Available online at www.sciencedirect.com

ScienceDirect

Expert Systems with Applications 34 (2008) 1954–1962

Expert Systems with Applications

www.elsevier.com/locate/eswa

Wavelet/mixture of experts network structure for EEG signals classification

Elif Derya Übeyli *

Department of Electrical and Electronics Engineering, Faculty of Engineering, TOBB Ekonomi ve Teknoloji Üniversitesi, 06530 Söğütözü, Ankara, Turkey

Abstract

Mixture of experts (ME) is a modular neural network architecture for supervised learning. This paper illustrates the use of ME network structure to guide model selection for classification of electroencephalogram (EEG) signals. Expectation-maximization (EM) algorithm was used for training the ME so that the learning process is decoupled in a manner that fits well with the modular structure. The EEG signals were decomposed into time–frequency representations using discrete wavelet transform and statistical features were calculated to depict their distribution. The ME network structure was implemented for classification of the EEG signals using the statistical features as inputs. To improve classification accuracy, the outputs of expert networks were combined by a gating network simultaneously trained in order to stochastically select the expert that is performing the best at solving the problem. Three types of EEG signals (EEG signals recorded from healthy volunteers with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free interval, and epilepsy patients during epileptic seizures) were classified with the accuracy of 93.17% by the ME network structure. The ME network structure achieved accuracy rates which were higher than that of the stand-alone neural network models.

Keywords: Mixture of experts; Expectation-maximization algorithm; Classification accuracy; Discrete wavelet transform; EEG signals classification

1. Introduction

There have recently been widespread interests in the use of multiple models for pattern classification and regression in statistics and neural network communities. The basic idea underlying these methods is the application of a so-called divide-and-conquer principle that is often used to tackle a complex problem by dividing it into simpler problems whose solutions can be combined to yield a final solution. Utilizing this principle, Jacobs, Jordan, Nowlan, and Hinton (1991) proposed a modular neural network architecture called mixture of experts (ME). The ME models the conditional probability density of the target output by mixing the outputs from a set of local experts, each of which separately derives a conditional probability density of the target output. The ME weights the input space by using the posterior probabilities that expert networks generated for getting the output from the input. The outputs of expert networks are combined by a gating network simultaneously trained in order to stochastically select the expert that is performing the best at solving the problem (Chen, Xu, & Chi, 1999; Hong & Harris, 2002). As pointed out by Jordan and Jacobs (1994), the gating network performs a typical multiclass classification task (Mangiameli & West, 1999; Hu, Palreddy, & Tompkins, 1997; Güler & Übeyli, 2005a).

Expectation-maximization (EM) algorithm have been introduced to the ME architecture so that the learning process is decoupled in a manner that fits well with the modular structure (Chen et al., 1999; Hong & Harris, 2002; Jordan & Jacobs, 1994). The EM algorithm can be extended to provide an effective training mechanism for the MEs based on a Gaussian probability assumption. Though originally the model structure is predetermined

* Tel.: +90 312 2924080; fax: +90 312 2924091.
E-mail address: edubeyli@etu.edu.tr

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.02.006

and the training algorithm is based on the Gaussian probability assumption for each expert model output, the ME framework is a powerful concept that can be extended to a wide variety of applications including medical diagnostic decision support system applications due to numerous inherent advantages such as (i) a global model can be decomposed into a set of simple local models, from which controller design is straightforward. Each model can represent a different data source with an associated state estimator/predictor. In this case the ME system can be viewed as a data fusion algorithm. (ii) The local models operate independently but provide output correlated information that can be strongly correlated with each other, so that the overall system performance can be enhanced in terms of reliability or fault tolerance. (iii) The global output of the ME system is derived as a convex combination of the outputs from a set of N experts, in which the overall system predictive performance is generally superior to any of the individual experts (Chen et al., 1999; Hong & Harris, 2002; Jordan & Jacobs, 1994; Mangiameli & West, 1999).

The electroencephalogram (EEG), a highly complex signal, is one of the most common sources of information used to study brain function and neurological disorders. Electroencephalography provides key information for the interictal definition of epilepsies and localizing ictal (epileptic seizure) onset (Agarwal, Gotman, Flanagan, & Rosenblatt, 1998; Adeli, Zhou, & Dadmehr, 2003; Hazarika, Chen, Tsoi, & Sergejew, 1997; Özdamar & Kalaycı, 1998). Large amounts of data are generated by EEG monitoring systems for epilepsy, and their complete visual analysis is not routinely possible. Computers have long been proposed to solve this problem and thus, automated systems to recognize EEG epileptiform activity have been under study for several years (Gotman & Wang, 1991, 1992; Dingle, Jones, Carroll, & Fright, 1993). There is a strong demand for the development of such automated devices, due to the increased use of prolonged and long-term video EEG recordings for proper evaluation and treatment of epilepsy and prevention of the possibility of the analyst missing (or misreading) information.

