

Original Paper

An Intelligent Mobile-Enabled System for Diagnosing Parkinson Disease: Development and Validation of a Speech Impairment Detection System

Liang Zhang^{1*}, PhD; Yue Qu^{2*}, MS; Bo Jin², PhD; Lu Jing³, MS; Zhan Gao⁴, PhD; Zhanhua Liang³, PhD

¹International Business College, Dongbei University of Finance and Economics, Dalian, China

²School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, China

³Department of Neurology, The First Affiliated Hospital of Dalian Medical University, Dalian, China

⁴Beijing Haoyisheng Cloud Hospital Management Technology Ltd, Beijing, China

* these authors contributed equally

Corresponding Author:

Lu Jing, MS

Department of Neurology

The First Affiliated Hospital of Dalian Medical University

No.222 Zhongshan Road

Dalian, 116011

China

Phone: 86 18098876262

Email: jinglu131129@126.com

Abstract

Background: Parkinson disease (PD) is one of the most common neurological diseases. At present, because the exact cause is still unclear, accurate diagnosis and progression monitoring remain challenging. In recent years, exploring the relationship between PD and speech impairment has attracted widespread attention in the academic world. Most of the studies successfully validated the effectiveness of some vocal features. Moreover, the noninvasive nature of speech signal-based testing has pioneered a new way for telediagnosis and telemonitoring. In particular, there is an increasing demand for artificial intelligence-powered tools in the digital health era.

Objective: This study aimed to build a real-time speech signal analysis tool for PD diagnosis and severity assessment. Further, the underlying system should be flexible enough to integrate any machine learning or deep learning algorithm.

Methods: At its core, the system we built consists of two parts: (1) speech signal processing: both traditional and novel speech signal processing technologies have been employed for feature engineering, which can automatically extract a few linear and nonlinear dysphonia features, and (2) application of machine learning algorithms: some classical regression and classification algorithms from the machine learning field have been tested; we then chose the most efficient algorithms and relevant features.

Results: Experimental results showed that our system had an outstanding ability to both diagnose and assess severity of PD. By using both linear and nonlinear dysphonia features, the accuracy reached 88.74% and recall reached 97.03% in the diagnosis task. Meanwhile, mean absolute error was 3.7699 in the assessment task. The system has already been deployed within a mobile app called No Pa.

Conclusions: This study performed diagnosis and severity assessment of PD from the perspective of speech order detection. The efficiency and effectiveness of the algorithms indirectly validated the practicality of the system. In particular, the system reflects the necessity of a publicly accessible PD diagnosis and assessment system that can perform telediagnosis and telemonitoring of PD. This system can also optimize doctors' decision-making processes regarding treatments.

(*JMIR Med Inform* 2020;8(9):e18689) doi: [10.2196/18689](https://doi.org/10.2196/18689)

KEYWORDS

Parkinson disease; speech disorder; remote diagnosis; artificial intelligence; mobile phone app; mobile health

Introduction

Parkinson disease (PD) is a long-term degenerative disorder of the central nervous system that mainly affects the motor system. In the early stages, the symptoms include tremor; rigidity; slowness of movement; and difficulty with walking, talking, thinking, or completing other simple tasks. Dementia becomes common in the later stages of the disease. More than a third of patients have experienced depression and anxiety [1]. Other symptoms include sensory and sleep problems. In 2017, PD affected more than 10 million people worldwide, making it the second-most common neurological condition after Alzheimer disease. Currently, there is no cure for PD [2]. Accurate diagnosis, prognosis, and progression monitoring remain nontrivial.

As reported in previous work [3,4], approximately 90% of patients with PD develop voice and speech disorders during the course of the disease, which can have a negative impact on functional communication, thus leading to a decline in the quality of life [5]. Reduced volume (ie, hypophonia), reduced pitch range (ie, monotone), and difficulty with the articulation of sounds or syllables (ie, dysarthria) are the most common speech problems [6]. At the same time, many patients gradually dislike communication because of their own language barriers, which will cause more serious speech disorders and then form a vicious circle. Note that the speech signal-based test is noninvasive and can be self-administered. Hence, it has been regarded as a promising approach in PD diagnosis, evaluation, and progression monitoring, especially in the telediagnosis and telemonitoring medical fields.

In this work, we built a publicly accessible real-time system to efficiently diagnose and assess the severity of PD via speech signal analysis. The most relevant works can be found in Lahmiri et al [7] and Wroge et al [8]. They utilize similar machine learning algorithms as those based on previously proposed audio features [9-13]; however, their work neither considered severity assessment of PD nor made a publicly accessible app that allows for real-time mobile-aided PD diagnosis or evaluation, which is actually a trend and even a necessity in the current 4G and future 5G era for telediagnosis and telemonitoring. For instance, the outbreak of coronavirus disease 2019 (COVID-19) highlights the importance of intelligent and accurate telehealth during disease epidemics.

More specifically, our system first collects the speech signals of the subjects and then utilizes speech signal processing techniques to extract a variety of speech impairment features; it further utilizes advanced machine learning algorithms to diagnose PD and analyze the disease severity. In our work, in the speech signal feature-extraction stage, we utilized many traditional and novel methods to obtain clinically meaningful voice signal features, such as jitter, fine-tuning, recurrence period density entropy, pitch period entropy, signal-to-noise ratio, harmonics-to-noise ratio (HNR), and the mel frequency

cepstral coefficient [9-11]. We regarded the PD diagnosis task as a classification problem and then utilized classical algorithms (eg, support vector machine [SVM] and artificial neural network [ANN]) to perform diagnosis. We formed the PD severity assessment task into a regression problem, with the Unified Parkinson Disease Rating Scale (UPDRS) score as the dependent variable; the UPDRS is the most widely employed scale for tracking PD symptom progression. Various regression algorithms (eg, support vector regression [SVR] and least absolute shrinkage and selection operator [LASSO] regression) were tested. We then obtained the most suitable model by comparing and blending different algorithms. In the end, we developed a mobile phone app for our system to realize remote diagnosis, severity evaluation, and progression monitoring of PD, which will significantly reduce detection and prevention costs.

The main structure of this paper is divided into four parts: (1) description of the methods used in our system: data collection, data preprocessing, feature extraction of speech signals, classification, and regression problem formulation, (2) analysis of our experimental results, (3) system description of our mobile app, and (4) final discussion.

Methods

Data Collection

The speech signal data used in the experiment came from two sources:

1. One part of the dataset came from the open data platform from the University of California Irvine (UCI) Machine Learning Repository, where three sets of parkinsonian speech data with different characteristics were obtained.
2. The other part of the dataset was collected in collaboration with the Department of Neurology, the First Affiliated Hospital of Dalian Medical University, China. The data recorded the voice signals of patients with PD.

