Epilepsy Research (2011) **96**, 29—38

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

Epileptic EEG classification based on extreme learning machine and nonlinear features

Qi Yuan [1], Weidong Zhou*, Shufang Li [2], Dongmei Cai [3]

School of Information Science and Engineering, Shandong University, Jinan 250100, China

Received 21 February 2011; received in revised form 19 April 2011; accepted 24 April 2011
Available online 25 May 2011

KEYWORDS
Epileptic EEG;
Approximate entropy (ApEn);
Hurst exponent;
Detrended fluctuation analysis (DFA);
Extreme learning machine (ELM);
Support vector machine (SVM)

Summary The automatic detection and classification of epileptic EEG are significant in the evaluation of patients with epilepsy. This paper presents a new EEG classification approach based on the extreme learning machine (ELM) and nonlinear dynamical features. The theory of nonlinear dynamics has been a powerful tool for understanding brain electrical activities. Nonlinear features extracted from EEG signals such as approximate entropy (ApEn), Hurst exponent and scaling exponent obtained with detrended fluctuation analysis (DFA) are employed to characterize interictal and ictal EEGs. The statistics indicate that the differences of those nonlinear features between interictal and ictal EEGs are statistically significant. The ELM algorithm is employed to train a single hidden layer feedforward neural network (SLFN) with EEG nonlinear features. The experiments demonstrate that compared with the backpropagation (BP) algorithm and support vector machine (SVM), the performance of the ELM is better in terms of training time and classification accuracy which achieves a satisfying recognition accuracy of 96.5% for interictal and ictal EEG signals.

Introduction

Epilepsy is a common and chronic neurological disorder which is characterized by the recurrence of sudden abnormal reactions of the brain as epileptic seizures (Niedermeyer and Lopes, 2005). The epileptic seizures resulting from sudden excessive electrical discharges in a group of neurons may be accompanied by loss of awareness or consciousness and disturbances of movement, sensation, mood, or mental function (Misra and Kalita, 2005; Sanei and Chambers, 2007). About 1% of the world population suffers from epilepsy (Lehnertz et al., 2003). The electroencephalogram (EEG) reflects the electrical activity of the brain recorded by amplifying voltage differences between electrodes placed on the scalp or cerebral cortex (Misra and Kalita, 2005). EEG contains a mass of physiological and pathological information, which plays an important role in the evaluation of patients with epilepsy such as determining epileptogenic zone for presurgical evaluations (Lehnertz et al., 2003; Gandhi et al., 2010). At present, analysis of EEG relies

* Corresponding author at: 27 Shanda Road, Jinan, PR China. Tel.: +86 531 86333629.
E-mail addresses: qiyuan@mail.sdu.edu.cn (Q. Yuan), wdzhou@sdu.edu.cn (W. Zhou), lishufang@mail.sdu.edu.cn (S. Li), caidongmei@mail.sdu.edu.cn (D. Cai).
[1] 27 Shanda Road, Jinan, PR China. Tel.: +86 15806609671.
[2] 27 Shanda Road, Jinan, PR China. Tel.: +86 13361020357.
[3] 27 Shanda Road, Jinan, PR China. Tel.: +86 15966304164.

0920-1211/$ — see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.eplepsyres.2011.04.013

on neurologists to check the EEG recordings visually. The visual scoring of long term EEG is very tedious and time-consuming. Therefore, the automatic detection technology is valuable for assisting neurologists to analyze the EEG recordings.

The methods for automatic epileptic seizure detection have been under study for several years. The first widely applicable technique and its improvements were developed by Gotman (1982, 1990). Gabor et al. (1996) used wavelet filters to extract the frequency features from EEGs and then combined with a self-organizing neural network for classification, while Weng and Khorasani (1996) applied the features proposed by Gotman (1982) to an adaptive structure neural network. As a self-adaptive pattern classification method, artificial neural network (ANN) makes few demands on statistical characteristics of signals, and is fit for complex nonlinear mapping (Übeyli, 2009). However, the conventional learning algorithms for ANN, such as the back-propagation (BP) algorithm, are prone to fall into a local minimum, and their learning speeds are too slow to meet requirements of real-time applications, which has been a major bottleneck for their development. The support vector machine (SVM) founded on statistical learning theory implements structural risk minimization and has relatively strong generalization capability, whose idea is to map nonlinear separable samples onto another higher-dimensional space, and then locate the maximum margin hyperplane in the projection space (Pan et al., 2008; Swiderski et al., 2008). But if the sample size is large, the learning speed of the SVM will also be very slow. The extreme learning machine (ELM), which not only avoids falling into local optima but also advances the learning speed of the network (Huang et al., 2006), recently has been used to train a single layer feedforward network (SLFN) for identifying mental tasks based on EEG signals (Liang et al., 2006). The ELM algorithm abandons the traditional strategy of adjusting all the parameters of the network iteratively, but determines the weights between hidden neurons and output neurons of the SLFN analytically (Huang et al., 2004, 2006).

The suitable features extracted from EEG signals are essential for an excellent seizure detection technology. Recently the nonlinear dynamics and deterministic chaos theory provide new ways for the analysis of EEG signals. Radhakrishnan and Gangadhar (1998) used approximate entropy (ApEn) to estimate regularity in epileptic seizure time series data. Then ApEn was applied to recognize epileptic activity in EEG recordings (Diambra et al., 1999). Hurst exponent has been used to quantify self-similarity in a time-series. Kannathal et al. (2005a) have shown that Hurst exponent is a good measure for discriminating normal and epileptic EEG signals. The detrended fluctuation analysis (DFA) provides a technique for studying long-range temporal correlations (LRTC) in a non-stationary EEG time series. It has been applied to sleep EEG by Lee et al. (2002) for calculating the scaling exponents which can characterize the variance of sleep stages. The nonlinear dynamics parameters, such as, ApEn, Hurst exponent and scaling exponent obtained with DFA can give information about rhythmicity, regularity and long-range correlation in EEG time series.

