

An Ensemble Method for Diagnosis of Parkinson's Disease Based on Voice Measurements

Abstract

Background: Parkinson's disease (PD) is the most common destructive neurological disorder after Alzheimer's disease. Unfortunately, there is no specific test such as electroencephalography or blood test for diagnosing the disease. In accordance with the previous studies, about 90% of people with PD have some types of voice abnormalities. Therefore, voice measurements can be used to detect the disease. **Methods:** This study presents an ensemble-based method for identifying patients and healthy samples by class label prediction based on voice frequency characteristics. It includes three stages of data preprocessing, internal classification and ultimate classification. The outcomes of internal classifiers next to primary feature vector of samples are considered the ultimate classifier inputs. **Results:** According to the results, the proposed method achieved 90.6% of accuracy, 95.8% of sensitivity, and 75% of specificity, admissible compared to those of other relevant studies. **Conclusion:** Current experimental outcomes provide a comparative analysis of various machine learning classifiers and confirm that using ensemble-based methods has improved medical diagnostic tasks.

Keywords: Classification, ensemble learning, medical diagnostics, parkinson's disease, voice measurements

Submitted: 17-Nov-18

Revision: 06-Mar-19

Accepted: 12-Apr-19

Published: 24-Oct-19

Introduction

Parkinson's disease (PD) is the most common destructive neurological disease after Alzheimer's disease. It was first reported by James Parkinson, a British scientist, in 1817. He named the disease a shaking palsy which is known today as PD.^[1] According to the European Parkinson Disease Association statistics, about 6.3 million people around the world are involved with this disease.^[2] PD is a chronic and progressive disease in which brain dopamine-secreting cells are destroyed. Due to the absence of dopamine, body movements become irregular. Nearly 70%–80% of dopamine-secreting cells are destroyed when primary Parkinson symptoms appear. Tremoring of hands and feet while resting is one of the most common symptoms of the disease. Other symptoms include muscle stiffening, voice changing, slowness of movements, and depression.^[3] The average age for symptoms

to appear is about 60, even though one among twenty samples demonstrates primary symptoms under the age of 50. Unfortunately, elderly people may associate these symptoms with age-related changes. It should be noted that PD is currently untreatable, but the available drugs significantly reduce the symptoms, mainly at the beginning of the disease.^[4] Based on previous studies, about 90% of people with PD have some types of voice abnormalities. Therefore, voice measurements can be used to diagnose and track the progression of PD's symptoms.^[5] In order to determine these symptoms, various verbal experiments are designed. These experiments are categorized into two main groups: running speech experiments and sustained phonation experiments. For the first group of experiments, patients are asked to say an expression that includes a sample of linguistic units and for the second one, they are asked to generate a single vowel.^[6,7] In many studies, mentioned verbal experiments are used as an effective element for determining patients and healthy samples. Sakar *et al.*^[8] used a

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Sheibani R, Nikookar E, Alavi SE. An ensemble method for diagnosis of Parkinson's disease based on voice measurements. *J Med Sign Sens* 2019;9:221-6.

**Razieh Sheibani,
Elham Nikookar,
Seyed Enayatollah
Alavi**

Department of Computer
Engineering, Faculty of
Engineering, Shahid Chamran
University of Ahvaz, Ahvaz, Iran

Address for correspondence:
Mrs. Elham Nikookar,
Department of Computer
Engineering, Faculty
of Engineering, Shahid
Chamran University of Ahvaz,
Ahvaz, Iran.
E-mail: e.nikookar@scu.ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_57_18

Quick Response Code:



dataset of 40 samples including 20 patients. Each person has been taught to say a set of 26 distinct terms, including sustained vowels, numbers, words, and short sentences. They applied the Summarized Leave One Out technique on each person's voice samples and achieved an accuracy of 77.5%. Zhang *et al.*^[9] applied a combination of Multi Edit Nearest Neighbor and an ensemble learning algorithm on the previous dataset. They obtained an improved classification accuracy of 29.44%. Tsanas *et al.*^[10] used a dataset of 42 participants including 6000 voice samples. They evaluated the Unified Parkinson Disease Rating Scale (UPDRS) which indicates the presence and severity of symptoms by applying linear and nonlinear regression analyses. Their study showed a useful accuracy of 7.5 UPDRS points different from the clinician's estimates. Gil and Johnson^[11] used a dataset of 31 samples including 23 patients. They applied artificial neural networks (ANNs) and support vector machines (SVMs) and obtained an accuracy of 90%. Other studies on this dataset are as follows: Ene^[12] applied three types of probabilistic neural networks including incremental search, Monte Carlo search, and hybrid search (HS). Their study reported that mentioned techniques are not significantly different, but the HS technique with an accuracy of 81% performs better. Ullah Khan,^[13] after data preprocessing by means of data cleaning and recovering missing values, applied three methods of k-nearest neighboring (k-NN), AdaBoost (AB), and random forest (RF) to classify the data. Finally, k-NN achieved the best accuracy of 90.2%. Khemphila and Boonjing^[14] applied ANN to classify patients and healthy samples. They obtained about 83.3% of accuracy. Ozcift and Gulden^[15] proposed classifier ensemble construction with a rotation forest approach that achieved a classification accuracy of 87.13%. In this study, the mentioned dataset of Gil and Johnson which is a collection of multiple voice recordings has been used.

