COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 109 (2013) 339–345

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal

Behshad Hosseinifard[a,*], *Mohammad Hassan Moradi*[a], *Reza Rostami*[b]

[a] *Department of Biomedical Engineering, Amirkabir University of Technology, Iran*
[b] *Department of Psychology, Tehran University, Iran*

ARTICLE INFO

Article history:
Received 19 November 2011
Received in revised form
20 September 2012
Accepted 5 October 2012

Keywords:
Depression
EEG
Correlation dimension
Detrended fluctuation analysis
Higuchi fractal
Lyapunov exponent

ABSTRACT

Diagnosing depression in the early curable stages is very important and may even save the life of a patient. In this paper, we study nonlinear analysis of EEG signal for discriminating depression patients and normal controls. Forty-five unmedicated depressed patients and 45 normal subjects were participated in this study. Power of four EEG bands and four nonlinear features including detrended fluctuation analysis (DFA), higuchi fractal, correlation dimension and lyapunov exponent were extracted from EEG signal. For discriminating the two groups, *k*-nearest neighbor, linear discriminant analysis and logistic regression as the classifiers are then used. Highest classification accuracy of 83.3% is obtained by correlation dimension and LR classifier among other nonlinear features. For further improvement, all nonlinear features are combined and applied to classifiers. A classification accuracy of 90% is achieved by all nonlinear features and LR classifier. In all experiments, genetic algorithm is employed to select the most important features. The proposed technique is compared and contrasted with the other reported methods and it is demonstrated that by combining nonlinear features, the performance is enhanced. This study shows that nonlinear analysis of EEG can be a useful method for discriminating depressed patients and normal subjects. It is suggested that this analysis may be a complementary tool to help psychiatrists for diagnosing depressed patients.

1. Introduction

Depression is one of the most common mental disorders that affects 121 million people worldwide. It is estimated by the World Health Organization that depression will be the second major disability causing disease in the world by 2020 [1]. Depression is more than a low mood and people with depression may experience lack of interest in daily activities, poor concentration, low energy, feeling of worthlessness, and at its worst, could lead to suicide [2]. The exact cause of depression is not known. Many researchers believe it is caused by chemical imbalances in the brain, which may be hereditary or caused by events in a person's life.

EEG is a medical test used to measure the electrical activities of brain and evaluate brain disorders. Among all methods that used for diagnosing brain disorders and studying of brain functions, EEG has been more popular due to its low-cost and comparatively easy recording process. There are some evidences that EEG may be a useful tool in studying of depression [3–5].

* *Corresponding author*. Tel.: +98 9173056417.
E-mail address: behshad.fard@gmail.com (B. Hosseinifard).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2012.10.008

During past years, theoretical and experimental studies of brain have shown that this system is best-characterized nonlinear dynamical process. The nonlinearity of brain limits the ability of linear analysis to provide full description of underlying dynamics [6]. The nonlinear analysis method such as correlation dimension or DFA effectively applied to EEG to study complexity and dynamics of brain behavior. There are few EEG studies on depression that based on nonlinear analysis and in most studies linear methods such as power and frequency have been used [4,5]. Fan et al. [7] used Lempel–Ziv complexity as a feature for classifying 26 normal persons, 62 schizophrenic and 48 depressed patients and applied BP-ANN for classification. The overall accuracy in this study was about 80%. Li et al. [8] studied wavelet entropy of EEG of 20 normal subjects and 20 depressed patients and they observed higher wavelet entropy in depression groups. In addition, the discriminant analysis and jackknife replication classification yield an accuracy of 80%. Detrended fluctuation analysis was used for 11 depressed patients and 11 normal subjects in [9]. In this study higher DFA value was obtained in depression groups.

The aim of this study is to classify depressed patients and control subjects based on nonlinear features and improving the accuracy of classification. The EEG of 19 channels is recorded for 45 depression patients and 45 healthy participants. Two groups of features are studied in this study. Power of four EEG bands: delta, theta, alpha and beta as frequency and linear features and DFA, higuchi, correlation dimension and maximum lyapunov exponent are four nonlinear features that are used. For evaluation these features, LDA (linear discriminant analysis), LR (logistic regression) and KNN (*k*-nearest neighbors) classifier are adopted to classify the two groups of depressed patients and normal subjects. For selecting the most important and discriminate features, genetic algorithm is used.

