Journal of King Saud University – Computer and Information Sciences (2014) 26, 1–6

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Classification of EEG Signals using adaptive weighted distance nearest neighbor algorithm

E. Parvinnia [a,*], M. Sabeti [a], M. Zolghadri Jahromi [b], R. Boostani [b]

[a] Department of Computer Engineering, Shiraz Branch, Islamic Azad University, Shiraz, Iran
[b] Department of Computer Science and Engineering, Shiraz University, Shiraz, Iran

Received 3 March 2012; revised 31 December 2012; accepted 2 January 2013
Available online 8 January 2013

KEYWORDS

Nearest neighbor;
Noisy training data;
EEG signal classification;
Band power;
Fractal dimension;
Autoregressive coefficients

Abstract Electroencephalogram (EEG) signals are often used to diagnose diseases such as seizure, alzheimer, and schizophrenia. One main problem with the recorded EEG samples is that they are not equally reliable due to the artifacts at the time of recording. EEG signal classification algorithms should have a mechanism to handle this issue. It seems that using adaptive classifiers can be useful for the biological signals such as EEG. In this paper, a general adaptive method named weighted distance nearest neighbor (WDNN) is applied for EEG signal classification to tackle this problem. This classification algorithm assigns a weight to each training sample to control its influence in classifying test samples. The weights of training samples are used to find the nearest neighbor of an input query pattern. To assess the performance of this scheme, EEG signals of thirteen schizophrenic patients and eighteen normal subjects are analyzed for the classification of these two groups. Several features including, fractal dimension, band power and autoregressive (AR) model are extracted from EEG signals. The classification results are evaluated using Leave one (subject) out cross validation for reliable estimation. The results indicate that combination of WDNN and selected features can significantly outperform the basic nearest-neighbor and the other methods proposed in the past for the classification of these two groups. Therefore, this method can be a complementary tool for specialists to distinguish schizophrenia disorder.

1. Introduction

Electroencephalogram (EEG) signals (Sanei and Chambers, 2007) are brain activities recorded using electrodes placed on the scalp. Although several methods for the brain function analysis such as megnetoencephalography (MEG), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have been introduced, the EEG signal is still a valuable tool for monitoring the brain activity due to its relatively low cost and being convenient for the patient.

There have been several EEG classification studies within the recent years. These studies used different classification techniques, compared their performance, and evaluated different combinations of feature sets. Among these classifiers, k-nearest neighbor (k-NN), linear discriminant analysis (LDA), support vector machine (SVM), artificial neural network

* Corresponding author. Tel/fax: +98 711 6191656.
E-mail addresses: parvinnia@iaushiraz.ac.ir, parvinn@shirazu.ac.ir, eparvinnia@gmail.com (E. Parvinnia).

Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578 © 2013 Production and hosting by Elsevier B.V. on behalf of King Saud University.
http://dx.doi.org/10.1016/j.jksuci.2013.01.001

(ANN) have been popular. Boostani et al. (2008) used five different classification algorithms including LDA, Boosted version of direct LDA (BDLDA), Adaboost, SVM, and fuzzy SVM to classify two schizophrenic and normal groups. Their result showed the BDLDA method achieved slightly better performance than the other classification methods. Hazarika et al. (1997) applied the three-layered ANN using wavelet transform as a feature extraction method for classifying of three groups: normal, schizophrenia, and obsessive compulsive disorder. Their results showed the wavelet transform can be used as a powerful technique for preprocessing EEG signals prior to classification. Li and Fan, 2005 studied the classification of three kinds of subjects (10 schizophrenic patients, 10 depressive patients and 10 normal controls) with EEG rhythms used as feature vectors. They used two ANN approaches, BP ANN and self-organizing competitive ANN for classification. Their results showed that BP ANN has a better comprehensive performance than the self-organizing competitive ANN technique.

