Available online at www.sciencedirect.com

ScienceDirect

Biomedical Signal Processing and Control 2 (2007) 122–134

Biomedical Signal Processing and Control

www.elsevier.com/locate/bspc

Selection of relevant features for EEG signal classification of schizophrenic patients

M. Sabeti [a,*], R. Boostani [a], S.D. Katebi [a], G.W. Price [b]

[a] *Department of Computer Science and Engineering, School of Engineering, Shiraz University, Shiraz, Iran*

[b] *School of Psychiatry and Clinical Neuroscience and Centre for Clinical Research in Neuropsychiatry, University of Western Australia and Graylands Hospital, Perth, Australia*

Received 8 December 2006; received in revised form 18 February 2007; accepted 21 March 2007
Available online 16 May 2007

Abstract

In this paper, EEG signals of 20 schizophrenic patients and 20 age-matched control participants are analyzed with the objective of determining the more informative channels and finally distinguishing the two groups. For each case, 22 channels of EEG were recorded. A two-stage feature selection algorithm is designed, such that, the more informative channels are first selected to enhance the discriminative information. Two methods, bidirectional search and plus-*L* minus-*R* (LRS) techniques are employed to select these informative channels. The interesting point is that most of selected channels are located in the temporal lobes (containing the limbic system) that confirm the neuro-phychological differences in these areas between the schizophrenic and normal participants. After channel selection, genetic algorithm (GA) is employed to select the best features from the selected channels. In this case, in addition to elimination of the less informative channels, the redundant and less discriminant features are also eliminated. A computationally fast algorithm with excellent classification results is obtained. Implementation of this efficient approach involves several features including autoregressive (AR) model parameters, band power, fractal dimension and wavelet energy. To test the performance of the final subset of features, classifiers including linear discriminant analysis (LDA) and support vector machine (SVM) are employed to classify the reduced feature set of the two groups. Using the bidirectional search for channel selection, a classification accuracy of 84.62% and 99.38% is obtained for LDA and SVM, respectively. Using the LRS technique for channel selection, a classification accuracy of 88.23% and 99.54% is also obtained for LDA and SVM, respectively. Finally, the results are compared and contrasted with two well-known methods namely, the single-stage feature selection (evolutionary feature selection) and principal component analysis (PCA)-based feature selection. The results show improved accuracy of classification in relatively low computational time with the two-stage feature selection.

Keywords: EEG signal classification; Schizophrenic; Genetic algorithm

1. Introduction

Schizophrenia is a mental disorder from which 1% of the whole population suffer. According to the diagnostic criteria of the American Psychiatry Association [1,2], patients show some characteristic symptoms including delusions, hallucinations or disorganized speech. Recently, much attention has been paid to the analysis of EEG signals of schizophrenic patients. In some research [3–5], non-linear methods have been applied to EEG signals of the two groups of schizophrenic patients and control participants. The results showed differences in dynamic process between the two groups. Hornero et al. [6] asked the participants to press space bar key randomly to generate time series. The results obtained showed that the time series generated by schizophrenic patients had a lower complexity than those for the control group. In an interesting test, for random number generation [7], participants were asked to choose a number from 1 to 10 several times. Numbers had to lack a generative rule, that is, to be as random as possible. They found that schizophrenic patients were more inclined to be repetitive. Pressman et al. [8] showed lack of synchronization alternation ability in the schizophrenic patients during working memory task. They indicated a difference in brain activity, especially in frontal and temporal channels. Cherif et al. [9] indicated that the schizophrenic patients, compared with healthy participants, presented abnormality in eye fixation

[*] Corresponding author. Tel.: +98 9151613674.
E-mail address: sabeti@shirazu.ac.ir (M. Sabeti).

1746-8094/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bspc.2007.03.003

tasks. Gaser et al. [10] depicted structural brain changes in schizophrenic patients. Moreover, Keil et al. [11] showed rhythmic finger oscillations in schizophrenic patients. Paulus et al. [12] carried out a simple choice task consisting of predicting 500 random right or left appearances of a stimulus in order to obtain binary response in patients with schizophrenia and control group. After applying mutual and cross-mutual information, they showed that the response sequences generated by patients exhibited a higher degree of interdependency than those in control group.

The objective of this study is to evaluate different feature extraction methods to classify the two mentioned groups. In this research, it is shown that the selection of EEG recording locations (channel selection) can be done robustly in the absence of prior knowledge of brain activity. Search methods including: bidirectional and plus-*L* minus-*R* techniques are employed to solve the problem of channel selection. After the channel selection phase, genetic algorithm is used to reduce the feature dimension by considering the classification error as fitness function.

The proposed approach is compared and contrasted with other methods, particularly, a single-stage feature selection and a PCA-based method. It is demonstrated that excellent results are obtained in an efficient computational time. The paper is structured as follows: in Section 2, data acquisition method is presented. Feature extraction methods are illustrated in Section 3. Then, channel selection methods are discussed in Section 4. In Section 5, feature selection methods are represented and in Section 6, the classifiers are reviewed and finally the paper will be closed with a description of results and discussion.

2. Data acquisition

Twenty patients with schizophrenia and 20 age-matched control participants (all male, 18–55 years old) participated in this study. They were recruited from the Center for Clinical Research in Neuropsychiatry, Perth, Western Australia. The patients were diagnosed according to DSM-IV criteria [1], and ICD-10 criteria [2] for a lifetime diagnosis of schizophrenia or schizophrenia spectrum disorder were recruited from consecutive admissions to a psychiatric hospital. Probands with comorbid organic brain disease or substance use disorder that could best account for the psychotic symptoms, or probands with language difficulties that were most likely to impede interviewing or testing were excluded. Normal control participants were recruited through advertisements in local community newspapers and at a Red Cross blood donor agency. Participants volunteering as control group were excluded if they showed a history of psychotic disorder, or if any of their first-degree relatives had been diagnosed with schizophrenia, schizophrenia spectrum or bipolar affective disorder.

Patients were receiving a variety of standard neuraleptics at the time of recording, but no effort was made to standardise the dosages. All patients were recruited from the admitted population. Each participant was seated upright with eyes open and the experiment lasted for 2 min. Electrophysiological data were recorded using a Neuroscan 24 Channel Synamps system, with a signal gain equal to 75 K (150× at the headbox). For EEG paradigms, 20 electrodes (Electrocap 10–20 standard system) were recorded plus left and right earlobes and mastoids, VEOG and HEOG. In the EEG paradigms, eye-blink artifacts were corrected using the technique proposed in ref. [13], and manually screened for artifact. According to the international 10–20 system, EEG data were recorded from 22 electrodes (Fpz, Fz, Cz, Pz, C_3, T_3, C_4, T_4, LFTMST, Fp_1, Fp_2, F_3, F_4, F_7, F_8, P_3, P_4, T_5, T_6, O_1, O_2, RHTMST). The sampling frequency f_s is set to 200 Hz based on Nyquist theorem, which states a sampling rate f_s should be at least twice as high as the highest frequency f_{high} (50 or 60 Hz). Therefore, the sampling frequency of more than 100 or 120 Hz is advisable.

3. Feature design and extraction

Four types of feature extraction methods are used in this study, these are autoregressive (AR) model coefficients [14,15], band power [16,17], fractal dimension [18,19] and wavelet energy [20–22]. The EEG signal is assumed to be a non-stationary time series [23] and the feature extraction methods are applicable to stationary signal. To deal with this problem, the time series are divided into a number of short windows and its dynamics is assumed to be approximately stationary within each window. The following feature extraction methods are applied to each 250 ms windowed signal.

