value of $r$ was selected as 3. The decomposition of the 1$^{st}$ IMF from seizure difference signal into the 1$^{st}$ to 8$^{th}$ and 9$^{th}$ to 16$^{th}$ sub-bands are shown in Fig. 6 and 7, respectively.

For CenCorrEn features computed from the 16$^{th}$ sub-band are shown in Table 4.

The computation of CenCorrEn was done using Laplacian kernel in ITL toolbox available online (http://www.esat.kuleuven.be/sista/lssvmlab/). The MATLAB function computes the parametric CenCorrEn between two IMF vectors, IMF$_x$ and IMF$_y$, where $x'$ and $y'$ are positive integers. In this work, the maximum entropy was obtained for IMF$_{34}$ for default parameter of [1, 0] and kernel size ($k$Size) = 1. The CenCorrEn was computed for these IMF pairs to form the final feature set which is of size $16 \times 1000$. To select optimum values of $Q$ and $J$ using CenCorrEn, experiments were performed to evaluate classification accuracy for various values of $Q$ and $J$ of TQWT. The LS-SVM classifier was implemented in MATLAB using LS-SVM toolbox available online (http://www.esat.kuleuven.be/sista/lssvmlab/).

The selection of optimum value of $Q$ and $J$ was experimented by taking only CenCorrEn features that produced the highest classification accuracy. Further, the values of $J$ are varied by keeping $Q=3$. The maximum possible value of $J$ with $Q=3$ is 16. Hence, the value of $J$ was tested from 4 to 18 and CenCorrEn was computed for each sub-band. Then, the classification was performed by varying value of $J$ from 4 to 18. The plot of classification accuracy versus values of $J$ is shown in Fig. 8.

It was observed that at $J=16$, maximum classification accuracy is obtained. The fixed value of $J=16$ was then used to test the value of $Q$ for maximum accuracy. The variation of accuracies with $Q$ is shown in Fig. 9.

<table>
<thead>
<tr>
<th>Feature</th>
<th>IMF</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>KraEn</td>
<td>IMF$_1$</td>
<td>0.0413</td>
</tr>
<tr>
<td></td>
<td>IMF$_2$</td>
<td>0.0365</td>
</tr>
<tr>
<td>PermEn</td>
<td>IMF$_1$</td>
<td>0.0371</td>
</tr>
<tr>
<td></td>
<td>IMF$_2$</td>
<td>0.0194</td>
</tr>
<tr>
<td>SampEn</td>
<td>IMF$_1$</td>
<td>0.000147</td>
</tr>
<tr>
<td></td>
<td>IMF$_2$</td>
<td>0.000026</td>
</tr>
<tr>
<td></td>
<td>IMF$_3$</td>
<td>0.000054</td>
</tr>
<tr>
<td></td>
<td>IMF$_4$</td>
<td>0.000135</td>
</tr>
</tbody>
</table>

IMF: intrinsic mode functions; KraEn: Kraskov entropy; PermEn: permutation entropy; SampEn: sample entropy.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Non-focal EEG signal</th>
<th>Focal EEG signal</th>
<th>$P$-value (average for all four signals of first IMF decomposed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$-channel</td>
<td>$y$-channel</td>
<td>$x$-channel</td>
</tr>
<tr>
<td>CenCorrEn</td>
<td>0.1582</td>
<td>0.1463</td>
<td>0.1763</td>
</tr>
</tbody>
</table>

CenCorrEn: centered correntropy; EEG: electroencephalogram.
The use of various transfer functions in MLPNN, namely, "tansig", "logsig", "purelin", and "elliotsig" resulted in a comparable performance. The features were randomized before being fed into MLP network and re-trained 10 times to calculate average MSE. From the preliminary study, a number of hidden neurons were set to 10 and "tansig" was set as an input-hidden transfer function. Four training functions, namely, conjugate gradient with Powell-Beale restarts (CGB), variable learning rate backpropagation (GDX), Levenberg-Marquardt (LM), and scaled conjugate gradient (SCG) were used for the experimental study from neural network toolbox of MATLAB (MathWorks, Natick, USA). Fig. 10 depicts average classification accuracy for CGB, GDX, LM, and SCG training functions for 2-class (normal-seizure) and 3-class (normal-pre-seizure-seizure) problems.

It was inferred from the simulation study that the tan-sigmoid and pure linear transfer functions were optimal for input-hidden layer and hidden-output layer respectively with LM training function for network learning. Four different classification tasks were considered to assess the performance of MLPNN.

