

## Effective Detection of Parkinson's Disease at Different Stages using Measurements of Dysphonia

Elmehdi Benmalek, Jamal Elmhamdi, Abdelilah Jilbab

Laboratory LRGE, ENSET, Mohamed V University, Rabat, Morocco

---

### Article Info

#### Article history:

Received May 25, 2018

Revised Jun 20, 2018

Accepted Sep 6, 2018

---

#### Keyword:

Dysphonia measures

LDA

Multiclass classification

Parkinson's disease

PSO

Random forest

---

### ABSTRACT

This paper addresses the problem of multiclass of Parkinson's disease by the characteristic features of person's voice. So we computed 22 dysphonia measures from 375 voice samples of healthy and people suffering from Parkinson's disease. We used the particle swarm optimization (PSO) feature selection method, with random forest and the linear discriminant analysis (LDA) along with the 4-fold cross validation analysis to classify the subjects in 4 classes according to the severity of symptoms. With a classification accuracy score of 95.2%, promisingly the proposed diagnosis system might serve as a powerful tool for diagnosing PD, and could also be extended for other voice pathologies.

Copyright © 2018 Institute of Advanced Engineering and Science.  
All rights reserved.

---

### Corresponding Author:

Elmehdi Benmalek,

Laboratory LRGE, ENSET, Mohamed V University,

Rabat, Morocco.

Email: elmehdi.benmalek@um5s.net.ma

---

## 1. INTRODUCTION

Parkinson's disease (PD) is a neurodegenerative disorder caused by a progressive damage of dopamine producing nerve cells in the midbrain [1]. We estimate according to the statistics by Parkinson's disease foundation that 7 to 10 million people around the world are living with PD [2]. PD is for the most part found in individuals over 50 years old, age is considered the main factor of PD. Several non-invasive methods have been suggested by scientists to detect the severity of PD employing acoustic analysis of physiological and voice signals.

Speech impairment is an early onset indicator of PD, with the disease progression, it was estimated that up to 90% of the patients develop speech symptoms [3], these vocal issues don't show up suddenly, and may get unnoticed in the early stages. Some investigations suggested strong interlink between the degradation in speech quality and the general PD severity [4]. For this purpose, speech processing has been considered to be an excellent tool for voice disorder detection. Recent studies are using acoustic measurements of dysphonia and machine learning tools for the detection of PD. Little in [5] have employed the Gaussian radial basis kernel functions and the SVM classifier to detect PD, and obtained an accuracy rate of 91.4%. Sakar in [6] got a classification accuracy of 92.75% using the mutual information feature selection methods integrated with the SVM classifier. Guo in [7] used genetic programming along with the expectation maximization algorithm (GP-EM) for the detection of PD, and obtained a classification score of 93.1%. Tsanas in [8] got an almost 99% accuracy using the SVM classifier and RELIEF feature selection algorithm, a significant improvement over the previous studies. All these studies have been performed for binary classification, so for an early diagnosis of PD a multiclass classification based on severity of symptoms has been achieved with different classifiers using the Local Learning-Based Feature Selection (LLBFS) and the cepstral analysis [9], [10]. We aim in this paper to classify 375 subjects on 4 groups according to the UPDRS scores; the first group has 55 subjects as healthy, the second one has 178 considered in early stage, the third

one has 118 subjects in intermediate stage and the last one has 24 subjects in advanced stages. Each subject pronounce at a comfort level the sustained vowel /a/. Then we extract acoustic features from each voice sample, and we apply the Particle swarm optimization (PSO) feature selection algorithm to reduce the number of these acoustic features and get only the most pertinent ones,. For classification, we used the random forest and discriminant analysis classifiers along with k-folds cross validation method.

## 2. MATERIALS AND METHODS

### 2.1. Dataset

The PVA-dryrun data set consists of brief voice recordings of sustained phonations [8], [11], 22 features extracted from the voice recordings, Parkinson's Disease Rating Scale (PDRS) as well as some demographic information from 620 individuals with PD; age with the mean age is 62.17, the max is 84 and min is 34 years old, years since first symptom, the gender of the subject and if he is on treatment or not. In this study we used 375 voice samples (duplicated and useless records have been taken away). All subjects were requested to record and maintain as possible the sustained vowel /a/. They also provided the following information; Along with the voice records we have PDRS scores which is an abbreviated version of the Unified Parkinson's Disease Rating Scale (UPDRS), this metric is used to evaluate PD severity. The 17 items PDRS questionnaire omits the clinical observation section present in the more comprehensive 42-item UPDRS. PDRS can be self-administered and completed quickly (~10 minutes). PDRS has a maximum score of 68 points. Each question is rated on a (0-4) scale with "0" representing no disability and "4" worst disability.