3.1. Autoregressive (AR) coefficients

AR model is a powerful tool used for signal modeling. In this model, each sample is considered to be a prediction of previous weighted samples. The number of weights (coefficients) determines the model order. Here, autoregressive coefficients are estimated by Burg method [24]. The Burg method fits a *p*th order AR model to the input signal, *x*, by minimizing (least squares) the forward and backward prediction errors while constraining the AR coefficients, a_i, to satisfy the Levinson-Durbin recursion [24].

$$x(t) = -\sum_{i=1}^{p} a_i x(t-i) \tag{1}$$

One of the important decisions that have to be made in this section is that of estimating the order. Several ideas are used for order determination. Some check the correlation or spectral flatness; other use decision rules based on Bayesian approach, maximum likelihood approach and amount of information measures. In this study, one of the well known methods, Broersen and Wensink's criterion [24] is used to select the best order of AR.

3.2. Band power

Normally, in most cases, most waves in the EEG can be classified as alpha, beta, theta and delta waves. The definition of the boundaries between the bands is somewhat arbitrary, however, in most of applications these are defined as:

delta = [less than 4 Hz], theta = [4–8 Hz], alpha = [8–13 Hz] and beta = [13–30 Hz]. Alpha waves are rhythmical waves that occur at frequencies between 8 and 13 Hz and are found in the EEGs of almost all adult people when they are awake. When the awake person's attention is directed to some specific type of mental activity, the alpha waves are replaced by asynchronous, higher frequency beta waves. Beta waves occur at frequencies greater than 13 Hz. Theta waves have frequencies between 4 and 8 Hz. They occur normally in parietal and temporal regions in children, but they also occur during emotional stress in some adults. Theta waves also occur in many brain disorders, often in degenerative brain states. Delta waves include all the waves of the EEG with frequencies less than 4 Hz, and they occur in very deep sleep, in infancy and in serious organic brain disease. Therefore, EEG contains different specific frequency components, which carry the discriminative information. This type of feature reflects the energy in four bands, which are particularly important to classify different brain states. At first, EEG signals are filtered by four Butterworth band pass filters (order five) in 0–4 Hz (delta), 4–8 Hz (theta), 8–13 Hz (alpha) and 13–30 Hz (beta). Then, the filtered signals are squared to get power of the signal in each band. Finally, by applying an average filter, each sample is an average of 250 ms of the last samples.

3.3. Fractal dimension

Fractal dimension has a relation with entropy, and entropy has a direct relationship with the amount of information inside a signal. Fractal dimension can be interpreted simply as the degree meandering (or roughness or irregularity) of a signal. Here, three methods of fractal dimension calculation are presented.

3.3.1. Higuchi method

Consider $x(1)$, $x(2)$, …, $x(N)$ the time sequence to be analyzed [25]. Construct k time series x_m^k as follow:

$$x_m^k = \left\{ x(m), x(m+k), x(m+2k), \ldots, x\left(m + \lfloor \frac{N-m}{k} \rfloor k \right) \right\} \tag{2}$$

where $m = 1, 2, \ldots, k$, m shows the initial time and k shows delay between the points. For each time series x_m^k, the average length $L_m(k)$ is computed as:

$$L_m(k) = \frac{(N-1)\sum_{i=1}^{\lfloor (N-m)/k \rfloor} |x(m+ik) - x(m+(i-1)k)|}{\lfloor (N-m)/k \rfloor k} \tag{3}$$

where N is the length of time sequence and $(N-1)/\lfloor (N-m)/k \rfloor k$ is a normalization factor. Total average length $L(k)$ is computed for all time series having the same delay k but different m as:

$$L(k) = \sum_{m=1}^{k} L_m(k) \tag{4}$$

This procedure is repeated for each k ranging from 1 to $k_{\max}$, the total average length for delay k, $L(k)$, is proportional to k^{-D}, where D is the fractal dimension by Higuchi's method. In the curve of $\ln(L(k))$ versus $\ln(1/k)$, the slope of the least-squares linear best fit, is the estimate of the fractal dimension.

3.3.2. Katz method

The fractal dimension of a signal can be defined as [26]:

$$D = \frac{\log_{10}(L)}{\log_{10}(d)} \tag{5}$$

where L is sum of distances between successive points and d is the diameter estimated as the distance between the first point of the sequence and the point of the sequence that provides the farthest distance.

3.3.3. Petrosian method

EEG signal is converted to the binary signal using four methods [27]: a, b, c and d. The fractal dimension of the binary sequences is then computed as:

$$D = \frac{\log_{10} n}{\log_{10} n + \log_{10}(n/(n+0.4N))} \tag{6}$$

where n is the length of the sequence and N is the number of sign changes in the binary sequence generated.

3.4. Wavelet transform

The discrete wavelet transform (DWT) is a versatile signal processing tool that finds many engineering and scientific applications. DWT analyzes the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are associated with low-pass and high-pass filters, respectively. The decomposition of the signal into the different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time domain signal. Selection of suitable wavelet and the number of levels of decomposition is very important in analysis of signals using DWT. The typical way is to visually inspect the data first, and if the data are discontinues, Haar or other sharp wavelet functions are applied [28]; otherwise, a smoother wavelet can be employed. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The number of levels of decomposition is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlates well with the frequencies required for classification of the signal are retained in the wavelet coefficients. Since the EEG signals do not have any useful frequency components above 30 Hz, the number of levels is chosen to be 5. Thus, the signal is decomposed into the details D1–D5 and one final approximation, A5. The range of various frequency bands are shown in Table 1.

Table 1
Frequencies corresponding to different levels

Decomposed signal	Frequency range (Hz)
D1	50–100
D2	25–50
D3	12.5–25
D4	6.25–12.5
D5	3.125–6.25
A5	0–3.125

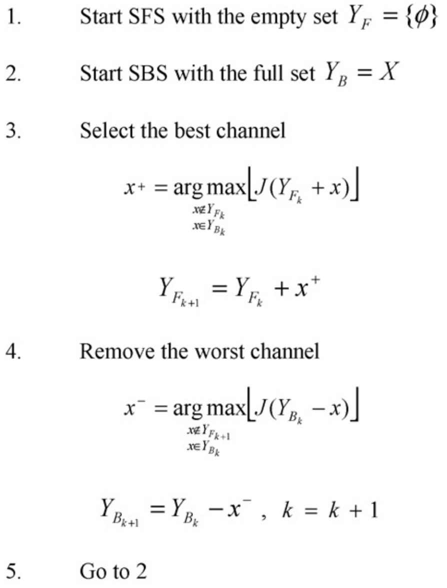
Wavelet transform acts like a mathematical microscope zooming into small scales to reveal compactly spaced events in time and zooming out into large scales to exhibit the global waveform patterns. The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EEG signal in time and frequency. In this study, energy at different frequency band is considered as a feature (six features corresponding to energy of each frequency band are arranged in the feature vector). Table 1 presents frequencies for different levels of decomposition for Daubechies order four wavelet with a sampling rate of 200 Hz. It can be seen from Table 1 that the components A5 decomposition are within the δ (<4 Hz), D5 decomposition are within the θ range (4–8 Hz), D4 decomposition are within the α range (8–13 Hz) and D3 decomposition are within the β range (13–30 Hz).

4. Channel selection

For the application of EEG channel selection [29], it is necessary to treat a certain kind of grouped features homogenously: numerical values belonging to one and the same EEG channel have to be dealt with in a congeneric way so that a spatial interpretation of the solution becomes possible. In this study, the state of the art feature selection methods, bidirectional search and plus-*L* minus-*R* search is adapted for channel selection paradigm. These two methods use the classification algorithm as a subroutine for the channel selection task, in order to evaluate the selected channel set.