2. Materials and methods

In the first stage, feature extraction is performed. The EEG signals of 45 depressive patients and 45 normal subjects are used as the input. Power of four EEG bands and four nonlinear features, correlation dimension, Higuchi, DFA and large lyapunov exponent are calculated for 19 EEG channels.

In the first experiment, each feature vector (included 19 features related to 19 channels) is applied to KNN, LDA and LR classifiers. To validate reliability and generalization of classifiers and datasets independent test is used in this paper. For the independent dataset test, each dataset is divided into two parts, a training set and a testing set. Two-third samples are chosen randomly as training set, and the remainder, one-third samples as testing set. Leave-One-Out Cross Validation (LOOCV) method is applied in classification of training data and genetic algorithm for feature selection. Finally, based on the selected features in classifying of training dataset, classification of test dataset is performed. The results of classifiers on the test datasets are shown in Sections 3.1 and 3.2. In the second experiment, all extracted features of each group, power and nonlinear, are used by the classifiers. In Section 3.3 the results of second experiment are given.

2.1. Data acquisition

The EEG data were obtained from Psychiatry Centre Atieh, Tehran, Iran. The data included 45 right-handed unmedicated depressed patients (23 of which were females), ranged in age from 20 to 55 years (33.5 ± 10.7 years; mean ± standard deviation (Std.)) and 45 normal persons right-handed (25 of which were females), ranged in age from 19 to 60 years (33.7 ± 10.2 years; mean ± Std.), who have no psychiatric disorders in past. Depression diagnosing is assessed prior to EEG data. For diagnosing depression symptoms and illness severity, two criteria were considered: DSM-IV interview resulting in a diagnosis of depression [2] and Beck Depression Inventory (BDI) [10] score of ≥10.

The EEG data were recorded in resting condition with eyes closed for 5 min. Each participant was seated in a comfortable chair in an electrically and acoustically shielded room. EEG recording were obtained from 19 surface electrodes placed on the scalp according to standard international 10/20 system (Fz, Cz, Pz, Fp1, Fp2, F3, F4, F7, F8, C3, C4, T3, T4, P3, P4, T5, T6, O1 and O2). The sampling frequency, *fs* is set to 256 Hz with 12 bit A/D convertor precision. All EEG signals were highpass filtered with 0.5 Hz cutoff frequency and lowpass filtered with 70 Hz cutoff frequency. Notch filter is used to remove the 50 Hz frequency. Artifacts were inspected visually and discarded. The software used for analysis was Matlab.

2.2. Feature extraction

2.2.1. EEG band power

The EEG signals are filtered with band-pass butterworth filter to extract four common frequency bands, delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (13–30 Hz). For each band of each channel, the size of 15,360 samples (1 min of signals) was selected and Welch method was applied to calculate power spectrum of EEG bands [11]. In the Welch method, time series is divided into segments (possibly overlapping) and then modified periodogram of all segments is averaged.

2.2.2. Detrended fluctuation analysis

DFA is a method for quantifying fractal scaling and correlation properties in the signal. The advantages of this method are that it distinguishes between intrinsic fluctuation generated by the system and those caused by external system [12]. In the DFA computation of a time series, $x(t)$ of finite length N, is integrated to generate a new time series $y(k)$ shown in (1).

$$y(k) = \sum_{i=1}^{k} [x(i) - \langle x \rangle] \tag{1}$$

where $\langle x \rangle$ is the average of x, is given by

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x(i) \tag{2}$$

Next, the integrated time series, $y(k)$ is divided into boxes of equal length and a least squares line is fit to the data of each box, represents by $y_n(k)$. Then, the time series $y(k)$ is detrended by subtracting the local linear fit $y_n(k)$ for each segment. The detrended fluctuation is given after removing the trend in the root-mean-square fluctuation

$$F(n)=\sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k)-y_n(k)]^2} \tag{3}$$

This computation is repeated for different box sizes (time scale) to characterize the relation between $F(n)$ and the box size n. A linear relation between logarithm of $F(n)$ and size of the box, indicates presence of power-law scaling: $F(n)\sim n^{\alpha}$. The scaling exponent, α, can be calculated as the slope of $\log F(n)$ versus $\log n$. This parameter represents the correlation properties of the time series.

