Diagnosis of Parkinson's Disease using Principal Component Analysis and Boosting Committee Machines

Indira Rustempasic, Mehmet Can

International University of Sarajevo, Faculty of Engineering and Natural Sciences
Hrasnicka Cesta 15, 71000 Sarajevo, Bosnia and Herzegovina
irustempasic@ius.du.ba, mcan@ius.edu.ba

Parkinson's disease (PD) has become one of the most common degenerative disorder of the central nervous system. In this study, our main goal was to discriminate between healthy people and people with Parkinson's disease. In order to achieve this we used artificial neural networks, and dataset taken from University of California, Irvine machine learning database, having 48 normal and 147 PD cases. We examine the performance of neural network systems with back propagation together with a majority voting scheme. In order to train examples we used boosting by filtering technique with seven committee machines, and principal component analysis is used for data reduction. The experimental results have demonstrated that the combination of these proposed methods has obtained very good results with correct positive value of 92% on the classification of PD.

Index Terms—**Artificial neural network, boosting by filtering, committee machine, Parkinson's disease, Principal Component Analysis**

I. INTRODUCTION

Parkinson's disease (PD) is one of the most common neurodegenerative disorder. It accounts for a variety of motor and non-motor deficits which are the result of the loss of dopamine-producing brain cells. Parkinson's primarily affects neurons in the area of the brain called the substantia nigra. These cells normally produce dopamine, a chemical (neurotransmitter) that transmits signals between areas in the brain that, when working normally, coordinate smooth and balanced muscle movement. As Parkinson Disease progresses, the amount of dopamine produced in the brain decreases, leaving a person unable to control movement normally. The exact cause of this deterioration is not known, but scientists are doing a lot of research to look for the answer.

The four primary symptoms of PD are tremor, or trembling in hands, arms, legs, jaw, and face; rigidity, or stiffness of the limbs and trunk; bradykinesia, or slowness of movement; and postural instability, or impaired balance and coordination. As these symptoms become more pronounced, patients may have

difficulty walking, talking, or completing other simple tasks. PD usually affects people over the age of 50. However, it can start earlier, and it is more common in men than in women. Early symptoms of PD are subtle and occur gradually. In some people the disease progresses more quickly than in others. As the disease progresses, the shaking, or tremor, which affects the majority of PD patients may begin to interfere with daily activities. Other symptoms may include depression and other emotional changes; difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions.

The disease can be difficult to diagnose accurately, particularly in the early stages of the disease when symptoms resemble other medical conditions, and misdiagnosis occurs occasionally. There are currently no blood or laboratory tests that have been proven to help in diagnosing PD, and the prognosis depends on the patient's age and symptoms. The diagnosis is based on the medical history and neurological examination conducted by interviewing and observing the patient. Brain scans or laboratory tests may be used to help doctors exclude other medical conditions that produce symptoms similar to those of Parkinson's disease.

After a Parkinson's diagnosis, Parkinson's disease treatments are given to help relieve symptoms. At present, there is no cure for PD, but a variety of medications considerably reduce the typical Parkinson movement disorders (i.e. tremor at rest, rigidity, akynesia and postural instability). In some cases, surgery may be appropriate if the disease doesn't respond to drugs. Medical treatment or surgery may just alleviate some of the most troubling symptoms, but there is no causal cure now available, and early diagnosis is critical for maximizing the effect of treatment and improving the quality of the patient's life.

Current research programs are trying to respond how the disease progresses and to develop new drug therapies. Scientists looking for the cause of PD continue to search for possible environmental factors, such as toxins, that may trigger the disorder, and study genetic factors to determine how defective genes play a role. Other scientists are working to develop new protective drugs that can delay, prevent, or reverse the disease. Recently, a group of experts found some features in the voices of the people with Parkinson's disease that can be used as discriminatory measures to differentiate those who have the disease from those who do not. Max Little, from the University of Oxford, has been developing software that learns to detect differences in voice patterns, in order to spot distinctive clues associated with Parkinson's. He used machine learning. He was collecting a large amount of data when he knew if someone has the disease or not and trained the database to learn how to separate out the true symptoms of the disease from other factors, and he succeeded to develop an algorithm to detect changes in voice purely associated with Parkinson's. He introduce a new

measure of dysphonia, pitch period entropy (PPE), which is robust to many uncontrollable confounding effects including noisy acoustic environments and normal, healthy variations in voice frequency. He collected sustained phonations from 31 people, 23 with PD, and using a kernel support vector machine (SVM) got overall correct classification performance of 91.4%.

