Expert Systems with Applications 40 (2013) 5477–5489

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Automatic recognition of epileptic EEG patterns via Extreme Learning Machine and multiresolution feature extraction

Yuedong Song [a,*], Jiaxiang Zhang [b]

[a] University of Cambridge, Computer Laboratory, JJ Thomson Avenue, Cambridge, United Kingdom
[b] MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

ARTICLE INFO

Keywords:
Epilepsy diagnosis
Electroencephalogram (EEG)
Multiresolution analysis
Feature extraction
Genetic algorithm (GA)
Extreme Learning Machine (ELM)

ABSTRACT

Epilepsy is one of the most common neurological disorders- approximately one in every 100 people worldwide are suffering from it. In this paper, a novel pattern recognition model is presented for automatic epilepsy diagnosis. Wavelet transform is investigated to decompose EEG into five EEG frequency bands which approximate to delta (δ), theta (θ), alpha (α), beta (β), and gamma (γ) bands. Complexity based features such as permutation entropy (PE), sample entropy (SampEn), and the Hurst exponent (HE) are extracted from both the original EEG signals and each of the frequency bands. The wavelet-based methodology separates the alterations in PE, SampEn, and HE in specific frequency bands of the EEG. The effectiveness of these complexity based measures in discriminating between normal brain state and brain state during the absence of seizures is evaluated using the Extreme Learning Machine (ELM). It is discovered that although there exists no significant differences in the feature values extracted from the original EEG signals, differences can be recognized when the features are examined within specific EEG frequency bands. A genetic algorithm (GA) is developed to choose feature subsets that are effective for enhancing the recognition performance. The GA is also examined for weight alteration for both sensitivity and specificity. The results show that the abnormal EEG diagnosis rate of the model without the involvement of the genetic algorithm is 85.9%. However, the diagnosis rate of the model increases to 94.2% when the genetic algorithm is integrated as a feature selector.

1. Introduction

Epilepsy is a sudden and recurrent brain malfunction associated with an enormous and hypersynchronous activity of the nerve cells. Approximately one in every 100 individuals is suffering from epilepsy (Iasemidis et al., 2003). An epileptic seizure is characterized by paroxysmal occurrence of synchronous oscillations, which is mainly divided into two classes in terms of the extent of connection of different brain fields: partial seizures and generalized seizures. Partial seizures begin from a circumscribed field of the brain, usually called the epileptic foci. Determined by their type, they may or may not impair consciousness. Generalized seizures involve most fields of the brain and may cause loss of consciousness and muscle contractions or stiffness.

Electroencephalography (EEG) is a non-invasive, low-cost and effective technique for examining electrical activity of the brain and diagnosing brain diseases in clinical setting (Stam, Pijn, Suffczynski, & Lopez da Silva, 1999). Recent developments in signal processing and data mining studies provide the possibility of extracting hidden patterns or associations in EEGs for better comprehension of brain functions from a system perspective. Automatic epileptic seizure analysis is generally modeled as an abnormal EEG recognition problem, which has important clinical implications. An automatic computer model which is able to discriminate between interictal EEG recordings and normal EEG recordings can be applied for diagnosing epilepsy in a clinical environment. However, in a clinical environment, the discriminations among these various EEG groups are usually not well defined (Adeli, Ghosh-Dastidar, & Dadmehr, 2010). Hence, it is crucial that the developed approaches are capable of recognizing EEG signals acquired from various conditions precisely and efficiently. Because of the differences of dynamic properties in EEGs under different brain states, the diagnosis and detection of brain disorders in EEGs can be regarded as pattern recognition problems, in which the signal could be spatial, temporal or spatio-temporal samples in feature spaces.

Contrary to ictal brain activity where nerve cells fire as rapid as about 500 times per second, nerve cells can only fire 80 times per second in the healthy brain. The pattern of neuronal firing steadily evolves from a normal state, first to an interictal state and then to onset of a seizure (Lopes da Silva, Pijn, & Wadman, 1994). The physiology behind the epileptiform discharges are attributed to chronic dysfunction or defect in the epileptic brain, that is to say,

* Corresponding author. Tel.: +44 07900274262.
E-mail address: syd918@gmail.com (Y. Song).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.04.025

the epileptic brain is not normal even in the period of seizure free. Epileptic seizures appear if the abnormal epileptogenic tissues in the brain remain untreated (Bromfield, 2012; Nurujjaman, Narayanan, & Iyengar, 2009). Diagnosis of neurological disorders is on the basis of the recognition of typical pathological patterns in EEG time series signals. Epilepsy could be caused by some pathological procedures of genetic or acquired origins. The occurrence of epileptiform activities such as spike waves and spike-and-slow waves in EEG signals supports the diagnosis of epilepsy (Wang, Zou, Zhang, Wang, & Wang, 2010). However, discriminating between normal brain activities and epileptiform activities for epilepsy diagnosis can be challenging, even from a visual inspection of the EEG by an experienced neurologist due to excessive presence of artifacts, interference or overlapping symptomatology with other neurological disorders. Even the most highly trained neurology experts are not able to differentiate between interictal EEG signals of epileptic patients and normal EEGs with over 80% of accuracy (Adeli et al., 2010). Overcoming these clinical difficulties through effective and robust automatic algorithms may have a great impact on the diagnosis and treatment of epilepsy.

In the last few years, a variety of signal processing approaches have been proposed for the purpose of epilepsy diagnosis. These approaches can be divided into the fast Fourier transform based approach (Polat & Güneş, 2007), frequency domain based approach (Nigam & Graupe, 2004; Srinivasan, Eswaran, & Sriraam, 2005; Übeyli, 2010), principle component analysis based approach (Polat & Güneş, 2008), crosscorrelation function based approach (Chandaka, Chatterjee, & Munshi, 2009) and entropy based approach (Kannathal, Acharya, Lim, & Sadasivan, 2005). However, most of the above studies developed approaches which can only be utilized to discriminate between normal brain state and brain state during seizure onset, which is much easier but clinically less practical than the problem of discrimination between normal brain state and the brain state during seizure free. So far, very limited studies have been conducted to develop quantitative signal processing approaches which can be utilized to differentiate between normal and interictal EEG signals. Wang et al. (2010) proposed a method for feature extraction and recognition of epileptiform activity in EEG signals, which combined principle component analysis and approximate entropy to reduce feature dimensions of EEG time series signals and to decorrelate normal EEG signals and interictal EEG signals. Neyman–Pearson criteria were investigated to discriminate these two types of EEG signals. Distinct difference was discovered between the approximate entropy values of epileptic and normal EEG signals.

The EEG time series signals are non-stationary, frequency alterations with time. Although EEG signals hold a great amount of information, only partial EEG elements are clinically useful. Noises may appear in background EEGs in the form of high frequency oscillations due to electromagnetic interference as well as very low frequency artifacts from eye blinks, muscle movements and electrocardiograms. Signal processing techniques are utilized for removing noises as well as extracting meaningful information from EEGs. The conventional signal analysis on the basis of fast Fourier transform (FFT) is only capable of acquiring frequency information, and no transient characteristics in FFT coefficients exist (Kiymi, Güler, Dizibüyük, & Akin, 2005). Wavelet transform (WT) is a powerful tool for time–frequency representation of time series signals. The appealing characteristic of wavelet transform is that it can capture appropriate frequency information at low frequencies as well as appropriate time information at high frequencies. This characteristic is useful in biomedical practice, since the majority of signals in this field, such as electroencephalography and electrocardiography, always contain low frequency elements during long time periods and high frequency elements during short time periods. Wavelet transform is particular suitable for analysis of non-stationary signals and can be employed to localize transient variations, which often appear during the period of seizure onset. The wavelet transform utilizes a multi-resolution analysis approach for decomposing EEG signals into several frequency bands. By means of WT, transient events in EEGs are accurately captured and mapped in both time and frequency framework. Because of these advantages, wavelet transform has been thought to be an effective method for EEG signal analysis (Adeli, Zhou, & Dadmehr, 2003). Complicated non-linear systems like EEG time series usually exhibit non-systematic, superficially random patterns, but actually may contain deterministic chaos. Deterministic chaos is different from randomness since it obeys a specific rule that controls the non-linearity of the system. The discriminability of the system relies on this underlying rule.

In this paper, a wavelet based pattern recognition model is proposed for analysis of EEG time series signals for epilepsy diagnosis. It is composed of four stages: (1) wavelet transform, (2) complexity estimation, (3) recognition model evaluation, (4) feature subset optimization. Each EEG is decomposed into five EEG frequency bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (30–60 Hz) by means of wavelet based filters. With the goal of recognizing deterministic rules in the system, the non-linear parameters of the original EEG signals are extracted in the form of permutation entropy (relating to system order structure), sample entropy (relating to system regularity), and the Hurst exponent (relating to system autocorrelation). The effectiveness of these non-linear features in identifying an abnormal brain state from normal brain states is examined on the basis of recognition model evaluation. Finally, the feature set is further selected by a genetic algorithm for performance optimization. The whole framework of the proposed methodology is showed in Fig. 1.

2. Theory and methods

2.1. Dataset

The EEG time series data employed in this study come from the Department of Epileptology, Bonn University, Germany (Andrzejak et al., 2001) and were used here to develop signal processing approaches for epilepsy diagnosis. The complete EEG data were taken from five healthy subjects and five epileptic patients. A total of 100 single-channel EEG datasets of 23.6 s duration were recorded. Set Z was taken from surface EEG signals recorded using extracranial electrodes of five healthy volunteers with eyes open. Major artefacts, for example, due to eye or hand movements, have been manually removed by the authors of the dataset. Set F originated from the EEG archive recorded using intracranial electrodes of presurgical diagnosis of five epilepsy patients. Signals in Set F were recorded from within the epileptogenic zone and it contains only brain activity measured during seizure free intervals. All EEG signals were recorded with the same 128-channel amplifier. The data were digitized at 173.6 samples per second at 12-bit resolution. Fig. 2 describes the electrode placement for recording of normal EEG signals. Fig. 3 describes examples of EEG signals for Set Z and Set F. A summary of the EEG data set is shown in Table 1.

