

Comparison of classification methods on EEG signals based on wavelet packet decomposition

Yong Zhang · Yuting Zhang · Jianying Wang · Xiaowei Zheng

Received: 17 September 2014 / Accepted: 9 December 2014 / Published online: 24 December 2014
© The Natural Computing Applications Forum 2014

Abstract EEG signals play an important role in both the diagnosis of neurological diseases and understanding the psychophysiological processes. Classification of EEG signals includes feature extraction and feature classification. This paper uses approximate entropy and sample entropy based on wavelet package decomposition as the feature extraction methods and employs support vector machine and extreme learning machine as the classifiers. Experiments are performed in epileptic EEG data and five mental tasks, respectively. Experimental results show that the combination strategy of sample entropy and extreme learning machine has shown great performance, which obtains good classification accuracy and low training time.

Keywords EEG · Wavelet packet · Classification · Approximate entropy · Extreme learning machine

1 Introduction

The electroencephalogram (EEG) signals play an important role in both the diagnosis of neurological diseases and understanding the psychophysiological processes. Although several methods for the brain function analysis such as

magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and positron emission tomography (PET) have been introduced, the EEG signal is still a valuable tool for monitoring the brain activity due to its relatively low cost and being convenient for the patient [1]. Effectively classifying EEG is the foundation for building usable Brain–Computer Interfaces (BCI) and improving the performance of EEG analysis software used in clinical and research settings.

Feature extraction and classification of EEG signals is a challenge for engineers and scientists. In recent years, there has been an increasing interest in the application of machine learning methods for EEG signal. Many methods have been proposed to classify and analyze EEG signal. Among these methods, support vector machine (SVM) and artificial neural network [2, 3] are two widely used classification techniques. Most of EEG signal classifications are built by time–frequency domain-based feature extraction followed by a variety of classification models. So, the classification performance of these methods depends on the feature extraction of the EEG signal.

Wavelet has played an important role in biomedical signal processing for its ability to capture localized spatial-frequency information of EEG signals [4]. In the feature extraction phase, this paper uses approximate entropy (ApEn) and sample entropy (SampEn) based on wavelet packet decomposition (WPD). Obtained feature values are employed as the inputs of a classifier, such as extreme learning machine (ELM) and SVM. Experiments are performed in four different combinations of feature extractions and classifiers. Experimental results in epileptic EEG data and five mental tasks indicate that the combination of sample entropy and ELM has shown great performance.

The remaining of the paper is organized as follows. In Sect. 2, we briefly presented the literature survey that has been performed. In Sect. 3, we described about the

Y. Zhang · Y. Zhang · X. Zheng (✉)
School of Computer and Information Technology, Liaoning Normal University, No. 1, Liushu South Street, Ganjingzi District, Dalian, Liaoning Province 116081, China
e-mail: zhengxw_ln@163.com

Y. Zhang
e-mail: zhyong@lnnu.edu.cn

J. Wang
School of Psychology, Liaoning Normal University,
Dalian 116029, China

materials and methods. In Sect. 4, we compared the classification results of four different combinations in epileptic EEG data and five mental tasks, respectively. And Sect. 5 concluded the paper.

2 Related work

In the feature extraction for EEG signals, Burke et al. [5] extend the usual autoregressive (AR) models for EEG feature extraction. The extension model is an AR with exogenous input (ARX) model for combined filtering and feature extraction. Polat and Günes [6] use fast Fourier transform as feature extraction method. Wavelet transform is also widely used in extracting EEG features [7, 8], in which EEG signals are decomposed by wavelet transform to calculate approximation and detail coefficients. ApEn is a statistical parameter that measures the predictability of the current amplitude values of a physiological signal based on its previous amplitude values. ApEn is used for the first time in the proposed system for the detection of epilepsy using neural networks [2]. A novel feature extraction method based on multi-wavelet transform and ApEn is proposed by Guo et al. [9]. The proposed method uses approximate entropy features derived from multi-wavelet transform and combines with an artificial neural network to classify the EEG signals. Song et al. [10, 11] use sample entropy (SampEn) as a feature extraction method for detecting epileptic seizures. In Ref. [10], SampEn is selected as a feature extraction method to classify the task of three different kinds of EEG signals and detecting epileptic seizures. Based on Ref. [10], the authors propose an optimized sample entropy algorithm to identify the EEG signals in [11].