In practice, the collected pronunciation content needs to be short and reflect the patient's speech disorder to a certain extent. On one hand, considering the need for different languages, dialects, and accents as well as unclear pronunciations, we adopted the continuous pronunciation method. Meanwhile, the control of the vocal cords and airflow is also weakened due to the weakening control of the pronunciation system of the nervous system. On the other hand, since the relationship between the vibration of the vocal cords and the speech disorder is relatively strong, the vowels can better reflect the degree of speech impairment [6,11,14]. Another fact is that the basic vowels in different regions of the world are very similar, so it is more reasonable to use vowels. The vowels used here are the five long vowels with the following English phonetic symbols: [a:], [i:], [e:], and [u:]; the subjects are required to pronounce them repeatedly. The collected syllables are shown in Table 1.

Table 1. Collected syllables.

International phonetic symbol	Duration (seconds)
[ɑ:]	3
[ɜ:]	3
[i:]	3
[ɔ:]	3
[u:]	3

The UPDRS [15] is the most commonly used severity indicator in clinical studies of PD. It is evaluated via filling out a form, which requires considerable medical expertise, so it is difficult for patients to perform self-testing using this scale. That explains why we need automatic and artificial intelligence-powered prediction tools. We collected the UPDRS score as the dependent variable in our regression task. At present, UPDRS version 3.0 is the most widely used version, and it can be divided into four parts:

1. Mentation, behavior, and mood, including a total of four questions (16 points).
2. Activities of daily living, including a total of 13 questions (52 points).
3. Motor examination, including a total of 14 questions (108 points).
4. Treatment complications, including a total of 11 questions (23 points).

In summary, UPDRS version 3.0 has a total of 42 questions and the highest score is 199 points. The higher the UPDRS score, the more serious the PD is. The third item, *motor examination*, can reflect the severity of speech disorder. In practice, when collecting the data, the doctor is required to evaluate the total UPDRS score as well as the value of the motor examination score.

Note that the first part of the data is open source, and we can easily download this data from UCI's official website. Therefore, the datasets were mainly used to train the machine learning models and verify the validity of the dysphonia features, all of which have been integrated in our app system. This part of the data will be introduced in detail in the Results section. The second part of the data requires us to work closely with local hospitals—we collected PD patients' vocal data in a local hospital; the data collection table is shown in [Multimedia Appendix 1](#). The Data Preprocessing section that follows describes how we processed the second part of the data, which has been implemented as a function in our app. We then extracted the dysphonia features, which have also been integrated as a function in our app system. Moreover, we tested

them with machine learning models, which have been trained based on the first part of data, and achieved good results in the PD diagnosis task. Until now, the number of collected Chinese speech signals is still not big enough to train an effective model. Therefore, our model within our app system was trained by the first part of the data: the first and third datasets were used in the diagnosis task and severity assessment task, respectively. However, this app is continuously collecting new data, including positive and negative samples. As the amount of data increases in the future, we will utilize advanced technology, such as transfer learning, to realize PD diagnosis and severity evaluation for people in various regions.

Data Preprocessing

The initially collected voice signals cannot be directly used; some preprocessing was required. This operation removed some of the interference factors and paved the way for subsequent feature extraction. The formats of different audio files were unified into the WAV file format, with 44,100 Hz sampling frequency and two channels. These audio files were then uploaded into the back-end server for storage.

The first step of data preprocessing is *sampling frequency conversion*, that is, resampling, which can uniformly record the speech frequency and reduce the amount of calculation by down-clocking. In our work, only one channel of the speech signal (ie, the left channel) is reserved, and then the sampling frequency is converted to 10 kHz.

The second step is *pre-emphasis*. Since the low-frequency part of speech signals tends to contain noise, we performed pre-emphasis to filter out the low frequencies and improve the resolution of the high-frequency part of speech signals. In our work, a first-order, finite impulse response, high-pass digital filter was used to achieve pre-emphasis [16]. The transfer function is defined in equation 1 of [Figure 1](#). In equation 1, a is the pre-emphasis coefficient; generally, $0.9 < a < 1.0$. Let $x(n)$ denote the voice sample value at time n . After the pre-emphasis processing, the result is $y(n) = x(n) - ax(n-1)$, where $a=0.9375$.

Figure 1. Equations 1-10. FN: false negative; FP: false positive; MAE: mean absolute error; MSE: mean square error; RMSE: root mean square error; TN: true negative; TP: true positive.

$$H(z) = 1 - az^{-1} \quad (1)$$

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad (2)$$

$$R_{xx}(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \quad (3)$$

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (4)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |T_i - P_i| \quad (8)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (T_i - P_i)^2 \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (T_i - P_i)^2} \quad (10)$$

The third step is *windowing and framing*. The speech signal was divided into some shorter signal segments (ie, frames) for processing, which is the framing process, such that the signal can be treated as stationary in the short-time window. In practice, to reduce the impact of segmenting on the statistical properties of the signal, we applied windowing to the temporal segments. The frame width in our work was set as 25 milliseconds long, the frame shift was 10 milliseconds long, and the Hamming window was leveraged as the window function.

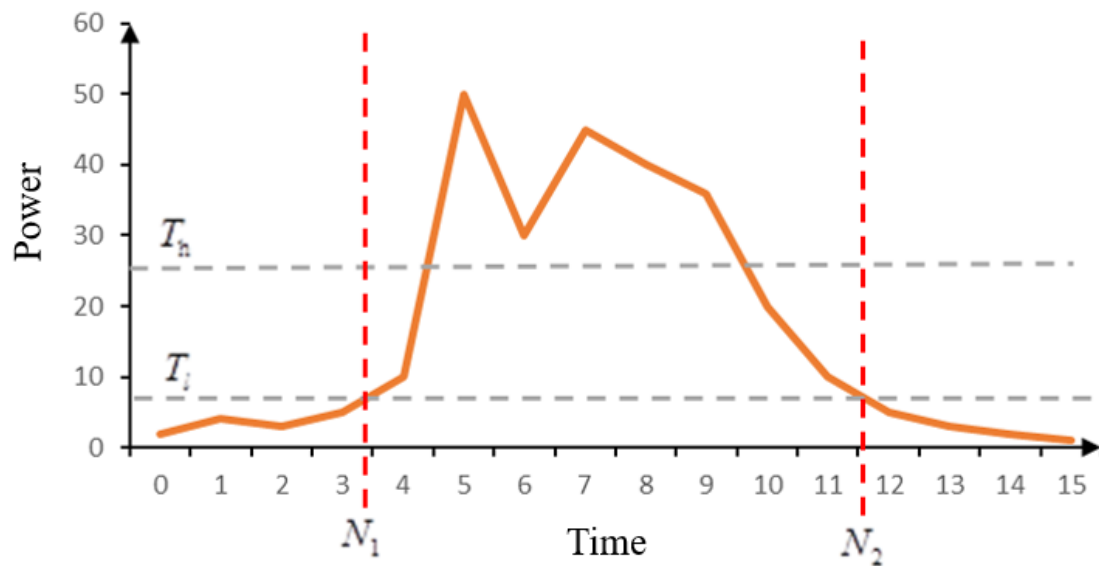
The fourth step is *silent discrimination*. Because there is no guarantee that the collected audio files will always have sound, it is necessary to filter out the blank periods of those sounds. Therefore, silent discrimination, also known as voice endpoint detection, was required. A common solution is to use double-threshold methods [17], which are based on the