So far, several studies have been reported focusing on PD diagnosis. In these studies different methods were applied to the given problems. Consequently, different of novel methods have been proposed. In generally, most of these methods are based on appliaction of some neural networks.

In [25], the Multi-Layer Perceptron (MLP) with Back-Propagation learning algorithm were used to classify to effective diagnosis Parkinsons disease(PD). The Artificial neural networks was used to classify the diagnosis of patients.The accuracy in training data set was 82.051% and in the validation data set 83.333%. Gil and Johnson [1] propose a hybrid system combining ANN (Artificial Neural Networks) and SVM (Support Vector Machines) classifiers. These two classifiers, which are widely used for pattern recognition, can provide a good generalization performance in the diagnosis task. They showed a high degree of certainty, above 90%. Furthermore, some of the parameters reach very high accuracy such as "Sensitivity" and "Negative predictive value" with 99.32% and 97.06% respectively. Also, in [5] two types of the ANNs were used for classification: Multilayer Perceptrons (MLP) and Radial Basis Function (RBF) Networks. The other method is Adaptive Neuro-Fuzzy Classifier (ANFC) with linguistic hedges, which is also used for feature selection from the dataset. Adaptive Neuro-Fuzzy Classifier with linguistic hedges gave the best recognition results with %95.38 training and %94.72 testing classifying performance. An evolutionary algorithms and RBF (Radial Basis Function) neural networks can be successful in diagnosis of Parkinsons's disease. Three different evolutionary algorithms; genetic algoritm, particle swarm optimization and artificial bee colony algorithm (ABC) are used for training different structures of RBF neural networks.. Artificial Bee Colony based RBF neural networks with four neurons yielded the best score either in terms of accuracy or speed. The best simulation result gained 93.7% and 91.2% classification accuracy respectively as train and test classification stages for Parkinson's disease diagnosis [20]. Paper [21] deals with the application of some probabilistic neural network (PNN) variants to discriminate between healthy people and people with Parkinson's disease. Three PNN types have been used in this classification process, related to the smoothing factor search: incremental search (IS), Monte Carlo search (MCS) and hybrid search (HS). The concrete application has provided diagnosis accuracies ranging between 79% and 81% for new, undiagnosed patients. A comparison of multiple classification methods for diagnosis of Parkinson disease is done in [19]. Four independent classification schemas were applied

and a comparative study was carried out. These are Neural Networks, DMneural, Regression and Decision Tree respectively. Various evaluation methods were employed for calculating the performance score of the classifiers. According to the application scores, neural networks classifier yields the best results. The overall classification score for neural network is 92.9%.

As it is shown, there are many methods that are in use today in medical research and public health for recognizing Parkinson's disease. Generally, in the field of health care, neural networks play an increasingly important role. In recent years, most of the researchers have been using neural networks for detecting PD in the early stage.

This paper deals with the application of Neural Networks with back propagation together with Principal Component Analysis to a medical dataset concerning PD with the aim of automatically classify patients in PD or non-PD depending on their medical attributes. In order to test the performance and efficiency of the proposed method, the classification accuracy, sensitivity and specificity were used. The paper is organized as follows. Section 1 defines the related works carried out in the Parkinson Disease area. Section 2 deals with the Parkinson Dataset that is used in this research work. Section 3 gives an overview of the Artificial neural networks and explain architecture used for classification of the patient as PD or non-PD. Section 4 and 5 describe committee machine and Principal Component Analysis, respectively. Section 6 is dealt with the experimental results of the Algorithms. And Section 7 concludes research paper and proposes the future work.

II. PARKINSON DATASET

The Parkinson database used in this study is taken from the University of California at Irvine (UCI) machine learning repository [15]. The features of dataset are given in Table 1.