2.2.3. *Higuchi*

Higuchi is an algorithm for measuring fractal dimension of time series and is used to quantify complexity and self-similarity of signal [13]. Higuchi method can be calculated in time domain directly so it is easy and fast. For time series $x[1]$, $x[2], \ldots, x[n]$, the higuchi fractal can be calculated as follow:

K new time series are constructed for $m=1, 2, \ldots, k$,

$$X_m^k=\left\{x[m], x[m+k], \ldots, x\left[m+\left\lfloor\frac{N-m}{k}\right\rfloor k\right]\right\} \tag{4}$$

where $\lfloor a \rfloor$ is integer part of a. The length of each k time series can be defined as:

$$L_m(k)=\frac{1}{k}\,\frac{\sum_{i=1}^{\lfloor(N-m/k)\rfloor}|x(m+ik)-x(m+(i-1)k)|\times(N-1)}{\lfloor(N-m/k)\rfloor} \tag{5}$$

An average length is computed for each time series having the same delay k, as the mean of k length $Ln(k)$ for $m=1, 2, \ldots, k$. The procedure is repeated for all k ranging from k_{min} to k_{max}. Yielding a sum of average length $L(k)$ for each k as indicated in

$$L(k)=\sum_{m=1}^{k}L_m(k) \tag{6}$$

The slope of least square linear fit in the curve of $\ln(L(k))$ versus $\ln(1/k)$ is estimated as fractal dimension. The method in [14] is used to determine k_{min} and k_{max} in this paper. A value of $k_{min}=1$ and $k_{max}=30$ were chosen for our study.

2.2.4. *Correlation dimension*

Dimension of a signal can give much information about nature of a system. Fractal dimension is one of nonlinear methods that is used to approximate dimension of a signal. Grssberger and Procaccia in 1983 [15], proposed an algorithm (GP) for computing fractal dimension that has become most widely used for estimating dimension of experimental data. The GP algorithm is based on embedding theory and phase space reconstruction. Assume a time series with N data points, $x=[x(1),x(2), \ldots, x(N)]$, by choosing time delay τ and embedding dimension m, a new m dimension vectors can be reconstructed as

$$X(i)=[x(i), x(i+\tau), \ldots, x(i+(m-1)\tau)] \quad i=1, 2, \ldots, N-(m-1)\tau \tag{7}$$

The probability that the points of the set are in the same cell of size r is presented by $C(r)$:

$$C(r)=\frac{2}{N(N-1)}\sum_{i\neq j}\theta(r-|X(i)-X(j)|) \tag{8}$$

where $C(r)$ is correlation integral, θ is the Heaviside step function which is defined as $\theta(x)=0$ for $x<0$ and $\theta(x)=1$ for $x>0$. The correlation dimension can be estimated from the slop of $\log(C(r))$ versus $\log r$ over linear region:

$$d=\lim_{r\to 0}\left[\frac{\log C(r)}{\log r}\right] \tag{9}$$

The procedure is repeated for increasing m. By increasing embedded dimension, the value of d will increase gradually until saturation. The saturation value of d is defined as GP correlation dimension.

In this study, correlation dimension is computed with the time delay τ that is determined by use of minimum mutual information method [16] and embedding dimension varying from 3 to 30.