principles of short-time energy, short-term average amplitude, and short-time zero-crossing rate. In our work, for the sake of simplicity and algorithm efficiency, we utilized only short-term energy as the principle for the double-threshold method. The definition of short-term average energy is shown Figure 1, equation 2.

For illustration, as is shown in Figure 2, we let T_h and T_l denote the upper and lower thresholds, respectively. The voiced part must have a section above T_h . The endpoint energy of the voiced part is equal to T_l . N_1 is the starting point, N_2 is the ending point, and w is the Hamming window. The fifth step is *fundamental frequency extraction*. The fundamental frequency refers to the lowest and theoretically strongest frequency in the sound, which reflects the vibration frequency of the sound source. In our work,

we adopted the most widely used autocorrelation method to extract the fundamental frequency.

Figure 2. Principle of the double-threshold method. N_1 : starting point; N_2 : ending point; T_h : upper threshold; T_l : lower threshold.



The short-term autocorrelation function is defined in Figure 1, equation 3. We need to obtain the first positive peak point, $R_{xx}(k_f)$, after crossing the zero point in sequence $R_{xx}(k)$, and $1/k_f$ is the extracted fundamental frequency.

Note that the audio files may be mixed with unknown noise, which can cause a sudden jump at some points. These points are called wild points or outliers. Therefore, it is necessary to initially remove the wild points. We first calculated the average

value of the fundamental frequency of the audio and then deleted the point that was too far from the average value.

Dysphonia Features

In 2012, Tsanas et al summarized 132 features of speech impairments [11]. Considering the speed requirement of the real-time system, the selected model cannot use all of the features. The final selected features [18-23] are illustrated in Table 2.

Table 2. Dysphonia features.

Classification and dysphonia features	Description
Pitch [18] (fundamental frequency)	
F_0 _mean	Mean of pitch
F_0 _max	Max of pitch
F_0 _min	Min of pitch
F_0 _median	Median of pitch
F_0 _std	SD of pitch
Jitter [18] (pitch period perturbation)	
Jitter	Jitter
Jitter_abs	Absolute jitter
Jitter_PPQ5	5 adjacent points' jitter
Jitter_rap	3 adjacent points' jitter
Jitter_ddp	Difference of 3 adjacent points' jitter
Shimmer [18] (amplitude perturbation)	
Shimmer	Shimmer: percentage
Shimmer_dB	Shimmer: decibels (dB)
Shimmer_APQ5	5 adjacent points' shimmer
Shimmer_APQ3	3 adjacent points' shimmer
Shimmer_dda	Difference of 3 adjacent points' shimmer
Shimmer_APQ11	11 adjacent points' shimmer
Harmonics-to-noise ratio (HNR) and noise-to-harmonics ratio (NHR) [19]	
HNR_mean	Mean of HNR
HNR_std	SD of HNR
NHR_mean	Mean of NHR
NHR_std	SD of NHR
Nonlinear feature	
DFA	Detrended fluctuation analysis [20]
RPDE	Recurrence period density entropy [21]
D2	Correlation dimension [22]
PPE	Pitch period entropy [23]

Problem Formulation

Diagnosis

Because the predicted value in PD diagnosis is discrete and binary, it can be regarded as a two-category classification problem. This paper chose the following classical classification algorithms: (1) SVM, (2) ANN, (3) Naive Bayes, and (4) logistic regression.

Severity Assessment

Because the predicted value (ie, the UPDRS score) is continuous in the assessment of the severity of speech impairment in PD,

it can be seen as a regression problem. This paper chose the following classical regression algorithms: (1) SVR, (2) linear regression, and (3) LASSO regression.

Results

Overview

We should initially introduce some indicators to evaluate the quality of the algorithms. First, for a two-category classification problem, there are usually the following classification results, as seen in [Table 3](#).

Table 3. Classification confusion matrix.

Class	Predictive class	Predictive negative class
Actual positive class	True positive (TP)	False negative (FN)
Actual negative class	False positive (FP)	True negative (TN)

Then, the indicators are generally employed to evaluate the classification effect (see [Figure 1](#), equations 4-7). The accuracy represents the proportion of subjects who are classified correctly out of the total number of subjects; precision indicates the proportion of real patients who are predicted to be sick; recall indicates the proportion of patients who are predicted to be sick; and the F1 value is the harmonic mean of the accuracy rate and the recall rate. In our PD diagnosis task, if a normal user is detected to be sick, the impact is usually not large, since we can continue to check the result using various clinical methods. However, if a model fails to detect PD, the impact is relatively large. Hence, the most important indicator is the recall rate.

Second, for a regression problem, if the total number of samples is N , the true value of the i -th sample is T_i , and the predicted

value is P_i , then the indicators in equations 8-10 (see [Figure 1](#)) are available. Among the indicators, mean absolute error (MAE) measures the average magnitude of the errors in a set of predictions, without considering their direction; mean square error (MSE) and root mean square error (RMSE) are quadratic scoring rules that also measure the average magnitude of the error. However, both MSE and RMSE give a relatively higher weight to large errors. As a result, they are more useful when large errors are particularly undesirable.

According to the characteristics of the dataset, different experiments were performed on the three kinds of datasets downloaded from UCI. The characteristics of these datasets are shown in [Table 4](#).

Table 4. Characteristics of three datasets from the University of California Irvine.