TABLE I
TABLE DESCRIBING THE ATTRIBUTES
THAT ARE REMOVED AFTER APPLYING THE CORRELATION FILTER

No.	Attribute Name	Rem.	Description
1	MDVP:Fo(Hz)	Yes	Average vocal fundamental frequency
2	MDVP:Fhi(Hz)	Yes	Maximum vocal fundamental frequency
3	MDVP:Flo(Hz)	Yes	Minimum vocal fundamental frequency
4	MDVP:Jitter(%)	Yes	Variation in fundamental frequency
5	MDVP:Jitter(Abs)	No	Variation in fundamental frequency
6	MDVP:RAP	Yes	Variation in fundamental frequency
7	MDVP:PPQ	Yes	Variation in fundamental frequency
8	Jitter:DDP	No	Variation in fundamental frequency
9	MDVP:Shimmer	Yes	Measures of variation in amplitude
10	MDVP:Shimmer(dB)	Yes	Measures of variation in amplitude
11	Shimmer:APQ3	Yes	Measures of variation in amplitude
12	Shimmer:APQ5	Yes	Measures of variation in amplitude
13	MDVP:APQ	No	Measures of variation in amplitude
14	Shimmer:DDA	No	Measures of variation in amplitude
15	NHR	No	Ratio of noise to tonal components
16	HNR	No	Ratio of noise to tonal components
17	status	No	The status of the patient (1)-Parkinson's Disease, (0)-Healthy
18	RPDE	No	Dynamic complex measurement
19	DFA	No	Signal fractal scaling exponent
20	spread1	Yes	Non-linear measure of fundamental frequency
21	spread2	Yes	Non-linear measure of fundamental frequency
22	D2	No	Dynamic complex measurement
23	PPE	No	Non-linear measure of fundamental frequency

The PD dataset was created by Max little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado. The data consists of 195 sustained vowel phonations from 31 male and female subjects, of which 23 were diagnosed with PD [5, 7, 8]. The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years. Averages of six phonations were recorded from each subject, ranging from 1 to 36 s in length. There are 195 instances comprising 48 normal and 147 PD cases in the dataset. The essential aim of processing the data is to discriminate healthy people from those with PD, according to the "status" attribute which is set to 0 for healthy and 1 for people with Parkinson's disease, which is a two-decision classification problem. Little applied a correlation filter and of these 23 attributes 12 are removed. Each correlation coefficient, which is less than 0.95 is considered not to contribute to classification accuracy, thus the attribute is removed. A total of 11 attributes are kept after the correlation filter has been applied. Table 2 indicates which features are kept. First 10 are used as inputs to the classifiers. The data set used for this implementation are also described in detail in [2], and at the UCI website [15].

TABLE II FEATURES USED IN THIS PAPER

Index	Attribute Name	Description
1	MDVP:Jitter(Abs)	Variation in fundamental frequency
2	Jitter:DDP	Variation in fundamental frequency
3	MDVP:APQ	Measures of variation in amplitude
4	Shimmer:DDA	Measures of variation in amplitude
5	NHR	Ratio of noise to tonal components
6	HNR	Ratio of noise to tonal components
7	RPDE	Dynamic complex measurement
8	DFA	Signal fractal scaling exponent
9	D2	Dynamic complex measurement
10	PPE	Non-linear measure of fundamental frequency
11	status	The status of the patient (1)-Parkinson's Disease, (0)-Healthy

However a 10 dimensional input data vector would be computationally expensive, therefore Principle Component Analysis (PCA) is used to reduce the dimensionality of the input data without losing accuracy of data representation.

III. ARTIFICIAL NEURAL NETWORKS

Brain is highly complex, nonlinear and parallel computer capable to perform certain computations (e.g., pattern recognition, perception, and motor control) many times faster than the fastest digital computer in existence today. A neural network is an artificial representation of the human brain that tries to simulate its information processing. It is an interconnected group of artificial neurons which may share some properties of biological neural networks. A neural network derives its computing power through its parallel distributed structure and its ability to learn and generalize. These two capabilities make it possible for neural networks to solve complex problems by decomposing problem into a number of relatively simple tasks, and neural networks are assigned a subset of the tasks that match their inherent capabilities. It is important to recognize, however, that we have a long way to go (if ever) before we can build a computer architecture that mimics a human brain.