2.2. Wavelet transform

The wavelet is a smooth and rapidly vanishing oscillating function with good localization in both time and frequency. A wavelet family $\psi_{a,b}$ is the set of elementary functions produced by dilations and translations of a unique mother wavelet $\psi(t)$:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}}\psi\left(\frac{t-b}{a}\right) \quad (1)$$

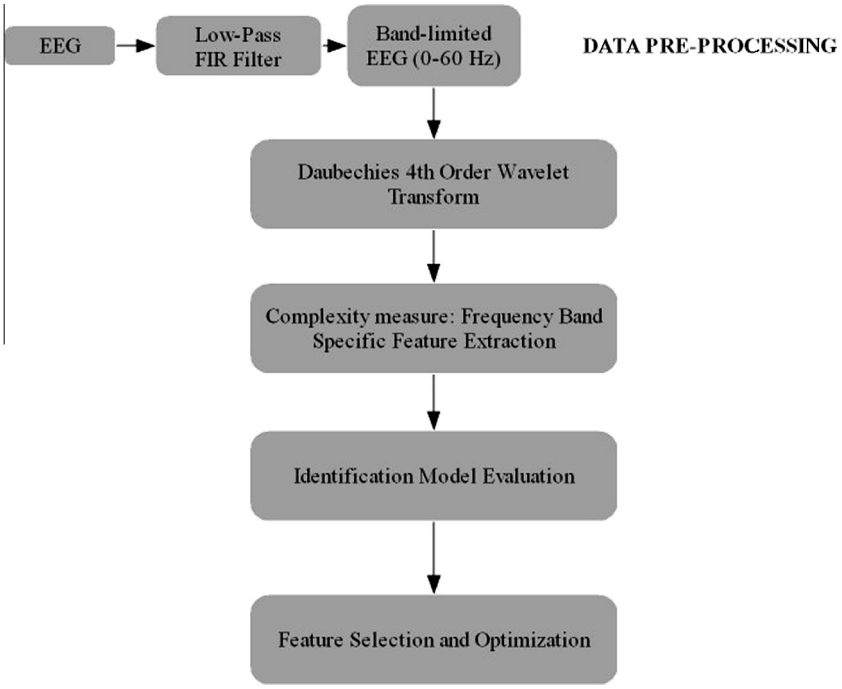

Fig. 1. Overview of the wavelet based complexity-genetic algorithm model for the abnormal EEG recognition problem.

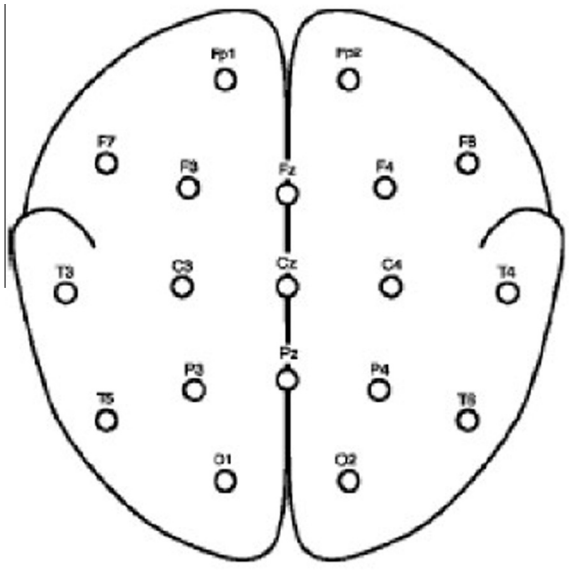

Fig. 2. Scheme of the locations of surface electrode positions are derived from their anatomical locations. Segments of Set Z were taken from all described electrodes (Andrzejak et al. 2001).

where $a, b \in \Re$, $a \neq 0$ are the scale and translation parameters, respectively, and t is the time. As a increases, the wavelet becomes narrower. Hence, there exists a unique analytic pattern at different scales. The continuous wavelet transform (CWT) of a signal $S(t)$ is defined as the correlation between the $S(t)$ and the wavelet family $\psi_{a,b}$ as follows:

$$W_{a,b} = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} S(t)\psi^*\left(\frac{t-b}{a}\right)dt = \langle S, \psi_{a,b}\rangle \tag{2}$$

For selection of the mother wavelet function $\psi(t)$ and the discrete set of parameters, $a_j = 2^j$ and $b_{j,k} = 2^j k$, *with* $j,k \in Z$, the family

$$\psi_{j,k}(t) = 2^{-\frac{j}{2}}\psi(2^{-j}t - k) \quad j, k \in Z \tag{3}$$

forms an orthonormal basis of square integrable space $L^2(\Re)$.

The discrete wavelet transform (DWT) supplies a non-redundant representation of the signal and its values constitute the coefficients in a wavelet series. These wavelet coefficients provide a direct estimation of local energies at different scales. Additionally, the information can be organized in a hierarchical scheme of nested subspaces regarded as multiresolution analysis in $L^2(\Re)$. In multiresolution analysis, $L^2(\Re)$ is decomposed into sum of the subspaces W_j, j from $-\infty$ to ∞ (Adeli et al., 2003).

$$L^2(\Re) = \cdots W_{-3} \oplus W_{-2} \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus W_2 \oplus W_3 \oplus \cdots \tag{4}$$

If the closed subspaces V_j are defined as

$$V_j = W_{j+1} \oplus W_{j+2} \oplus W_{j+3} \oplus \cdots \quad \textit{for all } j \in Z \tag{5}$$

where $\oplus$ represents direct sum, then the subspaces are a multiresolution approximation of the square integrable space $L^2(\Re)$. The subspaces W_j are the orthogonal complementary of the subspaces V_j:

$$V_{j-1} = V_j \oplus W_j \quad \textit{for all } j \in Z \tag{6}$$

Finally, the multiresolution decomposition of the original signal $S(t)$ with DWT is derived as follows:

$$S(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t-k) + \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi(2^j t - k) \tag{7}$$

where $\phi(.)$ represents the scaling function, $\psi(t)$ represents a set of wavelet functions, $d_{j,k}$ represents the wavelet coefficients, and c_k represents the scaling coefficients. The multiplication of the wavelet coefficients $d_{j,k}$ and the dilated and translated wavelet functions represents the detail of the original signal at scale j, which is defined as

$$r_j(k) = \sum_{k=-\infty}^{\infty} d_{j,k}\psi(2^j t - k) \tag{8}$$

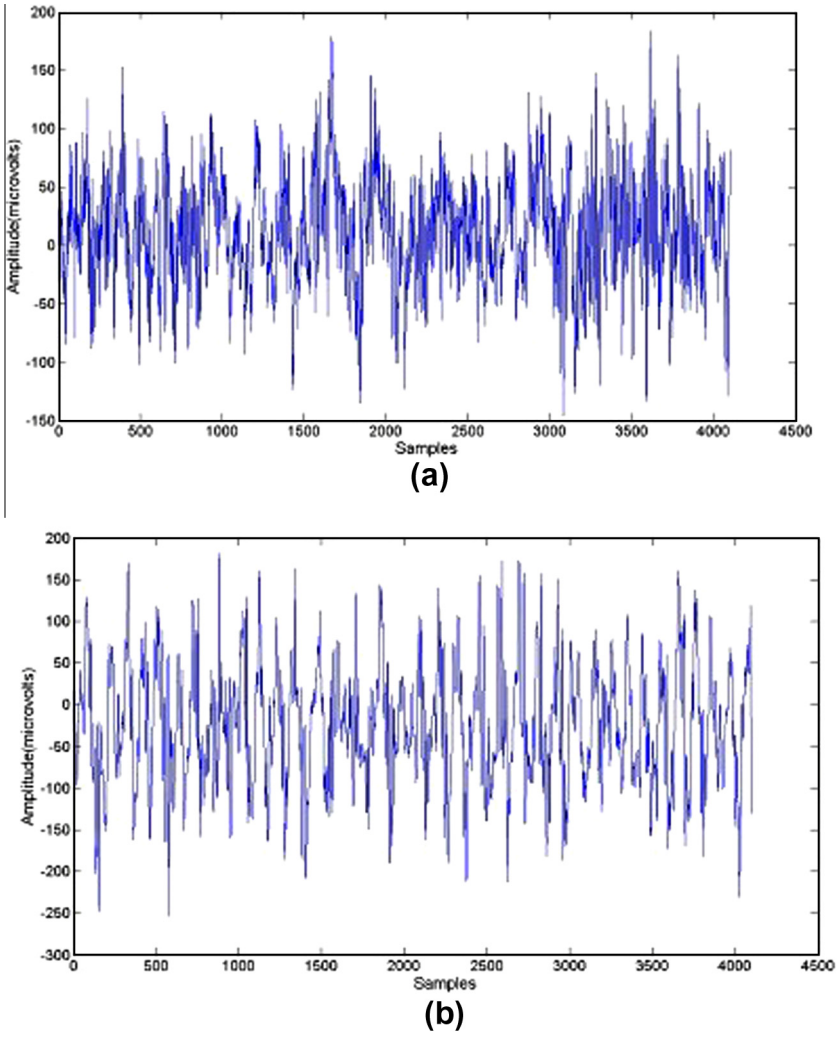

Fig. 3. Sample EEG recordings. (a) Normal EEG (Set Z). (b) Interictal EEG (Set F).

Table 1
Summary of the EEG data.

