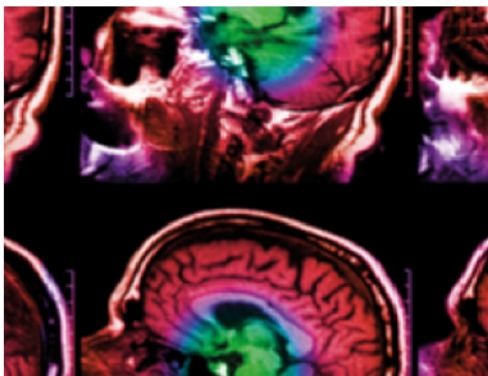


PAPER

A strategy combining intrinsic time-scale decomposition and a feedforward neural network for automatic seizure detection

To cite this article: Lijun Yang *et al* 2019 *Physiol. Meas.* **40** 095004

View the [article online](#) for updates and enhancements.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology

Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.



PAPER

A strategy combining intrinsic time-scale decomposition and a feedforward neural network for automatic seizure detection

Lijun Yang^{1,3}, Sijia Ding¹, Hao-min Zhou² and Xiaohui Yang¹¹ School of Mathematics and Statistics, Henan University, Kaifeng 475004, People's Republic of China² School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, United States of America³ Author to whom any correspondence should be addressed.E-mail: yanglijun@henu.edu.cn, sijiading@henu.edu.cn, hmzhou@math.gatech.edu and xhyang@henu.edu.cn

Keywords: EEG classification, seizure detection, intrinsic time-scale decomposition, feedforward neural network

RECEIVED
13 March 2019REVISED
18 August 2019ACCEPTED FOR PUBLICATION
23 August 2019PUBLISHED
26 September 2019**Abstract**

Epilepsy is a common neurological disorder which can occur in people of all ages globally. For the clinical treatment of epileptic patients, the detection of epileptic seizures is of great significance. *Objective:* Electroencephalography (EEG) is an essential component in the diagnosis of epileptic seizures, from which brain surgeons can detect important pathological information about patient epileptiform discharges. This paper focuses on adaptive seizure detection from EEG recordings. We propose a new feature extraction model based on an adaptive decomposition method, named intrinsic time-scale decomposition (ITD), which is suitable for analyzing non-linear and non-stationary data. *Approach:* Firstly, using the ITD technique, every EEG recording is decomposed into several proper rotation components (PRCs). Secondly, the instantaneous amplitudes and frequencies of these PRCs can be calculated and then we extract their statistical indices. Furthermore, we combine all these statistical indices of the corresponding five PRCs as the feature vector of each EEG signal. Finally, these feature vectors are fed into a feedforward neural network (FNN) classifier for EEG classification. The whole process of feature extraction proposed in this paper only involves one parameter and the role of the ITD method is based on a piecewise linear function, which makes the computation of the model simple and fast. More useful information for classification can be obtained since we take advantage of both instantaneous amplitude and instantaneous frequency for feature extraction. *Main results:* We consider the 17 classification problems which contain normal versus epileptic, non-seizure versus seizure and normal versus interictal versus ictal using a FNN classifier which only contains one hidden layer. Experimental results show that the proposed method can catch the discriminative features of EEG signals and obtain comparable results when compared with state-of-the-art detection methods. *Significance:* Therefore, the proposed system has a great potential in real-time seizure detection and provides physicians with a real-time diagnostic aid in their practice.

1. Introduction

Epilepsy is a chronic brain dysfunction syndrome, which is characterized by recurrent epileptic seizures caused by abnormal discharge of brain neurons. Nowadays, epilepsy is a common neurological disease, and its prevalence is only second to stroke. According to a new report from the World Health Organization, there are around 50 million people worldwide suffering from epilepsy (WHO 2019). Moreover, epilepsy can cause other health problems. Therefore, its prevention and diagnosis have become one of the primary problems in the medical community. An electroencephalogram (EEG) is a graph obtained by amplifying and recording the spontaneous biologic potential of the brain cells with sophisticated electronic instruments. EEG recordings reflect the spontaneous and rhythmic electrical activities of the brain cells, which makes it an essential component in the evaluation of epilepsy (Nolan *et al* 2004). However, reviewing a continuous EEG recording is tedious and time-consuming. In addition, when the number of EEG channels increases, the problem becomes even worse, which reduces the efficiency. To overcome these limitations, it is of great significance to design an automated EEG

recognition system to assist neurologists in classifying epileptic and non-epileptic EEG recordings (Ullah *et al* 2018). Automatic seizure detection, particularly if performed online, can be a valuable clinical tool to identify segments of the EEG likely to contain seizures (Li *et al* 2013).

In the last decade, much research work on automatic EEG detection has been carried out, and this can be mainly categorized into two research hotspots. One is feature extraction and the other is classifier design. In terms of EEG feature extraction, some researchers extracted features directly from the EEG signals (Zhou *et al* 2008, Minasyan *et al* 2010, Acharya *et al* 2012a). However, the features extracted only from the time domain are not comprehensive and inevitably ignore a lot of important information. Therefore, other researchers were more concerned with extracting features from the time-frequency combination by means of transform techniques. They transformed the EEG signals into the frequency domain and extracted the features in the frequency domain as well as in the time domain (Polat and Günes 2007, Tzallas *et al* 2009, Li *et al* 2013, Martis *et al* 2013, Pachori and Patidar 2014, Das and Bhuiyan 2016, Djemili *et al* 2016, Riaz *et al* 2016, Bhattacharyya and Pachori 2017, Li *et al* 2017, Gaur *et al* 2018). For multichannel EEG recordings, some researchers considered joint analysis-based techniques to extract the features (Zhou *et al* 2016, Bhattacharyya and Pachori 2017). Sparse representation-based techniques have also been used (Li *et al* 2014, Zhang *et al* 2019). Regularization-based techniques can prevent overfitting in EEG detection (Zhang *et al* 2016). Some researchers turned the problem of feature extraction into an optimization problem. They designed special penalized functions and solved them to obtain discriminative EEG features (Hosseini *et al* 2016, Hussein *et al* 2018). After extracting the features, one can choose proper classifiers for EEG classification. There are many classical classifiers to choose from. Classifiers frequently used in literature include the k-nearest neighbor (KNN), artificial neural network (ANN), decision tree (DT), random forest (RF) and support vector machine (SVM) (Martis *et al* 2013, Sharmila and Geethanjali 2016, Li *et al* 2017, Hussein *et al* 2018). Some researchers modified the classical classifiers; for example in Richhariya and Tanveer (2018), the authors proposed a novel model named universum SVM for EEG classification. Meanwhile, the authors in Jin *et al* (2018) proposed a sparse Bayesian extreme learning machine (ELM) to solve the EEG classification problem by integrating the Bayesian interference into the ELM. Some researchers tried to apply recent deep learning techniques to this field (Acharya *et al* 2018, Ullah *et al* 2018). However, training a deep model needs a large number of samples, which does not always work in reality. In this study, we provide an alternative means of seizure detection using a different feature extraction strategy that only involves one parameter and runs fast. In addition, using these features, a feedforward neural network (FNN) classifier which only has one hidden layer can achieve high accuracy. These characteristics make our system suitable for online recognition and big dataset operation.

The organization of this paper is as follows. The dataset used in this paper is described in section 2. We also present the literature review in this section. Section 3 contains the methodology. We describe the proposed method for EEG feature extraction. We also introduce the process of intrinsic time-scale decomposition (ITD) and explain how to obtain the instantaneous amplitude and instantaneous frequency of each proper rotation component (PRC). The descriptions of the statistical features are also presented. Section 4 is the experimental section. We compare our method with state-of-the-art methods. Finally, section 5 provides the conclusion of this paper.

2. Dataset description and literature review

2.1. Dataset description

The EEG dataset is taken from the public resource at the University of Bonn, which has been extensively used in epilepsy recognition⁴. The EEG signals were recorded by the same 128-channel amplifier system and digitized using a 12-bit analog-to-digital converter. The total dataset consists of EEG signals for healthy and epileptic subjects. It has five classes (A, B, C, D and E). Each class contains 100 single-channel EEG segments with sampling time 23.6 s and sampling rate 173.61 Hz. A summary is given in table 1. For a more detailed description of the dataset please refer to Andrzejak *et al* (2001).

2.2. Literature review

Transform is a common method in the problem of feature extraction. By means of a suitable transform, some discriminative features of EEG recordings can be extracted from the transform domain. In Polat and Günes (2007), the authors calculated power spectral density (PSD) by fast Fourier transform as EEG features and used the DT classifier for the two-class classification problem: A versus E. Finally, they obtained 98.68% classification accuracy using five-fold cross-validation and 98.72% accuracy using ten-fold cross-validation, respectively. In another study, the PSD of each EEG segment was calculated by means of short-time Fourier transform and time-frequency distributions for feature extraction (Tzallas *et al* 2009). Then an ANN classifier was used for EEG classification. For the two-class classification problem A versus E, they obtained an average accuracy of

⁴The data can be downloaded from <http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html>.