A. Architecture

A neural network consists of a certain number of layers, and each layer contains a certain number of units. There is an input layer, an output layer, and one or more hidden layers between the input and the output layer. In general, we may identify three different classes of network architectures [35]:

Recurrent networks unlike others, has at least one feedback loop. For example, a recurrent network may consist of a single layer of neurons with each neuron feeding its output signal back to the inputs of all the other neurons. The presence of the feedback loops, has impact on the learning capability of the network and on its performance

Single-Layer Feedforward networks is the simplest form of a layered network, where we have an input layer of source nodes that projects onto an output layer of neurons, but not vice versa. In other words, this network is strictly a feedforward, and single refers to one output layer of computation nodes.

Multilayer Feddforward Networks has one or more hidden layers. The function of hidden neurons is to intervene between the external input and the network output in some useful manner. In this kind of networks, there are no connections from any of the units to the inputs of the previous layers (no feedback information) nor to other units in the same layer, nor to units more than one layer ahead. Every unit only acts as an input to the immediate next layer. Obviously, this class of networks is easier to analyze theoretically than other general topologies because their outputs can be represented with explicit functions of the inputs and the weights. The architectural graph in Figure1 illustrates the layout of a multilayer feedforward neural network for the case of a two hidden layer.
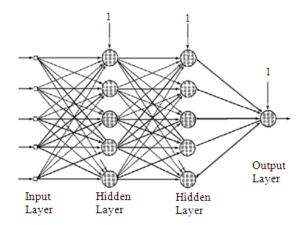

Fig.1 Feedforward neural network with two hidden layers

Each neuron (see figure 2) in the input and hidden layers is connected to all neurons in the next layer by weighted connections. These neurons (see figure 2) compute weighted sums of their inputs and adds a threshold. The resulting sums are used to calculate the activity of the neurons by applying a sigmoid activation function.
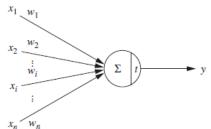

Fig.2 A neuron in the hidden or output layer in the MLP.

This process is defined as follows:

$$y = F\left(\sum w_i x_i - t\right) \qquad (1)$$

where variables $x_1, x_2, \ldots, x_i, \ldots, x_n$ are the inputs, t is a threshold, variables $w_1, w_2, \ldots, w_i, \ldots, w_n$ are the weights associated with the impulses/inputs, signifying the relative importance that is associated with the path from which the input is coming and F is the activation function of the neuron, and y is the output. In a neural networks each neuron has an activation function which specifies the output of a neuron to a given inputs. Activation functions for the hidden units are needed to introduce non-linearity into the networks. For classification tasks antisymmetric sigmoid tangent hyperbolic function is a common choice of the activation function. It is defined as:
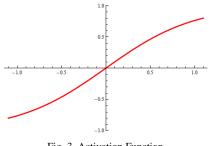
$$F(t) = \frac{1}{1 + e^{-t}} \qquad (2)$$

Fig. 3. Activation Function

The architecture used in this study is the multilayer feed-forward network consisting of an input layer and hidden layers, and an output layer (healthy or ill) which represents the classification result. For training multi-layer feed-forward network back-propagation is used. Back propagation algorithm is a generalization of the least mean squared algorithm that minimizes the mean squared error between the desired and actual output of the network by modifying its network weight. A back propagation algorithm uses supervised learning, where the network is trained using the known data of inputs and desired output. After the data is trained, the network weights are frozen to be used later for new input samples to compute an output values.

IV. COMMITTEE MACHINES

When a task is too complex, the best thing to do is to divide it into smaller and simpler tasks and combine solutions in order to solve the whole task. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces. The combination of experts is said to constitute a committee machine. Basically, it fuses knowledge acquired by experts to arrive at an overall decision that is supposedly superior to that attainable by any one of them acting alone. So, committee machines are expected to produce better results than using any expert individually, because they combine knowledge from several experts to reach a decision. Committee machines can be built in two different ways, using static and dynamic structures. [35]

In *dynamic structures* input data is involved in each expert output combination mechanism to generate the global output. This category includes 2 methods: mixture of experts, where answers from experts are nonlinearly linked by only one gating network; hierarchical mixture of experts, where answers from experts are nonlinearly linked by several gating networks arranged in a hierarchical fashion.