Hornero et al. (2006) used three nonlinear methods of time series analysis for analyzing the time series generated by 20 schizophrenic patients and 20 control subjects. Their results show that the ability of generating random time series between schizophrenic subjects and controls is different. The patient group is characterized by less complex neurobehavioral and neuropsychologic measurements. Rosenberg et al. (1990) studied a random number generation experiment. They asked the participant to choose a random number in interval [1..10] without any generative rule. They found that schizophrenic patients tended to be more repetitive. AlZoubi et al. (2009) evaluated three different classifier techniques to classify the EEG signals in a 10-class emotion experiment. Their results showed using the adaptive algorithm can improve the performance of the classification task.

We believe that the main problem in the classification of EEG signals is the quality of the recorded signal, which can be different during the experiment. These unwanted disturbances cannot be controlled since many activities are going on at the same time in the brain. Existence of artifacts at the time of recording the EEG signal, directly affects the reliability of the recorded signal. It seems that using adaptive classifiers can be useful for the biological signals such as EEG. In this paper, a general adaptive method named weighted adaptive nearest neighbor (WDNN) (Zolghadri et al., 2009) is applied for EEG signal classification. This classifier assigns a weight to each training sample that controls its influence in classifying test samples. When a large weight is assigned to a training sample, it will increase its influence in classifying many samples. On the other hand, reducing the weight of a training sample will decrease its influence in the classification task. The most important ability of this classifier is determining the quality of each EEG segment by assigning different weights for the classification task. Therefore if the training samples are changed, the weights of these samples will be recalculated.

To assess the performance of the WDNN classifier, EEG signals of thirteen schizophrenic patients and eighteen normal subjects are analyzed for the classification of the two groups. The EEG signals are recorded in the Center for Clinical Research in Neuropsychiatry, Perth, Western Australia.

This paper is structured as follows. Section 2 presents nearest neighbor (NN) classification with weighted training samples. In Section 3, feature extraction techniques are illustrated. Experimental results are discussed in Section 4 and Section 5 presents our conclusion.

2. Weighted adaptive nearest-neighbor classification

This method, by assigning a weight to each training sample, attempts to improve the performance of the 1-NN. WDNN tries to minimize the leave one out (LV1) classification error on the given training set by assigning the weights of training samples. These weights are used in the test phase for finding the nearest neighbor of a query sample. By assigning small weights to low quality training samples, their influence in feature space can be reduced.

Assume there is a problem with a set of training samples like (A_i, C_i) where $i = 1, \ldots, n$, A_i has f features, and C_i has M-classes. Different types of distance functions have been introduced by Wilson and Martinez (2000) for measuring the distance between two patterns for identifying the NN of a query pattern. Euclidean distance has been suggested, in most situations, for the distance between two samples A_i and A_j:

$$distance(A_i, A_j) = \sqrt{\sum_{k=1}^{f}(A_{ik} - A_{jk})^2} \qquad (1)$$

The similarity measure can be used instead of using the distance function as follows:

$$\lambda(A_i, A_j) = \frac{1}{distance(A_i, A_j)} \qquad (2)$$

The sample A_r that has the most similarity to a query sample Q can be mentioned as follows by using (2):

$$r = \underset{1\leqslant i\leqslant n}{\arg\max}\{\lambda(Q, A_i)\} \qquad (3)$$

The assumption of NN classifier is all of the training samples have the same weight. The WDNN believes that the quality of the stored samples is not equal. This is especially true when each sample represents an EEG sample recording. To take this into account, a weight w_k is allocated to each training sample A_k. In the test phase, these weights are used for finding the sample A_p that has the most similarity to a query sample Q.

$$p = \underset{1\leqslant i\leqslant n}{\arg\max}\{w_i.\lambda(Q, A_i)\} \qquad (4)$$

2.1. Learning algorithm for weighting training samples

The WDNN is a greedy method that tries to minimize the LV1 error rate of classification on the given training set by specifying the weights of training samples. Note that, a training sample with a large weight can increase its influence in classifying many samples in LV1 test. On the other hand, a training sample having zero weight is not used to classify any test samples and can be removed from the data set.

The main part of the WDNN learning method is a procedure that specifies the best weight for a training sample with respect to all other samples having fixed weights.