Methods

This study presents an ensemble-based method for diagnosing patients and healthy samples by class label prediction based on voice frequency characteristics. It includes three stages of data preprocessing, internal classification, and ultimate classification. In the first stage, for improving the processing time, dataset is separated into six subsets according to recorded voice types. Therefore,

each dataset sample has only one record in each subset. In the next stage, by applying a set of different internal classifiers, several prediction models are generated. Then, the result of each prediction model is calculated and considered as the next stage's input. At the end, ultimate classifiers determine the final class label of the sample. A schematic illustration of the proposed method is depicted in Figure 1.

Because performance evaluation is a vital task in all autodiagnostic systems, three evaluation metrics of accuracy, sensitivity, and specificity have been used to indicate the efficiency of different classifiers. Accuracy is a general metric that allows researchers to measure the efficiency of a method. Sensitivity and specificity refer to the level of sensitivity and specificity of their proposed method, and they are more common in the medical diagnostic domain. The quite description of mentioned metrics is shown in Table 1.

Results

The data for this study, developed by the University of Oxford in cooperation with the National Center for Voice and Speech, Denver, Colorado, include biomedical phonetic measurements of 31 men and 23 women with PD. The duration of diagnosis was from 0 to 28 years, and the age range was between 46 and 85 years, with an average age of 65.8 years. On an average, six voices were recorded for each person.^[16] For the analysis of people with PD,

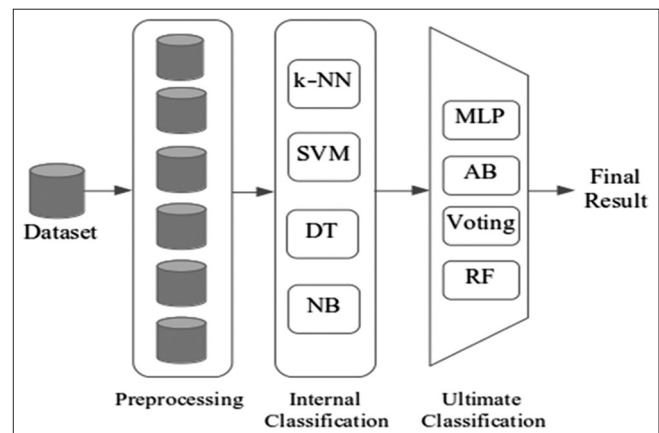


Figure 1: Schematic illustration of the proposed method. MLP: Multilayer perceptron, DT: Decision tree, NB: Naive Bayes

Table 1: Description of performance evaluation metrics

Metric	Formula	Description
Sensitivity	$\frac{TP}{TP + FN} \times 100$	TP refers to cases with PD label that are correctly classified as PD FN refers to cases with PD label that are classified as healthy controls
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN} \times 100$	TN refers to cases with healthy label that are correctly classified as healthy controls FP refers to cases with healthy label that are classified as PD
Specificity	$\frac{TN}{TN + FP} \times 100$	TN refers to cases with healthy label that are correctly classified as healthy controls FP refers to cases with healthy label that are classified as PD

PD – Parkinson's disease; TP – True positive; TN – True negative; FP – False positive; FN – False negative

23 voice frequency characteristics were measured that are demonstrated completely in Table 2.

In order to apply the proposed method, the WEKA software (Developed by University of Waikato, Hamilton, New Zealand) was used. WEKA includes a set of data mining and machine learning algorithms. After initial preprocessing of the data and creating six subsets, several well-known classification algorithms including the k-NN (*k* with values of 1, 3, and 5), SVM, DT, and NB were used. For a better prediction, all these classifiers were applied on the data using 10-fold cross validation. It is worth mentioning that the ABM1 method was used as a resampling method due to the small number of data samples. The obtained accuracy, sensitivity, and specificity measures of applying each internal classifier on the data are summarized in Tables 3-5, respectively.