In *static structures*, combination mechanism between experts does not depend on input data. This category can also be classified in: ensemble averaging, where global output is a result of linear combination of each specialist outputs; boosting, where a weak learning algorithm can learn how to reach a higher accuracy. In boosting machine the experts are trained on data sets with entirely different distributions. Boosting can be implemented in three different ways: boosting by filtering, subsampling and reweighting. In this paper we used boosting by filtering technique.

A. Boosting by filtering

In boosting by filtering, the committee machine consists of three experts, arbitrarily labeled as first, second and third. The three experts are individually trained as follows: the first expert is trained on a set consisting of N1 examples. The trained first expert is used to filter another set of examples by proceeding in the following manner: flip a fair coin to simulate a random guess. If the result is heads, pass new patterns through the first expert and discard correctly classified patterns until a pattern is misclassified. That misclassified pattern is added to the training set for the second expert. If the result is tails, do the opposite. Specifically, pass new patterns through the first expert and discard incorrectly classified patterns until a pattern is classified correctly. That correctly classified pattern is added to the training set for the second expert. Continue this process until a total N1 examples has been filtered by the first expert. This set of filtered examples constitutes the training set for the second expert. Once the second expert has been trained in the usual way, a third training set is formed for the third expert by proceeding in the following manner: pass a new pattern through both the first and second experts. If the two experts agree in their decision, discard that pattern. If, on the other hand, they disagree, the pattern is added to the training set for the third expert. Continue with this process until a total of N1 examples have been filtered jointly by the first and second experts. This set of jointly filtered examples constitutes the training set for the third expert. The third expert is then trained in the usual way, and the training of the entire committee machine is thereby completed. The three-point filtering procedure is illustrated in Figure2.
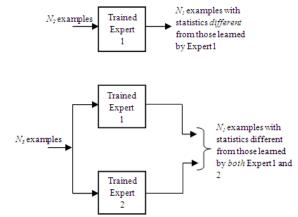

Fig.4 Illustration of boosting by filtering, (a) Filtering of examples performed by Expert 1, (b) Filtering of examples performed by Expert 2 and 3

Let N_2 denote the number of examples that must be filtered by the first expert to obtain the training set of N_1 examples for the second expert. Let N_3 denote the number of examples that must be jointly filtered by the first and second experts to obtain the training set of N_1 examples for the third expert. With N_1 examples also needed to train the first expert, the total size of data needed to train the entire committee machine is $N_4=N_1+N_2+N_3$. However, the computational cost is based on $3N_1$ examples because N_1 is the number of examples actually used to train each of the three experts. We may therefore say that the boosting algorithm described here is indeed smart in the sense that the committee machine requires a large set of examples for its operation, but only a subset of that data set is used to perform the actual training.

To evaluate the performance of the committee machine on test patterns, simple voting scheme was used in this paper. If the first and second experts in the committee machine agree in their respective decision, that class label is used. Otherwise, the class label discovered by the third expert is used.

V. PRINCIPAL COMPONENT ANALYSIS

Principal Components Analysis (PCA) is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. PCA is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, principal component analysis is used for data reduction, by compressing the data or mapping the data into a lower dimensional space. In general, reduction of dimensionality will be accompanied by a loss of some of the information. Thus, the main goal in dimensionality reduction is to preserve as much of the relevant information as possible.

A. Theory of Principal component Analysis

Principal component analysis is useful if we have obtained data on a number of variables (possibly a large number of variables), and we believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another. Because of this redundancy, it should be possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables.