4.1. Bidirectional search

Bidirectional search [30,31] is a parallel implementation of forward selection and backward selection. The forward selection tries to select the best channel in an incremental manner. It starts with an empty set (no channels). While the backward selection tries to exclude one redundant channel at a time from the current channel set. It starts with a full set (22 channels). In each step, forward selection adds the best channel to its set and backward selection removes the worst channel from its set. The criterion for selecting the best channel is the classification accuracy. This procedure continues until the forward set and the backward set are equal and this set will be the favored result. To guarantee that forward selection and backward selection converge to the same solution, the following conditions must be met. The first condition: features already selected by forward selection are not removed by backward selection. The second condition: features already removed by backward selection are not selected by forward selection. In this method, search in two directions causes the number of selected channels always to remain fixed (including 11 channels). The algorithm for bidirectional search is outlined in Fig. 1.



```
1.  Start SFS with the empty set $Y_F = \{\phi\}$
2.  Start SBS with the full set $Y_B = X$
3.  Select the best channel
        $x^{+} = \arg\max_{x \notin Y_{F_k}, x \in Y_{B_k}} \lfloor J(Y_{F_k} + x) \rfloor$
        $Y_{F_{k+1}} = Y_{F_k} + x^{+}$
4.  Remove the worst channel
        $x^{-} = \arg\max_{x \notin Y_{F_{k+1}}, x \in Y_{B_k}} \lfloor J(Y_{B_k} - x) \rfloor$
        $Y_{B_{k+1}} = Y_{B_k} - x^{-}$ ,  $k = k + 1$
5.  Go to 2
```

Fig. 1. The algorithm of bidirectional search.

4.2. Plus-L minus-R search

Plus-*L* minus-*R* search (LRS) [30] is a generalization of forward selection and backward selection. In this way, a compact set of channels that minimizes the error of classification, can be obtained. If $L > R$, LRS starts from the empty set and repeatedly adds *L* channels and removes *R* channels. If $L < R$, LRS starts from the full set and repeatedly removes *R* channels followed by adding *L* channels. This procedure continues until the accuracy of the selected channels becomes acceptable. LRS attempts to compensate the weaknesses of forward selection and backward selection with some backtracking capabilities. The main limitation is the lack of a theory to help the prediction of optimal values of *L* and *R*. In LRS method, the number of selected channels can vary when the *L* and *R* parameters are changed. In this study, different values for *L* and *R* are checked. The results show the $L = 3$, $R = 2$ are the best choices. The algorithm for bidirectional search is shown in Fig. 2.

5. Feature selection

Feature selection and dimension reduction [32] are important steps in a pattern recognition task. Here, we are faced with a large number of high dimensional feature vectors, which make the classification process complicated. Some features are redundant and decrease the classification rate, therefore, feature dimension should be reduced. Furthermore, reducing the number of features will also help the classifier to learn a more robust solution and achieve a better generalization

```
1.   If L>R then
        start with the empty set $Y = \{\phi\}$
     else
        start with the full set $Y = X$
        go to step 3
2.   Repeat L times
        $x^{+} = \arg\max_{x \notin Y_k} \lfloor J(Y_k + x) \rfloor$
        $Y_{k+1} = Y_k + x^{+}$ ,  $k = k + 1$
3.   Repeat R times
        $x^{-} = \arg\max_{x \in Y_k} \lfloor J(Y_k - x) \rfloor$
        $Y_{k+1} = Y_k - x^{-}$ ,  $k = k + 1$
4.   Go to 2
```

Fig. 2. The algorithm of LRS.

performance. This is due to the fact that irrelevant feature components are eliminated by the optimal subspace projection. Algorithms for feature selection can be divided into three main groups [33]. Embedded algorithm, where the selection is embedded within the induction algorithm; filter algorithms that select features before passing them to the classification algorithm; wrapper algorithms which performs feature selection around (and with) the classification algorithm. Wrapper methods use the classification algorithm as a subroutine for the feature selection task, in order to evaluate the selected feature set. These methods can be thought of as an optimization algorithm, which uses the results of the classification as the target function. In this study, genetic algorithm (GA) is used for solving the optimization problem. GA is a computational model inspired by evolution. As such, they encode a potential solution as a chromosome-like data structure and apply genetic operators on these structures. In this case, the chromosomes are combination of previously defined features.

5.1. Genetic algorithm

Genetic algorithm [34] is a stochastic search that mimics the natural biological evolution. It operates on a population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them using operators borrowed from natural genetics. This process leads to the evolution of individuals in the population that are better suited to their environment than the individuals that were created from, just as in natural adaptation.

Genetic algorithm models natural processes, such as selection, recombination, mutation, migration, locality and neighborhood. Fig. 3 shows the structure of a simple genetic algorithm. At the beginning of the computation, a number of individuals (the population) are randomly initialized and the objective function is computed for these individuals to produce the first generation. If the optimization criteria are not met, the creation for a new generation starts. Individuals are selected according to their fitness for the production of offspring. Parents are recombined to produce offsprings. Offsprings will be mutated with a certain probability. The fitness of the offsprings is then computed. The offsprings are inserted into the population replacing the parents, producing a new generation. This cycle is repeated until the optimization criteria are met.

6. Classifiers

It is expected that the features selected by the scheme outlined would have a good performance on various types of classifiers. To test this, two widely used classifiers, i.e., linear discriminant analysis (LDA) and support vector machine (SVM), which are the state of art of pattern recognition

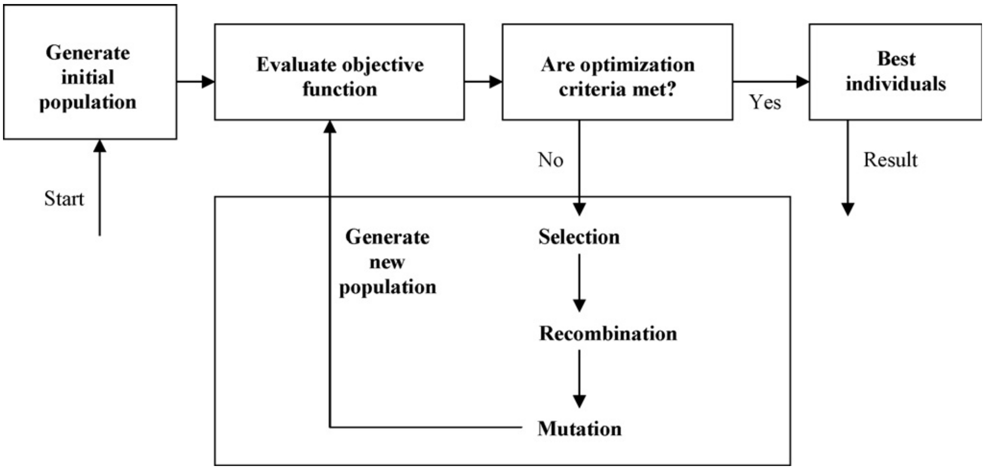


Fig. 3. The simple structure of genetic algorithm.

techniques, are considered. As one of the earliest classifiers, LDA [30] learns a linear classification boundary in the input feature space. SVM [35,36] is a more modern classifier that uses kernels to construct linear classification boundaries in higher dimensional space. Both classifiers are easy to implement and are moderate in computational demand. Additionally, it is straight forward to evaluate classification accuracy in both cases.