In this paper, we evaluate the representative ability of the chaotic measures for interictal and ictal EEG, and classification property of the ELM algorithm. ApEn, Hurst exponent and scaling exponent are combined to extract quantitative nonlinear features from EEG signals. In the proposed approach, these three features are chosen as input to a single hidden layer feedforward network (SLFN), and the ELM is adopted to train the SLFN for classifying clinical interictal and ictal EEG signals. The performance of the ELM, BP algorithm and SVM are compared in terms of classification performance, training and testing time. Different from the work of Song and Liò (2010), we investigate the discriminatory capability of the three nonlinear features other than sample entropy used in their method, and further more the classification method based on the combination of the features with the ELM classifier is assessed.

EEG data selection

The EEG signals used in this paper come from Department of Epileptology, Bonn University, Germany. The dataset has been described by Andrzejak et al. (2001) and employed in epileptic diagnosis and analysis research widely. The complete dataset contains five sets of data denoted sets A—E, each consisting of 100 single-channel EEG segments of 23.6 s. All the EEG segments were selected and cut out from continuous multichannel EEG recordings after visual inspection for artifacts (e.g., due to hand or eye movements). The EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were sampled at 173.61 Hz with 12-bit A/D resolution.

In this study, two sets (D and E) of the complete dataset are examined. Sets D and E were recorded from five patients intracranially. Set D contains only signals measured in seizure-free intervals, while set E only contains EEG recordings during seizure activity. Both the sets were from the epileptogenic zone. Each segment includes 4096 sampling points. The classification of interictal (Set D) and ictal (Set E) EEGs is more difficult to solve but closer to clinical applications than the other classification problems based on this dataset.

Feature extraction

Approximate entropy

Approximate entropy (ApEn) is a statistical parameter that provides a general understanding of the regularity of a time series. ApEn employs a non-negative number to quantify the complexity of data and the creation of information in a time series. The larger the value is, the more complex or irregular the time series data is (Kannathal et al., 2005b; Ki et al., 2009). The general process of estimating ApEn is given by Appendix A.

ApEn measures the difference in the time series between the probability of modes of length m that are close to each other and that of modes of length $m+1$ that are close to each other. Hence it reflects the probability of creating a new mode while the vector dimension in the embedding space is varying. The bigger the probability is, the more complex the time series is. Fig. 1(A) illustrates the ApEns of interictal and ictal EEG epochs each of which contains 1024 sampling points. From Fig. 1(A) we can see that the values of the ApEn

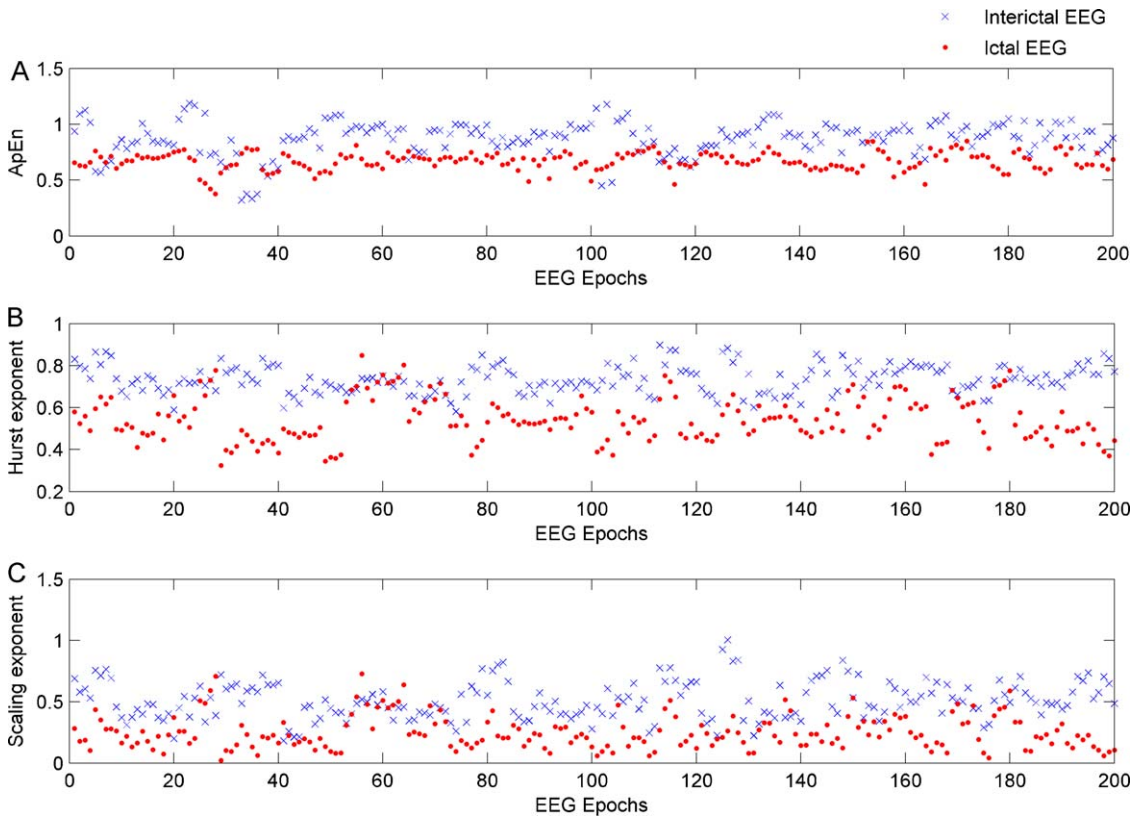

Figure 1 The comparison of the nonlinear features between interictal and ictal EEG signals. (A) ApEns of interictal and ictal EEGs. (B) Hurst exponents of interictal and ictal EEGs. (C) Scaling exponents of interictal and ictal EEGs.

of ictal EEGs are generally lower than those of interictal EEGs.