Multivariate statistics deals with the relation between several random variables. The sets of observations of the random variables are represented by a multivariate data matrix X. Multivariate statistics deals with the relation between several

random variables. The sets of observations of the random variables are represented by a multivariate data matrix X,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ x_{31} & x_{32} & \cdots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (3)$$

Each column vector u_k represents the data for a different variable. If c is an $p \times 1$ matrix, then

$$X_c = c_1 \begin{bmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{n1} \end{bmatrix} + c_2 \begin{bmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{n2} \end{bmatrix} + \dots + c_p \begin{bmatrix} x_{1p} \\ x_{2p} \\ x_{3p} \\ \vdots \\ x_{np} \end{bmatrix} \quad (4)$$

is a linear combinations of the set of observations. Descriptive statistics can also be applied to a multivariate data matrix X, the sample mean of the kth variable is

$$\overline{x}_k = \frac{1}{n}\sum_{i=1}^{n} x_{ik}, \ k=1,2,...,p, \quad (5)$$

the sample variance is defined by

$$s_k^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_{ik} - \overline{x}_k\right)^2, \ k=1,2,...,p. \quad (6)$$

Next we introduce a matrix that contains statistics that relate pairs of variables (x_i, x_k), sample covariance s_{ik}:

$$s_{ik} = \frac{1}{n}\sum_{j=1}^{n}\left(x_{ji} - \overline{x}_i\right)\left(x_{jk} - \overline{x}_k\right), \ i=1,2,...,p, \ k=1,2,...,p. \quad (7)$$

It follows that $s_{ik}=s_{ki}$ and $s_{ii}=s_i^2$, the sample variance. Matrix of sample covariances is symmetric.

$$S_n = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ s_{31} & s_{32} & \cdots & s_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{np} \end{bmatrix} \quad (8)$$

THEOREM: Let S_n be the $p \times p$ covariance matrix related to the multivariate data matrix X. Let eigenvalues of S_n be

$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$, and corresponding orthonormal eigenvectors be u_1, u_2, ..., u_p. Then ith principal component y_i is given by the linear combination of the original variables in the data matrix X [37]:

$$y_i = Xu_i, \ i=1,2,...,p \tag{9}$$

The variance of y_i is λ_i and $cov(y_i, y_j)=0$, $i \neq j$. The total variance of the data in X is equal to the sum of eigenvalues:

$$\sum_{j=1}^{p} s_{jj} = \sum_{j=1}^{p} \lambda_j \tag{10}$$

Proportion of the total variance covered by the kth principal component is given by:

$$\frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j} \tag{11}$$

If a large percentage of the total variance can be attributed to the first few components, then these new variables can replace the original variables without significant loss of information. Thus we can achieve significant reduction in data.

VI. RESULTS

As the performance measures, true positive, true negative, false positive and false negative values have been used in this paper. A confusion matrix contains information about actual and predicted classifications done by a classification system. The confusion matrix is shown in Table 3 (actual vs. predicted):

TABLE III
CONFUSION MATRIX

Actual	Predicted	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

The dataset is composed of a range of biomedical voice measurements from 31 people, 23 with PD. The data set used in this study is very unbalanced, where out of 195 samples, 147 are Parkinson's disease type and others represent healthy people. The main problem with imbalanced data set is that it is very difficult to train to predict the presence of Parkinson's disease, since the ratio between classes is 3:1. Figure5 gives 3D graph of Parkinson's disease data showing samples with and without Parkinson's disease with different colors.
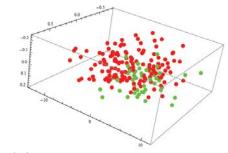

Fig. 5 Parkinson's dataset

As it can be seen from Figure 3 PD and non-PD samples coincide a lot. This makes difficulty in separating these two classes of data. Also, imbalanced data set may increase false positive rates. In order to increase performance of true recognition rates, parallel neural networks, boosted by filtering, in combination with majority rule based system are used.

Before the classification of Parkinson dataset, PCA is used to reduce the dimensionality of the input. After using PCA, the input dataset was randomly partitioned into train and test dataset. For neural networks classifier, the following adjustments were carried out: the back propagation learning algorithm has been used in the feed-forward, four hidden layer neural network. A tangent sigmoid transfer function has been used for both the hidden layers and the output layer. We used 7 committee machines. The initial weights were chosen randomly. All networks are run several times each, and training and testing performances are obtained. The obtained classification true/false counts for this case are listed in Table 4.