Among the 375 subjects from the data were recorded, based on UPDRS scores we consider the first 55 subjects as healthy, the second 178 as in early stage, the third 118 as in intermediate stage, and the last 24 are considered as in advanced stages. For the evaluation of voice disorders the pre-processing of the voice recordings alone is not adequate. Therefore, it is essential for speech analysis to use a set of acoustic features, represented as a feature vectors.

### 2.2. Features Extraction

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

Table 1. Extracted Features

Field	Description	Type, format/units/values
Feature 1	MDVP: Fo (Hz)	Average vocal fundamental frequency
Feature 2	MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
Feature 3	MDVP: Flo (Hz)	Minimum vocal fundamental frequency
Feature 4	MDVP: Jitter (%)	Several measures of variation in fundamental frequency
Feature 5	MDVP: Jitter(Abs)	
Feature 6	MDVP: Jitter:RAP	
Feature 7	MDVP: Jitter:PPQ5	
Feature 8	Jitter:DDP	
Feature 9	MDVP: Shimmer	Several measures of variation in amplitude
Feature 10	MDVP: Shimmer(dB)	
Feature 11	Shimmer:APQ3	
Feature 12	Shimmer:APQ5	
Feature 13	MDVP:APQ	
Feature 14	Shimmer:DDA	
Feature 15	NHR	Two measures of ratio of noise to tonal components in the voice
Feature 16	HNR	
Feature 17	RPDE	Two nonlinear dynamical complexity measures
Feature 18	D2	
Feature 19	DFA	Signal fractal scaling exponent
Feature 20	Speard1	Three nonlinear measures of fundamental F21 Spread2
Feature 21	Spread2	frequency variation
Feature 22	PPE	

Jitter (%): expressed as a percentage, is the division of the cycle-to-cycle variation of fundamental frequency by the average period, expressed as:

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

Where  $N$  is the total number of windows and  $T_i$  is the period of fundamental frequency of window number “ $i$ ”.

Jitter (ABS): Jitter absolute known as jitta, is the average absolute difference between consecutive periods, expressed as:

$$Jitter(ABS) = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i-1}| \quad (2)$$

Where  $T_i$  and  $N$  are the lengths and the number of extracted F0 period respectively.

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

Jitter (PPQ5) is defined as the ratio of disturbance within 5 periods, it represents the average absolute difference between a period and the average containing its four nearest neighbors periods, divided by the average period [12], [13].

Shimmer: represents the division of the average absolute difference between the amplitudes of two consecutive periods, by the average amplitude

$$Shimmer = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (3)$$

Shimmer (APQ5): It is defined as the ratio of perturbation amplitude of 5 periods, i.e., the division of the average absolute difference between the amplitude of a period and the average of the amplitudes of it containing its four closest neighbors, by the average amplitude.

RPDE: Recurrence Periodicity Density Entropy, is a measure based on the notion of recurrence [14], which can be considered as a generalization of periodicity [5]. By measuring the deviations from exact periodicity, it addresses the capacity of the vocal folds to support stable vocal fold oscillation.

PPE: Pitch Period Entropy, since people with PD have hard time to maintain stable pitch during sustained phonations [15], the PPE measures this impairment [16].

DFA: Detrended Fluctuation Analysis, is a measure based on scaling analysis which tend to overcome the problems of scaling analysis technique, that is only adapted for stationary signals by quantifying long range power-law autocorrelations in non-stationary signals[5], [17].

HNR: harmonics to noise ratio.

