Neurocomputing 175 (2016) 383–391

Contents lists available at ScienceDirect

Neurocomputing

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

Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine ☆

Jiang-Ling Song, Wenfeng Hu, Rui Zhang *

Medical Big Data Research Center, Northwest University, Xi'an 710069, China

ARTICLE INFO

Article history:
Received 16 August 2015
Received in revised form
10 October 2015
Accepted 16 October 2015
Communicated by G.-B. Huang
Available online 30 October 2015

Keywords:
Epileptic seizure detection
Mahalanobis distance
Discrete wavelet transformation (DWT)
Sample entropy
Fusion feature
Extreme learning machine (ELM)

ABSTRACT

Automated seizure detection using EEG has gained increasing attraction in recent years and appeared more and more helpful in both diagnosis and treatment. How to design an appropriate feature extraction method and how to select an efficient classifier are recognized to be crucial in the successful realization. This paper first proposes a new Mahalanobis-similarity-based feature extraction method on the basis of the Mahalanobis distance and discrete wavelet transformation (DWT). Then in order to further improve the performance, this paper designs a fusion feature (MS-SE-FF) in the feature-fusion level, where the Mahalanobis-similarity-based feature characterizing the similarity between signals and the sample-entropy-based feature characterizing the complexity of signals are combined together. Finally, an automated seizure detection method FF-ELM-SD has been built, which is integrated between the novel fusion feature MS-SE-FF and extreme learning machine (ELM). Experimental results demonstrate that the proposed method FF-ELM-SD does a good job in the epileptic seizure detection while preserving the efficiency and simplicity.

1. Introduction

Epilepsy is one of the most common neurological disorder that affects nearly 0.9% of the world's population [5]. It is characterized by transient and recurrent epileptic seizures, which are caused by hyper-synchronous discharges of an excessive group of cells in the brain, and typically manifest in muscle stiffness, staring and impaired consciousness etc. [31]. Electroencephalography (EEG) is the recording of electrical activities of the brain that carries a large amount of information about physiology and pathology. It has been verified that EEG is a valuable tool for the diagnosis and treatment follow-up in epilepsy patients during last 70 years [12]. However, the traditional visual inspection of epileptic EEGs by a trained neurologist is not only a time-consuming and subjective process, but also very challenging due to inevitable factors resulting in the presence of myogenic artifacts [16]. Hence, there has been an increasing interest in the study of the automated seizure detection using EEGs in recent years. In order to realize it successfully, how to design an appropriate feature extraction method, as well as how to select an efficient classifier, are then two important topics in the study.

Various feature extraction methods have been developed from different points of view, which can be summarized as the time domain, frequency domain, time-frequency domain, graph domain as well as nonlinear methods and so on. In the case of time-domain analysis, Chandaka et al. [4] has proposed a feature based on the cross-correlation between each present EEG segment and the given reference (inter-ictal) EEG segment. In [3], linear prediction error energy (LPEE) has been designed in terms of the error, which results from the approximation of EEG signals by applying the linear prediction filter. On the other hand, frequency-domain analysis mainly focuses on calculating the energy of EEG signals in the field of frequency. One of the most representative features refers to as power spectral density (PSD), which can be estimated by non-parametric methods (e.g., fast Fourier transform (FFT) [27]), or parametric methods (e.g., eigenvector (EV) and multiple signal classification (MUSIC) [14]). However, the above mentioned methodologies only reflect single characteristic of the signals (say, time or frequency), and it has to be assumed that each of the EEG segment satisfies being stationary or pseudo-stationary [21]. Therefore, the time-frequency analysis has been investigated, whose key idea is to apply different decomposition methods, including the wavelet transform (WT) [8,19,11], the empirical mode decomposition (EMD) [12,20], and the Hilbert-Huang transform [7], to re-express the EEG signals. Furthermore, graph-domain-based methods have been proposed through first representing EEG signals as a corresponding

☆This work was supported by the National Natural Science Foundation of China under Grant 61473223 and the Natural Science Foundation of Shaanxi Province, China under Grant 2014JM1016.

* Corresponding author.
E-mail addresses: sjl@stumail.nwu.edu.cn (J.-L. Song), meideno1@163.com (W. Hu), rzhang@nwu.edu.cn (R. Zhang).

http://dx.doi.org/10.1016/j.neucom.2015.10.070
0925-2312/

graph, and then extracting certain graphic characteristics as the final features [6,35].

Since recent studies show that brain is a nonlinear dynamical system by neurologists, different nonlinear methods have been applied to find a governing rule by which the system can be distinguished between a non-seizure state and a seizure state. With the recognition that the system in inter-ictal periods is more complex than that in ictal periods, feature extraction methods based on sample entropy [26], permutation entropy [15], and correlation sum [30] have been proposed respectively. Differently, due to the fact that the anti-persistence of the system in inter-ictal periods is weaker than that in ictal periods, feature extraction methods based on the Hurst exponent and Detrended fluctuation analysis have been designed in [33]. In addition, it is well known that the ictal EEGs present more synchronously than inter-ictal EEGs owing to the essence of epileptic seizure. Therefore, the phase synchronization and the mean phase coherence [13], which could characterize the synchronization of EEG signals, have been employed in the feature extraction. Moreover, the similarity (or dissimilarity) between the reference window and the present window of EEG signals has also been used to extract features. In view of this point, different features have been explored, including the dynamical similarity index and fuzzy similarity [17], as well as Bhattacharyya-based dissimilarity index [16].

With the appropriate extracted features, another main problem is to select an efficient classifier to complete the seizure detection. We summarize the commonly used classifiers in seizure detection in what follows. Artificial neural networks (ANNs), which is an information processing system imitating the biological nervous systems, has been applied to discriminate the EEGs in [8,19,2]. Decision tree (DT), which is established by a group of nodes generated by a sequence of queries along a path from root to leaf, has been employed to be the classifier in [10]. As a prototype-based supervised classification algorithm, learning vector quantization (LVQ), has also been used to differentiate normal and epileptic EEGs [18]. Meanwhile, support vector machine (SVM), which aims to find the optimal separating hyperplane on the basis of the statistical learning theory, is one of the most popular machine learning tools in the field of pattern recognition. It has been successfully used to differentiate ictal EEGs from inter-ictal EEGs in [4,11,15] respectively. As an emergent technology, extreme learning machine (ELM) studies a much wider type of generalized single-hidden layer feedforward networks (GSLFNs) whose hidden layer needs not be tuned. ELM has been attracting the attentions from more and more researchers due to its successful applications in regression and classification problems, including seizure detection in [26,33].

In this paper, we will first design a novel Mahalanobis-similarity-based feature extraction method, which can be evaluated in accordance with the following three steps. At the first step, with the consideration that certain change in EEG signals may be amplified when signals on each sub-band are analyzed separately [17], the whole EEG signal is decomposed into those on different sub-bands by the discrete wavelet transform (DWT). At the second step, the Mahalanobis similarity between the reference state (represented by the trajectory matrix of inter-ictal EEG signals) and the present state (represented by the trajectory matrix of present EEG signals) on each sub-band is calculated by Mahalanobis distance respectively. At the third step, the available sub-bands are selected and the Mahalanobis similarity vectors on such sub-bands are put into together to construct the Mahalanobis-similarity-based feature. We then propose a fusion feature (MS-SE-FF) combining the proposed Mahalanobis-similarity-based feature with the sample entropy, which characterize the complexity of the system, so as to further improve the detection capability. Finally, extreme learning machine (ELM) is employed to differentiate ictal EEGs from inter-ictal EEGs to complete the seizure detection.