Historically, detection techniques have reflected the technology of their era and thus improvements in detection techniques have occurred concomitantly with improvements in technology. For example, in the early years, analog approaches used band-pass filters with threshold detectors, and in the digital era, computer-based devices used software techniques as the primary technology for the detection of epileptiform activity. As digital techniques have developed over the years vastly different approaches have been employed, ranging from statistical methods to signal processing methods (Guedes de Oliviera, Queiroz, & Lopes Da Silva, 1983; Arakawa, Fender, Harashima, Miyakawa, & Saitoh, 1986; Glover, Raghaven, Ktonas, & Frost, 1989; Webber, Litt, Lesser, Fisher, & Bankman, 1993). Most of these approaches have used one or more parameters of the EEG transient events, such as duration, amplitude, sharpness, and different digital processing methods. However, all of these methods required the definition of a set of features which have to adequately describe the waveforms to be detected. Two different approaches can be used for representing the EEG signals to automated EEG event detection systems. The first approach is to use raw EEG signal as input to the automated detection systems. In this approach, raw EEG amplitudes from an analysis window are fed into the automated detection system in a time series format as input (Özdamar & Kalaycı, 1998). The second approach uses pre-extracted parameters as input data to the automated detection systems. Such an approach has been used by many researchers with fairly successful results (Güler & Übeyli, 2005b; Adeli et al., 2003; Hazarika et al., 1997; Gabor & Seyal, 1992; Webber, Litt, Wilson, & Lesser, 1994). Generally three types of parameters are used; (1) waveform parameters, such as amplitude, width, slope, and sharpness, which are extracted from epileptic discharges, (2) context parameters, such as EEG variance, and baseline crossings, which are extracted from the EEG surrounding the epileptiform discharges, and (3) transformation parameters, such as autocorrelation coefficients, Fourier coefficients, autoregressive coefficients, and wavelet coefficients, which are computed by application of the mathematical transformations to the EEG signals. Multiresolution wavelet representation provides a simple hierarchical framework for interpreting the signal information. At different resolutions, the signal details may characterize different physiological processes in the signal source. The ability of wavelets to extract and localize those specific transient patterns makes them a natural complement to the application of the automated detection systems (Adeli et al., 2003; Hazarika et al., 1997).

In the present study, in order to classify the EEG signals ME network structure was developed by feature vectors extracted from the EEG signals with the usage of wavelet transform (WT). In the configuration of ME, three local experts and a gating network were used, which were in the form of multilayer perceptron neural networks (MLPNNs), since there were three classes of the EEG signals (EEG signals recorded from healthy volunteers with eyes open, epilepsy patients in the epileptogenic zone during a seizure-free interval, and epilepsy patients during epileptic seizures). Significant improvement in accuracy was achieved by using the ME network structure compared to the stand-alone MLPNN.

The outline of this study is as follows. In Section 2, the sets of the EEG signals used in the present study are described. In Section 3, spectral analysis of signals using discrete wavelet transform (DWT) is explained in order to extract features characterizing the behavior of the signal under study. In Section 4, description of neural network models including MLPNN and ME architecture used in this study is presented. In Section 5, the application results of ME networks to the EEG signals are presented. Finally, in Section 6 the study is concluded.

2. Data selection

The data described in reference (Andrzejak et al., 2001), which is publicly available, was used. In this section, only a short description is presented and refer to reference (Andrzejak et al., 2001) for further details. The complete dataset consists of five sets (denoted A–E), each containing 100 single-channel EEG signals of 23.6 s. Each signal has been selected after visual inspection for artifacts and has passed a weak stationarity criterion. Sets A and B have been taken from surface EEG recordings of five healthy volunteers with eyes open and closed, respectively. Signals in two sets have been measured in seizure-free intervals from five patients in the epileptogenic zone (D) and from the hippocampal formation of the opposite hemisphere of the brain (C). Set E contains seizure activity, selected from all recording sites exhibiting ictal activity. Sets A and B have been recorded extracranially, whereas sets C, D, and E have been recorded intracranially. In the applications, performance degraded for a more detailed classification which further dissociated between sets A (healthy volunteer, eyes open) and B (healthy volunteer, eyes closed), and sets D (epileptogenic zone) and C (hippocampal formation of opposite hemisphere). Therefore, in the present study three dataset (A, D, E) of the complete dataset were classified.

3. Spectral analysis using discrete wavelet transform

Predicting the onset of epileptic seizure is an important and difficult biomedical problem, which has attracted substantial attention of the intelligent computing community over the past two decades (Agarwal et al., 1998; Adeli et al., 2003; Hazarika et al., 1997; Özdamar & Kalaycı, 1998). The ME network structure combined with signal wavelet decomposition was applied to the problem. The WT provides very general techniques which can be applied to many tasks in signal processing. One very important application is the ability to compute and manipulate data in compressed parameters which are often called features (Daubechies, 1990; Soltani, 2002; Unser & Aldroubi, 1996; Akay, 1997). Thus, the EEG signal, consisting of many data points, can be compressed into a few parameters (Adeli et al., 2003; Hazarika et al., 1997). These parameters characterize the behavior of the EEG signal. This feature of using a smaller number of parameters to represent the EEG signal is particularly important for recognition and diagnostic purposes. The WT can be thought of as an extension of the classic Fourier transform, except that, instead of working on a single scale (time or frequency), it works on a multi-scale basis. This multi-scale feature of the WT allows the decomposition of a signal into a number of scales, each scale representing a particular coarseness of the signal under study. The procedure of multiresolution decomposition of a signal $x[n]$ is schematically shown in Fig. 1. Each stage of this scheme consists of two digital filters and two downsamplers by 2. The first filter, $g[\cdot]$ is the discrete mother wavelet, high-pass in nature, and the second, $h[\cdot]$ is its mirror version, low-pass in nature. The downsampled outputs of first high-pass and low-pass filters provide the detail, D_1 and the approximation, A_1, respectively. The first approximation, A_1 is further decomposed and this process is continued as shown in Fig. 1.