	Dataset Z	Dataset F
Subjects	Five healthy subjects	Five epileptic patients
Electrode type	Scalp	Intracranial
Electrode placement	International 10–20 system	Within epileptogenic zone
Patients' state	Awake and eyes open (normal)	Seizure-free (interictal)
Number of epochs	100	100
Epoch duration (s)	23.6	23.6

DWT works on a multi-scale basis where each signal is decomposed into several scales, and each scale providing a particular coarseness of signal. Each phase of decomposition of a signal is composed of two down samplers by 2 and two digital filters. Fig. 4 describes the process of a three-level wavelet transform of a signal $X(n)$. In each phase, $h(.)$ is the high-pass filter of that phase which serves as the discrete mother wavelet, and $g(.)$ is the low-pass filter, which acts as the mirror version of the corresponding $h(.)$. The down-sampled outputs of the first low-pass and high-pass filters supply the approximation $A1$ and the detail $D1$, respectively. The first approximation, $A1$, is further decomposed and the procedure is continued. Each low-pass filter g satisfies the quadrature mirror condition given in Guler and Übeyli (2007).

Among a variety of wavelet bases, the Daubechies family of wavelets is renowned for its orthogonality characteristic and effective filter implementation. The Daubechies order 4 wavelet (db4) is thought to be the most suitable for EEG signal analysis (Adeli et al., 2003). Hence, in this study, Daubechies order 4 wavelet was utilized for discrete wavelet transform of the EEG time series signals.

2.3. Complexity based features

Appropriate feature extraction methods are necessary for an effective epilepsy diagnosis model. Li, Ouyang, and Richards (2007) used permutation entropy to predict the absence seizures in genetic absence epilepsy rats from Strasbourg through EEG time series recordings. The permutation entropy was also employed to identify epileptic activity in human EEG signals (Nicolaou & Georgiou, 2012). The Hurst exponent has been studied for

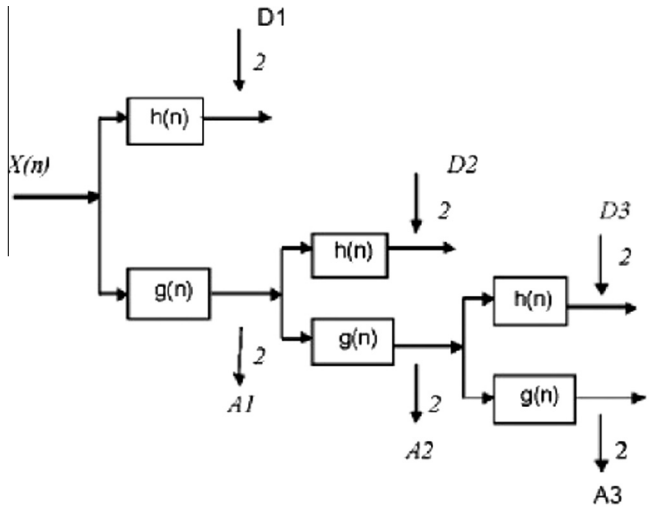

Fig. 4. Sub-band decomposition with three-level discrete wavelet transform.

quantifying self-similarity in EEG signals and has been shown to be a promising means for differentiating different epileptic EEG recordings (Kannathal et al., 2005). The complexity measures such as sample entropy, the Hurst exponent and permutation entropy can provide information about regularity, long-range autocorrelation and order structure in EEG time series recordings.

2.3.1. Permutation entropy

The complexity of an EEG series can be quantified through symbolic dynamics. A new permutation approach was presented in Bandt and Pompe (2002) for mapping a continuous time series onto a symbolic sequence; the statistical properties of the symbolic sequences were named permutation entropy (PE). The permutation entropy is simple and fast to calculate. These two characteristics enable PE to identify dynamic alterations in complex time series data effectively (Cao, Tung, Gao, Protopopescu, & Hively, 2004).

Given a time series (x_t, $t = 1, 2, \ldots, T$), an embedding process constructs a vector $X_t = [x_t, x_{t+1}, \ldots, x_{t+(n-1)l}]$ with embedding dimension, n, and lag, l. Then X_t is arranged in increasing order: $[x_{t+(j_1-1)l} \leqslant x_{t+(j_2-1)l} \leqslant \cdots \leqslant x_{t+(j_n-1)l}]$. For n different numbers, there exist $n!$ possible order patterns π, which are also called permutations. Let $f(\pi)$ represent its frequency in the time series, then its relative frequency is $p(\pi) = f(\pi)/(T-(n-1)l)$. The permutation entropy is defined as:

$$H_p(n) = -\sum_{n=1}^{n!} p(\pi)\ln p(\pi) \tag{9}$$

The corresponding normalized permutation entropy is

$$H_p = \frac{H_p(n)}{\ln(n!)} \tag{10}$$

The largest value of H_p is 1, indicating that the time series recording is totally random; the smallest value of H_p is 0, indicating that the time series recording is extremely regular. Permutation entropy measures the departure of the time series data from totally random data: the smaller the value of the PE, the more regular the time series. In brief, the permutation entropy measure is associated with local order structure of the time series data (Li et al., 2007). With respect to the embedding dimension n, if n is too large then it becomes too difficult to recognize alterations in the time series data. However, if n is too small, then there exist very few distinct states and the scheme will not work. For EEG time series recordings values of $n = 3, \ldots, 7$ have been recommended (Bandt & Pompe, 2002). In this study, order $n = 3$ was chosen during the computation of permutation entropy, as recommended by Nicolaou and Georgiou (2012). For the time lag, it is sufficient to employ a value of $l = 1$ to extract most of the information in the EEG time series recordings (Bruzzo et al. 2008; Olofsen, Sleigh, & Dahan, 2008); therefore this value is commonly selected in the EEG time series analysis.

2.3.2. The Hurst exponent

The Hurst exponent (HE) is an indicator which has been utilized for depicting the correlation characteristics and self-similarity of physiological time series recordings (Acharya, Faust, Kannathal, Chua, & Laxminarayan, 2005). The Hurst exponent is a measure of the smoothness of fractal time series data on the basis of the asymptotic behavior of the rescaled scope of the procedure, and it can be estimated using rescaled range analysis (R/S) which is described as follows:

Assume time series data $\{x(i), i = 1, \ldots, N\}$. The deviation from the mean $\bar{x}(n)$ of the first k data points is defined as

$$W_k = (x_1 + x_2 + x_3 + \cdots + x_k) - k\bar{x}(n) \tag{11}$$

where $1 \leqslant k \leqslant n$ and $1 \leqslant n \leqslant N$.

Then the difference between the maximum value and minimum value of the deviations corresponding to n is obtained by

$$R(n) = \max(0, W_1, \ldots, W_n) - \min(0, W_1, \ldots, W_n), \quad n = 1, \ldots, N \tag{12}$$

If $S(n)$ represents the standard deviation of the time series $\{x(i), i = 1, \ldots, n\}$, $R(n)/S(n)$ increases as a power law,

$$\frac{R(n)}{S(n)} = C \times n^H, \quad n = 1, \ldots N \tag{13}$$

where C is a constant and H is the estimated value of the Hurst exponent, i.e.,

$$HE = \frac{\log[R(n)/S(n)]}{\log(n)}, \quad n = 1, \ldots, N \tag{14}$$

With regard to the analysis of the dynamic cerebral procedures, HE can be utilized for characterizing the extent of long-range dependence in EEG time series signals. If $HE = 0.5$, then the behavior of the time series signal is similar to a random walk. If $0 < HE < 0.5$, then the behavior of the time series signal shows anti-persistence, namely if the time series increments, it is more likely that it will decrease in the next period. The time series covers less 'distance' than a random walk, and the closer is to zero HE, the stronger the extent of the anti-persistency. If $0.5 < HE < 1$, it will have persistent effects on the behavior of the time series. There exists a tendency that if the time series increments, then it is more likely that the time series will continue incrementing. Thus, the time series covers more 'distance' than a random walk (Kannathal et al., 2005).

2.3.3. Sample entropy

Richman and Moorman (2000) introduced a new family of statistics called Sample Entropy (SampEn) to avoid the bias of approximate entropy. SampEn quantifies the complexity of time series data and can be applied for short-length time series data. Additionally, it is resistant to short strong transient interferences (outliers) such as spikes. These characteristics make Sample Entropy an appealing tool for non-linear analysis of physiological signals. In Abásolo, Hornero, Espino, Grant, and Ludbrook (2006), sample entropy was used to analyse the electroencephalogram background activity of Alzheimer's disease patients for testing the hypothesis that the regularity of their EEGs is higher than that of age-matched controls. Aboy, Cuesta-Frau, Austin, and Micó-Tormos (2007) con-

ducted a characterization study of SampEn for supplying additional insights about the interpretation of this complexity metric in the context of biomedical signal analysis. Moreover, this entropy measure has been used to evaluate the signal complexity of the cyclic behavior of heart rate variability (HRV) in obstructive sleep apnea syndrome (Al-Angari & Sahakian, 2007).

For calculating the SampEn, the embedding dimension (m) and vector comparison threshold (r) must be specified. It is common to set the embedding dimension parameter m to be $m = 1, 2$ or 3 and to set the vector comparison threshold r to be some percentage of the standard deviation of the time series so as not to depend on the absolute amplitude of the signal (Richman & Moorman, 2000). SampEn is the negative logarithm of the conditional probability that two sequences similar for m points remain similar at the next point, where self-matches are not included in calculating the probability. Thus, a larger value often corresponds to more irregularity or complexity in the time series data. The value of the SampEn is determined as in the following steps:

(1) Given N data points from a time series $\{x(n)\} = x(1), x(2), \ldots, x(N)$, take m vectors $X_m(1), \ldots, X_m(N - m + 1)$ defined as $X_m(i) = [x(i), x(i + 1), \ldots, x(i + m - 1)]$, for $1 \leqslant i \leqslant N - m + 1$. These vectors stand for m consecutive x values, starting at the ith sample.
(2) Let r denote the noise filter level which is defined as

$$r = g \times Std \quad for\ g = 0.1, 0.2, \ldots, 0.9 \tag{15}$$

where Std represents the standard deviation of the data sequence X.

