Expert Systems with Applications 38 (2011) 14314–14320

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection

Deng Wang *, Duoqian Miao, Chen Xie

Department of Computer Science and Technology, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China
The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 201804, China

ARTICLE INFO

Keywords:
Electroencephalogram (EEG)
Feature extraction
Wavelet packet entropy
Epileptic detection
Hierarchical knowledge base

ABSTRACT

In this study, a hierarchical electroencephalogram (EEG) classification system for epileptic seizure detection is proposed. The system includes the following three stages: (i) original EEG signals representation by wavelet packet coefficients and feature extraction using the best basis-based wavelet packet entropy method, (ii) cross-validation (CV) method together with *k*-Nearest Neighbor (*k*-NN) classifier used in the training stage to hierarchical knowledge base (HKB) construction, and (iii) in the testing stage, computing classification accuracy and rejection rate using the top-ranked discriminative rules from the HKB. The data set is taken from a publicly available EEG database which aims to differentiate healthy subjects and subjects suffering from epilepsy diseases. Experimental results show the efficiency of our proposed system. The best classification accuracy is about 100% via 2-, 5-, and 10-fold cross-validation, which indicates the proposed method has potential in designing a new intelligent EEG-based assistance diagnosis system for early detection of the electroencephalographic changes.

1. Introduction

Brain is one of the most vital organs of humans, controlling the coordination of human muscles and nerves. Epilepsy, sometimes referred to as a seizure disorder, is a brain disorder that involves neuronal networks and is characterized predominantly by recurrent and unpredictable interruptions of normal brain function, called epileptic seizures (Fisher et al., 2005). It was estimated that approximately one in every 100 persons will experience a seizure at some time in their life (Iasemidis et al., 2003). Although the diagnosis of epilepsy is clinical, the scalp electroencephalogram (EEG) is the most widely accepted test for the diagnosis of epilepsy (Güler, Übeyli, & Güler, 2005; Güler & Übeyli, 2005; Polat & Güneş, 2007, 2008; Subasi, 2007). Careful analysis the EEG records can provide valuable insight into the detection of epileptiform discharges. Through the electroencephalographic records, developing a detection system using computers has long been under study for several years (Gabor & Seyal, 1992; Glover, Raghaven, Ktonas, & Frost, 1989; Nigam & Graupe, 2004; Webber, Litt, Lesser, Fisher, & Bankman, 1993). An ideal epileptic detection system is the one that has 100% seizure detection rate along with 0% false positive rate (i.e., the rate of mis-classified healthy volunteer).

General pattern recognition approach on EEG signals for detection of electroencephalographic changes includes pre-processing, feature extraction, feature selection/dimensionality reduction and classification. A feature extraction is the determination of a feature or a feature vector from a pattern vector (Cvetkovic, Übeyli, & Cosic, 2008). Since the EEG is a time-varying and space-varying nonstationary signal, this made both wavelet transform (WT) and wavelet packet transform (WPT) excellent candidates for feature extraction from such data. Many literatures have demonstrated that WPT is one of the most promising methods to extract features from the EEG signals (Adeli, Zhou, & Dadmehr, 2003; Ocak, 2009; Yang, Yan, Yan, & Wu, 2006; Yildiz, Akin, Poyraz, & Kirbas, 2009). The first challenge in this study about EEG signal processing is how to choose features that best characterize the nonstationary EEG signals. The best basis selection provides a means for choosing the features which are best for classification based on Shannon entropy. In this study, the best basis-based wavelet packet entropy feature extraction algorithm was first proposed.

In real life, fast or real-time decision-making is more important instead of getting global optimization in human cognition. The mental processing based on prior knowledge or experience always needs not complex computing. An interesting idea is that there are many different decision-making levels on which one can choose and utilize them to generate actions. Therefore, the second challenge is how to store the experiences or rules during learning or training stage, and which level is sufficient for decision-making in the testing stage. More recently, granular computing which

* Corresponding author at: Department of Computer Science and Technology, Tongji University, 4800 Cao'an Highway, Shanghai 201804, China.

E-mail addresses: w_deng208@hotmail.com (D. Wang), miaoduoqian@163.com (D. Miao), x_c_grace@163.com (C. Xie).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.05.096

recognizes and exploits the knowledge present in data at various levels of resolution has become an emerging computing paradigm of information processing (Lin, 2003; Yao, 2000; Yao & Yao, 2002).