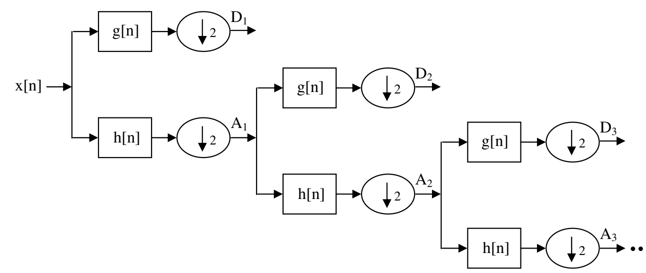

Fig. 1. Sub-band decomposition of discrete wavelet transform implementation; $g[n]$ is the high-pass filter, $h[n]$ is the low-pass filter.

All wavelet transforms can be specified in terms of a low-pass filter h, which satisfies the standard quadrature mirror filter condition:

$$H(z)H(z^{-1}) + H(-z)H(-z^{-1}) = 1, \tag{1}$$

where $H(z)$ denotes the z-transform of the filter h. Its complementary high-pass filter can be defined as

$$G(z) = zH(-z^{-1}). \tag{2}$$

A sequence of filters with increasing length (indexed by i) can be obtained.

$$\begin{aligned} H_{i+1}(z) &= H(z^{2^i})H_i(z), \\ G_{i+1}(z) &= G(z^{2^i})H_i(z), \quad i = 0, \ldots, I-1 \end{aligned} \tag{3}$$

with the initial condition $H_0(z) = 1$. It is expressed as a two-scale relation in time domain

$$\begin{aligned} h_{i+1}(k) &= [h]_{\uparrow 2^i} * h_i(k), \\ g_{i+1}(k) &= [g]_{\uparrow 2^i} * h_i(k), \end{aligned} \tag{4}$$

where the subscript $[\cdot]_{\uparrow m}$ indicates the up-sampling by a factor of m and k is the equally sampled discrete time.

The normalized wavelet and scale basis functions $\varphi_{i,l}(k)$, $\psi_{i,l}(k)$ can be defined as

$$\begin{aligned} \varphi_{i,l}(k) &= 2^{i/2}h_i(k - 2^i l), \\ \psi_{i,l}(k) &= 2^{i/2}g_i(k - 2^i l), \end{aligned} \tag{5}$$

where the factor $2^{i/2}$ is an inner product normalization, i and l are the scale parameter and the translation parameter, respectively. The DWT decomposition can be described as

$$\begin{aligned} a_{(i)}(l) &= x(k) * \varphi_{i,l}(k), \\ d_{(i)}(l) &= x(k) * \psi_{i,l}(k), \end{aligned} \tag{6}$$

where $a_{(i)}(l)$ and $d_i(l)$ are the approximation coefficients and the detail coefficients at resolution i, respectively (Daubechies, 1990; Soltani, 2002; Unser & Aldroubi, 1996; Akay, 1997).

4. Description of neural network models

4.1. Multilayer perceptron neural network

Artificial neural networks (ANNs) are computational tools utilizing a combination of many elementary processing units (cell). Each unit is connected to a number of network units to process information by transfer function. The relationship between the input and the output determine network behavior. Contrary to conventional computing methods, ANNs are trained to produce desired the input–output relationship. The principal applications of ANNs have been in the area of pattern recognition. The pattern is turned into a feature vector used as ANN input, and the output is interpreted as identifying the input to be a member of one of a number of classes of possible inputs. An important quality of neural networks (referred to as generalization) is that when they are correctly trained, neural networks can appropriately process data that have not been used for training. The MLPNNs are the most commonly used neural network arhitectures since they have features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation. One major property of these networks is their ability to find nonlinear surfaces separating the underlaying patterns which is generally considered as an improvement on conventional methods. A MLPNN consists of (i) an input layer with neurons representing input variables to the problem, (ii) an output layer with neurons representing the dependent variables (what is being modeled), and (iii) one or more hidden layers containing neurons to help capture the nonlinearity in the data. The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks (Haykin, 1994; Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000).