According to the results, k-NN algorithm (*k* with a value of 1) showed better performance than others with about 90% of accuracy. In the next stage, MLP, AB, RF, and voting were applied as the ultimate classifiers, using 10-fold cross validation. The main motive for using this method was to reduce the error rate. In ensemble learning methods, the probability of a mistake in determining a class label or position of a new sample is much less than the prediction mode with only one model. The obtained accuracy, sensitivity, and specificity measures of applying each ultimate classifier are summarized in Tables 6-8, respectively.

In addition, comparison between the results of ultimate classification stage is depicted in Charts 1-3.

In accordance with the results, both MLP (with SVM and NB kernels) and AB (with NB kernel) ultimate classifiers showed better performance and obtained 90.6% of classification accuracy, 95.8% of sensitivity, and 75.0% of specificity. However, the proposed method has achieved acceptable results by applying an ensemble-based approach; another vital challenge is to compare the current work with other previous studies. For this purpose, some previous studies with similar data are summarized in Table 9. Regarding classification accuracy which is a well-known evaluation metric that is used in related studies, the ensemble-based approach has achieved better performance.

Discussion

As mentioned previously, there is no specific test such as electroencephalography or blood test for diagnosing PD. Moreover, timely diagnosis of this disease is a vital task for clinical reasons. In these cases, developing a prediction method by means of machine learning techniques will be helpful. In order to identify PD patients, a combination of various machine learning classifiers has been investigated. The proposed method takes the advantage of gathering internal classifiers by sending their outcomes to the next classification stage. Initially, by applying internal

Table 2: Description of voice frequency characteristics

Name	Description
MDVP: Fo (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
MDVP: Flo (Hz)	Minimum vocal fundamental frequency
MDVP: Jitter (%)	Fundamental frequency variation measures
MDVP: Jitter (Abs)	
MDVP: RAP	
MDVP: PPQ	
Jitter: DDP	
MDVP: Shimmer	Amplitude variation measures
MDVP: Shimmer (dB)	
Shimmer: APQ3	
Shimmer: APQ5	
MDVP: APQ	
Shimmer: DDA	
NHR	Ratio of noise-to-tonal component measures
HNR	
RPDE	Nonlinear dynamical complexity measures
D2	
DFA	Signal fractal scaling exponent
Spread1	Nonlinear measures of fundamental frequency variation
Spread2	
PPE	
Status	Health status

Table 3: Accuracy measures of applying internal classifiers

	k-NN (<i>k</i> =1)	k-NN (<i>k</i> =3)	k-NN (<i>k</i> =5)	SVM	DT	NB
D1	90.6	62.5	81.2	81.2	87.5	87.5
D2	90.6	78.1	78.1	84.3	87.5	81.2
D3	87.5	84.3	81.2	90.6	93.7	81.2
D4	93.7	75.0	81.2	84.3	90.6	81.2
D5	78.1	75.0	78.1	84.3	93.7	78.1
D6	87.5	75.0	78.1	75.0	78.1	81.2

SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

Table 4: Sensitivity measures of applying internal classifiers

	k-NN (<i>k</i> =1)	k-NN (<i>k</i> =3)	k-NN (<i>k</i> =5)	SVM	DT	NB
D1	91.6	66.6	87.5	83.3	87.5	91.6
D2	95.8	79.1	87.5	87.5	91.6	83.3
D3	91.6	91.6	91.6	95.8	98.0	79.1
D4	95.8	79.1	87.5	87.5	91.6	75.0
D5	79.1	79.1	87.5	87.5	98.0	75.0
D6	91.6	75.0	83.3	83.3	75.0	79.1

SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

classifiers, different models are created on the training data. The outcomes of internal classifiers next to primary feature vector of samples are considered the ultimate classifier

Table 5: Specificity measures of applying internal classifiers

	k-NN (k=1)	k-NN (k=3)	k-NN (k=5)	SVM	DT	NB
D1	87.5	50.0	62.5	75.0	87.5	75.0
D2	75.0	75.0	50.0	75.0	75.0	75.0
D3	75.0	62.5	50.0	75.0	75.0	87.5
D4	87.5	62.5	62.5	75.0	87.5	98.0
D5	75.0	62.5	50.0	75.0	75.0	87.5
D6	75.0	75.0	62.5	50.0	87.5	87.5

SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

Table 6: Accuracy measures of applying ultimate classifiers

	k-NN (k=1)	k-NN (k=3)	k-NN (k=5)	SVM	DT	NB
MLP	78.1	78.1	71.8	90.6	84.3	90.6
AB	78.1	78.1	65.6	87.5	84.3	90.6
Voting	75.0	75.0	75.0	75.0	75.0	75.0
RF	87.5	81.2	71.8	84.3	87.5	87.5