Motivated by ideas from EEG signal processing and granular computing theory, in this study, we propose a system to classification of EEG signals. The proposed system has three stages. (1) Original signals are represented by WPT, and then the best basis-based wavelet packet entropy method was applied for feature extraction. (2) In the training stage, cross-validation (CV) method together with k-Nearest Neighbor (k-NN) classifier was used for hierarchical knowledge base (HKB) construction. During each validation process, the obtained optimal k-values with the best classification accuracies as the discriminative rules were stored and reorganized. (3) In the testing stage, to categorize a new sample into either epileptic or normal class, k-NN classifier using discriminative rules from HKB was used to calculate the similarity between the new sample and the corresponding k training process samples, respectively, and then uses the class labels of the k most similar neighbors to predict the class of the new sample.

In order to trade off between classification reliability level and classification performance to EEG signals, a threshold minimal confidence level (MCL) was introduced. In this study, the reject option in the testing stage was also been considered because by setting the threshold MCL, only samples with a high confidence (i.e., greater than or equal to MCL) are indeed classified. In other words, those samples outside the boundaries of the known classes or from the ambiguity region between classes are rejected. Experiments were carried out upon a publicly available EEG database. The objective was to discriminate recordings from healthy volunteers with eyes open and epilepsy patients during epileptic seizures. Our experimental results on the proposed method could able to achieve significant improvements.

The organization of the paper is as follows. In Section 2, we briefly describe the data set of the EEG signals used in our study, and then we explain the proposed method with subsections. In Section 3, experimental results and discussion are given. Finally, we conclude the study in Section 4.

2. Materials and methods

2.1. EEG data description

A publicly available EEG database (EEG time series) was experimented with in this study. This session will give a short description and refer to Andrzejak et al. (2001) for further details. The complete data set consists of five sets (denoted A-E) each containing 100 single channel EEG segments of 23.6 s duration. These segments were selected and cut out from continuous multi-channel EEG recordings after visual inspection for artifacts, e.g., due to muscle activity or eye movements. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme (see Fig. 1). Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D, and E originated from EEG archive of presurgical diagnosis. The EEGs from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone. Segments in set D were recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12 bit resolution. Band-pass filter settings were 0.53–40 Hz (12 dB/oct) (Andrzejak et al., 2001). In this study, we used two dataset (A and E) of the complete dataset. Typical EEGs from Set A and E are depicted in Fig. 2.

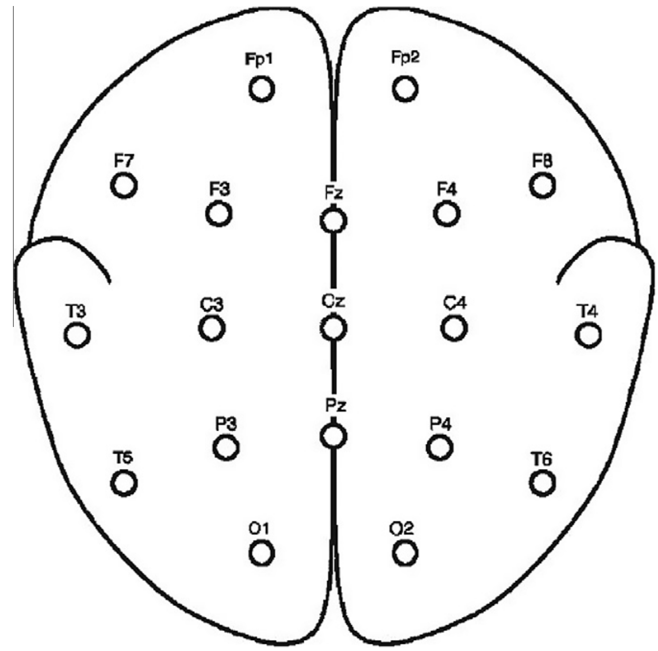

Fig. 1. Scheme of the locations of surface electrodes according to the international 10–20 system. Names of the electrode positions are derived from their anatomical locations (Andrzejak et al., 2001).

2.2. The proposed method

2.2.1. WPT representation of EEG signals

The wavelet packet transform (WPT) can be viewed as a generalization of the classical wavelet transform, which provides a multi-resolution and time–frequency analysis for non-stationary EEG data. The wavelet packet transform generates the full decomposition tree, as depicted in Fig. 3. A low (L) and high (H) pass filter is repeatedly applied to the function f, followed by decimation by 2, to produce a complete subband tree decomposition to some desired depth. The low- and high-pass filters are generated using