4.2. Mixture of experts

As illustrated in Fig. 2, the ME architecture is composed of a gating network and several expert networks. The gating network receives the vector **x** as input and produces scalar outputs that are partition of unity at each point in the input space. Each expert network produces an output vector for an input vector. The gating network provides linear combination coefficients as veridical probabilities for expert networks and, therefore, the final output of the ME architecture is a convex weighted sum of all the output vectors produced by expert networks. Suppose that there are N expert networks in the ME architecture. All the expert networks are linear with a single output nonlinearity that is also referred to as "generalized linear". The ith expert network produces its output $\mathbf{o}_i(\mathbf{x})$ as a generalized linear function of the input **x** (Jacobs et al., 1991; Chen et al., 1999; Hong & Harris, 2002):

$$\mathbf{o}_i(\mathbf{x}) = f(\mathbf{W}_i\mathbf{x}), \tag{7}$$

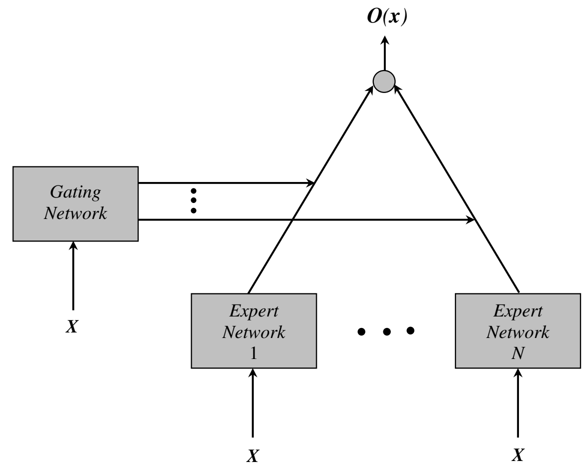

Fig. 2. The architecture of mixture of experts.

where $\mathbf{W}_i$ is a weight matrix and $f(\cdot)$ is a fixed continuous nonlinearity. The gating network is also generalized linear function, and its ith output, $g(\mathbf{x},\mathbf{v}_i)$, is the multinomial logit or softmax function of intermediate variables ξ_i:

$$g(\mathbf{x},\mathbf{v}_i) = \frac{e^{\xi_i}}{\sum_{k=1}^{N} e^{\xi_k}}, \tag{8}$$

where $\xi_i = \mathbf{v}_i^T\mathbf{x}$ and $\mathbf{v}_i$ is a weight vector. The overall output $\mathbf{o}(\mathbf{x})$ of the ME architecture is

$$\mathbf{o}(\mathbf{x}) = \sum_{k=1}^{N} g(\mathbf{x},\mathbf{v}_k)\mathbf{o}_k(\mathbf{x}). \tag{9}$$

The ME architecture can be given a probabilistic interpretation. Based on the probabilistic model, learning in the ME architecture is treated as a maximum likelihood problem. Jordan and Jacobs (1994) have proposed an EM algorithm for adjusting the parameters of the architecture. In this framework a number of relatively small expert networks can be used together with a gating network designed to divide the global classification task into simpler subtasks (Fig. 2) (Jacobs et al., 1991; Chen et al., 1999; Hong & Harris, 2002; Jordan & Jacobs, 1994). In the present study, both the gating and expert networks were MLPNNs consisting of neurons arranged in contiguous layers. This configuration occured on the theory that MLPNN has features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation. The ME network structure proposed for the classification of EEG signals were implemented by using the MATLAB software package.

5. Experimental results

5.1. Feature extraction using discrete wavelet transform

The EEG signals can be considered as a superposition of different structures occuring on different time scales at different times. One purpose of wavelet analysis is to separate and sort these underlying structures of different time scales. It is known that the WT is better suited to analyzing nonstationary signals, since it is well localized in time and frequency. The property of time and frequency localization is known as compact support and is one of the most attractive features of the WT. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrow at high frequencies, thus leading to an optimal time–frequency resolution in all frequency ranges. Therefore, spectral analysis of the EEG signals was performed using the DWT as described in Section 3.

Selection of appropriate wavelet and the number of decomposition levels is very important in analysis of signals using the WT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients. In the present study, the number of decomposition levels was chosen to be 4. Thus, the EEG signals were decomposed into the details $D_1 - D_4$ and one final approximation, A_4. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 2 made it more suitable to detect changes of the EEG signals. Therefore, the wavelet coefficients were computed using the Daubechies wavelet of order 2 in the present study. The wavelet coefficients were computed using the MATLAB software package.

Selection of the ANN inputs is the most important component of designing the neural network based on pattern classification since even the best classifier will perform poorly if the inputs are not selected well. Input selection has two meanings: (1) which components of a pattern, or (2) which set of inputs best represent a given pattern. The computed discrete wavelet coefficients provide a compact representation that shows the energy distribution of the signal in time and frequency. Therefore, the computed detail and approximation wavelet coefficients of the EEG signals were used as the feature vectors representing the signals. A rectangular window, which was formed by 256 discrete data, was selected so that it contained a single EEG segment. For each EEG segment, the detail wavelet coefficients (d^k, $k = 1, 2, 3, 4$) at the first, second, third and fourth levels ($129 + 66 + 34 + 18$ coefficients) and the approximation wavelet coefficients (a^4) at the fourth level (18 coefficients) were computed. Then 265 wavelet coefficients were obtained for each EEG segment. In order to reduce the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients were used. The following statistical features were used to represent the time–frequency distribution of the EEG signals:

1. Mean of the absolute values of the coefficients in each sub-band.
2. Maximum of the absolute values of the coefficients in each sub-band.
3. Average power of the wavelet coefficients in each sub-band.
4. Standard deviation of the coefficients in each sub-band.
5. Ratio of the absolute mean values of adjacent sub-bands.
6. Distribution distortion of the coefficients in each sub-band.

