Scientific Research

Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine

Mohammad Shahbakhi[1*], Danial Taheri Far[1], Ehsan Tahami[2]

[1]Department of Biomedical Engineering, Dezful Branch, Islamic Azad University, Dezful, Iran
[2]Department of Biomedical Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran
Email: [*]shahbakhti_m@yahoo.com

Abstract

Parkinson's disease (PD) is the most common disease of motor system degeneration that occurs when the dopamine-producing cells are damaged in substantia nigra. To detect PD, various signals have been investigated, including EEG, gait and speech. Since approximately 90 percent of the people with PD suffer from speech disorders, speech analysis is considered as the most common technique for this aim. This paper proposes a new algorithm for diagnosing of Parkinson's disease based on voice analysis. In the first step, genetic algorithm (GA) is undertaken for selecting optimized features from all extracted features. Afterwards a network based on support vector machine (SVM) is used for classification between healthy and people with Parkinson. The dataset of this research is composed of a range of biomedical voice signals from 31 people, 23 with Parkinson's disease and 8 healthy people. The subjects were asked to pronounce letter "A" for 3 seconds. 22 linear and non-linear features were extracted from the signals that 14 features were based on F0 (fundamental frequency or pitch), jitter, shimmer and noise to harmonics ratio, which are main factors in voice signal. Because changing in these factors is noticeable for the people with PD, optimized features were selected among them. Of the various numbers of optimized features, the data classification was investigated. Results show that the classification accuracy percent of 94.50 per 4 optimized features, the accuracy percent of 93.66 per 7 optimized features and the accuracy percent of 94.22 per 9 optimized features, could be achieved. It can be observed that the best classification accuracy may be achieved using Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5).

Keywords

Parkinson's Disease; Speech Analysis; Genetic Algorithm; Support Vector Machine

---

[*]Corresponding author.

1. Introduction

James Parkinson described Parkinson's disease (PD) in 1817 for the first time and the disease was named after him. Among neurological disorders, PD is the most common after Alzheimer and it is estimated that currently 4 to 6 million people suffer from it worldwide. Most people who are infected with PD are aged 50 or over but younger people can suffer it too [1]. Normally, there are brain cells (neurons) in the human brain that produce *dopamine*. These neurons concentrate in a particular area of the brain, called the **substantia nigra**. Dopamine is a chemical that relays messages between the substantia nigra and other parts of the brain to control movements of the human body. Dopamine helps humans to have smooth coordinated muscle movements [2]. Symptoms of PD typically begin appearing between the ages 50 and 60 and they develop slowly and often go unnoticed by the person who has them. Tremor is often the first symptom that people with PD or their family members notice. Initially, the tremor may appear in just one arm or leg or only on one side of the body. The tremor also may affect the chin, lips, and tongue. As the disease progresses, the tremor may spread to both sides of the body. Other symptoms may include depression and other motional changes: difficulty in swallowing, chewing, and speaking; urinary problems or constipation; skin problems; and sleep disruptions [3]. Currently, there is no cure for PD, although types of drugs called dopaminergic generally help reduce muscle rigidity, improve speed and coordination of movement and lessen tremor. Various signals, including EEG [4] speech [5]-[8] and gait, have been undertaken for diagnosis of PD. Voice signal recording is the earliest, easiest and most non-invasive technique for diagnosis of PD [9]. Since most of the people with PD suffer from speech disorders [10] [11], it could be considered as the most reasonable way for detection of PD [12] [13]. The PD dataset used in this article, has been studied by many professionals of voice analysis. It consists of 31 subjects: 23 suffer from PD and the rest are healthy. M. Ene [5] extracted 22 linear and nonlinear features out of the same data in this paper. Three types of probabilistic neural network (PNN), including incremental search (IS), Monte Carlo search (MCS) and hybrid search (HS), had been used for classification process. The concrete application had provided diagnosis accuracies ranging between 79% and 81%. The maximum classification accuracy of 81.28%, based on HS algorithm, was achieved. M. A. Little and his colleagues [6] extracted features similar to those used in [5]. They selected four optimized features based on the correlation equation and achieved a classification accuracy of 91.4% using the Support Vector Machine (SVM) method. M. F. Caglar and his colleagues [7] also selected four optimized features similar to those used in [5]. The process of selecting optimized features was based on Adaptive Neuro-Fuzzy Channel (ANFC) and the classification was investigated with Multi-Layer Perceptron (MLP), *radial basis function* (RBF) and ANFC networks. The classification accuracy for MLP was 89.69%, for RBF was 87.63% and for ANFC was 94.72%. D. Gil and M. Johnson [8] investigated ANN and SVM networks for diagnosis of PD with the same data in this paper. They could achieve classification accuracy of 90%. In this paper, a method based on combination of genetic algorithm (GA) and SVM, is investigated at which GA selects powerful features from all extracted features [14] and SVM network is used as the classifier. The GA is now widely recognized as an effective search paradigm in artificial intelligence, image processing, features extraction and many other areas. SVM is a computer algorithm that learns by example to assign labels to objects. SVMs have also been successfully applied to an increasingly wide variety of biological applications. A common biomedical application of support vector machines is the automatic classification of microarray gene expression profiles [15]. The remainder of the paper is organized as follows. In Section 2, genetic algorithm technique for selecting optimized features from the data, SVM network for classification, data acquisition and extracted features from the data are described, respectively. In Section 3, the effectiveness of our method with various numbers of optimized features is investigated. Afterwards, to evaluate the performance of classifier, three statistical parameters will be utilized. In the last section, we make a few concluding remarks.

