EPILEPTIC SEIZURE DETECTION USING A NEURAL NETWORK ENSEMBLE METHOD AND WAVELET TRANSFORM

Reza Ebrahimpour[*], *Kioumars Babakhani*[†], *Seyed Ali Asghar Abbaszadeh Arani*[‡], *Saeed Masoudnia*[§]

Abstract: This paper presents a new method to automate the process of epileptic seizure detection in electroencephalogram (EEG) signals using wavelet transform and an improved version of negative correlation learning (NCL) algorithm. An improved version of NCL is proposed by incorporating the capability of gating network, as a dynamic combining part of the mixture of experts (ME), into the combining outputs of base experts which are trained using negative correlation learning algorithm. The NCL training algorithm encourages the base experts to learn different parts or aspects of data set and the gating network provides the local competence of these base experts. Three types of normal (recorded from five healthy persons with eyes open), seizure-free (recorded from epileptogenic zoon of five patients) and epileptic EEG signals were decomposed into wavelet coefficients using discrete wavelet transform. Then the statistical features of the wavelet coefficients were computed representing them into the classifiers. Experimental results show that our proposed method classifies normal, seizure-free and epileptic EEG signals with the accuracy of 96.92% which is significantly better than previous combining methods.

Key words: *Epileptic seizure, electroencephalogram (EEG) signals, discrete wavelet transform, mixture of experts, negatively correlated learning*

Received: February 28, 2012
Revised and accepted: June 16, 2012

[*]Reza Ebrahimpour
School of Cognitive Sciences (SCS), Institute for Research in Fundamental Sciences (IPM) Niavaran, Tehran, Iran

[†]Kioumars Babakhani
School of Mathematics, Statistics and Computer Science, University of Tehran, Iran, E-mail: kbabakhani@ut.ac.ir

[‡]Seyed Ali Asghar Abbaszadeh Arani
Brain & Intelligent Systems Research Laboratory, Department of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University, Tehran, Iran

[§]Saeed Masoudnia
Islamic Azad University, Central Branch and Young Researchers Club, Tehran, Iran

1. Introduction

The electroencephalogram (EEG) is the non-invasive record of the electrical activity of the neurons in the brain. The EEG is one of the most common methods used to analyze the brain functions and neurological disorders, such as epilepsy. Epilepsy is characterized by sudden recurrent and transient disturbances of brain functions termed "seizure". An epileptic seizure is a sudden synchronous and repetitive discharge of brain cells with symptoms depending on the location within the brain of the seizure onset, and the spread of the seizure.

About one percent of the people in the world suffer from epilepsy and almost a third of epileptic seizures cannot be controlled by medication [1]. In these cases, one option could be surgery to remove the epileptic part of the brain. Newer methods where parts of the brain are electrically stimulated to avoid the onset of seizure are being developed. Automatic detection of epileptic seizures forms an integral part of such methods. Therefore, there is a strong demand to automate this process.

The typical procedure for epilepsy seizure detection is based on brain activity monitoring through EEG data. This usually involves identifying sharp repetitive waveforms or rhythmic patterns in the EEG data that indicate seizure onset. Careful analysis of the EEG records can provide valuable insight and improved understanding of the mechanisms causing epileptic disorders.

For many years, EEG analysis has been mainly based on two significant characteristics extracted from EEG: frequency and amplitude [2]. These approaches, which include EEG epoch analysis, spike detection, parametric models, quantitative analysis, and spectral EEG signal analysis, assume quasi-stationary, require long recordings and present relatively high false detection rates due to the presence of typical EEG artifacts [3-7]. These methods give frequency and energy information but they do not provide temporal information about when seizure discharges begin.

Wavelet Transform (WT) was proposed in the late 1980s to address the problem of poor temporal resolution in non-stationary EEG signals. It is particularly effective for representing various aspects of non-stationary signals such as trends, discontinuities, and repeated patterns where other signal processing approaches fail or are not as effective [8]. The main advantage of the WT is that it has a varying window size, being broad at low frequencies and narrows at high frequencies, thus leading to an optimal time-frequency resolution in all frequency ranges. Discrete Wavelet Transform (DWT) developed for recognizing and quantifying spikes, sharp waves and spike-waves. Through wavelet decomposition of the EEG records, transient features are accurately captured and localized in time and frequency context [9]. Therefore, the EEG signals were decomposed into time-frequency representations using discrete wavelet transform. Wavelet coefficients were used as feature vectors identifying characteristics of the signal that were not apparent from the original time domain signal.

Automatic detection of epileptic EEG seizures has been investigated for many years. Several methods have been proposed to characterize the EEG seizures based on spectral analysis [10-12], wavelet features [13-17], chaotic features [18] such as correlation dimension [19], entropy [20, 21] and Lyapunov exponents [22, 23]. Many different classifiers, such as nearest neighbor classifier [22], decision trees [8], support vector machines (SVMs) [24, 25] and artificial neural networks (ANNs) [13, 15-17, 26], then used the extracted features to identify epileptic seizures.

Due to having non-stationary signals, poor signal to noise ratio, highly overlapped classes, small training size and high dimensional feature sets, epileptic seizure detection of EEG signals can be categorized into complex problems [27]. Combining classifiers is an approach to improve the performance in classification particularly for complex problems [28].

Combining methods have two major components, i.e., a method for creating base classifiers which are briefly called *experts* and a method for combining the outputs of the experts such that the combination improves upon the performance of the single classifier. Both theoretical and experimental studies [29] have shown that combining procedure is most effective when the experts' estimates are negatively correlated.

The Bagging and Boosting [28] are two popular algorithms that use different training sets to train individual experts in an ensemble system. Negative Correlation Learning (NCL) [30] and Mixture of Experts (ME) [31], as two of more advanced ensemble learning methods, employ special error functions to simultaneously train experts whose errors are negatively correlated. While bagging and boosting create explicitly different training sets for different experts by probabilistically changing the distribution of the original training data, NCL and ME implicitly create different training sets by encouraging different experts to learn different parts or aspects of the training data. As the explicit and implicit partitioning of a training set between experts have complementary features, some authors proposed new methods to combine their strengths in integrated approaches [32, 33].

