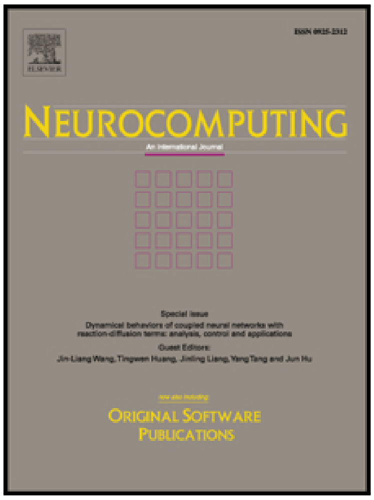
Communicated by Liang Lin

# Accepted Manuscript




Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier

Md Mursalin, Yuan Zhang, Yuehui Chen, Nitesh V Chawla

PII: S0925-2312(17)30344-2
DOI: 10.1016/j.neucom.2017.02.053
Reference: NEUCOM 18127

To appear in: *Neurocomputing*

Received date: 29 June 2016
Revised date: 29 December 2016
Accepted date: 19 February 2017

Please cite this article as: Md Mursalin, Yuan Zhang, Yuehui Chen, Nitesh V Chawla, Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.02.053

# Automated Epileptic Seizure Detection Using Improved Correlation-based Feature Selection with Random Forest Classifier

Md Mursalin[1], Yuan Zhang[1]*,Yuehui Chen[1], Nitesh V Chawla[2]
[1]Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China
[2]Department of Computer Science and Engineering, University of Notre Dame, USA
yzhang@ujn.edu.cn

***Abstract*—Analysis of electroencephalogram (EEG) signal is crucial due to its non-stationary characteristics, which could lead the way to proper detection method for the treatment of patients with neurological abnormalities, especially for epilepsy. The performance of EEG-based epileptic seizure detection relies largely on the quality of selected features from an EEG data that characterize seizure activity. This paper presents a novel analysis method for detecting epileptic seizure from EEG signal using Improved Correlation-based Feature Selection method (ICFS) with Random Forest classifier (RF). The analysis involves, first applying ICFS to select the most prominent features from the time domain, frequency domain, and entropy based features. An ensemble of Random Forest (RF) classifiers is then learned on the selected set of features. The experimental results demonstrate that the proposed method shows better performance compared to the conventional Correlation-based method and also outperforms some other state-of-the-art methods of epileptic seizure detection using the same benchmark EEG dataset.**

***Index Terms*—Electroencephalogram (EEG), Discrete Wavelet transformation (DWT), Correlation-based Feature Selection (CFS), Improved Correlation-based Feature Selection (ICFS), Random Forest (RF)**

## I. Introduction

Epilepsy is one of the most common disorders of the nervous system and affects people of all ages, races and ethnic backgrounds. Epileptic seizures are characterized by an unpredictable occurrence pattern and transient dysfunctions of the central nervous system, due to excessive and synchronous abnormal neuronal activity in the cortex [1]. This activity could include several neurons of different locations and sizes. The clinical symptoms of epileptic seizures might affect the motor, sensory, and automatic functions of the body along with the consciousness, cognition, and memory of the patient [2]. To diagnose epilepsy, EEG signal interpretation is considered as the most prominent testing tools due to the fact that it is painless, low cost, and has efficient temporal resolution of long-term monitoring [3]. However for long EEG recording the visual interpretation becomes an expensive, intensive and tedious error-prone exercise and also result can vary from different neurophysiologists in the same recording [4]. Thus, there is an ever-increasing need for developing an effective method for automatic seizure detection in order to prevent the possibility of any missing information so as to give a proper diagnosis and possible treatment plan for epilepsy. There are several factors which affect a classifier's behavior, such as the dimensionality of the feature space, the number of available patterns to learn from, imbalanced class labels and so on. However, to develop an efficient method, the main challenges are designing an appropriate feature extraction method and selecting the most prominent features because the quality of feature set plays an important role on the classification accuracy. An optimized feature selection using the gentle adaboost algorithm was recently proposed by Peng et al. [5], which also emphasizes on the quality of the features.

In this study we have developed a method to detect seizure from EEG signal using improved correlation based feature selection (ICFS) method which outperforms the conventional correlation based feature selection (CFS) methods. Our study shows that, ICFS requires on average three fewer features than the conventional CFS method. The Random Forest (RF) classifier is used in this study for detecting the seizure. The contributions and novelties of the proposed method are,

i) Extracting the most prominent features from time domain, frequency domain, and entropy based features.

ii) Developing an Improved Correlation based Feature Selection (ICFS) method for selecting the most distinguishing features.

iii) Classifying the output by applying random forest classifier based on ICFS.

iv) Analyzing the performance of both conventional CFS and ICFS method.

The rest of this paper is organized as follows. Section II discusses the review of prior work related to the different feature extraction and classification methods. The details of the EEG processing pipeline for our approach and its components are described in section III. In section IV the evaluation procedure and the experimental results are presented and concluded in section V.

## II. Related Work

Significant research has been done to detect the epileptic seizure from EEG signal. M. Z. Parvez et al. [6] proposed a generic seizure detection approaches for feature extraction of ictal and interictal signals using various established transformations and decompositions and least square support vector machine was applied on the features for classifications. They concluded that high frequency components of EEG signals contains the most prominent characteristics.

Rodriguez-Lujan et al. [7] proposed a feature selection method based on Quadratic Programming Feature Selection (QPFS). To minimize the computational complexity of the QPFS method, authors used the Nystrom method for approximate matrix diagonalization. In case of QPFS and our proposed ICFS, both methods used the conventional CFS for similarity measurement. However, there is some significant difference between these two methods. QPFS computed mutual information using principle component analysis while our proposed method simply analyzed the statistical measurement by calculating standard deviation with two-boundary threshold. QPFS introduced an objective function with quadratic and liner terms. Thus solving a quadratic programming problem can have high time and space complexity when large data sets are involved.

Song et al. [8] proposed a new Mahalanobis-similarity-based (MS) feature extraction method on the basis of the Mahalanobis distance and discrete wavelet transformation (DWT). In order to further improvement, the authors designed a fusion feature (MS-SE-FF) in the feature-fusion level, where the Mahalanobis-similarity-based feature characterizing the similarity between signals and the sample-entropy-based (SE) feature characterizing the complexity of signals are combined together.

Kumar et al. [9] applied DWT based fuzzy approximate entropy (fApEn) and support vector machine. The authors observed that the quantitative value of fuzzy approximate entropy drops during the ictal period which proves that the epileptic EEG signal is more ordered than the EEG signal of a normal subject. From all the data sets the fApEn values of different sub-bands are used to form feature vectors and these vectors are used as inputs to classifiers. They compared the classification accuracies between radial basis function based support vector machine (SVMRBF) and linear basis function based support vector machine (SVML).

Shen et al. [10] introduced an approach based on a cascade of wavelet-approximate entropy for the feature extraction in the EEG signal classification. They tested three existing methods, SVM, k-nearest neighbor (k-NN), and radial basis function neural network (RBFNN), and determined the classifier of best performance. In [11], the authors developed an efficient feature extraction method based on Hjorths mobility to reduce computational complexity while maintaining high detection accuracy. They proposed a new feature extraction method by computing the spectral power of Hjorths mobility components, which were effectively estimated by differentiating EEG signals in real-time. They suggested that the spectral features of Hjorths mobility components in EEG signals can represent seizure activity and may pave the way for developing a fast and reliable epileptic seizure detection method.

Tawfik et al. [12] introduced a new hybrid automated seizure detection model that integrates Weighted Permutation Entropy (WPE) and a SVM classifier model to enhance the sensitivity and precision of the detection process. They proposed that entropy based measures for the EEG segments during epileptic seizure are lower than in normal EEG. In [13], the authors proposed a combined method to classify normal and epileptic seizure EEG signals using wavelet transform (WT), phase-space reconstruction (PSR), and Euclidean distance (ED) based on a neural network with weighted fuzzy membership functions (NEWFM).