WDNN starts with an initial set of weights equal to one ($w_j = 1.0$). The weight of each training sample is adjusted in turn. Assuming a training sample A_k belongs to a sample class that is denoted by *ClassT*, the algorithm tries to specify the best weight w_k, that is a real number in the interval $[0, \infty]$, as follows:

At first, the weight of A_k is set to zero ($w_k = 0$) for removing it from the attribute space. There are some training samples that their classification correctness depends on the value of w_k. To describe the best weight of A_k, the algorithm required to identify these patterns. So, WDNN marks two groups of samples that their classification correctness does not relate to the value of w_k (Zolghadri et al., 2009). These groups are:

1) The samples of $ClassT$ that are classified correct with respect to the weight of A_k is set to zero ($w_k = 0$).
2) The samples of $\overline{ClassT}$ that are misclassified.

Now, the classification of unmarked samples is related to the value of w_k.

In the second step, the score S of any unmarked samples A_t is calculated by the definition as follows:

$$S(A_t) = \frac{\max_{1 \leqslant i \leqslant n} \{w_i.\ \lambda(A_t, A_i),\ i \neq t\}}{\lambda(A_t, A_k)} \quad (5)$$

The most important characteristic of the score of a sample A_t is that if A_k gets a weight $w_k > S(A_t)$, then A_t will select A_k as its nearest neighbor and so classified as $ClassT$. It can derive easily from (5) as follows:

$$w_k.\lambda(A_t, A_k) > \max_{1 \leqslant i \leqslant n} \{w_i.\ \lambda(A_t, A_i),\ i \neq t\} \quad (6)$$

In the last step, the score of unmarked samples is ranked in ascending order to select the best weight for A_k. Suppose that L sample is in the ranked score list. There are $L + 1$ values for w_k, because it is selected between two successive ranked scores to choose the best value. All samples that their scores are smaller than w_k will be classified as $classT$. WDNN chose the best value of $\mathrm{w_k}$ that minimizes the LV1 error for the samples in the list (Zolghadri et al., 2009). The algorithm is shown in Tables I and II (Zolghadri et al., 2009).

3. Feature extraction

Different approaches for extraction of quantitative features from the EEG signal were proposed more than 70 years ago where these methods are usually used to explore the information from EEG. In this paper, the autoregressive (AR) model coefficients, band power and fractal dimension (Boostani et al., 2008; Sabeti et al., 2007) are applied because they investigate the EEG signal in different aspects. They are related to power spectrum, frequency domain and complexity or irregularity of the EEG signals, respectively.

The EEG is inherently a non-stationary signal (Galka, 2000) and the feature extraction methods are only applicable to the stationary signal. In this paper, autocorrelation test as one of the stationary-test methods has been used to determine the size of shorter stationary time series (Chatfield, 1996). Then time series is divided into a number of short windows (one-second interval) and its dynamics is assumed to be approximately stationary within each window (Sabeti et al., 2009). The following feature extraction methods are applied to each one-second windowed signal for each channel.

Table I The procedure for finding the weight of training samples.

1. for $k = 1$ to No. of training samples
   a. $W_k = 0$ {assume A_k belongs to $ClassT$}
   b. Mark samples that have $ClassT$ and classified correctly.
   c. Mark samples that have $\overline{ClassT}$ and are misclassified.
   d. Rank the score of unmarked training samples in ascending order using (5).
   e. Choose the best value for weight of A_k by using the best-weight algorithm (see Table II).

Table II The best-weight algorithm for finding the best value of $\mathrm{w_k}$.