Hurst exponent

Hurst exponent is an indicator that has been used to describe the correlation properties and self-similarity of physiological time series data. Hurst exponent is a measure of the smoothness of a fractal time series based on the asymptotic behavior of the rescaled range of the process (Acharya et al., 2005; Wang et al., 2006), and it can be determined by R/S algorithm which is described in Appendix B.

In the analysis of the dynamic cerebral processes, Hurst exponent H can be employed to characterize the degree of long-range dependence in EEG time series. If $H = 0.5$, the analyzed time series is similar to a random walk. If $0 < H < 0.5$, the behavior of the time series exhibits antipersistence, i.e., if the time-series increases, it acquires a higher likelihood of decreasing in the future time, and vice versa. And closer to zero is H, the antipersistent density is stronger. If $0.5 < H < 1$, there are persistent effects in the behavior of the time series so that the process displays a tendency that if the time-series increases, it is more probable that then it will keep on increasing, and vice versa. And the degree that H is close to 1 reflects the persistent density (Acharya et al., 2005; Wang et al., 2006). The Hurst exponents of interictal and ictal EEG epochs which are also used for estimating ApEns are as shown in Fig. 1(B).

Detrended fluctuation analysis

As a technique for quantifying long-range temporal correlations in EEG time series, detrended fluctuation analysis (DFA) not only has robustness against non-stationarity but also can systematically remove trends of various order aroused by the noise of EEG time series (Jiang et al., 2005; Leistedt et al., 2007). The main idea of DFA is to compute the average root-mean-square fluctuation of the time series. Appendix C provides the procedure of DFA proposed by Peng et al. (1995).

The result of DFA is called the scaling exponent denoted by α. When $\alpha = 0.5$, the EEG signal is a random walk. When $0 < \alpha < 0.5$, there are power-law anticorrelations in the EEG signal. When $0.5 < \alpha < 1$, long-range power-law correlations are present (Leistedt et al., 2007; Lee et al., 2007). The scaling exponents of interictal and ictal EEG epochs which are also used for estimating ApEns and Hurst exponents are exhibited in Fig. 1(C).

Extreme learning machine

All the parameters of the traditional feedforward neural networks need to be tuned commonly by the gradient-based learning algorithms. However, with the goal of better learning performance, the procedure of training the neural networks is repeated so many times that the learning speed is extremely slow and it is easy to fall into local optima. Although adding a momentum term to the weight adjust-

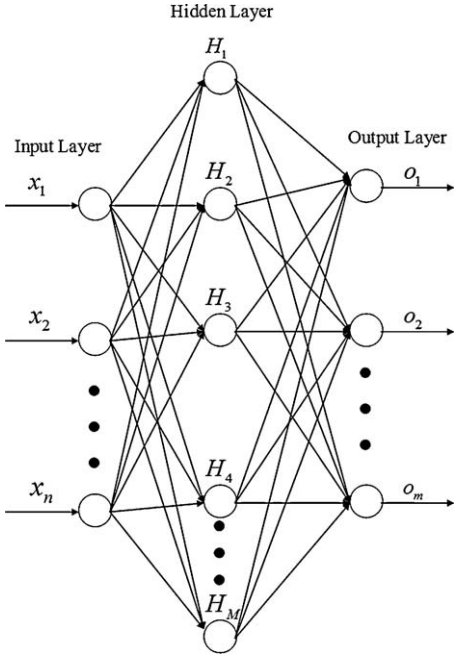

Figure 2 A diagram of a single hidden layer feedforward neural network (SLFN).

ment can reduce the risk of the network being trapped in local optima, the time spent on the training process is not decreased. The research indicates that a single hidden layer feedforward neural network (SLFN) with N hidden neurons can learn N different samples with any small error, whose input weights and hidden layer biases are assigned randomly. Thus only the output weights need to be adjusted, and the adjustments to the input weights and hidden layer biases make no contribution to the performance of the SLFN (Huang et al., 2004, 2006). The extreme learning machine (ELM) selects the input weights and hidden layer biases randomly, and then the output weights can be obtained analytically. Hence the ELM has much faster learning speed and better generalization performance on some tasks than conventional learning algorithms (Huang et al., 2004, 2006).

Single hidden layer feedforward neural networks

The SLFN structure is illustrated in Fig. 2. Given N distinct training samples $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^N$, where $\mathbf{x}_k = [x_{k1}, x_{k2}, \ldots, x_{kn}]^T$ is the kth sample and $\mathbf{t}_k = [t_{k1}, t_{k2}, \ldots, t_{km}]^T$ denotes the label of $\mathbf{x}_k$, the mathematical model of SLFNs with M hidden neurons can be defined as

$$\sum_{i=1}^{M} \beta_i g(w_i \cdot x_k + b_i) = o_k, k = 1, \ldots, N \tag{1}$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ and $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ are the input and output weights respectively, $\mathbf{o}_k = [o_{k1}, o_{k2}, \ldots, o_{km}]^T$ is the output vector of the network, and b_i is the threshold of the hidden neuron. $\mathbf{w}_i \cdot \mathbf{x}_k$ represents the inner product of $\mathbf{w}_i$ and $\mathbf{x}_k$, $g(\bullet)$ denotes the activation function.

If the SLFN approach the N training samples with zero error, i.e., $\sum_{K=1}^{N} \|\mathbf{o}_k - \mathbf{t}_k\| = 0$, there are $\mathbf{w}_i$, $\boldsymbol{\beta}_i$ and b_i satisfying

$$\sum_{i=1}^{M} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_k + b_i) = \mathbf{t}_k, k = 1, \ldots, N \tag{2}$$

The above N equations can be replaced by

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \tag{3}$$

where

$$\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_M, b_1, \ldots, b_M, \mathbf{x}_1, \ldots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_M \cdot \mathbf{x}_1 + b_M) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_M \cdot \mathbf{x}_N + b_M) \end{bmatrix}_{N \times M} \tag{4}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_M^T \end{bmatrix}_{M \times m} \quad T = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \tag{5}$$

$\mathbf{H}$ denotes the hidden layer output matrix of the neural network.