Classification of epileptic seizures in EEG signals based on phase space representation (PSRs) of intrinsic mode functions was proposed in [14]. In their work, firstly the EEG signals were decomposed using empirical mode decomposition (EMD) and then phase space were reconstructed for obtained intrinsic mode functions (IMFs). Two-dimensional (2D) and three-dimensional (3D) PSRs were used for the purpose of classification of epileptic seizure and seizure-free EEG signals.

Wang et al. [15] proposed a novel cost-effective active learning framework incorporate with the deep convolutional neural networks, which is capable of building a competitive classifier with optimal feature representation where the sample was selected via cost effective manner.

Bandarabadi et al. [16] proposed a method for epileptic seizure prediction using relative spectral power features. The relative combinations of sub-band spectral powers of EEG recordings across all possible channel pairs were utilized for tracking gradual changes preceding seizures. They used a specifically developed feature selection method and then a set of best candidate features were fed to support vector machines in order to discriminate cerebral state as preictal or non-preictal.

The aforementioned methods either use frequency domain or entropy based features to detect the abnormality in EEG signal. In this study we used time domain, frequency domain and entropy based features and selected the minimum features using our proposed ICFS method. In the following section we describe our proposed method in details.

## III. Proposed Method

In this work, EEG data sets are analyzed by DWT to decompose the signal in order to extract five physiological EEG bands, delta (0-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-60 Hz), using four levels with fourth-order Daubechies (db4). Then features were extracted from both wavelet decomposed and raw EEG data. All the extracted features were input to the improved correlation based feature selection (ICFS) step and then the most prominent features were selected to form the feature vector. This feature vector is used as input to the RF classifier for classifying the EEG signal as normal or seizure. The block diagram of the proposed approach is shown in Figure 1.

### A. Dataset

The EEG database used in this study is developed by the Department of Epileptology, University of Bonn, Germany [17]. The whole database consists of five EEG data sets (denoted as Set A-Set E). Each data set contains 100 single-channels and the duration of each channel is 23.6 s. There are 4097

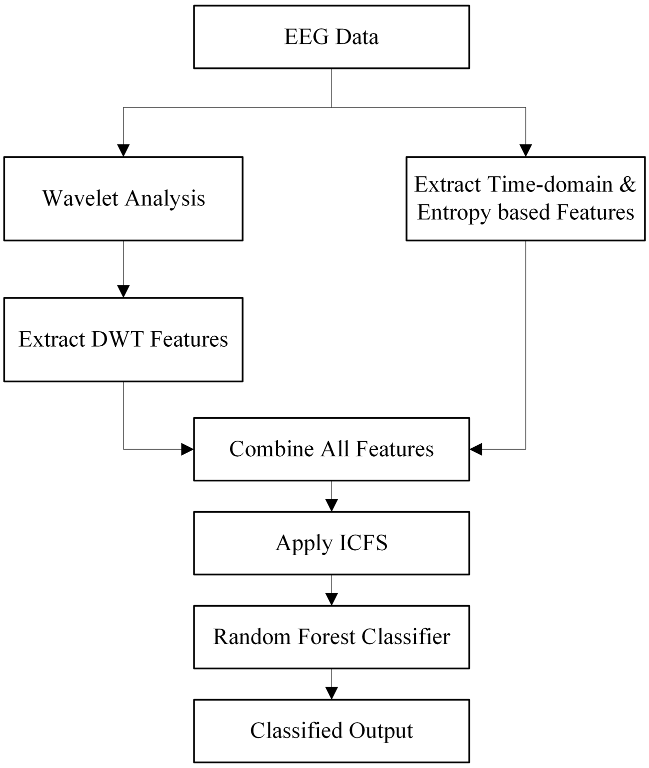



Fig. 1. The block diagram of proposed approach.

data points in each channel. Visual inspection for artifacts, such as causes of muscle activities or eye movements was done before choosing each signal. Using an average common reference, all EEG recordings were made with the same 128-channel amplifier system. The recoded data was digitized at 173.61 data points per second using 12-bit resolution where the band-pass filter settings were 0.5340 Hz (12 dB / oct). Set A and Set B were collected from surface EEG recordings of five healthy volunteers with eyes open and eyes closed, respectively. Sets C, D and E were collected from the EEG records of the pre-surgical diagnosis of five epileptic patients. Signals in Set C and Set D were recorded in seizure-free intervals from five epileptic patients from the hippocampal formation of the opposite hemisphere of the brain and from within the epileptogenic zone, respectively. Set E contains the EEG records of five epileptic patients during seizure activity. A summary description of the five set EEG data shown in Table I. Some examples of five EEG signals (Set A-Set E) are depicted in Figure 2.

### B. Feature extraction

The feature extraction process is a crucial step in processing non-stationary EEG signals because it reduces the space dimension and improves the accuracy of the classification process. We classify the features into three separate categories: (1) time domain features; (2) frequency domain features; and (3) entropy-based features. These features are described in the following subsections.

*1) Time domain features:* We extracted a total of fifteen time-domain features (divided into three groups) based on various studies [18], [19]. The first group consists of descriptive statistical features and the extracted features are mean ($X_{mean}$), median ($X_{med}$), mode ($X_{mean}$), minimum ($X_{min}$), maximum ($X_{max}$), skewness ($X_{sks}$), standard deviation ($X_{std}$), kurtosis ($X_{kur}$), first quartile ($Q_1$), third quartile ($Q_3$) and interquartile range ($Q_{ir}$).

The second group consists of two Hjorth parameters: mobility ($H_{mob}$) and complexity ($H_{com}$).

The third group consists of nonlinear features and the extracted features are Hurst exponent ($Hurst_{exp}$) and the detrended fluctuation analysis ($D_{fluc}$) consecutively.

*2) Frequency domain features:* EEG signal has non-stationary and transient characteristics. Thus, using only time domain feature is not enough to diagnosis. Moreover, frequency is one of the top fundamental characteristic of the EEG signal. When EEG waveform has a stable frequency, it is known as rhythmic [20]. The five basic frequency rhythms are Delta, theta, alpha, beta, and gamma. These rhythms are often related to various brain states, functions, or pathologies. In this paper we use DWT to extract the rhythms as explained below.

*Discrete Wavelet Transform:* As EEG signal is non-stationary and having transient characteristics so applying the Fourier transform directly is not suitable for processing any time-frequency information since they cannot be analyzed simultaneously. In order to overcome this disadvantage, Wavelet transforms (WT) are extensively applied in many engineering areas for explaining a variety of real-life problems. By using WT, most of the physiological signals having irregular patterns like impulses which are occurring at various points in the signals are generally analyzed. It can represent EEG sub bands as a weighted sum of shifted and scaled versions of the original wavelet, without any loss of information and energy. In this study, a discrete wavelet transform (DWT) was utilized to facilitate efficient time-frequency analysis. The DWT decomposes non-stationary signals at different frequency intervals with various resolutions. For a given wavelet function $\psi(t)$ that can be scale-shifted by the scaling parameter $a_j = 2^j$ and the translation parameter $b_{j,k} = 2^j k$. The DWT of signal can be defined as,

$$DWT\,(j,k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x\,(t)\,\psi\left(\frac{t-2^j k}{2^j}\right) dt \quad (1)$$

In the first step of the DWT, the signal is simultaneously passed through a low-pass (LP) and high-pass (HP) filters with the cut-off frequency being one fourth of the sampling frequency and the outputs are referred to as approximation (A1) and detail (D1) coefficients of the first level, respectively. The output signals with frequency bandwidth equal to half of the original signal can be down-sampled by a factor of two according to the Nyquist rule. The same procedure can be repeated for the first level approximation and the detail coefficients to get the next level coefficients. At each step of

TABLE I
SUMMARY OF THE EPILEPTIC EEG DATA OF FIVE SUBJECTS EACH.