Inputs: L unmarked patterns A_t, with ranked scores $S(A_t)$
{assume that A_t and A_{t+1} are two successive patterns in the L elements ranked list}
Output: the best value of w_k
1. *optimum_state* = the classification rate of training data when $w_k = 0$.
2. *best-threshold* = *0*
3. for $t = 1$ to *L-1*
  3.1. *threshold* = $(S(A_t) + S(A_{t+1}))/2$
  3.1. *current_state* = classification rate based on to the specified threshold
  {all samples A_t that have $S(A_t) <$ *threshold* are classified as $ClassT$ as seen in equation 5}
  3.2. if *current_state* > *optimum_state* then
    Optimum_state = *current_state*
    best-threshold = *threshold*
4. return w_k = *best-threshold*

3.1. Autoregressive coefficients

One of the powerful tools for signal modeling is AR model. In this model, each sample can be predicted from previous weighted samples where the number of coefficients denotes the model order.

$$x(t) = -\sum_{i=1}^{p} \hat{a}_i x(t-i) \quad (7)$$

where $\hat{a}_i$ denotes the AR model coefficients and p is the model order. In this paper, the Burg method (Stoica and Moses, 1997) is applied to estimate the AR coefficients based on forward and backward prediction error. In addition, finite sample criteria (FSC) (Broersen and Wensink, 1993) is used to select the best order of AR model based on the residual variance and the prediction error.

3.2. Band power

It is shown that the EEG contains different frequency components, which can show different brain states and contain the discriminative information. Normally, EEG is classified as delta = [less than 4 Hz], theta = [4–8 Hz], alpha = [8–13 Hz], beta = [13–30 Hz] and gamma = [more than 30 Hz]. Band power feature reflects the power in these five bands at each electrode position. First, the signal is filtered in determined frequency ranges using a band-pass filter (Butterworth filter of order five). Second, each sample is squared and is averaged over a one-second interval.

3.3. Higuchi fractal dimension

The fractal dimension can be interpreted simply as the degree of irregularity in a signal. It estimates the fractal dimension

Table III The Higuchi fractal dimension procedure.

1. Generate k time series x_m^k from $x(t)$ = {x(1), x(2), ..., x(N)} as $x_m^k = \{x(m), x(m+k), x(m+2k), \ldots x(m+\lfloor\frac{N-m}{k}\rfloor k)\}$ where k denotes the delay between the points, m = 1, 2, ..., k, shows the initial time and N denotes the length of time sequence.
2. Compute the average length $L_m(k)$ for each x_m^k as $L_m(k) = \frac{(N-1)\sum_{i=1}^{\lfloor\frac{N-m}{k}\rfloor}\lvert x(m+ik)-x(m+(i-1)k)\rvert}{\lfloor\frac{N-m}{k}\rfloor k}$
3. Compute the total average length $L(k)$for all x_m^k with same k and different m as $L(k) = \sum_{m=1}^{k} L_m(k), \quad k = 1, \ldots, k_{\max}$
4. Plot the curve of $\ln(L(k))$ versus ln(1/k), then estimate Higuchi fractal dimension as the slope of this curve using least square linear fit.

directly in the time domain where the original signal is considered as a geometric figure (Sabeti et al., 2009). The procedure used to estimate Higuchi fractal dimension (Esteller et al., 2001) is shown in Table III.

4. Experimental results

4.1. Data acquisition

Schizophrenia is a severe and persistent psychiatric disorder and it causes some characteristic symptoms including hallucinations, delusions, or disorganized speech (DSM-IV-TR, 2000; ICD-10, 2005). Thirteen schizophrenic patients (all male with mean age 33.3 and standard deviation (std) 9.52) and eighteen normal subjects (all male with mean age 33.4 and std 9.29) participated in this study. The EEG signals are recorded in the Center for Clinical Research in Neuropsychiatry, Perth, Western Australia. The patients were recruited from the admitted population of a psychiatric hospital and they were receiving standard neuraleptics medicine. Additionally, the normal subjects were selected carefully without a history of psychiatric disorder.