2. Materials and Methods

2.1. Overall Structure of the Proposed Method

In this section, we propose a new algorithm for detection of PD based on genetic algorithm and SVM network. In frist part, our strategy for selecting optimized features with genetic algorithm is described. In second part, SVM network and reasons that why it used for classification is explained.

2.2. Genetic Algorithm

Genetic Algorithm (GA) is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural

selection and genetic. It is one of the most influential methods in the process of data classification, which is effectively used to select optimized features. In genetic algorithm, the solution is called chromosome or string. This method requires a population of chromosomes (strings) representing a combination of features from the solution set, and requires a cost function (called an evaluation or fitness function). This function calculates the fitness of each chromosome. The algorithm manipulates a finite set of chromosomes (the population), based loosely on the mechanism of evolution. In each generation, chromosomes are subjected to certain operators, such as crossover, inversion and mutation, which are analogous to processes, which occur in natural reproduction. Crossover of two chromosomes produces a pair of offspring chromosomes, which are synthesis of the traits of their parents. Mutation of a chromosome produces a nearly identical chromosome with only local alternations of some regions of the chromosome. The optimization process is performed in cycles called generations. During each generation, a set of new chromosomes is created using crossover, inversion, mutation and other operators. Since the population size is fixed, only the best chromosomes are allowed to survive to the next cycle of reproduction. The crossover rate usually assumes quite a high value (on the order of 80%), while the mutation rate is small (typically 1% - 15%) for efficient search. The cycle repeats until the population "converges", that is all the solutions are reasonably the same and further exploration seems fruitless, or until the answer is "good enough [16]-[18].

Strategy for Selecting Optimized Features

The process of running this algorithm in order to select the optimized feature (pattern) is explained below [14].

1. Calculate each pattern's entropy by using Equation (1) and output (target) vector's entropy by using Equation (2)

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{1}$$

$$H(Y) = -\sum_{i=1}^{n} p(y_i) \log p(y_i) \tag{2}$$

where x is the vector of features and y is the vector of targets, $p(x)$ and $p(y)$ are respectively density probability function of features and targets.