### 2.3. Feature Selection and Validation

To enhance the classification accuracy and optimize the visualization plus the comprehension of the data, which help also the reduction of the storage space, CPU-expenditure and the time consumption, we aimed to apply a feature selection algorithm, so we can identify the most pertinent features, thus the redundant and useless information will be circumvented, so we have a better representation of the data. The main objective of the feature selection can be described by [18]. However, the classification error can be increased by the elimination of certain very relevant informations, considering this information if they are used can prove to be informative [19]. Our goal is to design efficient algorithms to select a solid set of pertinent features. In this study we used the Particle swarm optimization, a swarm intelligence method developed by [20], it is a population-based optimization algorithm. In PSO, each solution is considered a bird of the flock, that is, a particle in the search space. Each particle’s memory and knowledge gained by the swarm enable the algorithm to find the best solution:

Each particle has its own fitness value, evaluated by an optimized fitness function, and have own pace to manage its movement, and all particles adjust their positions according to their own as well as neighbors’ particles experiences, and use the best position. The swarm is initially created in sort of that the particles’ population is randomly distributed over the search space. For each iteration, by following the best values “pbest” and “gbest” every particle is updated and keeps track of its coordinates associated with the best fitness value “pbest” so far found. And every particle is associated with the best value that the whole swarm has achieved so far is called “gbest”. The PSO procedure is given below.

After the extraction of the features and the selection the more pertinent ones, we map the voice samples into four groups depending on the severity; Healthy subjects, and those with PD in early, intermediate and advanced stages. Using these parameters, we built a matrix; the columns and the rows represent the dysphonia measurements and the voice samples respectively. In this paper, we used 4-folds cross validation along with random forest classifier and discriminant analysis; The dataset is divided into 4 subsets, Each time 3 subset (75%) form the training set and one subset (25%) is used for the testing, then we calculate the average error across all 4 trials. So, it doesn’t matter how the dataset gets divided, each data point by using this method is used in the testing set exactly once, and 3 times in the training set

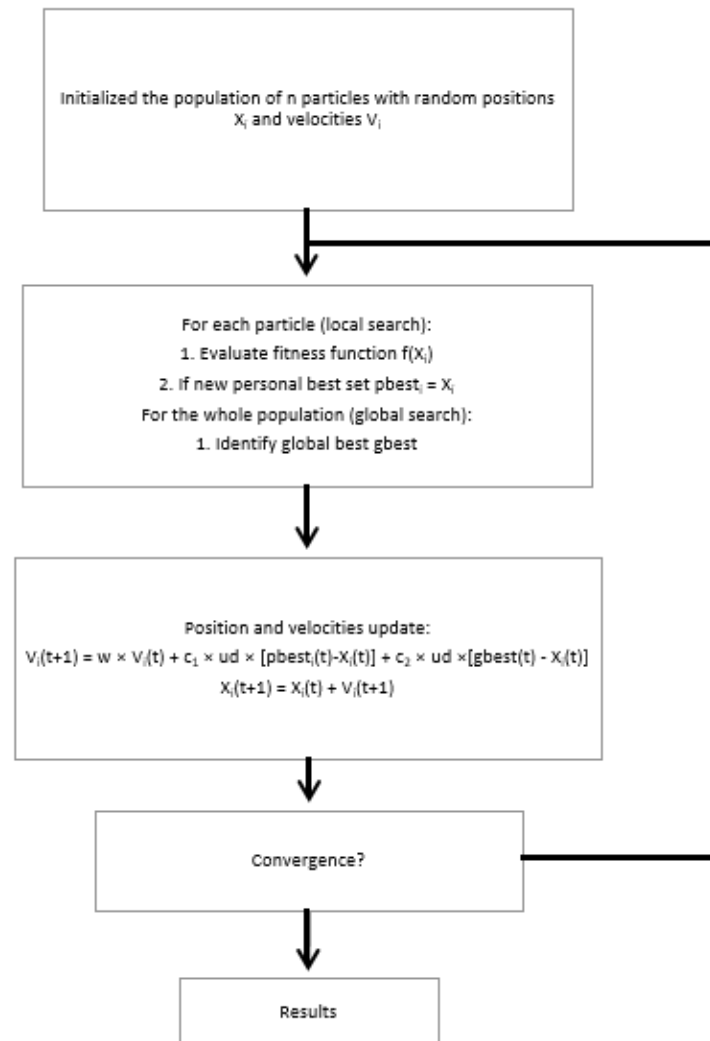


Figure 1. PSO algorithm

### 3. EXPERIMENTAL RESULTS

In order to demonstrate the detail of the feature selection procedure, we list the features selected and their rank according to PSO algorithm in Table 2.