Each subject was seated upright with eyes open and EEG signal was recorded for two minutes using a neuroscan 24 channel Synamps system, with a signal gain equal to 75 K (150x at the headbox). Based on the 10–20 system with reference to linked earlobes, 20 electrodes (Fpz, Fz, Cz, Pz, C_3, T_3, C_4, T_4, Fp_1, Fp_2, F_3, F_4, F_7, F_8, P_3, P_4, T_5, T_6, O_1, O_2) were used with a sampling frequency of 200 Hz for recording EEG signals. Elimination of muscle artifacts was performed off-line with visual inspections of EEG and the eye-blink artifacts were omitted by the methods mentioned in (Semlitch et al., 1986). In addition, the signals were filtered with a band pass filter (Butterworth filter of order 5) at 0.5–50 Hz to eliminate the very low and the power line frequency noises. Fig. 1 shows the sample EEG signal plot for normal and schizophrenic subjects on Cz channel.

4.2. Data analysis

The 20-channel EEG signal is partitioned to a number of one-second windows (with 50% overlap) where its dynamics is assumed to be approximately stationary within each window (i.e. for each subject 34 windows). Features were extracted from all channels of each window. In each window, 14 features were extracted that consist of AR coefficients (8), band power (5), and Higuchi fractal dimension (1). Therefore, the data set will have 280 features (20 channel *14 feature) for each window. We normalized the features of each window to the interval [0,1] and used the Euclidian distance function in the experiment. Fig. 2 shows the overall view of the feature extraction process.

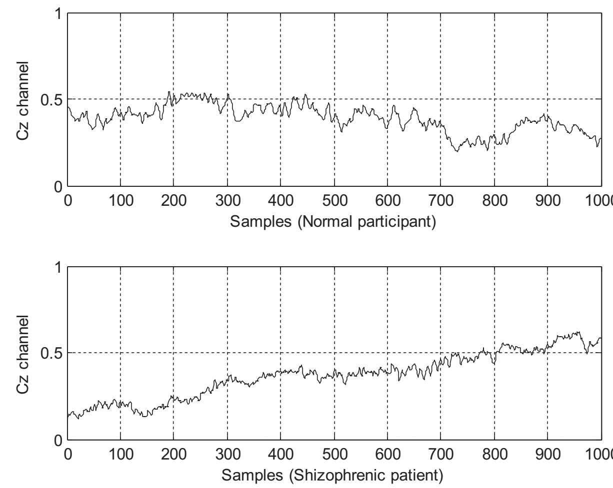

Figure 1 The sample EEG signal of normal and schizophrenic subjects on Cz channel.

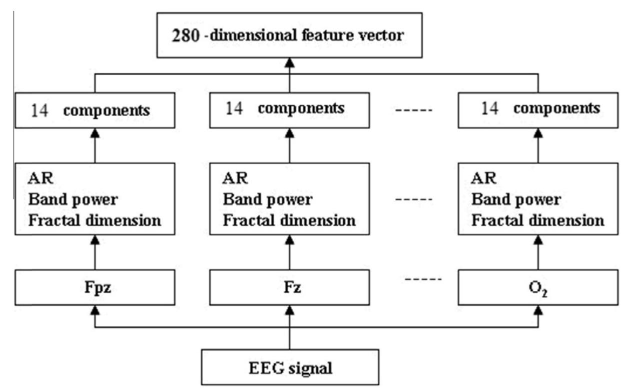

Figure 2 Preprocessing for feature extraction.

The WDNN is applied for EEG signal classification task. This classifier assigns a weight to each training sample that controls its influence in classifying test samples. The noisy window (or segment) is considered as outlier and their influence in classifying test samples is decreased.

Table IV gives the average LV1 generalization accuracy of WDNN classifier for this data set. In LV1, one subject is assigned to the test set and the others used for the training set. This procedure is repeated until all the subjects are used as test data. The average of classification rate on test set is calculated as the performance of classifier.

For comparison with WDNN algorithm, the classification rates of other methods in the literature are also reported in Table IV. The performance and the standard deviation of the five different classifiers are compared in Table IV. To show the improvement of basic NN by WDNN algorithm, the classification rate of basic NN is also shown in Table IV. As seen, the performance of basic NN is improved by WDNN.

Table IV Classification rates of Basic_NN, WDNN, SVM, NaiveBayes, BDLDA, ADM classifiers.