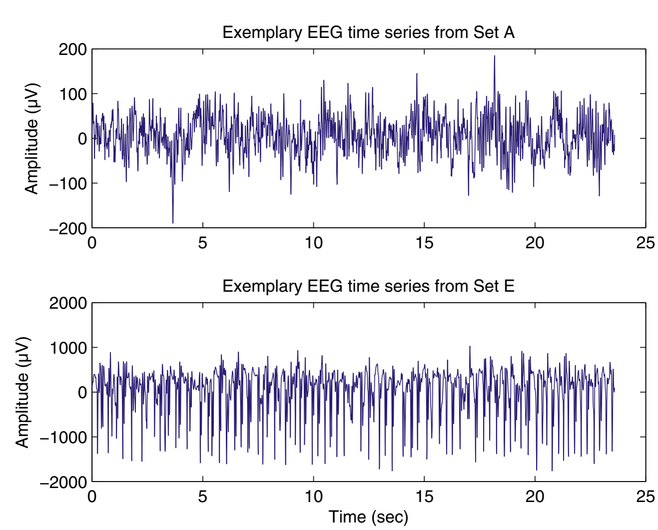

Fig. 2. Examples of two different sets of EEG signals taken from different subjects.

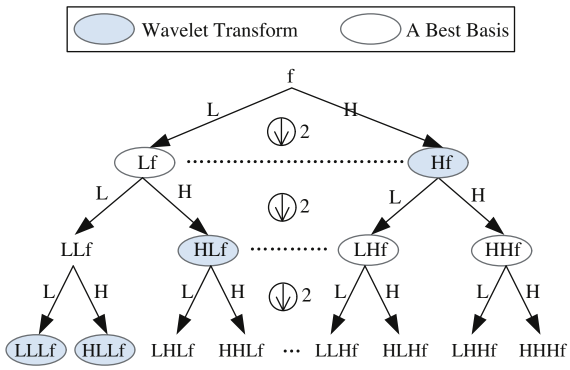

Fig. 3. Illustration of the wavelet packet decomposition. The wavelet transform basis is indicated by gray ovals. The orthogonality of the basis functions allows the use of an additive cost function to determine the optimal basis for data compression. An example of a possible best basis is shown using white ovals. (Figure adapted from Jones et al., 2002).

orthogonal basis functions (Jones, Begleiter, Porjesz, Wang, & Chorlian, 2002). Because WPT not only decomposes the approximations of the signal but also details, it holds the important information located in higher frequency components than WT in certain applications. Therefore, in this study the first representation studied is composed of the WPT coefficients.

A wavelet packet is represented as a function (Shinde & Hou, 2004):

$$\psi^i_{j,k}(t) = 2^{-j/2}\psi^i(2^{-j}t - k), \tag{1}$$

where i is the modulation parameter, j is the dilation parameter and k is the translation parameter. $i = 1,2,\ldots,j^n$, and n is the level of decomposition in wavelet packet tree.

The wavelet ψ^i is obtained by the following recursive relationships:

$$\psi^{2i} = \frac{1}{\sqrt{2}}\sum_{-\infty}^{\infty} h(k)\psi^i\left(\frac{t}{2} - k\right); \tag{2}$$

$$\psi^{2i+1} = \frac{1}{\sqrt{2}}\sum_{-\infty}^{\infty} g(k)\psi^i\left(\frac{t}{2} - k\right). \tag{3}$$

Here ψ^i is called as a mother wavelet and the discrete filters $h(k)$ and $g(k)$ are quadrature mirror filters associated with the scaling function and the mother wavelet function (Daubechies, 1992).

The wavelet packet coefficients $c^i_{j,k}$ corresponding to the signal $f(t)$ can be obtained as,

$$c^i_{j,k} = \int_{-\infty}^{\infty} f(t)\psi^i_{j,k}(t)\mathrm{d}t \tag{4}$$

provided the wavelet coefficients satisfy the orthogonality condition. The wavelet packet component of the signal at a particular node can be obtained as

$$f^i_j(t) = \sum_{k=-\infty}^{\infty} c^i_{j,k}\psi^i_{j,k}(t)\mathrm{d}t. \tag{5}$$

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. In order to further reduce the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients was used (Güler & Übeyli, 2005; Tzanetakis, Essl, & Cook, 2001). The following statistical features were used to represent the time–frequency distribution of the EEG signals in this study:

- Maximum of the absolute values of the coefficients in each subband.
- Minimum of the absolute values of the coefficients in each subband.
- Mean of the absolute values of the coefficients in each subband.
- Standard deviation of the coefficients in each subband.