Data characteristics	Dataset 1	Dataset 2	Dataset 3
Creation date (year/month/day)	2008/06/26	2014/06/12	2009/10/29
Number of subjects			
Parkinson disease	23	48	42
Non-Parkinson disease	8	20	0
Number of records (ie, samples)	195	1208	5875
Number of features	22	26	18
Task	Classification	Classification and regression	Regression

All results are based on experiments with 5-fold cross validation. To evaluate our models' efficiency and effectiveness, for the PD diagnosis (ie, classification task), the ratio of the training set to the validation set was 4:1 in the first two datasets. We then used a dataset collected from a local hospital as the test dataset. The data collection table is shown in [Multimedia Appendix 1](#). We collected a dataset that included 14 PD patients

and 30 non-PD patients in total. For the PD severity evaluation (ie, regression task), the ratio of training set to the validation set to the testing set was 4:1:1 in the third dataset. The testing results are shown in the following paragraphs.

For the first set of data [9], we conducted classification experiments according to a combination of linear and nonlinear features; the final result is shown in [Table 5](#).

Table 5. Classification results for the first set of data.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Support vector machine	88.74 ^a	88.89	97.03	92.55
Logistic regression	85.71	89.97	91.32	90.18
Neural network (single layer)	88.68	91.16	94.26	92.45
Neural network (double layer)	88.63	92.55	93.38	92.71
Naive Bayes	69.24	96.02	62.37	75.21

^aItalics represent the highest values.

We can see that the combination of linear and nonlinear features for the diagnosis of PD patients is feasible and effective. The SVM algorithm achieved higher accuracy and recall rate, and the Naive Bayes algorithm had the worst effect. According to the previous discussion, the recall rate is the most important

indicator. At the same time, considering the speed requirement of the mobile app, our system finally leveraged the SVM algorithm to perform the PD patient diagnosis. From [Multimedia Appendix 2](#), we can see that these features have small P values,

especially for the nonlinear features, which statistically show the effectiveness of these features.

For the second set of data [24], we conducted classification experiments using only linear features, and the final result is demonstrated in Table 6.

Table 6. Classification results for the second set of data.

Algorithm	Accuracy, %	Precision, %	Recall, %	F1 score, %
Support vector machine	66.71	66.37	<i>83.71</i> ^a	73.98
Logistic regression	66.56	67.68	79.08	72.84
Neural network (single layer)	70.78	71.13	81.54	75.89
Neural network (double layer)	70.29	<i>71.45</i>	80.81	75.40
Naive Bayes	59.36	61.80	73.78	67.19

^aItalics represent the highest values.

It can be clearly seen that using only linear features for PD diagnosis brings about a poor model performance, which is consistent with the conclusion from Tsanas et al [11] that feeding linear features into speech models is not very satisfactory. Meanwhile, some researchers claimed that nonlinear features are more effective [23], and another PD

speech dataset analysis study [24] also obtained similar results. In particular, our experimental results showed that the SVM algorithm achieved a relatively high recall rate.

For the third set of data (ie, regression) [25], we tested multiple regression algorithms on the third dataset. The final result is illustrated in Table 7.

Table 7. Regression results on the third dataset.

Algorithm	Mean absolute error	Mean square error	Root mean square error
Linear regression	8.0786	95.1344	9.7494
Support vector machine	<i>3.7699</i> ^a	<i>34.1202</i>	<i>5.8357</i>
Least absolute shrinkage and selection operator	8.0687	91.1600	9.7452

^aItalics represent the best values.

Experimental results showed that both linear and nonlinear features contribute to the severity assessment of PD patients. Among regression algorithms, the SVR algorithm achieved the best performance on each indicator, and the prediction results of LASSO and linear regressions were not much different; the reason for this is that LASSO regression is actually a variant of linear regression. Hence, the system finally adopted SVR as the severity evaluation method.

In particular, we selected the best results from each algorithm and observed the degree of fit. Figures 3-5 show the fitting

results of the aforementioned three methods. In each figure, the upper graph is the degree of fitting of the training set and the lower graph is the degree of fitting of the test set; the red line is the predicted value and the blue line is the true value. It can be seen from these three figures that SVR fits the best.

As we know, LASSO can perform feature selection [26] by setting the feature weights to zero. The five characteristics most relevant to the value are shown in Table 8.

Figure 3. Linear regression fitting. The red line is the predicted value and the blue line is the true value. UPDRS: Unified Parkinson's Disease Rating Scale.

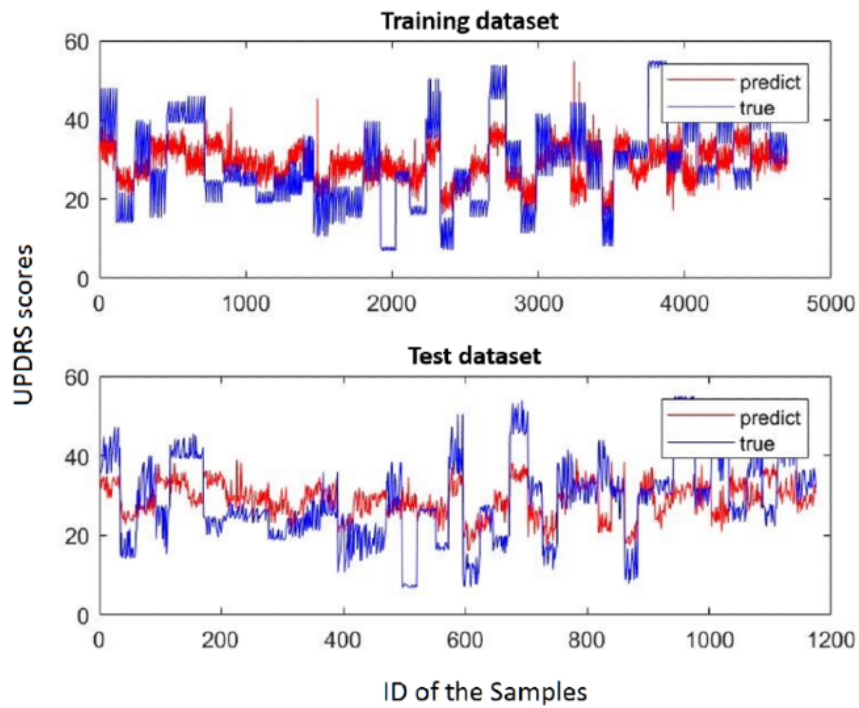


Figure 4. Support vector regression (SVR) fitting. The red line is the predicted value and the blue line is the true value. UPDRS: Unified Parkinson's Disease Rating Scale.