Features 1–3 represent the frequency distribution of the signal and the features 4–6 the amount of changes in frequency distribution. These feature vectors, calculated for the frequency bands $D_1 - D_4$ and A_4, were used for classification of the EEG segments.

5.2. Application of mixture of experts to EEG signals

The ME architecture used for the classification of EEG segments is shown in Fig. 2. Since three-group classification exclusively was investigated, the ME was configured with three local experts and a gating network which were in the form of MLPNNs. ANN architectures are derived by trial and error and the complexity of the neural network is characterized by the number of hidden layers. There is no general rule for selection of appropriate number of hidden layers. A neural network with a small number of neurons may not be sufficiently powerful to model a complex function. On the other hand, a neural network with too many neurons may lead to overfitting the training sets and lose its ability to generalize which is the main desired characteristic of a neural network. The most popular approach to finding the optimal number of hidden layers is by trial and error. The architecture studies confirmed that for the EEG segments, a minimal network has better generalization properties and results in higher classification accuracy. Apart from one hidden layered network architecture, two, four, and six hidden layered network architectures were also tried. Total classification accuracy obtained for each network architecture is presented in Table 1. From Table 1, one can see that for this data, network architectures with one hidden layer were superior to models with two, four and six hidden layers. The most suitable network configuration found was 10 neurons for the hidden layer and the number of output was 3. Samples with target outputs set A (EEG signals recorded from healthy volunteers with eyes open), set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free

Table 1
Total classification accuracy obtained for each network architecture

Network architecture	Total classification accuracy
One hidden layered	93.17%
Two hidden layered	92.50%
Four hidden layered	91.75%
Six hidden layered	90.88%

interval), and set E (EEG signals recorded from epilepsy patients during epileptic seizures) were given the binary target values of $(0,0,1)$, $(0,1,0)$, and $(1,0,0)$, respectively.

The adequate functioning of neural networks depends on the sizes of the training set and test set. In the ME, the 2400 vectors (800 vectors from each class) were used for training and the 2400 vectors (800 vectors from each class) were used for testing. A practical way to find a point of better generalization is to use a small percentage (around 20%) of the training set for cross validation. For obtaining a better network generalization 480 vectors (160 vectors from each class) of training set, which were selected randomly, were used as cross validation set. Beside this, in order to enhance the generalization capability of the ME, the training and the test sets were formed by data obtained from different subjects. For all of the segments, waveform variations were observed among the vectors belonging to the same class.

The waveforms of three different EEG segments classified in the present study are shown in Fig. 3a–c. For the three diagnostic classes (sets A, D, and E) training and test sets were formed by 4800 vectors (1600 vectors from each class) of 29 dimensions (extracted feature vectors). The detail wavelet coefficients at the first decomposition level of the three types of EEG segments are given in Fig. 4a–c, respectively. It can be noted that the detail wavelet coefficients of the three types of EEG segments are different from each other. In order to extract features, the wavelet coefficients corresponding to the $D_1 - D_4$ and A_4 frequency bands of the three types of EEG segments were computed.

The training holds the key to an accurate solution, so the criterion to stop training must be very well described. When the network is trained too much, the network memorizes the training patterns and does not generalize well. Cross validation is a highly recommended criterion for stopping the training of a network. When the error in the cross validation increases, the training should be stopped because the point of best generalization has been reached. Training of the ME was done in 800 epochs since the cross validation errors began to rise at 800 epochs. Owing to the values of mean square error (MSE) converged to small constants approximately zero in 800 epochs, training of the ME was determined to be successful. However, the stand-alone MLPNN (wavelet coefficients used as inputs) trained with the backpropagation algorithm had a slow convergence and MSE converged to a small constant of approximately zero in 4500 epochs. The backpropagation algorithm searched global optimal solution for the classification problem so that the number of epochs required for convergence increased. However, in the ME the