(3) The distance between vectors $X_m(i)$ and $X_m(j)$, $d[X_m(i), X_m(j)]$, is defined as the maximum absolute difference between their scalar components:

$$d[X_m(i), X_m(j)] = \max_{k=0,\ldots,m-1}(|x(i+k) - x(j+k)|) \tag{16}$$

(4) For a given $X_m(i)$, count the number of j ($1 \leqslant j \leqslant N - m, j \neq i$), such that $d[X_m(i), X_m(j)] \leqslant r$. This number is represented as B_i, Then, for $1 \leqslant i \leqslant N - m$,

$$B_i^m(r) = \frac{1}{N - m - 1} B_i \tag{17}$$

Here, note that only the first $N - m$ vectors of length m are considered in order to ensure that for $1 \leqslant i \leqslant N - m$, the vector $X_{m+1}(i)$ is also defined.

(5) Define $B^m(r)$ as

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^m(r) \tag{18}$$

(6) Increment the dimension to $m = m + 1$ and compute A_i as the number of $X_{m+1}(i)$ within r of $X_{m+1}(j)$, where j ranges from 1 to $N - m (j \neq i)$. Then define $A_i^m(r)$ as

$$A_i^m(r) = \frac{1}{N - m - 1} A_i \tag{19}$$

(7) Define $A^m(r)$ as

$$A^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i^m(r) \tag{20}$$

Thus, $B^m(r)$ represents the probability that two sequences will match for m points, whereas $A^m(r)$ represents the probability that two sequences will match for $m + 1$ point.

The sample entropy is defined by

$$SampEn(m, r) = \ln\left[\frac{B^m(r)}{A^m(r)}\right] \tag{21}$$

2.4. Genetic algorithms

Genetic algorithms (GAs) are a family of optimization methods which are based on the notion of "the survival of the fittest" in Darwin's classical theory of evolution (Eiben & Smith, 2003). They were originally proposed by Holland (1975). The main characteristic feature of genetic algorithms is that they can search for an optimum point in a multi-dimensional optimization surface by iteratively refining a single solution, and can deal with nonlinear, non-continuous and non-differentiable optimization problems (Huang & Wang, 2006). In addition, genetic algorithms operate on a set of candidate solutions in parallel. As a result, GAs possess a higher likelihood of positioning a global optimum point than the conventional approaches which are more likely to get stuck at a local optimum around the initial guess. A traditional genetic algorithm mainly consists of the following steps:

(1) Use a chromosome with a settled length for denoting the problem variable domain, and specify some parameters such as the size of a chromosome population, the crossover probability and the mutation probability.
(2) Apply a fitness function for setting up the standard of selecting chromosomes. In terms of the value of the fitness function, some chromosomes will be selected for mating in the process of reproduction.
(3) Create an initial population of chromosomes randomly.
(4) Compute the fitness of each individual chromosome.
(5) Use genetic operators- selection, crossover and mutation- for generating a pair of offspring chromosomes.
(6) Add these new chromosomes to the new population.
(7) Reiterate step (5) until the size of the new chromosome population becomes the same as that of the old population.
(8) Take the place of the old chromosome population with the new population.
(9) Return to step (4), and reiterate the procedure until the ending condition is contented.

The flow chart of a basic genetic algorithm is depicted in Fig. 5.

2.5. Extreme Learning Machine

A new learning algorithm for single-hidden layer feed-forward neural networks which is named Extreme Learning Machine (ELM) was proposed recently (Huang, Chen, & Siew, 2006; Huang, Zhu, & Siew, 2006). The research shows that a single-hidden layer neural network with N hidden neurons can learn N different observations for nearly any non-linear activation function with any small error. Unlike the common opinion that the parameters of network have to be tuned, one may refrain from tuning the input weights and hidden layer biases but assign them at random. Thus, only the output weights are required to be adjusted, and the adjustments of the input weights and hidden layer biases do not make any contribution to the performance of the SLFN. ELM transforms the learning problem into a simple linear system whose output weights can be determined analytically by means of a Moore–Penrose generalized inverse operation of the weight matrices in hidden layers. Such a learning scheme achieves much faster learning speed and better generalization ability with the smallest training error and weights of SLFN than traditional learning algorithms (Huang et al., 2006).

2.5.1. Moore–Penrose generalized inverse

The solution of a general linear system of equations

$$Ax = y$$

where A is a singular or rectangular matrix can be achieved using the Moore–Penrose generalized pseudo inverse.

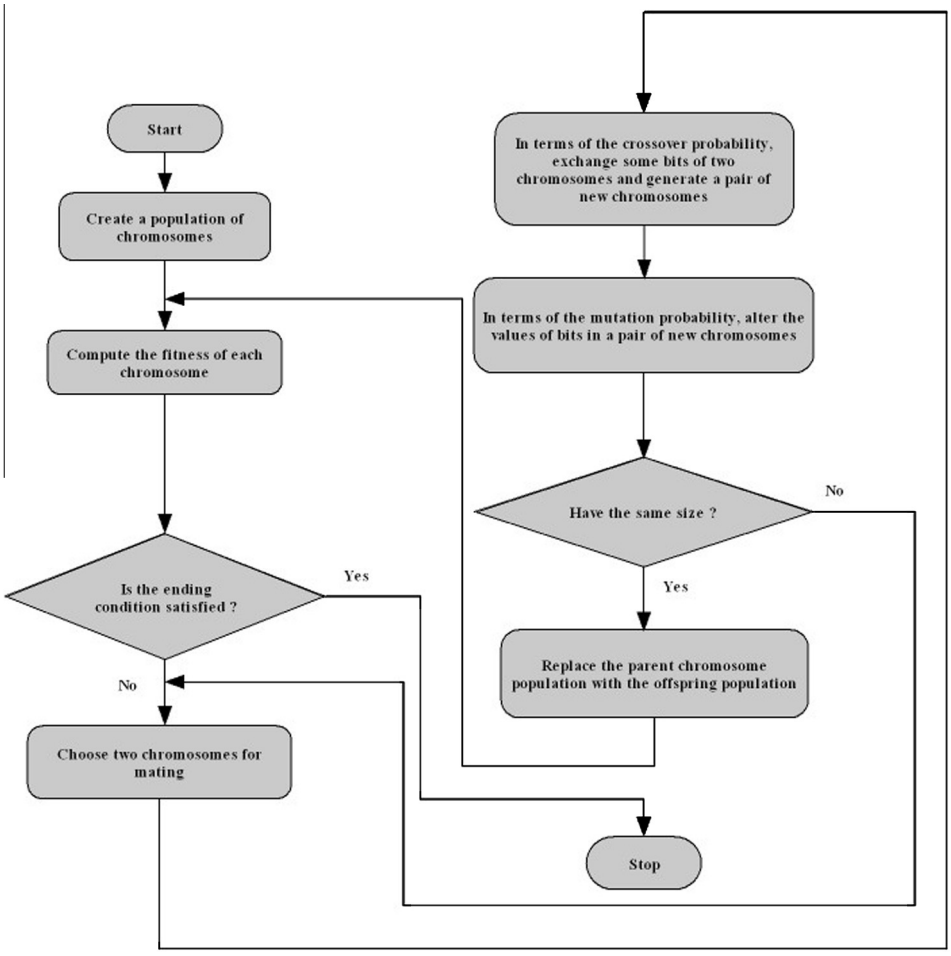

Fig. 5. Flow chart of a basic genetic algorithm.

Definition 1 Rao & Mitra, 1971. A matrix G of size $n \times m$ is called the Moore–Penrose generalized pseudo inverse of a given matrix A of size $m \times n$, if

$$AGA = A, \quad GAG = G, \quad (AG)' = AG, \quad (GA)' = GA.$$

In this case, we represent G by $A^{\dagger}$.

Definition 2. For the given general linear system of equations $Ax = y$ where A is a matrix of size $m \times n$ and y is a vector in $\Re^m$, a vector x^* in $\Re^n$ is regarded as a least squares solution if

$$||Ax^* - y|| = \min_x ||Ax - y||$$

Definition 3. The vector x^* in $\Re^n$ is regarded as a minimum norm least squares solution of the general linear system $Ax = y$, if x^* must be a least squares solution and further among all least squares solutions x in $\Re^n$

$$||x^*|| \leqslant ||x||$$

must be true.

Theorem 1 Rao & Mitra, 1971. *Let G be a matrix of size $n \times m$. Then $x^* = Gy$ is a minimum norm least squares solution of the general linear system $Ax = y$ if and only if $G = A^{\dagger}$, the Moore–Penrose generalized inverse of A. It is clear that A^*y is the unique minimum norm least squares solution.*

2.5.2. Single hidden layer feedforward neural networks

Suppose learning N arbitrary different instances (x_j, t_j), where $x_j = [x_{j1}, x_{j2}, \ldots, x_{jn}]^T \in \Re^n$, and $t_j = [t_{j1}, t_{j2}, \ldots, t_{jm}]^T \in \Re^m$, standard single-hidden layer feed-forward networks with M hidden neurons and activation function $g(x)$ are mathematically modeled as a linear system

$$\sum_{i=1}^{M} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \ldots, N \tag{22}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ denotes the weight vector connecting the ith hidden neuron and the input neuron, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ denotes the weight vector connecting the ith hidden neuron and output neurons, $o_j = [o_{j1}, o_{j2}, \ldots, o_{jm}]^T$ is the output vector of the SLFN, and b_i represents the threshold of the ith hidden neuron. $w_i \cdot x_j$ represents the inner product of w_i and x_j. If the SLFN with N hidden neurons with activation function $g(x)$ is able to approximate N distinct instances (x_j, t_j) with zero error means that

$$H\beta = T \tag{23}$$

where

$$H(w_1, \ldots, w_M, b_1, \ldots, b_M, x_1, \ldots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_M \cdot x_1 + b_M) \cdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_M \cdot x_N + b_M) \end{bmatrix}_{N \times M} \tag{24}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \cdot \\ \beta_M^T \end{bmatrix}_{M \times m} \qquad T = \begin{bmatrix} t_1^T \\ \cdot \\ t_N^T \end{bmatrix}_{N \times m} \tag{25}$$

H is the hidden layer output matrix of the SLFN.