Huang et al. (2006) have proved that given arbitrarily small value $\varepsilon > 0$, if the hidden layer activation function of the SLFN is infinitely differentiable, and the number of hidden neurons is $M \leq N$, the input weights and hidden layer biases can be evaluated randomly, then the SLFN approximates the N training samples $(\mathbf{x}_i, \mathbf{t}_i)$ with ε error, i.e., $\left\|\mathbf{H}_{N \times M}\boldsymbol{\beta}_{M \times m} - \mathbf{T}_{N \times m}\right\| \leq \varepsilon$.

Extreme learning machine (ELM) algorithm

For the SLFN with the infinitely differentiable activation function, the ELM randomly choose the input weights $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^N$ and hidden layer biases b_i, i.e., the hidden layer output matrix $\mathbf{H}$ has been randomly fixed, while the procedure of training an SLFN is equivalent to seeking a least-squares solution of the linear system $\mathbf{H}\,\boldsymbol{\beta} = \mathbf{T}$:

$$\left\|\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_M, b_1, \ldots, b_M)\hat{\boldsymbol{\beta}} - \mathbf{T}\right\| = \min_{\boldsymbol{\beta}} \left\|\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_M, b_1, \ldots, b_M)\boldsymbol{\beta} - \mathbf{T}\right\| \tag{6}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{H}^{+}\mathbf{T}$ is the smallest norm least-squares solution of $\mathbf{H}\,\boldsymbol{\beta} = \mathbf{T}$, and $\mathbf{H}^{+}$ denotes the Moore-Penrose generalized inverse of $\mathbf{H}$. The procedure of the ELM is described in Appendix D.

For feedforward networks, the smaller their weights are, the stronger generalization ability they can have (Bartlett, 1998). The norm of $\hat{\boldsymbol{\beta}}$ is the smallest among all the least-squares solutions of the linear system $\mathbf{H}\,\boldsymbol{\beta} = \mathbf{T}$, that is to say

$$\left\|\hat{\boldsymbol{\beta}}\right\| = \left\|\mathbf{H}^{+}\mathbf{T}\right\| \leq \|\boldsymbol{\beta}\|, \forall \boldsymbol{\beta} \in \left\{\boldsymbol{\beta} : \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\| \leq \|\mathbf{H}\mathbf{z} - \mathbf{T}\|, \forall \mathbf{z} \in R^{M \times m}\right\} \tag{7}$$

Thus the ELM can not only achieve the smallest training error but also gain better generalization performance than the conventional gradient-based learning algorithms.

And because of the uniqueness of Moore-Penrose generalized inverse $\mathbf{H}^{+}$, the special solution $\hat{\boldsymbol{\beta}}$ is unique.

Results

All the experiments were carried out in MATLAB7.0 environment running in an Intel Pentium processor with 3.00 GHz. Every EEG segment of 4096 sampling points for each class (ictal and interictal) was divided into equal four epochs. Then there were 800 epochs (400 epochs from each class) containing 1024 samples. Half of the data per class were randomly chosen for training and the remaining for testing. Fifty diverse training-test sets were randomly built for the classification.

The performance of the three classifiers can be assessed by the statistical measures of sensitivity, specificity and recognition accuracy.

- Sensitivity: number of true positives/the total number of ictal samples labeled by the EEG experts. True positive represents the ictal EEG identified by the computer programs and by the EEG experts.
- Specificity: number of true negatives/the total number of interictal samples labeled by the EEG experts. True negative represents the interictal EEG identified by the computer programs and by the EEG experts.
- Recognition accuracy: number of correctly identified patterns/total number of patterns.

Both the ELM-trained and the BP-trained SLFNs that can achieve the best classification results in this experiment have 10 hidden neurons, and the sigmoid function is chosen as the hidden layer activation function for the two SLFNs. There is one node in the output layer whose target value is defined as 1 or -1 for the two networks. The interictal EEG is denoted as 1 and the ictal EEG is denoted as -1. The parameters of the BP network were tuned 200 times iteratively with batch training in which weights and biases are only updated after all the training samples are presented. For the SVM classifier, radial basis function (RBF) was chosen as kernel function and optimal parameters were found in training step for every trial. The definition of RBF is $\exp(-(1/\sigma^2)\|\mathbf{x}-\mathbf{x}_i\|^2)$.

The ApEn, Hurst exponent and scaling exponent of EEG data used for training were respectively input to the SLFN and SVM classifier as the nonlinear features. The conventional back-propagation (BP) algorithm with self-adaptive learning rate (Bhattacharya and Parui, 1995) and extreme learning machine (ELM) were used for training the SLFN. The three features of the test data set were separately employed to investigate the performance of the three different classifiers. The average results of fifty classification implementations for the three different nonlinear features are listed in Tables 1–3.

It can be noticed from the three tables that the recognition accuracies are almost same when using ELM, BP network and SVM classifiers, respectively. However, the time consumed by the SVM classifier is the longest, and the time spent by BP network is over 10 times longer than the time by the ELM. The performance of the ELM is better than the other two classifiers in terms of efficiency.

Further, feature vectors are composed with the ApEns, Hurst exponents and scaling exponents of EEG and fed into ELM, BP and SVM classifiers for the EEG classification evaluation, respectively. The mean results of fifty executions under the same experimental conditions are shown in Table 4.