Table 1. Description of the public dataset.

	Set A	Set B	Set C	Set D	Set E
Subjects	Healthy	Healthy	Epileptic	Epileptic	Epileptic
State	Eyes open	Eyes closed	Interictal	Interictal	Ictal
Electrode placement	International 10–20 system	International 10–20 system	Opposite to epileptogenic zone	Within epileptogenic zone	Within epileptogenic zone
Number	100	100	100	100	100
Time duration	23.6 s	23.6 s	23.6 s	23.6 s	23.6 s
Sample rate	173.61 Hz	173.61 Hz	173.61 Hz	173.61 Hz	173.61 Hz

94.27%; for the three-class classification problem A versus C versus E, they obtained an average accuracy of 94.68%. More and more researchers are trying to apply wavelet transform to EEG feature extraction (Acharya *et al* 2012b, Li *et al* 2017). The authors in Acharya *et al* (2012b) applied the wavelet packet decomposition to EEG signals and extracted EEG features by calculating the eigenvalues from the obtained wavelet coefficients. A Gaussian mixture model classifier was used in their study and they reached 99% classification accuracy for the three-class classification problem A versus C versus E. Li *et al* (2017) considered the problem of detecting normal, interictal and epileptic signals, and the authors used discrete wavelet transform and envelope analysis for EEG feature extraction. The experimental results showed that the scheme achieved 98.78% classification accuracy when using the neural network ensemble as the classifier. Since the EEG recordings are typical non-linear and non-stationary signals (Acharya *et al* 2012a), the adaptive decomposition method can reveal the signal composition better. Hilbert Huang transform (HHT) is an adaptive signal transform and suitable for such signals (Huang *et al* 1998). The fundamental part of HHT is the empirical mode decomposition (EMD) method. Through this process, the input signal can be decomposed adaptively into subcomponents named intrinsic mode functions (IMFs) whose frequencies range from high to low. The instantaneous amplitude and instantaneous frequency of each IMF are calculated by Hilbert transform. Then the time-frequency distribution is obtained consequently. Many authors have analyzed the EEG signals using the EMD and Hilbert transform method (Li *et al* 2013, Pachori and Patidar 2014, Das and Bhuiyan 2016, Djemili *et al* 2016, Riaz *et al* 2016, Gaur *et al* 2018). Riaz *et al* (2016) used HHT to extract the temporal and spectral features. The authors in that study tested the performance of four classifiers (KNN, DT, ANN and SVM). It was found that the accuracies of these four classifiers were higher than 99% for dealing with the two-class EEG classification problem, but the results on the three-class EEG classification problem were less than satisfactory. Li *et al* (2013) also used the EMD method to extract EEG features. After obtaining the IMFs, the authors chose the first five IMFs and calculated their coefficients of variation and fluctuation as features. Finally, they achieved satisfactory classification accuracies for the two-class problem. Although the EMD technique is self-adaptive and used in many applications, it has several drawbacks, such as overenveloping (Yang *et al* 2013, 2014), mode mixing (Hu *et al* 2012) and end point effects (Zheng *et al* 2013), that bring down its practical utility. Instead, ITD, another self-adaptive decomposition method (Frei and Osorio 2007), uses a novel process of constructing the baseline. The ITD method can strictly control the end point effect at the end points and does not propagate the whole data. Moreover, the ITD method constructs the baseline signals through a piecewise linear function and the whole algorithm has monolevel iteration which makes the algorithm run fast. This method has been a useful tool in many recognition problems since it was proposed (Martis *et al* 2013, Duan *et al* 2016, Xing *et al* 2017). However, there are few articles about using the ITD method in EEG signal analysis. In Martis *et al* (2013), the authors used the ITD method to extract features for EEG signal classification. After this, they computed the energy, Higuchi fractal dimension and sample entropy of the PRCs as features. Then they fed the features into the DT classifier to verify the proposed features and the experimental results showed that the sensitivity, specificity and classification accuracy were all more than 90%. Empirical wavelet transform (EWT) is another decomposition technique, which is a fusion of wavelet transform and EMD, and also used in EEG feature extraction (Bhattacharyya and Pachori 2017, Bhattacharyya *et al* 2018). For example, the authors in Bhattacharyya and Pachori (2017) analyzed multichannel EEG recordings via EWT. The EEG recording in each selected channel was decomposed into ten mode components using EWT. Three characteristics were extracted from the joint instantaneous amplitude to form a feature vector. Finally, their model showed good performance. However, the EWT technique is relatively complex and has many parameters to be adjusted, which limit its running speed. Therefore, in this paper, we choose the ITD technique to decompose the EEG recordings. Meanwhile, it is noticed that Bhattacharyya and Pachori (2017) extracted the features from the joint instantaneous amplitude. Inspired by this, we consider both the instantaneous amplitudes and instantaneous frequencies and extract the EEG features from both of them for classification.

3. Methodology

Here we present the proposed method of feature extraction from EEG recordings. Our method can be boiled down to the following three steps.

Step 1 Using the ITD technique, each EEG recording is decomposed into several PRCs (more than five because of the complex neural system). Moreover, the EEG waveforms are often subdivided into five bandwidths depending on the range of frequencies (Tatum 2014). Based on this, we choose the first five PRCs for further feature extraction.

Step 2 The instantaneous amplitude and instantaneous frequency of each PRC is calculated. Then, we extract the statistical indices (mean, standard deviation, kurtosis and skewness) from the instantaneous amplitudes and frequencies of these PRCs. For each EEG signal, we combine all these indices of its five PRCs as the feature vector. Obviously, the feature vector has 40 dimensions for each EEG signal.

Step 3 We choose a classifier for the classification of EEG signals. In the experimental part, we tested multiple classifiers: KNN, DT, RF, SVM and FNN. We find that the FNN classifier achieves the optimal performance.

A brief illustration of our scheme is given in figure 1.

3.1. Intrinsic time-scale decomposition (ITD)

As a self-adaptive signal decomposition technique, ITD is applied to the field of signal analysis successfully (Martis *et al* 2013, Restrepo *et al* 2014, Xing *et al* 2017). Through the ITD process, the input signal can be decomposed adaptively into several PRCs, with frequencies ranging from high to low, and a trend component. In this subsection, we briefly review the ITD technique. For a more detailed description please refer to Frei and Osorio (2007).

For a given signal $x(t)$, let $\mathcal{L}$ be the baseline extracting operator, and let $\mathcal{H}$ be the PRC extracting operator. Then in the first step of ITD, $x(t)$ is decomposed into two components:

$$x(t) = \mathcal{L}x(t) + \mathcal{H}x(t) = L(t) + H(t), \quad (1)$$

where $L(t)$ is a baseline and $H(t)$ is a PRC.

After the first step, the process can be re-applied using the baseline signal as the new input signal. We iterate this procedure until the resulting baseline has only two extrema, or is a constant. Finally, the input signal $x(t)$ can be decomposed into a sequence of PRCs with decreasing instantaneous frequencies. If the iteration has S steps, the decomposition has the following form:

$$L^0(t) := x(t) = L^S(t) + \sum_{j=1}^S H^j(t). \quad (2)$$

PRCs and baselines satisfy

$$L^j(t) = L^{j+1}(t) + H^{j+1}(t), \quad j = 0, 1, 2, \dots, S. \quad (3)$$

Let $\{\tau_k^j, k = 1, 2, \dots, K\}$ be the extrema points of $L^j(t)$, and we define $\tau_0^j = 0$ for convenience. If there are several successive data points with the same extremal value, we take τ_k^j to the rightmost time of these extremal values. To simplify, we define $L_k^j := L^j(\tau_k^j)$. Then the baseline $L^{j+1}(t)$ is constructed by a piecewise linear formula: in the interval $t \in (\tau_k^j, \tau_{k+1}^j]$, between successive extrema,

$$L^{j+1}(t) = L_k^{j+1} + \frac{(L_{k+1}^{j+1} - L_k^{j+1})}{(L_{k+1}^j - L_k^j)} (L^j(t) - L_k^j), \quad (4)$$

where the knots in the above formula are

$$L_{k+1}^j := L^j(\tau_{k+1}^j) = \alpha \left[L_k^j + \left(\frac{\tau_{k+1}^j - \tau_k^j}{\tau_{k+2}^j - \tau_k^j} (L_{k+2}^j - L_k^j) \right) \right] + (1 - \alpha) L_{k+1}^j, \quad (5)$$

where α is a tunable parameter, and $0 < \alpha < 1$. In general, $\alpha = 1/2$.