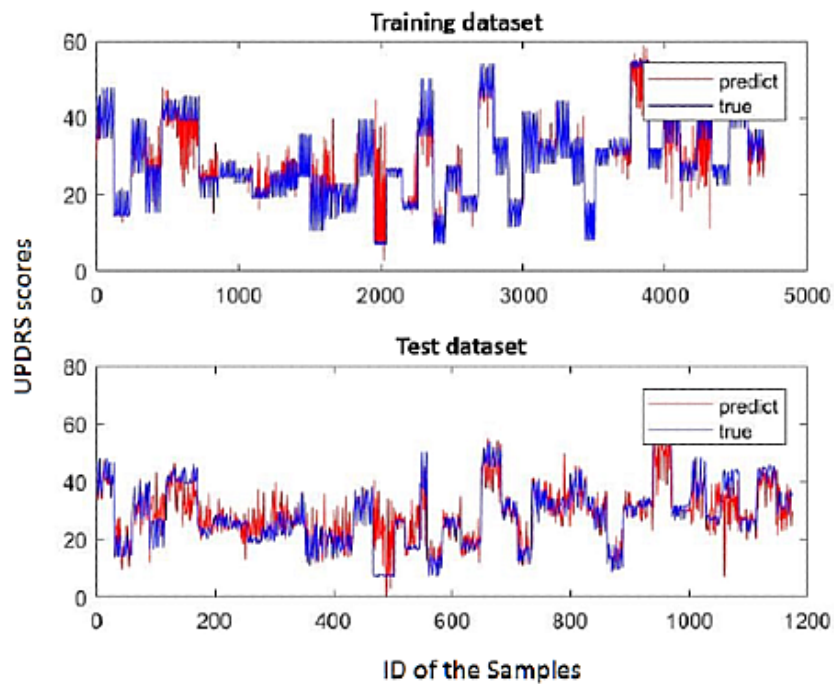


Figure 5. Least absolute shrinkage and selection operator (LASSO) fitting. The red line is the predicted value and the blue line is the true value. UPDRS: Unified Parkinson's Disease Rating Scale.

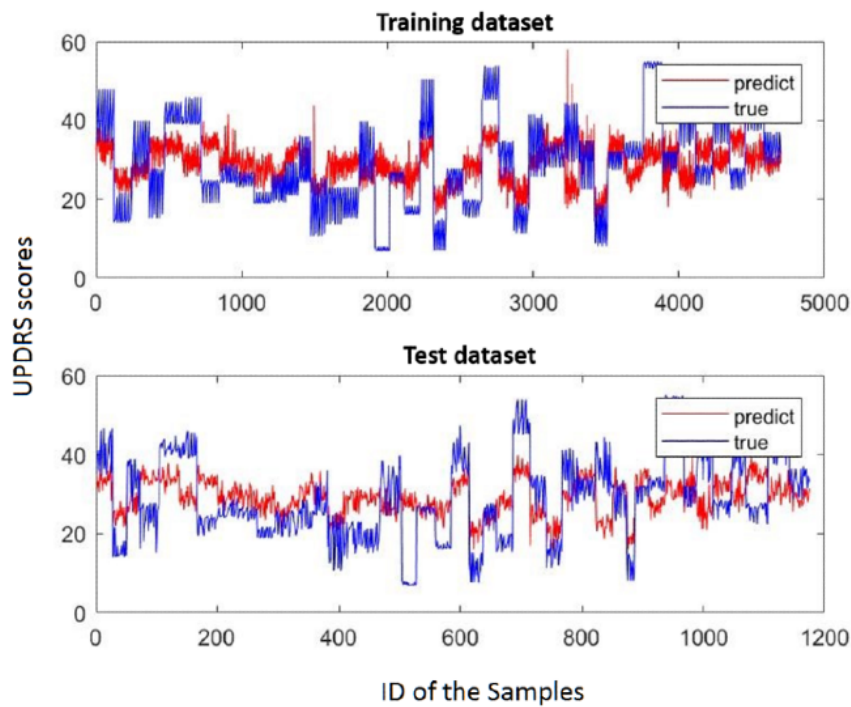


Table 8. Top five principal characteristics.

Feature	Corresponding weighted value
Age	2.84
Harmonics-to-noise ratio mean	-2.66
Absolute jitter	-2.18
Detrended fluctuation analysis	2.14
Pitch period entropy	1.51

It can be considered that these five characteristics are highly correlated with the UPDRS score. Age itself is highly related to PD, and the rest of the characteristics have three nonlinear features—HNR mean is also a nonlinear feature—indicating the importance of nonlinear features. Gender also explains why the regression result of the second set of data was relatively poor.

From [Multimedia Appendix 2](#), we see that these features all have small P values—the features of the Jitter series may be a bit higher than others—which proves that we need these features for our system.

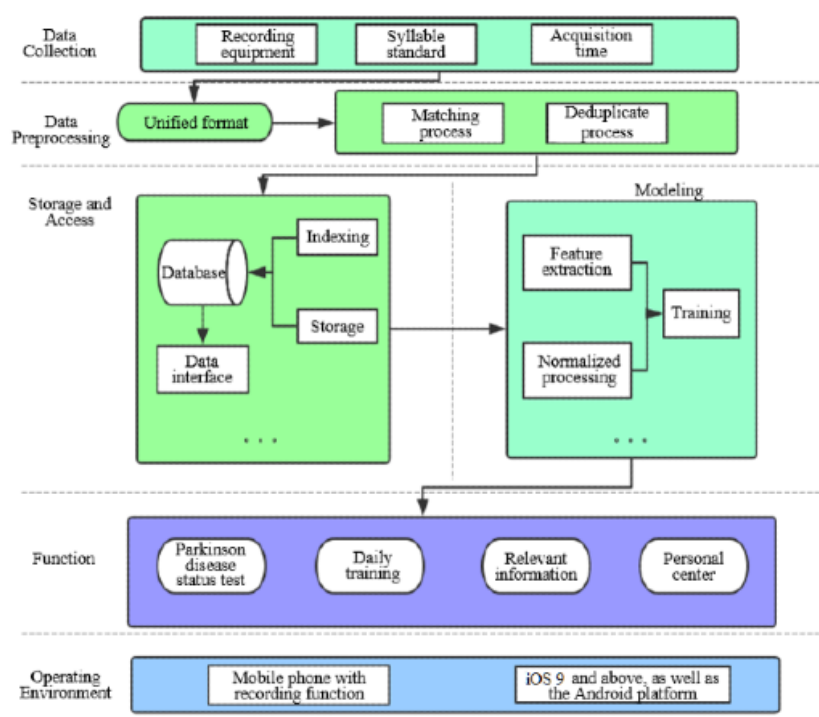
In summary, the PD speech detection system uses SVM and SVR for PD speech diagnosis and severity assessment, respectively.