In the present work, the strengths and limitations of the NCL and ME methods are investigated and based on the complementary features of both methods, a novel hybrid method is proposed.

In this study, EEG signals which contain three types: normal (EEG signals recorded from healthy persons with eyes open (A)), seizure-free (EEG signals recorded from epilepsy patients in the epileptogenic zone during a seizure-free interval (D)) and epileptic seizure (EEG signals recorded from epilepsy patients during epileptic seizures (E)), were classified using our proposed method. The experimental results show that the proposed ensemble method improves the classification accuracy rather than the previous combining techniques.

The rest of the paper is as follows. In Section 2, first the sets of EEG signals used in the study are briefly described and then discrete wavelet transform is illustrated as the feature extraction method. In Section 3, the training algorithms of NCL and ME are presented. In Section 4, first the strengths and weaknesses of NCL and ME are investigated and compared with each other. Then, based on the complementary features of both methods, a novel ensemble method is proposed. In Section 5, the results of the experiments are presented and compared with the results of previous methods are reported earlier. Finally, in Section 6, the study is concluded.

2. Materials and Methods

2.1 EEG data description

In this application, we studied the data set described in [34]. 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 stationary 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. All EEG signals were recorded with the same 128-channel amplifier system, at sampling rate of 173.61 Hz, and filtered using band-pass filter with settings 0.53–40 Hz.

In our study, three types of normal (set A), seizure-free (set D) and epileptic (set E) EEG signals were used for the classification. Fig. 1 shows an exemplary of raw normal, seizure-free and epileptic seizure EEG signals.

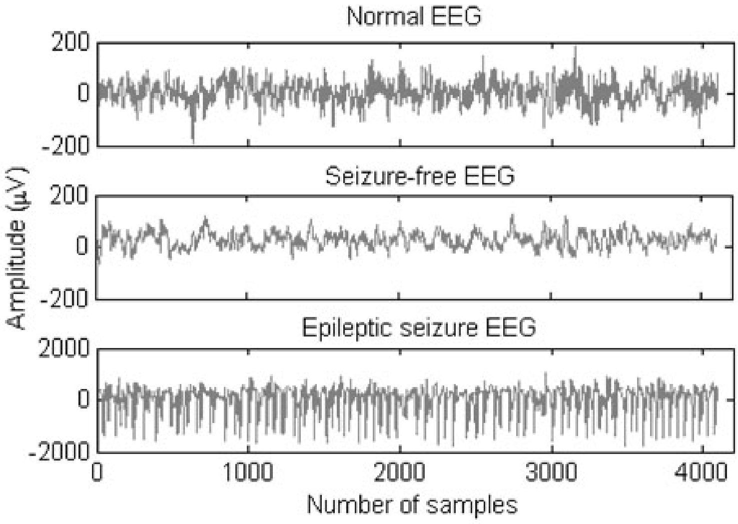


Fig. 1 *An exemplary of raw normal, seizure-free and epileptic seizure of EEG signals.*

2.2 Discrete wavelet transform

Due to the non-stationary nature of EEG signal, it should be analyzed in both time and frequency. Therefore, the wavelet transform, as a time-frequency analysis tool, is a suitable choice [6]. The wavelet transform is quite similar to the Short Time Fourier Transform (STFT) except the window is not fixed as in STFT. Wavelet decomposition overcomes the shortcomings of the classical short time Fourier transform for the analysis of non-stationary signals, permitting higher time resolution of higher frequencies, as well as temporal localization of non-stationary signals.

The Discrete Wavelet Transform (DWT) [9] is a versatile signal processing tool that finds many engineering and scientific applications. One area in which the DWT has been particularly successful is analyzing non-stationary EEG signals, such as the epileptic seizure detection, because it captures transient features and localizes them in both time and frequency content accurately.

The DWT is a representation of signal $x(t)$ using an orthonormal basis consisting of a countably infinite set of wavelets. DWT employs two functions, $\varphi(t)$ the scaling function and $\psi(t)$, the wavelet function, which are associated with low and high pass filters, respectively. Both of these functions are shifted and scaled as shown below:

$$\forall k, n, k \wedge n \in Z : \varphi_{k,n}(t) = 2^{-k/2}\varphi(2^{-k}t - n) \tag{1}$$

$$\forall k, n, k \wedge n \in Z : \psi_{k,n}(t) = 2^{-k/2}\psi(2^{-k}t - n) \tag{2}$$

The wavelet representation of a signal $x(t)$ in terms of the scaling and wavelet functions is given by:

$$x(t) = \sum_{n=-\infty}^{\infty} a_{k_0,n}\varphi_{k_0,n}(t) + \sum_{k=k_0}^{\infty} ((d_{k,n}\psi_{k,n}(t))), \tag{3}$$

where $a_{k_0,n}$ and $d_{k,n}$ are called approximation and detailed coefficients, respectively. The frequency up to which the approximation coefficients are used for representation of the signal is determined by k_0.

The decomposition of a signal into the different frequency bands as accomplished by the process detailed above is shown in Fig. 2. It is simply high and low-pass filtering of the time domain signal yielding detailed and approximation coefficients, respectively. The low pass filter's output is further subjected to the same process of high- and low-pass filtering.

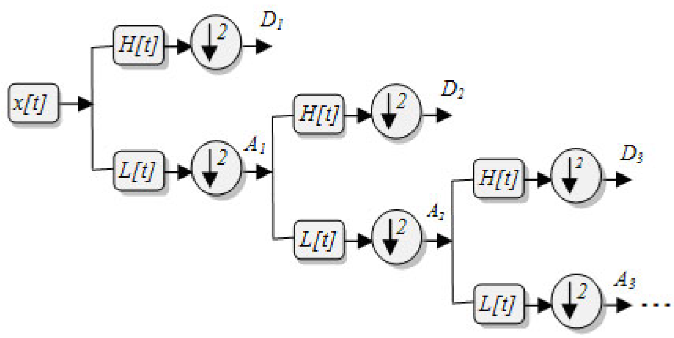


Fig. 2 *Subband decomposition of DWT implementation. H[t] and L[t] are the high and low pass filters, respectively. D_i and A_i are the detailed and approximation coefficients of level i, respectively.*

This is repeated until the number of desired decomposition is reached. The outputs of both filters are down sampled at each stage. For this reason, it is to be ensured that the sampling frequency of the signal is at least two times that of the maximum frequency to be analyzed. Selection of suitable wavelet function and the number of decomposition levels is very important in the analysis of signals using DWT. The wavelet can be chosen depending on how smooth the signal is and also on the basis of the amount of computation involved. The number of decomposition levels is chosen based on the dominant frequency components of the signal.