6.1. Linear discriminant analysis

This classifier is a statistical binary classifier, which is based on the within and between class scatter matrices. The general formula for LDA is given below:

$$g_i(\underline{x}) = \underline{w}_i^T \underline{x} + w_{i0}, \quad i = 1, 2, \ldots, c \tag{7}$$

where x is the input vector and W is the weight vector. Within class scatter and between class scatter are described as follows:

$$S_i = \sum_{x_i \in C_i} (\underline{x} - \underline{m}_i)(\underline{x} - \underline{m}_i)^T \tag{8}$$

$$S_W = S_1 + S_2 \tag{9}$$

$$S_B = (\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^T \tag{10}$$

The weights of this classifier are optimized based on Fisher's criterion. These weights try to maximize the *between class scatter* and minimize the *within class scatter* to make a better discrimination. Finally, the weights are calculated as follows:

$$\underline{w} = S_W^{-1}(\underline{m}_1 - \underline{m}_2) \tag{11}$$

6.2. Support vector machine

The main idea of SVM is to construct a hyper-plane as a decision surface in such a way that the margin of separation between positive and negative examples is maximized. The support vector machine is an approximate implementation of the method of structural risk minimization. The SVM, given labeled training data

$$D = \{(x_i, y_i)\}_{i=1}^{l}, \quad x_i \in X \subset R^d, \quad y_i \in Y = \{-1, +1\} \tag{12}$$

constructs a maximal margin linear classifier in a high dimensional feature space $\phi(x)$ defined by a positive definite kernel function $k(x,x')$ specifying an inner product in the feature space,

$$\phi(x) \cdot \phi(x') = k(x, x') \tag{13}$$

A common kernel is the Gaussian radial basis function (RBF),

$$k(x, x') = \mathrm{e}^{-||x-x'||^2/2\sigma^2} \tag{14}$$

The function implemented by a support vector machine is given by

$$f(x) = \left\{ \sum_{i=1}^{l} \alpha_i y_i k(x_i, x) \right\} - b \tag{15}$$

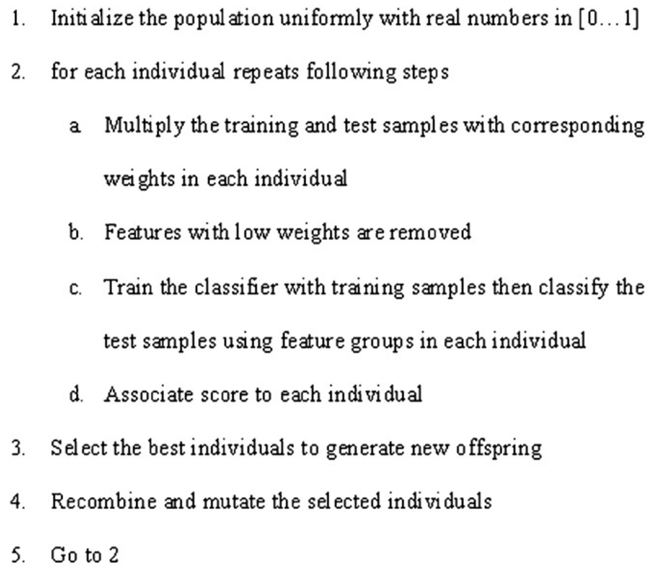
To find the optimal coefficients α of this expansion, it is sufficient to maximize the function,

$$W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \tag{16}$$

7. Computational procedure (two-stage feature selection)

At first, feature extraction phase is done. The EEG signals (22 channels) of 20 schizophrenic patients and 20 normal participants are used as input. The recorded signal of each channel is divided into short windows to validate the assumption of stationary. The window length is 250 ms. Several features including the AR model coefficients (with order eight), band power, fractal dimension and wavelet energy are extracted from each window. If all of these features are arranged in one feature vector, for each windowed signal, there will be 21 features for each channel (8 for AR coefficients, 4 for band power, 3 for fractal dimension and 6 for wavelet energy). The final data set contains 31,240 samples or patterns and the number of features characterizing each pattern is 462. Without any channel selection and feature reduction, the classifiers are faced with rather long feature vector and need high computational time. Reducing the number of features will also help the classifiers to learn a more robust solution and achieve a better generalization performance. One of the objectives of this study is to determine which channels have more discriminative information and finally choose the best subset of features for classification. In this research, two-stage feature selection is suggested and compared with two well-known feature selection methods namely, PCA-based feature selection [37] and single-stage (evolutionary) feature selection. PCA is a projective method, which is attempted to find low dimensional projection that extract useful information from the data, by maximizing a suitable objective function. The dominant eigenvectors (those with the largest associated eigenvalues) of covariance matrix are chosen and the rest of eigenvectors are deleted. In this study, PCA-based feature selection is applied to all 22 channels (462 features). The different choices for reduced dimension are checked and the best choice is 100 features.

In single-stage feature selection, all channels are used by genetic algorithm to reduce the dimension of search (the individual length is 462 doubles). The genetic algorithm is applied to all channels without any channel elimination. In genetic algorithm, each individual in the population has a fixed size and contains weights for corresponding features. In the evolutionary process, the weights are multiplied with the both training and test samples. The features that have weights less than a predefined threshold (T) may be removed to reduce the dimension. Only 10% of the data set (3124 samples) is selected randomly to build training samples and another 10% (3124 samples) is selected randomly to build test samples. Only a fraction of data is used because of large and long computations, using the whole data set. The training and test samples are used



```
1. Initialize the population uniformly with real numbers in [0...1]
2. for each individual repeats following steps
    a. Multiply the training and test samples with corresponding
       weights in each individual
    b. Features with low weights are removed
    c. Train the classifier with training samples then classify the
       test samples using feature groups in each individual
    d. Associate score to each individual
3. Select the best individuals to generate new offspring
4. Recombine and mutate the selected individuals
5. Go to 2
```

Fig. 4. The genetic algorithm.

by two different classifiers, LDA and SVM to evaluate individuals, where each classifier is trained with training samples (with reduced features) and the error of classification on the test samples is used as the fitness value.

Finally, the proposed methodology consists of two stages, first stage is channel selection and the second stage is feature selection and finally classification task is used to evaluate the performance of subset of features. In the first stage, two different methods are used to select the best channels, bidirectional and LRS search. These methods are used to determine which channels have more discriminative information. Both methods investigate the different combination of channels and try to select the best combination. These methods use the accuracy of classification between the two groups as a measure to select the better channels. In these methods, all data is used by five-fold cross validation to split the data set to the training and test samples. A classifier is chosen to train the training samples, the accuracy of classifier on test samples is used as a measure to select the best channel in each iteration.

In the second stage, the features from the selected channels are considered as input. The same procedure as for the single stage is carried out for the feature vector extracted from the selected channels. The procedure for genetic algorithm is outlined in Fig. 4. The global architecture of the proposed approach is outlined in Fig. 5.

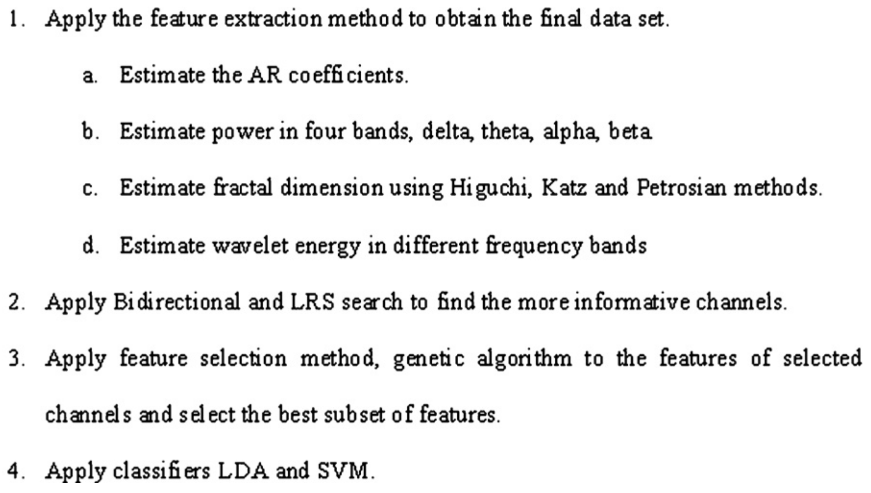
8. Results

In this study, the EEG signal is divided into short windows to validate the assumption of stationary. The window length is 250 ms. At first, two different classifiers, namely LDA and SVM are applied to classify the two groups for each channel. The SVM is selected due to its high accuracy and LDA due to low computational cost. In this experiment, accuracy is calculated using the five-fold cross-validation method on all the data set. Results of this experiment are shown in Tables 2 and 3, respectively. These classifiers are trained with the training set, and then the accuracy of classification on test set is calculated.

According to the results given by Tables 2 and 3, the accuracy of the classification based on each channel is not remarkably good. The summary of results of Tables 2 and 3 is given in Table 4. The mean ± S.D. of all channels classification accuracy, by LDA and SVM, are shown in Table 4. It is shown that the AR coefficients have more discrimination information than other features.