System

System Overview

[Figure 6](#) shows the architecture of our app system—called the No Pa app—including voice signal collection, data preprocessing, data storage and access, and signal modeling. At its core, the PD diagnosis model is SVM trained by the first set of data and the PD severity assessment model is SVR trained by the third set of data. Meanwhile, [Figure 6](#) displays the four key functions and the operating environment in the application layer.

Figure 6. Architecture overview of the No Pa app system.

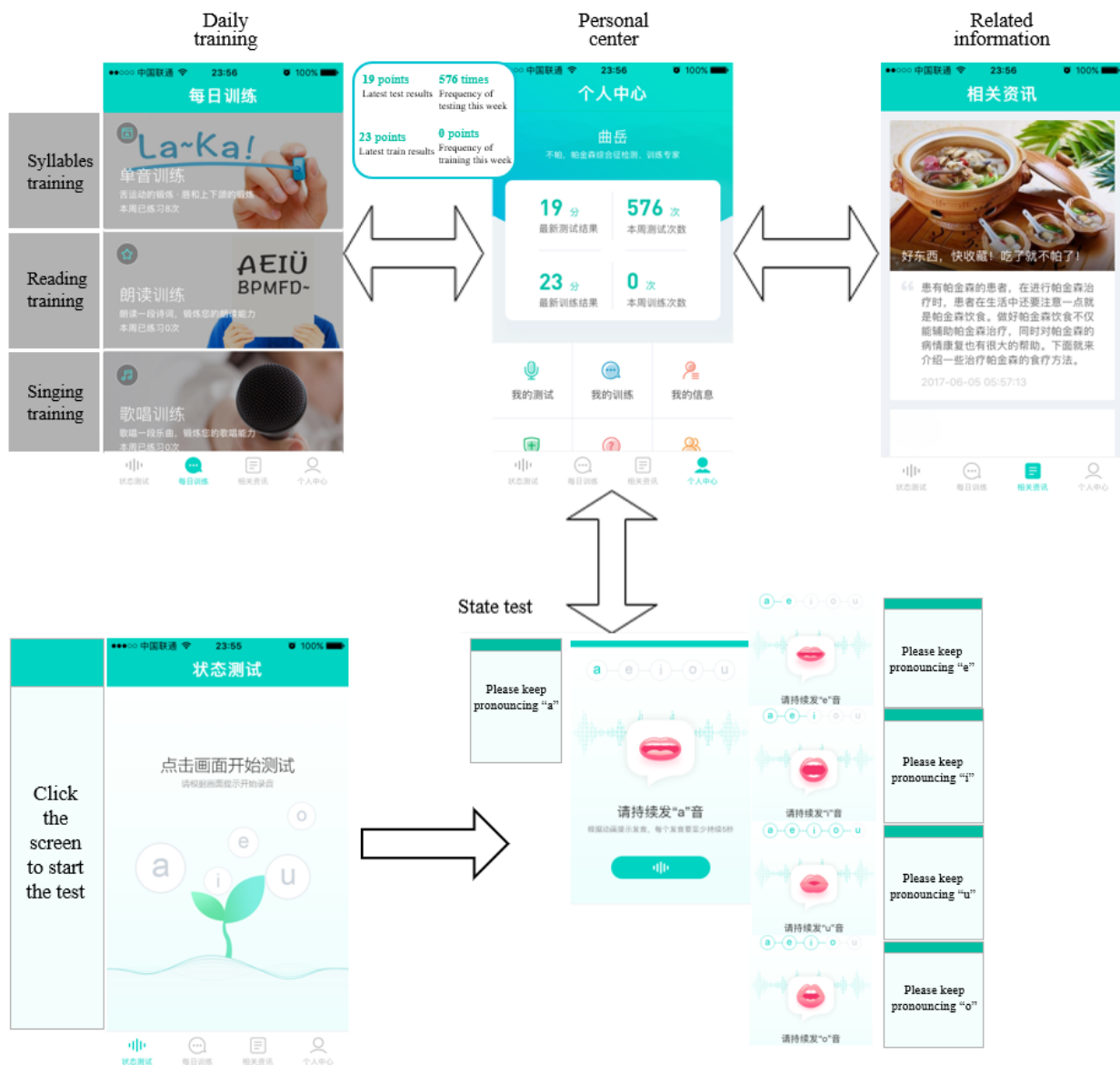


The Main Function

Android and iOS versions of the No Pa mobile app are currently available online. The app includes four functions—state test, daily training, related information, and personal center—which are shown as follows (see [Figure 7](#) for a few screen captures):

1. State test: the subject pronounces five long vowels according to the voice guidance, and each long vowel sound lasts for 5 seconds. Then, our system will calculate the current speech impairment severity status.
2. Daily training: the daily training function aims to improve subjects' speech impairment status by encouraging them to speak. It includes monophonic training, reading training, and singing training. Monophonic training includes the user's pronunciation training according to some specific single syllables; during reading training, the user reads ancient poetry; and singing training improves the user's daily training interest via singing songs. Note that each training function will give a corresponding feedback score according to our speech signals model. However, since the calculation is not based on the five long vowels, the scores may not be accurate, but it is acceptable since our aim is to attract subjects' attention to daily training in speaking.
3. Related information: this function provides users with some advice about PD and physical health.
4. Personal center: this function helps the user view their testing history and some personal information.

Figure 7. Screen captures from the No Pa app showing four functional modules.



Back-End Configuration

The back-end server of the No Pa app is the Alibaba Cloud Server. Its configuration is as follows: 4-core central processing unit (CPU), 8 GB RAM, 64-bit Ubuntu system, and 200 GB disk space.

Algorithm Acceleration

The original system’s computational cost can range from 20 to 30 seconds without any acceleration techniques. Experimental results showed that autocorrelation calculation is the most time-consuming unit, so the C++ programming language was used to accelerate the autocorrelation calculation. To speed up the system, we adopted MEX (MATLAB executable) technology [27] as the acceleration scheme. In the end, the computational cost for predicting UPDRS scores was compressed from 20 seconds to only about 1 second. This response time is acceptable for an app.

Guide and Interaction

For better a user experience, we provided voice-guided navigation that can offer step-by-step instructions. Meanwhile, considering that PD patients may suffer from hand tremors, we designed big buttons in this app. Moreover, if they do not click the recording function button or the system fails to record an effective sound, the system will give them a reminder.