Artificial neural networks (ANNs) have been widely applied to classify EEG signals [2, 9]. Kumar et al. [7] use discrete wavelet transform (DWT)-based ApEn and ANN to detect epileptic seizures. SVM is also one of the classification methods for EEG signals [3, 4]. Zhang et al. [12] present an improved method to calculate phase locking value (PLV) based on Hilbert–Huang transform, and the PLVs are used as features for a least squares support vector machine (LS-SVM) to recognize normal and hypoxia EEG. Bajaj et al. [13] present a new method for classification of EEG signals using empirical mode decomposition (EMD) method. The proposed method for classification of EEG signals is based on the bandwidth features and employs LS-SVM for classifying seizure and non-seizure EEG signals. Wu et al. [14] propose a multiscale cross-approximate entropy method to quantify the complex fluctuation between R–R intervals series and pulse transit time series. Ahangi et al. [15] use a multiple classifier system for classification of EEG signals. The proposed method uses DWT decomposition, and a variety of classifier

combination methods along with genetic algorithm feature selection is evaluated.

3 Materials and methods

In this section, we give a description of proposed method for EEG signal classification problem, present feature extraction based on the WPD, and give a briefly review of two classifiers, ELM and SVM.

Figure 1 shows the block diagram of the proposed method based on WPD. In our method, we first use the WPD to decompose EEG data into sub-band signals. Then, ApEn values or SampEn values are calculated by approximation and detail coefficients. Obtained feature vectors are used as the inputs of classifiers, such as SVM and ELM. Last, we evaluate the classification accuracy.

3.1 Feature extraction based on wavelet packet decomposition

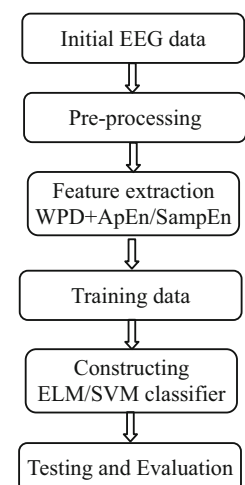
Feature extraction of EEG signals includes two phases. In the first phase, EEG signals are decomposed by the WPD to calculate approximation and detail coefficients. In the second phase, approximate entropy values or sample entropy values of the approximation and detail coefficients are calculated, which form feature vectors. These feature vectors are used as the inputs of a classifier.

In this subsection, we first simply analyze the characteristic of the WPD, and then, give a briefly review of approximate entropy and sample entropy.

3.1.1 Wavelet packet decomposition

Feature extraction is very important to EEG signals analysis and processing. WPD is a wavelet transform where

Fig. 1 Block diagram of the proposed method



the discrete-time signal is passed through more filters than the DWT, which can provide a multi-level time–frequency decomposition of signals [16]. In the DWT, each level is calculated by passing only the previous wavelet approximation coefficients, which are decomposed to yield approximation and detail information at a higher level. However, in the WPD, both the detail and approximation coefficients are decomposed to create the full binary tree.

For n levels of decomposition, the WPD produces 2^n different sets of coefficients. Figure 2 gives the three-level WPD. In Fig. 2, the symbol “s” represents an original signal. Symbols “A” and “D” are the approximation and detail coefficients, respectively. It can be seen in Fig. 2 that in the wavelet packet analysis, both the approximation and details at a certain level are further decomposed into the next level, which means the wavelet packet analysis can provide a more precise frequency resolution than the wavelet analysis.

There are many wavelets that can analyze the EEG signal and extract the feature vector. The most important wavelets families are Haar, Daubechies, Symlets, Coiflets, and biorthogonal [16]. It is very important to select the suitable wavelet and the number of decomposition levels in the analysis of EEG signals. Among the various wavelet bases, the Daubechies family of wavelets is known for its orthogonality property and efficient filter implementation.

3.1.2 Approximate entropy and sample entropy

Approximate entropy was first proposed by Pincus [17]. It is a formulated statistical parameter to quantify the regularity of a time series data of physiological signals. EEG signals are time series data essentially, in which raw data values are from measurement equally spaced in time.

Let m be an integer which represents the length of a vector, and r be a similar limit. Given a time series $u(1), u(2), \dots, u(N)$, where N is the total number of data points.

1. Use the time series $u(1), u(2), \dots, u(N)$ to form a sequence of m -length vectors $x(1), x(2), \dots, x(N - m + 1)$, where $x(i) = [u(i), u(i + 1), \dots, u(i + m - 1)]$, $i = 1, \dots, N - m + 1$.