The ITD technique is self-adaptive and can be used to analyze EEG signals, which are typically non-linear and non-stationary signals. Figure 2 illustrates the decomposition results of an EEG signal using the ITD method.

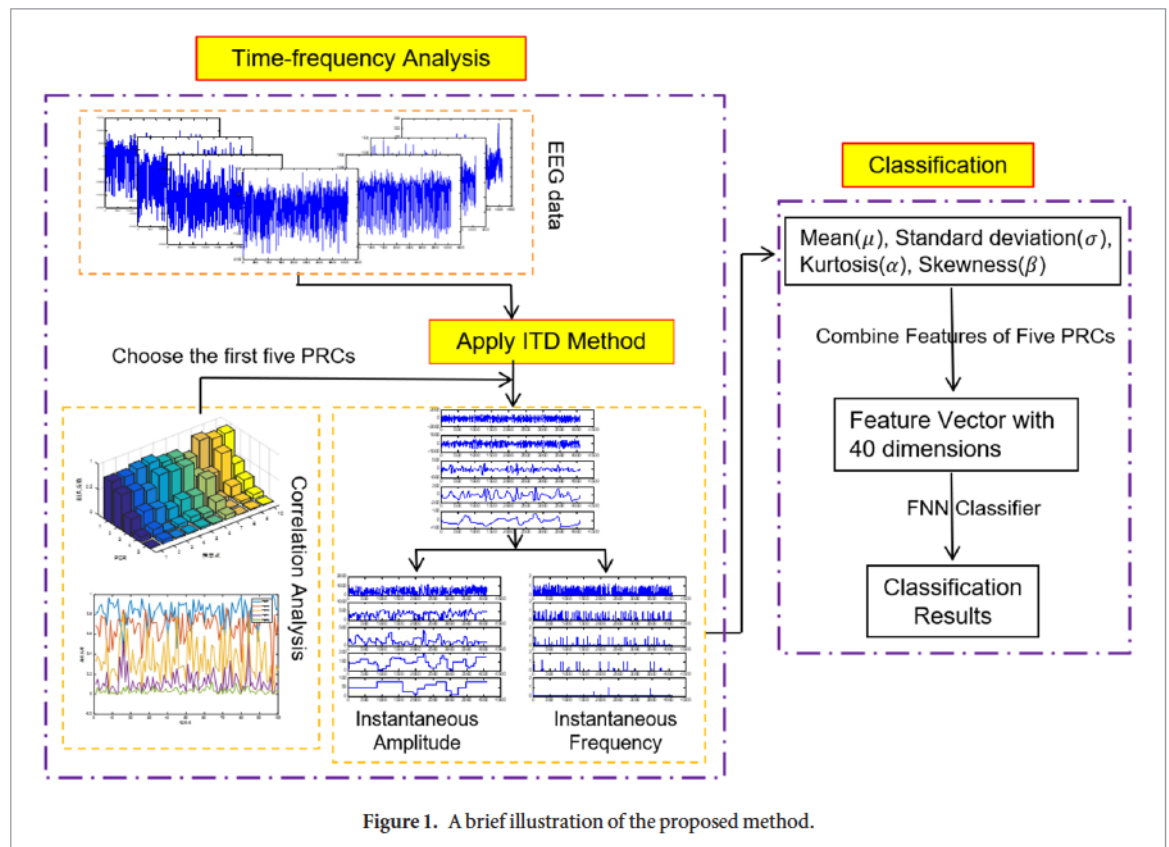


Figure 1. A brief illustration of the proposed method.

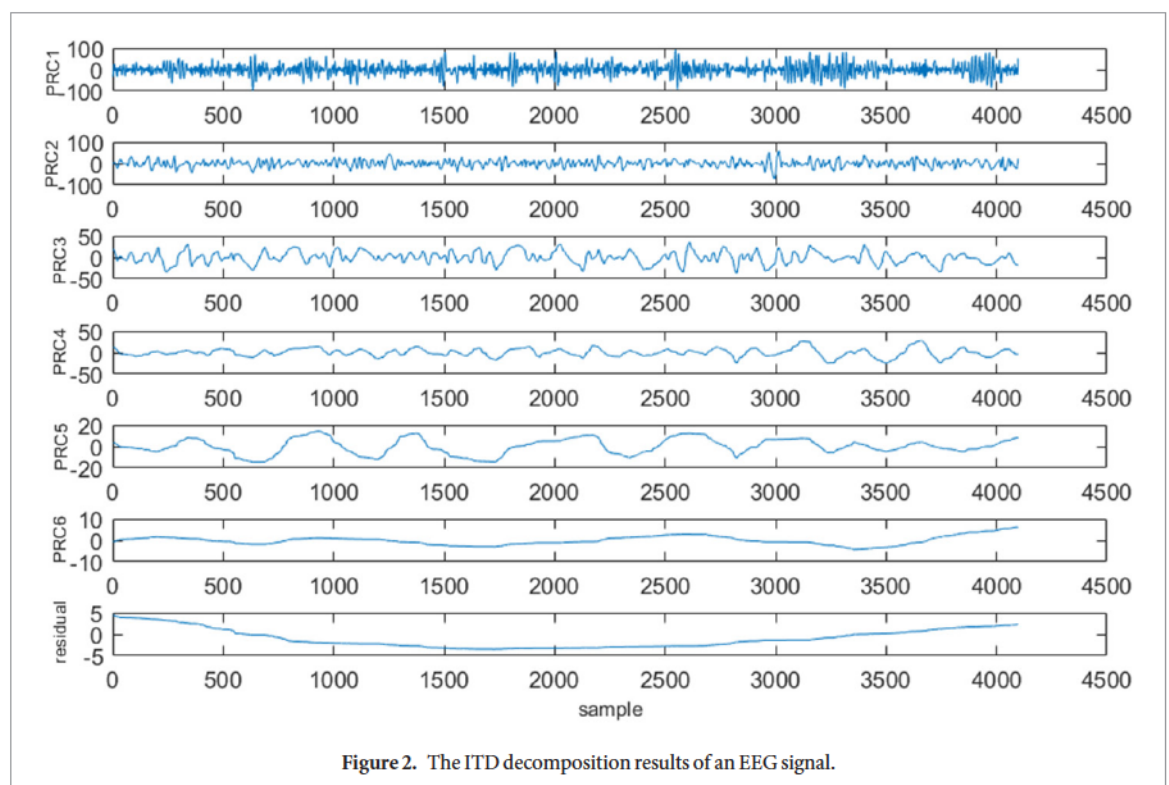
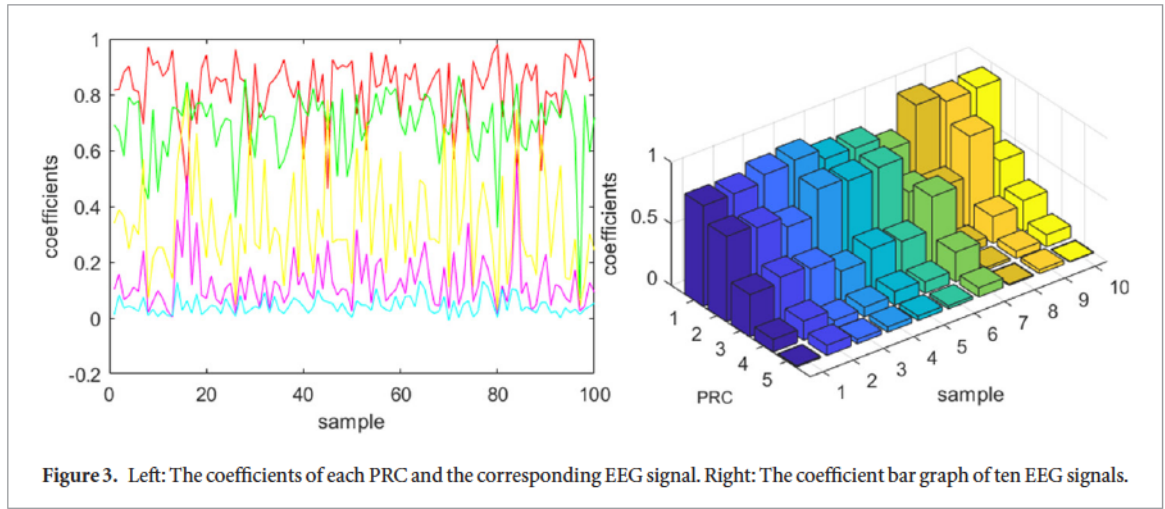


Figure 2. The ITD decomposition results of an EEG signal.