The rest of this paper is organized as follows. Section 2 systematically describes the integrated seizure detection method, including the Mahalanobis-similarity-based feature, the sample-entropy-based feature, the fusion feature, as well as ELM. The data set and experimental results are presented in Section 3, followed by some concluding remarks in the last section.

2. Methods

This section will present the integrated seizure detection method FF-ELM-SD proposed in this paper, including a new Mahalanobis-similarity-based feature (MS-F), the sample-entropy-based feature (SE-F), the fusion feature MS-SE-FF combining MS-F and SE-F together, as well as extreme learning machine (ELM).

2.1. Mahalanobis-similarity-based feature

In this subsection, we will first review the definitions of trajectory matrix, Mahalanobis distance and the discrete wavelet transform (DWT) separately. Then the novel Mahalanobis-similarity-based feature will be introduced.

2.1.1. Trajectory matrix

Given a time series $\{s_1, s_2, \ldots, s_N\}$, we first reconstruct its trajectories according to the time delay method proposed by Takens [28] as

$$\mathbf{p}_j = (s_j, s_{j+\tau}, \ldots, s_{j+(l-1)\tau})^T, j = 1, 2, \ldots, N-(l-1)\tau$$

on the phase space. Here, l is the embedding dimension, τ is the time delay, and T represents the transpose of the matrix.

Then the trajectory matrix $\mathbf{A}$ of time series can be defined by

$$\mathbf{A} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{N-(l-1)\tau})_{l\times(N-(l-1)\tau)}. \quad (1)$$

Obviously, the trajectory matrix $\mathbf{A}$ contains the complete recording of patterns that have occurred within a window of size l [17].

2.1.2. Mahalanobis distance

In 1936, P.C. Mahalanobis defined a new distance from the statistical point of view, which is referred to as the Mahalanobis distance (also called the classed quadratic distance). It takes into account the standard deviation of random variables and their mutual correlations.

Given two random variables ξ and η, the Mahalanobis distance between them is defined as [21]

$$d(\xi, \eta) = \sqrt{(\xi-\eta)^T \mathbf{B}^{-1} (\xi-\eta)}$$

where

$$\mathbf{B} = \begin{pmatrix} \sigma_\xi^2 & \sigma_{\xi\eta} \\ \sigma_{\eta\xi} & \sigma_\eta^2 \end{pmatrix}$$

is the variance-covariance matrix of ξ and η. Here, $\sigma_\xi^2, \sigma_\eta^2$ denote the variances of ξ and η respectively, and $\sigma_{\xi\eta}$ $(=\sigma_{\eta\xi})$ denotes the covariance of ξ and η.

Furthermore, Mahalanobis distance can measure the separation of two random groups. Suppose that there are two groups $\mathcal{X}$ and $\mathcal{Y}$ with same number of members. Let $\overline{\mathbf{x}}$ and $\overline{\mathbf{y}}$ denoting the mean vectors of $\mathcal{X}$ and $\mathcal{Y}$, then the Mahalanobis distance of these two groups is given by [29]

$$d_M(\mathcal{X}, \mathcal{Y}) = \sqrt{(\overline{\mathbf{x}}-\overline{\mathbf{y}})^T \mathbf{C}^{-1} (\overline{\mathbf{x}}-\overline{\mathbf{y}})}. \quad (2)$$

Here, **C** denotes the covariance matrix of groups $\mathcal{X}$ and $\mathcal{Y}$. Note that the numbers of random variables in two groups must be same, but the dimensions of them are not required to be same.

Remark 1. Mahalanobis distance is a generalization of Euclidean distance. In the case where the variables are uncorrelated and their standard deviations are equal, Mahalanobis distance is degenerated to be Euclidean distance [21].

2.1.3. Discrete wavelet transform (DWT)

Niknazar et al. [17] has revealed that a tiny modification in EEG signals may be amplified when signals on each sub-band are analyzed separately. With this recognition, different from the traditional methods to calculate the similarity between signals in the whole frequency field directly, we first decompose the raw EEG signal into different sub-signals according to discrete wavelet transform (DWT). Hence, the difference between ictal EEGs and inter-ictal EEGs will be much easier to be captured, which may contribute to the subsequent feature extraction.

DWT is a powerful tool in time-frequency analysis. It can provide a more flexible representation of signals in time-frequency domain. One of the most attractive features of DWT is that it uses long-time windows at low frequencies and short-time windows at high frequencies, leading to good time-frequency localization [13]. Given a prototype wavelet $\psi(t)$ (called mother wavelet), it can be scaled and shifted by $a_j = 2^j$ and $b_{j,k} = 2^j k (j, k \in Z)$. Here, $a_j, b_{j,k}$ are called the scaled parameter and the translation parameter respectively. Then the DWT of signal $x(t)$ can be formulated by

$$DWT(j,k) = 2^{-j/2} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-2^j k}{2^j}\right) dt$$

where ψ^* denotes the complex conjugation of ψ.

In 1989, Mallat developed an efficient way to implement DWT by passing a signal through a set of low-pass and high-pass filters [25], where the cut-off frequency is defined to be one-forth of the sampling rate. Meanwhile, the corresponding outputs of low pass and high pass filters are named as the approximate coefficient and the detail coefficient respectively. In this way, the signal is then decomposed into a family of signals on different sub-bands through a series of filters. At each step of the decomposition, the frequency resolution gets doubled by filtering and time resolution gets halved by down-sampling [11].

It should be noted that the selection of wavelet function and the number of decomposition level are rather important in DWT. In general, the decomposition level depends on the range of the dominant frequencies, which are determined by signals we concern most in applications. In this paper, the Daubechies wavelet with order 4 is selected as the mother wavelet function, which is recommended by other work [8,11,34]. The number of decomposition level is chosen to be 5 since the range of frequencies in EEG is generally set to be [0.5, 30] Hz both in clinical and physiological interests [2].

2.1.4. Mahalanobis-similarity-based feature

With the proceeding preliminaries, we give the Mahalanobis-similarity-based feature (MS-F) extraction method in detail here.

Let $\mathbf{S} = S_1, S_2, \ldots, S_\Gamma$ be the set of all EEG signals where Γ denotes the total number of EEG segments.

For EEG segment S_i, we first decompose it by DWT into K sub-band signals, denoted by $S_i(D_k)$ where D_k is the kth sub-band $(k = 1, \ldots, K)$.

Then we collect EEG segments in the kth sub-band together as

$$\mathbf{S}(D_k) = \{S_i(D_k) : i = 1, 2, \ldots, \Gamma\}.$$

In $\mathbf{S}(D_k)$, we randomly select one inter-ictal EEG segment as the reference signal, denoted by $S^r(D_k)$, and the rest EEG segments are all considered to be the present ones.

Next, for each $S_i(D_k)$, we calculate its corresponding trajectory matrix $A_i(D_k)$ according to Eq. (1). We name the trajectory matrices of the reference and present signals as the reference state and the present state respectively in this paper.