| | Set A | Set B | Set C | Set D | Set E |
|---|---|---|---|---|---|
| Subjects | healthy | healthy | epileptic | epileptic | epileptic |
| Patients state | Awake and eyes open (normal) | Awake and eyes closed (normal) | Seizure-free (interictal) | Seizure-free (interictal) | Seizure activity (ictal) |
| Electrode type | Surface | Surface | Intracranial | Intracranial | Intracranial |
| Electrode placement | International 1020 system | International 1020 system | Opposite to epileptogenic zone | Within epileptogenic zone | Within epileptogenic zone |
| Number of channels | 100 | 100 | 100 | 100 | 100 |
| Time duration (s) | 23.6 | 23.6 | 23.6 | 23.6 | 23.6 |

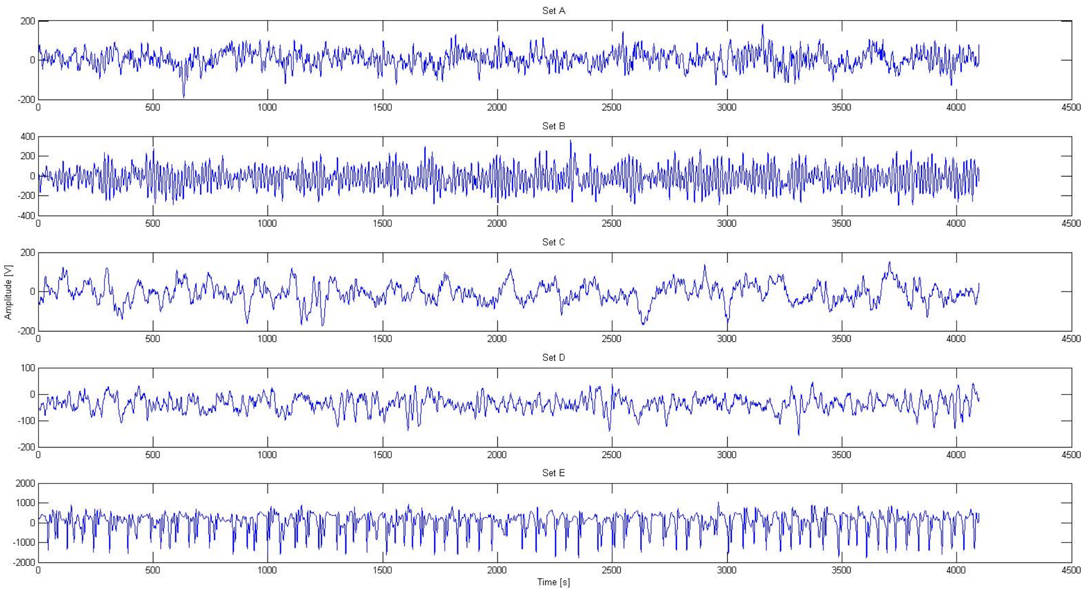



Fig. 2. Example of five different sets of EEG signals taken from different subjects.

this decomposition procedure, the frequency resolution is doubled through filtering and the time resolution is halved through down sampling. In this study, to extract five physiological EEG bands, delta (0-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-60 Hz), four levels DWT with fourth-order Daubechies (db4) wavelet function were used. Since our dataset is in range of 0-60 Hz, coefficients D1, D2, D3, D4 and A4 were extracted corresponding to 30-60 Hz, 15-30 Hz, 8-15 Hz, 4-8 Hz and 0-4 Hz, respectively. Table II shows the frequencies corresponding to different levels of wavelet decomposition for the EEG signals with a 173.6 Hz sampling rate. Figure 3 shows the sample of these sub-band signal from set E and set A respectively. The DWT features are extracted from (D1, D2, D3, D4 and A4) for each and every channel. The next subsection describe about the DWT features.

*Discrete wavelet-based features:* The extracted features, shown in Table III were used to represent the frequency distribution of the wavelet coefficients of the EEG signals.

TABLE II
FREQUENCY BAND OF EEG SIGNAL USING 4TH LEVEL DECOMPOSITION.

| Band name | Frequency band(Hz) | Sub-band signal | Decomposition level |
|---|---|---|---|
| gamma | 30-60 | D1 | 1 |
| beta | 15-30 | D2 | 2 |
| alpha | 8-13 | D3 | 3 |
| theta | 4-8 | D4 | 4 |
| delta | 0-4 | A4 | 4 |

These features were calculated for the decomposed signals, shown in Table II. Therefore total $5 \times 4$ values generated for each and every channel (here the number of coefficient and features are 5 and 4 respectively).

*3) Entropy-based features:* We used two entropy based features in this study, namely Sample entropy (SmpE) and

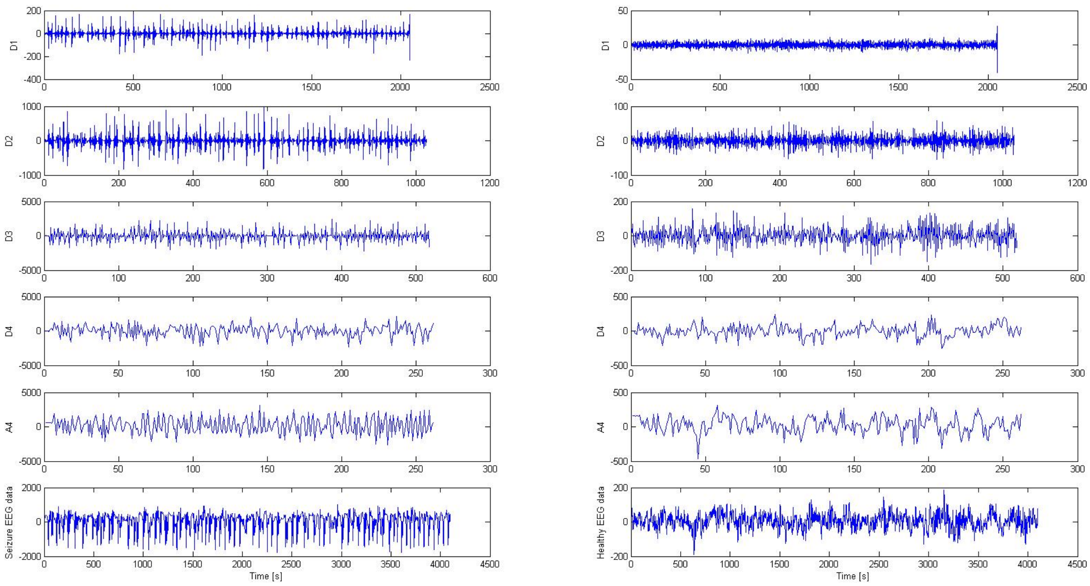



Fig. 3. Approximation and detail coefficients taken from set E (left side) and set A (right side)

TABLE III
DESCRIPTION OF THE EXTRACTED FEATURES FROM DECOMPOSED SIGNALS.

| No | Extracted features in each sub-band |
|---|---|
| 1 | Maximum of the wavelet coefficients ($W_{max}$) |
| 2 | Minimum of the wavelet coefficients ($W_{min}$) |
| 3 | Mean of the wavelet coefficients ($W_{mean}$) |
| 4 | Standard deviation of the wavelet coefficients ($W_{std}$) |

Shannon entropy (SE). SmpE is given by the formula,

$$SmpE(k, r, N) = ln(\frac{A(k)}{B(k-1)}) \quad (2)$$

where $B(0) = N$, the length of the input series, $k$ is the embedding dimension, $r$ is the scale or tolerance parameter (typically $0.2 \times std$).

The Shannon entropy equation provides a way to estimate the average minimum number of bits needed to encode a string of symbols, based on the frequency of the symbols as follows,

$$SE = -\sum_i s_i^2 \log_2(s_i)^2 \quad (3)$$

where $s$ is the signal and $(s_i)_i$ the coefficients of $s$ in an orthonormal basis. In this study SmpE value is calculated directly from EEG data while SE value is calculated from DWT coefficients (for each and every sub-band).