Measure mutual information between each pattern and every single output (target) via Equation (3)

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) = H(Y - H(Y|X)) \\
&= H(X) + H(Y) - H(X,Y) = H(X,Y) - H(X|Y) - H(Y|X)
\end{aligned} \tag{3}
$$

In Equation (3), the patterns' entropy ($H(X)$), the target vector's entropy ($H(Y)$) and $H(X,Y)$ are calculated by using Equation (4)

$$H(X,Y) = \sum_{i=1}^{N} \sum_{k=1}^{N} -\log(p(x_i, y_k)) p(x_i, y_k) \tag{4}$$

And

$$H(Y|X) = H(X,Y) - H(X) \tag{5}$$

And ultimately $H(Y|X)$ is measured via Equation (6)

$$
\begin{aligned}
H(Y|X) &= \sum_{x \in X} p(x) H(Y|X = x) = \sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log\left(\frac{1}{p(y|x)}\right) \\
&= -\sum_{x \in X} \sum_{y \in Y} P(x,y) \log(y|x) = -\sum_{x \in X, y \in Y} P(x,y) \log(y|x)
\end{aligned} \tag{6}
$$

2. Initial population of genetic algorithm is produced randomly using $200 \times n$ chromosomes, n is the number of features that need to be selected. Thus, each chromosome consists of n genes where the feature's number is placed randomly and it is possible for the feature number to be repeated randomly in a chromosome.

3. Measure the amount of relevance between patterns and targets for each chromosome using Equation (7)

$$V = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i; Y_i\right) \tag{7}$$

where *I* is the mutual information between features and targets. The amount of redundancy among patterns and targets is measured for each chromosome using Equation (8)

$$P = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} I\left(X_i, Y_j\right) \tag{8}$$

4. Assign the fitness value to each chromosome via Equation (9)

$$\phi = V - P \tag{9}$$

Purpose of the suggested genetic algorithm is to maximize the fitness function of Equation (9).

5. Rearrange the chromosomes according to the given fitness function.

6. Select elite chromosomes as a parent.

7. Apply crossover and mutation and produce a new population.

The chromosomes which can maximize the fitness function will remain and the rest will be removed and then Steps 1 - 5 are repeated and this process continues as long as the changes in chromosomes' fitness is less than 0.02 or the algorithm reaches the predetermined number of iterations which is supposed to be 80 in this paper. Finally, the chromosome with the maximum fitness is chosen and the number of features in that chromosome is considered as selected features.

2.3. Support Vector Machine

The most representative example of local neural network is the Support Vector Machine (SVM) of the Gaussian kernel function. It is a two layer neural network employing hidden layer of radial units and one output neuron. The procedure of creating this network and learning its parameters is organized in the way in which we deal only with kernel functions instead of direct processing of hidden unit signals [19]. Basic SVM is linear but it can be used for non-linear data by using kernel function to first indirectly map non-linear data into linear feature space. Basic SVM is also a two-class classifier however; with some modification, multiclass classifier can be obtained. If we consider a set of L linearly separable data and its class $\left\{(x_1, y_1), (x_2, y_2), \cdots, (x_l, y_l)\right\}$ where $x_i \in R^d$ and $y_i \in \{\pm 1\}$ the maximum margin classifier is $f(x) = \mathrm{sgn}(w \cdot x + b)$ where w and b are parameters that maximize the margin with respect to the two classes. A new form of the classifier, expressed with input and output vectors information is as follows:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{L} \alpha_i y_i \left(x_i \cdot x\right) + b\right) \tag{10}$$

Here parameter α for each corresponding input vectors needs to be found in order to find the maximal margin classifier. The α values are mostly zero and those inputs with non-zero α's are called the support vectors and they contribute strongly towards the decision function. If the data set is not linear, the decision function used is:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{L} \alpha_i y_i K \left(x_i \cdot x\right) + b\right) \tag{11}$$

where a kernel *K* is used in the mapping of the non-linear input space into a linear space. The maximal margin classifier is found in the linear space. There are a few possible kernels that can be chosen: Linear, Polynomial, Radial basis function, Hyperbolic tangent kernels [20]. Since the decision region is dependent on the data set, by using prior knowledge of the data and the characteristics of various kernels, we can achieve better performance. For example, if a data set is known to need closed decision regions, it is better to use an RBF kernel rather than a linear or a low order polynomial kernel [21], thus, RBF kernel has been preferred here.

$$k\left(x_i, x\right) = \exp\left(-\frac{\|x_i - x\|}{2\sigma^2}\right) \tag{12}$$

To define RBF kernel simpler, definition includes a parameter:

$$\lambda = -\frac{1}{2\sigma^2} \tag{13}$$

where $\|x_i - x\|$ is recognized as squared Euclidean distance between the two feature vectors and Λ is a free parameter which is estimated empirically [22].