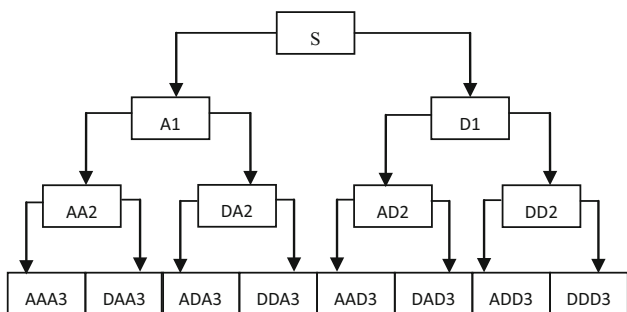


Fig. 2 Processing of 3-level wavelet packet decomposition

2. Define the distance between each vector $x(i)$ and $x(j)$, $d[x(i), x(j)]$, as the maximum difference between their scalar components:

$$d[x(i), x(j)] = \max_{k=1,2,\dots,m} (|u(i + k - 1) - u(j + k - 1)|). \tag{1}$$

3. For a given $x(i)$, count the number of j , such that the distance between $x(i)$ and $x(j)$ is less than or equal to r . Then, for each i , compute

$$C_i^m(r) = \frac{\sum_j \{1 | d[x(i), x(j)] \leq r\}}{N - m + 1} \tag{2}$$

4. Define $\Phi^m(r)$ and $\Phi^{m+1}(r)$ as

$$\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r), \tag{3}$$

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

5. ApEn(m, r, N) is computed by $\Phi^m(r)$ and $\Phi^{m+1}(r)$ as follows:

$$\begin{aligned} \text{ApEn}(m, r, N) &= \Phi^m(r) - \Phi^{m+1}(r) \\ &= \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln \frac{C_i^m(r)}{C_i^{m+1}(r)}. \end{aligned} \tag{5}$$

Two input parameters of approximate entropy must be specified. Usually, parameter m is set to be 1, 2, or 3. Parameter r is set to be some percentage of the standard deviation of the amplitude of time series.

ApEn allows the self-matching ($i = j$ in Eq. 2) to avoid logarithm of zero, thus inducing the bias in estimates. Sample entropy [18] was developed to reduce the bias caused by the self-matching. Mathematically, the computation of SampEn has the same steps 1–3 explained for ApEn. Note that the self matched templates are not computed in step 3. Then, define

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} C_i^m(r), \tag{6}$$

$$B^{m+1}(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} C_i^{m+1}(r), \tag{7}$$

Finally, sample entropy can be defined as

$$\text{SampEn}(m, r, N) = \lim_{N \rightarrow \infty} \left\{ -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right] \right\}. \tag{8}$$

And it is estimated by the statistic

$$\text{SampEn}(m, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right]. \tag{9}$$

The parameters m and r are critical in determining the outcome of SampEn. Usually, SampEn is estimated using the widely established parameter values where m is set 1 or 2 and r is set between 0.11 and 0.25 times the standard deviation of the original time series.

3.2 Classification methods

3.2.1 Extreme learning machine

The ELM [19, 20] is a newly proposed algorithm for generalized single-hidden layer feedforward neural networks (SLFNs) with randomly chosen hidden nodes and analytically determined output weights.

Given N distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ is the i th sample with n -dimensional features and $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ represents the actual labels of \mathbf{x}_i . Standard SLFNs with M hidden neurons can be defined as:

$$\sum_{j=1}^M \beta_j g(w_j \cdot x_i + b_j) = o_i, \quad i = 1, 2, \dots, N, \tag{10}$$

where $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$ represents the weight vector connecting the j th hidden neuron and the input nodes, β_j is the output weight connecting the j th hidden neuron and the output nodes, b_j is the bias of the j th hidden neuron, o_i is the i th output neuron, and $g(\bullet)$ denotes the activation function.

If there exist β_j , w_j , and b_j , an SLFN with M hidden neurons can approximate these N samples with zero error, we have

$$\sum_{j=1}^M \beta_j g(w_j \cdot \mathbf{x}_i + b_j) = \mathbf{t}_i, \quad i = 1, 2, \dots, N \tag{11}$$

Equation (11) can be rewritten in a matrix form:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{12}$$

where

$$\mathbf{H} = \begin{bmatrix} g(w_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(w_M \cdot \mathbf{x}_1 + b_M) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(w_M \cdot \mathbf{x}_N + b_M) \end{bmatrix}_{N \times M}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times m}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_M^T \end{bmatrix}_{N \times m}$$