### C. Feature selection

Feature selection is a very important task in machine learning where it selects the best set of features from all the available features. This paper presents a new approach to feature selection, called ICFS (Improved Correlation-based Feature Selection) that uses a correlation based heuristic to evaluate the utility of selected features. This algorithm is simple, fast to execute and extends easily to continuous class problems by applying suitable correlation measures. The following subsections describe both CFS and our proposed ICFS method.

*1) Correlation-based Feature Selection (CFS):* CFS is a feature filtering method and is used to find the subset of features, which is potentially most relevant to the given classification task. The main part of the CFS algorithm is a heuristic for evaluating the utility or merit of a subset of features, as provided in Equation 4. This heuristic shows the usefulness of individual features for predicting the class label along with the level of intercorrelation among them.

$$Merit_S = \frac{k\bar{r_{cf}}}{\sqrt{k + k(k-1)\bar{r_{ff}}}} \quad (4)$$

where $Merit_s$ is the heuristic "merit" of a feature subset $S$ containing $k$ features, $\bar{r_{cf}}$ is the average feature class correlation, and $\bar{r_{ff}}$ is the average feature-feature intercorrelation. The heuristic aims to remove irrelevant and redundant features as they will be poor predictors of the class. In order to apply equation 4, it is necessary to compute the correlation between features. First of all CFS discretizes numeric features using the technique of Fayyad and Irani [21] and then uses symmetrical uncertainty (SU) to estimate the degree of association between discrete features. Equations 5 and 6 give the entropy of Y before and after observing $X$ where $X$ and $Y$ are discrete random variables.

$$H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y) \quad (5)$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2 p(y|x) \quad (6)$$

Information gain is the amount by which the entropy of Y decreases reflects the additional information about Y provided by X [22]. Information gain is given by,

$$Gain = H(Y) - H(Y|X) = H(X) - H(X|Y) \quad (7)$$
$$= H(Y) + H(X) - H(X, Y) \quad (8)$$

However, information gain is biased in favor of features with more values, that is, features with greater numbers of values will appear to gain more information than those with fewer values even if they are actually no more informative. Moreover the correlations in equation 4 should be normalized to ensure they are comparable and have the same effect. Thus the symmetrical uncertainty (SU) [23] compensates for information gain's bias toward attributes with more values and normalizes its value to the range of [0, 1]:

$$SU = 2.0 \times [\frac{Gain}{H(Y) + H(X)}] \quad (9)$$

First of all CFS calculates a matrix of feature-class and feature-feature correlations from the training data and then searches the feature subset space using various heuristic search strategies. We used a best first search technique. This search starts with an empty set of features and generates all possible single feature expansions. The subset with the maximum evaluation is chosen and expanded in the same manner by adding single features. If expanding the subset results in no improvement, then the search drops back to the next best unexpanded subset and continues from there. The best first search can explore the entire feature subset space as time goes, so it is common to limit the number of subsets expanded that result in no improvement. After terminating the search, the best subset found is returned. In this paper the stopping criterion of CFS is five consecutive fully expanded non-improving subsets.

*2) Improved Correlation-based Feature Selection (ICFS):* Feature selection methods should eliminate as much features as possible. Our main objective is to improve the performance of conventional CFS method and select the minimum number of features from a large feature space. We assume that statistical measurements such as standard deviation (SD) can play an important role in feature selection because a low standard deviation indicates that the data points tend to be close to the mean of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. Considering this fact our intention is to apply this technique in conventional CFS and modify algorithm in such a way that the performance will be optimized. To establish our assumption, we first calculated the SD value for each feature and checked the distribution of the data points in the feature. We found two types of distribution: for one case the points are largely distributed in one area while in the other case the points are distributed closer to the SD value area. Figure 4 shows the distribution for the different features. We can see that $W_{med2}$, $W_{med3}$, $W_{med4}$ and $W_{mean3}$ demonstrate the second type of distribution and all other features show the first type of distribution. From the performance analysis using except type one and except type two features, we discovered that except type two performed better. This implies that the data points with more variations have more important aspect for classification. Figure 5 shows the distribution after removing type two distributions of features. We described the ICFS algorithm in the following paragraph (The algorithm of ICFS method is showed in Algorithm 1).

**Algorithm 1** ICFS Algorithm

**Input:** *EEG signal*
**Output:** *Classified output*
Decompose the original signal: four levels DWT with fourth-order Daubechies (db4)
Extract both time domain and frequency domain features to form a feature vector
Calculates a matrix of feature class and feature-feature correlations from the training data using equation 4. The features are ranked in descending order based on their SU values
Let, $OPEN$ list contain $start\ sate, CLOSED$ list is empty, and $BEST \leftarrow start\ state$
**while** $OPEN$ is not empty **do**
 i. Let, $s = arg\ max(e(x))$ (get the state from OPEN with the highest evaluation)
 ii. Remove $s$ from $OPEN$ and add to $CLOSED.$
 **if** $e(s) \geq (BEST)$ **then**
  $BEST \leftarrow s$
 **end if**
 iii. For each child $t$ of $s$ that is not in the $OPENorCLOSED$ list, evaluate and add to $OPEN$
 **if** $BEST$ changed in the last set of expansions, **then**
  go to step i
 **end if**
 iv. $return\ BEST$
**end while**
Calculate the $\varphi$ value from the equation 6 for each $BEST$ value and put in the $Features$ list
call the $Th\_value$ selection Algorithm to get the $Th\_value1$, $Th\_value2$
**for** $i = 1\ to\ N'$ (where $N'$ is the number of features from BEST list) **do**
 **if** $Th\_value1 \leq Features[i] \leq Th\_value2$, **then**
  remove $Features[i]$ from the feature vector
 **end if**
**end for**
Fed these selected Features to the $RFclassifier$

In ICFS, we first calculate a matrix of feature-class and feature-feature correlations from the training data and then

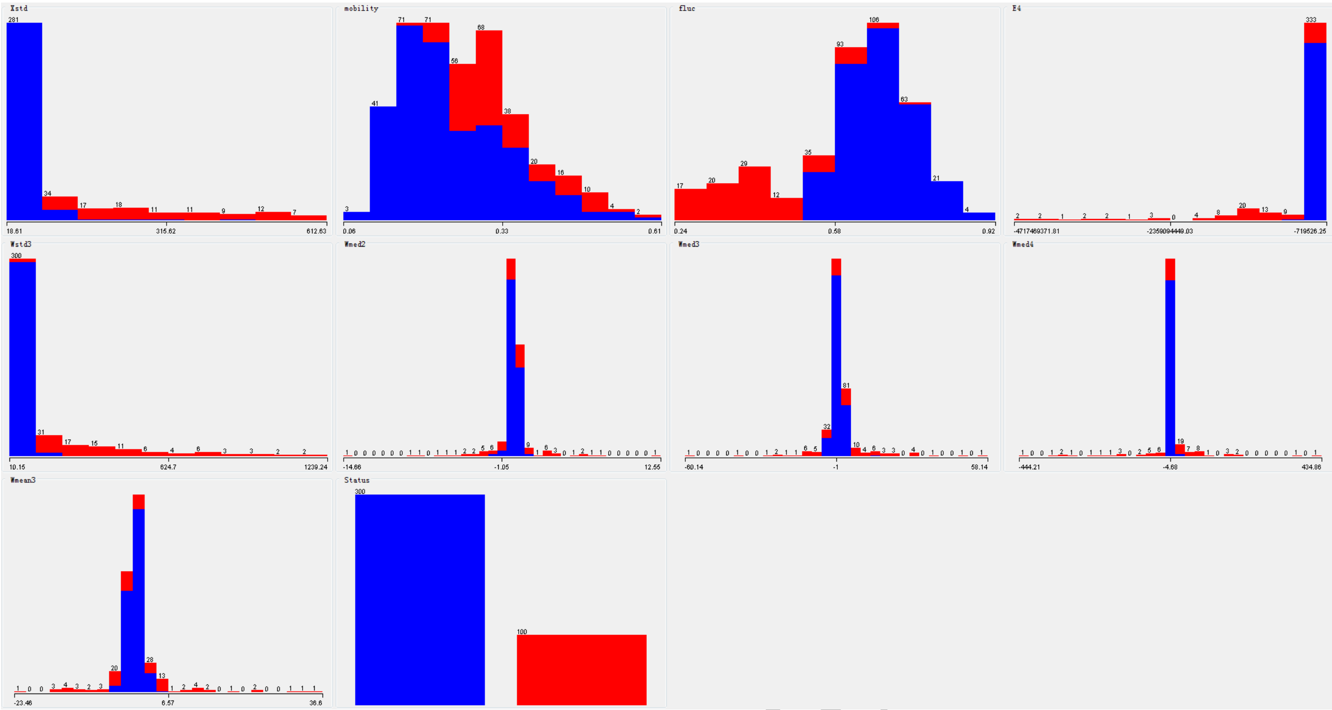


Fig. 4. The visualization of standard deviation of different features after conventional CFS (case 5, red color is for epileptic data while blue is for non-epileptic).