2.2.2. Wavelet packet entropy and best basis selection

The number of decompositions from a signal in different ways may be very large. An exhaustive search for the optimal decomposition is not feasible, given the large number of possible binary subtrees decompositions. Therefore, it is necessary to find an optimal decomposition by using a convenient algorithm. The best basis selection provides a means for choosing the features which are best for classification based on various criteria. The selection process is based on either (1) best representation of a given class of signals, or (2) best distinction between classes (Ubeyli & Guler, 2007). Entropy is a common method in many fields, especially in signal processing applications. Commonly, there are several useful entropy types such as Shannon, log energy, sure, threshold, etc. (Coifman & Wickerhauser, 1992) for calculating the lowest cost basis. Here only one of the most attractive cost functions Shannon entropy (Shannon, 1948) was employed which is a measure of signal complexity to wavelet coefficients generated by WPT where larger entropy values represent higher process uncertainty and therefore higher complexity. In Rosso et al. (2001), wavelet entropy can provide useful information about the underlying dynamical process associated with the signal. The entropy 'E' must be an additive information cost function such that $E(0) = 0$ and

$$E(s) = \sum_i E(s_i). \tag{6}$$

The entropy for the observed signal in l^p norm with $p \geqslant 1$ can be expressed as

$$E(s_i) = |s_i|^p, \tag{7}$$

and

$$E(s) = \sum_i |s_i|^p. \tag{8}$$

The Shannon entropy is defined as

$$E(s) = -\sum_i s_i^2 log(s_i^2), \tag{9}$$

where s_i represents coefficients of signal s in an orthonormal basis. If the entropy value is greater than one, the component has a potential to reveal more information about the signal and it needs to be decomposed further in order to obtain simple frequency component of the signal (Ekici, Yildirim, & Poyraz, 2008; Shinde & Hou, 2004). By using the entropy, it gave a useful criterion for comparing and selection the best basis.

2.2.3. The procedure of feature extraction

Given the EEG signals, the best basis-based wavelet packet entropy feature extraction is obtained by performing the following steps:

(*a*) Select a wavelet function W and specify the decomposition level L.
(*b*) Calculate the sample mean SM.
(*c*) Decompose SM at the specified level with the selected wavelet function, and return a wavelet packet tree T. Let $B_{l,k}$ be the set of WPT basis vector, $0 \leqslant l < L$, $1 \leqslant k \leqslant 2^L - 1$.
(*d*) Calculate energies $E_{l,k}$ for all subbands using Eq. (9).

(*e*) Set the initial basis $B = \{B_{L-1,1}, B_{L-1,2}, \ldots, B_{l,k}, \ldots, B_{L-1,2^{L-1}}\}$ related to the subbands at the bottom level.
(*f*) Compare the entropy of a parent node $E_{l,k}$ with the sum of the entropy of two child nodes $(E_{l+1,2k-1} + E_{l+1,2k})$. If $E_{l,k} \leqslant (E_{l+1,2k-1} + E_{l+1,2k})$, then replace $B_{l+1,2k-1}$ and $B_{l+1,2k}$ by $B_{l,k}$ in B; else set $E_{l,k} = (E_{l+1,2k-1} + E_{l+1,2k})$, i.e., assign the sum of the children's entropy to the parent node.
(*g*) Repeat (*f*) for the next higher level until the root is reached.
(*h*) Select a sample from training set.
(*i*) Decompose the sample to L using W.
(*j*) Calculate wavelet coefficients in the corresponding best basis B.
(*k*) Calculate *Max*, *Min*, *Mean* and *Standard deviation* of the wavelet coefficients to form a 4-dimension feature.
(*l*) Repeat steps (*h*)–(*k*) for all samples.

The wavelet packet decomposition of the EEG signals was implemented by using the MATLAB software package (MATLAB version 7.0 with WPT toolbox). Using the above procedure, 4-dimension features are extracted from EEG signals.

2.2.4. k-NN classifier

In pattern recognition community, the k-Nearest Neighbor (k-NN) classifier is a popular method owing to its simplicity, interpretability and good performance (Cunningham & Delany, 2007). This method is a supervised learning method which is to find the set of k nearest neighbors in the training set to a test sample x_0 and then classifier x_0 as the most frequent class among the k neighbors. In the present study, k-NN classifier is used to categorize each new sample into either healthy subjects or subjects suffering from epilepsy diseases class.

2.2.5. Hierarchical classification system

Using the same features for classification of decision-making in different surroundings may result in different results. For example, imagine that you are a teacher evaluating a student. Test subject achievement of each student may be one of the important features to put forward the accomplishment evaluation. But in reality, only this information is insufficient because other students' studying status, i.e. subject achievements, are indispensable. It means there must have other information as the background references. Therefore, a distinguishing object must have a set of features and the corresponding background reference information. Moreover, it should be note that different feature set and the corresponding background reference information would result in different confidence for decision-making. Based on the above observation, we shall give definitions to characterize.