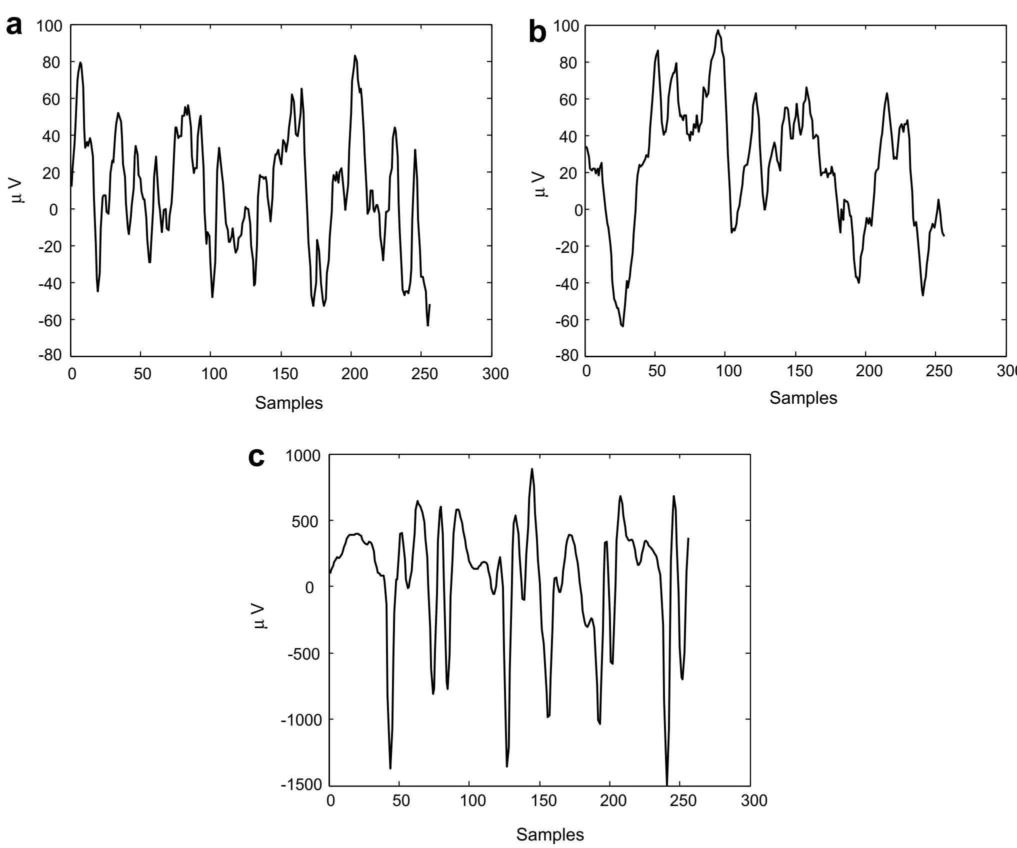

Fig. 3. Waveforms of the EEG segments: (a) set A (EEG signals recorded from healthy volunteers with eyes open), (b) set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval) and (c) set E (EEG signals recorded from epilepsy patients during epileptic seizures).

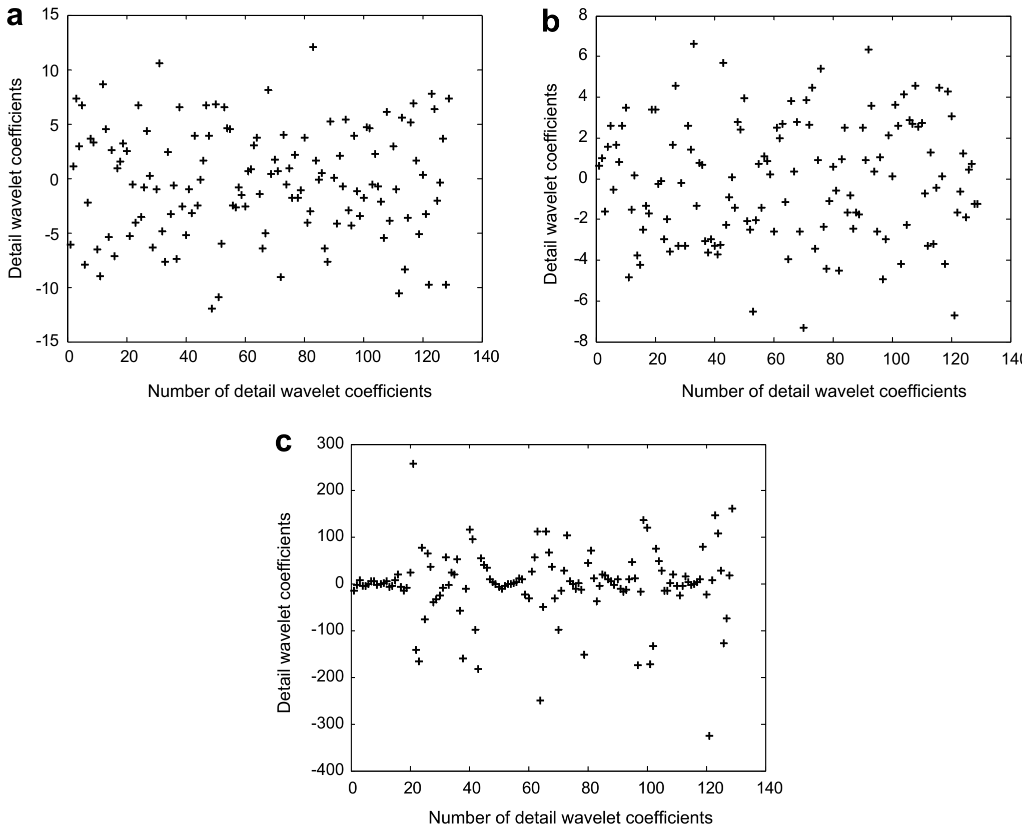

Fig. 4. The detail wavelet coefficients at the first decomposition level of the EEG segments: (a) set A (EEG signals recorded from healthy volunteers with eyes open), (b) set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval) and (c) set E (EEG signals recorded from epilepsy patients during epileptic seizures).

classification problem was divided into simpler problems and then each solution was combined. In addition to this, the training algorithm of the ME is a general technique for maximum likelihood estimation that fits well with the modular structure and enables a significant speed up over the backpropagation algorithm. Thus, the convergence rate of ME presented in this study was found to be higher than that of the stand-alone MLPNN.