EEG signals are typically non-linear and non-stationary signals and obtain multiple modes of oscillation (Acharya *et al* 2012a). Therefore, each EEG signal has more than five components after ITD processing. From the procedure of the ITD method, we know the sum of these PRCs constitutes the input signal. Figure 3 illustrates the correlativity between each PRC and the original signal. The left figure plots the coefficients between the first five PRCs and the input EEG signal, and the right figure is the bar graph of these coefficients, in which we choose ten EEG signals. From figure 3 it can be seen that the correlations decrease in turn and the fifth coefficients are very small. This also implies that it is reasonable to choose the first five components for further analysis.



3.2. Instantaneous amplitude and instantaneous frequency

After having decomposed the input signal into a set of PRCs and a residual using ITD, the next step is to consider the instantaneous amplitude, instantaneous phase and instantaneous frequency for further analysis. Unlike in the HHT method, the authors in Frei and Osorio (2007) did not use Hilbert transform but instead used a wave-based method to calculate the instantaneous phase and instantaneous amplitude of PRCs, which guarantees a monotonically increasing phase angle. The formula of the instantaneous phase is as follows:

$$\theta(t) = \begin{cases} \left(\frac{X(t)}{A_1}\right) \frac{\pi}{2}, & t \in [t_1, t_2]; \\ \left(\frac{X(t)}{A_1}\right) \frac{\pi}{2} + \left(1 - \frac{X(t)}{A_1}\right) \pi, & t \in [t_2, t_3]; \\ \left(-\frac{X(t)}{A_2}\right) \frac{3\pi}{2} + \left(1 + \frac{X(t)}{A_2}\right) \pi, & t \in [t_3, t_4]; \\ \left(-\frac{X(t)}{A_2}\right) \frac{3\pi}{2} + \left(1 + \frac{X(t)}{A_2}\right) 2\pi, & t \in [t_4, t_5], \end{cases} \quad (6)$$

where t_1 and t_5 are two successive zero up-crossing points, $t_3 \in [t_1, t_5]$ is the zero down-crossing point, $t_2 \in [t_1, t_3]$ is the maximum point and $t_4 \in [t_3, t_5]$ is the minimum point. A_1 is the value at t_2 (i.e. the maximum on the positive half-wave) and $-A_2$ is the value at t_4 (i.e. the minimum on the negative half-wave). The instantaneous amplitude is defined as follows:

$$A(t) = \begin{cases} A_1, & t \in [t_1, t_3]; \\ A_2, & t \in [t_3, t_5]. \end{cases} \quad (7)$$

Obviously, $A(t)$ is a piecewise constant and determined by the extrema of the PRCs. Then, the instantaneous frequency can be calculated by the following formula:

$$f = \frac{1}{2\pi} \frac{d\theta}{dt}. \quad (8)$$

According to the decomposition results in figure 2, the instantaneous amplitudes and instantaneous frequencies of the first five PRCs are illustrated in figure 4.

3.3. Feature extraction

Once the PRCs of each EEG signal are obtained, we can calculate the instantaneous amplitudes and instantaneous frequencies of these PRCs, which contain a large amount of physiological and pathological information and can be used to extract the features. Since the EEG waveforms are often subdivided into five bandwidths depending on the range of frequencies (Tatum 2014), we choose the first five PRCs for further feature extraction. For each PRC, we extract the following statistical indices of each instantaneous amplitude and instantaneous frequency to contribute feature vectors for EEG classification:

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i, \quad (9)$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2}, \quad (10)$$

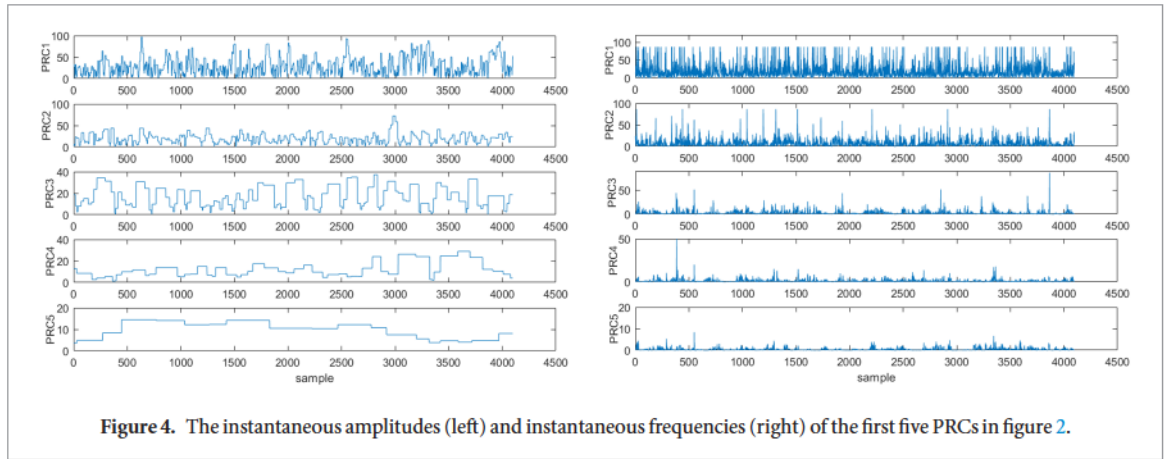


Figure 4. The instantaneous amplitudes (left) and instantaneous frequencies (right) of the first five PRCs in figure 2.

$$\alpha = \frac{1}{\sigma^3} \sum_{i=0}^{N-1} (x_i - \mu)^3, \quad (11)$$

$$\beta = \frac{1}{\sigma^4} \sum_{i=0}^{N-1} (x_i - \mu)^4, \quad (12)$$

where x_i corresponds to the i th value of the instantaneous amplitude or the instantaneous frequency of each PRC. N is the length of the signal.

The mean, standard deviation, kurtosis and skewness are important numerical features reflecting the distribution form. The reasons why we choose them for EEG classification are as follows.

- Mean μ reflects the arithmetic mean of the instantaneous amplitude or the instantaneous frequency. Observing the normal and epileptic EEG signals, we can find their ranges are different in most cases. Hence, the mean index can distinguish normal and abnormal signals to some extent and can be used as an EEG feature.
- Standard deviation σ represents the degree of dispersion of data points and is mathematically defined as the square root of the variance. Since the epileptic EEG signals have a more dramatic fluctuation, it is reasonable to have standard deviation as a feature.
- Skewness α is a measurement of reflecting the asymmetry of the data distribution. Right skewness (also called positive skewness) is represented by a long tail on the right side of the data. At this time, most values are distributed on the left side, and a small part of the values are distributed on the right side.
- Kurtosis β is an index used to measure the deviate degree of outlier data. The higher the kurtosis is, the more extreme values are in the data series. The different modes of oscillation between normal and abnormal EEG recordings make a significant difference to their kurtosis values, which motivates us to use the kurtosis index as a feature for EEG signal classification.

For each PRC of an EEG recording, we calculate its instantaneous amplitude and instantaneous frequency and then extract the four statistical features of them. Sequentially, we combine all these indices of the five PRCs as the feature vector. Obviously, the feature vector has 40 dimensions. For quantitative analysis, we apply the analysis of variance (ANOVA) on these extracted features for their statistical tests. The detailed results are shown in table 2. The tests compute an F measure (F-value) and a probability value (p-value) among the eight indices of the five sets A, B, C, D and E. The higher F-value and lower p-value show the good discrimination of the features. Since there are five PRCs for each EEG recording, we compute the F-values and p-values of each feature via arithmetic average. It can be seen from table 2 that the indices μ , σ of the instantaneous amplitudes and σ , α , β of the instantaneous frequencies provide good discrimination. Although the p-values of α , β of the instantaneous amplitudes and μ of the instantaneous frequencies are not lower, they work in a complementary fashion. The proposed method exploits this combination.

In addition, to illustrate the effectiveness of these features, we present two detailed examples. Considering the EEG signals in set A and E, figure 5 plots the box and scatter diagrams of the two features (skewness and kurtosis) of the instantaneous frequencies on these two classes. It is shown that the kurtosis and skewness values of the two different classes are significantly different and obviously separable, which confirms the statistical analysis in table 2.

Table 2. The ANOVA test based on the extracted features.