In the case of learning an arbitrary function with zero training error, Baum (1988) proposed some constructions of single-hidden layer feed-forward neural networks with adequate hidden neurons. But practically, much less hidden neurons are needed to obtain appropriate generalization ability on unseen data samples, and the corresponding learning error might not approach to zero. Huang, Chen, and Siew, (2006) proved that, given arbitrarily small value $\varepsilon > 0$, if the activation function of the hidden layer in the single-hidden layer feed-forward network is infinitely differentiable, and the number of hidden neurons is less than or equal to the number of data samples, then the input weights and hidden layer biases can be assessed at random, and the single-hidden layer feed-forward network approximates the N training data with ε error, that is to say, $||H\beta - T|| \leqslant \varepsilon$. The ELM algorithm randomly assigns the input weights w_i and hidden layer biases b_i for the SLFN with the infinitely differentiable activation function, thus, the hidden layer output matrix H has been fixed at random. Hence, the process of training a single-hidden layer feed-forward neural network is equal to solving a well-established linear system optimization problem which is shown as follows:

$$||H\hat{\beta} - T|| = \min_{\beta} ||H\beta - T|| \tag{26}$$

Its unique least-squares solution with minimum norm is

$$\hat{\beta} = H^{\dagger}T \tag{27}$$

where $H^{\dagger}$ is the Moore–Penrose generalized inverse of H. According to the study in Barlett (1998), as for feed-forward networks, the smaller their output weights are, the better generalization ability they have. The norm of $\hat{\beta}$ is the smallest among all the least-squares solutions of the linear system $H\beta = T$, thus ELM obtains not only the minimum square training error but also the best generalization ability on unseen data samples.

The ELM algorithm is outlined as follows:

Input: Training set $\{(x_j, t_j)\}_{j=1,2,\ldots,N}$ where $x_j \in \Re^n$ and $t_j \in \Re^m$; M is the number of hidden neurons and the activation function $g(.)$.

(1) For $M = 1, 2, \ldots, M$, choose arbitrary value for input weights $w_i \in \Re^n$ and biases $b_i \in \Re$ of hidden neurons.
(2) Calculate hidden layer output matrix H.
(3) Obtain the optimal $\hat{\beta}$ in the light of equation $\hat{\beta} = H^{\dagger}T$.

Output: For any input data $x \in \Re^n$ the output value t can be obtained through the following formula:

$$t = \sum_{i=1}^{M} \hat{\beta}_i g(w_i \cdot x + b_i) \tag{28}$$

3. Experimental results and discussion

EEG time series recordings contain a wealth of information that can be extracted for generating marvellous insight into the dynamics of the human brain. Traditional visual inspection of EEG recordings by human experts involves the recognition of the spatial scope (local or generalized) and temporal persistence of the abnormalities, and involves inspection of wavelength, amplitude, waveform stability and hyperventilation and so on (Adeli et al., 2010). Although it is possible for neurology experts to recognize such features from a visual inspection, other information, such as complicated hidden patterns underlying the EEG waveform, which may contain more useful sources is not visible and must be mined through analytical signal processing approaches.

As can be seen from Fig. 3, normal EEG signals and interictal EEG signals in the time domain look similar; there is not much outstanding information that can be extracted to accurately differentiate between epileptic brain states and normal brain states simply by means of EEG waveform examination. In order to overcome this problem, non-linear features, such as permutation entropy, sample entropy and the Hurst exponent, were investigated to examine underlying information in EEG time series signals.

3.1. Wavelet decomposition of eeg into different frequency bands

The goal of wavelet analysis is to decomposed signals into a number of frequency bands. The choice of appropriate wavelet and the number of decomposition levels are very important in the analysis of EEG signals through discrete wavelet transform. The five EEG frequency bands, delta, theta, alpha, beta, and gamma, cover the 0–60 Hz frequency scope. EEG signals have little useful information above a frequency of 60 Hz (Adeli et al., 2010). With the goal of correlating the wavelet decomposition with the frequency scopes of physiological brain rhythms, the frequency content of the wavelet filter is supposed to be limited to the scope of 0–60 Hz in this study. To satisfy the above-mentioned conditions, the EEG time series signals were band-limited to the desired 0–60 Hz scope by using a low-pass finite impulse response (FIR) filter. The energy of the frequency band removed by the FIR filter can be neglected in comparison with that of the saved rhythm waves within the scope of 0–60 Hz. After testing various types of traditional wavelets, Daubechies wavelet of order 4 was found to be more appropriate for identifying alterations in EEG time series signals because of its smoothing characteristics (Subasi, 2007). Hence, the Daubechies order 4 wavelet was chosen in this study.

The original EEG signals were then subjected to a level four decomposition utilizing order 4 Daubechies wavelet transform. After the first level of decomposition, the EEG signal (0–60 Hz) was decomposed into its higher resolution component, d1

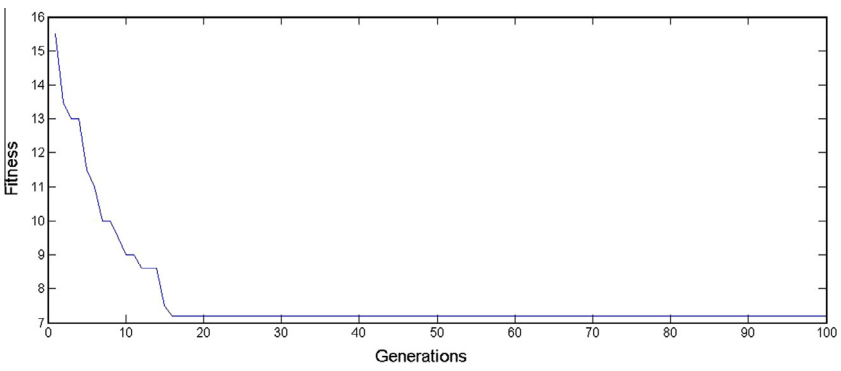

Fig. 6. Fitness values of the best population chromosomes in each generation.

(30–60 Hz), and lower resolution, a1 (0–30 Hz). The a1 component was further decomposed into its higher resolution component, d2 (15–30 Hz), and lower resolution component, a2 (0–15 Hz), through the second level of decomposition. After four levels of decomposition, the EEG time series signal was decomposed into five EEG frequency bands approximating to delta, theta, alpha, beta, and gamma frequency bands.

3.2. Nonlinear feature extraction

Feature extraction is a highly effective approach for dimension reduction and is often applied to complex, high dimensional, multivariate data. It is the most significant step in EEG signal recognition. The recognition performance greatly relies on the features which are being utilized to characterize the primary EEG signals. Hence, optimal extraction and selection of the features plays a very crucial role in the performance of the diagnosis system. In the feature extraction phase, EEG time series signals are commonly analyzed by inspecting frequency bands related to various conscious states or mental activities. Some feature extraction methods have been proposed and include a measure of signal similarity in the correlation dimension (Quyen, Martinerie, Baulac, & Varela, 1999), a measure of energy variation in EEGs (Esteller et al., 2005) and a measure of accumulated energy (Gigolab, Ortiza, Dattellis, Silvac, & Kochenc, 2004).

Huge amounts of nerve cells in the human brain are connected with axons and synapses for building a complex neuronal network. It is known that the straightforward superposition of the functions of all nerve cells does not equal to the function of the neuronal network. Therefore brain electrical activities possess rather complicated dynamic characteristics and properties. As an effective technique for measuring brain electrical activities, EEG time series demonstrate typically complex dynamics. Even though feature extraction based on linear approaches have achieved satisfactory results for the investigation of EEG time series recordings, they could simply describe linear properties of EEGs and are not capable of describing the non-linear characteristics of the neuronal network. Hence, feature extraction on the basis of non-linear analysis has become popular in defining the dynamic characteristics of EEG signals. Adeli, Ghosh-Dastidar, and Dadmehr (2007) presented a chaos based analysis methodology which used two chaotic measures, namely largest Lyapunov exponent (LLE) and correlation dimension (CD), to quantify the non-linear dynamic characteristics of EEG time series data. Hsu and Yu (2010) also measured LLE and CD for EEG time series signals and combined a genetic optimization operation to detect epileptic seizure signals. Ocak (2008) extracted approximate entropy values from each EEG frequency band and then fed them into a learning vector quantization (LVQ)-based normal and epileptic EEG classifier to maximize EEG classification performance. Although these above-mentioned approaches obtained satisfactory results, non-linear methods such as Lyapunov exponent and correlation dimension often need more than ten thousand data points to obtain right results (Eckmann & Ruelle, 1992), which will have a great effect on the efficiency of the system. On the other hand, the approximate entropy algorithm counts each vector as matching itself in order to prevent the appearance of ln(0) in the computations, which results in the bias of approximate entropy. In order to avoid these limitations, in this study, three relatively new complexity based non-linear measures, namely sample entropy, Hurst exponents and permutation entropy were examined to measure the underlying non-linear dynamic characteristics of EEG time series recordings for epilepsy diagnosis. Sample entropy avoids the bias of approximate entropy and can be applied to short-length time series data (Richman & Moorman, 2000). Additionally, it is resistant to short strong transient interferences (outliers) such as spikes. The embedding dimension (m) and vector comparison distance (r) were set as 2 and 0.1 times the standard deviation of the coefficients, respectively, in light of several tests on feature extraction performance and suggestions by Richman and Moorman (2000). Permutation entropy is known for its conceptual simplicity and computational efficiency, which make it an outstanding candidate for an effective, fast and robust screener and detector of abnormal patterns in complex time series recordings. For permutation entropy extraction, order $n = 3$ was chosen during the computation of permutation entropy, as recommended by Nicolaou and Georgiou (2012). For the time lag, it is sufficient to employ a value of $l = 1$ to extract most of the information in the EEG time series recordings (Bruzzo et al., 2008; Olofsen et al., 2008); therefore this value was selected in the current EEG time series analysis. The Hurst exponent is an indicator which has been utilized for depicting the correlation properties and self-similarity of physiological signal data. In Nurujjaman et al. (2009), the Hurst exponent has been shown to be useful in characterizing normal and abnormal brain states.