The mean accuracy in Table 4 is used and the channels with an accuracy level higher than the mean value are specified in Table 5. According to the mean of accuracy, most channels on (or close to) temporal lobes have an accuracy higher than the mean. For each classifier, the more accurate channels are shown in Table 5. The bold channels are located on temporal lobes.

Next, bidirectional search as well as plus-*L* minus-*R* techniques are used for channel selection. Using all channels will lead to a large dimension of search, therefore, selecting those channels with a better discrimination ability, can improve



```
1. Apply the feature extraction method to obtain the final data set.
    a. Estimate the AR coefficients.
    b. Estimate power in four bands, delta, theta, alpha, beta
    c. Estimate fractal dimension using Higuchi, Katz and Petrosian methods.
    d. Estimate wavelet energy in different frequency bands
2. Apply Bidirectional and LRS search to find the more informative channels.
3. Apply feature selection method, genetic algorithm to the features of selected
   channels and select the best subset of features.
4. Apply classifiers LDA and SVM.
```

Fig. 5. The simple structure of the proposed approach.

Table 2
Accuracy of classification by LDA for each channel

Channel	AR	Katz	Petrosian	Higuchi	Wavelet energy	Band power
Fpz	0.6812	0.6631	0.5668	0.5760	0.5679	0.5286
Fz	0.6677	0.4521	0.5566	0.5645	0.5004	0.5450
Cz	0.6625	0.5617	0.5210	0.5104	0.5335	0.5463
Pz	0.6587	0.5439	0.5198	0.5199	0.5093	0.5173
C_3	0.6477	0.5657	0.5144	0.4861	0.4958	0.5435
T_3	0.6886	0.5030	0.5251	0.5239	0.5743	0.5750
C_4	0.7025	0.5466	0.5221	0.5118	0.5076	0.5422
T_4	0.6828	0.5219	0.5084	0.5165	0.5065	0.5677
LFTMST	0.6170	0.5398	0.5516	0.5708	0.5295	0.6300
Fp_1	0.6562	0.5472	0.4923	0.5079	0.5469	0.5277
Fp_2	0.6492	0.5205	0.5190	0.5136	0.5567	0.5303
F_3	0.6515	0.5460	0.5076	0.5020	0.5007	0.5120
F_4	0.6700	0.5360	0.5078	0.4942	0.5137	0.5410
F_7	0.6679	0.5147	0.5150	0.4922	0.4907	0.5284
F_8	0.7004	0.5210	0.5231	0.5057	0.5045	0.5328
P_3	0.6643	0.5366	0.5005	0.5063	0.5109	0.5801
P_4	0.6726	0.5641	0.5398	0.5196	0.5166	0.5652
T_5	0.6844	0.4974	0.5300	0.5245	0.5043	0.5222
T_6	0.6975	0.5463	0.5412	0.5324	0.4868	0.4682
O_1	0.6529	0.5356	0.5423	0.5604	0.5043	0.4990
O_2	0.6455	0.5483	0.5538	0.5519	0.5125	0.5111
RHTMST	0.6360	0.4851	0.5278	0.5495	0.5149	0.5343

Table 3
Accuracy of classification by SVM for each channel

Channel	AR	Katz	Petrosian	Higuchi	Wavelet energy	Band power
Fpz	0.7155	0.6624	0.5626	0.5666	0.5746	0.5729
Fz	0.7011	0.6847	0.5626	0.5832	0.5633	0.5770
Cz	0.6852	0.6174	0.5715	0.5852	0.5714	0.5690
Pz	0.6962	0.6287	0.5848	0.5916	0.5632	0.5689
C_3	0.6757	0.6136	0.5626	0.5846	0.5657	0.5779
T_3	0.7188	0.6389	0.5626	0.5626	0.5851	0.5771
C_4	0.7280	0.6171	0.5626	0.5818	0.5651	0.5681
T_4	0.7265	0.6353	0.5626	0.5792	0.5678	0.5920
LFTMST	0.6832	0.5626	0.5626	0.5726	0.5628	0.6245
Fp_1	0.6982	0.5693	0.5626	0.5767	0.5724	0.5814
Fp_2	0.6872	0.5732	0.5626	0.5790	0.5755	0.5807
F_3	0.6978	0.5773	0.5626	0.5795	0.5703	0.5665
F_4	0.7053	0.5679	0.5626	0.5641	0.5744	0.5645
F_7	0.7048	0.5698	0.5626	0.5687	0.5658	0.5630
F_8	0.7340	0.6100	0.5626	0.5830	0.5652	0.5680
P_3	0.6950	0.6101	0.5681	0.5829	0.5630	0.5828
P_4	0.7026	0.6301	0.5686	0.5854	0.5645	0.5651
T_5	0.7260	0.5880	0.5626	0.5732	0.5645	0.5669
T_6	0.7223	0.6465	0.5769	0.5872	0.5660	0.5802
O_1	0.7037	0.5709	0.5802	0.5822	0.5645	0.6570
O_2	0.6907	0.5961	0.5778	0.5810	0.5632	0.5626
RHTMST	0.7257	0.5833	0.5626	0.5832	0.5636	0.5761

the classification accuracy. The results of these two methods are shown in Tables 6 and 7.

Shaded channels in Tables 6 and 7 are related to the temporal lobes. Most of the selected channels are located in the temporal lobes or close to them, which contain the limbic system (Hypocamp and Amigdal) area. This result confirms the neuropsychological differences between normal and schizophrenic participants in the limbic areas [1]. These results also show that the selected channels in the limbic area carry more discrimination information between schizophrenic and normal participants. Voting operator (the channel will be chosen if it is selected by at least four features from six) is used to combine results of channel selection methods for each feature. In this way, the final selected channels are found as shown in Tables 8 and 9.

Genetic algorithm is applied to features of all selected channels to reduce the dimension of feature vectors using two evaluation functions—LDA and SVM. Classification error is

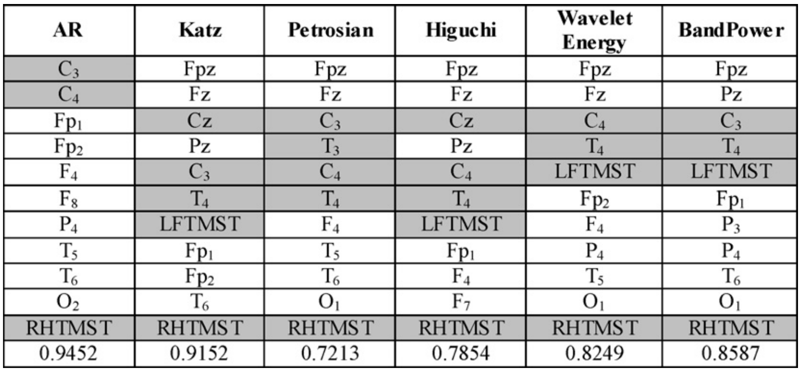
Table 4
The mean ± S.D. of all channels classification accuracy

	AR	Katz	Petrosian	Higuchi	Wavelet energy	Band power
LDA	0.6662 ± 0.0216	0.5362 ± 0.0393	0.5266 ± 0.0194	0.5245 ± 0.0264	0.5177 ± 0.0241	0.5385 ± 0.0326
SVM	0.7056 ±0.0167	0.6070 ± 0.0341	0.5667 ± 0.0070	0.5788 ± 0.0077	0.5678 ± 0.0057	0.5792 ± 0.0219