**Algorithm 2** $Th\_value$ selection Algorithm

**Input:** *List of features with* $\varphi$ *value*
**Output:** $Th\_value1$, $Th\_value2$
  **for** $i = 1$ *to* $n$ (where $n$ is the number of features from BEST list) **do**
    **if** more than 80% of data points have the value closed to $\varphi$ value **then**
      select $\varphi$ and put this value to a list[].
    **end if**
  **end for**
  Find the minimum and maximum value from the list[].
  $Th\_value1$= minimum, $Th\_value2$= maximum.
  return $Th\_value1$, $Th\_value2$

search the feature subset space using best first search. The subset with the maximum evaluation is chosen and expanded in the same manner as CFS by adding single features. If no improvement in expanding a subset results, the search drops back to the next best unexpanded subset and continues from there. After five consecutive non-improving subsets, the search is terminated and the best subset is returned. Then $\varphi$ (see equation 8) value is calculated for each and every selected feature.

$$\varphi = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(A_i - \mu)^2} \qquad (10)$$

where $N$ is the number of data channel and $A$ represents the value of selected attribute and $\mu$ is the mean value of all selected attributes which is calculated by following equation

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \qquad (11)$$

After calculating the $\varphi$ value for each selected features, two threshold values have been defined as $Th\_value1$ and $Th\_value2$. The threshold selection method is shown in Algorithm 2. Here we compared the each data points with the $\varphi$ value to see whether the data points were closer to the $\varphi$ value. If 80% of data points have nearest value of $\varphi$ then we selected that $\varphi$ value. In this way, we checked all the features and finally calculated the minimum and maximum $\varphi$ value. We assigned the value as $Th\_value1$= minimum and $Th\_value2$= maximum. All the $\varphi$ values presented between these two thresholds were removed from the feature vector.

## D. Classification

The Random Forest (RF) classifier proposed by L. Breiman [24] added an additional layer of randomness to bagging. RF consists of a collection or ensemble of simple tree predictors where each tree is capable of producing a response when presented with a set of predictor values. Furthermore to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. In case of standard trees, each node is split using the best split among all variables while in a random forest, each node is split using the best among a subset of predictors randomly chosen at that node. This strategy turns out to perform very well compared to many other classifiers,

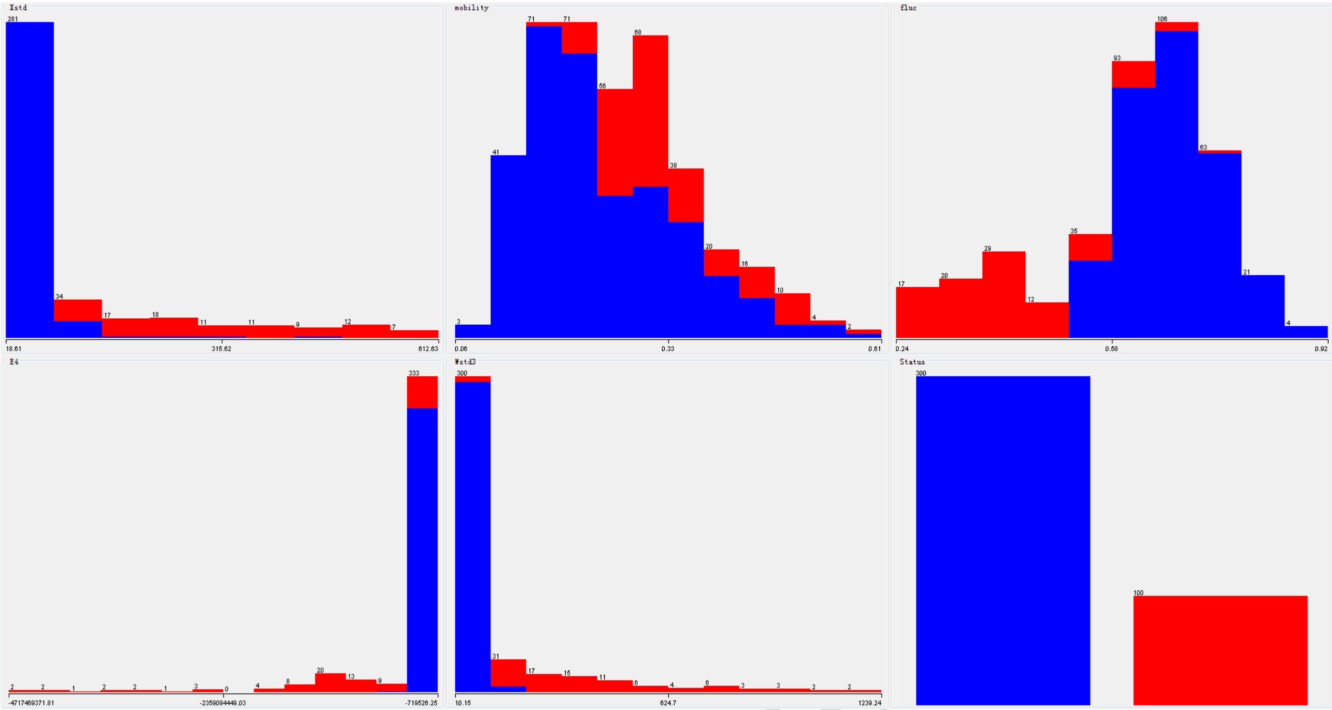



Fig. 5. The visualization of standard deviation of different features after conventional ICFS (case 5, red color is for epileptic data while blue is for non-epileptic).

| Case 1 | | Case 2 | | Case 3 | | Case 4 | | Case 5 | | Case 6 | | Case 7 | | Case 8 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value |
| Q1 | 101.9 | Qir | 205.6 | Q1 | 99.4 | Q1 | 100.0 | Xstd | 136.4 | Q3 | 98.4 | Xstd | 149.2 | Q3 | 90.9 |
| Qir | 215.4 | Xstd | 161.9 | Xstd | 165.9 | Q3 | 120.9 | mobility | 0.1 | Xstd | 134.1 | Xb2 | 1.7 | Xstd | 124.2 |
| Xstd | 169.4 | Xb1 | 0.5 | Xb1 | 0.6 | Qir | 210.2 | fluc | 0.1 | Xmin | 414.2 | Xmin | 461.9 | Xmin | 383.3 |
| Xb1 | 0.5 | Xmax | 469.8 | Xb2 | 0.9 | Xstd | 164.768 | E4 | 7.23E+08 | hus | 0.1 | mobility | 0.1 | E2 | 1.2E+08 |
| Xmin | 532.1 | hus | 0.1 | complexity | 0.2 | mobility | 0.1 | Wstd3 | 209.9 | E2 | 1.33E+08 | fluc | 0.2 | E4 | 6.54E+08 |
| Xmax | 497.2 | E2 | 18158612 | hus | 0.1 | fluc | 0.2 | Wmed2 | 1.6 | E4 | 7.21E+08 | E4 | 8.15E+08 | Wstd5 | 311.8 |
| saen | 0.3 | Wstd5 | 413.6 | saen | 0.2 | E3 | 9.42E+08 | Wmed3 | 7.9 | Wstd5 | 335.8 | Wstd3 | 236.1 | Wmed2 | 1.5 |
| E3 | 9.42E+08 | Wmin5 | 1091.8 | E3 | 9.43E+08 | E4 | 9.49E+08 | Wmed4 | 59.5 | Wmed2 | 1.7 | Wmed1 | 0.2 | Wmean3 | 4.5 |
| Wstd5 | 416.0 | | | Wmin2 | 354.3 | Wstd2 | 84.7 | Wmean3 | 4.9 | Wmean3 | 5.0 | Wmed2 | 1.9 | | |
| Wmin4 | 964.9 | | | Wmin4 | 960.4 | Wstd3 | 266.8 | | | | | Wmed4 | 68.7 | | |
| Wmin5 | 1105.8 | | | Wmean3 | 6.9 | Wmean3 | 6.9 | | | | | Wmean2 | 0.4 | | |
| | | | | Wmean4 | 23.4 | | | | | | | | | | |