2.2.5. *Lyapunov exponent*

Lyapunov exponent is a useful nonlinear dynamic measure that quantifies the exponential divergence or convergence of initially nearby trajectories in phase space. Also, LE can characterize instability or predictability of a system. A d-dimensional dynamical system has d lyapunov exponents but in most applications, largest lyapunov exponent (LLE) is computed instead of all exponents [17]. The positive LLE indicates that divergence among initially nearby trajectories grows exponentially in time and therefore, the nonlinear system is chaotic. The maximum lyapunov exponent, λ_1 for a dynamical system is defined as,

$$d_j(i)=d_j(0)e^{\lambda_1 i\Delta t} \tag{10}$$

where $d_j(i)$ is the mean Euclidian distance between two neighbor trajectories in phase space at time t_i and $d_{j(0)}$ is Euclidian distance between the jth pair of initially nearest neighbors after i time step. Taking the algorithm of both side of Eq. (11), we obtain:

$$\ln\ d_{j(i)}=\lambda_1(i\Delta t)+\ln\ d_i(0) \tag{11}$$

The maximum lyapunov exponent is calculated by the slop of linear fit to the average log divergence curve defined by,

$$y(i)=\frac{1}{\Delta t}\langle \ln\ d_j(t)\rangle \tag{12}$$

where $\langle\cdot\rangle$ is average over all value of j [18].

2.3. **Feature selection**

Feature selection is a very important step in pattern recognition. The idea of feature selection is to choose a subset of features that improve the performance of the classifier especially when we are dealing with high dimension data. Finding significant features that produce higher classification accuracy is an important problem.

In this study, we have 90 training data and features which made our data very sparse. In this case, appropriate feature selection method can significantly improve the accuracy of classification by selecting the most informative features. Genetic algorithm (GA) as the feature selection technique is used in this experiment. A vector of length k defines each chromosome 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, other feature reduction methods such as PCA are used. However, GA as the feature selection technique significantly improves the accuracy of classifiers.

2.4. *Classification*

Three classification techniques are used in our experiments. We also tested other classification techniques such as SVM with nonlinear kernel and Naive Bayes [19] using different types of features. However, our results with LDA, LR and KNN are superior to those obtained with other methods.

2.4.1. *LDA*

Linear discriminate analysis, known as Fisher's linear discriminant, is a statistical method that is commonly used for data classification. LDA finds linear combination of features to classify two or more classes [20]. The LDA function for two classes (c_1, c_2) is defined as:

$$g(x) = w^t(x) + w_0 \tag{13}$$

where x is the input feature vector, w is the weight vector and w_0 is threshold value. The goal is to find optimum w and w_0 based on the linear combination of features. Parameters w and w_0 are determined by maximizing the ratio of between-class variance to within-class variance to guarantee maximal separability. After optimizing the parameters, we classify the input instance as class c_1 if $g(x) > 0$, otherwise x is classified as class c_2.

2.4.2. *LR*

Logistic regression is another classification technique that we have used in our experiments. Logistic regression is used as a powerful technique for classification. The classification is done by fitting the training data to a logistic function [21]. Logistic function is a continuous function between 0 and 1 that defined as,

$$\pi(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n}} \tag{14}$$

where x is the input vector and β is the parameter vector.

Table 1 – Results of classification accuracy for power EEG bands.

Classifier	Feature			
	Delta (%)	Theta (%)	Alpha (%)	Beta (%)
KNN	66.6	70	70	66.6
LDA	66.6	70	73.3	70
LR	70	70	73.3	70

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 positive or negative classes.

2.4.3. *KNN*

KNN classifier is one of simple classification that is based on a distance function for pairs of observations. In KNN algorithm, k nearest training sample for a test sample is found. Then, test sample is assigned to particular class which is most frequent class among k nearest training data. This algorithm only requires an integer value for k and a metric to measure closeness [20].

3. Experimental results

3.1. *Result of classification based on power EEG bands*

Table 1 summarized the experimental results when power bands are applied to classifiers as input. In this table, each row shows the result of three classifiers. The number of selected features is about 8 in three classifiers.

Table 1 shows the highest accuracy, 73.3%, has been achieved when alpha power is applied to the classifiers as the input. For better comparison the bar chart of results has been provided in Fig. 1.