Comparing Tables 1–3 with Table 4, we can observe that the combination features of ApEn, Hurst exponent and scaling exponent can remarkably enhance the recognition accuracy. The union of the three nonlinear features has stronger discriminatory power for interictal and ictal EEG signals, and makes the three different classifiers obtain relatively ideal results. It is found that the classification accuracy of the SVM is a little worse than that of the ELM and BP network, when the united feature vectors are used. The ELM outperforms BP and SVM with the best classification performance and the highest speed.

Table 5 presents a comparison on recognition accuracy between the methods proposed in this work and different methods based on autoregressive (AR) model of the EEG signals. AR model has been widely employed for EEG analysis (Pardey et al., 1996). The coefficients of the sixth order AR model for each EEG segment are calculated as the input features for the three classifiers. The results in the table show that our method improves recognition accuracy comparing the methods based on AR model, indicating the better performance of our method.

Discussion

Billions of neurons in the human brain are connected together with axons and synapses to construct a complex neural system. It is well known that the function of the neural system is not equivalent to the simple linear superposition of the functions of all the neurons. Hence brain electrical activities have very complex dynamic properties. EEG, as recording of brain electrical activity, exhibits typically complicated dynamics. Although linear methods for the analysis of EEG signals have obtained relatively good results, they could only provide linear features of EEG signals and could not reflect the nonlinear behavior of the neural system. Therefore, nonlinear analysis methods have attracted more attention for improving characterizations of EEG signals in recent years. In this work, three nonlinear measures namely ApEns, Hurst exponents and scaling exponents were evaluated as EEG features and applied for EEG recognition.

Because of synchronous discharge of large groups of neurons during an epileptic seizure, the loss of complexity in ictal EEG results in the decrease of ApEns (Diambra et al., 1999), which can be observed in Fig. 1(A). It can be found from Fig. 1(B) that the Hurst exponents of ictal EEG are generally lower than interictal EEG's. This phenomenon shows that the antipersistent density of the EEG signal becomes stronger and the persistent density turns weaker during an epileptic seizure. We can clearly see smaller values of the scaling exponents obtained with DFA for the most of ictal EEG epochs in Fig. 1(C). The decrease of the scaling exponent means the increase of rhythmicity and regularity in EEG time series during an epileptic activity. Those three nonlinear features can provide different types of nonlinear properties of EEG signals, and the combined features could complement each other to achieve stronger discrim-

Table 1 Experimental results using ApEn.

Classifier	Sensitivity (%)	Specificity (%)	Recognition accuracy (%)	Time (s)	
				Training	Testing
ELM	93.00 ± 1.00	83.50 ± 1.00	88.00 ± 0.75	0.0714	0.0131
BP	93.00 ± 0.50	82.00 ± 1.00	88.00 ± 0.75	1.5369	0.0256
SVM	93.75 ± 0.25	82.50 ± 1.00	88.00 ± 0.75	11.9764	3.5236

Table 2 Experimental results using Hurst exponent.

Classifier	Sensitivity (%)	Specificity (%)	Recognition accuracy (%)	Time (s)	
				Training	Testing
ELM	81.00 ± 2.00	95.00 ± 1.00	88.00 ± 0.50	0.0729	0.0120
BP	81.00 ± 0.50	94.00 ± 1.00	88.00 ± 0.50	1.5864	0.0230
SVM	79.75 ± 0.75	96.00 ± 1.00	88.00 ± 0.50	11.5000	3.5250

Table 3 Experimental results using DFA scaling exponent.

Classifier	Sensitivity (%)	Specificity (%)	Recognition accuracy (%)	Time (s)	
				Training	Testing
ELM	73.50 ± 0.50	90.00 ± 0.50	82.00 ± 0.50	0.0736	0.0129
BP	73.50 ± 0.50	89.75 ± 0.25	81.75 ± 0.50	1.5476	0.0243
SVM	73.50 ± 0.50	89.75 ± 0.25	81.75 ± 0.50	10.6000	3.5000

Table 4 Experimental results using the united features of ApEn, Hurst exponent and scaling exponent.

Classifier	Sensitivity (%)	Specificity (%)	Recognition accuracy (%)	Time (s)	
				Training	Testing
ELM	92.50 ± 2.00	96.00 ± 2.50	96.00 ± 0.50	0.0803	0.0135
BP	91.50 ± 3.00	94.00 ± 3.50	95.50 ± 0.50	1.6363	0.0256
SVM	95.00 ± 2.00	93.75 ± 0.25	95.25 ± 0.25	12.6410	3.6406

inatory capability for EEG classification. Our experimental results show the highest accuracy of 96.5% was obtained with those features composed by the ApEns, Hurst exponents and scaling exponents of EEG.

To gain further insight of EEG nonlinear features, one-way analysis of variance (ANOVA) which is a powerful and common statistical technique used to compare means of two or more groups is employed to examine whether the ApEns, Hurst exponents and scaling exponents differ significantly between interictal and ictal EEGs. A one-way ANOVA for each nonlinear feature is performed by SPSS version 13.0 software. The mean values and standard deviations of those nonlinear features are illustrated in Fig. 3.

Fig. 3 confirms that the means of all the nonlinear features of interictal EEG signals are higher than those of ictal EEG signals. The *p*-values resulting from the one-way ANOVA

Table 5 A comparison of recognition accuracy obtained by our method and AR model method.