**Discussion**

In the RF classifier, the increase in the number of decision trees decreases the classification errors. The number of trees was less than 50 and the errors varied with minor changes when the number of trees was larger than 250. For the RF classifier, the minimum samples for the terminal node were taken 15 and the
maximum depth of tree was tuned to the cross-validation. The maximum number of terminal nodes and maximum features to consider for split was selected based on the value of square root of the total features. However, values up to 30%–40% of the total number of features were tested to check over-fitting. The 10-fold cross validation and split of dataset into 70% train and 30% test group for both the LS-SVM and RF classifiers resulted in classification accuracies shown in Table 5.

It was observed that the tuning of RF classifier for 3-class problem outperformed LS-SVM classifier on real time dataset. The receiver operating characteristics (ROC) curves for these two classifiers is shown in Fig. 11. The ROC curves depict that the proposed method improved the RF classifier performance significantly from 84% to 96% for multichannel data samples. The computation time for the entropy calculation of all four entropy parameters is shown in Table 6. Multichannel EEG analysis from all 19 channels involves a large number of samples and hence computational complex. Hence, multi-channel signals are grouped in 5 regions and the average of signals from these regions is considered without reducing the number of channels. Table 6 shows computation time of 3 entropy parameters for average of signals from brain regions.

It was observed that average computation time of 3.54, 2.62 and 3.32 seconds is required for SampEn, PermEn and KraEn. As shown in Table 7, the computation time for CenCorrEn is relatively small (0.47 seconds) as it is computed only for the 1st and 16th sub-band of TQWT decomposed IMFs.

In Table 8, the results of recent algorithms in EEG research are compared with the proposed method. In the comparison experiment, it was found that the proposed joint algorithm shows approximately 2% of improvement in classification accuracy. Besides, a comparative experiment using a basic feature extraction method using discrete wavelet transform (DWT) based entropy features fed to RF classifier results in 87% classification accuracy, while the proposed joint EEMD-TQWT results an improvement of 9.2% accuracy. This study was carried to verify the effectiveness of the proposed model with baseline method.

Finally, the joint EEMD-TQWT and RF algorithms for feature extraction and classification respectively have been implemented using an embedded system.

The EEG signal is captured with a low cost 5-channel EMOTIV INSIGHT EEG sensor. The sensor kit records EEG from 5 channels, namely, AF3, AF4,
Epilepsy is a chronic brain disorder with recurrent seizures and associated disorders. The AEDs and surgical procedures are conventional ways to control seizures in case of drug resistant epilepsy. In such cases, the non-invasive seizure detection plays a vital role in automated diagnosis of epilepsy and needs systematic study and development of signal processing algorithms. In this work, a technique to categorize seizure and normal EEG signals using the tuning of $Q$-factor in the wavelet transform is explored. Initially, EEMD is used to decompose raw EEG signals into band-limited IMFs. Based on highest values of three entropy parameters namely, KraEn, PermEn, and SampEn, IMFs are selected for further decomposition using TQWT. The CenCorrEn is then computed for the 1st and 16th sub bands of selected IMFs. This ensures a more relevant feature space in classifying bivariate signals. The values of TQWT parameters $Q=3$, $r=3$, and $J=16$ have produced optimum classification performance. The joint EEMD-TQWT method has generated highest classification accuracy of 96.2% for RF classifier when compared to MLPNN and LSSVM. A compact low cost embedded system developed using optimum feature extraction and classification algorithms may find application in automated seizure diagnosis.

### Acknowledgments
The authors are grateful to Dr. M. D. Mohire, MD, DM (Neurology), Neuro-physician, Mohire's Neurology Center & Research, Kolhapur, Maharashtra, INDIA for observing the results and providing the EEG dataset for experimentation. We also thank the developers of open source ITL toolbox and TQWT toolbox in MATLAB.

### References


### Table 8 Comparison of classification performance

<table>
<thead>
<tr>
<th>Authors</th>
<th>Features</th>
<th>Classifier</th>
<th>Accuracy (%)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rajeev Sharma et al[2]</td>
<td>Entropy</td>
<td>LS-SVM (RBF kernel)</td>
<td>86.00</td>
<td>88.00</td>
<td>84.00</td>
</tr>
<tr>
<td>Vipin Gupta et al[3]</td>
<td>Entropy</td>
<td>LS-SVM (RBF kernel)</td>
<td>94.41</td>
<td>93.25</td>
<td>95.57</td>
</tr>
<tr>
<td>Baseline method</td>
<td>DWT-entropy</td>
<td>RF</td>
<td>87.00</td>
<td>86.23</td>
<td>83.12</td>
</tr>
<tr>
<td>Proposed work</td>
<td>Entropy</td>
<td>RF</td>
<td>96.20</td>
<td>94.50</td>
<td>95.60</td>
</tr>
</tbody>
</table>

LS-SVM: least squares support vector machine; RBF: radial basis function; DWT: discrete wavelet transform; RF: random forest.