Fig. 6. Standard deviation for each case after conventional CFS.

including discriminant analysis, support vector machines and neural networks, and is robust against over-fitting [24].

### *E. Cross-validation design*

The choices of dividing the data into training and test sets have many options [25]. In order to reduce the bias of training and test data, this study proposes employing k-fold cross-validation technique considering k = 10. This k-fold technique is implemented to create the training set and testing set for evaluation. Generally, with k-fold cross validation the feature vector set is divided into k subsets of equal size. Of the k subsets, a single subset is retained as the validation data for testing the model, and the remaining (k-1) subsets are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsets used exactly once as the validation data. Then, the average accuracy across all k trials is computed for consideration.

### *F. Performance measurements*

This paper assesses the performance of the proposed classifiers using criteria that are usually used in biomedical research such as sensitivity (proportion of the correctly classified ictal EEGs out of the total number of labeled ictal EEGs), specificity (proportion of the correctly classified inter-ictal EEGs out of the total number of labeled inter-ictal EEGs) and classification accuracy (proportion of the correctly classified

| Case 1 | | Case 2 | | Case 3 | | Case 4 | | Case 5 | | Case 6 | | Case 7 | | Case 8 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value | Features | SD Value |
| Q1 | 101.9 | Qir | 205.6 | Xstd | 165.9 | Q3 | 120.9 | Xstd | 136.4 | Xstd | 134.1 | Xstd | 149.2 | Xstd | 124.2 |
| Qir | 215.4 | Xstd | 161.9 | complexity | 0.2 | Qir | 210.2 | mobility | 0.1 | Xmin | 414.2 | Xmin | 461.9 | Xmin | 383.3 |
| Xstd | 169.4 | Xmax | 469.8 | hus | 0.1 | Xstd | 164.768 | fluc | 0.1 | hus | 0.1 | mobility | 0.1 | E2 | 1.2E+08 |
| Xmin | 532.1 | hus | 0.1 | saen | 0.2 | mobility | 0.1 | E4 | 7.23E+08 | E2 | 1.33E+08 | fluc | 0.2 | E4 | 6.54E+08 |
| Xmax | 497.2 | E2 | 18158612 | E3 | 9.43E+08 | fluc | 0.2 | Wstd3 | 209.9 | E4 | 7.21E+08 | E4 | 8.15E+08 | Wstd5 | 311.8 |
| saen | 0.3 | Wstd5 | 413.6 | Wmin2 | 354.3 | E3 | 9.42E+08 | | | Wstd5 | 335.8 | Wstd3 | 236.1 | | |
| E3 | 9.42E+08 | Wmin5 | 1091.8 | Wmin4 | 960.4 | E4 | 9.49E+08 | | | | | Wmed1 | 0.2 | | |
| Wstd5 | 416.0 | | | | | Wstd3 | 266.8 | | | | | Wmean2 | 0.4 | | |
| Wmin4 | 964.9 | | | | | | | | | | | | | | |
| Wmin5 | 1105.8 | | | | | | | | | | | | | | |

Fig. 7. Standard deviation for each case after ICFS.

EEGs out of the total number of EEGs). These criteria allow estimating the behavior of the classifiers on the extracted feature set. The definitions of these parameters are as follows:

$$SEN = \frac{TP}{TP + FN} \times 100\% \tag{12}$$

$$SPE = \frac{TN}{TN + FP} \times 100\% \tag{13}$$

$$CA = \frac{TP + TN}{TP + FN + TN + FP} \times 100\% \tag{14}$$

where, TP= True Positive, FN= False Negative, TN= True Negative, FP= False Positive, SEN= Sensitivity, SPE= Specificity, CA= Classification Accuracy.

## IV. Experimental results and discussion

### A. Results

All the 500 channels of EEG data sets are decomposed into different sub-bands using DWT of level four with fourth-order Daubechies (db4). The frequency ranges of these sub-bands are shown in Table II. Approximation and detail coefficients of the sample EEG channel taken from data sets A and E are shown in Figure 3. The wavelet features that are shown in Table III have been extracted from D1, D2, D3, D4 and A4. The fifteen time domain features are extracted from the raw EEG data channel. Sample entropy value is calculated directly from the raw EEG data while shannon entropy value is calculted from the five DWT coefficients (Table II). After extracting the features, total eight cases have been considered which is showed in Table IV.

TABLE IV
Different cases for classification.

| Case | Class 1 | Class 2 |
|---|---|---|
| 1 | A | E |
| 2 | B | E |
| 3 | C | E |
| 4 | D | E |
| 5 | ACD | E |
| 6 | BCD | E |
| 7 | CD | E |
| 8 | ABCD | E |

For each and every case we have applied ICFS and got the feature vector for classification. In this study the value of $Th\_value1$ and $Th\_value2$ were calculated according to the Algorithm 2. This algorithm first calculated all the standard deviation and then checked the data point. If all the data points situated close to the SD value then we discarded the feature from the feature set. This way we calculated the threshold value. It was observed that the value situated between .5 and 100, showed the above criteria. So we assigned the $Th\_value1$ as .5 and $Th\_value2$ as 100. For selecting these two values, we consider to remove only those features that would not lower the accuracy compare to conventional CFS method. Figure 6 shows the SD value of all cases with the selected features after conventional CFS and Figure 7 shows the SD value of all cases after our proposed ICFS method. We can see that the values present between the thresholds are discarded in Figure 7. The comparison between proposed method and the conventional CFS is shown in Figure 8. It is clearly observed that for each and every case the number of features is less than the conventional CFS method. We observed that the lowest number of features were selected for Case 5 and 8. After the ICFS step, the selected feature vector is then fed to the RF classifier where the number of trees to be generated is 100 and the random number seed is 1. Figure 9 shows the accuracy comparison between ICFS and CFS. For cases 4, 5 and 7, ICFS provides better performance with minimum features as 8, 5 and 8 respectively. Table VI shows the selected features from the case 4 and 5. The performance of the ICFS method for all cases is shown in Table V. It has been observed that case 1 shows the maximum result while case 8 shows the minimum among all the cases. This happens because for case 1, the data is either healthy or epileptic (set A vs set E) while on the other hand for case 8, the data consists of various classes (Table I) from set A to set D and another data containes seizure activity (set E). The average accuracy of the proposed method is 98.45% (Table V). The following subsection discusses the accuracy comparison with other existing methods.

