Classifying Depression Patients and Normal Subjects using Machine Learning Techniques

Behshad Hosseinifard*, Mohammad Hassan Moradi **, and Reza Rostami***
* Amirkabir University of Technology, behshad.h@aut.ac.ir
** Amirkabir University of Technology , mhmoradi@aut.ac.ir
***Tehran University , rrostami@ut.ac.ir

Abstract: *Depression is one of the most common mental disorder that at its worst can lead to suicide. Diagnosing depression in the early curable stage is very important. In this paper we study performance of different classification techniques for classifying depression patients from normal subjects. For this aim, power spectrum of three frequency band (alpha, beta, theta) and the whole bands of EEG are used as features. We have shown that Support Vector Machine (SVM) classifier using Genetic algorithm for feature selection can achieve accuracy of 88.6% on classifying depression patients.*

Keywords: Depression, EEG, Power Spectrum, Support Vector Machine, Linear discriminant analysis, genetic algorithm.

1. Introduction

Depression is one of the most common mental disorders that affect about 121 million people worldwide. Lifetime prevalence of depression is about 17% [1]. It is estimated by the World Health Organisation that depression will be the second major disability causing disease in the world by 2020. Depression presents with low interest and motivation, low energy and poor concentration and at its worst, lead canto suicide. This disorder can be diagnosed at primary care.

Researchers have shown that electroencephalogram (EEG) signals can be used to distinguish between depressive patients and normal healthy persons [2,3,4,5]. The EEG signals reflect electrical activity of the brain during various states. Li and Fan [2] extracted power of time series for four frequency bands (alpha, beta, theta, delta) and used Back-Propagation ANN (BP-ANN) and self-organizing competitive ANN for discriminating normal controls, depressive patients, and schizophrenic patients. In the other experiment, Fan et al. [3] used Lempel-ziv complexity as a feature and applied BP-ANN for classification. Li et al. [4] studied wavelet entropy of EEG of 20 normal subjects and 20 depressive patients. Knott et al. [5] derived relative and absolute power, frequency, asymmetry and coherence measures from spectrally analyzed EEG of 70 depressive male and 23 normal controls.

The aim of this paper is to discriminate the depressive patients and control subjects. For this purpose, EEG data of 19 channels for 30 depressive patients and 30 normal persons are recorded. Total power spectrum of the whole band of EEG and sub-bands are extracted as features. Logistic Regression (LR) model [6] and Support Vector Machine (SVM) classifier [6] are used to classify depression patients from normal subjects. In the last stage, for selecting the most important features, genetic algorithm (GA) [7] is used. Our experimental result show that SVM classifier using 15 features and genetic algorithm as the feature selection technique can achieve accuracy of 88.6%.

The main contributions of this paper are: (1) studying performance of different machine learning techniques for classifying depression patients, and (2) studying performance of different features and using feature selection technique to select the best features among them. In the next sections we describe our method and the experimental results.

2. Methods

2.1. Data acquisition

The EEG data are obtained from Psychiatry Centre Atieh . The data includes 40 depressive patients (33.58 ± 10.74 year; mean ± standard deviation) who met the DSM-IV criteria and 40 normal persons who have no psychiatric disorders in the past. The EEG data are recorded in resting condition with eyes closed for 5 minutes and obtained from 19 surface electrodes placed on the scalp according to the standard international 10/20 system (Fz, Cz, Pz, Fp_1, Fp_2, F_3, F_4, F_7, F_8, C_3, C_4, T_3, T_4, P_3, P_4, T_5, T_6, O_1 and O_2). The sampling frequency fs is set to 256 Hz with 12 bit A/D convertor precision. All EEG signals are highpass filtered with 0.5Hz cutoff frequency and lowpass filtered with 70Hz cutoff frequency. Notch filter is used to remove the 50Hz frequency. Artifacts were inspected visually and discarded.

2.2. Feature Extraction

First, the EEG signals are filtered with bandpass butterworth filter to extract 3 common frequency band, Alpha (8-13 Hz), Beta (4-13 Hz) and theta (4-8 Hz). Then, for each band and the whole band of EEG, the size of 15360 samples (1 minute of signals) are selected and power spectrum analysis are performed by Welch method [7]. In the Welch method, time series is divided into segments (possibly overlapping) and modified periodogram of all segments is averaged. The i'th modified periodogram is defined as the following:

$$\widehat{P}_{xx}^{(i)}(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} x_i(n)w(n)e^{-2j\pi fn}\right|^2 \qquad (1)$$

In the Eq.1 M is the length of time series *w(n)* represents window function and *U* is a normalizing factor defined as

$$U = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \qquad (2)$$

Finally, power spectrum density estimation by Welch method is:

$$P_{xx}^{w}(f) = \frac{1}{L}\sum_{i=0}^{L-1} P_{xx}^{(i)}(f) \qquad (3)$$

After calculating power spectrum, total power for each band was obtained.