MLP – Multilayer perceptron; AB – AdaBoost; RF – Random forest; SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

Table 7: Sensitivity measures of applying ultimate classifiers

	k-NN (k=1)	k-NN (k=3)	k-NN (k=5)	SVM	DT	NB
MLP	79.1	79.1	83.3	95.8	87.5	95.8
AB	79.1	83.3	75.0	91.6	87.5	95.8
Voting	95.0	95.0	95.0	95.0	95.0	95.0
RF	91.6	83.3	83.3	83.6	91.6	91.6

MLP – Multilayer perceptron; AB – AdaBoost; RF – Random forest; SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

Table 8: Specificity measures of applying ultimate classifiers

	k-NN (k=1)	k-NN (k=3)	k-NN (k=5)	SVM	DT	NB
MLP	75.0	75.0	37.5	75.0	75.0	75.0
AB	75.0	62.5	37.5	75.0	75.0	75.0
Voting	60.0	60.0	60.0	60.0	60.0	60.0
RF	75.0	75.0	37.5	75.0	75.0	75.0

MLP – Multilayer perceptron; AB – AdaBoost; RF – Random forest; SVM – Support vector machine; DT – Decision tree; NB – Naive Bayes; k-NN – k-nearest neighboring

inputs. Then, the final classifiers determine the sample class label which results in a lower error rate. In accordance with the results, both MLP and AB classifiers showed better performance and obtained 90.6% of classification accuracy, 95.8% of sensitivity, and 75% of specificity, admissible compared to those of other relevant studies.

Conclusion

Due to the difficult medical diagnostic of PD and its prevalence, presenting a simple and inexpensive solution

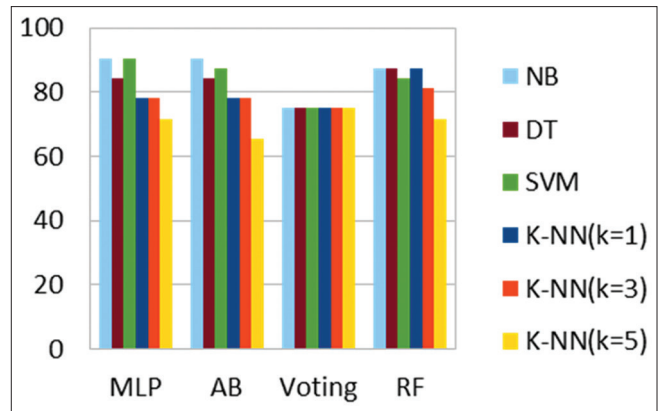


Chart 1: Comparison between accuracy measures of ultimate classification stage

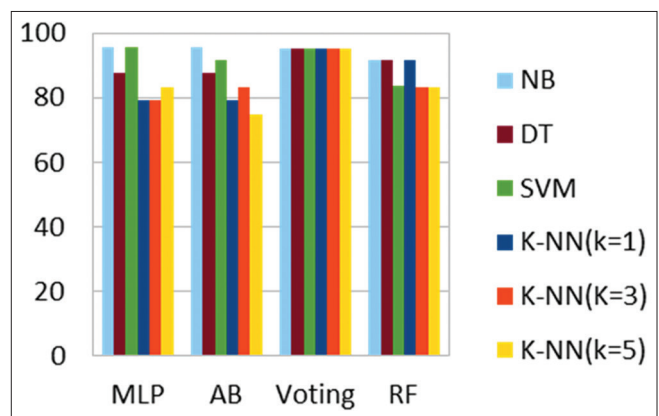


Chart 2: Comparison between sensitivity measures of ultimate classification stage

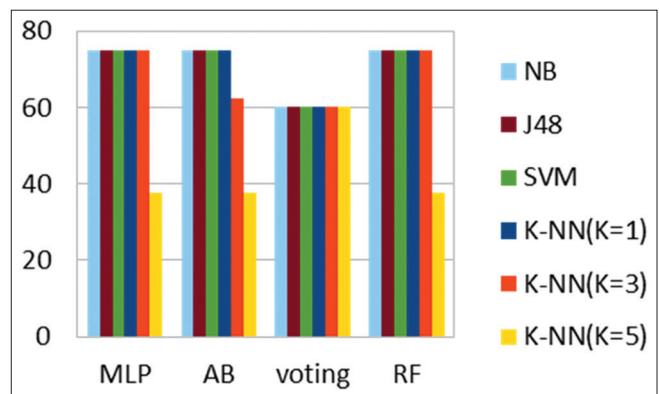


Chart 3: Comparison between specificity measures of ultimate classification stage

for its correct and timely diagnosis is very important. An effective way for diagnosing PD is to compare the voice frequency characteristic of people under controlled conditions. This study presents an ensemble method for detecting patients and healthy samples by class label prediction based on mentioned voice frequency characteristics. It evaluates the idea of merging primary feature vector of samples and predicted class labels. The current experimental results confirm that the idea of using