3. Investigated of NCL and ME methods

In this section, the NCL and ME methods are investigated and reviewed.

3.1 NCL

In neural network ensemble methods, the individual networks (or experts) are usually trained independently. This approach leads to loss of interactions among the experts during the learning process. Consequently, it may be that some of individual experts contribute little to the whole ensemble.

Liu and Yao [30] proposed the negative correlation learning (NCL) method that trains experts in the ensemble simultaneously and interactively through the correlation penalty terms in their error functions. In NCL, the error function of the i-th expert is expressed by the equation:

$$E_i = \frac{1}{2}\left(O_i - y\right)^2 + \lambda P_i \tag{4}$$

where O_i and y are the actual and desired outputs of the i-th expert, respectively. The first term in Eq. 4 is the empirical risk function of the i-th expert. The second term P_i is the correlation penalty function, which can be expressed as:

$$P_i = -(O_i - \mathrm{O}_{ens})^2 \tag{5}$$

where O_{ens} is the average of outputs of experts in the ensemble. Here, P_i can be regarded as a regularization term which provides a convenient way to balance the bias-variance-covariance trade-off [30]. This term is meant to quantify the amount of error correlation, so it can be minimized explicitly during training, which leads to negatively correlated experts. The term λ is a scaling parameter that controls the trade-off between the objective and penalty functions. The interaction and correlation among the experts of the ensemble is controlled explicitly by the value of λ. This penalty function encourages different individual experts in an ensemble to learn different parts or aspects of the training data so that the ensemble can better learn the whole training data set.

3.2 Mixture of experts

The mixture of experts (ME) method was introduced by Jacobs et al. [31] in 1991. ME is composed of a two neural networks (NNs) model: a number of separate NNs called *experts* and a trainable combiner called *gating network.* During a competitive learning process, gating network learns for each training case to assign a prior probability for each expert according to its performance on various regions of input space. Jacobs et al. proposed making experts local in different distributions of data space; as a result, the increased diversity among the experts led to improvements in the performance of this method.

The gating network is used to complete a system of competing local experts. The learning rule for the gating network attempts to maximize the likelihood of the training set by assuming a Gaussian mixture model in which each expert is responsible for one component of the mixture.

The ME method has special characteristics that distinguish it from the other combining methods. This method differs from the others due to its dynamic combination method. In the literature on combining methods, ME refers to the methods in which complex problems based on a "divide and conquer" approach are partitioned into a set of simpler subproblems and are distributed among the experts. In this method, instead of assigning a set of fixed combinational weights to the experts, as described previously, an extra gating network is used to compute these weights dynamically from the inputs. For further details on the implementation of this method, please refer to [35].

4. Proposed Hybrid Ensemble Method

In this section, first the properties of NCL and ME are investigated and compared. Then, based on the similar ensemble structures and strategies used in both the NCL and ME methods and due to their complementary features, an improved hybrid ensemble method is proposed.

4.1 NCL versus ME

In this part, we compare the features of ME and NCL, discussing their advantages and disadvantages. First, the similar features of the two methods are discussed. Both of these ensemble algorithms train experts simultaneously and interactively. As mentioned before, the different and unique error functions of the two methods have specific properties that encourage the experts to learn different parts or aspects of the training data, so that the ensemble can learn the entire training data set efficiently. By implicitly assigning different distributions of data space to different experts, these two methods produce biased individual experts with negatively correlated estimations.

Nevertheless, there are some differences between the ME and NCL methods that arise from their specific characteristics in comparison with other ensemble algorithms. One of the advantages of ME over other combining methods is its distinct technique for combining the outputs of the base experts. ME uses a trainable combiner that, according to the input x, dynamically selects the best expert(s) and combines its/their outputs to create the final output. The combining function of ME includes a dynamic weighted average in which the local competences of the experts with respect to the input are estimated by the weights produced by the gating network. The outputs of all experts responsible for input x are then fused.

As mentioned before, combining systems have two major components. Regarding the first component, the creation of individual experts, based on a considered comparison, NCL has the better efficiency. Its superiority comes from its use of a regularization term that provides a convenient way to balance the bias-variance-covariance trade-off and thus improves the generalization ability, whereas ME does not include such control over the trade-off. In contrast, ME provides a better approach for the second component of combining systems, the combination of base experts.

As it is clear from the analysis of the features of both methods and their advantages and disadvantages, the two methods have complementary features. In the next part, we present a proposed approach that attempts to combine the features of both methods.

4.2 Using a gating network to combine NCL experts

As described earlier, NCL shows better efficiency in creating individual experts, whereas the combining functionality of ME provides higher performance in the combination of base experts. Based on this idea, our proposed hybrid combining system consists of two stages. In the first stage, the base experts are trained using the NCL training algorithm; in the second stage, the gating network, i.e., the combining algorithm in ME, is employed to combine the base NCL experts. In one point of view, this proposed method can be regarded as an improvement on the combination method of NCL because the special error function of the NCL training algorithm encourages each expert to learn different parts or aspects of the training data. Thus, the local competence of the experts should be considered in the combining approach. The gating network, as the combining part of ME, provides a way to support this needed functionality for combining the NCL experts. In the second step, after training the NCL experts, a gating network is employed to model the local competence of the experts. Therefore, we call this proposed method Gated NCL (G-NCL). To implement this idea, a gating network should be trained on the targets that can be used to measure the local efficiency of each expert for different parts of the training data. Here, we suggested h_{G-NCL}, similarly to ME, as the proportional measure of competence for each expert:

$$h_{G\text{-}NCL,i} = \frac{\exp(-\frac{1}{2}(y - O_i)^2)}{\sum\limits_{j=1}^{L} \exp(-\frac{1}{2}(y - O_j)^2)}. \tag{6}$$