\mathbf{H} is called the hidden layer output matrix of the network. In many real applications, the training error cannot be made exactly zero. To obtain the smallest nonzero training error, a simple representation of the solution of Eq. (13) is given explicitly by Huang et al. [19] as

$$\tilde{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T}, \tag{13}$$

where \mathbf{H}^+ is the Moore–Penrose generalized inverse of \mathbf{H} . The general procedure of ELM is as follows:

- Step 1: Assess the input weights w_j and hidden layer biases b_j randomly;
- Step 2: Compute the hidden layer output matrix \mathbf{H} ;
- Step 3: Obtain the output weight $\tilde{\boldsymbol{\beta}}$ according to Eq. (13);

In addition, SVM is applied for comparison with ELM. The following is the simple description of SVM.

3.2.2 SVM classifier

SVM developed by Vapnik [21] has received considerable attention. Based on statistics learning theory, it has shown promising empirical results in many fields. SVM can prevent over-fitting based on the structural risk minimization principle.

Given a training data set $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, where $x_i \in \mathbb{R}^n$ denotes a sample with n -dimension and $y_i \in \{+1, -1\}$ is the class of sample x_i . An SVM training algorithm constructs an optimal separating hyperplane in a high-dimensional space to classify new samples. In order to find the optimal hyperplane, the training samples are first transformed into a higher dimensional feature space by a mapping function ϕ . Then, a possible separating hyperplane can be represented by [21]

$$w \cdot \phi(x) + b = 0 \tag{14}$$

The above classification problem can be formalized as the following optimization problem:

$$\min_{w, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \tag{15}$$

subject to

$$y_i(w \cdot \phi(x_i) + b) + \xi_i - 1 \geq 0 \quad \xi_i \geq 0, \quad i = 1, \dots, l \tag{16}$$

where parameter C is a user-specified positive parameter representing the penalty of misclassifying the training samples. It should be noted that the larger the C , the more the error is penalized. The parameter C can be chosen based on the model’s performance on the validation set. The slack variables $\xi_i > 0$ provide an estimate of the error of the decision boundary on the training samples x_i , which

are mapped into a higher dimensional space by the function ϕ .

This constrained optimization problem can be solved by constructing a Lagrangian representation and transforming it into the dual problem. When we obtain the optimal solution (w^*, b^*) of the decision boundary, the SVM decision function can be given by

$$f(x) = \text{sign} \left(\sum_{i=1}^l y_i K(x, x_i) + b \right) \tag{17}$$

where l is the number of support vectors and $K(x_i, x_j)$ is the kernel function as follows:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j). \tag{18}$$

Thus, the kernel function is the key that determines the performance of the SVM. Normally, several suggested kernel functions are the linear kernel ($K(x_i, x_j) = x_i \cdot x_j$), the polynomial kernel with degree d ($K(x_i, x_j) = (x_i \cdot x_j + 1)^d$), the Gaussian kernel ($K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$), and the sigmoid kernel ($K(x_i, x_j) = \tan h(\eta \cdot x_i \cdot x_j + \theta)$).

4 Experimental results

This section describes the details of the experiments performed in order to evaluate classification capabilities of the proposed method. The experimental study is divided in two sets of experiments. One is tested by the epileptic EEG data, and the other is implemented in the data set of five mental tasks.

In the experiments, we first use wavelet package to decompose EEG signals and then extract features using approximate entropy or sample entropy. ELM and SVM are selected as classification methods. To evaluate classification capabilities, we combine feature extraction with classification methods. The combination leads to four cases, namely ApEn + SVM, ApEn + ELM, SampEn + SVM, and SampEn + ELM.

Before performance evaluation, the parameters of SVM achieved the best generalization performance are pre-estimated. SVMs in this paper use Gaussian kernel and involve the penalty parameter C as well as kernel parameter γ to be optimized. We determined the tuning parameters γ and C with standard tenfold cross validation. So, the original data is randomly partitioned into ten equal size subsets. Of the ten subsets, a single subset is retained as the validation data for testing the model, and the remaining nine subsets are used as training data. The initial search for optimal parameters γ and C was done on a 10×10 uniform coarse grid in the (γ, C) space, namely $\gamma = [2^{-10}, 2^{-8}, \dots, 2^6, 2^8]$, $C = [2^{-10}, 2^{-8}, \dots, 2^6, 2^8]$. We used LIBSVM [22] in our experiments.