SVM training involves solving a convex quadratic programming (QP) problem with equality and inequality constraints obtained by the objective of margin maximization. The solution solves for nonzero parameters α's introduced in the formulation and extracts the support vectors corresponding to it. The values of α's obtained is constraint to be positive for perfectly separable case and between 0 and C in the case of non-linearly separable data. The value C is the penalty term and needs to be chosen prior to training the SVM. Other issues in SVM training include finding the best training model using appropriate kernel and the hyper parameters. In multiclass SVM, many two class SVMs are trained and in classification, voting schemes are used for selecting the correct class. SVM is preferred here because it can directly measure the extent to which people with Parkinson can be discriminated from healthy controls on the basis of measures of dysphonia, Addressing the problem of classifying subjects as healthy or PD. With such classification method, it is also possible to combine measures to create more effective discrimination in practice [6].

3. Dataset

The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals. The original study published the feature extraction methods for general voice disorders [23].

The data consists of 195 sustained vowel phonations from 31 male and female subjects, of which 23 were diagnosed with PD. The time since diagnoses ranged from 0 to 28 years, and the ages of the subjects ranged from 46 to 85 years (mean 65.8, standard deviation 9.8). Averages of six phonations were recorded from each subject, ranging from one to 36 seconds in length. The phonations were recorded in an IAC sound-treated booth using a head-mounted microphone (AKG C420) positioned at 8 cm from the lips. The voice signals were recorded directly to computer using CSL 4300B hardware (Kay Elemetrics), sampled at 44.1 kHz, with 16 bit resolution. Although amplitude normalization affects the calibration of the samples, the study is focused on measures insensitive to changes in absolute speech pressure level. Thus, to ensure robustness of the algorithms, all samples were digitally normalized in amplitude prior to calculation of the measures. **Figure 1** illustrates speech signals of healthy and subject with PD, respectively [6].

Features Extraction

In this dataset, 22 linear and non-linear features were extracted from the data. **Table 1** contains all the features and the brief descriptions [22].

14 features are based on four factors: F0 (fundamental frequency or pitch), jitter, shimmer and noise to harmonics ratio, which are the most important factors of the voice signal.

It was concluded that the change in these factors is remarkable in people with Parkinson's disease compared to healthy people, therefore, optimized features are selected among them. Each feature is described below:

Fo (Hz): Average vocal fundamental frequency.

Fhi (Hz): Maximum vocal fundamental frequency.
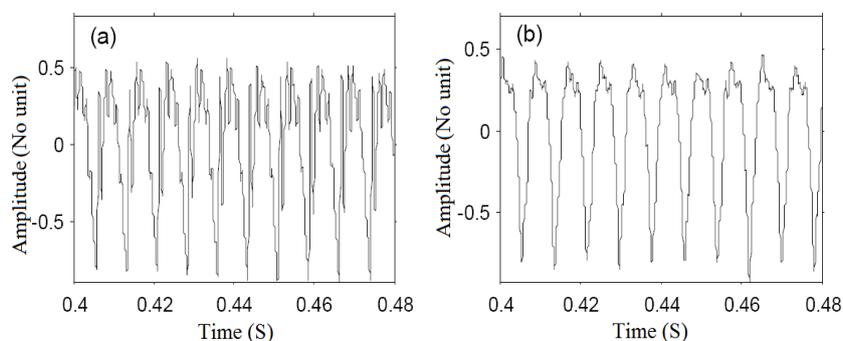
Flo (HZ): Minimum vocal fundamental frequency.

Figure 1. (a) healthy; (b) subject with Parkinson's disease.

Table 1. List of extracted features and their description.