Furthermore, the Mahalanobis distance between the reference state $A^r(D_k)$ and $A_i(D_k)$ (here including each present state and the reference state itself) can be computed in accordance with Eq. (2) as follows:

$$d_i^r(k) = d_M(A^r(D_k), A_i(D_k)), \quad i = 1, 2, \ldots, \Gamma.$$

Then the vector

$$F_{MS}(k) \triangleq (d_1^r(k), d_2^r(k), \ldots, d_\Gamma^r(k))$$

is defined as the Mahalanobis similarity in terms of the reference state in the kth sub-band. Note that the less the distance is, the more the similarity will be.

Finally, for all the sub-bands, we define the original Mahalanobis-similarity-based feature as

$$\tilde{F}_{MS} = (F_{MS}(1)^T, F_{MS}(2)^T, \ldots, F_{MS}(K)^T)^T.$$

However in practice, the frequency range of epileptic seizures is suggested to be 3-30 Hz by neurologists [32]. Hence, for all the sub-bands used in DWT, only those approximately corresponding to the range [3, 30] will be most helpful for the seizure detection. Under this situation, we further define the available decomposed signal as $S_i(D_{k_j})$ where $\{k_j\}_{j=1}^{J^*}$ is a subset of $\{1, 2, \ldots, K\}$.

To sum up, we define the Mahalanobis-similarity-based feature as

$$\mathbf{F}_{MS} = (F_{MS}(k_1)^T, F_{MS}(k_2)^T, \ldots, F_{MS}(k_{J^*})^T)^T.$$

2.2. Sample-entropy-based feature

In this subsection, we introduce the sample-entropy-based feature (SE-F), which can be used to characterize the complexity or irregularity of EEG signals.

As an improvement of approximate entropy (ApEn), Sample entropy (SampEn) was proposed by Richman and Moorman in [22]. It is a new family of statistics that could overcome the bias produced in the calculation of ApEn. Due to the fact that the complexity of EEG signals in ictal periods is much less than that in inter-ictal periods, SampEn has been applied to be the feature in seizure detection in [26]. The calculation of SampEn can be described as follows [26].

SampEn Algorithm: Given N data points $\{x(1), x(2), \ldots, x(N)\}$.

Step 1: Define a sequence of vectors (say, template) as

$$X_i^m = (x(i), x(i+1), \ldots, x(i+m-1))^T, \quad i = 1, 2, \ldots, N-m+1$$

where m is a parameter.

Step 2: Define the distance d_{ij}^m between two vectors X_i^m and X_j^m as

$$d_{ij}^m = d(X_i^m, X_j^m) = \max_{0 \le k \le m-1} |x(i+k) - x(j+k)|.$$

Step 3: Define the similarity degree of vector X_i^m.

According to the Heaviside function, we have

$$\theta(r - d_{ij}^m) = \begin{cases} 0, & d_{ij}^m > r; \\ 1, & d_{ij}^m \le r. \end{cases}$$

where r is the comparison threshold whose value is recommended to be taken as the product of the standard deviation of data and a

scale number ε in [0.1, 0.25] [7].

If $\theta(r-d_{ij}^m)=1$, then we say that vectors X_i^m and X_j^m are similar. Denote by $\omega(i)$ the total number of vectors $X_j^m(1\le j\le N-m, j\ne i)$ which are similar to X_i^m, then we further define

$$\Omega_r^m(i)=\frac{1}{N-m-1}\omega(i), 1\le i\le N-m.$$

as the similarity degree between X_i^m and other vectors in terms of r.

Step 4: Define the group similarity as

$$\phi^m(r)=\frac{1}{N-m}\sum_{i=1}^{N-m}\Omega_r^m(i).$$

Step 5: Increase m to $m+1$, and compute $\phi^{m+1}(r)$ by repeating steps 1-4.

Step 6. Finally, the sample entropy is formulated by

$$\mathrm{SampEn}(m,r)=-\ln\left(\frac{\phi^{m+1}(r)}{\phi^m(r)}\right).$$

Remark 2. For vectors X_i^m and X_j^m, we can further obtain the corresponding $(m+1)$-dimensional vectors $X_i^{(m+1)}=((X_i^m)^T, x(i+m))^T$ and $X_j^{(m+1)}=((X_j^m)^T, x(j+m))^T$. In this case, SampEn can be explained as the negative logarithm of the sum of conditional probabilities. Each of them presents the probability of $X_i^{(m+1)}$ and $X_j^{(m+1)}$ remaining similar under the condition that $X_i^{(m)}$ and $X_j^{(m)}$ are similar. Note that parameters m and r must be specified before calculating the sample entropy.

In the end, given the set of EEG signals $\mathbf{S}=S_1,S_2,\ldots,S_\Gamma$, the sample-entropy-based feature can be expressed by

$$\mathbf{F}_{SE}=(F_{SE}^1,F_{SE}^2,\ldots,F_{SE}^\Gamma)_{1\times\Gamma}$$

where

$$F_{SE}^i=\mathrm{SampEn}_i(m,r).$$

Here, $\mathrm{SampEn}_i(m,r)$ is the sample entropy of the ith EEG segment in $\mathbf{S}$ computed by the above algorithm.

2.3. The fusion feature combining the Mahalanobis-similarity-based feature and sample-entropy-based feature (MS-SE-FF)

This subsection will present the novel fusion feature in the feature-fusion level, whose idea is originated from the following observation. The discriminative ability only with Mahalanobis-similarity-based feature may be limited due to the random selection of the reference signal in the calculation. In order to overcome the limitation, we propose a novel fusion feature in this paper. It combines the Mahalanobis-similarity-based feature (denoted by $\mathbf{F}_{MS}$) and the sample-entropy-based feature (denoted by $\mathbf{F}_{SE}$) together so that the seizure detection capability can be enhanced greatly. We abbreviate the proposed fusion feature as MS-SE-FF. The details of MS-SE-FF can be summarized as the following algorithm.

MS-SE-FF Algorithm: Given EEG signals $\mathbf{S}=S_1,S_2,\ldots,S_\Gamma$. The embedding dimension l, the time delay τ, the optimal number K^* of sub-bands and the mother wavelet $\psi(t)$ in $\mathbf{F}_{MS}$, the parameter m and the comparison threshold r in $\mathbf{F}_{SE}$.

Step 1: Compute the sample-entropy-based feature. For $i=1:\Gamma$,

(a) Compute $\mathrm{SampEn}_i(m,r)$ for the ith EEG segment S_i according to SampEn Algorithm;

(b) Write $F_{SE}^i=\mathrm{SampEn}_i(m,r)$. Then we have

$$\mathbf{F}_{SE}=(F_{SE}^1,F_{SE}^2,\ldots,F_{SE}^\Gamma)_{1\times\Gamma}.$$

Step 2: Compute the Mahalanobis-similarity-base feature.

2.1: Decompose $\mathbf{S}$ into sub-band signals $\{S_i(D_k): i=1,\ldots,\Gamma; k=1,\ldots,K^*\}$ through DWT. Write $\mathbf{S}(D_k)=\{S_i(D_k): i=1,\ldots,\Gamma\}$ to be the set of signals on the same sub-band D_k.