In classification, the aim is to assign the input patterns to one of several classes, usually represented by outputs restricted to lie in the range from 0 to 1, so that they represent the probability of class membership. While the classification is carried out, a specific pattern is assigned to a specific class according to the characteristic features selected for it. In this application, there were three classes: set A (EEG signals recorded from healthy volunteers with eyes open), set D (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval), and set E (EEG signals recorded from epilepsy patients during epileptic seizures). Classification results of the ME were displayed by a confusion matrix. The confusion matrix showing the classification results of the ME is given below.

According to the confusion matrix, 30 healthy segments were classified incorrectly by the ME as seizure-free epileptogenic zone segments, 18 healthy segments were classified as epileptic seizure segments, 41 seizure-free epileptogenic zone segments were classified as healthy segments, 19 seizure-free epileptogenic zone segments were classified as epileptic seizure segments, 24 epileptic seizure segments were classified as healthy segments, 32 epileptic seizure segments were classified as seizure-free epileptogenic zone segments.

Confusion matrix

Output/desired	Result (set A – healthy segment)	Result (set D – seizure-free epileptogenic zone segment)	Result (set E – epileptic seizure segment)
Result (set A - healthy segment)	752	41	24
Result (set D – seizure-free epileptogenic zone segment)	30	740	32
Result (set E – epileptic seizure segment)	18	19	744

The test performance of the ME was determined by the computation of the following statistical parameters:

Specificity: number of correct classified healthy segments/number of total healthy segments.
Sensitivity (seizure-free epileptogenic zone segments): number of correct classified seizure-free epileptogenic zone segments/number of total seizure-free epileptogenic zone segments.
Sensitivity (epileptic seizure segments): number of correct classified epileptic seizure segments/number of total epileptic seizure segments.
Total classification accuracy: number of correct classified segments/number of total segments.

The values of these statistical parameters are given in Table 2. As it is seen from Table 2, the ME classified healthy segments, seizure-free epileptogenic zone segments, epileptic seizure segments with the accuracy of 94.00%, 92.50%, 93.00%, respectively. The healthy segments, seizure-free epileptogenic zone segments, epileptic seizure segments were classified with the accuracy of 93.17%. The total classification accuracy of the stand-alone MLPNN (trained with the backpropagation algorithm, wavelet coefficients used as inputs) was 84.83%. Thus, the accuracy rates of the ME network structure presented for this application were found to be higher than that of the stand-alone MLPNN.

The performance of a test can be evaluated by plotting a ROC curve for the test. For a given result obtained by a classifier system, four possible alternatives exist that describe the nature of the result: (i) true positive (TP), (ii) false positive (FP), (iii) true negative (TN), and (iv) false negative (FN) (Zweig & Campbell, 1993). In this study, a TP decision occured when the positive detection of the ME coincided with a positive detection of the physician. A FP decision occured when the ME made a positive detection that did not agree with the physician. A TN decision occured when both the ME and the physician suggested the absence of a positive detection. A FN decision occured when the ME made a negative detection that did not agree with the physician. A good test is one for which sensitivity rises rapidly and 1-specificity hardly increases at all until sensitivity becomes high. ROC curves which are shown in Fig. 5 represent performances of the stand-alone MLPNN (trained with the backpropagation algorithm, wavelet coefficients used as inputs) and the ME network structure on the EEG segments test file. Fig. 5 shows that the performance of the ME is higher than that of the stand-alone MLPNN.

Table 2
The values of statistical parameters of the ME used for classification of the EEG segments

Statistical parameters	Values (%)
Specificity	94.00
Sensitivity (seizure-free epileptogenic zone segments)	92.50
Sensitivity (epileptic seizure segments)	93.00
Total classification accuracy	93.17

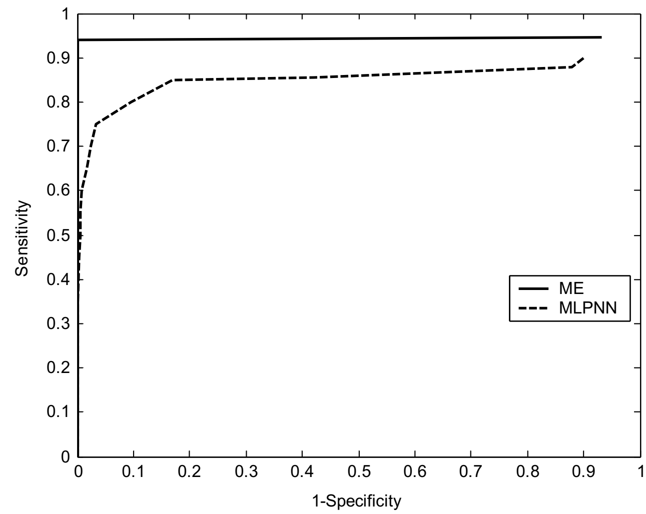

Fig. 5. ROC curves of the stand-alone MLPNN and ME network structure used for the classification of EEG segments.