Classifiers	Accuracy ± STD_DEV
WDNN	95.32 ± 4.12
Basic NN	91.08 ± 8.43
SVM	85.02 ± 16.18
NaiveBayes	88.19 ± 9.90
BDLDA	87.51 ± 16.98
ADM	92.75 ± 8.14

AlZoubi et al. (2009) used three adaptive classifiers for the classification of EEG signals. KNN as a classical sample based algorithm with $k = 3$, Naïve Bayes as a standard probabilistic classifier predicts the class of the samples using the maximum estimated posterior probability, and SVM that combines a maximal margin strategy with a kernel method to choose the best boundary in the feature space. The Naïve Bayes and SVM classifiers were applied on our data set using WEKA. These classifiers were set to their default parameter values as implemented in WEKA. Table IV shows their classification rates. As seen, WDNN has 10.20% and 7.13% improvement compared to SVM and Naïve Bayes classifiers, respectively.

Boostani et al. (2009) applied different classifiers for EEG signals, and they reported BDLDA is an efficient classifier for EEG signal classification. In this study, the same data set is used to compare our results with BDLDA. As seen in Table IV, WDNN has 8.81% improvement compared to BDLDA for EEG signal classification.

For comparison of WDNN with other adaptive methods, a locally adaptive distance measure (ADM) (Wang et al., 2007) is used. ADM like WDNN assigns a weight to each training sample, but the parameters of the distance function are specified by a simple heuristic. ADM can be effective in improving the performance of the basic NN. Table IV shows the classification accuracy of ADM on our data set. As seen, ADM improves the classification rate of basic NN by 1.67%. But, WDNN has 2.57% improvement compared to ADM for EEG signal classification.

4.3. Robustness

In order to further verify the robustness of WDNN classifier on noisy data, a noise as a disturbance is considered for the vectors and the classification rate is calculated. We added white noises with different amplitudes to the test vectors. The noise amplitude is based on 10%, 20%, 30%, and 40% of maximum amplitude in each dimension. The classification rates of some classifiers against different amplitudes of the noise were shown in Fig. 3. As seen the slop of the WDNN classifier curve is comparable or lower than other methods. Also, the performance of WDNN is better than other methods. This characteristic of WDNN algorithm shows better reliability and robustness.

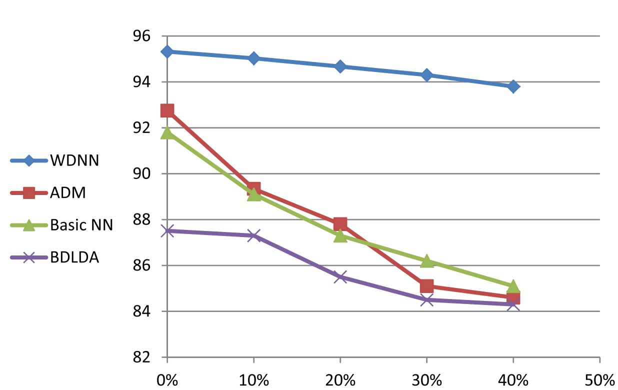

Figure 3 Comparison of robustness of BDLDA, Basic NN, ADM, and WDNN classifiers.

5. Conclusion

The main problem in the classification of EEG signals is the quality of the recorded signal, which can be different during the experiment. These unwanted disturbances cannot be controlled since many activities are going on at the same time in the brain. Changes in the environment can distract the attention of the patient at the time of recording the EEG signal, which directly affects the quality of the recorded signal. In this paper, WDNN is applied for EEG signal classification task. This classifier assigns a weight to each training sample that controls its influence in classifying test samples. When a large weight is assigned to a training sample, it will increase its influence in classifying many samples. In contrast, reducing the weight of a training sample will decrease its effect in classification task.

To show the effectiveness of WDNN for biological signals, EEG signals of eighteen normal subjects and thirteen schizophrenic patients are analyzed with the objective of classifying these two groups. The EEG signals are recorded in the Center for Clinical Research in Neuropsychiatry, Perth, Western Australia. Several features like Higuchi fractal dimension, band power and AR coefficients are extracted from EEG signals. Our results showed that this scheme could improve the generalization accuracy for EEG signal classification task. Therefore, this classifier can be a complementary tool for specialists to distinguish schizophrenia disorder.