If this measure is used as the target vector to train the gating network, the local competence of each expert can be estimated by the outputs of the gating network. So considering the h_{G-NCL} measure, the modified error function of the gating network can be expressed as:

$$E_{G\text{-}NCL} = \frac{1}{2}(h_{G\text{-}NCL} - O_g)^2 \tag{7}$$

where O_g is the output of the MLP layer of the gating network. Similar to conventional ME method, the gate process is composed of two layers: the first layer is an

MLP network, and the second layer is a softmax nonlinear operator, then applies the softmax function to obtain:

$$g_i = \frac{\exp(O_{g,i})}{\sum_{j=1}^{L} \exp(O_{g,j})} \quad i = 1, ..., L \tag{8}$$

where L is the number of expert and $O_{g,i}$ is i-th value of the gate output. Here, the g_i values are nonnegative and sum to unity, and they can be interpreted as estimates of the prior probability that expert i can generate the desired output y.

To implement the MLP training algorithm in a gating network based on the error function given by Eq. 6, the weights are learned using the error BP algorithm according to the following rules:

$$\Delta w_{yg} = \eta_g (h_{G\text{-}NCL} - O_g)(O_g(1 - O_g))O_{hg}^T \tag{9}$$

$$\Delta w_{hg} = \eta_g w_{yg}^T (h_{G\text{-}NCL} - O_g)(O_g(1 - O_g))O_{hg}(1 - O_{hg})x_i. \tag{10}$$

where η_g is the learning rate and g is the output of each gating network after applying the softmax function, w_{hg} and w_{yg} are the weights of the inputs to the hidden and the hidden to output layers of the gating network, respectively. Also, O_{hg}^T is the transposes of O_{hg}, the outputs of the hidden layer of gating networks.

To combine the outputs of the experts, the gate assigns a weight g_i as a function of x to each of the experts' output, O_j and the final mixed output of the ensemble is O_T:

$$O_T = \sum_{j=1}^{L} O_j g_j. \tag{11}$$

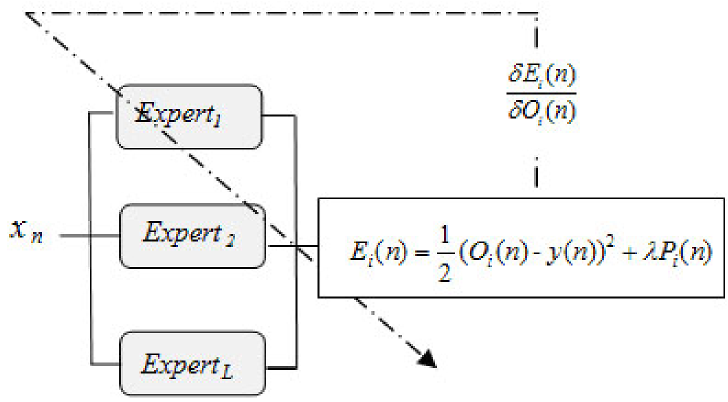
The two training stages of the G-NCL algorithm are shown in Fig. 3.

This approach provides an efficient tool for combining experts based on their local competence. In this approach, the combining weights are estimated dynamically from the inputs based on the different competences of each expert regarding different parts of the problem. Hence, the combination of NCL experts using this approach is superior to the previous static methods [30].

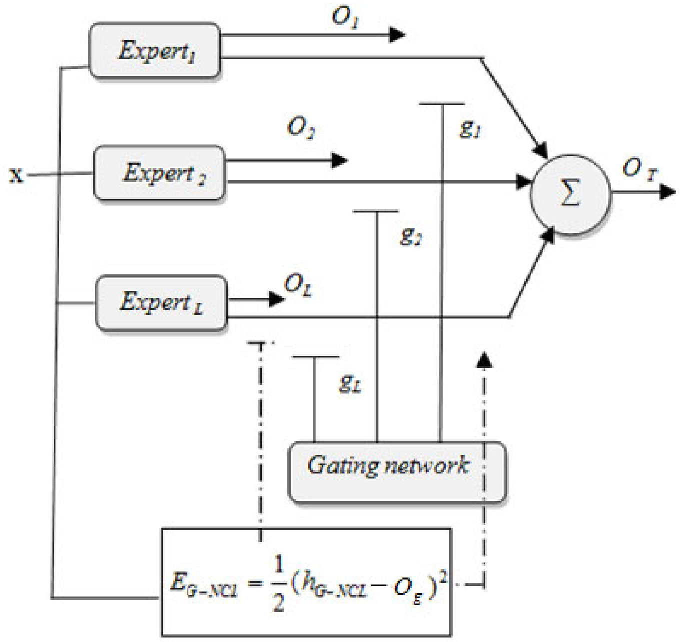
5. Experimental Results and Discussion

The EEG signals can be considered as a superposition of different structures occurring on different time scales at different times. One purpose of wavelet analysis is to separate and sort these underlying structures of different time scales. In this study, EEG signals were segmented by a rectangular window with size 256 so that the EEG signal considered being stationary in that interval. Fig. 4 shows the waveform of normal (set A), seizure-free (set D) and epileptic seizure (set E) EEG segments.