Definition 1 (*Discriminative rule*). A discriminative rule DR is a 3-tuple defined as follows:

$$DR = (F, K, R), \tag{10}$$

where F is the set of features, K is the number of nearest neighbors in a k-NN classifier which corresponding to the background reference, and R represents the degree of confidence for classification under F and K, $0 \leqslant R \leqslant 1$.

Definition 2 (*Hierarchical knowledge base*). Let $DR = (F,K,R)$ be a discriminative rule. A hierarchical knowledge base HKB is defined as a sequence:

$$HKB = \langle \ldots, DR_i, DR_j, \ldots \rangle, \tag{11}$$

such that for $1 \leqslant i, j \leqslant n$, $Acc(DR_i) \geqslant Acc(DR_j)$, where $Acc(DR)$ indicates the classification accuracy by using DR. n is the number of discriminative rules at HKB. From the above definitions, we can see that a hierarchical knowledge base is composed of all discriminative rules reorganized on different classification accuracy levels.

There is another phenomena linked to training stage, in practice not all samples in a training set are useful for classifiers; maybe some are noisy or redundant. Therefore, it is convenient to discard irrelevant instances from training set by setting R which represents degree of confidence for classification in Definition 1. This process is similar to prototype selection (Sanchez, Pla, & Ferri, 1997). Through this process, the size of training set is cut down which could be useful for reducing the time consuming in the training stage, particularly for large datasets. It is known that people making decision in real life not always base on full confidence, especially in some emergency situations or given imperfect information. In order to describe quantitatively the degree of confidence, we introduce a concept: minimal confidence level (MCL). It is a threshold (decimal between 0 and 1). Since the feature information from training samples rarely meets the standard of perfection and complete, what we need to do is to ignore the imperfections, that is, to continue training operation as if the information were sound and complete. This implies that the decision-making from the training set might be inaccurate, namely, few 100% confidence for classification. However, if the confidence level is too low or below a certain threshold, the result can be discarded. Therefore, setting the value of MCL needs more careful. Obviously, the higher MCL, the stricter requirements with the correct result is, and vice versa.

Based on above study, we proposed a hierarchical classification system illustrated in Fig. 4. The system consists of three main parts: (i) extract features by using the best basis-based wavelet packet entropy feature extraction, (ii) during the training stage, the discriminative rules and the corresponding background reference so-called k-domain (the optimal value of k) are extracted and reorganized (sorting all discriminative rules by R in descending order) using cross-validation for constructing a hierarchical knowledge base (HKB), (iii) in the testing stage, by selecting the top-ranked discriminative rules from the HKB according to the up-

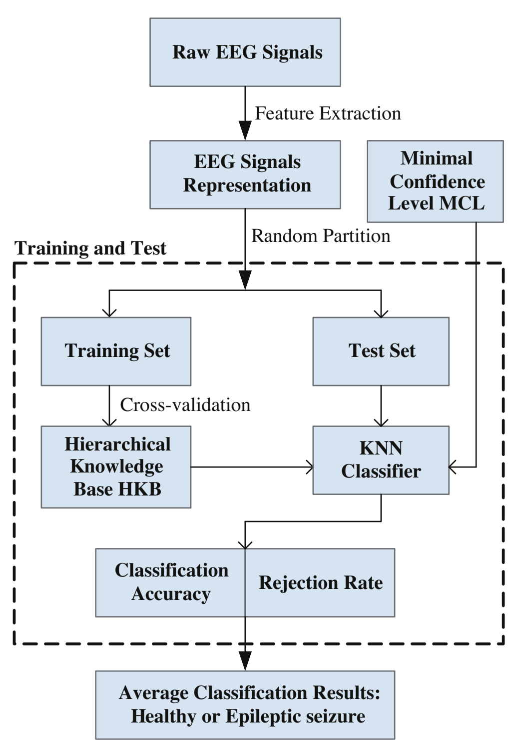

Fig. 4. The flow chart of the proposed system.

bottom search strategy, compute classification accuracy for those recognizable samples, and rejection rate for those unrecognizable samples. This process repeated until the classification accuracy is greater than or equal to minimal confidence level (MCL). Finally, the Training and Test part repeat in Fig. 4, and the average classification results are used to evaluate the performance of the proposed system.

3. The experimental results

In this section, we present the performance measure methods used to evaluate the proposed method. Finally, we give the experimental results and discuss our observations.

3.1. Performance measure

Two performance measures are used for the evaluation of EEG signals classification. One is commonly used average classification accuracy, which is defined as

$$\overline{Acc} = \left(\frac{N_{correct}}{N_{total}}\right), \tag{12}$$

where $N_{correct}$ is the number of correctly classified samples for those recognizable samples, and N_{total} is the number of all testing samples.

