Application of Principal Component Analysis and Neural Network Methods for Data reduction and Parkinson Disease Detection

Ayşegül Güven[1], Fatma Latifoğlu[1], Hayriye Hisar[1]
[1]Erciyes University, Faculty of Engineering, Department of Biomedical Engineering, 38039 Kayseri, Turkey
aguven@erciyes.edu.tr, flatifoglu@erciyes.edu.tr, hayriye_hisar@yahoo.com

ABSTRACT: *Data reduction techniques are proposed extensively in literature with different perspectives and understanding. One potential area of application for data reduction techniques is on Parkinson's disease detection. In the current work, we have used the combination of Principles component analysis (PCA) and Artificial Neural Networks (ANN) to distinguish healthy and PD using speech signals data set. We have found that the classification rates obtained was high and significant. The final stage, results are categorized in to two divisions, viz., 'healthy' and 'diseased'. Testing results were compatible with the expected results that are derived from the physician's direct diagnosis. The experimental results obtained reveal that the proposed method is capable of developing newer intelligent assistance diagnosis systems.*

Keywords: Artificial neural network, Parkinson disease detection, Principal component analysis

Received: 14 March 2010, Revised 2 April 2010, Accepted 5 April 2010

1. Introduction

Parkinson's disease (PD) is a degenerative illness of the brain that usually destroys the motor skills, speech, and other functions. The substantia nigra area of brain degenerates this condition. PD is called motor system defectiveness's, which are the result of the loss of dopamine-producing brain cells. Trembling in hands, arms, legs, jaw, and face; rigidness of the limbs and trunk; slowness of movement are the primary symptoms of PD and impaired balance and coordination. As these symptoms become more significant, patients may be unable to walking, talking, or completing other simple daily activities. PD occurs about 1.5 times more common in men than in women [1].

Physical observations and questioning of the patients are usually basis diagnosis of PD. Though sometimes scanning methods are used.

Positron emission tomography (PET) and single photon emission tomography (SPECT) have been used to monitorize changing in dopaminergic function in PD and these methods show decreasing of dopamine neurons in the striatum of brain. Parkinson's Disease measurement tool were using the Unified Parkinson Disease Rating Scale (UPDRS) to assist the diagnose [2.- 4]. The PD can be difficult to early diagnose exactly. PD symptoms increase gradually in time. There numbers of studying about PD using computational methods. Recent studies in literature involve image processing methods and classification using UPDRS. Acton,Paul D. and coworkers studied "Artificial neural network classifier for the diagnosis of PD using image processing methods". Stephen L. Smith and coworkers studied "an immune network inspired evolutionary algorithm for the diagnosis of PD". C. Okan Sakar and coworkers studied "telediagnosis of PD using measurements of dysphonia". Athanasios Tsanas and coworkers studied "accurate telemonitoring of PD progression by non-invasive speech tests". Marius Ene studied "neural network-based approach to discriminate healthy people from those with PD" using voice data set [5-8].

To the best of the authors knowledge no research has been published on an integrated and simultaneous application of PCA-ANN architecture for voice data set to classify PD.

Currently, automated detection of disease techniques based on Artificial Intelligence are needed to increase the diagnosis accuracy of illness and to help clinician make accurate decisions.

In this study, we have facilitated Artificial Neural Network (ANN) that will not only simplify the diagnosis but also enable the physician to make a quicker judgment about the existence of PD in confidence.

Our primary research motivation was to advance the research of diagnosing PD from voice measurements. We have employed the some features of medical data set from voice recordings, then Principles component analysis (PCA) for the aim of data reduction and Artificial Neural Networks (ANN) in order to distinguish between PD and healthy subjects [9].

2. Methods

2.1. Patients Data Set

The data set were obtained, from Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. The original paper "Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease" named has been published in IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, named with [10].

In this study, the data consist of 195 continuously vowel phonations from 31 male and female subjects, of which 23 were patient with PD. Diagnosing time is between 0 to 28 years, and the ages of the subjects is between from 46 to 85 years (mean 65.8, standard deviation 9.8).

The feature information of this data set consists of the following technical details:

1. Average vocal fundamental frequency
2. Maximum vocal fundamental frequency
3. Minimum vocal fundamental frequency
4. Voice Jitter as a percentage
5. Voice Absolute jitter in microcesonds
6. Relative Amplitude Perturbation,
7. Five Point Period Perturbation Quotient
8. Average absolute difference of differences between cycles divided by the average period.
9. Local shimmer
10. Local shimmer in decibels
11. Three point amplitude perturbation Quotient
12. Five point amplitude perturbation Quotient
13. 11. point amplitude perturbation Quotient
14. Average absolute difference between consecutive differences between the amplitudes of consecutive periods
15. Noise to harmonics ratio
16. Harmonics to noise ratio
17. Recurrence period density entropy
18. Detrended fluctuation analysis
19. Spread1
20. Correlation dimension
21. Spread2
22. Pitch period entrphy

2.2. Feature Selection-Principal Component Analysis

The PCA method was used to make an ANN system more efficient For this aim, before classifying with ANN, PCA method was used for data reduction of feature vector. It describes the data set in the sense of its variance. Each principal component identifies a percentage of the total variance of a data set and elaborates loadings or weights that each variate contributes to this variance.One of the main advantage of PCA technique is founding patterns in the data, and data has been compressed by reducing the number of dimensions, without much loss of information [11,12].