Method	Recognition accuracy (%)
Coefficients of the AR model + BP network	91.75
Coefficients of the AR model + SVM	91.50
Coefficients of the AR model + ELM	91.75
Nonlinear features + ELM (this work)	96.50

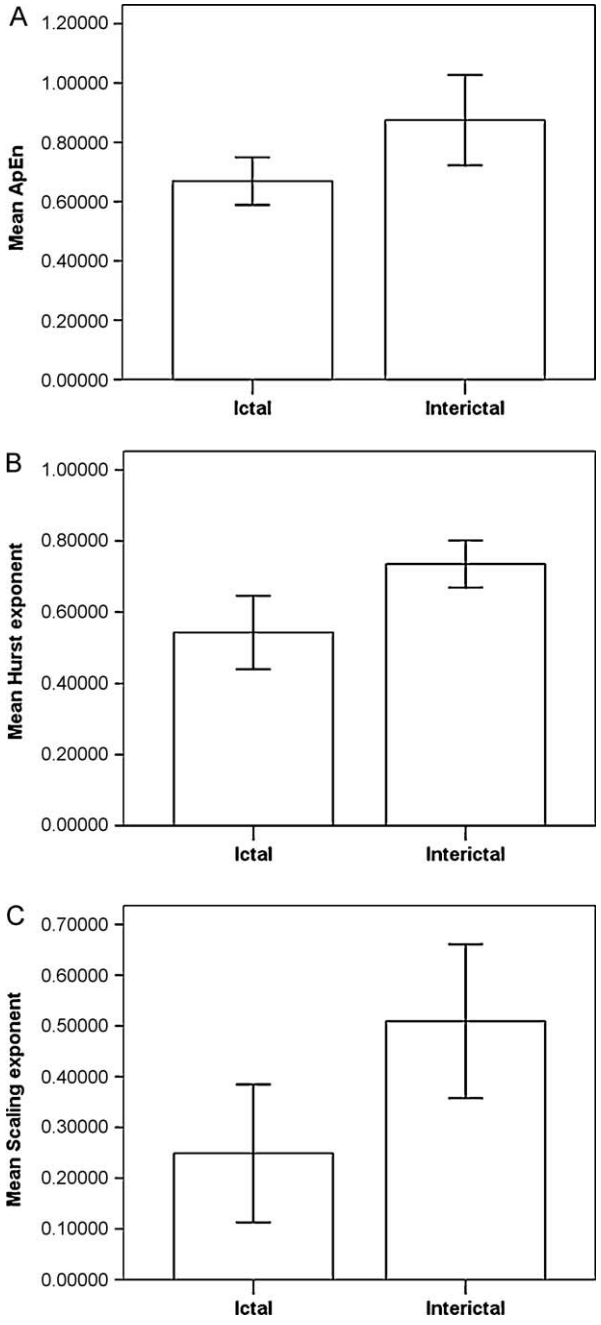

Figure 3 The mean values and standard deviations of the nonlinear features extracted from the EEG samples. (A) The mean values and standard deviations of ApEns. (B) The mean values and standard deviations of Hurst exponents. (C) The mean values and standard deviations of scaling exponents.

test for the ApEns, Hurst exponents and scaling exponents are all less than 0.01, which is a strong indication that the differences of those nonlinear features between interictal and ictal EEGs are statistically significant.

Besides, it can be noticed from Tables 1—3 that the values of sensitivity are much higher than those of specificity for ApEn, but the cases are opposite for Hurst exponent and scaling exponent. This phenomenon indicates that the complexity measured by the ApEn is more sensitive to the ictal EEG and the correlation property and self-similarity quantified by the Hurst exponent and scaling exponent is more responsive to the interictal EEG. To achieve better recognition accuracy, the feature vectors composed by the ApEn, Hurst exponent and scaling exponent of the EEG signals could remarkably enhance the recognition accuracy. Taking both sensitivity and specificity into account, we can observe from comparing Tables 1—3 with Table 4 that the union of the three nonlinear features makes the three different classifiers obtain better results than the separate one.

As observed from Tables 1—3 the recognition accuracies are almost same when using the ELM, BP network and SVM classifiers, respectively. However, the time consumed by the ELM classifier is the shortest. The main reasons why the ELM is better than the other two classifiers in terms of efficiency lie in: unlike the conventional gradient-based learning algorithms which tune the parameters of the networks iteratively, the ELM randomly chooses the input weights and hidden layer biases, and then analytically determines the output weights without needing complex iterative operations. The results from Table 4 verify that the ELM achieves the fastest learning speed among the three classifiers. For the real-time detection system for ambulatory EEG recordings, the time spent by calculation of the features and seizure activity judgment for one analyzed EEG segment should be not longer than time span of the EEG segment. In this work, the time duration of an EEG epoch used in the experiments is 5.9392 s. The sum of computing time of the nonlinear features and testing time of the ELM classifier for one EEG epoch is 0.7813 s which is far shorter than its time duration. The low computational burden makes it possible to build a real-time detection system based on the presented method.

Although our work focuses on the classification of EEG segments, future research is under way to devise an online automatic seizure detection system for clinical use based on the proposed method. For a detection system, firstly a moving-window technique is employed to divide long term EEG into non-overlapping EEG segments of length equal to that used for training. Then the proposed classification method is carried out for each EEG segment. However, there will be a heavily class-imbalanced problem with long term EEG detection, i.e., seizure segments are extremely rare compared to non-seizure ones. This may cause the classifier to generate too many false-positives to be practically useful. Although the class-imbalanced problem is not involved in the present framework, the following technical scheme can be suggested. In many cases, EEG data are recorded with several electrodes in different sites of the focal area. If a seizure detection system can take the classification results from all channels into account and make the decision with a predefined criterion, the number of false-positives is likely to decrease due to the integration of information obtained from multichannel EEGs.

Conclusions

This paper investigates nonlinear measures of EEG, which are ApEn, Hurst exponent and scaling exponent, and com-

bines those nonlinear features with the ELM classifier to discriminate the interictal and ictal EEG signals. The ApEn, Hurst exponent and scaling exponent obtained with DFA are suitable for extracting the nonlinear and non-stationary properties of EEG time series. The one-way ANOVA test for the ApEns, Hurst exponents and scaling exponents indicates that the differences of those nonlinear features between interictal and ictal EEG are statistically significant. The union of the three nonlinear features can make a distinction between the interictal and ictal EEG signals.