According to Fig. 1, alpha power has highest accuracy in discriminating depressed and normal groups in all classifiers. To study the power in alpha band of two hemispheres between depressed patients and healthy controls, t-test were carried out for the mean values of the EEG bands power in left and right hemispheres in each electrode. Alpha power band of five electrodes in left hemisphere (C3, P3, O1, F7, T3) and one electrode in right hemisphere (O2) differ significantly between

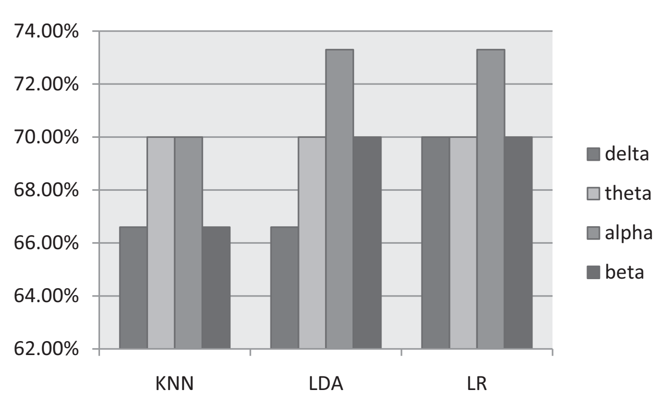

Fig. 1 – Comparison of classifiers' accuracy for power bands.

Table 2 – Results of classification for nonlinear features.

Classifier	Feature			
	DFA (%)	Higuchi (%)	Correlation dimension (%)	Lyapunov (%)
KNN	70	73.3	76.6	70
LDA	76.6	73.3	80	73.3
LR	76.6	76.6	83.3	73.3

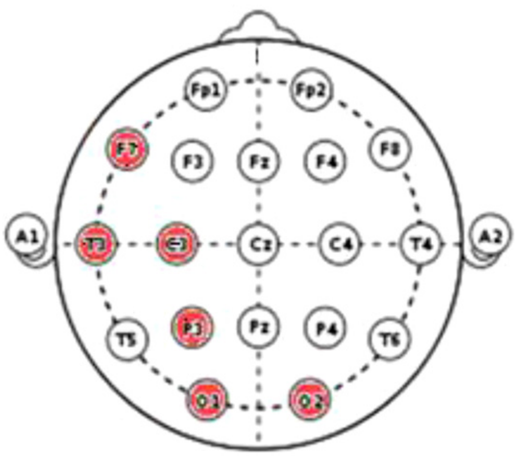

Fig. 2 – Location of electrodes with $p < 0.05$ between depressed patients and healthy subjects.

depressed patients and normal controls ($p < 0.05$). Fig. 3 shows the location of these electrodes.

3.2. *Results of classification based on nonlinear features*

In Table 2 the result of nonlinear features, DFA, higuchi, correlation dimension and large lyapunov exponent are given for three classifiers. It can be seen that correlation dimension has been achieved the highest accuracy, 83.3%, when is used as the input of LR classifier among all features. At last, about 9 features are selected by GA in each classifier when 19 features are used as input. Fig. 2 indicates bar chart of these results.

The DFA and higuchi have approximately the same accuracy in classifying two groups in three classifiers and accuracy has the lowest value when lyapunov exponent is used as the input of classifiers. In addition, Fig. 2 shows LDA and LR classifier have better accuracy in all features in compare to KNN classifier.

Table 3 – Results of classifiers accuracy with combined features in each group.

Classifier	Feature	
	Nonlinear features (%)	Power features (%)
KNN	80	73.3
LDA	86.6	76.6
LR	90	76.6

3.3. *Results of classification based on combining features*

In the second experiment, all features of each group are combined and applied to classifiers. GA is employed for selecting the best features and removing redundant ones. Table 3 shows the classifiers results on testing dataset.

Fig. 4 shows the accuracy of all classifiers for all nonlinear features and the accuracy of correlation dimension and LR classifier, as the accuracy among other nonlinear features and classifiers.

According to these results, the best accuracy, 90%, is achieved when all nonlinear features used as input of LR classifier. In addition, the accuracy of all classifiers is higher than the best result in Section 3.2, where the highest accuracy is related to correlation dimension. Fig. 5 summarized the results of accuracy for all power bands as features for three classifiers and the accuracy of alpha power band as the best result among other power bands.

In this experiment, the best accuracy is obtained by LR and LDA classifiers and all power bands as the input of classifiers.