2.3. Classification

Two linear classification techniques are used in our experiments. We also tested other classification techniques such as SVM with non-linear kernel, Naive Bayes [6] , etc using different types of features. However our results with linear SVM and logistic regression are superior to those obtained with other methods.

Logistic regression is used as powerful technique for classification. The classification is done by fitting the training data to a logistic function. For the binary classification, the input to the logistic function is a feature vector and the output is the probability of classifying the input data to the positive or the negative classes. SVM is a popular machine learning technique that constructs a hyperplane (or a set of hyperplanes) to classify the data points. The classification is done by maximizing the margin between the seperators and the training data points. In the linear case, the margin is defined by the distance of the hyperplane to the nearest data point in the positive and negative training data points.

In the first experiment, each EEG band that includes 19 features is used for classifying two classes: normal subjects and depressive patients. Leave-One-Out Cross Validation (LOOCV) method is used to test the classifier. In LOOCV, each of the participants is considered as the test data and the others as the training data. At each step, one of the data is tagged as the test data and the classifier is trained on the rest. Then the result of the classifier is tested on the tagged data. This process is continued for all the possible ways of leaving out one of the data. We have reported the average accuracy of the classifier for all iterations.

2.4. Feature selection and classification

Feature selection is a very important step that can improve the performance of the classifier, especially when we are dealing with the high dimension data. We have 80 training data and 57 features which have made our data very sparse. In these cases, feature selection can significantly improve the accuracy of classifier by selecting the most informative features.

We have used GA as the feature selection technique in our experiments. Each chromosome is defined by a vector of length k where k is the number of features. Each bit in the chromosome corresponds to one of the features and it indicates if the correspondence feature is selected in the feature selection process. In each generation of GA, selected features are given to the classifier and the overall accuracy of classifier is determined as the fitness function for the next generation. In our experiments, the size of population for GA is set to 50, crossover rate to 80% and mutation rate to 5%. In our experiments we show that the accuracy of classifier is significantly improved by using GA as the feature selection technique.

3. Experimental Results

Table I shows the experimental result when features of 3 bands are used as the input for the classifier. The first row shows the result of logistic regression classifier and the second row shows the result when SVM is used as the classifier. We can observe that the alpha band has achieved the highest accuracy result when is used as the input of the classifier. SVM has achieved the accuracy of 69% when alpha band is only used as the input of the classifier. The experimental result show that the accuracy of SVM classifier can be improved from 69% to 71% when we use all the Alpha, Theta and Beta bands as the input of the classifier.

Table II shows the experimental results when we use GA as the feature selection technique. It is shown that the accuracy of both classifiers for all the bands is significantly improved when we use feature selection. The best result is obtained when we combine all the features of 3 bands and apply GA for the feature selection. Table II shows that SMV can achieve the accuracy of 88% when all the bands are used as its input. The accuracy of classifier is improved approximately by 23% when we use feature selection technique comparing to the results in Table I.

TABLE I: Results of LR and SVM Classifier without Feature Selection.

Classifier	Alpha	Theta	Beta	Alpha + Theta +Beta
LR	68.8%	62.26%	49.6%	61.32%
SVM	69%	65.09%	51%	71.70%

TABELII: Results of LR and SVM Classifier with GA Feature Selection

Classifier	Alpha	Theta	Beta	Alpha +Theta +Beta
LR	80%	73%	67%	86.7%
SVM	82%	71.6%	70%	88.6%

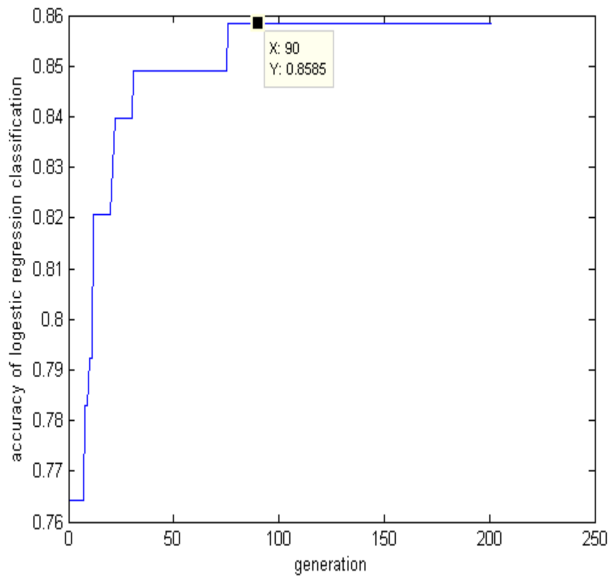

Fig.1: Accuracy of logistic regression classification with genetic algorithm for feature selection