2.2: For $k=1:K^*$,

(a) Randomly select an inter-ictal EEG segment in $\mathbf{S}(D_k)$ as the reference signal, denoted by $S^r(D_k)$; For $i=1:\Gamma$,

(a1) Calculate the trajectory matrices $A_i(D_k)$ of $S_i(D_k)$ according to Eq. (1);

(a2) Calculate the Mahalanobis distance $d_i^r(k)$ between $A^r(D_k)$ and $A_i(D_k)$ according to Eq. (2);

(b)write $F_{MS}(k)=(d_1^r(k),d_2^r(k),\ldots,d_\Gamma^r(k))$.

Then we have

$$\mathbf{F}_{MS}=\begin{pmatrix}F_{MS}(1)\\F_{MS}(2)\\\vdots\\F_{MS}(K^*)\end{pmatrix}_{K^*\times\Gamma.}$$

Step 3: The MS-SE-FF is achieved by

$$\mathbf{F}_{fusion}=\begin{pmatrix}\mathbf{F}_{SE}\\\mathbf{F}_{MS}\end{pmatrix}_{(K^*+1)\times\Gamma.}$$

2.4. Extreme learning machine

Extreme learning machine (ELM), which was first proposed by Huang et al. [9], is a family of fresh learning algorithms which studies a much wider type of "generalized" SLFNs (GSLFNs) where much wider type of formulations are available for the hidden nodes. The essence of ELM is that: different from the common understanding of learning, the hidden layer of SLFNs need not be tuned. This subsection briefly reviews the basic ELM, whose implementation is to apply random computational nodes in the hidden layer, which may be independent of the training data.

Without loss of generality, we introduce a GSLFN with d input nodes, L hidden nodes and one linear output nodes as follows:

$$\boldsymbol{f}_L(\mathbf{x})=\sum_{i=1}^{L}\boldsymbol{\beta}_i G(\mathbf{a}_i,b_i,\mathbf{x}) \tag{3}$$

where $G(\mathbf{a}_i,b_i,\mathbf{x})$ denotes the output of the ith hidden node with hidden-node parameters $(\mathbf{a}_i,b_i)\in R^d\times R$ and $\beta_i\in R$ is the weight connecting the ith hidden node and the output node. It should be noted that the hidden-node output $G(\mathbf{a}_i,b_i,\mathbf{x})$ can be neuron alike, such as additive nodes $G(\mathbf{a}_i,b_i,\mathbf{x})=g(\mathbf{a}_i\cdot\mathbf{x}+b_i)$ where g is the hidden-node activation function, and radius basis function (RBF) nodes $G(\mathbf{a}_i,b_i,\mathbf{x})=\exp(-b_i\|\mathbf{x}-\mathbf{a}_i\|^2)$ [9], or non-neuron alike, such as multiquadratic nodes $G(\mathbf{a}_i,b_i,\mathbf{x})=(\|\mathbf{x}-\mathbf{a}_i\|^2+b_i^2)^{\frac{1}{2}}$ [9].

Given a set of training data $\{(\mathbf{x}_i,t_i)\}_{i=1}^N\subset\mathbf{R}^d\times\mathbf{R}$, if the output of the network is equal to the target, then we have

$$f_L(\mathbf{x}_j)=\sum_{i=1}^{L}\beta_i G(\mathbf{a}_i,b_i,\mathbf{x}_j)=t_j,\quad j=1,\ldots,N$$

which can be equivalently expressed in the matrix form

$$\mathbf{H}\boldsymbol{\beta}=\mathbf{T} \tag{4}$$

where

$$\mathbf{H}=\begin{pmatrix}G(\mathbf{a}_1,b_1,\mathbf{x}_1) & \ldots & G(\mathbf{a}_L,b_L,\mathbf{x}_1)\\G(\mathbf{a}_1,b_1,\mathbf{x}_2) & \ldots & G(\mathbf{a}_L,b_L,\mathbf{x}_2)\\\vdots & \vdots & \vdots\\G(\mathbf{a}_1,b_1,\mathbf{x}_N) & \ldots & G(\mathbf{a}_L,b_L,\mathbf{x}_N)\end{pmatrix}_{N\times L},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{pmatrix}_{L\times 1}, \quad \mathbf{T} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{pmatrix}_{N\times 1}.$$

Here, $\mathbf{H}$ is called the hidden-layer output matrix of the network. In ELM, the hidden-layer output matrix $\mathbf{H}$ is randomly generated. Thus, training SLFNs simply amounts to getting the solution of the linear system (4) with respect to the output weight vector $\boldsymbol{\beta}$. Specifically, we have

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T} \tag{5}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of $\mathbf{H}$.

3. Results and discussion

3.1. Dataset

The EEG database used in this study are taken from the Department of Epileptology, University of Bonn. It consists of single-channel scalp and intracranial EEGs with sampling rate 173.61 Hz and 12-bit A-D resolution [1]. There is no artifact in all EEGs. The whole database is composed of five data sets with each including 100 EEG segments of 23.6 s duration, denoted by Z, O, N, F and S respectively. EEGs in sets Z and O are scalp EEGs and collected from five healthy people in awake and relaxed state with eye open and close respectively. Differently, EEGs in sets N, F and S are intracranial EEGs and recorded from five epilepsy patients. Set N comprises the segments taken from the hippocampal formation of opposite hemisphere of the brain. Sets F and S are constituted of inter-ictal and ictal EEG signals respectively, which are both recorded within the epileptogenic zone [10]. In this paper, we verify the performance of the proposed seizure detection method FF-ELM-SD using the fusion feature MS-SE-FF, as well as extreme learning machine on datasets F and S rather than the whole database, because it is most difficult for seizures to be detected in these two sets [31]. Fig. 1 illustrates two EEG segments in set F and set S.

3.2. Experiments and results

This subsection verifies the performance of the proposed method FF-ELM-SD from three aspects: (1) performance comparison between the proposed fusion feature MS-SE-FF and other two single features MS-F and SE-F with the same classifier ELM; (2) performance comparison between ELM and SVM with the same fusion feature MS-SE-FF; (3) performance comparison between the proposed method FF-ELM-SD with other existing seizure detection methods in [11,15,33,35,24].

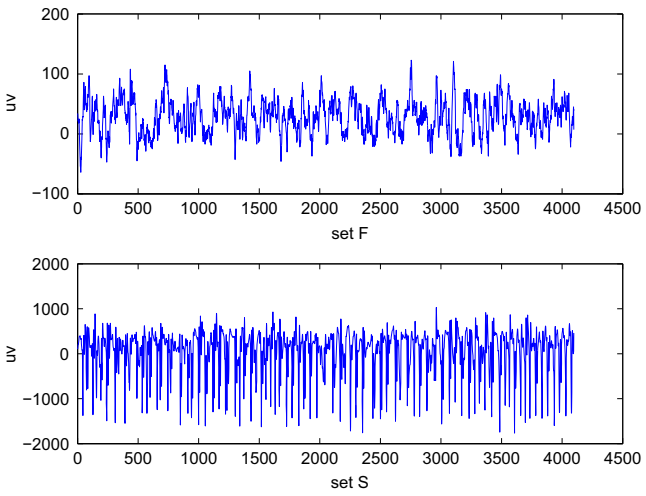

Fig. 1. Sample EEG recordings of set F (inter-ictal) and set S (ictal).