All EEG segments were decomposed into wavelet coefficients using discrete wavelet transform. The number of decomposition levels is chosen based on the dominant frequency components of the signal. In the present study, the number of decomposition levels was chosen to be 4. Ubelyli in [26] showed that the smoothing feature of the Daubechies wavelet of order 2 (db2) is more suitable to detect changes



(a)



(b)

Fig. 3 *Diagram of the two training stages of G-NCL. In the first training stage (a), the expert networks are trained using the NCL error function. In the second stage of the G-NCL algorithm, after training the NCL experts, a gating network is trained to model the local competence of the NCL experts.*

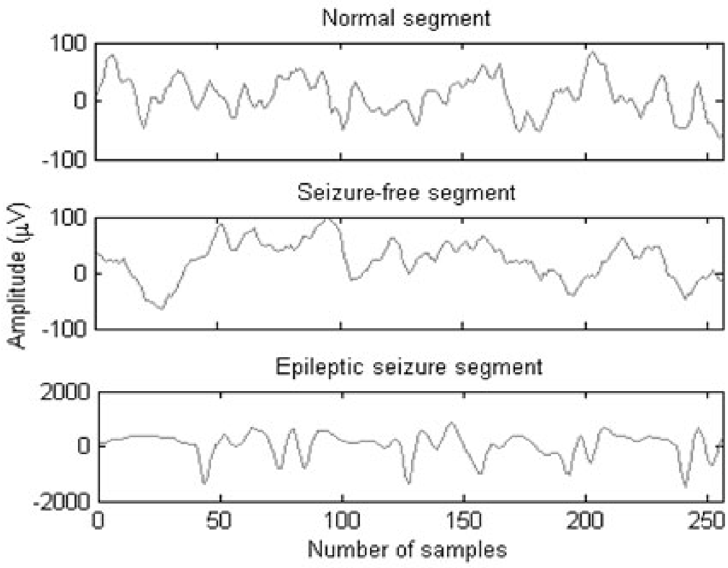


Fig. 4 *Waveform of three EEG signals normal (set A), seizure-free (set D) and epileptic seizure (set E) which are windowed by a rectangular window with size 256. Vertical line shows the amplitude value of signal. Horizontal line shows the number of samples.*

of the EEG signals. Therefore, the wavelet coefficients were computed using the db2 in the present work. 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. The detail wavelet coefficients at first decomposition level of normal, seizure-free and epileptic EEG segments are presented in Figs. 5a-5c, respectively.

Using a smaller number of features to represent the EEG signals is particularly important for recognition and diagnostic purposes. In order to reduce the dimensionality of the feature vectors, the following statistical features were used:

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

Thus, the data set was formed from 4800 vectors (1600 vectors per class) of dimension 20 (four statistical features for each subband). The whole data set was divided into two groups; training and testing sets. The 2400 vectors (800 vectors from each class) where used for training the classifiers and 2400 vectors (800 vectors from each class) were used for testing.

5.1 Experimental setup

Different experiments have been conducted to evaluate the performance of the proposed method. In the first experiment, the performance of the G-NCL was compared to Majority Voting (MV) and Averaging (AVG), as the two previous combining techniques, and the conventional mixture of experts (ME) method. In the second experiment, the effect of ensemble size on the performance of the proposed method was investigated. In all experiments, MLP network with one hidden layer was used as the expert. In all the methods, the number of neurons in the hidden layer of experts was set at 20. Also, in all the experiments (except for the second part of the second experiment), the number of hidden neurons for the gating network of ME and G-NCL methods was set at 10. All the methods were trained using the BP training algorithm. In all the methods, the experts were trained with the learning rate values of $\eta_e = 0.1$. Also, the learning rate values for gating network in the G-NCL and ME methods were set $\eta_g = 0.05$.

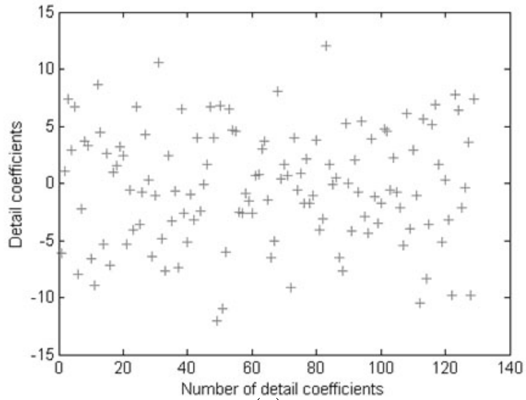
For the G-NCL, MV-NCL (NCL experts with majority voting combining rule) and AVG-NCL (NCL experts with average combining rule) methods, λ^*, the optimum value of λ in terms of the maximum performance, was determined using a trial-and-error procedure for each classification problem in the interval [0.1: 0.1: 1.5] (The numbers range from 0.1 to 1.5 with steps 0.1).

5.2 G-NCL in comparison with previous combining techniques

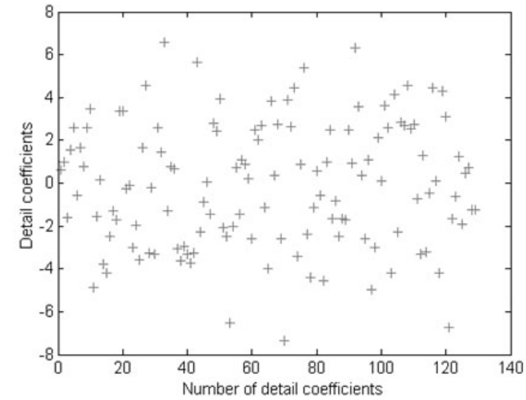
In the first part of this experiment, the performance of the proposed method, G-NCL, was compared to the MV-NCL, AVG-NCL and ME methods. In this part, all the methods were evaluated with five experts. Also, the value of parameter λ varied in the range [0.1:0.1:1.5]. The whole data set is divided into two disjoint equal-sized subsets, training and testing sets. Twenty five percent of the training set is randomly selected for validation set to employ early stopping technique to ensure the generalization ability and the remained samples were used for training the classifiers. After training and validating the classifiers, the testing set was used to verify the accuracy and effectiveness of the implemented classifiers. In this application, there were three types of EEG sets; set A (normal EEG segments from five healthy persons, eyes open), set D (seizure-free EEG segments of five patients from epileptogenic zone), and set E (epileptic seizure segments). The classification results of the G-NCL, MV-NCL, AVG-NCL and ME methods were displayed by a confusion matrix in Tab. I.

According to the results of the Tab. I, 26 normal EEG segments (set A) were classified incorrectly by our proposed method, G-NCL, as seizure-free EEG segments (set D), four normal segments were classified as epileptic seizure segments (set E), 20 seizure-free segments were classified as normal segments, 13 seizure-free segments were classified as epileptic seizure segments, and 14 epileptic seizure segments were classified as seizure-free segments. As seen from Tab. I, the total number of patterns which are incorrectly classified by our proposed method is significantly fewer than the misclassifications number of the other classifiers.