Features	Description
MDVP: Fo (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
MDVP: Flo (Hz)	Minimum vocal fundamental frequency
Jitter (%) Jitter (Abs) MDVP: RAP MDVP: PPQ Jitter: DDP	Several measures of variation in fundamental frequency
Shimmer Shimmer (dB) Shimmer: APQ3 Shimmer: APQ5 MDVP: APQ Shimmer: DDA	Several measures of variation in amplitude
NHR HNR	Two measures of ratio of noise to tonal components in the voice
RPDE D 2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
spread 1 spread 2 PPE	Three nonlinear measures of fundamental frequency variation

Jitter (%): This is the average absolute difference between consecutive periods of fundamental frequency, divided by the average period (expressed as a percentage)

$$\text{Jitter}(\%) = \frac{\frac{1}{N}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}T_i} \tag{14}$$

where T_i is the period of fundamental frequencies of window number "i" and N is the total number of windows.

Jitter (ABS): Jitter absolute is the cycle-to-cycle variation of fundamental frequency, *i.e.* the average absolute difference between consecutive periods, expressed as:

$$\text{Jitter}(\text{ABS}) = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}| \tag{15}$$

where T_i are the extracted F0 period lengths and N is the number of extracted F0 periods.

Jitter (RAP): it is defined as the Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbours, divided by the average period.

Shimmer: This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude

$$\text{Shimmer} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N}A_i} \tag{16}$$

Shimmer (APQ5): It is defined as the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of it and its four closest neighbours, divided by the average amplitude.

HNR: harmonics to noise raito.

4. Results

Before the classification process between healthy people and people with Parkinson, various numbers of optimized features have been selected by genetic algorithm. In order to implement GA and SVM classification method, MATLAB software has been used. In classification process, after using genetic algorithm for determining powerful features, training and testing procedures are applied with them. Each column of the data has 195 different properties which divided into 2 parts, 75% for training and 25% for testing. The network was used 100 times for the classification per different numbers of features. The experimental results from **Table 2** indicate that the maximum amount of classification accuracy, 94.50%, has been achieved with having Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5) features. It is also shown that a classification accuracy of 93.66% for Fhi (Hz), Fho (Hz), Flo (hz), jitter (RAP), shimmer (APQ5), Jitter (ABS), shimmer features and a classification accuracy of 94.22% for Fhi (Hz), Fho (Hz), Flo (hz), jitter (RAP), shimmer (APQ5), Jitter(ABS), shimmer, Jitter (%), HNR features. **Table 2** contains data classification accuracy per different numbers of optimized features with SVM classifier.

To evaluate the performance of the proposed method, 3 statistical parameters: specificity, sensitivity and total classification accuracy, per various numbers of features, have been calculated.

$$\text{Specificity} = \frac{\text{NCPD}}{\text{NTPD}} \times 100 \tag{17}$$

$$\text{Sensitivity} = \frac{\text{NCH}}{\text{NTH}} \times 100 \tag{18}$$

$$\text{Total classification accuracy} = \frac{\text{NCCP}}{\text{NTP}} \times 100 \tag{19}$$

where NCPD, NTPD, NCH, NTH, NCCP and NTP are number of correct classified PD, number of total PD, number of correct classified healthy, number of total healthy, number of correct classified persons and number of total persons respectively. **Table 3** shows the value of statistical parameters per various numbers of features. **Figure 2** shows the plots of pair of first four prominent features which were extracted by genetic algorithm.

5. Discussion

The aim of this research was diagnosis of Parkinson's disease using voice analysis. In this paper, the dataset consisted of 31 people at which 23 subjects suffer from PD and the rest are healthy. Various linear and non-linear features were extracted from the data which among them, 14 features with emphasis on four main speech factors: fundamental frequency (pitch), jitter, shimmer and noise to harmonics ratio were extracted from the data. The recent researches showed that changing in these 14 factors is notable for people with Parkinson's disease compared to healthy people, thus, the process of extracting optimized features was done among them. Selecting powerful features is a primary step to improve classification accuracy. GA produces successive populations of alternate solutions that are represented by a chromosome, a solution to the problem, until acceptable results are

**Table 2.** Classification accuracy per different number of features.

Features	$N = 4$	$N = 7$	$N = 9$
Accuracy	94.50 ± 3.54	93.66 ± 3.61	94.22 ± 3.66

Table 3. Statistical parameters: specificity, sensitivity and total classification accuracy, per various numbers of features.