TABLE IV
CONFUSION MATRIX FOR DISEASE DIOGNOSIS

Confusion matrix	Committee machine
True positive (%)	92
True negative (%)	70.67
False positive (%)	8
False negative (%)	29.33
Success (%)	81.33

It can be seen from results that this method yields very good classification result, having true positive rate of 92% and true negative rate of 70.67%, and very small false classification rates. Comparing the presented results with those reported in other studies one can notice that the proposed method gives excellent results, considering the fact that is applied on vary imbalanced dataset with small number of samples. It can be concluded from this study that parallel neural networks, when boosted by filtering, in combination with majority rule based system can increase performance of true recognition rates in an imbalanced data set.

VII. Conclusion

In this research we aimed to discriminate between healthy speakers and speakers within early stages of PD. We have proposed a system for classification of Parkinson's disease based on a neural network with back propagation together with a majority voting scheme. The focus was on static structure of committee machine, known as boosting by filtering, and the total seven committee machines were used. We examined the performance and based on the results obtained, the proposed method could be confidently used in the classification of Parkinson's dataset. Namely, this method improves classification true positive rate. With this system we achieved a true recognition rate of 92 %. The experimental results show that the proposed method based on neural networks is comparable to other algorithms existing in the literature and demonstrates the effectiveness and computational efficiency of the mechanism.

References

[1] A. David Gil, B. M. Johnson, Diagnosing Parkinson by using Artificial Neural Networks and Support Vector Machines, *Global Journal of Computer Science and Technology*, pp. 63-71

[2] Max A. Little, Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease, *IEEE Transactions on Biomedical Engineering,* Vol. 56, No. 4, April 2009

[3] T. J. Ross, Fuzzy Logic with Engineering Applications, Third Edition, ISBN: 978-0-470-74376-8, John Wiley & Sons, 2010

[4] F. Åström, R. Koker, A parallel neural network approcah to prediction of Parkinson´s Disease, *Expert systems with applications*, 38(10): 12470-12474, 2011.

[5] M. F. Caglar, B. Cetisli, I. B. Toprak, Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier, *Journal of Engineering Science and Design*, Vol.1 No.2 pp.59-64, 2010

[6] P-F. Guo, P. Bhattacharya, N. Kharma, Advances in Detecting Parkinson's Disease, pp. 306–314, 2010.

[7] E. Micheli-Tzanakou, P. Wojnicki, J. Hamilton, Neuronal Signal Processing in Parkinson's Disease, Neural Networks, 2002. *IJCNN '02. Proceedings of the 2002 International Joint Conference*, 0-7803-7278-6/02, Vol.1 pp. 77 – 81, IEEE , 2002

[8] K. Polat, Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering, *International Journal of Systems Science,* 43:4, 597-609

[9] T. Bocklet, E. Noth, G. Stemmer, H. Ruzickova, J. Rusz, Detection of Persons with Parkinson's Disease by Acoustic, Vocal, and Prosodic Analysis, *Automatic Speech Recognition and Understanding (ASRU), IEEE Workshop*, 978-1-4673-0367-5/11, 478-483, 2011

[10] H. K. Rouzbahani, M. R. Daliri, Diagnosis of Parkinson's Disease in Human Using Voice Signals, Basic and Clinical Neuroscience, 2008126X, Vol. 2 Issue 3 pages/rec.No.12-20, 2011

[11] P. Verma, R.D.S. Yadava, Fuzzy C-Means Clustering Based Uncertainty Measure for Sample Weighting Boosts Pattern Classification Efficiency, *Computational Intelligence and Signal Processing (CISP), 2012 2nd National Conference,* 978-1-4577-0720-9/12,pp. 31–35, 2012, IEEE

[12] T. Khan, Parkinson's Disease Assessment using Speech Anomalies:A Review.