4.1 Epileptic EEG data

4.1.1 Data description

The epileptic EEG data from the Clinique of Epileptology of Bonn University [23] are used as the experimental data for our proposed method. The epileptic EEG data include five data sets denoted by A–E, respectively. Each data set contains 100 single-channel EEG segments at a sampling rate of 173.61 Hz for 23.6 s, and each segment contains 4,096 samples. These EEG data are obtained from three kinds of different people, namely healthy people, epileptic patients during seizure-free interval, and epileptic patients during seizure interval.

Data sets A and B are attained from five healthy volunteers in an awoken state with eyes open and eyes closed, respectively. Data sets C and D are obtained from five epileptic patients during seizure-free intervals (interictal). Data set E is obtained from five epileptic patients during seizure activity (ictal).

All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. After 12-bit analog-to-digital conversion, the data were written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz [23].

4.1.2 Experiment analysis

We first consider a two-classification problem, where the seizure activity (E) is labeled as positive class, while the other data (A, B, C, and D) form negative class. Our experiments are implemented in four different combinations, namely ApEn + SVM, ApEn + ELM, SampEn + SVM, and SampEn + ELM. In our experiments, data sets randomly divide into two sets, 80 % of data from positive class and negative class, respectively, are used to train classifiers, and the rest for testing the model. The r value is changed from 0.11 to 0.25, while m value is equal to 1, 2, or 3. The performances of each combination are evaluated by overall accuracy defined in the following equation

$$\begin{aligned} \text{Overall accuracy (\%)} &= \frac{\text{The total number of correctly detected patterns}}{\text{The total number of applied patterns}} \times 100 \end{aligned} \tag{19}$$

In order to select suitable Daubechies wavelet function and the level of decomposition, we take epileptic EEG data as experimental data set and give the experimental results as shown in Fig. 3. The left of Fig. 3 shows the classification accuracy and the right shows the time of feature extraction.

From Fig. 3, we can see that the experiment obtains the best accuracy in the case of Daubechies “db2” wavelet

function with 5-level decomposition, while the time of feature extraction is the shortest. So, in this experiment, we select the Daubechies “db2” wavelet function with 5-level decomposition to analyze the signal by wavelet package transform.

Experiments are repeated ten times, and we compute their average. Figure 4 shows the overall accuracies of four different combination strategies in the case of different parameters m and r .

Based on the experimental results in Fig. 4, ApEn + SVM strategy achieves higher accuracy values, in comparison with other strategies. Especially, the overall accuracy of ApEn + SVM strategy achieves 100 % in the case of $m = 2, r = 0.15$ and $m = 3, r = 0.23$. On the other hand, SampEn + ELM strategy shows very good result

which achieved 99.6 % overall accuracy in the case of $m = 2$ and $r = 0.17$. ApEn + ELM strategy achieves 99.4 % overall accuracy in the case of $m = 1$ and $r = 0.19$, while SampEn + SVM strategy only obtains 96.3 % overall accuracy in the case of $m = 2$ and $r = 0.19$. Generally, the overall accuracy using ApEn is higher than the overall accuracy using SampEn, when the experiment adopts the same classification method, SVM or ELM.

In order to further analyze the proposed methods, we employ four different classification problems presented by Tzallas et al. [24] in our experiments. The first and second problems are binary-classification problems, while the third and fourth are three-classification problems. The first problem includes normal (A) and seizure (E). The second problem includes non-seizure (A, B, and C) and seizure (E). We select out normal (A), seizure-free (C) and seizure (E) in the third problems. In the fourth problem, EEG data are classified into three categories including normal (A and B), seizure-free (C and D), and seizure (E). The above classification problems are described in Table 1.

Due to SampEn is superior to ApEn in the training time, we select out two combination strategies to experimental comparison, namely SampEn + SVM and SampEn + ELM. Referring to the experimental results in Fig. 4, we set $m = 2$ and $r = 0.19$ for the combination of SampEn and SVM, and set $m = 2$ and $r = 0.17$ for the