TABLE V
PERFORMANCE OF THE RANDOM FOREST CLASSIFIER ON THE ICFS SCHEME

| Case | Sensitivity (%) | Specificity (%) | Classification accuracy(CA) (%) | Average CA(%) |
|---|---|---|---|---|
| 1 | 100 | 100 | 100 | |
| 2 | 98.0 | 98.0 | 98.0 | |
| 3 | 99.0 | 99.0 | 99.0 | |
| 4 | 98.5 | 98.5 | 98.5 | 98.45 |
| 5 | 98.5 | 98.5 | 98.5 | |
| 6 | 97.5 | 97.5 | 97.5 | |
| 7 | 98.7 | 98.7 | 98.67 | |
| 8 | 97.4 | 97.5 | 97.4 | |

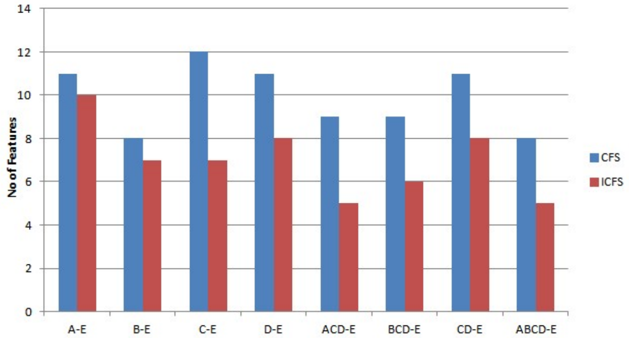



Fig. 8. Number of features in ICFS vs CFS.

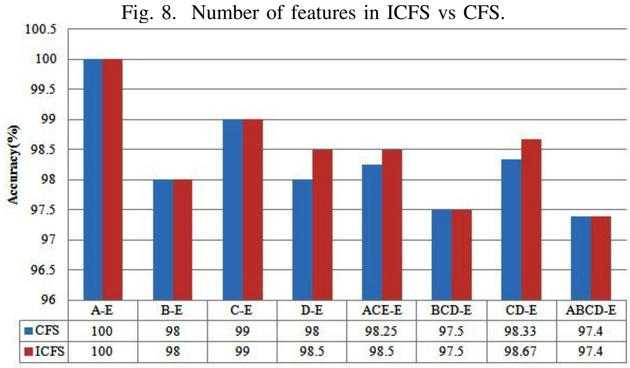



Fig. 9. Accuracy comparison between ICFS and CFS.

TABLE VI
THE SELECTED FEATURES FROM THE CASE 4 AND 5.

| Case 4 | Case 5 |
|---|---|
| $Q_3$,$Q_{ir}$,$X_{std}$,$H_{mob}$, $D_{fluc}$,$W_{std3}$,$W_{SE3}$,$W_{SE4}$ | $X_{std}$,$H_{mob}$,$D_{fluc}$,$W_{std3}$, $W_{SE4}$ |

### *B. Comparison with existing state-of-the-art work*

Table VII shows the comparison between this method and other recent methods ([9], [12], [26], [27], [28], [29]).

Case 1 shows the best classification accuracy estimated from this work as 100%. The similar result is also presented in Y. Kumar et al. [9] work, which was achieved from the Fuzzy approximate entropy with SVM classifier.

In cases 2 - 4, this work obtains the classification accuracies of 98%, 99.0% and 98.5%, which are the best compared to the work of [27] and [12] where they reported the accuracy as 82.8%, 88.0% , 79.94% (using permutation entropy with SVM), and 85%, 93.5%, 96.5% (using Weighted Permutation Entropy with SVM) respectively. It was observed that case 4 showed the maximum result among the other state-of-the-art work. However, results of cases 2 and 3 were 2% and 0.6% less than the corresponding cases in [9] due to the combination of Fuzzy approximate entropy with SVM classifier.

In case 5, the accuracy achieved from this work is 98.5% which is best compared to [9], where their result showed the accuracy of 98.15% by using SVM classifier.

In case 6, the classification accuracy obtained from the proposed method is 97.5% which is 0.72% less than [9] work due to the Fuzzy approximate entropy with SVM classifier.

In case 7, the classification accuracy obtained in this work is 98.67% which is the best presented for this data set. The result also presented in Kumar et al. [28] used Gabor filter with K-nearest neighbor classifier to achieve a 98.3% accuracy.

In case 8, the classification accuracy achieved from this study is 97.4% which is better than that presented in [9] (97.38%). However, the result is 0.87% less than that presented in [30] which used approximate entropy based feature with multi-wavelet transform.

## V. CONCLUSION

Accurate and perfect detection of epileptic seizure from EEG signals is one of the complex problems which depend on the features quality. The main contribution of this paper lies on developing an automatic, efficient and scalable ICFS based algorithm to detect the unpredictable occurrence of epileptic seizures in a reasonable time. In this work both time and frequency domain features are used to fed the proposed ICFS method which select the most prominent features for automatic seizure detection using random forest classifier. We have compared the both conventional CFS method and our proposed method and showed that our proposed method provides better performance. It was observed that, ICFS requires on average almost 3 features less than the conventional CFS method. The effectiveness of the proposed method is verified by comparing the performance of classification problems as addressed by

TABLE VII
COMPARISON BETWEEN OUR PROPOSED METHOD AND OTHER RECENT METHODS.

| Authors | Methods | Cases | Accuracy(%) |
|---|---|---|---|
| Ubeyli(2010) [31] | LS-SVM model-based method coefficients | A-E | 99.56 |
| Guo et al.(2010)[30] | Multiwavelet transform - approximate entropy feature-MLPNN | A-E | 99.85 |
| | | ABCD-E | 98.27 |
| Wang et al.(2011) [29] | Wavelet packet entropy - hierarchical EEG classification | A-E | 99.44 |
| Guo et al. (2011) [28] | GP-based feature extraction - KNN classifier | A-E | 99.2 |
| Nicolaou et al.(2012) [26] | Permutation entropy and SVM | A-E | 93.55 |
| | | B-E | 82.88 |
| | | C-E | 88.0 |
| | | D-E | 79.94 |
| Yatindra Kumar(2014)[9] | Fuzzy approximate entropy and SVM | A-E | 100 |
| | | B-E | 100 |
| | | C-E | 99.6 |
| | | D-E | 95.85 |
| | | ACD-E | 98.15 |
| | | BCD-E | 98.22 |
| | | ABCD-E | 97.38 |
| Noha S. Tawfik(2015)[12] | Weighted Permutation Entropy (WPE) and a SVM | A-E | 98.5 |
| | | B-E | 85.0 |
| | | C-E | 93.5 |
| | | D-E | 96.5 |
| T. S. Kumar (2015) [27] | Gabor filter and K-nearest neighbor | CD-E | 98.3 |
| Proposed model | Improved Correlation Feature selection and Random forest classifier | A-E | 100 |
| | | B-E | 98.0 |
| | | C-E | 99.0 |
| | | D-E | 98.5 |
| | | ACD-E | 98.5 |
| | | BCD-E | 97.5 |
| | | CD-E | 98.67 |
| | | ABCD-E | 97.4 |

other researchers. It can be concluded that using the proposed method for analyzing the EEG signal associated with epilepsy would help physicians, clinicians and neurophysiologists in making their work more efficient and their decisions more reasonable and accurate.

## Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61572231, 61173079 and 61472163. The corresponding author is Yuan Zhang.

## References

[1] A. Ngugi, S. Kariuki, C. Bottomley, I. Kleinschmidt, J. Sander, and C. Newton, "Incidence of epilepsy a systematic review and meta-analysis," *Neurology*, vol. 77, no. 10, pp. 1005–1012, 2011.

[2] W. H. Organization *et al.*, "Epilepsy management at primary health level in rural china: Who/ilae/ibe," *A Global Campaign Against Epilepsy Demonstration Project*, 2009.

[3] S. B. Akben, A. Subasi, and D. Tuncel, "Analysis of eeg signals under flash stimulation for migraine and epileptic patients," *Journal of medical systems*, vol. 35, no. 3, pp. 437–443, 2011.

[4] L. N. Ranganathan, S. A. Chinnadurai, B. Samivel, B. Kesavamurthy, and M. M. Mehndiratta, "Application of mobile phones in epilepsy care," *International Journal of Epilepsy*, vol. 2, no. 1, pp. 28–37, 2015.