Figure I and II show the result of GA during the feature selection process. The X axis shows the generation number in GA. At each generation, a set of features are candidate by GA. The classifier is trained using these features and its accuracy is used as an evaluation function. The result shows that GA improves the accuracy of the classifier during its process.

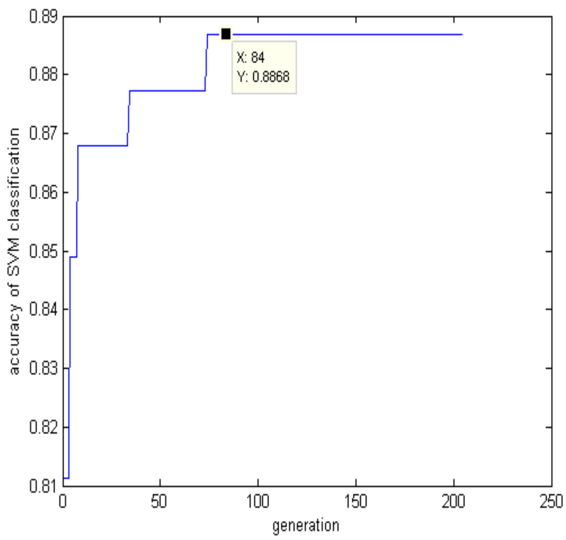

Fig.2: Accuracy of SVM classification with genetic algorithm for feature selection

4. Conclusion

In this paper we showed that classification techniques can be used to classify depressive patients with a high accuracy. In our experiments, power spectrum of 3 frequency bands of EEG (Alpha, Theta and Beta) are used as the input features for the classifier. We showed experimental results of using both LR and SVM classifiers using LOOCV method for experiments. Furthermore, GA was applied for feature selection. The experimental result showed that SVM with feature selection can achieve accuracy of 88.6% on test data. We studied performance of feature selection in our domain and showed that it can improve the accuracy of the classifier by about 23%. Moreover, the results showed that the power of alpha band is significantly more informative comparing to other EEG bands.

Acknowledgment

The authors would like to thank Atieh Psychiatry Centre for helping to obtain EEG data of depression patients.

Reference

[1] B. J .Sadock, V. A. Sadock, and P. Ruiz, *Kaplan & Sadock's Comprehensive Textbook of Psychiatry*, 9th ed. United States: Lippincott Williams and Wilkings, 2008, p. 1647.

[2] Y. Li and F. Fan, "Classification of Schizophrenia and Depression by EEG with ANNs," in *Proc. 2005 IEEE Engineering in Medicine and Biology Conf.,* pp. 2679-2682.

[3] F. Fan, Y. Li, Y. Qiu, and Y. Zhu, "Use of ANN and Complexity Measures in Cognitive EEG Discrimination," in *Proc. 2005 IEEE Engineering in Medicine and Biology conf.*, pp. 4638-4641.

[4] Y. Li, Yi. Li, Sh. Tong, Y. Tang, and Y. Zhu," More normal EEGs of depression patients during mental arithmetic than rest," in *Proc. 2007 IEEE of NFSI & ICFBI conf.*, pp. 165-168.

[5] V. Knott, C. Mahoney, S. Kennedy, and K. Evans, "EEG power, frequency, asymmetry and coherence in male depression," *Psychiatry Res.*, vol. 106, no. 2, pp. 123-140, 2001.

[6] T. M. Mitchell, *Machine Learning,* McGraw-Hill, New York, 1997.

[7] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and Anil K. Jain, ''Dimensionality Reduction Using Genetic Algorithms," *IEEE Trans. Evolutionary Computation,* vol. 4, no.2, pp. 164-171, July. 2000.

[8] A. Alkan and M. K. Kiymik, "Comparison of AR and Welch Methods in Epileptic Seizure Detection," Journal of Medical Systems, vol. 30, no. 6, pp. 413-419, 2006.