Table 5
The accurate channels for different features

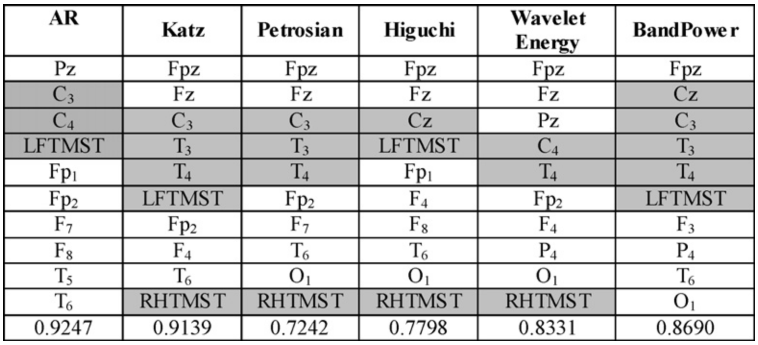
AR coefficients	LDA	Fpz, Fz, **T3**, **C4**, **T4**, F4, F7, F8, P4, T5 and T6
	SVM	Fpz, **T3**, **C4**, **T4**, F8, T5, T6 and **RHTMST**
Katz	LDA	Fpz, **Cz**, Pz, **C3**, **C4**, **LFTMST**, Fp1, F3, P3, P4, T6 and O2
	SVM	Fpz, Fz, **Cz**, Pz, **C3**, **T3**, **C4**, **T4**, F8, P3, P4 and T6
Petrosian	LDA	Fpz, Fz, **LFTMST**, P_4, T_5, T_6, O_1, O_2 and **RHTMST**
	SVM	Cz, Pz, P_3, P_4, T_6, O_1 and O_2
Higuchi	LDA	Fpz, Fz, **LFTMST**, T_5, T_6, O_1, O_2 and **RHTMST**
	SVM	Fz, **Cz**, Pz, **C_3**, **C_4**, **T_4**, Fp_2, F_3, F_8, P_3, P_4, T_6, O_1, O_2 and **RHTMST**
Wavelet energy	LDA	Fpz, **Cz**, **T_3**, **LFTMST**, Fp_1 and Fp_2
	SVM	Fpz, **Cz**, **T_3**, **T_4**, Fp_1, Fp_2, F_3, F_4
Band power	LDA	Fz, **Cz**, **C_3**, **T_3**, **C_4**, **T_4**, **LFTMST**, F_4, P_3, P_4
	SVM	**T_4**, **LFTMST**, Fp_1, Fp_2, P_3, T_6, O_1

Table 6
Channels selected by bidirectional search



AR	Katz	Petrosian	Higuchi	Wavelet Energy	BandPower
C_3	Fpz	Fpz	Fpz	Fpz	Fpz
C_4	Fz	Fz	Fz	Fz	Pz
Fp_1	Cz	C_3	Cz	C_4	C_3
Fp_2	Pz	T_3	Pz	T_4	T_4
F_4	C_3	C_4	C_4	LFTMST	LFTMST
F_8	T_4	T_4	T_4	Fp_2	Fp_1
P_4	LFTMST	F_4	LFTMST	F_4	P_3
T_5	Fp_1	T_5	Fp_1	P_4	P_4
T_6	Fp_2	T_6	F_4	T_5	T_6
O_2	T_6	O_1	F_7	O_1	O_1
RHTMST	RHTMST	RHTMST	RHTMST	RHTMST	RHTMST
0.9452	0.9152	0.7213	0.7854	0.8249	0.8587

Table 7
Channels selected by plus-L minus-R ($L = 3$, $R = 2$)



AR	Katz	Petrosian	Higuchi	Wavelet Energy	BandPower
Pz	Fpz	Fpz	Fpz	Fpz	Fpz
C_3	Fz	Fz	Fz	Fz	Cz
C_4	C_3	C_3	Cz	Pz	C_3
LFTMST	T_3	T_3	LFTMST	C_4	T_3
Fp_1	T_4	T_4	Fp_1	T_4	T_4
Fp_2	LFTMST	Fp_2	F_4	Fp_2	LFTMST
F_7	Fp_2	F_7	F_8	F_4	F_3
F_8	F_4	T_6	T_6	P_4	P_4
T_5	T_6	O_1	O_1	O_1	T_6
T_6	RHTMST	RHTMST	RHTMST	RHTMST	O_1
0.9247	0.9139	0.7242	0.7798	0.8331	0.8690

taken as the fitness function for genetic algorithm. The simulation parameters of GA are shown in Table 10.

Two experiments are carried out. In the first experiment, the genetic algorithm is applied to the selected channels by the bidirectional technique. In this experiment, the individual length is 210 doubles (10 channels × 21 features). The best fitness versus mean fitness of population (fitness is error of classification on the test set) during the evolutionary process are shown in Figs. 6 and 7, respectively.

In the second experiment, the genetic algorithm is applied to the selected channels by the LRS technique. In this experiment, the individual length is 189 doubles (9 channels × 21 features). The best fitness and mean fitness of population (fitness is the error of classification on the test set) during the evolutionary process are shown in Figs. 8 and 9, respectively. In these two experiments, the features with very low weights are removed and the search dimension is decreased. Genetic algorithm decreased the number of features approximately to half, where threshold T is set to mean of weights on each individual. Generally, genetic algorithm reduced the classification error because of decreasing the dimensionality and redundancy.

After using the genetic algorithm to reduce the dimension, the two described classifiers are applied to all data set to classify the two groups. Prediction success of the classifier may be evaluated by examining the confusion matrix. In order to analyze the output data obtained from the application, sensitivity (true positive ratio) and specificity (true negative ratio) are calculated by using confusion matrix, where *sensitivity* = TP/(TP + FN), *specificity* = TN/(TN + FP) and *accuracy* = (TP + TN)/(TP + TN + FP + FN) (TP = true positive; TN = true negative; FP = false positive; FN = false negative). Also, the results are compared by receiver operating characteristic (ROC) [38] analysis. ROC analysis is an

Table 8
The final selected channels by bidirectional search

Fpz	Fz	C_3	C_4	T_4	LFTMST	Fp_1	F_4	T_6	RHTMST

Table 9
The final selected channels by LRS search

Fpz	Fz	C_3	T_4	LFTMST	Fp_2	T_6	O_1	RHTMST

Table 10
The simulation parameters of GA

Population size	100
Stopping criteria	Maximum generation
Crossover operator	Scattered
Crossover probability	0.8
Initialization	Uniform in [0, . . ., 1]
Selection	Stochastic uniform
Mutation operator	Gaussian
Mutation probability	0.05

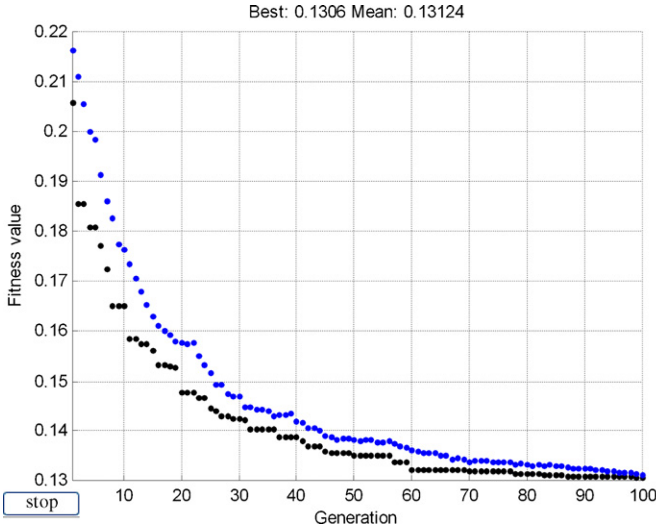


Fig. 6. Best fitness vs. mean fitness (LDA).