In the testing stage, reject rate for unrecognizable samples are also used as the other performance measure because the new sample may belong to one of the classes but is hard to classify with a minimal confidence level MCL.

$$\overline{Rr} = \left(\frac{N_{unrecogniable}}{N_{total}}\right), \tag{13}$$

where $N_{unrecogniable}$ is the number of unrecognizable samples.

3.2. Results and discussion

All the experiments of this section were done over 100 EEG time series of 4096 samples for each class from Set A and E mentioned in Section 2.1. There were two diagnosis classes: healthy and a patient who is subject to possible epileptic seizure. Classification results of the system were reported by using m-fold cross-validation method. That is to say, the data from each set was randomly partitioned into m complementary subsets, one of the subsets was used as the validation set and the other nine subsets were put together to form a training sets. Then in order to reduce variability, this procedure was repeated m times and the average classification results were computed. In our experiment study, all the obtained results were presented by using 2-, 5- and 10-fold cross-validation. The class distribution of the samples in the training set and testing set is summarized in Table 1.

In this present work, five wavelet functions (represented in the MATLAB Wavelet toolbox) in common use, such as Daubechies, Coiflets, Symlets, Discrete Meyer Wavelet, Biorthogonal and its reverse version, were examined and compared with decomposition level of 3. Our experiments showed that: (1) in Daubechies family of wavelets, the Db1 type of wavelet provided the best classification results; (2) Coiflet with order 4 (coif4) and the Symlet with order 7 (sym7) showed best results among the Coiflet and Symlet families of orthogonal wavelets, respectively; (3) in case of biorthogonal types of wavelets, bior1.1 and bior 2.6 provided best compression results where as in case of reverse type of biorthogonal wavelets rbio3.1 provided better compression results; (4) Dmey wavelet of Discrete Meyer was also examined. In order to investigate the effect of these wavelets, tests were carried out and the results including number of wrongly classified and rejected for different wavelet by using 2-fold cross-validation which are shown in Fig. 5. One can see that the dmey wavelet offers lower wrongly classified and rejected than the others, and the coif4 is marginally higher than the dmey. Hence, the dmey wavelet was chosen for this application.

From Tables 2–4, the number of correctly classified, wrongly classified and rejected were tested by varying the MCL from 0 to 1 with 2-, 5- and 10-fold cross-validation, respectively. We can see that the best result is 100% correctly classified with 0 rejected samples. As expected, the number of rejected samples reaches its maximum value when MCL equals to 1. The possible explanation is that the confidence level is the highest in this point in which only samples with the highest probability of being correctly classified are indeed classified (i.e., it prefers to reject more samples rather than classify them). What's more, the peak values of the number of correctly classified appear at the middle region instead of both ends. In Table 5, the average performance was achieved by using 2-, 5-, and 10-fold cross-validation, respectively. It can be seen that 10-fold cross-validation achieves the best performance. The reason for this could be there are more training samples in this test. However, the difference is small and the average accuracy is above 99% which demonstrated the generalization ability of our method. Besides, it can be seen that, unlike other classification systems, the proposed system not only provides the accuracy of the classifier, but also the corresponding degree of confidence.

Finally, we have compared our results with previous results reported by earlier methods. Table 6 gives the classification accuracies of our method and previous methods. As we can see from these results, our method obtained better classification accuracy.

It should be noted that we chose k-Nearest Neighbor (k-NN) as classifier in this study. One of the important considerations is that k is a variant of k-NN classifier which can be used to represent the background reference simply for each cross-validation procedure. Moreover, compared to other classifier methods such as Bayesian classifier, k-NN does not rely upon prior probability, and it is

Table 1
Class distribution of the samples in the training and test data sets.

Class	m-fold cross-validation	Training set	Testing set	Total
Epileptic		800	800	1600
Normal	$m = 2$	800	800	1600
Total		1600	1600	3200
Epileptic		1280	320	1600
Normal	$m = 5$	1280	320	1600
Total		2560	640	3200
Epileptic		1440	160	1600
Normal	$m = 10$	1440	160	1600
Total		2880	320	3200

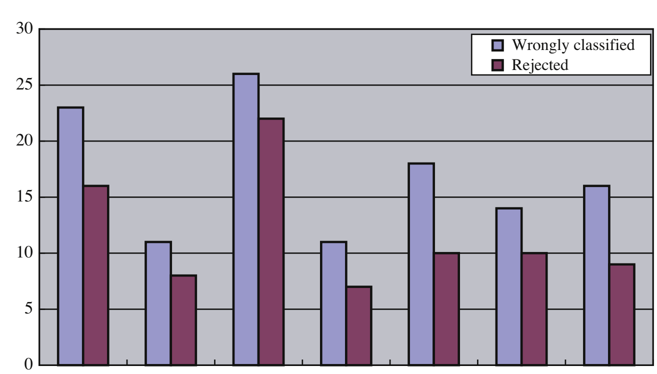

Fig. 5. The number of wrongly classified and rejected obtained for different wavelet when the EEG signals were classified using the proposed method.