In our experiments, we randomly select ten EEG segments from set F as the reference signals respectively, and all the corresponding rest EEG signals in sets F and S are considered to be the present ones. Then we complete the feature extraction and classification ten times based on such random selection separately. The average of ten results are finally shown in this paper. All the EEG signals are decomposed according to DWT where db4 wavelet is adopted as the mother wavelet. The resulting spectral sub-bands are D_1(43.4–86.8 Hz), D_2(21.7–43.4 Hz), D_3(10.8–21.7 Hz), D_4(5.4–10.8 Hz), D_5(2.7–5.4 Hz), and A_5(0–2.7 Hz) respectively. Fig. 2 illustrates the decomposed signals on six sub-bands $D1$–$D5$ and $A5$ of one segment from set S.

In Fig. 2, $d1$–$d5$ are called the detail coefficients with respect to $D1-D5$, and $a5$ is called the approximate coefficient with respect to $A5$. It can be observed that range $[2.7, 43.4]$ Hz of decomposed components D_2–D_5 approximately corresponds to 3–30 Hz. Therefore, only the decomposed signals on sub-bands D_2–D_5, which are most helpful for the seizure detection, are applied in our experiments.

In the calculation of Mahalanobis-similarity-based feature $\mathbf{F}_{MS}$, the embedding dimension l and the time delay τ are set to be 16 and 6 respectively, which are recommended by literature [17]. In the calculation of sample-entropy-based feature $\mathbf{F}_{SE}$, r is chosen with the scale number ε being 0.2. For determining parameter m, we employ an alternative estimation criterion δ defined in [26] as

$$\delta(\%) = \left|\frac{\overline{m}_a - \overline{m}_b}{\overline{m}_a}\right| \times 100$$

where $\overline{m}_a$ and $\overline{m}_b$ are the average values of sample entropy computed from inter-ictal EEGs and ictal EEGs respectively. In fact, the criterion δ reflects the varying range of the values of sample entropy from inter-ictal state to ictal state, and hence, the larger δ is, the better discrimination ability will be [26]. Table 1 shows the comparison results of $\overline{m}_a$, $\overline{m}_b$ and δ in the cases where $m=2$ and $m=3$. It can be seen from Table 1 that the result (say, $\delta = 0.3103$) with $m=2$ is better than the result (say, $\delta = 0.2667$) with $m=3$. Therefore, parameter m is set to be 2 in this study.

In ELM, all the inputs (features) have been normalized into the range $[-1, 1]$ while the outputs (targets) have been normalized into $[0, 1]$. Neural networks with sigmoid type of additive hidden nodes $G(\mathbf{a}, b, \mathbf{x}) = 1/(1+\exp(-(\mathbf{a}\cdot\mathbf{x}+b)))$ are tested. The hidden node parameters $\mathbf{a}$ and b are randomly chosen from the range $[-1, 1]$ based on a uniform sampling distribution probability. The number of hidden nodes are gradually increased by an interval of 5 in $[5, 30]$ and the nearly optimal number of nodes for ELM are then selected by trial and error. Fig. 3 illustrats the averaged training accuracy of ELM with different numbers of hidden nodes in ten detections. It is obvious that the performance of ELM is almost stable on the range $[15, 30]$. With the consideration of avoiding the over-fitting and keeping the fast learning speed, the number of hidden nodes is selected to be 15 in our experiments. Fifty trials have been conducted with training and testing data sets randomly generated for each trial where the size of training and testing data sets is equal.

SVM is performed using the latest LIBSVM software package 3.20 version [11], with radial basis function (RBF) kernel. In SVM, the regularization parameter C and the kernel width g are selected by the grid search method (e.g., $C = \cdots, 2^{-8}, 2^{-7}, \ldots, 2^{0}, \ldots, 2^{7}, 2^{8}, \cdots; g = \cdots, 2^{-8}, 2^{-7}, \ldots, 2^{0}, \ldots, 2^{7}, 2^{8}, \cdots$) based on ten-fold cross validation. Specifically, we find that the best (C,g) are $(2^5, 2^1)$

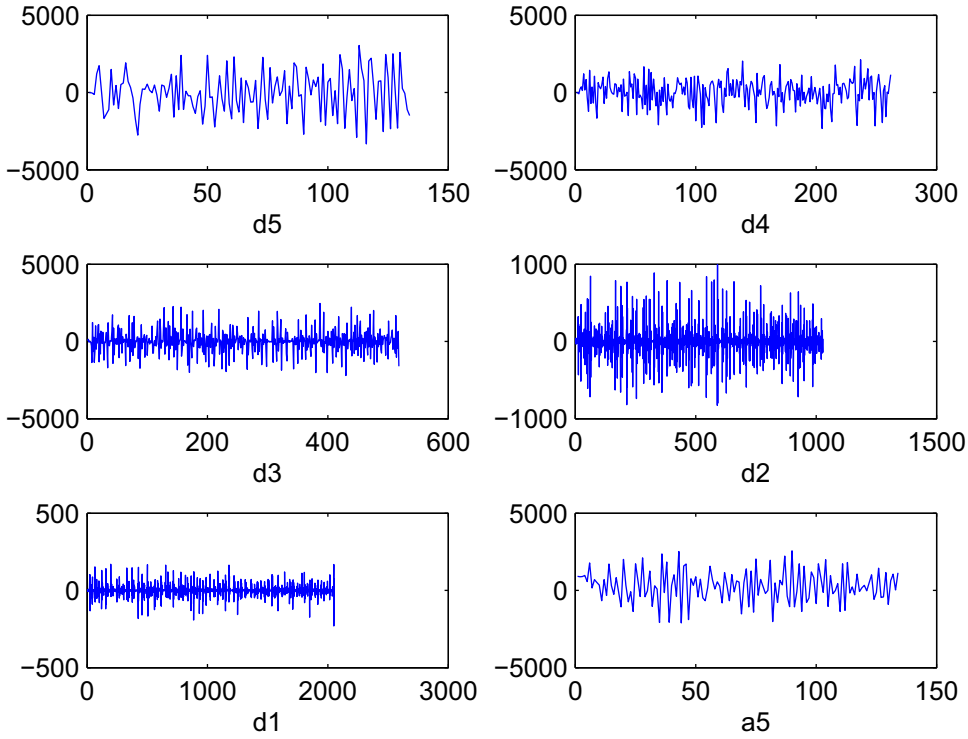

Fig. 2. The decomposed signals of one segment from set S on six sub-bands (d1-d5 are detail coefficients and a5 is the approximate coefficient.)

Table 1
Results of parameter selection for m.

m	$\overline{m}_a$	$\overline{m}_b$	δ
$m=2$	0.4786	0.6271	0.3103
$m=3$	0.4508	0.5711	0.2667

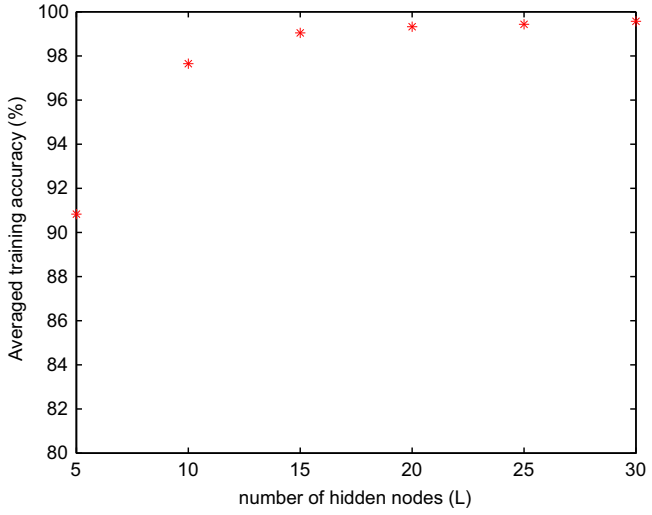

Fig. 3. The averaged training accuracy of ELM with different number of hidden nodes.

according to the averaged best accuracy 98.8% in ten detections as shown in Fig. 4.