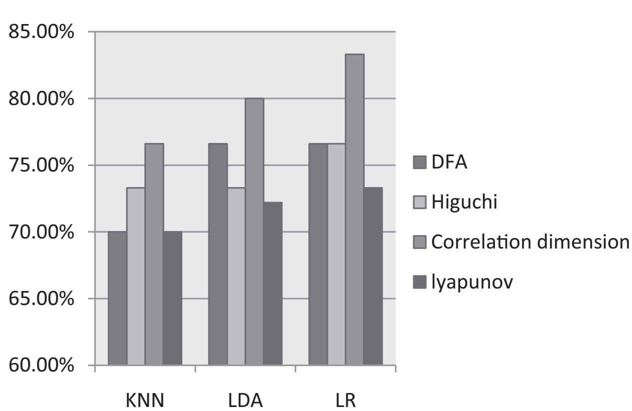

Fig. 3 – Comparison of classifiers accuracy for each nonlinear feature.

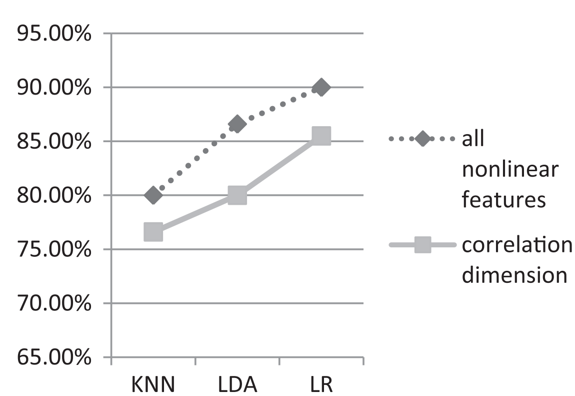

Fig. 4 – The accuracy of classifiers for combined nonlinear features and correlation dimension as input.

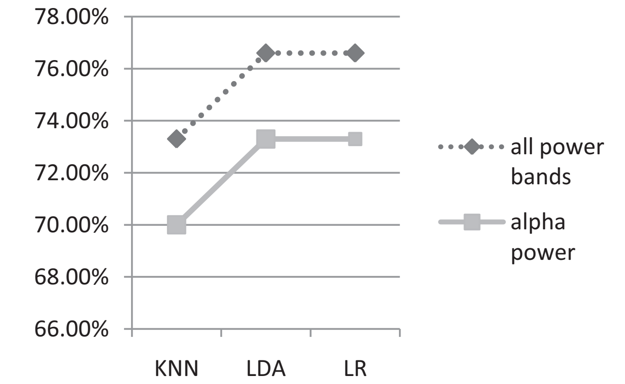

Fig. 5 – Accuracy of classifiers for combined power bands features and alpha power band as the input.

The results show the accuracy has no significant change when all power bands used as input in compare to best accuracy in Section 3.1, when alpha power is applied to LR classifier. For better comparison, results of nonlinear features and power bands for three classifiers are illustrated in Fig. 6.

Fig. 6 indicates that the accuracy of three classifiers are higher for all nonlinear features as the input in compare to power bands features. In this study, it is shown that the accuracy of all classifiers is significantly increased when nonlinear features are used as the input of classifiers. This result shows LR classifier can achieve the accuracy of 90% when all nonlinear features are applied to this classifier. The final subset of features that is selected by genetic algorithm in LR classifier is about 14 and 16 features for nonlinear features and power band respectively. In the last subset of nonlinear features that leads to best accuracy of 90%, the most features are selected from correlation dimension feature and the less is related to lyapunov exponent. According to the best accuracy in Table 2, related to correlation dimension, the accuracy of classifying depressed patients and normal persons is enhanced approximately by 6.7% when all features are combined. In this experiment, LR classifier has better results comparing to other classifiers. When all power bands are used as inputs, the accuracy has no considerable change in all classifiers in comparison with result of alpha power as input of classifiers. The highest accuracy of all power bands is 76.6% in

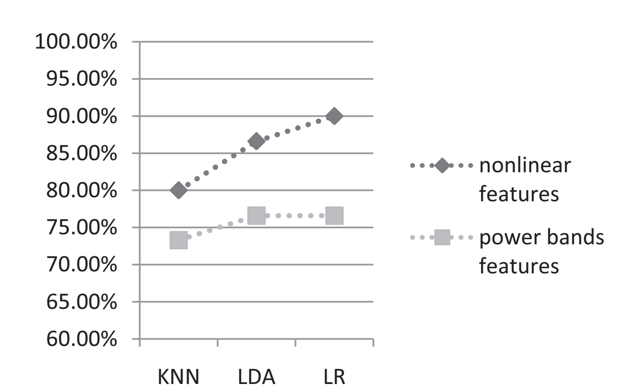

Fig. 6 – Comparison between accuracy of combined nonlinear features and combined power bands for three classifiers.