Table 2
Classification of test samples by 2-fold cross-validation for the different MCL.

MCL	2-fold cross-validation			
	Total samples	Correctly classified	Wrongly classified	Rejected
0	1600	1595	5	0
0.1	1600	1595	5	0
0.2	1600	1600	0	0
0.3	1600	1593	7	0
0.4	1600	1595	5	0
0.5	1600	1598	2	0
0.6	1600	1593	7	1
0.7	1600	1597	3	0
0.8	1600	1600	0	0
0.9	1600	1590	10	9
1	1600	1528	72	71

Table 3
Classification of test samples by 5-fold cross-validation for the different MCL.

MCL	5-fold cross-validation			
	Total samples	Correctly classified	Wrongly classified	Rejected
0	640	638	2	0
0.1	640	638	2	0
0.2	640	637	3	0
0.3	640	640	0	0
0.4	640	639	1	0
0.5	640	639	1	0
0.6	640	639	1	0
0.7	640	637	3	0
0.8	640	639	1	0
0.9	640	640	0	0
1	640	608	32	31

Table 4
Classification of test samples by 10-fold cross-validation for the different MCL.

MCL	10-fold cross-validation			
	Total samples	Correctly classified	Wrongly classified	Rejected
0	320	319	1	0
0.1	320	318	2	0
0.2	320	318	2	0
0.3	320	319	1	0
0.4	320	320	0	0
0.5	320	319	1	0
0.6	320	319	1	0
0.7	320	319	1	0
0.8	320	319	1	0
0.9	320	319	1	0
1	320	307	13	12

Table 5
The obtained average classification accuracy and the corresponding reject rate for classification of EEG signals with m-fold cross-validation.

Measures	*m*-fold cross-validation		
	$m = 2$	$m = 5$	$m = 10$
Accuracy (%)	99.355	99.420	99.449
Reject rate (%)	0.469	0.457	0.381

computationally efficient. As for the issue of poor run-time performance when need to determine value of parameter k (number of nearest neighbors), if the training set is large, is not such a problem these days with the computational power that is available. In this paper, we varied k-value from 1 to 10 for reducing the computation cost. For another parameter similarity metric, Euclidean distance was used to measure the similarity between two objects.

Table 6
Our method's classification accuracy for classification of EEG signals with classification accuracies obtained by other methods.

Author (Year)	Method	Acc (%)
Güler et al. (2005)	Recurrent neural networks	96.79
Subasi (2007)	Wavelet-ME	95
Subasi (2007)	Wavelet-MLPNN	93.6
Güler and Übeyli (2005)	Wavelet-ANFIS	98.68
Polat and Güneş (2007)	FFT-decision tree classifier (10 x FC)	98.72
Polat et al. (2008)	AIRS-PCA-FFT	100
Our study	Proposed system based on HKB	$99 \sim 100$

The introduced parameter MCL gives a flexible implementation of the discriminative rule selecting from HKB for classification, and it also provides a measure of the confidence of the classification, especially when this classification is used in critical health applications such as epileptic detection. However, there is another open question that is how to choose the optimal MCL, our consideration is that it relies on the actual applications data. In general, our observation is the peak values of the best accuracy appear at the middle region instead of both ends. It means that not the lower or higher of MCL, the better the accuracy is.

4. Conclusion

In this paper, we implemented the best basis-based wavelet packet entropy feature extraction method in the training stage to acquire fewer feature spaces of EEG signals, and in combination with the cross-validation, a hierarchical knowledge base (HKB) was constructed. In the testing stage, the discriminative rules from HKB were chose for final classification according to minimal conference level (MCL). The proposed method was successfully applied to EEG signals for the epileptic detection. The results including classification accuracy and reject rate showed that the proposed system can effectively detect possible epileptic seizure patients from the healthy. The average classification performance was generally superior to or near to some literatures. Especially, using cross-validation the best accuracy can achieve 100%. As we have known, the diagnosis of epilepsy by experts' complete visual analysis is a tedious and costly approach. Our proposed system using EEG can provide an important assistant to physicians, thus to make their decisions on their patients. Future work will consider other more exotic classifiers such as support vector machines, neural networks or other hybrid pattern classifiers.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (No. 60970061, No. 61075056). Thanks are also given to the anonymous reviewers for their kind comments improving the quality and presentation of the paper.