All of the specified parameters in our experiments are summarized in Table 2.

Experiments for ELM and SVM are all carried out in MATLAB 8.1.0 environment and the same PC with Intel Core i7 3.60 GHZ CPU and 12 GB RAM. We summarize the experiment results in Tables 4–7. The apparent better results are shown in boldface. Table 3 gives the definition and formula of the evaluation criterion for classification performance, including sensitivity, specificity,

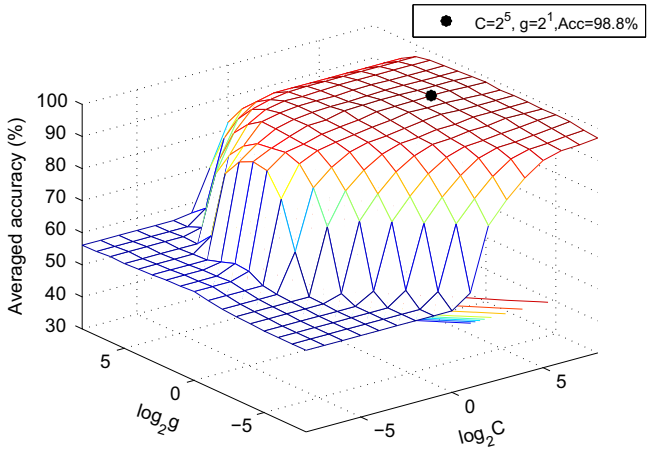

Fig. 4. Parameter selection of SVM using grid search.

Table 2
Specified parameters in the experiments.

Methods	Parameters	Symbols	Specified values
DWT	The optimal number of sub-band	K^*	4
Trajectory matrix	Embedding dimension	l	16
	Time delay	τ	6
Sample entropy	Template parameter	m	2
	Comparison threshold	r	$0.2 \times std$
ELM	Number of hidden nodes	L	15
SVM	Regularization parameter	C	2^5
	Kernel width	g	2

accuracy, positive predictive value (PPV), negative predictive value (NPV) and Matthews correlation coefficient (MCC).

3.2.1. Performance verification of the proposed fusion feature MS-SE-FF

We first compare the classification capability between the proposed fusion feature MS-SE-FF and other two single features MS-F and SE-F with the same ELM. Table 4 shows the experimental results including the averaged sensitivity, specificity, and accuracy. It can be seen from Table 4 that the performance of the proposed Mahalanobis-similarity-based feature MS-F in this paper has been greatly improved comparing with the sample-entropy-based feature SE-F. Moreover, the performance of the proposed fusion feature MS-SE-FF has been further enhanced, with the results of false alarm rate 1.11% and omission factor 3.82%. In detail, the false alarm rate of MS-SE-FF decreases to one half of that obtained by MS-F, while it reduces almost 40 times of that obtained by SE-F. Also, the omission factor of MS-SE-FF drops to nearly one fifth of the other two.

The discriminating capability of different features also can be shown through the Receiver Operating Characteristic Curve (ROC) analysis, which gives us an intuitive view of whole spectrum of sensitivity against 1-specificity [7]. Fig. 5 illustrates the ROC curves for three features MS-F, SE-F and MS-SE-FF with the averaged results of ten detections. The performance of ROC analysis is evaluated by the area under ROC curve (abbreviated by AUC). Obviously, the larger the AUC is, the better the classification performance will be [23]. We can see from Fig. 5 that the AUC achieved by MS-SE-FF is much bigger than those by other two single features MS-F and SE-F.

In order to verify that the range [3, 30] Hz is really the most helpful part for seizure detection, we further compare the classification performance between the fusion feature MS-SE-FF on each sub-band (that is, $D_k(k=1,\ldots,5)$) and MS-SE-FF on sub-bands D_2-D_5. The corresponding results can be shown in Table 5. As observed from Table 5, all the results obtained by MS-SE-FF extracted on D_2-D_5 is better than those extracted on each sub-band separately, which is consistent with our previous analysis.

We finally demonstrate the discriminating capability of our proposed Mahalanobis similarity measure in Fig. 6, which illustrates the Mahalanobis similarity extracted on the 3rd sub-band from sets F and S respectively. Here, the Mahalanobis similarity is computed by taking one of the ten randomly selected EEG segments as the reference signal. In Fig. 6, 100 red circles represent the values of 100 components in $F_{MS}(3)$ on set S (denoted by SD3), and 100 blue stars represent the values of 100 components in F_{MS} (3) on set F (denoted by FD3). Obviously, the proposed Mahalanobis similarity measure can discriminate the ictal EEG from inter-ictal EEG successfully. Meanwhile, we can also observe from Fig. 6 that the values of Mahalanobis similarity on set S are much larger than those on set F in general. It reveals that the inter-ictal EEG signals are more similar with the reference one than ictal EEG signals, which is reasonable due to the reference signal comes from the inter-ictal periods.

3.2.2. Performance comparison between ELM and SVM with the same fusion feature MS-SE-FF

In this part, we compare the performance of ELM with SVM using the same fusion feature MS-SE-FF. The experimental results including the accuracy, standard deviation (Std.) and time (training and parameter selection) are shown in Table 6. As observed from Table 6, the classification accuracy obtained by ELM is better than that obtained with much smaller standard deviation, which means a better and stabler performance of epileptic seizure detection can be achieved by using ELM and MS-SE-FF. Furthermore, the time spent by SVM is much longer than that spent by ELM.

Table 4
Performance comparison among MS-SE-FF, MS-F and SE-F with same ELM.

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)
SE-F+ELM	77.97	51.14	64.00
MS-F+ELM	82.62	96.34	81.38
MS-SE-FF+ELM	**96.18**	**98.89**	**97.53**

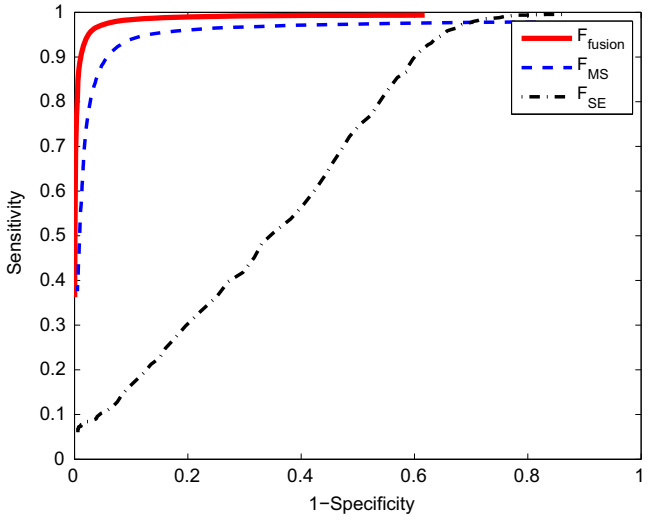

Fig. 5. ROC curves for three features MS-F, SE-F and MS-SE-FF.