LR and LDA classifiers. According to these results, nonlinear features give much better results in classification of depressed patients and normal subjects.

4. Discussion

In this study, we analyzed resting EEG of 45 depressed patients and 45 normal subjects by power of EEG bands and four nonlinear features. In classification based on EEG bands power, the highest accuracy achieved by alpha power, suggesting that depressed patients and normal subjects differ in alpha band more significantly than other power bands like delta, theta and beta. Furthermore, we observed that alpha power had significant difference in T3, F7, O1, P3, C3 in the left hemisphere and O2 in the right hemisphere. The mean alpha powers in these electrodes were higher in depressed patients in comparison with normal subjects. These results were similar to the results obtained in [3]. In this study, Henriques and Davidson reported that left hemisphere of depressed patients had higher alpha power than left hemisphere of normal subjects. Also, this study showed alpha power was higher in left hemisphere of depressed patients than right hemisphere of this group.

In classifying depression patients and normal healthy subjects with nonlinear features, the highest accuracy was achieved when correlation dimension used as input of classification compared with DFA, higuchi and maximum lyapunov exponent. This experiment showed in discriminating depressed and normal persons, correlation dimension was a powerful feature for analyzing EEG signals. For further improved, we combined the features and used them as one feature vector for classifying. Combination of power bands had no considerable changes in accuracy of classifiers but nonlinear features improved the accuracy of classification significantly. The highest accuracy was 90% by combination of nonlinear features and LR classifier. The number of features, which GA selected for achieving this accuracy, was 14. Most of these features are related to correlation dimension. Compared to previous researches that based on linear and nonlinear analysis of depressed patients and normal subjects EEG signals, this study can achieve considerable accuracy, according to the fact that in this study, independent test is used. The prediction accuracy obtained from the unknown set shows the performance of classification and datasets more precisely. Knott et al. [5] reported the accuracy 91.3% for classifying 70 depressed patients and 23 normal subjects using linear features such as relative power and absolute power. In [8] wavelet entropy was used for analyzing EEG signals of 26 depressed patients and normal subjects and 80% accuracy was achieved in this experiment. Lee et al. [9] applied DFA to EEG of 11 depressed patients and 11 normal subjects and they obtained that DFA of depressed patients are higher than normal subjects but classification were not used in this study. In our study DFA value of both depressed patients and normal subjects were between 0.5 and 1 similar to the results in [9] but no significant difference were found between two groups.

In this study three classifiers were used. Among this classification, LR classifier performed better compared to LDA and KNN classifiers. This study suggests that more nonlinear features should be studied for analyzing EEG of depressed patients. Also, instead of using EEG in rest condition, EEG in different conditions and tasks can be recorded and analyzed for depressed patients and normal subjects. However, future investigation should focus on finding the regions of brain that involved in depression. Finally, an increase in EEG data would make it possible to validate the reliability of these features and classifiers.

5. Conclusion

In this paper we showed EEG signal can be a useful tool in studying depression and discriminating depressed patients and normal subjects. EEG signal of 45 depressed patients and 45 normal persons were recorded and linear features such as EEG bands power and nonlinear features such as DFA, higuchi, correlation dimension and maximum lyapunov exponent were extracted from EEG. Three well-known classifiers, KNN, LDA and LR were employed for classification. Leave-One-Out method was used for training data sets and the results were examined on testing datasets. Furthermore, GA was applied for selecting more informative and significant features for training datasets.

In power bands of EEG, highest classification accuracy was achieved by power of alpha band and in this band we can observe significant difference between electrodes in left hemisphere of depressed patients and normal subjects. These results indicated that depressed patients and normal subject differ in alpha band more than other bands especially in the left hemisphere.