Feature Number	$N = 4$	$N = 7$	$N = 9$
Specificity	95.63	89.57	92.8
Sensitivity	88.46	80.04	70.12
Total classification accuracy	96.06	93.58	93.61
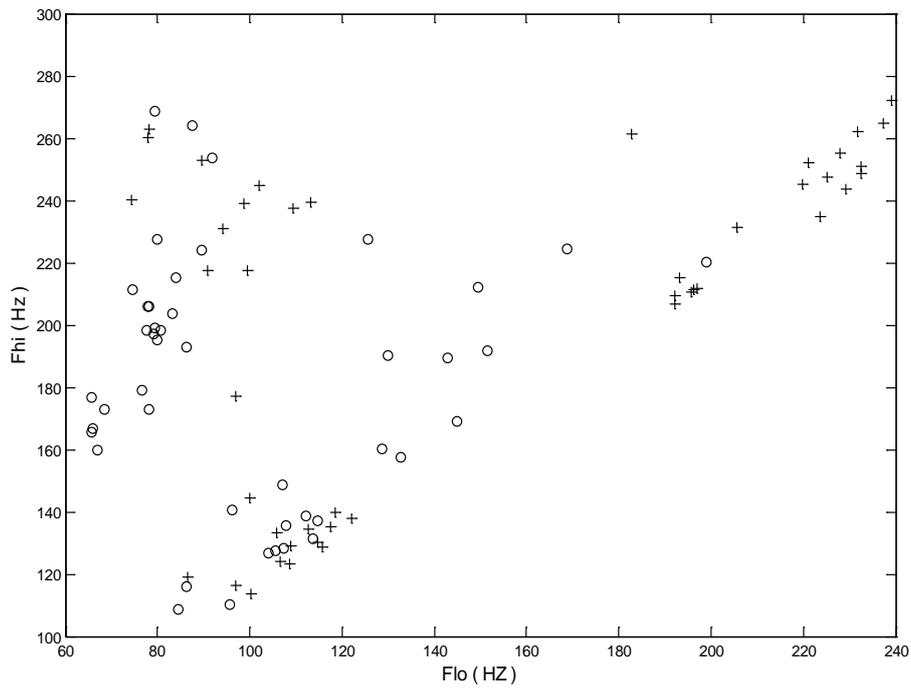
(a)
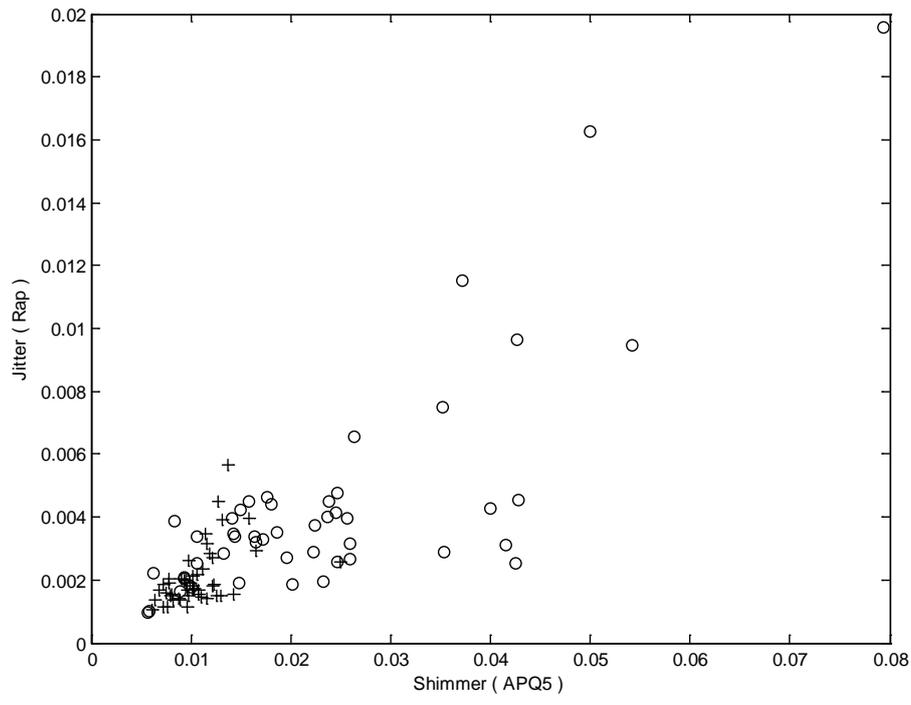
(b)