For our future work, we decide to use preprocessing methods such as wavelet or principal component analysis instead of using the raw signals. Also, we decide to modify WDNN to assign weights to the features as WDNN is changed to feature-weighing algorithm.

Acknowledgment

The authors express their gratitude to Dr. Gregory W. Price (senior scientist of Center for Clinical Research in Neuropsychiatry, Perth, Western Australia) who provided us with the database of the EEG signals.

References

AlZoubi, O., Calvo, R.A., Stevens, R.H., 2009. In: Nicholson, A., Li, X. (Eds.), Classification of EEG for affect recognition an adaptive approach, Lecture Notes in Computer Science 5866, pp. 52–61.

Boostani, R., Sadatnezhad, K., Sabeti, M., 2008. An efficient classifier to diagnose of schizophrenia based on the EEG signals. Expert Systems with Applications 36, 6492–6499.

Broersen, P.M.T., Wensink, H.E., 1993. On finite sample theory for autoregressive model order selection. IEEE Transaction on Signal Processing 41, 194–204.

Chatfield, C., 1996. The Analysis of Time Series: An Introduction. Chapman & Hall, London, U.K..

Diagnostic and statistical manual of mental disorders (DSM-IV-TR), 2000. American Psychiatric Association, Washington DC.

Esteller, R., Vachtsevanos, G., Echau, J., Litt, B., 2001. A comparison of waveform fractal dimension algorithms. IEEE Transaction on Circuits and Systems I: Fundamental Theory and Applications 48, 177–183.

Galka, A., 2000. Topics in Nonlinear Time Series Analysis with Implication for EEG Analysis. World Scientific, Singapore.

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

Hornero, R., Abasolo, D., Jimeno, N., Sánchez, C.I., Poza, J., Aboy, M., 2006. Variability regularity and complexity of time series generated by schizophrenic patients and control subjects. IEEE Transaction on Biomedical Engineering 53, 210–218.

International Statistical Classification of Diseases and Health Related Problems (ICD-10), 2005. World Health Organization, Geneva.

Li, Y., Fan, F., 2005. Classification of Schizophrenia and Depression by EEG with ANNs. Proceedings of the 2005 IEEE Engineering in Medicine and Biology, 27th Annual Conference Shanghai, September 1–4, (China).

Rosenberg, S., Weber, N., Crocq, M.A., Duval, F., Macher, J.P., 1990. Random number generation by normal alcoholic and schizophrenic subjects. Psychological Medicine 20, 953–960.

Sabeti, M., Boostani, R., Katebi, S.D., Price, G.W., 2007. Selection of relevant features for EEG signal classification of schizophrenic patients. Biomedical Signal Processing and Control 2, 122–134.

Sabeti, M., Katebi, S.D., Boostani, R., 2009. Entropy and complexity measures for EEG signal classification of schizophrenic and control participants. Artificial Intelligence in Medicine 47, 263–274.

Sanei, S., Chambers, J.A., 2007. EEG Signal Processing. John Wiley & Sons.

Semlitch, H.V., Anderer, P., Schuster, P., Presslich, O., 1986. A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. Psychophysiolog 23, 696–703.

Stoica, P., Moses, R.L., 1997. Introduction to Spectral Analysis. Prentice Hall, USA.

Wang, J., Neskovic, P., Cooper, L.N., 2007. Improving nearest neighbor rule with a simple adaptive distance measure. Pattern Recognition Letters 28, 207–213.

Wilson, D.R., Martinez, T.R., 2000. Reduction techniques for exemplar-based learning algorithms. Machine Learning 38, 257–286.

Zolghadri jahromi, M., Parvinnia, E., John, R., 2009. A method of learning weighted similarity function to improve the performance of nearest neighbor. Information Sciences 179, 2964–2973.