Among nonlinear features correlation dimension had more ability to classifying two groups of depressed patients and normal subjects. Also, LR classifiers performed better than two other classifiers: LDA and KNN. In other experiment, with combination of nonlinear features, we can improve the accuracy of classification by 6.7% and obtain highest accuracy 90% in this study. These results confirmed that nonlinear features are potentially effective methods to analyze EEG signal.

Acknowledgment

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

REFERENCES

[1] World Health Organization, 2011, Available from: http://www.who.int/mental_health/management/depression/definition/en
[2] B.J. Sadock, V.A. Sadock, P. Ruiz, Kaplan & Sadock's Comprehensive Textbook of Psychiatry, 9th ed., Lippincott Williams and Wilkings, United States, 2008, p. 1647.
[3] J.B. Henriques, R.J. Davidson, Left frontal hypoactivation in depression, Journal of Abnormal Psychology 100 (1991) 535–545.
[4] V.P. Omel'chenko, V.G. Zaika, Changes in the EEG-rhythms in endogenous depressive disorders and the effect of pharmacotherapy, Journal of Human Physiology 28 (3) (2002) 275–281.
[5] V. Knott, C. Mahoney, S. Kennedy, K. Evans, EEG power, frequency, asymmetry and coherence in male depression, Psychiatry Research 106 (2) (2001) 123–140.
[6] C.J. Stam, Nonlinear dynamical analysis of EEG and MEG: review of an emerging field, Journal of Clinical Neurophysiology 116 (2005) 2266–2301.
[7] F. Fan, Y. Li, Y. Qiu, Y. Zhu, Use of ANN and complexity measures in cognitive EEG discrimination, in: Proceedings of the 27th Annual Conference of the IEEE Engineering in Medicine and Biology, 2005, pp. 4638–4641.
[8] Y. Li, Yi. Li, Sh. Tong, Y. Tang, Y. Zhu, More normal EEGs of depression patients during mental arithmetic than rest, in: Proceedings of NFSI & ICFBI, 2007, pp. 165–168.
[9] J.S. Lee, B.H. Yang, J.H. Lee, J.H. Choi, I.G. Choi, S.B. Kim, Detrended fluctuation analysis of resting EEG in depressed outpatients and healthy controls, Clinical Neurophysiology 118 (2007) 2486–2489.
[10] A. Beck, C. Ward, M. Mendelson, J. Mock, J. Erbaugh, An inventory for measuring depression, Archives of General Psychiatry (1961) 561–571.
[11] A. Alkan, M.K. Kiymik, Comparison of AR and Welch methods in epileptic seizure detection, Journal of Medical Systems 30 (6) (2006) 413–419.
[12] C. Peng, J. Hausdorff, A. Goldberger, Fractal Mechanisms in Neural Control: Human Heartbeat and Gait Dynamics in Health and Disease, Cambridge University Press, Cambridge, 2000.
[13] R. Esteller, G. Vachtsevanos, J. Echauz, B. Litt, A comparison of waveform fractal dimension algorithms, IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications 48 (2) (2001) 177–183.
[14] C. Go'mez, A. Mediavilla, R. Hornero, D. Aba'solo, A. Ferna'ndez, Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients, Medical Engineering and Physics 31 (2009) 306–313.
[15] P. Grssberger, I. Procaccia, Measuring the strangeness of attractors, Physica D: Nonlinear Phenomena 9 (1983) 189–208.
[16] M. Akay, Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling, Wiley-IEEE Press, New York, 2000.
[17] U. Parlitz, Nonlinear time series analysis, in: Proceedings of the NDES'95, Ireland, 1995.
[18] F. Mormanna, T. Kreuza, Ch. Riekea, R.G. Andrzejaka, A. Kraskovc, P. David, Ch.E. Elgera, K. Lehnertza, On the predictability of epileptic seizures, Journal of Clinical Neurophysiology 116 (2005) 569–587.
[19] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[20] A. Webb, Statistical Pattern Recognition, Oxford University Press Inc, New York, 1999.
[21] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.