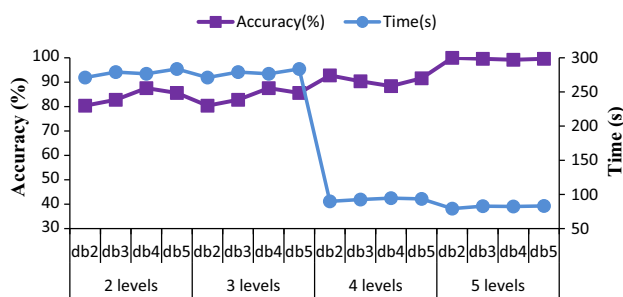


Fig. 3 Accuracy and time in different dbN and levels

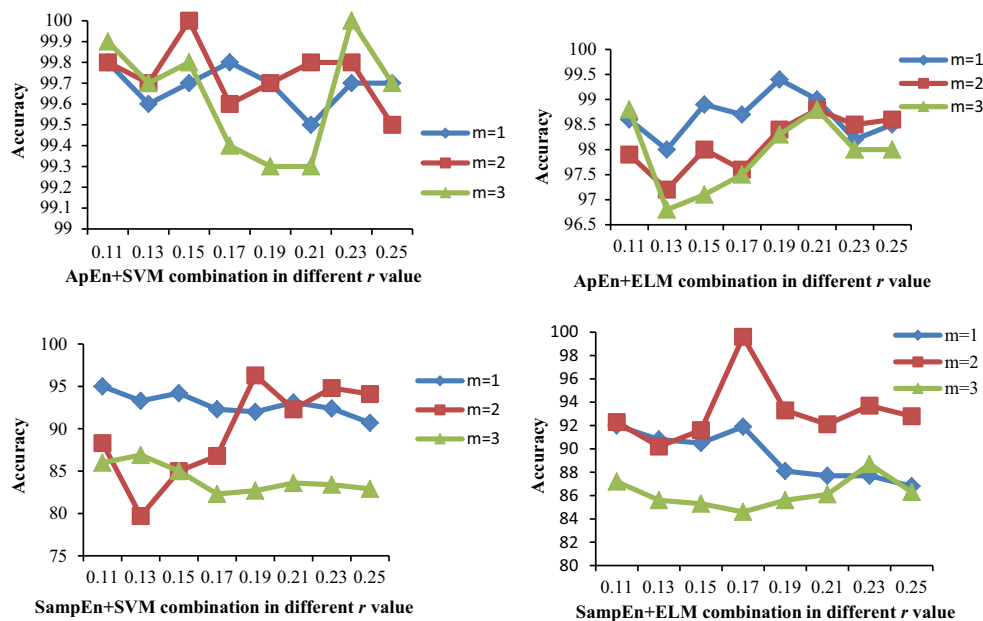


Fig. 4 Comparison of overall accuracies in four different combinations

Table 1 Description of different classification problems

Problems	Classes	Number of EEG segments
1	Normal (A)	100
	Seizure (E)	100
2	Non-seizure (A, B, C)	300
	Seizure (E)	100
3	Normal (A)	100
	Seizure-free (C)	100
	Seizure (E)	100
4	Normal (A and B)	200
	Seizure-free (C and D)	200
	Seizure (E)	100

combination of SampEn and ELM. Table 2 shows the classification accuracy for four classification problems by SampEn + SVM and SampEn + ELM approaches in this paper. Based on the results in Table 2, SampEn + ELM achieved higher accuracy values for classification problems 1, 3, and 4, in comparison with SampEn + SVM, while they achieved the similar accuracy in the second classification problem.

4.2 Five mental tasks

4.2.1 Data description

In this paper, we used EEG data from Colorado State University recorded by Keirn and Anderson [25] as the experimental data set. The EEG data collected signals from seven participants for five different mental tasks, each of which is 7 rows by 2,500 columns. The 7 rows represent the corresponding seven channels, namely c3, c4, p3, p4, o1, o2, and EOG. Each task is trialed no <5 times, each of which is 10 s. The acquired signal is sampled at a frequency of 250 Hz. Recording of these EEG data was performed with a bank of Grass 7P511 amplifiers whose bandpass analog filters were set at 0.1–100 Hz. Information of five mental tasks in the experiment is shown as Table 3 [26].

Before feature extraction, EEG signals are first normalized with zero mean and variance one and then divided into 0.5-s segments with an overlap of 0.25-s segments

Table 3 Description of five mental tasks

Serial number	Task	Description
1	Baseline (B)	The participants relaxed their mind and did not perform any mental tasks
2	Multiplication (M)	The participant was asked to perform a multiplication mentally without vocalizing
3	Letter composition (L)	The participants were asked to mentally compose a simple letter in mind
4	Rotation (R)	The participant was given a short time to observe a complex object and was instructed to visualize the object being rotated about an axis after the object was removed
5	Counting (C)	The participants performed mentally counting number from 1 to 9 repeatedly by visualizing the number on a blackboard in their mind

between adjacent segments. Thus, 39 segments are produced.