Figure 2. (a) demonstration of FHi (Hz)-Flo (Hz) database; (b) demonstration of Jitter (RAP)-Shimmer (APQ5) database. The "O" marks are for healthy subjects, the "+" marks for Parkinson's subjects.

obtained. GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms. In first step, to extract most useful features for classification between normal people and people with PD, genetic algorithm had been undertaken. Afterwards, the classification process was done

with various numbers of optimized features, which were extracted in the last step. Among different types of classifiers, support vector machine (SVM) was chosen. Indeed, by introducing the kernel, SVM gain flexibility in the choice of the form of the threshold separating healthy from people with Parkinson, which needs not be linear and even needs not have the same functional form for all data, since its function is non-parametric and operates locally. Furthermore, prior visual inspection of the layout and clustering of pairs of measures shows that it might be difficult to separate people with Parkinson from healthy people with linear or hyperplanes kernels thus, the kernel-SVM formulation, with Gaussian radial basis kernel functions was chosen [24]. As it was described, ʎ parameter should be determined experimentally, therefore, we systematically increased the value of ʎ from 0 to 1 and best classification accuracies were achieved when the value of ʎ was 0.452.

Per various numbers of features the classification process, had been investigated and the best classification accuracy was achieved with first four extracted features: Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5). However, it was possible to improve the classification accuracy if a combination of linear and non-linear features would be carried out. To this aim, we are working on a method, which uses both linear and non-linear features for classification. Moreover, a fusion SVM network will be used for the classification between healthy and PD subjects.

6. Conclusion

Parkinson's disease is known as the second common neurological disorder after Alzheimer. It influxes several aspects of human's functions in which speech disorder is the most prominent. Several researches have been proposed for diagnosis of PD with voice analysis [5]-[8]. In this paper, a method based on combination of genetic algorithm and SVM network, for classification of healthy people and people with Parkinson of various numbers of features, was investigated. Results showed that the highest accuracy was achieved with extracting 4 optimized features: Fhi (Hz), Fho (Hz), jitter (RAP) and shimmer (APQ5). It is observed that there is no major difference between accuracy of our technique and Reference [7].

References

[1] Parkinson's UK (2013) http://www.parkinsons.org.uk/about_parkinsons/what_is_parkinsons.aspx

[2] National Parkinson Foundation (2013) http://www.parkinson.org/parkinson-s-disease.aspx

[3] Webmd (2013) http://www.webmd.com/parkinsonsdisease/tc/parkinsons-disease-symptoms

[4] Pezard, L., Jech, R. and RuÊzÏicϊka, E. (2001) Investigation of Non-Linear Properties of Multichannel EEG in the Early Stages of Parkinson's Disease. *Clinical Neurophysiology*, **122**, 38-45.

[5] Ene, M. (2008) Neural Network-Based Approach to Discriminate Healthy People from Those with Parkinson's Disease. *Mathematics and Computer Science Series*, **35**, 112-116.

[6] Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J. and Ramig, L.O. (2008) Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE Transactions on Biomedical Engineering*, **56**, 1015-1022. http://dx.doi.org/10.1109/TBME.2008.2005954

[7] Caglar, M.F., Cetisli, B. and Toprak, I.B. (2010) Automatic Recognition of Parkinson's Disease from Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier. *Journal of Engineering Science Design*, **1**, 59-64.