The advantages of wavelet-complexity analysis can be explained in two ways. First, on the basis of the dynamic properties of the human brain, the method generates detailed information from different frequency bands over two sets of EEG signals, which has a considerable effect on the following feature extraction and selection procedures. The epileptic EEG signal analysis problem can be treated from a multi-resolution angle by wavelet based decomposition. Second, high dimensional feature space is significantly reduced to construct a more manageable input space. Before wavelet-complexity feature extraction, the dimensions of the feature space are $M \times 6 \times 2L$ (M samples per EEG signals or EEG frequency band, six EEGs and EEG frequency bands and two sets of L EEGs). Following the wavelet-complexity feature extraction, each EEG signal is denoted as R dimensional samples within the new feature space K. The dimensions of the input space are $R \times 2L$, where $R \in \{1, 2, \ldots, 18\}$ is chosen over three features acquired from both the original EEG signal and its given frequency bands. In this study, there are 18 nodes in a 4th-level wavelet decomposition framework. Hence the feature vector extracted from a single EEG signal is composed of 18 features. Both normal and interictal EEG groups are composed of 100 EEG signals.

3.3. Feature normalization

Since different features extracted from the original EEG signals and their five frequency bands have different scales, a normalization procedure is needed to tailor all non-linear features to the same level. The normalization is defined as follows:

$$x'_{ij} = \frac{x_{ij} - x_j^{\min}}{x_j^{\max} - x_j^{\min}} \tag{29}$$

where x_{ij} is the jth component of the ith feature vector, and $x_j^{\max}$ and $x_j^{\min}$ are the maximum and minimum feature values, respectively, of the jth component of the feature vectors.

The Extreme Learning Machine recognition model was used for EEG signal recognition. The sigmoid function was selected as the hidden layer activation function. The interictal EEG signals were represented as 1 and the normal EEG signals were represented as 0. The weight vector w and the threshold b of ELM were derived at random in the range of [0 1]. Among these 200 EEG signals (100 normal signals and 100 interictal signals), half of the normal EEG signals and half of the interictal EEG signals were utilized for training, and the rest for testing. In order to calculate the average test performance, the classification experiment was repeated ten times and all simulation results were averaged over ten trials.

3.4. Genetic algorithm based optimal feature space selection

Searching for the best feature subset space, which maximizes the differentiation between normal brain states and brain states with seizure free, is a hard task without an effective and robust optimization approach. A genetic algorithm was employed to search for the best feature subset. The initial population size was set as 26. Each individual chromosome was composed of 18 genes. Each gene standard for a complexity based feature of each EEG frequency band and the primary EEG signal. The values of genes were set to either 0 or 1. A value 0 means that the corresponding complexity based feature would be excluded from the feature subset space. The corresponding feature would be included in the feature subset space if the value of the gene was 1. Each gene was derived from the initial 26 chromosomes with value 0 or 1 at random. The fitness function of a single chromosome was defined as follows:

$$fitness = (m_1 + m_2)/2 \tag{30}$$

where m_1 and m_2 represent the misclassification rates of the ELM classifier in normal and interictal EEG time series data, respectively, in light of the corresponding feature subset that the chromosomes are standing for. With the goal of minimizing the fitness function evaluation, the conditions of the genetic algorithm were set as follows: two of the chromosomes with the best fitness values in the present generation were regarded as elites and they remained in the next generation with no variations in their gene values. During each generation, with the exception of two elite chromosomes, 75% of chromosomes in the population were generated by means of crossover operation. The rest of 25% of chromosomes were derived by means of mutation operation. The genetic algorithm was set to be run 100 generations in total, and the stopping condition was determined so that the optimization procedure could stop if no improvements in the fitness values existed for twenty successive generations. The crossover rate was 0.9 and the mutation rate was 0.01.

With the goal of identifying abnormal EEG signals from normal brain states, all features were utilized by ELM. Half of the signals were chosen for training and the rest for testing. This procedure was repeated ten times; the average sensitivities (interictal EEGs), average specificities (normal EEGs) and average accuracies are outlined in Table 2. The accuracy rate for the whole feature set was 85.9% with 86.6% and 85.2% for the interictal and normal EEG epochs recognitions, respectively. The genetic algorithm was then run for a maximum of 100 generations to select the optimal feature subset. The testing results are shown in Table 2. In comparison with the results obtained from extracting the whole feature set, the recognition accuracy of the ELM by genetic algorithm optimization was improved from 85.9% to 94.2%, together with an increase in sensitivity from 86.6% to 94.4% and an increase in specificity from 85.2% to 94.0%, respectively. Each of the performance criteria is the average value of ten trials with randomly chosen training and test data. The improvement in the results obtained from ten trials of ELMs, with and without feature subset selection by a genetic algorithm, were analysed statistically by means of a Student's t-test. All results demonstrated statistically significant improvements in accuracies with confidence of over 99% ($p < 0.01$), which further showed the advantage of incorporating a genetic algorithm as a feature subset selector for epilepsy diagnosis in EEG time series recordings.

3.5. Genetic algorithm with weight alteration

The objective of the epilepsy diagnosis approach used in the current study is to select possible abnormal EEG signals for further analysis and management. Hence, the number of false interictal EEG signal diagnoses should remain as few as possible, and at the same time the number of false normal EEG signal diagnoses should not be increased greatly. For the purpose of epilepsy diagnosis, the sensitivity value for identifying abnormal brain states may be more significant than the specificity value for identifying normal brain states. The weights of specificity and sensitivity can be altered through improving the fitness function of the genetic algorithm. Here, the weights of specificity and sensitivity were set as one and two, respectively, shown as follows:

$$fitness' = (1 \times m_1 + 2 \times m_2)/2 \tag{31}$$

where m_1 and m_2 represent the misclassification rates of the ELM classifier in normal and interictal EEG time series data, respectively, in light of the corresponding feature subset which the chromosomes stand for. The genetic algorithm with weight alteration in the fitness function was designed for screening the best feature set for abnormal EEG diagnosis and the results are shown in Table 3. The ELM recognition model was examined together with the genetic algorithm.

The recognition sensitivity of ELM, ELM with a genetic algorithm for feature selection (ELM + GA), and ELM with a genetic algorithm and weight alteration for feature selection (ELM + GA_weight) were 86.6%, 94.4% and 96.0%, respectively. The recognition sensitivity of ELM + GA_weight was higher than that of ELM and ELM + GA. The recognition accuracy of the interictal EEG group in ELM + GA_weight (94.8%) was higher than that in ELM and ELM + GA (85.9% and 94.2%, respectively). These results showed that the weight alteration strategy indeed increased the sensitivity of the recognition model despite a slight decrease in specificity.

3.6. Importance of genetic algorithm operators

Determining the genetic algorithm variables reasonably is significant for finding the best solution in a complicated optimization surface as it is in this study. Dependence on simple selection operation on the basis of fitness evaluation with no crossover and mutation procedures is inclined to update the population through the best single chromosome from the whole population and will have a very limited impact on the algorithm convergence. Only dependence on selection and crossover operations is inclined to result in the optimization procedure achieving a local optimal solution rather than a global optimum. Simply running the mutation operation without selection and crossover operations is more like a random searching process over the searching surface. In addition, applying only selection and mutation operation, without crossover operation generates a parallel, hill climbing procedure (Ocak, 2008). This study sets the crossover and mutation ratio for the genetic algorithm as 75% and 25%, respectively, which means 75% of the chromosomes in the population, excluding the elite

Table 2
Performance of ELM and ELM combined with GA.

	Sensitivity (%)	Specificity (%)	Accuracy (%)
ELM	86.6 ± 4.1150	85.2 ± 5.0067	85.9 ± 4.5570
ELM + GA	94.4 ± 3.5024	94.0 ± 5.4160	94.2 ± 4.4422
p-Value	2.4025e−4	0.0014	6.3684e−4

Table 3
Performance of ELM combined with GA with weight alteration for feature selection.

	Sensitivity (%)	Specificity (%)	Accuracy (%)
ELM	86.6	85.2	85.9
ELM + GA	94.4	94.0	94.2
ELM + GA_weight	96.0	93.6	94.8

chromosomes, were generated during each generation by running the crossover operation, and the rest of 25% of chromosomes were created by running the mutation operation. Using mutation operation produces the random addition of new candidates throughout the searching surface. Nevertheless, raising the mutation probability will lead to the increase of random behavior in the searching procedure and hence has passive effects on the convergence. Here, the variable setting for the genetic algorithm was found to be the relatively optimal option for the abnormal EEG recognition problem (see Fig. 6).