4.2.2 Experiment analysis

In this study, we make a comparison for each subject in different combinations of mental tasks. From the view of overall accuracy and training time, we select the combination of SampEn and ELM to perform this experiment. The experiment selects the Daubechies “db2” wavelet function with 5-level decomposition and sets parameters $m = 2$ and $r = 0.17$. Tables 4, 5, 6, and 7 show the classification accuracy of 2, 3, 4, and 5 metal tasks, respectively, for each subject.

Tables 4 and 5 show the classification results with the better accuracies of the employed strategy, in the case of the combinations of two and three mental tasks, respectively. However, for the combinations of four mental tasks shown in Table 6, subject 3 achieves the lowest accuracy value 66.05 % in the case of B, C, M, and R mental tasks, and subject 6 also achieves the lowest accuracy value 70.59 % in the case of B, L, M, and R mental tasks. The other

Table 2 Comparison of classification accuracy for four classification problems

	1		2		3			4		
	A	E	ABC	E	A	C	E	AB	CD	E
SampEn + SVM	76.66	77.33	100.00	96.67	80.00	97.67	97.00	34.08	98.58	97.33
SampEn + ELM	100.00	100.00	99.87	100.00	97.50	100.00	100.00	97.75	100.00	100.00

Table 4 The classification accuracy of two mental tasks for each subject (%)

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
B&C	99.21	97.92	96.32	99.00	97.76	98.05	97.40
B&L	99.00	97.60	97.16	98.42	99.20	97.89	97.50
B&M	98.89	98.33	97.00	98.74	99.02	97.58	98.54
B&R	99.05	98.02	97.89	99.00	98.88	97.37	98.65
C&L	98.47	97.92	97.68	97.32	98.43	97.26	98.44
C&M	98.58	97.81	98.26	98.26	98.29	96.00	98.33
C&R	98.74	96.87	98.21	98.63	98.64	96.84	98.54
L&M	97.53	97.92	98.63	99.05	98.92	96.58	98.54
L&R	98.53	97.71	98.95	98.53	98.99	96.89	98.23
M&R	98.68	97.71	99.26	98.37	98.81	98.68	98.65

Table 5 The classification accuracy of three mental tasks for each subject (%)

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
B&C&L	97.72	96.94	91.58	96.35	92.51	94.82	96.49
B&C&M	97.96	97.57	91.51	96.56	92.69	93.86	98.07
B&C&R	98.39	97.71	91.26	97.33	93.22	93.60	97.89
B&L&M	98.81	96.39	96.42	97.02	97.25	91.67	98.60
B&L&R	98.67	96.25	96.70	96.21	98.01	93.16	97.72
B&M&R	98.91	96.88	66.74	97.75	97.66	82.63	98.77
C&L&M	97.58	96.04	96.63	94.39	98.01	91.49	97.19
C&L&R	98.42	95.97	96.67	95.40	97.49	92.02	95.96
C&M&R	97.86	96.04	79.05	97.23	97.31	82.63	96.14
L&M&R	98.81	96.32	93.16	97.23	97.78	85.35	97.19

Table 6 The classification accuracy of four mental tasks for each subject (%)

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
B&C&L&M	97.57	93.68	91.78	94.80	92.15	93.22	96.45
B&C&L&R	97.24	94.34	92.37	92.83	93.42	92.57	95.13
B&C&M&R	97.96	96.84	66.05	95.13	92.68	73.88	95.79
B&L&M&R	97.96	91.32	73.09	94.61	95.61	70.59	95.79
C&L&M&R	96.78	89.87	73.09	94.01	96.49	71.38	95.00

Table 7 The classification accuracy of five mental tasks for each subject (%)

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7
B&C&L&M&R	96.74	90.63	70.32	92.58	92.14	79.42	95.58

subjects remain better classification accuracy for the combinations of four mental tasks. From Table 7, we can see that subject 3 and subject 6 achieve less than 80 % accuracy, while the other subjects perform with

over 90 % accuracy. Especially, subject 1 obtains 96.74 % accuracy.

Figure 5 gives an intuitional description of average classification accuracy in different tasks for each subject.