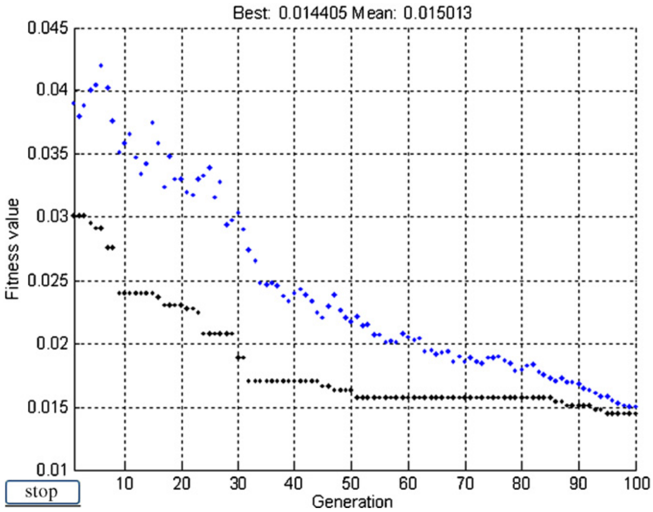


Fig. 7. Best fitness vs. mean fitness (SVM).

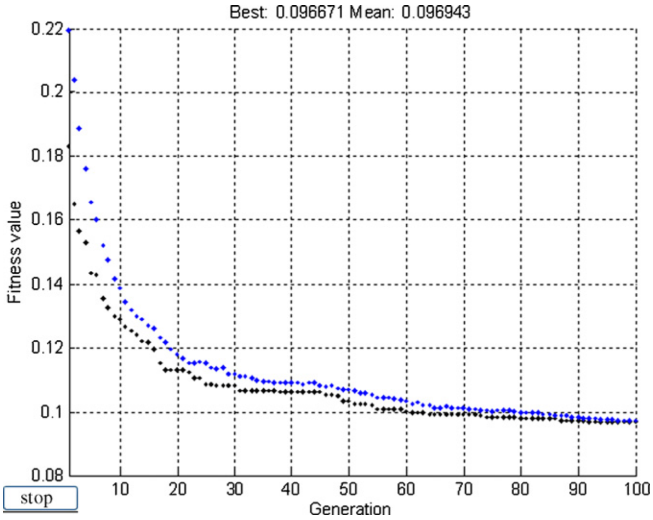


Fig. 8. Best fitness vs. mean fitness (LDA).

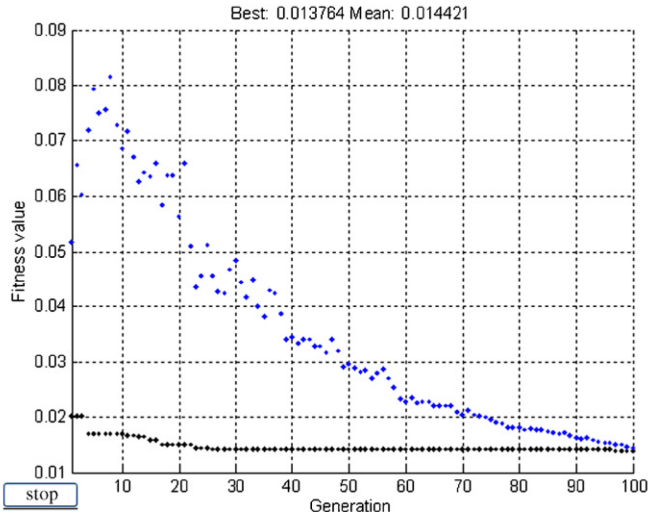


Fig. 9. Best fitness vs. mean fitness (SVM).

appropriate means to display sensitivity and specificity relationships when a predictive output for two possibilities is continuous. In its tabular form, the ROC analysis displays true and false positive and negative totals and sensitivity and specificity for each listed cutoff value between 0 and 1. In order to perform the performance measure of the output classification graphically, the ROC curve is calculated by analyzing the output data obtained from the test. Furthermore, the performance of the model may be measured by calculating the region under the ROC curve.

The ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificty) for each possible cutoff. To quantify the performance characteristics of each classifier, the area under ROC curve (AUC) is computed for validation data ROC curve. The final results for two classifiers to discriminate schizophrenic and normal participants using five-fold cross validation are listed in Tables 11–15.

Finally, the two-stage feature selection approach is compared with a single-stage feature selection and PCA-based feature selection. The result of comparison is presented in Tables 16 and 17. PCA-based feature selection is applied to all channels. Although, almost formidable computational time is required, the classification results are not satisfactory. In the single-stage feature selection, the classification accuracy is not as good as the two-stage feature selection, but more importantly it needs much higher computational time (because of larger feature vector).

9. Discussion

In this research, EEG signals of 20 schizophrenic patients and 20 age-matched control participants are analyzed. Well-known features including AR coefficients, band power, fractal dimension and wavelet energy are extracted from the signals. The results showed that AR coefficients are the most consistent feature for discrimination of the two groups. Katz's method is also a consistent feature for discrimination of these two groups, likely due to its exponential transformation of fractal dimension

Table 11
Test results on all data set by different classifiers

Different classifiers	Sensitivity		Specificity		Accuracy		Area under ROC curve	
	Bidirectional	LRS	Bidirectional	LRS	Bidirectional	LRS	Bidirectional	LRS
LDA	0.8315	0.8665	0.8651	0.9026	0.8462	0.8823	0.9261	0.9520
SVM	0.9931	0.9953	0.9946	0.9956	0.9938	0.9954	0.9995	0.9999

values and relative insensitivity to noise. Higuchi's method, however, yields a more accurate estimation of signal fractal dimension, but is more sensitive to noise. The performance of Petrosian method depends on the type of binary sequence used and the band power and wavelet energy had good information for discrimination (the accuracy of band power and wavelet energy is the same). After feature extraction phase, two different classifiers are applied for the classification task for each channel. According to Tables 1 and 2, the accuracy of classification based on each channel was not remarkable. Table 5 shows that the most channels on (or close to) temporal lobes have an accuracy rate higher than the mean accuracy. For each classifier, the more accurate channels are shown in Table 5. The bold channels are located on temporal lobes. Next, bidirectional search and plus-*L* minus-*R* techniques are used for channel selection. Using all channels led to a large dimension, therefore, these techniques are used to select some channels for better discrimination ability. Bidirectional search is a parallel implementation of SFS and SBS, and this property lowers the cost of search for the best channels. Plus-*L* minus-*R* technique attempts to compensate the weaknesses of SFS and SBS with some backtracking capabilities. The main limitation is the lack of a theory to help predict the optimal values of *L* and *R* (a proposed method to choose the best values for *L* and *R* can be obtained by trial and error). The results in Tables 8 and 9 show the final selected channels by bidirectional and LRS techniques. Most of the selected channels by these two techniques were located on or around the temporal lobes containing the limbic system (Hypocamp and Amigdal). It is shown by PET and MRI images that schizophrenic patients have some changes in Hypocamp and Amigdal parts of their brain compared to normal population [1]. Therefore, the results of channel selection firmly confirm the neuro-psychological differences between the two groups in the temporal lobes. After channel selection, the genetic algorithm is used to reduce the complexity and redundancy by feature selection. It is shown that the genetic algorithm reduced the classification error because it decreased the dimension and redundancy. The final result obtained from two different classifiers is shown in Tables 11–15. For both classifiers, sensitivity, specificity, accuracy and area under ROC curve are computed using five-fold cross validation on all the data set (with reduced features). Both classifiers differentiate the two groups with an accuracy rate above 80%, but SVM gives the best results. The reason is that SVM maps the features in high dimension and according to cover's theorem [35], in the mapped space, classification accuracy can be improved. Finally, the proposed two-stage feature selection is compared with PCA-based feature selection and a single-stage feature selection. The result is presented in Tables 16 and 17. The two-stage feature selection presents better results in comparison with PCA-based and the single-stage feature selection. The results show that PCA-based feature selection, although almost formidable computational time is required, the classification results are not satisfactory. In the classification procedure, the aim is to represent the features which possess the maximum discriminatory information between given classes and not to faithfully represent each class by itself. There may be indeed cases where the two classes share the same important features, but also have some different features (which may be less important in terms of representing each class). If the dimension of the classes by keeping only the important features, is reduced, some of the discriminatory information is lost. In the single-stage feature selection, the classification accuracy is not as good as the two-stage feature selection, but it needs higher computational time (because of larger feature vector). This