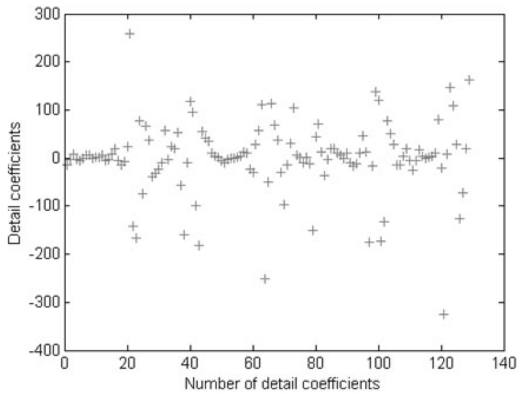
The rightmost column in Tab. I shows the average classification accuracy over the three EEG sets normal, seizure-free and epileptic seizure for each method. The



(a)



(b)



(c)

Fig. 5 *The detail wavelet coefficients at the first decomposition level of EEG segments: (a) normal EEG (set A), (b) seizure-free EEG segments (set D) and (c) epileptic seizure of EEG segments (set E).*

Methods	Desired	Output			Classification accuracy (%)
		Set A	Set D	Set E	
G-NCL	Set A	770	26	4	96.79
	Set D	20	777	13	
	Set E	0	14	786	
MV-NCL	Set A	720	75	5	91.63
	Set D	77	705	18	
	Set E	1	25	774	
AVG-NCL	Set A	725	71	4	92.17
	Set D	70	713	17	
	Set E	0	26	774	
ME	Set A	758	48	5	94.46
	Set D	44	746	18	
	Set E	0	19	781	

Tab. I *The confusion matrix and average classification accuracy of the classifiers.*

average classification accuracy can be obtained by dividing the trace of confusion matrix to the total number of the testing patterns. The average classification accuracy of G-NCL was 96.79% which is significantly better than ME, AVG-NCL and MV-NCL with the accuracies of 94.46%, 92.17% and of 91.61%, respectively.

In the second part of this experiment, the methods are listed in Tab. I, have been assessed in terms of accuracy measures ratio; specificity, sensitivity, selectivity which are defined as follows:

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{12}$$

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{13}$$

$$Selectivity = \frac{TP}{TP + FP} \times 100\%, \tag{14}$$

where TP-true positive, FN-false negative, TN-true negative and FP-false positive.

A true negative occurs when both the classifier and the physician suggested the absence of positive detection. A true positive occurs when the positive detection of the classifier coincided by the positive detection of the physician. A false negative (false positive) occurs when the classifier incorrectly suggested the absence (presence) of positive detection. The performance measures ratio (specificity, sensitivity and selectivity) on the test samples of three types; normal (set A), seizure-free (set D) and epileptic (set E) EEG segments are presented in Tab. II.

As seen from Tab. II, our proposed method classified the normal segments (set A), seizure-free segments of epileptogenic zone (set D) and epileptic segments (set E) with the accuracy of 96.80, 96.28 and 98.25, respectively.

Methods	EEG data sets	Statistical parameters (%)		
		Sensitivity	Specificity	Selectivity
G-NCL	Set A	96.80	98.76	97.47
	Set D	96.28	97.50	95.10
	Set E	98.25	98.94	97.88
MV-NCL	Set A	90.00	95.13	90.23
	Set D	88.12	93.75	87.58
	Set E	97.11	98.38	96.75
AVG-NCL	Set A	90.63	95.63	91.19
	Set D	89.13	93.74	88.02
	Set E	96.75	98.69	97.36
ME	Set A	93.46	97.26	94.51
	Set D	92.10	95.84	91.76
	Set E	97.62	98.58	97.14

Tab. II *The values of performance measures ratio (sensitivity, specificity and selectivity) of the classifiers.*

5.3 The effect of ensemble size on the performance of G-NCL

This experiment also consisted of two parts. In the first part, the effect of ensemble size on the performance of the G-NCL and other implemented methods was investigated. Apart from five experts' case which was evaluated above, all the methods were also tested with three and seven experts. Also, the value of parameter λ varied in the range [0.1:0.1:1.5]. Tab. III shows the classification accuracy of the G-NCL, MV-NCL, AVG-NCL and ME methods for different ensemble size and near optimum value of parameter λ (λ^*) in mean (standard deviation) format.

Number of experts						
Methods	3		5		7	
	$\lambda*$	Acc. (%)	$\lambda*$	Acc. (%)	$\lambda*$	Acc. (%)
G-NCL	0.9	91.33 (0.61)	0.9	**96.79 (0.88)**	1	94.58 (1.09)
MV-NCL	0.9	89.65 (0.45)	0.8	91.63 (0.92)	0.9	90.79 (1.34)
AVG-NCL	0.8	90.06 (0.74)	0.8	92.17 (0.96)	0.7	91.38 (1.29)
ME	–	91.89 (0.83)	–	94.46 (1.15)	–	93.25 (1.67)

Tab. III *The classification accuracy of implemented methods for the EEG classification problem for different ensemble sizes.*

According to the results in Tab. III, the G-NCL with three, five and seven experts classified the EEG signals with the accuracies of 91.33%, 96.79% and 94.58%, respectively.

As shown in Tab. III, the G-NCL obtained higher classification accuracies in respect to the MV-NCL and AVG-NCL methods for all of the tested ensemble sizes. Also, for both cases five and seven experts, our proposed method outperformed the ME method.

In the second part of this experiment, the effect of different number of hidden neurons for the gating network of the proposed method was investigated. In this case, the near-optimum number of hidden neurons in terms of maximum classification accuracy was determined using trial and error procedure in the interval [2:2:20]. (The number of neurons range from two to twenty with steps two). The classification accuracies of the G-NCL with five experts for different number of hidden neurons are presented in Fig. 6.