6. Conclusion

This paper presented the use of ME network structures to improve classification accuracy of the EEG signals since the overall structure predictive performance is generally superior to any of the individual experts. Toward achieving the classification of EEG signals three local experts and a gating network, which were in the form of MLPNNs, were used in the configuration of ME architecture. EM algorithm was used for training the ME network so that the learning process is decoupled in a manner that fits well with the modular structure. The ME used for the classification of EEG signals was trained, cross validated and tested with the extracted features using DWT of the EEG signals. The classification results, the values of statistical parameters, and ROC curves were used for evaluating performances of the classifiers. The accuracy rates achieved by the ME network structure presented for the classification of the EEG signals were found to be higher than that of the stand-alone MLPNN trained with the backpropagation algorithm (wavelet coefficients used as inputs). Based on the accuracies of the ME network structure, it can be concluded that the ME network structures trained by the EM algorithm are encouraging methods for classification task of the EEG signals.

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.

Agarwal, R., Gotman, J., Flanagan, D., & Rosenblatt, B. (1998). Automatic EEG analysis during long-term monitoring in the ICU. *Electroencephalography and Clinical Neurophysiology, 107*(1), 44–58.

Akay, M. (1997). Wavelet applications in medicine. *IEEE Spectrum, 34*(5), 50–56.

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 seires of brain electrical activity: dependence on recording region and brain state. *Physical Review E, 64*, 061907.

Arakawa, K., Fender, D., Harashima, H., Miyakawa, H., & Saitoh, Y. (1986). Separation of a nonstationary component from the EEG by a nonlinear digital filter. *IEEE Transactions on Biomedical Engineering, 33*(7), 724–726.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods, 43*(1), 3–31.

Chaudhuri, B. B., & Bhattacharya, U. (2000). Efficient training and improved performance of multilayer perceptron in pattern classification. *Neurocomputing, 34*, 11–27.

Chen, K., Xu, L., & Chi, H. (1999). Improved learning algorithms for mixture of experts in multiclass classification. *Neural Networks, 12*(9), 1229–1252.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory, 36*(5), 961–1005.

Dingle, A. A., Jones, R. D., Carroll, G. J., & Fright, W. R. (1993). A multistage system to detect epileptiform activity in the EEG. *IEEE Transactions on Biomedical Engineering, 40*(12), 1260–1268.

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 transients in the EEG: elimination of false positives. *Transactions on Biomedical Engineering, 36*(5), 519–527.

Gotman, J., & Wang, L. Y. (1991). State-dependent spike detection: concepts and preliminary results. *Electroencephalography and Clinical Neurophysiology, 79*(1), 11–19.

Gotman, J., & Wang, L. Y. (1992). State dependent spike detection: validation. *Electroencephalography and Clinical Neurophysiology, 83*(1), 12–18.

Guedes de Oliviera, P., Queiroz, C., & Lopes Da Silva, F. (1983). Spike detection based on a pattern recognition approach using a microcomputer. *Electroencephalography and Clinical Neurophysiology, 56*(1), 97–103.

Güler, İ., & Übeyli, E. D. (2005a). A mixture of experts network structure for modelling Doppler ultrasound blood flow signals. *Computers in Biology and Medicine, 35*(7), 565–582.

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

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan.

Hazarika, N., Chen, J. Z., Tsoi, A. C., & Sergejew, A. (1997). Classification of EEG signals using the wavelet transform. *Signal Processing, 59*(1), 61–72.

Hong, X., & Harris, C. J. (2002). A mixture of experts network structure construction algorithm for modelling and control. *Applied Intelligence, 16*(1), 59–69.

Hu, Y. H., Palreddy, S., & Tompkins, W. J. (1997). A patient-adaptable ECG beat classifier using a mixture of experts approach. *IEEE Transactions on Biomedical Engineering, 44*(9), 891–900.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*(1), 79–87.

Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixture of experts and the EM algorithm. *Neural Computation, 6*(2), 181–214.

Mangiameli, P., & West, D. (1999). An improved neural classification network for the two-group problem. *Computers & Operations Research, 26*(5), 443–460.

Özdamar, Ö., & Kalaycı, T. (1998). Detection of spikes with artificial neural networks using raw EEG. *Computers and Biomedical Research, 31*(2), 122–142.

Soltani, S. (2002). On the use of the wavelet decomposition for time series prediction. *Neurocomputing, 48*, 267–277.

Unser, M., & Aldroubi, A. (1996). A review of wavelets in biomedical applications. *Proceedings of the IEEE, 84*(4), 626–638.

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.

Webber, W. R. S., Litt, B., Wilson, K., & Lesser, R. P. (1994). Practical detection of epileptiform discharges (EDs) in the EEG using an artificial neural network: a comparison of raw and parameterized EEG data. *Electroencephalography and Clinical Neurophysiology, 91*(3), 194–204.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry, 39*(4), 561–577.