Table 3
Collection of characterizing classification performance.

Evaluation criteria	Definition	Formula
Sensitivity	Proportion of the correctly classified ictal EEGs out of the total number of labeled ictal EEGs	$\frac{TP}{TP+FN}\times 100(\%)$
Specificity	Proportion of the correctly classified inter-ictal EEGs out of the total number of labeled inter-ictal EEGs	$\frac{TN}{TN+FP}\times 100(\%)$
Accuracy	Proportion of the correctly classified EEGs out of the total number of EEGs	$\frac{TP+TN}{TP+FN+TN+FP}\times 100(\%)$
Positive predictive value (PPV) [23]	Proportion of the correctly classified ictal EEGs out of the total number of detected ictal EEGs	$\frac{TP}{TP+FP}\times 100(\%)$
Negative predictive value (NPV) [23]	Proportion of the correctly classified inter-ictal EEGs out of the total number of detected inter-ictal EEGs	$\frac{TN}{TN+FN}\times 100(\%)$
Matthews correlation coefficient (MCC) [23]	a correlation coefficient between the observed and predicted binary classifications	$\frac{TP\cdot TN-FN\cdot FP}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}}\times 100(\%)$

Table 5
Performance comparison of the fusion feature MS-SE-FF on different frequency bands.

Frequency bands	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	MCC (%)
D1	88.21	91.34	89.68	91.05	88.73	79.63
D2	93.00	97.13	95.01	96.92	93.37	90.19
D3	95.01	97.81	96.32	97.65	95.16	92.75
D4	90.95	91.65	91.21	91.70	91.12	82.70
D5	87.86	89.63	88.64	90.30	87.98	78.10
D2-D5	**96.18**	**98.89**	**97.53**	**98.64**	**96.26**	**94.81**

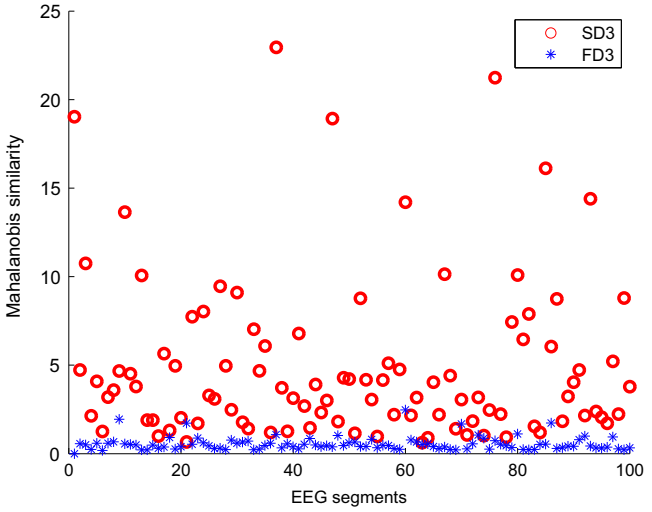

Fig. 6. The Mahalanobis similarity on sub-band D3 of EEG segments in sets F and S.

Table 6
Performance comparison between ELM and SVM with the same feature MS-SE-FF.

Methods	Accuracy (%)	Std.	Time (training+parameter selection)
MS-SE-FF+ELM	**97.53**	**0.0169**	**0.0010+0.06**
MS-SE-FF+SVM	93.67	3.6075	0.0037+6.7571

3.2.3. Performance comparison between the proposed method FF-ELM-SD with other existing methods

Finally, we compare the performance between the proposed method FF-ELM-SD with other existing methods. For convenience of the comparison, we have only listed those studies which used the same datasets (F and S in Bonn database) with ours.

Table 7 presents the classification accuracy (CA) comparison between the proposed method FF-ELM-SD and seven existing methods, which includes Approximate entropy + ELM, Hurst exponent + ELM and DFA + ELM (Yuan et al. [33]); Permutation entropy + SVM (Nicolaou and Georgiou [15]); Degree and Strength of HVG + KNN (Zhu et al. [35]); Clustering + SVM (Siuly and Wen [24]); Fuzzy approximate entropy + SVM (Kumar et al. [11]). As observed from Table 7, it is obvious that the proposed method FF-ELM-SD performs much better than others. Even for the best and the most recent result obtained in [11], the accuracy obtained by our method increases from 95.85% to 97.53%.

4. Conclusion

In this paper, we have first proposed a new Mahalanobis-similarity-based feature extraction method which is accomplished by the following steps: (1) the signals are decomposed by DWT; (2) the Mahalanobis distances between the reference signal and other signals on each sub-band are computed to characterize the similarity between them, which is referred to as the Mahalanobis similarity; (3) select the available sub-bands and put the Mahalanobis similarity vectors on these sub-bands into together to construct the Mahalanobis-similarity-based feature. Then in order to further improve the performance, we proposed a novel fusion feature MS-SE-FF in the feature-fusion level. Specifically, we combine the Mahalanobis-similarity-based feature (MS-F), which characterizes the similarity between signals, with the sample-entropy-based feature (SE-F), which characterizes the complexity (or irregularity) of signals. Finally, an automated seizure detection method FF-ELM-SD has been built, which is integrated between the novel fusion feature MS-SE-FF and extreme learning machine (ELM). The performance of the proposed method FF-ELM-SD has been verified from three aspects: (1) performance comparison between the proposed fusion feature MS-SE-FF and other two single features MS-F and SE-F with the same classifier ELM; (2) performance comparison between ELM and SVM with the same fusion feature MS-SE-FF; (3) performance comparison between the proposed method FF-ELM-SD with other existing methods. All the experimental results demonstrate that the proposed method FF-ELM-SD does a good job in the epileptic seizure detection while preserving the efficiency and simplicity.

Table 7
A comparison between the proposed FF-ELM-SD method with other existing methods.

Authors	Year	Methods	CA (%)
Qi yuan et al. [33]	2012	Approximate entropy + ELM	$88.00 \pm 0.75\%$
	2012	Hurst exponent + ELM	$88.00 \pm 0.5\%$
	2012	DFA+ELM	$82.00 \pm 0.5\%$
Nicolaou et al. [15]	2012	Permutation entropy + SVM	83.13%
Zhu et al. [35]	2014	Degree and strength of HVG+KNN	93%
Siuly et al. [24]	2011	Clustering + SVM	93.94%
Kumar et al. [11]	2014	Fuzzy approximate entropy+SVM	95.85%
This paper		Proposed FF-ELM-SD method (MS-SE-FF+ELM)	**97.53**%

References

[1] U.R. Acharya, F. Molinari, V.S. Subbhuraam, S. Chattopadhyay, Automated diagnosis of epileptic eeg using entropies, Biomed. Signal Process. Control 7 (2012) 401–408.
[2] A. Alkan, E. Koklukaya, A. Subasi, Automated seizure detection in eeg using logistic regression and artificial neural network, J. Neurosci. Methods 148 (2005) 167–176.
[3] S. Altunay, Z. Telatar, O. Erogul, Epileptic eeg detection using the linear prediction error energy, Expert Syst. Appl. 37 (2010) 5661–5665.
[4] S. Chandaka, A. Chatterjee, S. Munshi, Cross-correlation aided support vector machine classifier for classification of eeg signals, Expert Syst. Appl. 36 (2009) 1329–1336.