Compared with conventional learning algorithms of SLFNs, the ELM calculates output weights analytically. Due to avoiding complex iterative operation, the ELM has extremely fast speed, and cannot fall into oscillation or local optima. In the experiments, EEG feature vectors consisted of the ApEn, Hurst exponent and scaling exponent are employed to train and test the SVM, BP and ELM classifiers. The results show that the ELM could reach higher recognition accuracy, and spend far less time than the BP network and SVM classifier. Considering excellent performance of the ELM, the proposed approach demonstrates its potential for real-time seizure detection.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 30870666), the National Key Technology R&D Program of China (No. 2008BAI52B03), the Independent Innovation Foundation of Shandong University (No. 2009JC004), and the Program of Development of Science and Technology of Shandong (No. 2010GSF10243).

Appendix A. Approximate entropy

Suppose the original time series with N data points $\{x(l), l = 1, \ldots, N\}$. The calculation of the ApEn is shown in the following steps (Pincus, 1991; Wang et al., 2009).

(1) Construct a series of vectors in the embedding space R^m defined by

$$X(i) = [x(i), x(i+1), \ldots, x(i+m-1)] \tag{A.1}$$

where $1 \le i \le N-m+1$.

(2) Compute the distance between $X(i)$ and $X(j)$,

$$d_{ij} = \max_{k=0\sim m-1} \left|x(i+k) - x(j+k)\right| \tag{A.2}$$

where $0 \le k \le m-1$ and $1 \le i, j \le N-m+1$.

(3) Given the vector comparison distance r ($r > 0$), for each $X(i)$, count the number of $d_{ij} \le r$, denoted as $N_i^m(r)$. And then define the ratio between the number $N_i^m(r)$ and the total number of the vectors as $C_i^m(r)$,

$$C_i^m(r) = \frac{N_i^m(r)}{N-m+1} \tag{A.3}$$

(4) Compute the natural logarithm of the ratio $C_i^m(r)$, then average it over i,

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \tag{A.4}$$

(5) Increase m by one and repeat steps (1)–(4). Thereby, $C_i^{m+1}(r)$ and $\Phi^{m+1}(r)$ are obtained.

(6) ApEn is given by $\Phi^m(r)$ and $\Phi^{m+1}(r)$ as follows:

$$ApEn(m, r) = \lim_{N\to\infty} [\Phi^m(r) - \Phi^{m+1}(r)] \tag{A.5}$$

(7) If the number of data point N is limited, ApEn is estimated by the statistic values, i.e.,

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{A.6}$$

Based on the works of Pincus (1991) and Srinivasan et al. (2007), the embedding dimension m and vector comparison distance r were respectively set to 2 and 0.05 times the standard deviation of the EEG time series.

Appendix B. R/S algorithm for Hurst exponent

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

$$W_k = (x_1 + x_2 + x_3 + \cdots + x_k) - k\bar{x}(n) \tag{B.1}$$

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

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

$$R(n) = \max(0, W_1, \ldots, W_n) - \min(0, W_1, \cdots, W_n), n = 1, \ldots, N \tag{B.2}$$

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

$$\frac{R(n)}{S(n)} = C \times n^H, n = 1, \ldots, N \tag{B.3}$$

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

$$H = \frac{\log\ [R(n)/S(n)]}{\log(n)}, n = 1, \ldots, N \tag{B.4}$$

H can be calculated by least-squares linear fit in the $\frac{\log[R(n)/S(n)]}{\log(n)}$ plot.

Appendix C. Detrended fluctuation analysis

Given a discrete time series $\{x(i), i = 1, \ldots, N\}$, the process of DFA as follows:

(1) The integrated series is calculated by

$$z(k) = \sum_{i=1}^{k} [x(i) - \bar{x}]\ k = 1, \ldots, N, \tag{C.1}$$

where $\bar{x}$ is the mean of the time series.

(2) The new time series $z(k)$ is divided into N_s s-length segments without overlap, where $N_s = \mathrm{int}(N/s)$.

(3) Least-squares linear fit is used to determine the linear regression $z_{fit}(k)$ as the local trend for each segment. The mean square error corresponding to s is got by

$$F_s^2(v) = \frac{1}{s}\sum_{i=1}^{s}\{z[(v-1)s+i] - z_{fit}(i)\}^2 \; v = 1, \ldots, N_s, \tag{C.2}$$

(4) The average root-mean-square fluctuation is defined as

$$F(s) = \sqrt{\frac{1}{N_s}\sum_{v=1}^{N_s} F_s^2(v)} \tag{C.3}$$

(5) Repeat steps (1)—(4) for different scales s, $F(s)$ is proportional to a power of s,

$$F(s) \propto s^{\alpha} \tag{C.4}$$

where α is called the scaling exponent which can be estimated by a linear fit. The slope of the log-log plot of $F(s)$ in function of s is the scaling exponent α.

Appendix D. Extreme learning machine

The procedure of the ELM can be described as follows (Huang et al., 2004; Liang et al., 2006):

Suppose that there are N training samples $\{(\mathbf{x}_k, \mathbf{t}_k)\}_{k=1}^{N}$, hidden layer activation function $g(x)$, and hidden neuron number M, Step 1: Input weights $\mathbf{w}_i$ and hidden layer biases b_i are randomly evaluated.

Step 2: Hidden layer output matrix $\mathbf{H}$ is computed.

Step 3: Output weight $\hat{\boldsymbol{\beta}}$ is computed according to $\hat{\boldsymbol{\beta}} = \mathbf{H}^{+}\mathbf{T}$.

References

Acharya, U.R., Faust, O., Kannathal, N., Chua, T.L., Laxminarayan, S., 2005. Non-linear analysis of EEG signals at various sleep stages. Comput. Methods Programs Biomed. 80, 37—45.

Andrzejak, R., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C., 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, 061907-1—061907-8.

Bartlett, P.L., 1998. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. IEEE Trans. Inf. Theory 44, 525—536.