In this paper, the use of PCA for the characterization and feature selection of medical data set. Principal components calculate using eigenvectors and eigenvalues of covariance matrixes or correlation matrix.

$$\text{Cov (Xij)} = \frac{1}{n-1}\sum_{k=1}^{n}(X_{ik}-\bar{X}_i)(X_{jk}-\bar{X}_j) \quad (1)$$

$$i=1, 2, \ldots, m, j=1,2,\ldots\ldots, m$$

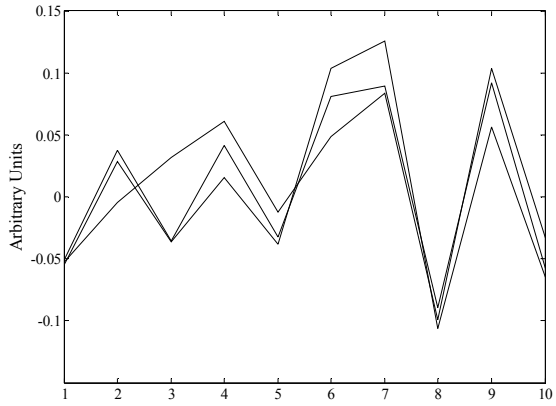
Where, n is the total number of signals and m is the signal dimension. $\overline{X}_i$ is the average signal of the population. First eight principal component were computed by the solution of described as;

$$Cwp=\lambda p*wp \qquad p = 1,2,..,8. \tag{2}$$

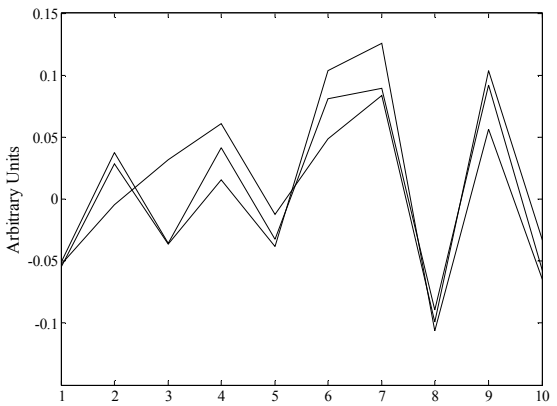
Where C is the covariance matrix, wp is the pth principal component (eigenvector) and λp is the corresponding eigen value. The λp are positive values proportional to the fraction of the total variance calculated using each component wp which have the significant property of forming an orthogonal set. The coefficients of the principal components of qth data feature set are then given by

$$a_{qp} = \sum_{i=1}^{m} x_{qi} w_{ip} \qquad p=1,2,3.\ q=1,2\ldots,n. \tag{3}$$

After PCA of data vector, each feature vector was represented a vector consists of 10 samples as seen figure1 for three healthy and PD subjects. This feature selected vector applied to ANN to classify PD and healthy subjects.



PD Feature Set for ANN Input
(a)



Healthy Feature Set for ANN Input
(b)

Figure 1. Examples of feature data set for ANN

2.3. Artificial Neural Networks

The ANN has been used in a number of different ways in medicine recently. The advantages of ANNs are abling to generalize, adapting to signal distortion and noise. In our study Levenberg-Marquartd training algorithm in the feed forward multilayer perceptron was used for ANN architecture. This is standard direction for resolving pattern recognition problems where superintended learning with back-propagation of errors is implemented. During supervised learning, an ANN is trained to correlate the input vectors with the related output vectors. The network is iterated for single and double hidden layers with combinations of one to twenty neurons in each layer.T he values of the internal weights and biases are adjusted so as to minimize the error between the actual output of the network and the desired output during training[13,14].

ANN underwent supervised learning to perform successful pattern recognition of the PD selected feature sets. The train input data set consisted of 163 subjects and 24 of train data set is healthy subjects, while the test data set was made of 8 healthy and 24 PD subjects. The best results were accomplished with the combination of double hidden layers consisting of nineteen and fiveteen neurons consecutively. Each hidden layer sigmoid function and output layer linear function was used. Weight learning functions is the Levenberg-Marquardt algorithms and the performance function is mean square error. We trained ANN till mean square error achieved 0.001. In the present study, 10-fold cross validation was used to evaluate the accuracy of each PD data sets. Training set was randomly divided into 10 test partitions of about equal size. For each of the 10 training state, one set of the test partitions was used as the test set, and the samples in other partition were used to train the neural network.

3. Results and Discussion

In this study the problem of improving the classification accuracy is attacked from two different ways. 1) designing an appropriate feature space using PCA and 2) designing a classifier that can accurately classify PD using ANN. We studied the feature set (195 voice feature) randomly selected as the test and the train groups by the ANN and the four layered MLP structure have been built to simplify the diagnosis and enable the physician to make a judgment about the existence of PD.