ANOVA	Features							
	Instantaneous amplitude				Instantaneous frequency			
	μ	σ	α	β	μ	σ	α	β
F-value	140.05	126.82	41.02	12.04	1.85	65.62	72.86	64.38
p-value	9.94×10^{-53}	6.38×10^{-54}	0.13	0.16	0.36	2.51×10^{-13}	1.46×10^{-13}	3.04×10^{-13}

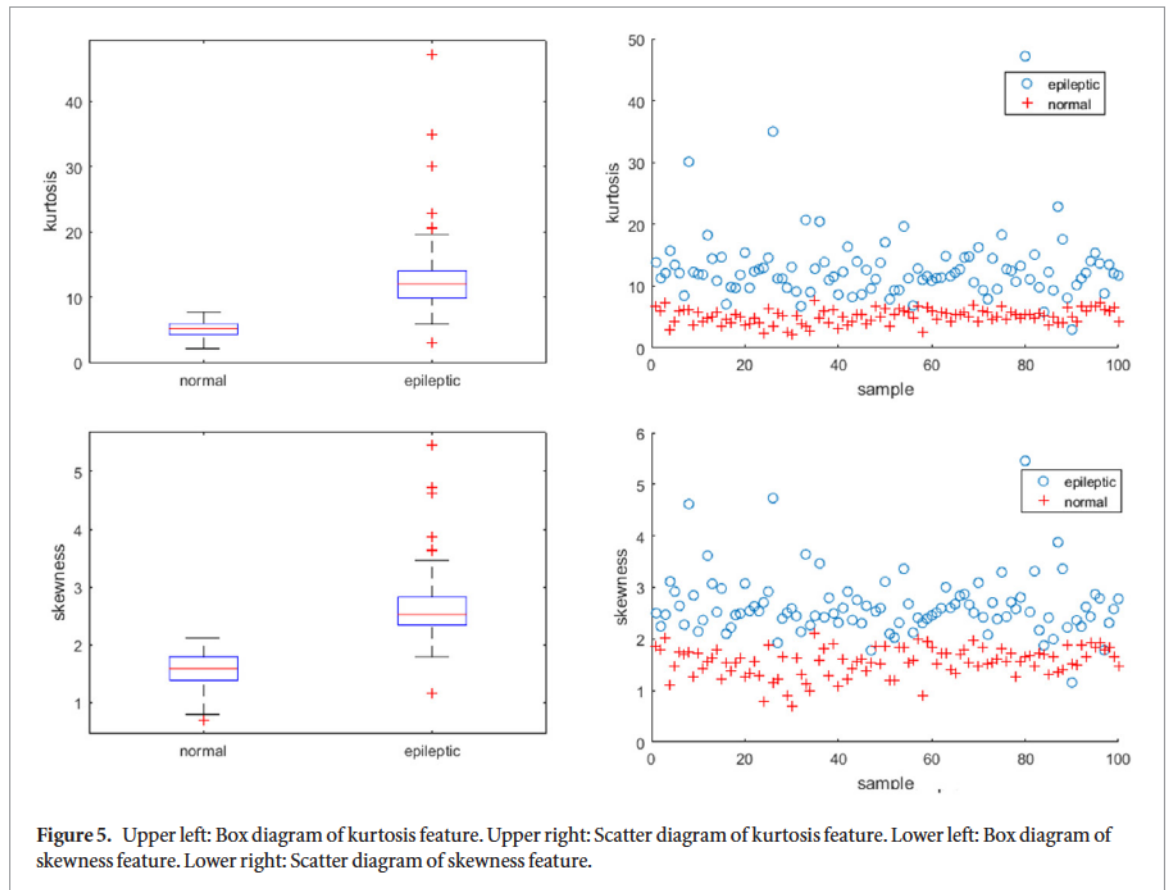


Figure 5. Upper left: Box diagram of kurtosis feature. Upper right: Scatter diagram of kurtosis feature. Lower left: Box diagram of skewness feature. Lower right: Scatter diagram of skewness feature.

Considering the first PRC of each EEG recording as an example, we randomly select 40 signals from A and E, separately. Figure 6 plots the features extracted from the instantaneous frequency of these PRCs. In the figure, we can easily see the feasibility of these features.

3.4. EEG classification

The classifier selection is the last step of EEG classification. In our study, we tested the performance of several classifiers, i.e. KNN, DT, RF, SVM and FNN. The results show that FNN achieves the optimal performance. Therefore, we will use the results based on the FNN classifier for the comparison with state-of-the-art methods (see table 9). In the experimental part, there is only one hidden layer with ten nodes in our FNN structure. The active function in the hidden layer is the logarithmic sigmoid transfer function and that in the output layer is the linear transfer function.

4. Results and discussion

Classification experiments are performed in this section. We adopt sensitivity, specificity and accuracy for evaluation. Most of the state-of-the-art methods for epilepsy detection also employ these metrics. Their definitions are given below:

$$\text{Sensitivity} = \frac{TP}{FN + TP}, \quad (13)$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \quad (14)$$

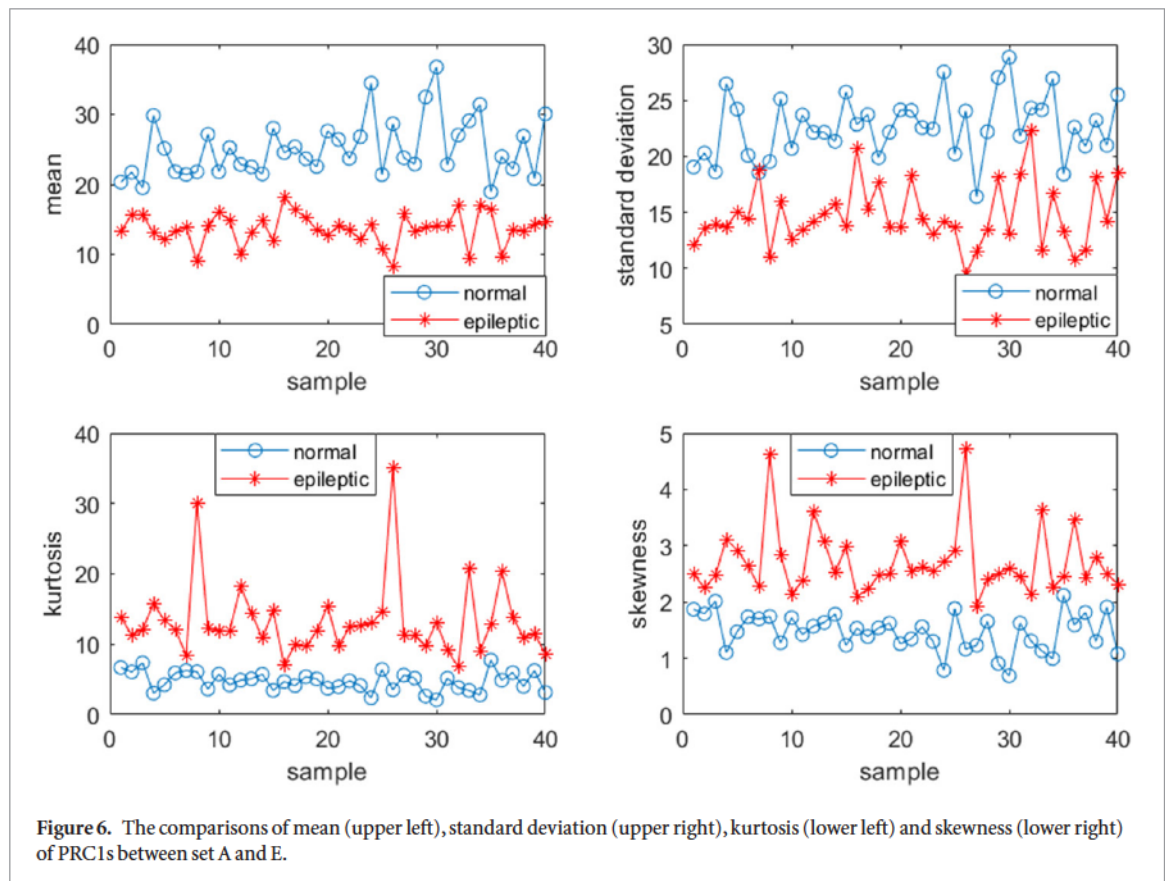


Figure 6. The comparisons of mean (upper left), standard deviation (upper right), kurtosis (lower left) and skewness (lower right) of PRC1s between set A and E.

$$Accuracy = \frac{TP + TN}{TotalSamples}, \quad (15)$$

where TP, true positive, represents the number of samples belonging to the positive class that are also identified as positive by the classifier; TN, true negative, represents the number of samples belonging to the negative class that are also identified as negative; FP, false positive, represents the number of samples belonging to the negative class that are identified as positive; and FN, false negative, represents the number of samples belonging to the positive class that are identified as negative.

For the five data sets of EEG recordings described in section 2, we have considered different experimental cases for classification: (1) non-seizure versus seizure (A versus E, B versus E, AB versus E, C versus E, D versus E, CD versus E, ABCD versus E); (2) normal versus epileptic (AB versus CD, AB versus CDE); (3) normal versus inter-ictal versus ictal (AB versus CD versus E, A versus C versus E). In the following experiments, we will abbreviate accuracy, sensitivity and specificity as Acc, Sen and Spe for convenience. All experiments will be performed using ten-fold cross-validation. All experiments are run using MATLAB R2017a on a laptop with Intel^(R) Core^(TM) i7-8550U CPU @1.80 GHz having 16 GB RAM.