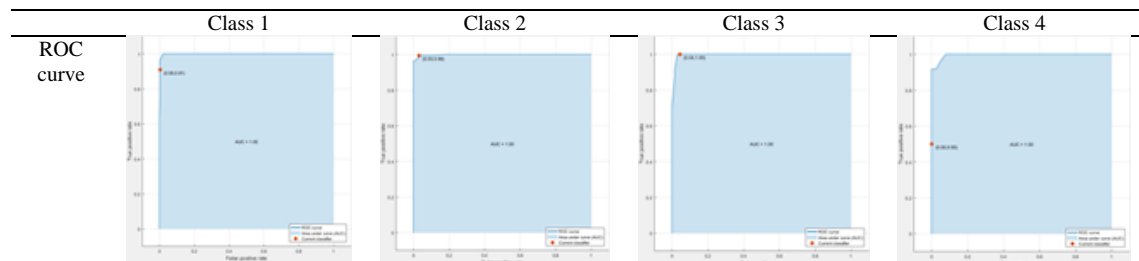
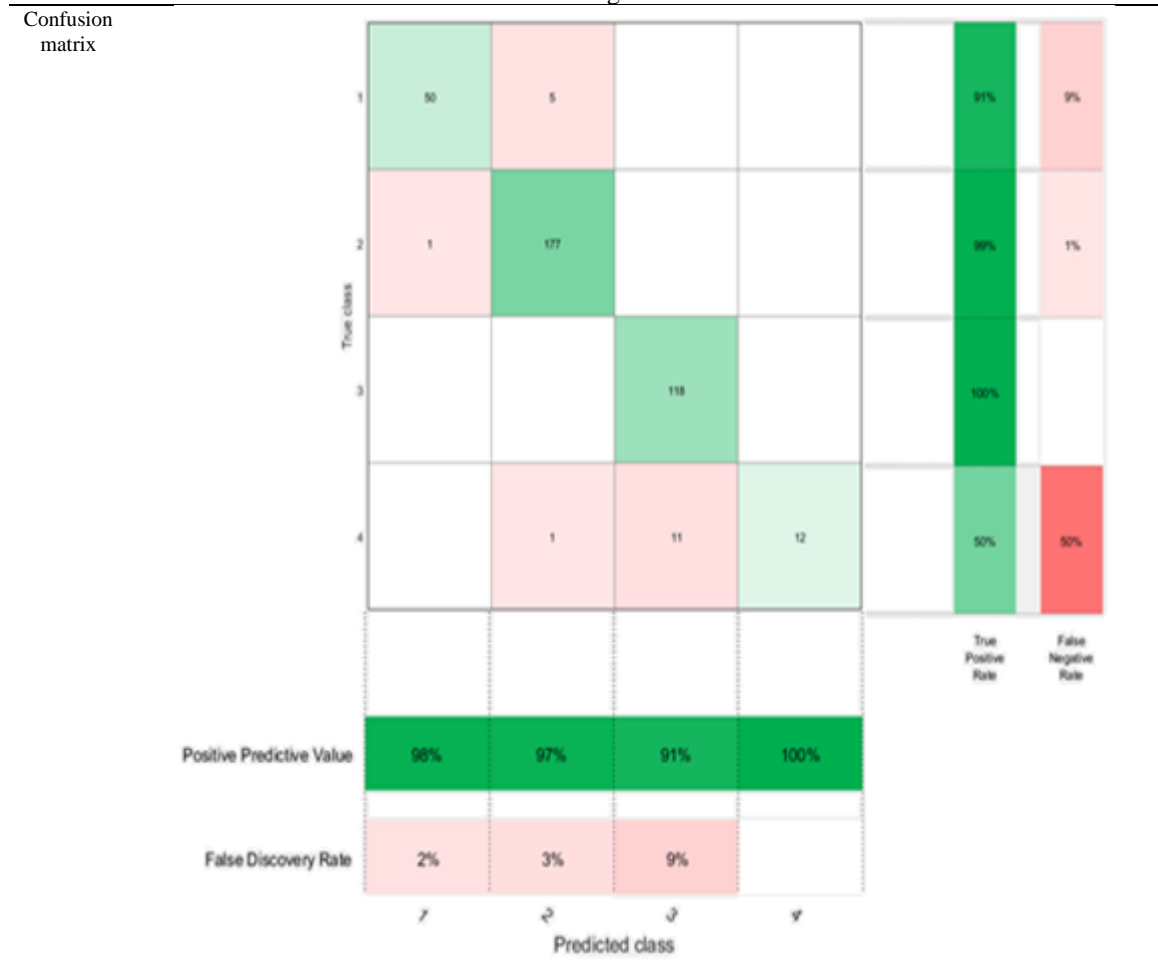
Feature no.	Rank (according to PSO)
MDVP:Fo (Hz)	1
MDVP: Jitter (%)	2
MDVP: Jitter(Abs)	3
MDVP: Jitter:RAP	4
Shimmer:APQ3	5
HNR	6
RPDE	7
D2	8
PPE	9

After ranking the dysphonia measures according to the PSO features selection algorithm, we used the random forest and Discriminant analysis along with the k-fold cross validation method to classify the subjects based on the 9 selected features.