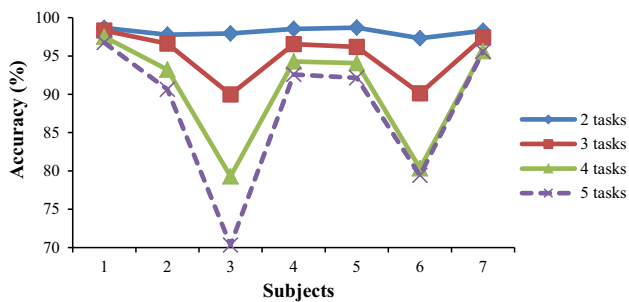


Fig. 5 Comparison of classification accuracy in different tasks for seven subjects

5 Conclusions

Wavelet has played an important role in biomedical signal processing for its ability to capture localized spatial-frequency information of EEG signals. In this paper, we have evaluated the performance of two classifiers, namely SVM and ELM on two EEG signals. The proposed scheme uses approximate entropy and sample entropy based on WPD to calculate the feature values, which are acted as the inputs of two classifiers. The combination of SampEn and ELM has shown great performance in classification accuracy and training time.

Acknowledgments This work is partly supported by National Natural Science Foundation of China (No. 61373127), the China Postdoctoral Science Foundation (No. 20110491530), and the University Scientific Research Project of Liaoning Education Department of China (No. 2011186).

References

- Parvinnia E, Sabeti M, Zolghadri Jahromi M, Boostani R (2014) Classification of EEG signals using adaptive weighted distance nearest neighbor algorithm. *J King Saud Univ-Comput Inform Sci* 26(1):1–6
- Srinivasan V, Eswaran C, Sriraam N (2007) Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans Inform Technol Biomed* 11(3):288–295
- Guo L, Wu Y, Cao T, Yan W, Shen X (2011) Classification of mental task from EEG signals using immune feature weighted support vector machines. *IEEE Trans Magn* 47(5):866–869
- Murugavel ASM, Ramakrishnan S (2012) Multi-class SVM for EEG signal classification using wavelet based approximate entropy. *CCSIT 2012*, pp 335–344
- Burke DP, Kelly SP, de Chazal P, Reilly RB, Finucane C (2005) A parametric feature extraction and classification strategy for brain–computer interfacing. *IEEE Trans Neural Syst Rehabil Eng* 13(1):12–17
- Polat K, Günes S (2007) Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Appl Math Comput* 187(2):1017–1026
- Kumar Y, Dewal ML, Anand RS (2014) Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network. *Signal Image Video Proc* 8(7):1323–1334
- Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst Appl* 32(4):1084–1093
- Guo L, Riveero D, Pazaos A (2010) Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *J Neurosci Methods* 193:156–163
- 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* 6:556–567
- Song Y, Crowcroft J, Zhang J (2012) Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. *J Neurosci Methods* 210:132–146
- Zhang J, Wang N, Kuang H, Wang R (2014) An improved method to calculate phase locking value based on Hilbert-Huang transform and its application. *Neural Comput Appl* 24(1):125–132
- Bajaj V, Pachori RB (2012) Classification of seizure and non-seizure EEG signals using empirical mode decomposition. *IEEE Trans Inform Technol Biomed* 16(6):1135–1142
- Wu HT, Lee CY, Liu CC, Liu AB (2013) Multiscale cross-approximate entropy analysis as a measurement of complexity between ECG R–R interval and PPG pulse amplitude series among the normal and diabetic subjects. *Comput Math Methods Med*, vol. 2013, Article ID 231762
- Ahangi A, Karamnejad M, Mohammadi N, Ebrahimpour R, Bagheri N (2013) Multiple classifier system for EEG signal classification with application to brain–computer interfaces. *Neural Comput Appl* 23(5):1319–1327
- Hu DY, Li W, Chen X (2011) Feature extraction of motor imagery EEG signals based on wavelet packet decomposition. In: *Proceedings of the 2011 IEEE/ICME international conference on complex medical engineering*, pp 694–697
- Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA* 88(6):2297–2301
- Richman J, Moorman J (2000) Physiological time series analysis using approximate entropy and sample entropy. *Am J Physiol* 278(6):2039–2049
- Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501
- Huang G-B, Zhou H, Ding X, Zhang R (2011) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern* 42(2):513–529
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York
- Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2(3):1–27. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE (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 Stat Nonlinear Soft Matter Phys* 64:061907-1–061907-8
- Tzallas AT, Tsipouras MG, Fotiadis DI (2007) Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Comput Intell Neurosci*. Article ID 80510
- EEG Data. http://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon
- Geethanjali P, Mohan YK, Jinisha S (2012) Time domain Feature extraction and classification of EEG data for brain computer interface. In: *Proceedings of the 9th international conference on fuzzy systems and knowledge discovery (FSKD 2012)*, May 29–31, 2012, IEEE Press, pp 1136–1139