4.1. Experiment 1: effectiveness of the proposed feature extraction model

In the first experiment, we test four classifiers (KNN, RF, SVM and FNN) with the extracted features based on the proposed model. We note that KNN, RF and FNN are trained by the built-in functions in MATLAB toolboxes, and SVM is trained by the Libsvm package. Table 3 gives the classification accuracies of these classifiers. It is shown that the features are effective in detecting normal and ictal EEG signals. Based on the extracted features, most classifiers show good performance. From table 3, we can see that the FNN classifiers achieve the optimal performance compared with the other three classifiers.

4.2. Experiment 2: comparison of two methods both using ITD

In Martis *et al* (2013), the authors also used the ITD method to extract features for EEG classification and the three-class problem A versus C versus E was considered. Different from the present study, these authors decomposed each EEG recording once using the ITD technique and obtained the high pass signal (the first PRC) and the low pass signal (the remaining signal). Then they computed the energy, Higuchi fractal dimension and sample entropy of the first PRC and the remaining signal as EEG features. They tested several classifiers and found the DT classifier achieved the best performance, yielding 99.0% sensitivity, 99.5% specificity and 95.67%

Table 3. Classification results using different classifiers with ten-fold cross-validation (Sen–Spe–Acc) (%).

Type of experiment	Classifiers			
	KNN	RF	SVM	FNN
A versus E	99–100–99.5	100–100–100	100–100–100	100–100–100
B versus E	99–98–98.5	100–100–100	100–100–100	100–100–100
AB versus E	100–98–99.33	99.5–100–99.67	100–100–100	100–100–100
C versus E	99–99–99	99–100–99.5	100–100–100	100–100–100
D versus E	95–94–94.5	98–96–97	95–96.7–95.5	100–100–100
CD versus E	96–97–96.67	98.5–97–98	97.24–97.31–97.33	100–98–99.33
AB versus CD	94–91–92.25	96.5–94.5–95.6	99.5–97.8–98.75	99–100–99.5
ABCD versus E	96.5–100–97.2	98.5–97–98.2	98.25–96.75–98.2	99–100–99.75
AB versus CDE	93–92.5–93.8	93.5–97.33–95.8	99.34–97.97–98.8	99.33–99–99.2

classification accuracy. This example is a three-class classification problem. In order to compute the sensitivity and specificity, which are concepts traditionally defined for two-class classification problems, researchers usually transfer a multiclass problem into several two-class problems and then obtain the sensitivity and specificity for the multiclass problem through the arithmetic average, which is exactly what we did here. We transferred the problem A versus C versus E into three two-class problems: A versus C, A versus E and C versus E. The values of sensitivity, specificity and accuracy for the two-class problems were computed and then these metrics of the problem A versus C versus E were obtained through the arithmetic average. Finally, we yielded 99.33% sensitivity, 99.67% specificity and 99.5% classification accuracy with the FNN classifier. We found that the classification results for the problem A versus C were less accurate compared with A versus E and C versus E. The detailed results are listed in table 4. In order to compare features proposed by Martis *et al* (2013) with our features, we also tested the DT classifier. When using the ten-fold cross-validation, we reached 98.33% sensitivity, 95.67% specificity and 97.0% classification accuracy. According to the mathematical formulas for sensitivity, specificity and accuracy, the value of accuracy should be varying between the values of sensitivity and specificity considering the data used. Our results in table 4 follow this pattern. Since (Martis *et al* 2013) did not provide the formulas that were used to compute their values, we only compare the final accuracies. Obviously, the classification accuracies of the proposed features are higher than that of Martis *et al* (2013) whether using the DT classifier or using FNN classifier. Although both Martis *et al* (2013) and this paper use the ITD technique in the process of feature extraction, the classification accuracies are different. Comparing the two processes of feature extraction we can see that the authors in Martis *et al* (2013) extracted the features directly from the first PRC and the baseline signal, while we extract the features from the instantaneous amplitude and the instantaneous frequency. Table 4 shows that our method achieves higher classification accuracy which verifies that there is indeed more useful information contained in the instantaneous amplitudes and frequencies.

4.3. Experiment 3: comparison of EWT, EMD and ITD

EWT, EMD and ITD are used to decompose non-stationary and non-linear signals. After the decomposition process, the instantaneous amplitude and instantaneous frequencies of each component can be calculated by different approaches. Among these three techniques, the ITD method is relatively simple since the baselines in it are based on a piecewise linear operation. Furthermore, there is only one layer of iteration and one parameter (α) in the ITD processing. Therefore, the ITD process has a faster decomposition speed. Table 5 compares the decomposition rate of the three methods. It shows that, compared with EWT and EMD, the ITD method can decompose more rapidly, at nearly two orders of magnitude faster. As the quantity of data increases, this advantage becomes more obvious.

In addition, we compare the performances of the three decomposition techniques (EWT, EMD and ITD) with the same statistical indices. This means that EWT, EMD and ITD separately decompose the EEG recordings and extract the same statistical indices as the EEG feature vectors from the instantaneous amplitude and instantaneous frequency of each corresponding component (modes, IMFs and PRCs). Finally, the three kinds of EEG feature vectors are fed into the same classifier for seizure detection. We present the classification results of the common classification problems: A versus E, C versus E, AB versus CDE and ABCD versus E. Table 6 shows the detailed results. Obviously, the combination strategy of ITD and statistical indices is optimal.

In Li *et al* (2013), every EEG signal was decomposed into several IMFs through the EMD process. The authors also used the first five components (IMFs) for feature extraction. They calculated the coefficients of variation and fluctuation indexed from the IMFs as features and fed them into the SVM classifier. In order to compare the effectiveness of the two feature extraction models, we also used the SVM classifier. In addition, like Li *et al* (2013), we also calculated the sensitivity and specificity when using the features of every single component or

Table 4. Several two-class classification results and the comparison with Martis *et al* (2013) (Sen–Spe–Acc) (%).

Type of experiment	Classifiers	
	DT	FNN
A versus C	96–92–94	98–99–98.5
A versus E	100–98–99	100–100–100
C versus E	99–97–98	100–100–100
A versus C versus E (Proposed method)	98.33–95.67–97	99.33–99.67–99.5
A versus C versus E (Martis <i>et al</i> (2013))	99–99.5–95.67	—

Table 5. Speed comparison of three decomposition methods: EWT, EMD and ITD.

Data length	Method	Decomposition time
4097(from data set A)	EWT	2.416 201 s
	EMD	0.175 299 s
	ITD	0.003 557 s
7305(from data of length-of-day)	EWT	11.059 461 s
	EMD	0.203 845 s
	ITD	0.005 567 s

Table 6. Classification results using EWT, EMD and ITD with a ten-fold cross-validation (Sen–Spe–Acc) (%).

Decomposition method	Type of experiments			
	A versus E	C versus E	AB versus CDE	ABCD versus E
EWT	99–97–98	99–95–97	81–83.33–82.4	99–99.75–99.38
EMD	99–100–99.5	98–97–97.5	95.5–97.33–96.6	98.75–100–99.38
ITD	100–100–100	100–100–100	99.33–99–99.2	99–100–99.75

their combination. All the comparison results are summarized in table 7. In this table, IMF1 means that the corresponding features are calculated solely over the single IMF1, IMF2 and IMF3 are similar, while IMF1–3 represents the combined features based on the first three IMFs used for EEG classification. From table 7, we can see that the combined features are indeed more effective than the features of single IMFs. The results are the same as features based on PRCs. However, our method always shows higher sensitivity and specificity, whether the features of a single component or the features of the combined components are used. For example, when the features of PRC1 are used, the sensitivity and specificity are 100% and 99.00%, respectively, which is obviously better than the previous results obtained by IMF1 (93.25% sensitivity and 96.90% specificity).

The authors in Djemili *et al* (2016) also used the EMD technique to preprocess the EEG signals before feature extraction. Instead of directly using EMD on EEG datasets, they divided each EEG recording into segments with a length of 256, and applied EMD on these segments. They chose the first four IMFs to calculate the minimum, maximum, standard deviation and the mean of absolute values as EEG features. Then the multilayer perceptron neural network classifier was used for seizure detection. They also tested the classification results using the features of every single IMF or their combination. Performance results are listed in table 8. It can be seen that, on the two-class problem (i.e. A versus E), regardless of whether the features of a single component or the features of the combined components are used, our scheme always shows the better performance. It is worth mentioning that the method in Djemili *et al* (2016) achieves a superior classification accuracy of 100% using the combined IMF1–4, while our model reaches the same superior classification accuracy only by using the combined PRC1–2. Table 8 also gives the classification results of the two-class problem (i.e. D versus E). The results in table 8 show that globally higher accuracies are obtained for the first component, no matter what decomposition method is used (EMD or ITD), and the classifier performance begins to decrease for the rest of the components, whereas the results improve when using features from a combination of components. It can be seen that the sensitivity, specificity and accuracy results obtained by ITD are superior to those obtained by EMD.