[5] Z. Peng, Y. Li, Z. Cai, and L. Lin, "Deep boosting: Joint feature selection and analysis dictionary learning in hierarchy," *Neurocomputing*, vol. 178, pp. 36–45, 2016.

[6] M. Z. Parvez and M. Paul, "Epileptic seizure detection by analyzing eeg signals using different transformation techniques," *Neurocomputing*, vol. 145, pp. 190–200, 2014.

[7] I. Rodriguez-Lujan, R. Huerta, C. Elkan, and C. S. Cruz, "Quadratic programming feature selection," *Journal of Machine Learning Research*, vol. 11, no. Apr, pp. 1491–1516, 2010.

[8] J.-L. Song, W. Hu, and R. Zhang, "Automated detection of epileptic eegs using a novel fusion feature and extreme learning machine," *Neurocomputing*, vol. 175, pp. 383–391, 2016.

[9] Y. Kumar, M. Dewal, and R. Anand, "Epileptic seizure detection using dwt based fuzzy approximate entropy and support vector machine," *Neurocomputing*, vol. 133, pp. 271–279, 2014.

[10] C.-P. Shen, C.-C. Chen, S.-L. Hsieh, W.-H. Chen, J.-M. Chen, C.-M. Chen, F. Lai, and M.-J. Chiu, "High-performance seizure detection system using a wavelet-approximate entropy-fsvm cascade with clinical validation," *Clinical EEG and neuroscience*, vol. 44, no. 4, pp. 247–256, 2013.

[11] J.-H. Kang, Y. G. Chung, and S.-P. Kim, "An efficient detection of epileptic seizure by differentiation and spectral analysis of electroencephalograms," *Computers in biology and medicine*, vol. 66, pp. 352–356, 2015.

[12] N. S. Tawfik, S. M. Youssef, and M. Kholief, "A hybrid automated detection of epileptic seizures in eeg records," *Computers & Electrical Engineering*, 2015.

[13] S.-H. Lee, J. S. Lim, J.-K. Kim, J. Yang, and Y. Lee, "Classification of normal and epileptic seizure eeg signals using wavelet transform, phase-space reconstruction, and euclidean distance," *Computer methods and programs in biomedicine*, vol. 116, no. 1, pp. 10–25, 2014.

[14] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in eeg signals based on phase space representation of intrinsic mode functions," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1106–1117, 2015.

[15] K. Wang, D. Zhang, Y. Li, R. Zhang, and L. Lin, "Cost-effective active learning for deep image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[16] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology*, vol. 126, no. 2, pp. 237–248, 2015.

[17] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E.

Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.
[18] A. Aarabi, F. Wallois, and R. Grebe, "Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis," *Clinical Neurophysiology*, vol. 117, no. 2, pp. 328–340, 2006.
[19] O. Farooq, Y. U. Khan *et al.*, "Automatic seizure detection using higher order moments," in *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*. IEEE, 2010, pp. 159–163.
[20] S. Sanei and J. A. Chambers, *EEG signal processing*. John Wiley & Sons, 2013.
[21] K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," 1993.
[22] J. Quinlan, "C4. 5: Program for machine learning morgan kaufmann," *San Mateo, CA, USA*, 1993.
[23] W. Press and B. Flannery, "Sa teukolsky, wt vetterling.(1988). numerical recipes in c," *Cambridge [ua]*.
[24] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
[25] W. A. Chaovalitwongse, Y.-J. Fan, and R. C. Sachdeo, "On the time series k-nearest neighbor classification of abnormal brain activity," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 37, no. 6, pp. 1005–1016, 2007.
[26] N. Nicolaou and J. Georgiou, "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines," *Expert Systems with Applications*, vol. 39, no. 1, pp. 202–209, 2012.
[27] T. S. Kumar, V. Kanhangad, and R. B. Pachori, "Classification of seizure and seizure-free eeg signals using local binary patterns," *Biomedical Signal Processing and Control*, vol. 15, pp. 33–40, 2015.
[28] L. Guo, D. Rivero, J. Dorado, C. R. Munteanu, and A. Pazos, "Automatic feature extraction using genetic programming: an application to epileptic eeg classification," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10 425–10 436, 2011.
[29] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical eeg classification for epileptic detection," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14 314–14 320, 2011.
[30] L. Guo, D. Rivero, and A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," *Journal of neuroscience methods*, vol. 193, no. 1, pp. 156–163, 2010.
[31] E. D. Übeyli, "Least squares support vector machine employing model-based methods coefficients for analysis of eeg signals," *Expert Systems with Applications*, vol. 37, no. 1, pp. 233–239, 2010.

**Md Mursalin** received his B.Sc. in Computer Science and Information Technology from Islamic University of Technology (IUT), Bangladesh, in 2010. He is presently a research scholar in the Ubiquitous Sensing Lab (USLab), which is part of the School of Information Science & Engineering at University of Jinan, China. Previously he was a lecturer of Computer Science and Engineering Department at Pabna University of Science and Technology.

**Yuan Zhang** received the M.S. degree in Communication Systems and Ph.D. degree in Control Theory & Engineering both from Shandong University, China, in 2003 and 2011 respectively. He is currently an Associate Professor at University of Jinan (UJN), China. Dr. Zhang was a visiting professor at Computer Science Department, Georgia State University, USA, in 2014. As the first author or corresponding author he has published more than 30 peer reviewed papers in international journals and conference proceedings, 1 book chapters, and 5 patents in the areas of mHealth. He has served as Leading Guest Editor for four special issues of IEEE, Elsevier, Springer and InderScience publications, and has served on the technical program committee for numerous international conferences. He is an associate editor for IEEE Access. Dr. Zhangs research interests are in biomedical signal processing and mHealth, currently focusing on wearable sensing and big data analytics in healthcare domain. His research has been extensively supported by the Natural Science Foundation of China, China Postdoctoral Science Foundation, and Natural Science Foundation of Shandong Province with total grant funding over one million RMB. Dr. Zhang is a Senior Member of both IEEE and ACM. For more information, please refer to http://uslab.ujn.edu.cn/index.html.

**Yuehui Chen** received his Ph.D. degree in electrical engineering from the Kumamoto University of Japan in 2001. During 2001-2003, he had worked as the Senior Researcher of the Memory-Tech Corporation at Tokyo. Since 2003 he has been a member at the Faculty of Electrical Engineering in University of Jinan, where he is currently head of the Laboratory of Computational Intelligence. His research interests include Evolutionary Computation, Neural Networks, Fuzzy Logic Systems, Hybrid Computational Intelligence and their applications in time-series prediction, system identification, intelligent control, intrusion detection systems, web intelligence and bioinformatics. He is the author and co-author of more than 200 technique papers. Professor Yuehui Chen is a member of IEEE, the IEEE Systems, Man and Cybernetics Society and the Computational Intelligence Society, a member of Young Researchers Committee of the World Federation on Soft Computing, and a member of CCF Young Computer Science and Engineering Forum of China.

**Nitesh Chawla** is the Frank M. Freimann Professor of Computer Science and Engineering at University of Notre Dame. He is the director of the Notre Dame Interdisciplinary Center for Network Science (iCeNSA), which is at the frontier of network and data science (Big Data) with a strong multidisciplinary focus. He has received numerous awards for research, scholarship and teaching. He is the recipient of the 2015 IEEE CIS Outstanding Early Career Award for 2015. He received the IBM Watson Faculty Award in 2012, and the IBM Big Data and Analytics Faculty Award in 2013. He received the National Academy of Engineering New Faculty Fellowship. In recognition of the societal and community driven impact of his research, he was recognized with the Rodney Ganey Award in 2014 and Michiana 40 Under 40 in 2013. He has also received and nominated for a number of best paper awards. He serves on editorial boards of a number of journals and organization/program committees of top-tier conferences. He is also the director of ND-GAIN Index, Fellow of the Reilly Center for Science, Technology, and Values, Fellow of the Institute of Asia and Asian Studies, and Fellow of the Kroc Institute for International Peace Studies at Notre Dame.