Study performance was determined with comparison of correct classification rate, testing mean square error, specificity and sensitivity for each ANN input data sets.

In our study the ANN structure, for 10 fold cross validation the testing mean square error of 0.0317 was observed for our optimized MLP feed forward network with a training mean square error of 2.6668*10-7. The final stage, results are categorized as healthy and diseased. There has been one false classification in the positive group, while 23 subjects were correctly recognized as PD. In the healthy group, any subject was misclassified and 8 subjects were accurately classified as healthy. It is seen in Figure 2 the desired network example output was given pattern characterized as diseased except one which has the value as healthy.

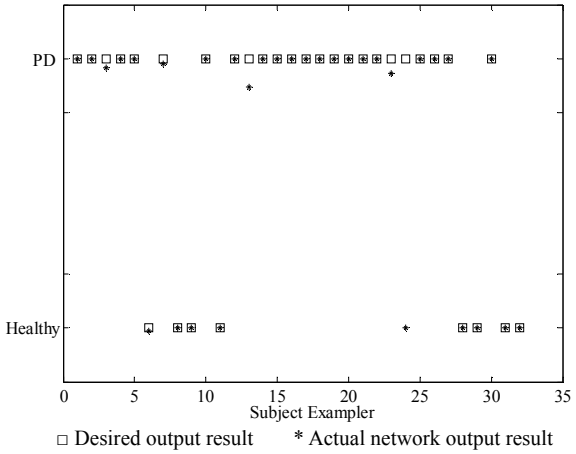


Figure 2. ANN Output results for third fold

Overall results 96.87% correct classification was achieved, whereas 1 false classifications have been observed for the group of 32 people in total. With these results, this network has about 100% sensitivity and specificity is calculated to be 88.88% (Table 1).

Statistical parameters	Values (%)
Specificity	88.88
Sensitivity	100
Correct Classification	96.87

Table 1. Performance results

ANN method has a generally robust structure, is tolerant of faults and noise, able to generalize well and capable of solving nonlinear problems.

The result of this study leads to the conclusion that all two key components of the PCA-ANN methodology are important for improving the accuracy of PD classification.

4. Acknowledgement

The data set were obtained, from Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado.

References

[1] Lang, A.E., Lozano, A.M. (1998). Parkinson's disease – First of two parts, *New England Journal Medicine*, 339.1044-1053.

[2] Schrag, A., Ben-Schlomo, Y., Quinn, N (2002). How valid is the clinical diagnosis of Parkinson's disease in the community. *Journal of Neurology, Neurosurgery Psychiatry.* 73. 529-535.

[3] Booij, J., Tissingh, G., Winogrodzka, A., Royen, E A (1999). Imaging of the dopaminergic neurotransmission system using single-photon emission tomography and positron emission tomography in patients with Parkinsonism, *Eur. J. Nucl. Med.* 26. 171–82.

[4] Martínez-Martín, P. et all "Unified Parkinson's disease rating scale characteristics and structure" *Movement Disorders*, vol 9(1), pp.76–83, 2004

[5] Acton, P D et al (2006). Artificial neural network classifier for the diagnosis of Parkinson's disease using [99mTc]TRODAT-1 and SPECT", *Physics in Medicine and Biology*, V. 51 (12) p. 3057-3066 .

[6] Stephen, L. S et al (2008). An immune network inspired evolutionary algorithm for the diagnosis of Parkinson's disease", *Biosystems -- Seventh International Workshop on Information Processing in Cells and Tissues - IPCAT 2007,* V. 94. 1-2, p. 34-46.

[7] Sakar, O et al (2009). Telediagnosis of Parkinson's Disease Using Measurements of Dysphonia, *Journal of Medical Systems*.

[8] Ene, M. (2008). Neural network-based approach to discriminate healthy people from those with Parkinson's disease, *Annals of the University of Craiova, Math. Comp. Sci. Ser.*V. 35, p. 112–116.

[9] Miller, A.S., Blott, B.H., Hames, T.K. (1992). Review of neural network applications in medical imaging and signal processing, *Med. Biol. Eng. Comput*., 30. 449-464.

[10] Little, M A. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, *Biomedical Engineering, IEEE Transactions* on, In Press.

[11] Smith, L.I. (2002). A tutorial on Principal Components Analysis, http://kybele.psych.cornell.edu/~edelman/Psych-465-Spring- 2003/PCA-tutorial. Accessed 26 February 2002

[12] Mutihac, R., Van Hulle, M. M (2004). Comparison Of Principal Component Analysis And Independedent Component Analysis For Blind Source Separation", *Romanian Reports In: Physics*, 56 (I) 20-32.

[13] Haykin, S. (1994). Neural networks: a comprehensive foundation, New York: Macmillan College Publishing Company Inc., New York.

[14] Simpson, P.K. (1989). Artificial Neural Systems, Pergamon Press.