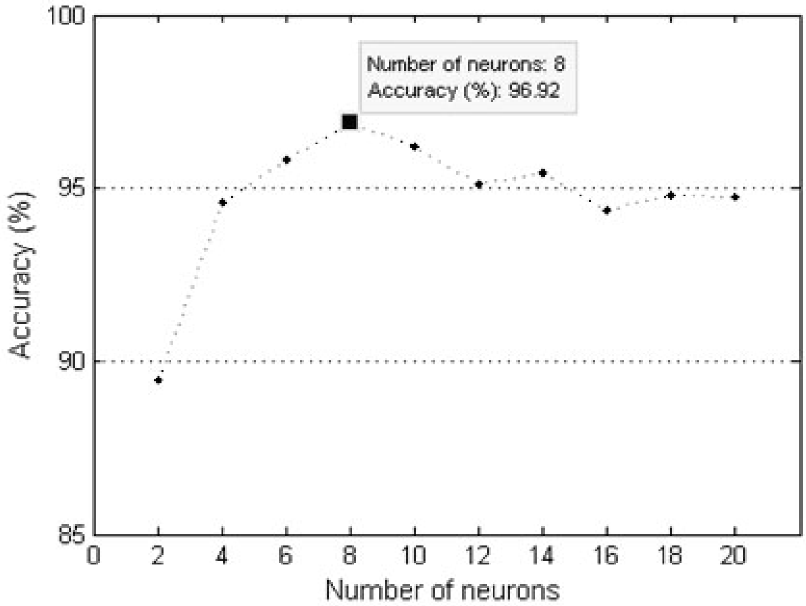


Fig. 6 *The classification accuray of G-NCL for different number of neurons in the hidden layer of gating network. Horizontal line shows the number of neurons varied in the range [2:2:20]. Vertical line shows the classification accuracy.*

As shown in Fig. 6, the classification accuracy of G-NCL appeared to increase initially with increasing number of neurons and reached the best classification accuracy of 96.92% with eight neurons. As also shown in this figure, when the number of neurons was too large, due to increasing the complexity of the gating network, the performance of G-NCL begins to reduce slightly.

As mentioned earlier, the WT is particularly effective for representing various aspects of signals such as trends, discontinuities and repeated patterns where other signal processing approaches fail or are not as effective. It is especially powerful for analyzing non-stationary signals, such as EEG. So the wavelet transform was used to analyze EEG signals in several researches. In this part, we have reviewed some previous methods that were proposed based on neural network classifiers and wavelet transform using same EEG data set. Tab. IV shows a brief description of these methods and their accuracies.

Author(s)	**Features extraction and classification methods**	**Accuracy (%)**
Sadati et. al [36]	DWT + ANFN	85.90
Übeyli [14]	DWT+ ME	93.17
Übeyli [26]	DWT + SG	94.83
	DWT+ MLPNN	84.83
Subasi [16]	DWT+ ME	95.00
	DWT+ MLPNN	93.60
Guler et. al [37]	DWT+ ANFIS	98.68
In this work	DWT + NCL experts with Gating network (G-NCL)	**97.17**

Tab. IV *Description of some NN ensemble methods in the literature for epileptic EEG classification.*

The wavelet coefficients of EEG signals were presented as feature vectors into an adaptive neural fuzzy network (ANFN) to classify normal and epileptic EEG signals was proposed by Sadati et. al [36]. They found that AFFN classified EEG signals with the classification accuracy of 85.90%. Übeyli [14] used the statistical features of wavelet coefficients extracted by discrete wavelet transform in order to train and evaluate a mixture of experts model. Her proposed method classified three types of normal, seizure-free and epileptic EEG sets with the accuracy of 93.17%. Also, in order to improve the generalization ability, Übeyli [26] proposed a two-level neural network structure called stacked generalization (SG). She showed that the proposed method classified three EEG sets A, D and E with the accuracy of 94.83%, which was so much better than stand-alone MLPNN with the accuracy of 84.83%. Subasi [16] also found that the mixture of experts with EM training algorithm identified epileptic seizures of EEG signals with the accuracy of 95.00%, which was moderately better than MLPNN with the accuracy of 93.60%. Güler and Übeyli [37] used an adaptive neuro-fuzzy inference system (ANFIS) model to combine the outputs of five ANFIS classifiers. Their proposed method classified five types of EEG sets (A-E) with the classification accuracy of 98.68%.

In this application, we have proposed a new combining scheme that employs a gating network to combine the experts which are trained using NCL algorithm. The training algorithm of NCL encourages each base expert to learn different part of the training data and the gating network provides a way to support the local competence of base experts. Then, the gating network combines the outputs of the base experts localized in different parts of the training data. The classification accuracy of our proposed method over the three types of normal, seizure-free and epileptic EEG segments was 96.92%.

6. Conclusion

In this paper, first the properties of ME and NCL, as two of more advanced frameworks for ensemble learning machines, were investigated. Then, based on their complementary features, a new method was proposed to improve the classification accuracy. The proposed method, G-NCL, may be considered as a modified version of conventional NCL which uses a dynamic training rule (gating network) to combine the outputs of the NCL experts. NCL training algorithm encourages the experts to learn different parts or aspects of training data and the gating network, as a dynamic combiner in ME, provides the local competence of the NCL experts. Three types of EEG signals (normal, seizure-free and epileptic seizure) were decomposed into wavelet coefficients using discrete wavelet transform and statistical features of the coefficients were used as input patterns, representing them into classifiers. The performance of G-NCL was compared to the MV-NCL, AVG-NCL and conventional ME methods. Experimental results showed that the classification accuracy of the G-NCL is significantly better than conventional ME, AVG-NCL and MV-NCL methods.

References

[1] Mormann F., Andrzejak R. G., Elger C. E., Lehnertz K.: Seizure prediction: the long and winding road, Brain, **130**, 2007, p. 314.

[2] Kay S. M., Marple Jr. S. L.: Spectrum analysis—a modern perspective, Proceedings of the IEEE, **69**, 1981, pp. 1380-1419.

[3] Barlow J. S.: Methods of analysis of nonstationary EEGs, with emphasis on segmentation techniques: a comparative review, Journal of Clinical Neurophysiology, **2**, 1985, p. 267.