3.7. Features selected by genetic algorithm

The optimal feature subset selected by the genetic algorithm is described in Fig. 7. The shaded blocks contributed the optimal feature subset while the rest of blocks were excluded from the feature set. In general, most of the features in the optimal feature subset come from the alpha frequency band and the beta frequency band (each of these two frequency bands contains two optimal non-linear features) for ELM. One feature was selected from the original EEG time series and one feature was selected from the delta frequency band. No features were selected from the theta frequency band or the gamma frequency band. The extraction of features from different brain rhythm bands becomes feasible with the use of wavelet decomposition prior to feature computation. The feature analysis based on genetic algorithm optimization further demonstrated the relative significance of single features extracted from the specific EEG frequency bands to the generalization capability of the recognition model. In terms of the current study, it was more effective and appropriate to utilize the non-linear features selected by the genetic algorithm for epilepsy diagnosis instead of all features. It was obvious that the non-linear parameters from the original EEG, alpha, beta, and delta frequency bands were more valuable for abnormal EEG recognition in comparison with the features extracted from the theta and gamma frequency bands.

The decomposition of the original EEG into its five frequency bands changes the primary non-linear dynamic characteristics of EEG time series and results in new properties which are not directly associated with the original EEG time series signal, i.e. each frequency band possesses its own non-linear dynamic characteristics. When the observation is made on the basis of the whole EEG time series, it may be concluded that only the sample entropy (and not the Hurst exponent or permutation entropy) could be utilized as a differentiating parameter between normal EEG signals and abnormal EEG signals. However, when the analysis was conducted over the EEG frequency bands, it was found that both the Hurst exponent and permutation entropy, extracted within certain physiological frequency bands, could also be used to discriminate between these two groups of EEG signals. Hence, it was concluded that alterations in the brain dynamics may not be spread out in an equal way across the spectrum of the EEG time series, but instead, were limited to certain brain frequency bands.

The proposed feature analysis and recognition framework tested the assumption that EEG time series signals denote the dynamic properties of the whole brain as a combined system and thus, should be considered as a whole. The primary EEG signals were decomposed into five frequency bands, each standing for a sub-procedure underlying the entire brain dynamics. From the results, it can be seen that more valuable characteristics which can be used to discriminate pathological brain states from normal brain states tend to appear in these EEG frequency bands, rather than the primary EEG signals. This means that contrary to the original assumption, EEG time series actually stands for the impact of the accumulation of a variety of procedures within the brain. Single EEG frequency band may be more representative of brain dynamics than the whole EEG time series, and certain EEG frequency bands may generate more precise information regarding interaction activities of cerebral neuronal networks underlying EEG time series signals. Hence, it is likely that some alterations in the EEG time series which are very difficult to observe in the primary EEG signals are intensified when each EEG frequency band is investigated.

Fig. 8 demonstrated the distributions of the Hurst exponent, permutation entropy and sample entropy values calculated from the delta frequency band, beta frequency band and alpha frequency band, respectively, for normal and interictal EEG signals. These features have been selected by the genetic algorithm as representative features for identifying abnormal brain states. As can be seen in Fig. 8, there was some overlapping between values from normal EEG signals and from interictal EEG signals for all three non-linear features. This means that the single non-linear parameter may not be effective to accurately discriminate between abnormal brain states and normal brain states, as can be learnt from the selection result of the genetic algorithm. However, in terms of the obtained high recognition accuracy using the 6-dimensional feature space, it is shown that the combination of these different non-linear parameters from certain EEG frequency bands can express the separations between these two EEG time series signals very well. The proposed feature extraction methodology was able to find these representative features effectively.

It can also be noted from Fig. 8 that values of all three non-linear parameters decreased in the abnormal brain states, implying that there was decline in variability within EEG time series during the transition between the normal brain state and the abnormal brain state.

Sample entropy quantifies the regularity in an EEG time series signal. Values of sample entropy (0.7339 ± 0.2134) were higher in the normal brain states, and became lower (0.4584 ± 0.1565) in the abnormal brain states, which indicated more self-similarity when there were abnormal epileptogenic tissues in the brain. According to Mormann et al. (2003), during interictal state, the EEG signals from the epileptogenic zone and neighboring zones become desynchronized. Therefore, there is a drop in connection among nerve cells within the epileptogenic zone, which leads to a reduction in the complexity of EEG time series data. The decrease in signal complexity reduces the value of sample entropy in the alpha frequency band. Due to the reduction in connection among epileptic nerve cells, these nerve cells are isolated, and thus ictal state starts.

Permutation entropy provides a measure of departure of the EEG time series signal under study from a totally random time series. The smaller the value of permutation entropy, the more regular the EEG time series signal. Here, normal EEG signals (0.6018 ± 0.11) had higher values of PE compared with the interictal EEG signals (0.3367 ± 0.0985). From Fig. 8(b), a decrease in

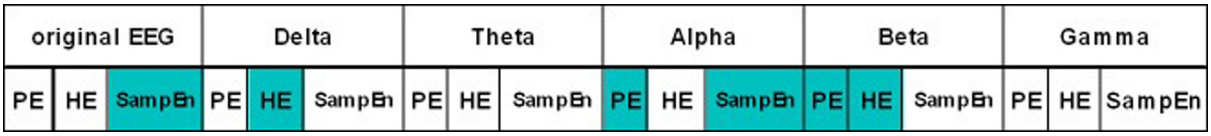

Fig. 7. The optimal feature subset select by genetic algorithm. Shaded blocks contributed the optimal feature subset while the rest were excluded from the feature set.

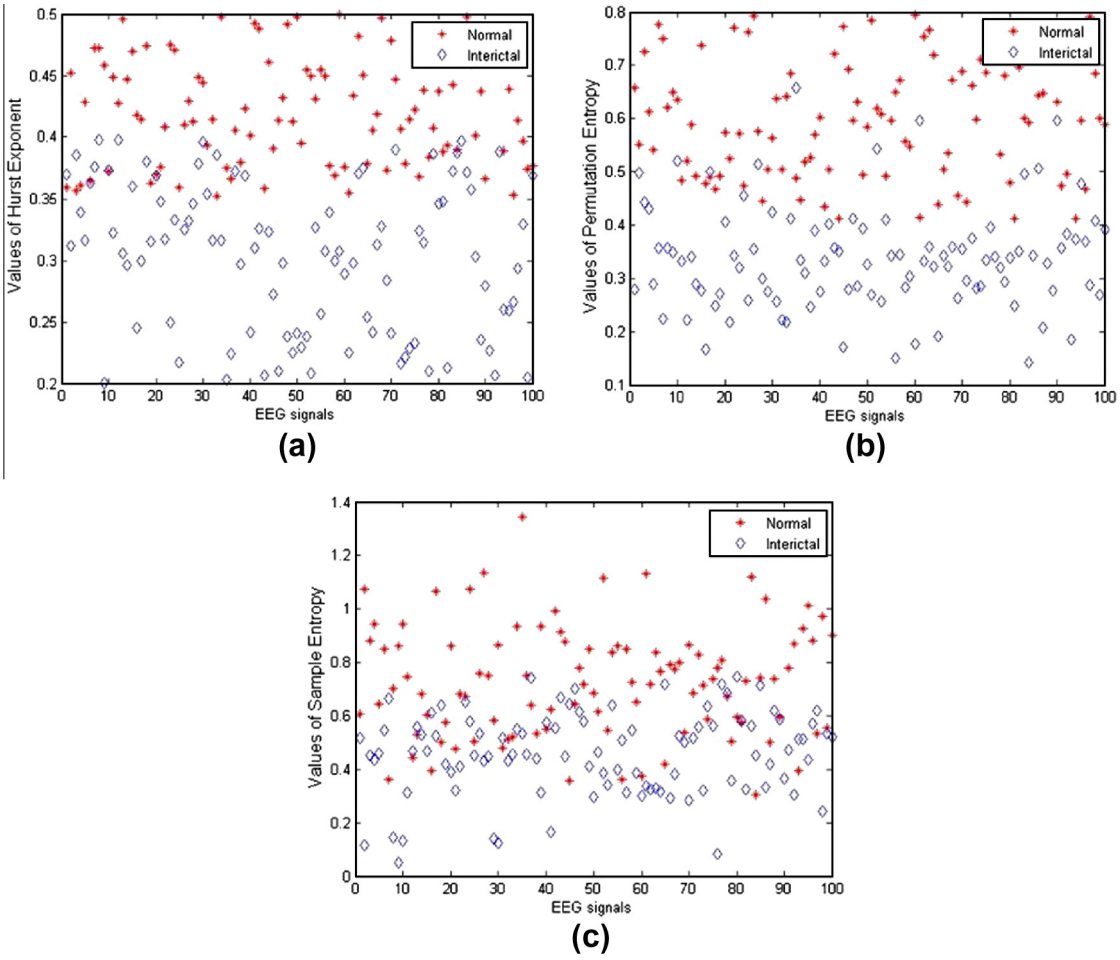

Fig. 8. Distributions of three non-linear feature values from certain EEG frequency bands between normal and interictal EEG signals. (a) The Hurst exponent values of normal and interictal EEG signals in delta frequency band. (b) Permutation entropy values of normal and interictal EEG signals in beta frequency band. (c) Sample entropy values of normal and interictal EEG signals in alpha frequency band. Feature values extracted from normal EEG signals and from interictal EEG signals were marked with 'star' and 'diamond', respectively.

permutation entropy values in the EEG beta frequency band can be seen, which indicated that brain neuronal activities in pathological states had more regular patterns than those in normal brain states, since they consist of repetitive activities with similar patterns.