Discussion

Principal Findings

Traditionally, PD patients need to be diagnosed by physical examination. We can now use a mobile app to help conduct straightforward and rapid detection. For PD patients or healthy people, instant detection and consistent monitoring of disease conditions are extremely important. For doctors, the app can be used as a decision-support tool to provide assistance in treatment and diagnosis.

We have built this mobile app by embedding a voice-oriented system. At the core of the system are machine learning algorithms. Experimental results showed that SVM and SVR achieved the best performance for the diagnosis (ie, classification task) and severity evaluation (ie, regression task) of PD, respectively. The recall rate of the classification task can reach 97.03% (ie, the patient's recognition ability), and the absolute average error of the regression task can reach 3.7699, which is acceptable since the value of UPDRS scores range from 0 to 199.

Finally, we will summarize the contributions of our work. We have built a voice-oriented system that can remotely and conveniently diagnose PD. The system first collects a user's five long vowels and then efficiently extracts dysphonia features, such that machine learning algorithms can be applied to the classification or regression of PD-related tasks. First, the system has been integrated into an app for public use. Second, our experiments have validated the effectiveness of voice signal-related features proposed by mainstream studies. Third, our system incorporates voice signal collection, feature extraction, and an algorithm interface, which can be regarded as a standard open-source platform for new algorithm development in voice signal-oriented disease identification tasks.

Comparison With Prior Work

There have been various studies utilizing vocal features for PD diagnosis or severity evaluation. More specifically, Lahmiri et al [7] proposed a study about diagnosing PD based on dysphonia

measures. They chose the same dataset as our first dataset and their results are similar to ours. However, our method achieved a higher recall value on this dataset. Wroge et al [8] also focused on PD diagnosis by speech signal analysis. After some speech signal processing, they extracted two groups of features—Audio-Visual Emotion recognition Challenge (AVEC) [12] and Geneva Minimalistic Acoustic Parameter Set (GeMAPS) features [13]—which were then fed into some machine learning models. However, their feature extraction process relied on some existing tools, which are not easily integrated into an app. In particular, their work needs to extract 1262 features while our work only extracts 24 features. Moreover, the accuracy of their results based on SVM and ANN were both lower than ours. Similar work that is based on the above features can be found in Tracy et al [28]. Deep learning methods have also been leveraged to learn patterns from vocal feature sets [29]. However, their model lacks explanations due to the inherent nature of deep learning models and achieves an inferior performance compared with our model. Moreover, besides PD diagnosis, our system realizes PD severity evaluation, which may be more helpful for patients and doctors.

Limitations

Our data were collected from healthy people and patients with PD from Dalian, China; the quantity of data is still not big enough. In the future, we plan to collect more disease-related data from different regions worldwide to improve the generalization of the model. At the same time, we will use deep learning methods to study the speech signals of patients with PD to avoid cumbersome manual extraction of speech signals.

Acknowledgments

We are deeply indebted to people from the Department of Neurology, the First Affiliated Hospital of Dalian Medical University, China, who supported us with data collection and offered us a lot of professional advice. In addition, we owe our thanks to all volunteers, who allowed us to collect their voice data with patience, and all users of the No Pa app. This study was supported by the National Key R&D Program of China (2018YFC0116800), the National Natural Science Foundation of China (No. 61772110), the China Education and Research Network (CERNET) Innovation Project (NGII20170711), the Program of Introducing Talents of Discipline to Universities (Plan 111) (No. B20070), and the Liaoning Provincial Hospital Reform Focus on Diagnosis and Treatment in Clinical Departments: Parkinson's Disease and Related Movement Disorders in Individualized Diagnosis and Neural Control Platform Construction Project (LNCCC-C06-2015).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data collection table.

[[XLSX File \(Microsoft Excel File\), 10 KB-Multimedia Appendix 1](#)]

Multimedia Appendix 2

P values for linear and nonlinear features.

[[XLSX File \(Microsoft Excel File\), 13 KB-Multimedia Appendix 2](#)]

References

1. Kano O, Ikeda K, Cridebring D, Takazawa T, Yoshii Y, Iwasaki Y. Neurobiology of depression and anxiety in Parkinson's disease. *Parkinsons Dis* 2011;2011:143547 [[FREE Full text](#)] [doi: [10.4061/2011/143547](https://doi.org/10.4061/2011/143547)] [Medline: [21687804](https://pubmed.ncbi.nlm.nih.gov/21687804/)]