[8] Gil, D. and Johnson, M. (2009) Diagnosing Parkinson by Using Artificial Neural Networks and Support Vector Machines. *Global Journal of Compute Science and Technology*, **9**, 63-71.

[9] Duffy, R.J. (2005) Motor Speech Disorders: Substrates, Differential Diagnosis and Management. 2nd Edition, Elsevier Mosby, St. Louis.

[10] Ho, A.K., Iansek, R., Marigliani, C., Bradshaw, J.L. and Gates, S. (1998) Speech Impairment in a Large Sample of Patients with Parkinson's Disease. *Behaviour Neurology*, **11**, 131-137. http://dx.doi.org/10.1155/1999/327643

[11] Logemann, J.A., Fisher, H.B., Boshses, B. and Blonsky, E.R. (1978) Frequency and Co-Occurrence of Vocal-Tract Dysfunctions in Speech of a Large Sample of Parkinson Patients. *Journal of Speech Hearing Disorder*, **43**, 47-57.

[12] Sapir, S., Spielman, J.L., Ramig, L.O., Story, B.H. and Fox, C. (2007) Effects of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals with Idiopathic Parkinson Disease: Acoustic and Perceptual Findings. *Journal of Speech Lang Hearing Research*, **50**, 899-912. http://dx.doi.org/10.1044/1092-4388(2007/064)

[13] Rahn, D.A., Chou, M., Jiang, J.J. and Zhang, Y. (2007) Phonatory Impairment in Parkinson's Disease: Evidence from Nonlinear Dynamic Analysis and Perturbation Analysis. *Journal of Voice*, **21**, 64-71.

http://dx.doi.org/10.1016/j.jvoice.2005.08.011

[14] Ludwig, O., Nunes, U., Araujo, R., Schnitman, L. and Lepikson, H.A. (2009) Applications of Information Theory, Genetic Algorithms, and Neural Models to Predict Oil Flow. *Communications in Nonlinear Science and Numerical Simulation*, **14**, 2870-2885. http://dx.doi.org/10.1016/j.cnsns.2008.12.011

[15] Nobel, W. (2006) What Is a Support Vector Machine? *Nature Biotechnology*, **24**, 1565-1568.

[16] Holland, J. (1975) Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor.

[17] Goldberg, D. (1989) Genetic Algorithms in Search, Optimization and Machine Learning. Addison Wesley, New York.

[18] Jong, K.A. and Spears, W.M. (1992) A Formal Analysis of the Role of Multi-Point Crossover in Genetic Algorithms. *Annals of Mathematics and Artificial Intelligence*, **5**, 1-26. http://dx.doi.org/10.1007/BF01530777

[19] Osowski, S., Siwek, K. and Markiewicz, T. (2004) MLP and SVM Networks—A Comparative Study. *Proceedings of the* 6*th Nordic Signal Processing Symposium*, Espoo, 37-40.

[20] Ahmad, A.H., Viard-Gaudin, C., Khalid, M. and Poisson, E. (2004) Comparison of Support Vector Machine and Neural Network in Character Level Discriminant Training for Online Word Recognition, UNITEN Students Con on Research and Development, Malaisie.

[21] Ganapathiraju, A., Hamaker, J. and Picone, J. (1998) Support Vector Machines for Speech Recognition. *The Proceedings of the* 5*th International Conference on Spoken Language Processing*, *Incorporating the* 7*th Australian International Conference on Speech Science and Technology*, Sydney.

[22] Jean-Philippe, V., Tsuda, K. and Schölkopf, B. (2004) Kernel Methods in Computational Biology. MIT Press, Cambridge.

[23] Center for Machine Learning and Intelligent Systems (2008) http://archive.ics.uci.edu/ml/datasets/Parkinsons

[24] Hastie, T., Tibshirani, R. and Friedman, J.H. (2001) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York. http://dx.doi.org/10.1007/978-0-387-21606-5