Table 12
Confusion matrix for LDA classifier (bidirectional)

	Actual positive	Actual negative
Predicted positive	2922.8	368.8
Predicted negative	592.2	2364.2

Table 13
Confusion matrix for SVM classifier (bidirectional)

	Actual positive	Actual negative
Predicted positive	3490.8	14.8
Predicted negative	24.2	2718.2

Table 14
Confusion matrix for LDA classifier (LRS)

	Actual positive	Actual negative
Predicted positive	3045.6	266.2
Predicted negative	469.4	2466.8

Table 15
Confusion matrix for SVM classifier (LRS)

	Actual positive	Actual negative
Predicted positive	3498.4	12
Predicted negative	16.6	2721

Table 16
The comparison of different methods with LDA

	Sensitivity	Specificity	Accuracy	Area under ROC curve
LRS channel selection followed by feature selection	0.8665	0.9026	0.8823	0.9520
Bidirectional channel selection followed by feature selection	0.8315	0.8651	0.8462	0.9261
Single-stage feature selection	0.7960	0.8569	0.8226	0.9039
PCA-based feature selection (with 100 features)	0.7617	0.8355	0.7940	0.8820

Table 17
The comparison of different methods with SVM

	Sensitivity	Specificity	Accuracy	Area under ROC curve
LRS channel selection followed by feature selection	0.9953	0.9956	0.9954	0.9999
Bidirectional channel selection followed by feature selection	0.9931	0.9946	0.9938	0.9995
Single-stage feature selection	0.9912	0.9898	0.9906	0.9995
PCA-based feature selection (with 100 features)	0.9442	0.9079	0.9283	0.9485

analysis can be a complementary tool to help psychiatrists to diagnose schizophrenic patients.

References

[1] American Psychiatric Association: diagnostic and statistical manual of mental disorders: DSM-IV, Washington DC, 1994.

[2] International Statistical Classification of Diseases and Health Related Problems (The) ICD-10, second ed., World Health Organization, 2005.

[3] J. Roschke, J. Fell, P. Beckmann, Nonlinear analysis of sleep EEG data in schizophrenia: calculation of the principal Lyapunov exponent, Psychiatry Res. 56 (1995) 257–268.

[4] M. Koukkou, D. Lehmann, J. Wackermann, Dimensional complexity of EEG brain mechanisms in untreated schizophrenia, Biol. Psychiatry 33 (1993) 397–407.

[5] J. Jeong, D.J. Kim, J.H. Chae, S.Y. Kim, H.J. Ko, I.H. Paik, Nonlinear analysis of the EEG of schizophrenics with optimal embedding dimension, Med. Eng. Phys. 20 (1998) 669–676.

[6] R. Hornero, D. Abásolo, N. Jimeno, C.I. Sánchez, J. Poza, M. Aboy, Variability, regularity and complexity of time series generated by schizophrenic patients and control subjects, IEEE Trans. Biomed. Eng. 53 (2) (2006) 210–218.

[7] S. Rosenberg, N. Weber, M.A. Crocq, F. Duval, J.P. Macher, Random number generation by normal, alcoholic and schizophrenic subjects, Psychol. Med. 20 (1990) 953–960.

[8] A. Pressman, A. Peled, A.B. Geva, Synchronization analysis of multichannel EEG of schizophrenic during working-memory tasks, in: IEEE proceedings of the 21st convention of the electronics engineering, Israel, (2000), pp. 337–341.

[9] R. Cherif, A. Naït-Ali, J.F. Motsch, M.O. krebs, A parametric analysis of eye tremor movement during ocular fixation, applied to schizophrenia, in: Proceeding of the 25th annual international conference of the IEEE EMBS cancun, Mexico, (2003), pp. 2710–2713.

[10] C. Gaser, H.P. Volz, S. Kiebel, S. Riehemann, H. Sauer, Detecting structural changes in whole brain based on nonlinear deformations—application to schizophrenia research, Neuroimage 10 (1999) 107–113.

[11] A. Kiel, T. Elbert, B. Rockstroh, W.J. Ray, Dynamical aspects of motor and perceptual processes in schizophrenic patients and healthy controls, Schizophr. Res. 33 (1998) 169–178.

[12] M.P. Paulus, M.A. Geyer, D.L. Braff, Long-range correlations in choices sequences of schizophrenics patients, Schizophr. Res. 35 (1999) 69–74.

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

[14] H. Schwilden, Concepts of EEG processing: from power spectrum to bispectrum, fractals, entropies and all that, Best Pract. Res. Clin. Anaesthesiol. 20 (1) (2006) 31–48.

[15] Y. Shen, E. Olbrich, P. Achermann, P.F. Meier, Dimensional complexity and spectral properties of the human sleep EEG, Clin. Neurophysiol. 114 (2003) 199–209.

[16] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain–computer communication, Proc. IEEE 89 (7) (2001) 1123–1134.

[17] B. Obermaier, C. Neuper, C. Guger, G. Pfurtscheller, Information transfer rate in a five-classes brain-computer interface, IEEE Trans. Rehabil. Eng. 9 (3) (2001) 283–288.

[18] R. Esteller, G. Vachtsevanos, J. Echauz, B. Litt, A comparison of waveform fractal dimension algorithms, IEEE Trans. Circuits Syst. I: Fund. Theory Appl. 48 (2) (2001) 177–183.

[19] J.Z. Liu, Q. Yang, B. Yao, R.W. Brown, G.H. Yue, Linear correlation between fractal dimension of EEG signal and handgrip force, Biol. Cybernetics 93 (2) (2005) 131–140.

[20] A. Subasi, Epileptic seizure detection using dynamic wavelet network, Expert Syst. Appl. 29 (2005) 343–355.

[21] H. Adeli, Z. Zhou, N. Dadmehr, Analysis of EEG records in an epileptic patient using wavelet transform, J. Neurosci. Methods 123 (2003) 69–87.

[22] A. Subashi, E. Ercelebi, Classification of EEG signals using neural network and logistic regression, Comput. Methods Programs Biomed. 78 (2005) 87–99.

[23] A. Galka, Topics in Nonlinear Time Series Analysis With Implication for EEG Analysis, World Scientific, 2000.

[24] P. Stoica, R.L. Moses, Introduction to Spectral Analysis, Prentice Hall, 1997

[25] T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Physica D 31 (1988) 277–283.

[26] M. Katz, Fractals and the analysis of waveforms, Comput. Biol. Med. 18 (3) (1988) 145–156.

[27] A. Petrosian, Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns, in: Proceeding of the IEEE Symposium on Computer-Based Medical Systems, 1995, 212–217.

[28] M. Akay, Wavelet applications in medicine, IEEE Spect. 34 (5) (1997) 50–56.

[29] T.N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, B. Schölkopf, Support vector channel selection in BCI, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1003–1010.

[30] A.R. Webb, Statistical Pattern Recognition, second ed., John Wiley and Sons Ltd., 2002.

[31] P. Pudil, J. Novovicova, J. Kittler, Floating search methods in feature selection, Pattern Recognit. Lett. 15 (1994) 1119–1125.

[32] E. Yom-Tov, G.F. Inbar, Feature selection for the classification of movements from single movement-related potentials, IEEE Trans. Neural Syst. Rehabil. Eng. 10 (3) (2002) 170–177.

[33] A.L. Blum, P. Langley, Selection of relevant features and examples in machine learning, Artif. Intell. 97 (1997) 245–271.

[34] D. Whitley, A genetic algorithm tutorial, Stat. Comput. 4 (2) (1994) 65–85.

[35] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines (and Other Kernel-Based Learning Methods), Cambridge University Press, 2000.

[36] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.

[37] I.T. Jolliffe, Principal Component Analysis, second ed., Springer Series in Statistics, 2002.

[38] M.S. Pepe, The Statistical Evaluation of Medical Tests For Classification and Prediction (Oxford Statistical Science Series), Oxford University Press, 2003.