[4] Gabor A. J., Seyal M.: Automated interictal EEG spike detection using artificial neural networks, Electroencephalography and C-clinical Neurophysiology, **83**, 1992, pp. 271-280.

[5] Güler I., Kiymik M. K., Akin M., Alkan A.: AR spectral analysis of EEG signals by using maximum likelihood estimation, Computers in Biology and M-medicine, **31**, 2001, pp. 441-450.

[6] Muthuswamy J., Thakor N. V.: Spectral analysis methods for neurological signals, Journal of Neuroscience Methods, **83**, 1998, pp. 1-14.

[7] Liu G., Zhang D., Meng J., Huang G., Zhu X.: Unsupervised adaptation of electroencephalogram signal processing based on fuzzy C-means algorithm, International Journal of Adaptive Control and Signal Processing, 2011.

[8] Daubechies I.: The wavelet transform, time-frequency localization and signal analysis, Information Theory, IEEE Transactions, **36**, 1990, pp. 961-1005.

[9] Walker J. S.: A primer on wavelets and their scientific applications: CRC press, 1999.

[10] Naghsh-Nilchi A. R., Aghashahi M.: Epilepsy seizure detection using eigen-system spectral estimation and Multiple Layer Perceptron neural network, Biomedical Signal Processing and Control, **5**, 2010, pp. 147-157.

[11] Du X., Dua S., Acharya R. U., Chua C. K.: Classification of Epilepsy Using High-Order Spectra Features and Principle Component Analysis, Journal of Medical Systems, 2010, pp. 1-13.

[12] Chua K., Chandran V., Acharya U., Lim C.: Automatic identification of epileptic electroencephalography signals using higher-order spectra, Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, **223**, 2009, pp. 485-495.

[13] Iik H., Sezer E.: Diagnosis of epilepsy from electroencephalography signals using multilayer perceptron and Elman artificial neural networks and wavelet transform, Journal of Medical Systems, **10**, 2010.

[14] Übeyli E. D.: Wavelet/mixture of experts network structure for EEG signals classification, Expert Systems with Applications, **34**, 2008, pp. 1954-1962.

[15] Wang D., Miao D., Xie C.: Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection, Expert Systems with Applications, 2011.

[16] Subasi A.: EEG signal classification using wavelet feature extraction and a mixture of expert model, Expert Systems with Applications, **32**, 2007, pp. 1084-1093.

[17] Patnaik L., Manyam O. K.: Epileptic EEG detection using neural networks and post-classification, Computer Methods and Programs in Biomedicine, **91**, 2008, pp. 100-109.

[18] Iasemidis L. D., Sackellares J. C.: Chaos theory and epilepsy, The Neuroscientist, **2**, 1996, pp. 118-125.

[19] Lerner D. E.: Monitoring changing dynamics with correlation integrals: Case study of an epileptic seizure* 1, Physica D: Nonlinear Phenomena, **97**, 1996, pp. 563-576.

[20] Acharya U. R., Molinari F., Sree S. V., Chattopadhyay S., Ng K. H., Suri J. S.: Automated diagnosis of epileptic EEG using entropies, Biomedical Signal Processing and Control, 2011.

[21] Nicolaou N., Georgiou J.: Detection of epileptic electroencephalogram based on Permutation Entropy and Support Vector Machines, Expert Systems with Applications, 2011.

[22] Güler N. F., Übeyli E. D., Güler I.: Recurrent neural networks employing Lyapunov exponents for EEG signals classification, Expert Systems with Applications, **29**, 2005, pp. 506-514.

[23] Übeyli E.: Fuzzy similarity index employing Lyapunov exponents for discrimination of EEG signals, Neural Network World, **16**, 2006, p. 421.

[24] Güler I., Übeyli E. D.: Multiclass support vector machines for EEG-signals classification, Information Technology in Biomedicine, IEEE Transactions, **11**, 2007, pp. 117-126.

[25] Faust O., Acharya U. R., Min L., Sputh B. H. C.: Automatic identification of epileptic and background EEG signals using frequency domain parameters, International Journal of Neural Systems, **20**, 2010, pp. 159-176.

[26] Übeyli E. D.: Combined neural network model employing wavelet coefficients for EEG signals classification, Digital Signal Processing, **19**, 2009, pp. 297-308.

[27] Lotte F., Congedo M., Lécuyer A., Lamarche F., Arnaldi B.: A review of classification algorithms for EEG-based brain-computer interfaces, Journal of Neural E-engineering, **4**, 2007, p. R1.

[28] Kuncheva L. I.: Combining pattern classifiers: methods and algorithms: Wiley-Interscience, 2004.

[29] Tumer K., Ghosh J.: Error correlation and error reduction in ensemble classifiers, Connection Science, **8**, 1996, pp. 385-404.

[30] Liu Y., Yao X.: Ensemble learning via negative correlation, Neural Networks, **12**, 1999, pp. 1399-1404.

[31] Jacobs R. A., Jordan M. I., Nowlan S. J., Hinton G. E.: Adaptive mixtures of local experts, Neural Computation, **3**, 1991, pp. 79-87.

[32] Waterhouse S., Cook G.: Ensemble methods for phoneme classification, Advances in Neural Information Processing Systems, 1997, pp. 800-806.

[33] Avnimelech R., Intrator N.: Boosted mixture of experts: an ensemble learning scheme, Neural Computation, **11**, 1999, pp. 483-497.

[34] Andrzejak R. G., Lehnertz K., Mormann F., Rieke C., David P., Elger C. E.: 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**, 2001, p. 061907.

[35] Ebrahimpour R., Kabir E., Yousefi M. R.: Face detection using mixture of MLP experts, Neural Processing Letters, **26**, 2007, pp. 69-82.

[36] Sadati N., Mohseni H. R., Maghsoudi A.: Epileptic seizure detection using neural fuzzy networks. In: Proceeding IEEE International Conference on Fuzzy Systems, Canada, 2006, pp. 596-600.

[37] Güler I., Übeyli E. D.: Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients, Journal of Neuroscience Methods, **148**, 2005, pp. 113-121.