4.4. Experiment 4: comparison with published results

There are also lots of state-of-the-art studies using methods other than EMD or ITD to extract EEG features. In this subsection, a comparison will be shown between our method and several other studies without using EMD or ITD. These studies all use the same dataset. The comparison results are summarized in table 9. We can see that all the methods achieved superior accuracy 100% in the two-class problem normal versus seizure

Table 7. The performance comparison of components obtained by EMD (Li *et al* 2013) and ITD acting on the normal and ictal EEG data (A versus E). The classifier is SVM.

Method	Features on	Sen (%)	Spe (%)	Method	Features on	Sen (%)	Spe (%)
EMD	IMF1	93.25	96.90	ITD	PRC1	100	99.00
	IMF2	87.50	94.50		PRC2	100	99.33
	IMF3	85.25	94.40		PRC3	99.00	99.60
	IMF1-3	97.75	99.40		PRC1-3	100	100
	IMF1-4	97.75	99.40		PRC1-4	100	100
	IMF1-5	98.00	99.40		PRC1-5	100	100

Table 8. Performance comparison of components obtained by EMD (Djemili *et al* 2016) and ITD acting on the classification problems A versus E and D versus E.

Methods	Features on	A versus E			D versus E		
		Sen (%)	Spe (%)	Acc (%)	Sen (%)	Spe (%)	Acc (%)
EMD	IMF1	97.6	98.6	98.1	97.7	97.7	97.7
	IMF2	94.3	99.2	96.7	90.2	96.4	93.3
	IMF3	89.7	98.8	94.2	78.1	87.2	82.6
	IMF4	83.4	92.9	88.2	72.6	77.8	75.2
	IMF1-2	99.3	99.8	99.6	97.2	96.7	96.9
	IMF1-3	98.2	99.9	99.2	97.7	96.6	97.1
	IMF1-4	100	100	100	96.0	94.3	95.2
ITD	PRC1	100	99.0	99.5	98.0	97.0	97.5
	PRC2	100	100	100	94.0	97.0	95.5
	PRC3	98.0	99.0	98.5	90.0	89.0	89.5
	PRC4	92.0	87.0	89.5	86.0	82.0	84.0
	PRC1-2	100	100	100	98.0	99.0	98.5
	PRC1-3	100	100	100	99.0	97.0	98.0
	PRC1-4	100	100	100	96.0	99.0	97.5

(A versus E). Sharma *et al* (2017) employed analytic time-frequency flexible wavelet transform (ATFFWT) to deal with EEG recordings, and calculated each sub-band's fractal dimension as EEG features. The chosen classifier is least-squares SVM (LS-SVM) which has a superior performance in the two-class problems A versus E, B versus E and AB versus E, all with an accuracy of 100%. However, in other binary class problems, the performance of the model in Sharma *et al* (2017) is reduced. Notably, it reached the accuracy of 92.5% in the experiment AB versus CD while our proposed model yielded a superior accuracy of 99.5%. Swami *et al* (2016) decomposed the EEG signals using dual-tree complex wavelet transform (DTCWT) to extract features and used the general regression neural network (GRNN) for classification. Their classification accuracies vary from 93.3% to 99.2% with the exception of the A versus E experiment, while our accuracies all are higher than theirs, which shows that our model outperforms theirs. The authors in Sharmila and Geethanjali (2016) employed discrete wavelet transform (DWT) to extract EEG features and naive Bayes (NB) and KNN for EEG classification. In table 9, we can see that their accuracies vary from 96.4% to 100% for the different classification cases, while the proposed model achieves the better performance in all these cases. Most studies focus on the two-class problem to detect seizures (such as A versus E, ABCD versus E, etc.). However, there is less research on recognizing normal, interictal and ictal EEG recordings (Zhang *et al* 2017). The authors in Zhang *et al* (2017) considered three-class problems (AB versus CD versus E). They extracted EEG features by fusing variational mode decomposition (VMD) and auto-regression (AR), which achieved an accuracy of 97.35% when using the RF classifier. Our model achieves 1.4% higher accuracy than that of Zhang *et al* (2017) in this three-class problem. Ullah *et al* (2018) constructed a pyramidal 1D-CNN (P-1D-CNN) which contained three convolution layers for EEG classification. Since CNN needs a large amount of data, the authors in Ullah *et al* (2018) used a window sliding (WS) through the EEG data for augmentation. Finally, they tested the performances of a single P-1D-CNN and ensemble of three P-1D-CNNs. The accuracies of Ullah *et al* (2018) in table 9 were obtained by using the P-1D-CNN ensemble, which was better than single P-1D-CNN. Compared with their results, our accuracies are higher in 11 of the 17 classification problems. Moreover, the proposed model in this study is simple and runs fast, so it is suitable for online detection and also provides a diagnostic aid for physicians in their practice.

Table 9. Performance comparison of the proposed model with state-of-the-art studies.

Type of experiment	Method used	State-of-the-art	Acc (%)	Our Acc (%)
A versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	100	100
	DWT+NB/KNN	Sharmila and Geethanjali (2016)	100	
	DTCWT+GRNN	Swami <i>et al</i> (2016)	100	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	100	
B versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	100	100
	DTCWT+GRNN	Swami <i>et al</i> (2016)	98.9	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.8	
AB versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	100	100
	DTCWT+GRNN	Swami <i>et al</i> (2016)	99.2	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.8	
C versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	99	100
	DTCWT+GRNN	Swami <i>et al</i> (2016)	98.7	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.1	
D versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	98.5	100
	DTCWT+GRNN	Swami <i>et al</i> (2016)	99.3	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.4	
CD versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	98.8	99.33
	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	98.7	
	DTCWT+GRNN	Swami <i>et al</i> (2016)	95.2	
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.7	
AB versus CD	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	92.5	99.5
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.9	
AC versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	99.6	100
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.7	
BC versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	98.3	99.67
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.5	
BD versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	96.5	98.67
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.6	
BCD versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	96.4	99.5
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.3	
ABCD versus E	ATFFWT+LS-SVM	Sharma <i>et al</i> (2017)	99.2	99.75
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.7	
	FT+MLP	Samiee <i>et al</i> (2015)	98.1	
AB versus CDE	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.5	99.2
ABC versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	98.7	100
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.97	
ACD versus E	DWT+NB/KNN	Sharmila and Geethanjali (2016)	97.3	99.67
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.8	
A versus C versus E	ITD+DT	Martis <i>et al</i> (2013)	95.67	99.5
AB versus CD versus E	VMD+AR+RFTQWT	Zhang <i>et al</i> (2017)	97.4	98.8
	WS+P-1D-CNN Ensemble	Ullah <i>et al</i> (2018)	99.1	

5. Conclusion

This paper provides an alternative seizure detection strategy. The proposed model extracts the features based on the ITD technique and detects seizures using the FNN classifier. The ITD is a self-adaptive decomposition technique and suitable for analyzing non-linear and non-stationary signals (such as EEG signals). Using ITD every EEG recording is decomposed into several PRCs. Then the inherent information in the EEG signals can be obtained from the instantaneous amplitudes and instantaneous frequencies of PRCs. Therefore, we extract the statistical indices mean, standard deviation, kurtosis and skewness from these instantaneous amplitudes and instantaneous frequencies as EEG feature vectors to detect normal, interictal and ictal EEG recordings. Experimental results show a good performance of the proposed system. Compared with the latest references, our model reaches a comparable or even better detection accuracy. The proposed model is particularly effective in detecting normal and ictal EEG recordings, and achieves superior sensitivity of 100%, superior specificity

of 100% and superior classification accuracy of 100% from the public dataset. The ITD method only contains monolayer iteration, and the baseline is based on a piecewise linear model, which makes the ITD algorithm run very fast and makes it suitable for big data processing and online processing. Therefore, the proposed system has great potential in real-time seizure detection.

In the future, we will continue our research on this paper in two aspects. Firstly, we will consider extending the proposed strategy to the classification problems of multichannel EEG recordings. Secondly, the robustness of the proposed method will be tested since EEG recordings are often disturbed by various noises in their acquisition process. Through the experiments we conducted in this research, we feel that the neural network can be an effective tool for medical signal classifications. However, different from the typical applications of neural networks in machine learning where a large amount of data is available, obtaining large-scale medical signals for training is practically challenging. This is a significant burden. One of the future research projects that we plan to tackle is designing strategies to train the neural networks with relatively small datasets.