**3.1. Obtained Results using Random Forest**

Random forest/forests is an ensemble that fits many decision trees classifier, and outputs the class that is the mode of the class's output by individual trees. They are among the most accurate models yet invented. Developed by [21], the Random forest algorithm combines Breiman's "bagging" idea and random features selection introduced independently in order to build an ensemble of decision trees with controlled variation [22]. The Random tree forests are as easy to establish as single tree models, but often have a degree of accuracy that cannot be obtained using this one.

Table 3. Results using Random Forest



The tables 3 represents the classification results obtained using the random forest after the extraction of the voice features and the selection of the more relevant one by using the PSO algorithm. In addition, the ROC curve for each class. The accuracy score of 95.2% was obtained. In this model we have:

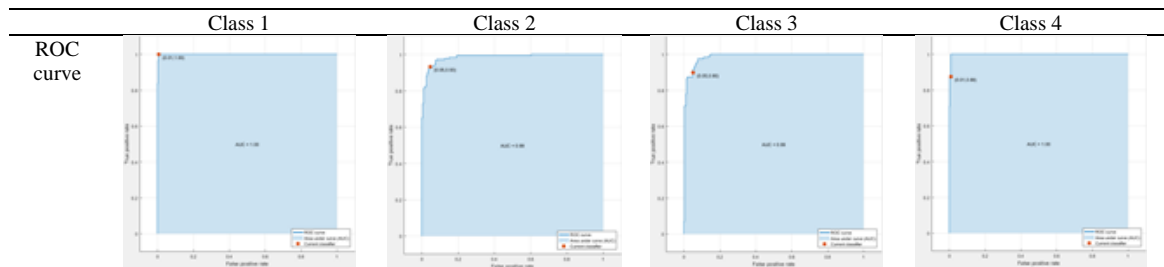
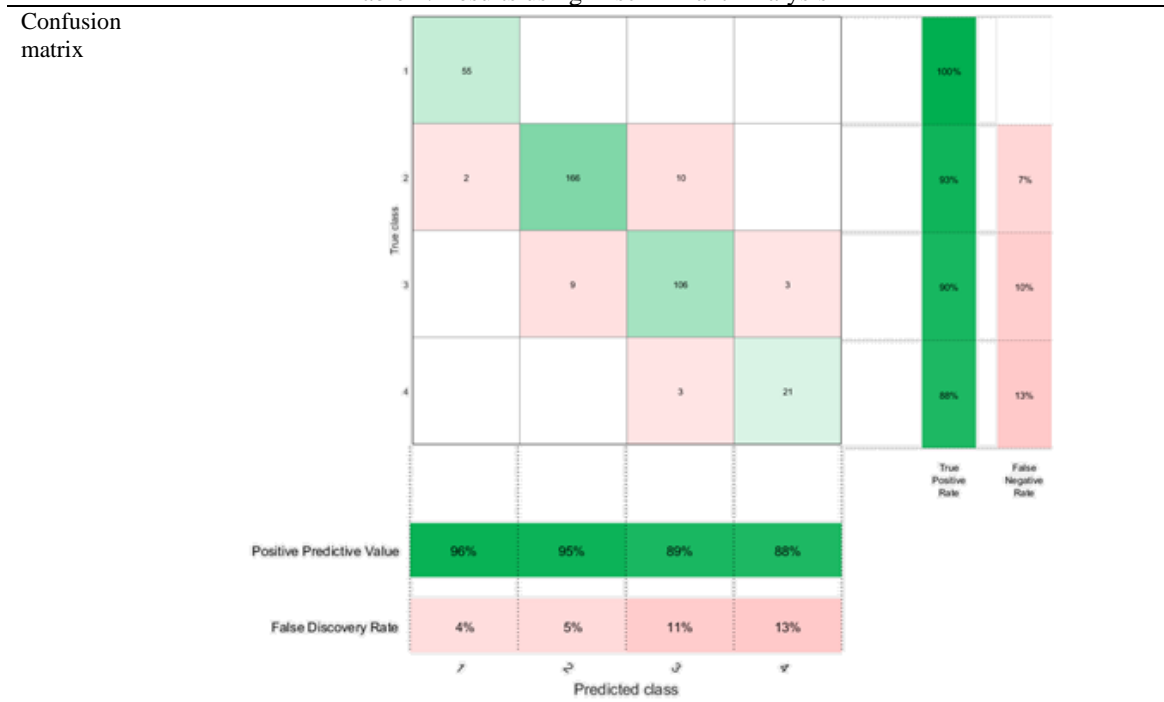
1. For the healthy subjects, we have 50 that have been correctly classified, 5 were misclassified (all considered as subjects in early stage), with a true positive rate of 91%;
2. For the subjects in early stage we have 177 that have been correctly classified, 1 was misclassified and considered as a healthy person, with a true positive rate of 99%;
3. All the 118 subjects in intermediate stage have been correctly classified, with a true positive rate of 100%;