Table 9: Comparison between the proposed method and other works

Author	Method	Accuracy (%)
Gil and Johnson ^[11]	ANN and SVM	90
Ene ^[12]	IS, MCS and HS	81
Ullah Khan ^[13]	k-NN, AB and RF	90.2
Khemphila and Boonjing ^[14]	ANN	83.3
Ozcift and Gulten ^[15]	Classifier ensemble construction with a rotation forest approach	87.13
Proposed method	Ensemble based	90.6

ANN – Artificial neural network; SVM – Support vector machine; IS – Incremental search; MCS – Monte Carlo search; HS – Hybrid search; AB – AdaBoost; RF – Random forest; k-NN – k-nearest neighboring

the proposed method has improved the machine learning outcomes and also medical diagnostic tasks.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

- DeMaagd G, Philip A. Parkinson's disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis. *P T* 2015;40:504-32.
- Coomber R, Alshameeri Z, Masia AF, Mela F, Parker MJ. Hip fractures and Parkinson's disease: A case series. *Injury* 2017;48:2730-5.
- Müller T. Drug therapy in patients with Parkinson's disease. *Transl Neurodegener* 2012;1:10.
- Singh N, Pillay V, Choonara YE. Advances in the treatment of Parkinson's disease. *Prog Neurobiol* 2007;81:29-44.
- Gupte C, Gadewar S. Diagnosis of Parkinson's disease using acoustic analysis of voice. *Int J Sci Res Netw Secur Communication* 2017;5:14-18.
- Baken RJ, Orlikoff RF. *Clinical Measurement of Speech and Voice*. 2nd ed. US: Cengage Learning; 1999.
- Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, *et al.* A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the committee on phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol* 2001;258:77-82.
- Sakar BE, Isenkul ME, Sakar CO, Sertbas A, Gurgen F, Delil S, *et al.* Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J Biomed Health Inform* 2013;17:828-34.
- Zhang HH, Yang L, Liu Y, Wang P, Yin J, Li Y, *et al.* Classification of Parkinson's disease utilizing multi-edit nearest-neighbor and ensemble learning algorithms with speech samples. *Biomed Eng Online* 2016;15:122.
- Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests. *IEEE J Biomed Health Inform* 2009;57:884-93.
- Gil D, Johnson M. Diagnosing Parkinson by using artificial neural networks and support vector machines. *Global J Comput Sci Tech* 2009;9:63-71.
- Ene M. Neural network-based approach to discriminate healthy people from those with Parkinson's disease. *Comput Sci Ser* 2008;35:112-6.
- Ullah Khan S. Classification of Parkinson's disease using data mining techniques. *J Parkinsons Dis Alzheimer Dis* 2015;2:4.
- Khemphila A, Boonjing V. Parkinson's disease classification using neural network and feature selection. *Int J Math Comput Sci* 2012;6:377-80.
- Ozcift A, Gulden A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed* 2011;104:443-51.
- Little MA, McSharry PE, Roberts SJ, Costello DA, Moroz IM. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomed Eng Online* 2007;6:23.

BIOGRAPHIES



Razieh Sheibani has received her B.Sc. in computer engineering from Shahid Chamran University of Ahvaz, Iran in 2018. Her research interest includes data mining, medical image processing and computational intelligence.

Email: raziieh.sh.agi@gmail.com



Elham Nikookar has received her B.Sc. degree in computer engineering from Shahid Chamran University, Ahvaz, Iran, and her M.Sc. degree in computer engineering from University of Tehran, Tehran, Iran, in 2012. She is currently an instructor and a faculty member in Computer Engineering Department of Shahid Chamran University, Ahvaz, Iran.

Email: e.nikookar@scu.ac.ir



Seyed Enayatollah Alavi has received his B.Sc. degree in computer engineering from Isfahan University, Isfahan, Iran, in 1991, his M.Sc. degree in computer engineering from Shiraz University, Shiraz, Iran, in 1996 and his PhD degree in computer engineering from Belarusian State University, Minsk, Belarus, in 2011. He is currently an assistant professor in Computer Engineering Department of Shahid Chamran University, Ahvaz, Iran.

Email: se.alavi@scu.ac.ir