References

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods, 123*(1), 69–87.
Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E, 64*(6), 61907.
Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory, 38*(2), 713–718.
Cunningham, P., & Delany, S.J. (2007). *k-Nearest neighbour classifiers*. Technical Report UCD-CSI-2007-4, Dublin: Artificial Intelligence Group.
Cvetkovic, D., Übeyli, E. D., & Cosic, I. (2008). Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study. *Digital Signal Processing, 18*(5), 861–874.
Daubechies, I. (1992). *10 lectures on wavelets*. Philadelphia: Capital City Press.

EEG time series are available under <http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>.

Ekici, S., Yildirim, S., & Poyraz, M. (2008). Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition. *Expert Systems with Applications, 34*(4), 2937–2944.

Fisher, R. S., van Emde Boas, W., Blume, W., Elger, C., Genton, P., Lee, P., et al. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia, 46*(4), 470–472.

Gabor, A. J., & Seyal, M. (1992). Automated interictal EEG spike detection using artificial neural networks. *Electroencephalography and Clinical Neurophysiology, 83*(5), 271–280.

Glover, J. R., Jr., Raghaven, N., Ktonas, P. Y., & Frost, J. D. Jr., (1989). Context-based automated detection of epileptogenic sharp transientsin the EEG: Elimination of false positives. *IEEE Transactions on Biomedical Engineering, 36*(5), 519–527.

Güler, İ., & Übeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of Neuroscience Methods, 148*(2), 113–121.

Güler, N. F., Übeyli, E. D., & Güler, İ. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications, 29*(3), 506–514.

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

Jones, K., Begleiter, H., Porjesz, B., Wang, K. M., & Chorlian, D. (2002). Complexity measures of event related potential surface Laplacian data calculated using the wavelet packet transform. *Brain Topography, 14*(4), 333–344.

Lin, T. Y. (2003). Granular computing-Structures, representations, and applications. *Lecture Notes on Artificial Intelligence, LNAI 2639*, ISSN 0302-9743. pp. 16–24.

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

Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications, 36*(2), 2027–2036.

Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation, 187*(2), 1017–1026.

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

Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schurmann, M., et al. (2001). Wavelet entropy: A new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods, 105*(1), 65–75.

Sanchez, J. S., Pla, F., & Ferri, F. J. (1997). Prototype selection for the nearest neighbour rule through proximity graphs. *Pattern Recognition Letters, 18*(6), 507–513.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379–423, pp. 623–656.

Shinde, A., & Hou, Z. K. (2004). *A wavelet packet based sifting process and its application for structural health monitoring*. Master thesis, Faculty of Worcester Polytechnic Institute.

Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications, 32*(4), 1084–1093.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Audio analysis using the discrete wavelet transform. In *Proceedings of the conference in acoustics and music theory applications, Skiathos, Greece* (pp. 318–323).

Ubeyli, E. D., & Guler, I. (2007). Features extracted by eigenvector methods for detecting variability of EEG signals. *Pattern Recognition Letters, 28*(5), 592–603.

Webber, W. R. S., Litt, B., Lesser, R. P., Fisher, R. S., & Bankman, I. (1993). Automatic eeg spike detection – What should the computer imitate. *Electroencephalography and Clinical Neurophysiology, 87*(6), 364–373.

Yang, B. H., Yan, G. Z., Yan, R. G., & Wu, T. (2006). Feature extraction for EEG-based brain-computer interfaces by wavelet packet best basis decomposition. *Journal of Neural Engineering, 3*(4), 251–256.

Yao, J. T., & Yao, Y. Y. (2002). Induction of classification rules by granular computing, rough sets and current trends in computing. In *Proceedings of the Third International Conference(RSCTC 2002), Lecture Notes in Artificial Intelligence, LNAI 2475*, ISSN 0302-9743 (pp. 331–338).

Yao, Y. Y. (2000). Granular computing: basic issues and possible solutions. In *Proceedings of the 5th Joint Conference on Information Sciences, New Jersey, USA, vol. I* (pp. 186–189).

Yildiz, A., Akin, M., Poyraz, M., & Kirbas, G. (2009). Application of adaptive neuro-fuzzy inference system for vigilance level estimation by using wavelet-entropy feature extraction. *Expert Systems with Applications, 36*(4), 7390–7399.