Bhattacharya, U., Parui, S.K., 1995. Self-adaptive learning rates in backpropagation algorithm improve its function approximation performance. Conf. Proc. IEEE Neural Netw., 2784—2788.

Diambra, L., Bastos, de Figueiredo, J.C., Malta, C.P., 1999. Epileptic activity recognition in EEG recording. Phys. A: Stat. Mech. Appl. 273, 495—505.

Gabor, A.J., Leach, R.R., Dowla, F.U., 1996. Automated seizure detection using a self-organizing neural network. Electroencephalogr. Clin. Neurophysiol. 99, 257—266.

Gandhi, T., Panigrahi, B.K., Bhatia, M., Anand, S., 2010. Expert model for detection of epileptic activity in EEG signature. Expert Syst. Appl. 37, 3513—3520.

Gotman, J., 1982. Automatic recognition of epileptic seizures in the EEG. Electroencephalogr. Clin. Neurophysiol. 54, 530—540.

Gotman, J., 1990. Automatic seizure detection: improvements and evaluation. Electroencephalogr. Clin. Neurophysiol. 76, 317—324.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. Conf. Proc. IEEE Neural Netw., 985—990.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing 70, 489—501.

Jiang, Z., Ning, Y., An, B., Li, A., Feng, H., 2005. Detecting mental EEG properties using detrended fluctuation analysis. Conf. Proc. IEEE Eng. Med. Biol. Soc., 2017—2020.

Kannathal, N., Acharya, U.R., Lim, C.M., Sadasivan, P.K., 2005a. Characterization of EEG-A comparative study. Comput. Methods Programs Biomed. 80, 17—23.

Kannathal, N., Choo, M.L., Acharya, U.R., Sadasivan, P.K., 2005b. Entropies for detection of epilepsy in EEG. Comput. Methods Programs Biomed. 80, 187—194.

Ki, C., Scully, C., Sheng, L., 2009. Approximate entropy for all signals. IEEE Eng. Med. Biol. Mag. 28, 18—23.

Lee, J.M., Kim, D.J., Kim, I.Y., Park, K.S., Kim, S.I., 2002. Detrended fluctuation analysis of EEG in sleep apnea using MIT/BIH polysomnography data. Comput. Biol. Med. 32, 37—47.

Lee, J.S., Yang, B.H., Lee, J.H., Choi, J.H., Choi, I.G., Kim, S.B., 2007. Detrended fluctuation analysis of resting EEG in depressed outpatients and healthy controls. Clin. Neurophysiol. 118, 2489—2496.

Lehnertz, K., Mormann, F., Kreuz, T., Andrzejak, R.G., Rieke, C., David, P., Elger, C.E., 2003. Seizure prediction by nonlinear EEG analysis. IEEE Eng. Med. Biol. Mag. 22, 57—63.

Leistedt, S., Dumont, M., Lanquart, J.P., Jurysta, F., Linkowski, P., 2007. Characterization of the sleep EEG in acutely depressed men using detrended fluctuation analysis. Clin. Neurophysiol. 118, 940—950.

Liang, N.Y., Saratchandran, P., Huang, G.B., Sundararajan, N., 2006. Classification of mental tasks from EEG signals using extreme learning machine. Int. J. Neural Syst. 16, 29—38.

Misra, U.K., Kalita, J., 2005. Clinical Electroencephalography, first ed. Elsevier, a division of Reed Elsevier India Private Limited, Noida, pp. 130.

Niedermeyer, E., da Silva F., L., 2005. Electroencephalography: Basic Principles, Clinical Applications and Related Fields, fifth ed. Lippincott Williams & Wilkins, Philadelphia, pp. 526.

Pan, Y., Ge, S.S., Al, Mamun, A., Tang, F.R., 2008. Detection of seizures in EEG signal using weighted locally linear embedding and SVM classifier. Conf. Proc. IEEE Cybern. Intell. Syst., 358—363.

Pardey, J., Roberts, S., Tarassenko, L., 1996. A review of parametric modelling techniques for EEG analysis. Med. Eng. Phys. 18, 2—11.

Peng, C.K., Havlin, S., Stanley, H.E., Goldberger, A.L., 1995. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 5, 82—87.

Pincus, S., 1991. Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. U.S.A. 88, 2297—2301.

Radhakrishnan, N., Gangadhar, B.N., 1998. Estimating regularity in epileptic seizure time-series data. IEEE Eng. Med. Biol. Mag. 17, 89—94.

Sanei, S., Chambers, J.A., 2007. EEG Signal Processing. John Wiley & Sons Ltd., Chichester, pp. 161.

Song, Y., Liò, P., 2010. A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine. J. Biomed. Sci. Eng. 3, 556—567.

Srinivasan, V., Eswaran, C., Sriraam, N., 2007. Approximate entropy-based epileptic EEG detection using artificial neural networks. IEEE Trans. Inf. Technol. Biomed. 11, 288—295.

Swiderski, B., Osowski, S., Cichocki, A., Rysz, A., 2008. Single-class SVM classifier for localization of epileptic focus on the basis of EEG. Conf. Proc. IEEE Neural Netw., 189—194.

Übeyli, E.D., 2009. Combined neural network model employing wavelet coefficients for EEG signals classification. Digit. Signal Process. 19, 297—308.

Wang, C., Zou, J., Zhang, J., Zhang, Z., Zhang, C., 2009. Classifying detection of epileptic EEG based on approximate entropy in wavelet domain. Conf. Proc. IEEE Med. Eng. Inform., 1—5.

Wang, Z., Guo, D., Li, X., Fei, Y., 2006. Estimating Hurst exponent with wavelet packet. Conf. Proc. IEEE Comput. Aided Ind. Des. Concep. Des., 1—4.

Weng, W., Khorasani, K., 1996. An adaptive structure neural networks with application to EEG automatic seizure detection. Neural Netw. 9, 1223—1240.