[5] A.G. Correa, P. LorenaOrosco, Diez, E. Laciar, Automatic detection of epileptic seizures in long-term eeg records, Comput. Biol. Med. 57 (2015) 66–73.
[6] E. Fernandez-Blanco, D. Rivero, J. Rabunal, J. Dorado, A. Pazos, C.R. Munteanu, Automatic seizure detection based on star graph topological indices, J. Neurosci. Methods 209 (2012) 410–419.
[7] K. Fu, J. Qu, Y. Chai, Y. Dong, Classification of seizure based on the time-frequency image of eeg signals using hht and svm, Biomed. Signal Process. Control 13 (2014) 15–22.
[8] L. Guo, D. Rivero, J. Dorado, J.R. Rabunal, A. Pazos, Automatic epileptic seizure detection in eegs based on line length feature and artificial neural network, J. Neurosci. Methods 191 (2010) 101–109.
[9] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[10] Z. Iscan, Z. Dokur, T. Demiralp, Classification of electroencephalogram signals with combined time and frequency features, Expert Syst. Appl. 38 (2011) 10499–10505.
[11] Y. Kumar, M.L. Dewal, R.S. Anand, Epileptic seizure detection using dwt based fuzzy approximate entropy and support vector machine, Neurocomputing 133 (2014) 271–279.
[12] S. Li, W. Zhou, Q. Yuan, S. Geng, D. Cai, Feature extraction and recognition of ictal eeg using emd and svm, Comput. Biol. Med. 43 (2013) 807–816.
[13] F. Mormann, K. Lehnertz, P. David, C.E. Elger, Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients, Physica D 144 (2000) 358.
[14] A.R. Naghsh-Nilchi, M. Aghashahi, Epilepsy seizure detection using eigen-system spectral estimation and multiple layer perceptron neural network, Biomed. Signal Process. Control 5 (2010) 147–157.
[15] N. Nicolaou, J. Georgiou, Detection of epileptic electroencephalogram based on permutation entropy and support vector machines, Expert Syst. Appl. 39 (2012) 202–209.
[16] M. Niknazar, S.R. Mousavi, A new dissimilarity index of eeg signals for epileptic seizure detection, Control Signal Process. (2010).
[17] M. Niknazar, S. Mousavi, M. Shamsollahi, B.V. Vahdat, M. Sayyah, S. Motaghi, A. Dehghani, S. Noorbakhsh, Application of a dissimilarity index of eeg and its sub-bands on prediction of induced epileptic seizures from rat's eeg signals, IRMB 33 (2012) 298–307.
[18] H. Ocak, Optimal classification of epileptic seizures in eeg using wavelet analysis and genetic algorithm, Signal Process. 88 (2008) 1858–1867.
[19] U. Orhan, M. Hekim, M. Ozer, Eeg signals classification using the k-means clustering and a multilayer perceptron neural network model, Expert Syst. Appl. 38 (2011) 13475–13481.
[20] R.B. Pachori, S. Patidar, Epileptic seizure classification in eeg signals using second-order difference plot of intrinsic modefunctions, Comput. Methods Programs Biomed. (2014).
[21] H. Rabal, N. Cap, C. Criado, N. Alamo, Holodiagrams using mahalanobis distance, International Journal for Light and Electron Optics 123 (2012) 1725–1731.
[22] J.S. Richman, J.R. Moorman, Physiological time series analysis using approximate entropy and sample entropy, Am. J. Physiol. Heart Circ. Physiol. 278 (2000) 2039–2049.
[23] R. Sharma, R.B. Pachori, Classification of epileptic seizures in eeg signals based on phase space representation of intrinsic mode functions, Expert Syst. Appl. 42 (2014) 1106–1117.
[24] Siuly, Y. Li, Wen, Clustering technique-based least square support vector machine for eeg signal classification, Comput. Methods Progr. Biomed. 104 (2011) 358–372.
[25] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE. Trans. Pattern Anal. March. Intell. 11 (7) (1989) 674–693.
[26] Y. Song, J. Crowcroft, J. Zhang, Automated epileptic seizure detection in eegs based on optimized sample entropy and extreme learning machine, J. Neurosci. Methods 210 (2012) 132–146.
[27] A. Subasi, E. Ercelebi, A. Alkan, E. Koklukaya, Comparison of subspace-based methods with ar parametric methods in epileptic seizure detection, Comput. Biol. Med. 36 (2006) 195–208.
[28] F. Takens, in: D.A. Rand, L.-S. Young (Eds.), Detecting Strange Attractors in Turbulence, vol. 898, 1981.
[29] K. Teknomo, ⟨http://people.revoledu.com/kardi/tutorial/Similarity/MahalanobisDistance.html⟩.
[30] M. Tito, M. Cabrerizo, M. Ayala, A. Barreto, I. Miller, P. Jayakar, M. Adjouadi, Classification of electroencephalographic seizure recordings into ictal and interictal files using correlation sum, Comput. Biol. Med. 39 (2009) 604–614.
[31] P. van Mierlo, M. Papadopoulou, E. Carrette, P. Boon, S. Vandenberghe, K. Vonck, D. Marinazzo, Functional brain connectivity from eeg in epilepsy: seizure prediction and epileptogenic focus localization, Progr. Neurobiol. 121 (2014) 19–35.
[32] A. Yan, W. Zhou, Q. Yuan, S. Yuan, Q. Wu, X. Zhao, J. Wang, Automatic seizure detection using stockwell transform for long-term eeg, Epilepsy Behav. 45 (2015) 8–14.
[33] Q. Yuan, W. Zhou, S. Li, D. Cai, Epileptic eeg classification based on extreme learning machine and nonlinear features, Epilepsy Res. 96 (2011) 29–38.
[34] S. Yuan, W. Zhou, Q. Yuan, Y. Zhang, Q. Meng, Automated seizure detection using diffusion distance and blda in intracranial eeg, Epilepsy Behav. 31 (2014) 339–345.
[35] G. Zhu, Y. Li, P.P. Wen, Epileptic seizure detection in eegs signals using a fast weighted horizontal visibility algorithm, Comput. Biol. Med. (2014) 64–75.

Jiangling Song received the B.Sc. degrees in financial mathematics and statistics from Northwest University, China, in 2009 and 2013, respectively. She has been with the School of Mathematics, Northwest University, China, since 2009, where she is now a second year graduated school student. Her current research interests include epileptic seizure detection and prediction, and epileptogenic focus localization.

Wenfeng Hu received the B.Sc. in mathematics from Northwest University, China, in 2009 and 2013. She is a second year graduated school student of Northwest University, China. Her current research interests include extreme learning machine, neural networks, machine learning, epileptic seizure detection and prediction.

Rui Zhang received the B.Sc. and M.Sc. degrees in mathematics from Northwest University, China, in 1994 and 1997, respectively, and Ph. D. degree in Applied Mathematics from Xian Jiaotong University, China, in 2011 and Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2012. She has been with the School of Mathematics, Northwest University, China, since 1997, where she is now a Professor. From August 2004 to January 2005, she was a Visiting Scholar in the Department of Mathematics, University of Illinois at Champaign-Urbana, and from August 2013 to August 2014, she was a Visiting Scholar in the Department of Neurology, Harvard Medical School/Massachusetts General Hospital. Her current research interests include extreme learning machine, neural networks, machine learning, epileptic seizure detection and prediction, and epileptogenic focus localization.