The Hurst exponent is an indicator which has been utilized for depicting the correlation characteristics of physiological time series recordings. In light of Fig. 8(a), it can be found that the values of Hurst exponents of interictal EEGs (0.3037 ± 0.0605) were generally lower than those of normal EEG signals (0.4217 ± 0.0436). This phenomenon demonstrated that the anticorrelation intensity of the EEG signals in the delta frequency band turns stronger and the correlation intensity becomes weaker in abnormal brain states. The increased anticorrelation may be due to epileptiform discharges during seizure free states. The cause behind the epileptiform discharge is probably due to the prolonged dysfunction in the epileptic brain, which means that the epileptic brain is not normal even during seizure-free periods (Nurujjaman et al., 2009). From the statistical analysis results, it was concluded that these three non-linear parameters from certain specific EEG frequency bands which were selected by the genetic algorithm were capable of supplying different kinds of non-linear dynamic properties of the EEG signals, and the combined features were able to complement each other to obtain a better ability to differentiate between normal and abnormal brain states.

4. Conclusions

A novel feature extraction framework was presented for analysis of EEG time series data for epilepsy diagnosis. In light of experimental results, we found that all crucial components of the wavelet-complexity feature extraction methodology are of great importance for enhancing the performance of abnormal EEG signal analysis. Wavelet transform decomposed the original EEG time series into five frequency bands. Three complexity based features, namely permutation entropy, Hurst exponent and sample entropy were extracted from the original EEG time series and its five frequency bands, respectively. Although these features were not capable of identifying the epileptic EEG signals from normal ones individually, we found that the combination of some of these features from certain EEG frequency sub-bands improved the recognition performance greatly. A genetic algorithm was developed for feature selection. The performances of the recognition models were significantly improved after the genetic algorithm was integrated into the framework as a feature selector. The abnormal EEG diagnosis rate of the ELM without the use of the genetic algorithm was 85.9%. The diagnosis rate of ELM increased to 94.2% through using the genetic algorithm for feature selection. The genetic algorithm based feature optimization not only greatly reduced the feature space dimension (from 18 features to 6 features), but also played a key role in demonstrating the relative

importance of non-linear features extracted from different EEG frequency bands. The proposed approach was able to discriminate between the interictal brain states of epileptic subjects and the normal brain states effectively, which is considered to be both difficult and clinically important.

References

Abásolo, D., Hornero, R., Espino, P., Grant, C., & Ludbrook, G. (2006). Entropy analysis of the EEG background activity in Alzheimer's disease patients. *Physiological Measurement, 27*, 241–253.

Aboy, M., Cuesta-Frau, D., Austin, D., & Micó-Tormos, P. (2007). Characterization of sample entropy in the context of biomedical signal analysis. In *Proceedings of the 29th annual international conference of the IEEE engineering in medicine and biology* (pp. 5942–5945).

Acharya, U. R., Faust, O., Kannathal, N., Chua, T. L., & Laxminarayan, S. (2005). Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine, 80*, 37–45.

Adeli, H., Ghosh-Dastidar, S., & Dadmehr, N. (2007). A wavlet-chaos methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy. *IEEE Transactions on Biomedical Engineering, 54*(2), 205–211.

Adeli, H., Ghosh-Dastidar, S., & Dadmehr, N. (2010). *Automated EEG-based diagnosis of neurological disorders: Inventing the future of neurology*. CRC Press, Taylor & Francis Group.

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods, 123*, 69–87.

Andrzejak, R. G., Lehnertz, K., Rieke, C., Mormann, F., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E, 64*, 061907.

Bandt, C., & Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters, 88*(17), 174102.

Barlett, P. L. (1998). The sample complexity of pattern classification with neural networks: The size of the weight is more important than the size of the network. *IEEE Transactions on Neural Networks, 5*, 989–993.

Baum, E. B. (1988). On the capabilities of multilayer perceptrons. *Journal of Complexity, 4*, 193–215.

Bromfield, E. B. (2012). Epileptiform discharges. http://emedicine.medscape.com/article/1138880-overview.

Bruzzo, A. A., Gesierich, B., Santi, M., Tassinari, C. A., Birbaumer, N., & Rubboli, G. (2008). Permutation entropy to detect vigilance changes and preictal states from scalp EEG in epileptic patients: A preliminary study. *Neurological Science, 29*, 3–9.

Cao, Y., Tung, W., Gao, J. B., Protopopescu, V. A., & Hively, L. M. (2004). Detecting dynamical changes in time series using the permutation entropy. *Physical Review E, 70*, 046217.

Chandaka, S., Chatterjee, A., & Munshi, S. (2009). Cross-correlation aided support vector machine classifier for classification of EEG signals. *Expert Systems with Applications, 36*, 1329–1336.

Eckmann, J. P., & Ruelle, D. (1992). Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems. *Physica D, 56*, 185–187.

Eiben, A. E., & Smith, J. E. (2003). *Introduction to evolutionary computing*. Berlin, Heigelberg: Springer-Verlag.

Esteller, R., Echauz, J., Alessandro, M. D., Worrell, G., Cranstoun, S., Vachtsevanos, G., et al. (2005). Continuous Energy Variation During the Seizure cycle: Towards an online accumulated energy. *Clinical Neurophysiology, 116*, 517–526.

Gigolab, S., Ortiza, F., Dattellis, C. E., Silvac, W., & Kochenc, S. (2004). Prediction of epileptic seizures using accumulated energy in a multiresolution framework. *Journal of Neuroscience Methods, 138*, 107–111.

Guler, I., & Übeyli, E. D. (2007). Multiclass support vector machines for EEG signals classification. *IEEE Transactions on Information Technology in Biomedicine, 11*(2), 117–126.

Al-Angari, Haitham M., & Sahakian, A. V. (2007). Use of sample entropy approach to study heart rate variability in obstructive sleep apnea syndrome. *IEEE Transactions on Biomedical Engineering, 54*, 1900–1904.

Holland, J. H. (1975). *Adaption in natural and artificial systems*. Cambridge, MA: MIT Press.

Hsu, K. C., & Yu, S. N. (2010). Detection of seizures in EEG using subband nonlinear parameters and genetic algorithm. *Computers in Biology and Medicine, 40*, 823–830.

Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks, 17*(4), 879–892.

Huang, C. L., & Wang, C. J. (2006). A GA-based feature selection and parameters optimization for support vector machines. *Expert Systems with Applications, 31*, 231–240.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501.

Iasemidis, L. D., Shiau, D. S., Chaovalitwongse, W., Sackellares, J. C., Pardalos, P. M., Principe, J. C., et al. (2003). Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering, 50*(5), 616–627.

Kannathal, N., Acharya, U. R., Lim, C. M., & Sadasivan, P. K. (2005). Characterization of EEG-A comparative study. *Computer Methods and Programs in Biomedicine, 80*(1), 17–23.

Kiymi, M. K., Güler, I., Dizibüyük, A., & Akin, M. (2005). Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application. *Computers in Biology and Medicine, 35*(7), 603–616.

Li, X. L., Ouyang, G. X., & Richards, D. A. (2007). Predictability analysis of absence seizures with permutation entropy. *Epilepsy Research, 77*, 70–74.

Lopes da Silva, F. H., Pijn, J. P., & Wadman, W. J. (1994). Dynamics of local neuronal networks: Control parameters and state bifurcations in epileptogenesis. *Progress in Brain Research, 102*, 359–370.

Mormann, F., Kreuz, T., Andrzejak, R. G., Peter, D., Lehnertz, K., & Elger, C. E. (2003). Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Research, 53*, 173–185.

Nicolaou, N., & Georgiou, J. (2012). Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Systems with Applications, 39*, 202–209.

Nigam, V. P., & Graupe, D. (2004). A neural-network-based detection of epilepsy. *Neurological Research, 26*(1), 55–60.

Nurujjaman, M., Narayanan, R., & Iyengar, A. S. (2009). Comparative study of nonlinear properties of EEG signals of normal persons and epileptic patients. *Nonlinear Biomedical Physics, 3*(6). http://dx.doi.org/10.1186/1753-4631-3-6.

Ocak, H. (2008). Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal Processing, 88*, 1858–1867.

Olofsen, E., Sleigh, J. W., & Dahan, A. (2008). Permutation entropy of the electroencephalogram: A measure of anaesthetic drug effect. *British Journal of Anaesthesia, 101*, 810–821.

Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform. *Applied Mathematics and Computation, 187*, 1017–1026.

Polat, K., & Güneş, S. (2008). Artificial immune recognition system with fuzzy resource allocation mechanism classier, principle component analysis and FFT method based new hybrid automated identification system for classification of EEG signals. *Expert Systems with Applications, 34*(3), 2039–2048.

Van Quyen, M. L., Martinerie, J., Baulac, M., & Varela, F. (1999). Anticipating epileptic seizure in real time by non-linear analysis of similarity between EEG recordings. *NeuroReport, 10*(10), 2149–2155.

Rao, C. R., & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. Wiley.

Richman, J. S., & Moorman, J. R. (2000). Physiological time series analysis using approximate entropy and sample entropy. *American Journal of Physiology, 278*(6), 2039–2049.

Srinivasan, V., Eswaran, C., & Sriraam, N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. *Journal of Medical Systems, 29*(6), 647–660.

Stam, C. J., Pijn, J. P., Suffczynski, P., & Lopez da Silva, F. H. (1999). Dynamics of the human alpha rhythm: Evidence for non-linearity. *Clinical Neurophysiology, 110*(10), 1801–1813.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications, 32*, 1084–1093.

Übeyli, E. D. (2010). Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. *Expert Systems with Applications, 37*, 233–239.

Wang, C. M., Zou, J. Z., Zhang, J., Wang, M., & Wang, R. B. (2010). Feature extraction and recognition of epileptiform activity in EEG by combining PCA with ApEn. *Cognitive Neurodynamics, 4*, 233–240.