For the subjects in advanced stage, half of them have been correctly classified, and 12 were misclassified (11 considered as in intermediate stage and one subject as in early stage), with a true positive rate of 50%.

**3.2. Obtained Results using Discriminant Analysis**

Linear discriminant analysis (LDA) also known as Fisher’s linear discriminant analysis [23] is one of the most used method aimed at finding a linear combinations of best observed features that describe or separate two or more classes of objects. The results of the combinations are used for discrimination, dimensionality reduction and classification. For each class this method consists of calculating statistical properties of the data. For a single input variable “x” this is the variance and the mean of the variable for each class. For multiple variables, this is the same properties calculated over the multivariate Gaussian, namely the means and the covariance matrix. The discriminant analysis has shown an excellent results in previous multiclass classification, so we took it as reference with a different feature selection algorithm in this study and compare it with the random forest.

Table 4. Results using Discriminant Analysis



The table 4 represents the the classification results by using the discriminant analysis along with the PSO feature selection algorithm, we obtained a classification accuracy of 92.8%, in addition to the ROC curve for each class, and the results are as follow:

1. All the 55 healthy subjects have been correctly classified, with a true positive rate of 100%;
2. For the subjects in early stage we have 166 that have been correctly classified, 12 were misclassified (2 considered as healthy persons and 10 subject as in intermediate stage), with a true positive rate of 93%;
3. For the subjects in intermediate stage we have 106 that have been correctly classified, 12 were misclassified (9 considered as in early stage and 3 subject as in advanced stage), with a true positive rate of 93%;

For the subjects in advanced stage, 21 have been correctly classified, and 3 was misclassified all of them were considered as in intermediate stage, with a true positive rate of 88%.

The highest classification rate of 95.2% was achieved using the random forest along with the PSO feature selection algorithm. From all these results, we conclude that the feature selection has a huge impact on the classification optimization. However, it should be bear in mind that for a more precise estimation of speech symptoms, a more comprehensive set of features is required for categorizing the severity levels of speech symptoms in PD. The limitation associated with this work is that, the speech tests were recorded in a silent rooms, in a real-life environments, the processing of noisy signals to quantify speech symptoms could be challenging.

#### 4. CONCLUSION

In this study we aimed to test the effectiveness of using different classifiers and optimization algorithm for detecting PD. A comparative study of two different classifiers on the dataset was performed; First of all, we extracted 22 different types of voice features, afterwards we applied the PSO algorithm for the selection of the more relevant among these features, subsequently 2 supervised classifiers are implemented, the random forest classifier presents an accuracy of 95.2%. The error rate can be explained by the limited number of features used, and also the relativity of the UPDRS for precisely determining the disease progression degree, but these results are encouraging and may help with other voice pathologies, especially in the early detection, it yields in treating the patient well ahead and preventing the risk of the disease's gradation.

#### ACKNOWLEDGEMENTS

These Datasets were generated through collaboration between Sage Bionetworks, PatientsLikeMe and Dr. Max Little as part of the Patient Voice Analysis study (PVA). They were obtained through Synapse ID [syn2321745].