Acknowledgments

This research was partially supported by the NSFC (Grant Nos. 11701144 and 11471101), the Higher Education School Young Backbone Teacher Training Program of Henan Province (Grant No. 2017GGJS020), the Key Scientific Research Projects of The Educational Department of Henan Province (Grant No. 16A120002), NSF of Henan Province (Grant No. 162300410061) and the Project of Emerging Interdisciplinary of Henan University (Grant No. xxjc20170003). Haomin Zhou was partially supported by NSF awards DMS-1419027 and DMS-1620345, and ONR award N000141310408.

Interest disclosure

The authors declare no conflict of interest.

References

- Acharya U R, Molinari F, Sree S V, Chattopadhyay S, Ng K H and Suri J S 2012a Automated diagnosis of epileptic EEG using entropies *Biomed. Signal Process. Control* **7** 401–8
- Acharya U R, Sree S V, Alvin A P C and Suri J S 2012b Use of principal component analysis for automatic classification of epileptic EEG activities in wavelet framework *Expert Syst. Appl.* **39** 9072–8
- Acharya U R, Oh S L, Hagiwara Y, Tan J H and Adeli H 2018 Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals *Comput. Biol. Med.* **100** 270–8
- Andrzejak R G, Lehnertz K, Mormann F, Rieke C, David P and Elger C E 2001 Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state *Phys. Rev. E* **64** 061907
- Bhattacharyya A and Pachori R B 2017 A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform *IEEE Trans. Biomed. Eng.* **64** 2003–15
- Bhattacharyya A, Sharma M, Pachori R B, Sircar P and Acharya U R 2018 A novel approach for automated detection of focal EEG signals using empirical wavelet transform *Neural Comput. Appl.* **29** 47–57
- Das A B and Bhuiyan M I H 2016 Discrimination and classification of focal and non-focal EEG signals using entropy-based features in the EMD-DWT domain *Biomed. Signal Process. Control* **29** 11–21
- Djemili R, Bourrouba H and Korba M C A 2016 Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals *Biocybernetics Biomed. Eng.* **36** 285–91
- Duan L X, Yao M C, Wang J J, Bai T B and Yue J J 2016 Integrative intrinsic time-scale decomposition and hierarchical temporal memory approach to gearbox diagnosis under variable operating conditions *Adv. Mech. Eng.* **8** 1–14
- Frei M G and Osorio I 2007 Intrinsic time-scale decomposition: time-frequency-energy analysis and real-time filtering of non-stationary signals *Proc. R. Soc. A* **463** 321–42
- Gaur P, Pachori R B, Wang H and Prasad G 2018 A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry *Expert Syst. Appl.* **95** 201–11
- Hosseini M P, Hajisami A and Pompili D 2016 Real-time epileptic seizure detection from EEG signals via random subspace ensemble learning *Proc. 2016 IEEE Int. Conf. on Autonomic Computing* pp 209–18
- Hu X Y, Peng S L and Huang W L 2012 EMD revisited: a new understanding of the envelope and resolving the mode-mixing problem in AM-FM signals *IEEE Trans. Signal Process.* **60** 1075–86
- Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C and Liu H H 1998 The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis *Proc. R. Soc. A* **454** 903–95
- Hussein R, Elgendi M, Wang J and Ward R K 2018 Robust detection of epileptic seizures based on L_1 -penalized robust regression of EEG signals *Expert Syst. Appl.* **104** 153–67
- Jin Z, Zhou G, Gao D and Zhang Y 2018 EEG classification using sparse Bayesian extreme learning machine for brain-computer interface *Neural Comput. Appl.* (<https://doi.org/10.1007/s00521-018-3735-3>)
- Li M Y, Chen W Z and Zhang T 2017 Classification of epilepsy EEG signals using DWT-based envelope analysis and neural network ensemble *Biomed. Signal Process. Control* **31** 357–65
- Li Y, Yu Z, Bi N, Xu Y, Gu Z and Amari S 2014 Sparse representation for brain signal processing *IEEE Signal Process. Mag.* **31** 96–106
- Li S F, Zhou W D, Yuan Q and Cai D M 2013 Feature extraction and recognition of ictal EEG using EMD and SVM *Comput. Biol. Med.* **43** 807–16

- Martis J M, Acharya U R, Tan J H, Petznick A, Tong L, Chua C K and Ng E Y K 2013 Application of intrinsic time-scale decomposition (ITD) to EEG signals for automated seizure prediction *Int. J. Neural Syst.* **23** 1350023
- Minasyan G R, Chatten J B, Chatten M J and Harner R N 2010 Patient-specific early seizure detection from scalp EEG *J. Clin. Neurophysiol.* **27** 163–78
- Nolan M A, Redoblado M A, Lah S, Sabaz M, Lawson J A, Cunningham A M, Bleasel A F and Bye A M E 2004 Memory function in childhood epilepsy syndromes *J. Paediatrics Child Health* **40** 20–7
- Pachori R B and Patidar S 2014 Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions *Comput. Methods Programs Biomed.* **113** 494–502
- Polat K and Günes S 2007 Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform *Appl. Math. Comput.* **187** 1017–26
- Restrepo J M, Venkataramani S, Comeau D and Flaschka H 2014 Defining a trend for time series using the intrinsic time-scale decomposition *New J. Phys.* **16** 085004
- Riaz F, Hassan A, Rehman S, Niazi I K and Dremstrup K 2016 EMD-based temporal and spectral features for the classification of EEG signals using supervised learning *IEEE Trans. Neural Syst. Rehabil. Eng.* **24** 28–35
- Richhariya B and Tanveer M 2018 EEG signal classification using universum support vector machine *Expert Syst. Appl.* **106** 169–82
- Samiee K, Kovacs P and Gabbouj M 2015 Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform *IEEE Trans. Biomed. Eng.* **62** 541–52
- Sharma M, Pachori R and Acharya U 2017 A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension *Pattern Recognit. Lett.* **94** 172–9
- Sharmila A and Geethanjali P 2016 DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers *IEEE Access* **4** 7716–27
- Swami P, Gandhi T, Panigrahi B, Tripathi M and Anand S 2016 A novel robust diagnostic model to detect seizures in electroencephalography *Expert Syst. Appl.* **56** 116–30
- Tatum W O 2014 Ellen R. Grass lecture: extraordinary EEG *Neurodiagn. J.* **54** 3–21
- Tzallas A T, Tsipouras M G and Fotiadis D I 2009 Epileptic seizure detection in EEGs using time-frequency analysis *IEEE Trans. Inf. Technol. Biomed.* **13** 703–10
- Ullah I, Hussain M, Qazi E and Aboalsamh H 2018 An automated system for epilepsy detection using EEG brain signals based on deep learning approach *Expert Syst. Appl.* **107** 61–71
- WHO 2019 'Epilepsy: fact sheets, detail' World Health Organization (Accessed: 20 June 2019) www.who.int/news-room/fact-sheets/detail/epilepsy
- Xing Z Q, Qu J F, Chai Y, Tang Q and Zhou Y M 2017 Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine *J. Mech. Sci. Technol.* **31** 545–53
- Yang L J, Yang Z H, Yang L H and Zhang P 2013 An improved envelope algorithm for eliminating undershoots *Digital Signal Process.* **23** 401–11
- Yang L J, Yang Z H, Zhou F and Yang L H 2014 A novel envelope model based on convex constrained optimization *Digital Signal Process.* **29** 138–46
- Zhang Y, Zhou G, Jing J, Zhao Q, Wang X and Cichocki A 2016 Sparse Bayesian classification of EEG for brain-computer interface *IEEE Trans. Neural Netw. Learn. Syst.* **27** 2256–67
- Zhang Y, Zhang H, Chen X, Liu Ming, Zhu X, Lee S and Shen D 2019 Strength and similarity guided group-level brain functional network construction for MCI diagnosis *Pattern Recognit.* **88** 421–30
- Zhang T, Chen W and Li M 2017 AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier *Biomed. Signal Process. Control* **31** 550–9
- Zheng J D, Cheng J S and Yang Y 2013 Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis *Mech. Syst. Signal Process.* **40** 136–53
- Zhou S M, Gan J Q and Sepulveda F 2008 Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface *Inf. Sci.* **178** 1629–40
- Zhou G, Zhao Q, Zhang Y, Adali T, Xie S and Cichocki A 2016 Linked component analysis from matrices to high-order tensors: applications to biomedical data *Proc. IEEE* **104** 310–31