[13] B. K. Kihel, M. Benyettou, Parkinson's Disease Recognition Using Artificial Immune System, *Journal of Software Engineering and Applications*, 2011, 4, 391-395

[14] R. G. Ramani, G. Sivagami, Parkinson Disease Classification using Data Mining Algorithms, *International Journal of Computer Applications*,Vol. 32, No.9, October 2011

[15] UCI Machine Learning Database, ftp://ftp.ics.uci.edu/pub/machine-learning-databases

[16] T. K. Koutouzis, D. Stone, Parkinson's Disease, http://www. emedicinehealth.com/parkinson_disease/article_em.htm, eMedicineHealth Archives

[17] N. Sietske Heyn, Parkinson's Disease, http://www.medicinenet.com/ parkinsons_ disease/article.htm

[18] N. M. Tahir, H. H. Manap, Parkinson Disease Gait Classification based on Machine Learning Approach. Journal of Applied Sciences, 12: 180-185, 2012.

[19] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, *Expert Systems with Applications*, 2009

[20] Y. Delican, L. Ozilmaz, T. Yildirim, Evolutionary Algorithms Based RBF Neural Networks For Parkinson s Disease Diagnosis, *ELECO 2011 7th International Conference on Electrical and Electronics Engineering*, 290-294, December 2011, Bursa, Turkey

[21] M. Ene, Neural network-based approach to discriminate healthy people from those with Parkinson's disease, Vol. 35, 112-116, ISSN: 1223-6934, January 2008.

[22] V. Cutsuridis, S. Perantonis, A neural network model of Parkinson's disease bradykinesia, Neural Networks 19 , 354–374, 2006

[23] S.Wu, J. Guo, A Data Mining Analysis of the Parkinson's Disease, *iBusiness*, 3, 71-75, 2011.

[24] D.S.V.G.K.Kaladhar, Confusion Matrix Analysis for Evaluation of Speech on Parkinson Disease using Weka and Matlab, *International Journal of Engineering Science and Technology*, Vol. 2 (7), 2734-2737, 2010.

[25] A.Khemphila, V. Boonjing, Parkinsons Disease Classification using Neural Network and Feature selection, World Academy of Science, Engineering and Technology 64, 15-18, 2012.

[26] A.H. Hadjahmadi, T. J. Askari, A Decision Support System for Parkinson's Disease Diagnosis using Classification and Regression Tree, *The Journal of Mathematics and Computer Science*, Vol. 4 No.2, 257 – 263, 2012.

[27] National Institute of Neurological Disorders and Stroke, Parkinson's Disease Information Page, http://www ninds. nih.gov /disorders /parkinsons_disease/parkinsons_disease.htm

[28] E. Kaya, O. Findik, I. Babaoglu, A. Arslan, Effect of Discretization Method on the Diagnosis, International Journal of Innovative Computing, Information and Control, ISSN 1349-4198, Vol. 7, No. 8, pp. 4669-4678, August 2011.

[29] D. Wu, K. Warwick, Z. Ma, M. N. Gasson, J.G. Burgess, S. Pan, T. Z. Aziz, Prediction of Parkinson's Disease Tremor Onset using a Radial Basis Function Neural Network based on Particle Swarm Optimization, International Journal of Neural Systems, Vol. 20, No. 2, 109–116, 2010.

[30] N. M. Tahir, H. H. Manap, Parkinson Disease Gait Classification based on Machine Learning Approach, Journal of Applied Sciences, pp 1-6, 2012.

[31] S. M. Kamruzzaman, A. R. Hasan, A. B. Siddiquee, E. H. Mazumder, Medical Diagnosis using Neural Network, *3rd International Conference on Electrical & Computer Engineering, ICECE 2004*, 28-30 December 2004, Dhaka, Bangladesh.

[32] U.S. National Library of Medicine, PubMed Health, http://www.ncbi. nlm.nih.gov/pubmedhealth/PMH0001762/.

[33] M. El-Rashidy, T.Taha, N. Ayad, H. Sroor, An Effective Intelligent Model for Medical Diagnosis, *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 5, No. 1, September 2011.

[34] J. Jankovic, Parkinson's Disease: Clinical Features and Diagnosis, *J Neurol Neurosurg Psychiatry 2008*, Vol. 79, 368-376

[35] S. Haykin, Neural Networks, A Comprehensive Foundation, Second Edition, Pearson Prentice Hall, 1999.

[36] C. M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1995.

[37] B. Kolman, and D. R. Hill, Elementary Linear Algebra, Pearson, New Jersey, 2004.