#### REFERENCES

- [1] Olanow, C.W., Stern, M.B., Sethi, K. The scientific and clinical basis for the treatment of Parkinson disease (2009). *Neurology*, 72(21 Supplement 4), 2009, S1-S136.
- [2] Statistics on Parkinson's, <http://www.pdf.org/en/parkinson-statistics> (accessed on January, 2013) 2013.
- [3] Harel, B., Cannizzaro, M., Snyder, P. J. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: a longitudinal case study. *Brain and cognition* 56(1), 2004, 24-29
- [4] Skodda, S., Rinsche, H., Schlegel, U. Progression of dysprosody in Parkinson's disease over time—a longitudinal study. *Movement Disorders* 24(5), 2009, 716-722.
- [5] M.A. Little, P. E. McSharry, S. J. Roberts, D. Costello and I. M. Moroz, —Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection, *Biomedical Engineering Online*, vol. 6 (23), 2007.
- [6] C.O. Sakar, O. Kursun, Telediagnosis of Parkinson's disease using measurements of dysphonia, *Journal of Medical Systems*, 2010, 34 (4) 1–9.
- [7] P.F. Guo, P. Bhattacharya, N. Kharna, Advances in detecting Parkinson's disease, *Medical Biometrics*, 2010, 306–314.
- [8] A. Tsanas, M.A. Little, P.E. McSharry, J. Spielman, L.O. Ramig Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease *IEEE Transactions on Biomedical Engineering*, 2012, 59(5):1264-1271
- [9] Benmalek, E., Elmhamdi, J., & Jilbab, A. Multiclass classification of Parkinson's disease using different classifiers and LLBFS feature selection algorithm. *International Journal of Speech Technology*, 2017, 20(1), 179-184.
- [10] Benmalek, E., Elmhamdi, J., & Jilbab, A. Multiclass classification of Parkinson's disease using cepstral analysis. *International Journal of Speech Technology*, 2017, 1-11.

- [11] A. Tsanas, M.A. Little, P.E. McSharry, L.O. Ramig Using the cellular mobile telephone network to remotely monitor Parkinson's disease symptom severity IEEE Transactions on Biomedical Engineering, 2012,
- [12] Farrús, Mireia, Javier Hernando, and Pascual Ejarque. "Jitter and shimmer measurements for speaker recognition." INTERSPEECH. 2007.
- [13] Shirvan, R. Arefi, and E. Tahami. "Voice analysis for detecting Parkinson's disease using genetic algorithm and KNN classification method." Biomedical Engineering (ICBME), 2011 18th Iranian Conference of. IEEE, 2011,
- [14] H. Kantz, T. Schreiber, Nonlinear time series analysis, Cambridge University Press, 2nd edition, 2004
- [15] L. Cnockaert, J. Schoentgen, P. Auzou, C. Ozsancak, L. Defebve and F. Grenez, —Low frequency vocal modulations in vowels produced by Parkinsonian subjects," Speech Communication, Vol. 50, pp. 288-300, 2008
- [16] M.A. Little, et al., Suitability of dysphonia measurements for telemonitoring of Parkinson' disease, IEEE Transactions on Biomedical Engineering, 2009.
- [17] Z. Chen, P.C. Ivanov, K. Hu, and H. E. Stanley, —Effect of nonstationarities on detrended fluctuation analysis, Physical Review E, Vol. 65(4), 041107, 2002
- [18] Guéris S. Réduction de dimension en apprentissage numérique non supervisée. PhD thesis, Université Paris 13 2006. p. 420 148. 421
- [19] Ferchichi SE, Zidi S, Laabidi K, Maouche S. Feature selection using an SVM learning machines. In Proceedings of the 422 3rd International Conference on Signals, Circuits and Systems (SCS 2009). 2009; 1-6.
- [20] J. Kennedy, R.C. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE International Conference on Neural Network, Vol. 4, 1995, pp. 1942–1948
- [21] Breiman, L. Random forests. Machine learning, 2001, 45(1), 5-32. DTREG. <http://www.dtreg.com> Software for Predictive Modeling and Forecasting.
- [22] Amit, Y., Geman, D., & Wilder, K. Joint induction of shape features and tree classifiers. IEEE transactions on pattern analysis and machine intelligence, 1997, 19(11), 1300-1305.
- [23] Fisher, R. A. "The Use of Multiple Measurements in Taxonomic Problems". Annals of Eugenics. 7 (2): 179–188. Doi :10.1111/j.1469-1809.1936.tb02137.x. 1936