2. Davie CA. A review of Parkinson's disease. *Br Med Bull* 2008;86:109-127 [FREE Full text] [doi: [10.1093/bmb/ldn013](https://doi.org/10.1093/bmb/ldn013)] [Medline: [18398010](https://pubmed.ncbi.nlm.nih.gov/18398010/)]
3. Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S. Speech impairment in a large sample of patients with Parkinson's disease. *Behav Neurol* 1999;11(3):131-137 [FREE Full text] [doi: [10.1155/1999/327643](https://doi.org/10.1155/1999/327643)]
4. Mahler LA, Ramig LO, Fox C. Evidence-based treatment of voice and speech disorders in Parkinson disease. *Curr Opin Otolaryngol Head Neck Surg* 2015;23(3):209-215 [FREE Full text] [doi: [10.1097/moo.0000000000000151](https://doi.org/10.1097/moo.0000000000000151)]
5. Miller N, Noble E, Jones D, Burn D. Life with communication changes in Parkinson's disease. *Age Ageing* 2006 May;35(3):235-239. [doi: [10.1093/ageing/afj053](https://doi.org/10.1093/ageing/afj053)] [Medline: [16540492](https://pubmed.ncbi.nlm.nih.gov/16540492/)]
6. Fox CM, Morrison CE, Ramig LO, Sapir S. Current perspectives on the Lee Silverman Voice Treatment (LSVT) for individuals with idiopathic Parkinson disease. *Am J Speech Lang Pathol* 2002 May;11(2):111-123 [FREE Full text] [doi: [10.1044/1058-0360\(2002/012\)](https://doi.org/10.1044/1058-0360(2002/012))]
7. Lahmiri S, Dawson DA, Shmuel A. Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. *Biomed Eng Lett* 2018 Feb;8(1):29-39 [FREE Full text] [doi: [10.1007/s13534-017-0051-2](https://doi.org/10.1007/s13534-017-0051-2)] [Medline: [30603188](https://pubmed.ncbi.nlm.nih.gov/30603188/)]
8. Wroge TJ, Özkanca Y, Demiroglu C, Si D, Atkins DC, Ghomi RH. Parkinson's disease diagnosis using machine learning and voice. In: *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. 2018 Presented at: IEEE Signal Processing in Medicine and Biology Symposium (SPMB); December 1, 2018; Philadelphia, PA p. 1-7. [doi: [10.1109/spmb.2018.8615607](https://doi.org/10.1109/spmb.2018.8615607)]
9. Little M, McSharry P, Hunter E, Spielman J, Ramig L. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat Precedings* 2008 Sep 12:1-27 [FREE Full text] [doi: [10.1038/npre.2008.2298.1](https://doi.org/10.1038/npre.2008.2298.1)]
10. Guo PF, Bhattacharya P, Khanna N. Advances in detecting Parkinson's disease. In: *Proceedings of the 2nd International Conference on Medical Biometrics*. 2010 Jun Presented at: 2nd International Conference on Medical Biometrics; June 28-30, 2010; Hong Kong, China p. 306-314 URL: https://doi.org/10.1007/978-3-642-13923-9_33 [doi: [10.1007/978-3-642-13923-9_33](https://doi.org/10.1007/978-3-642-13923-9_33)]
11. Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. *IEEE Trans Biomed Eng* 2012 May;59(5):1264-1271 [FREE Full text] [doi: [10.1109/tbme.2012.2183367](https://doi.org/10.1109/tbme.2012.2183367)]
12. Valstar M, Schuller B, Smith K, Eyben F, Jiang B, Bilakhia S, et al. AVEC 2013: The continuous audio/visual emotion and depression recognition challenge. In: *Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '13)*. 2013 Presented at: 3rd ACM International Workshop on Audio/Visual Emotion Challenge (AVEC '13); October 21-25, 2013; Barcelona, Spain p. 3-10 URL: <https://dl.acm.org/doi/10.1145/2512530.2512533> [doi: [10.1145/2512530.2512533](https://doi.org/10.1145/2512530.2512533)]
13. Eyben F, Scherer KR, Schuller BW, Sundberg J, Andre E, Busso C, et al. The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for voice research and affective computing. *IEEE Trans Affect Comput* 2016 Apr 1;7(2):190-202. [doi: [10.1109/taffc.2015.2457417](https://doi.org/10.1109/taffc.2015.2457417)]
14. Tsanas A. *Accurate Telemonitoring of Parkinson's Disease Symptom Severity Using Nonlinear Speech Signal Processing and Statistical Machine Learning* [doctoral thesis]. Oxford, UK: University of Oxford; 2012 Jun. URL: https://ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74/download_file?file_format=pdf&safe_filename=DPhil%2Bthesis_post_viva_v8.pdf&type_of_work=Thesis [accessed 2012-12-30] [WebCite Cache ID [ora.ox.ac.uk/objects/uuid:2a43b92a-9cd5-4646-8f0f-81dbe2ba9d74](https://www.webcitation.org/1011452530)]
15. Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. The Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations. *Mov Disord* 2003 Jul;18(7):738-750. [doi: [10.1002/mds.10473](https://doi.org/10.1002/mds.10473)] [Medline: [12815652](https://pubmed.ncbi.nlm.nih.gov/12815652/)]
16. Cohen I, Benesty J, Gannot S, editors. *Speech Processing in Modern Communication: Challenges and Perspectives*. Berlin, Germany: Springer-Verlag; 2010.
17. Sakhnov K, Verteletskaya E, Simak B. Dynamical energy-based speech/silence detector for speech enhancement applications. In: *Proceedings of the World Congress on Engineering (WCE 2009)*. Volume I. 2009 Jul Presented at: World Congress on Engineering (WCE 2009); July 1-3, 2009; London, UK p. 801 URL: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=6AACCA0D6A768D38D72BF9E501C027C8?doi=10.1.1.149.3572&rep=rep1&type=pdf>
18. Boersma P, Weenink D. Praat: Doing phonetics by computer. *Phonetic Sciences*, University of Amsterdam. 2002. URL: <http://webcache.googleusercontent.com/search?q=cache:EwTtYfw9KBAJ:fon.hum.uva.nl/praat/+&cd=1&hl=en&ct=clnk&gl=ca> [accessed 2020-09-02]
19. Ferrer CA, González E, Hernández-Díaz ME. Evaluation of time and frequency domain-based methods for the estimation of harmonics-to-noise-ratios in voice signals. In: *Proceedings of the Iberoamerican Congress on Pattern Recognition (CIARP 2006)*. 2006 Nov Presented at: Iberoamerican Congress on Pattern Recognition (CIARP 2006); November 14-17, 2006; Cancun, Mexico p. 406-415 URL: https://doi.org/10.1007/11892755_42 [doi: [10.1007/11892755_42](https://doi.org/10.1007/11892755_42)]
20. Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 1994 Feb;49(2):1685-1689. [doi: [10.1103/physreve.49.1685](https://doi.org/10.1103/physreve.49.1685)] [Medline: [9961383](https://pubmed.ncbi.nlm.nih.gov/9961383/)]

21. Kim H, Eykholt R, Salas JD. Nonlinear dynamics, delay times, and embedding windows. *Physica D* 1999 Mar;127(1-2):48-60 [FREE Full text] [doi: [10.1016/s0167-2789\(98\)00240-1](https://doi.org/10.1016/s0167-2789(98)00240-1)]
22. Janjarasjitt S, Scher MS, Loparo KA. Nonlinear dynamical analysis of the neonatal EEG time series: The relationship between sleep state and complexity. *Clin Neurophysiol* 2008 Aug;119(8):1812-1823. [doi: [10.1016/j.clinph.2008.03.024](https://doi.org/10.1016/j.clinph.2008.03.024)] [Medline: [18486543](https://pubmed.ncbi.nlm.nih.gov/18486543/)]
23. 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(1):23 [FREE Full text] [doi: [10.1186/1475-925x-6-23](https://doi.org/10.1186/1475-925x-6-23)]
24. 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 Jul;17(4):828-834 [FREE Full text] [doi: [10.1109/jbhi.2013.2245674](https://doi.org/10.1109/jbhi.2013.2245674)]
25. Tsanas A, Little MA, McSharry PE, Ramig LO. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans Biomed Eng* 2010 Apr;57(4):884-893 [FREE Full text] [doi: [10.1109/tbme.2009.2036000](https://doi.org/10.1109/tbme.2009.2036000)]
26. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 2018 Dec 05;58(1):267-288 [FREE Full text] [doi: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x)]
27. Eshkabilov S. MEX files, C/C++, and standalone applications. In: *Beginning MATLAB and Simulink: From Novice to Professional*. New York, NY: Apress; 2019:259-273.
28. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R. Investigating voice as a biomarker: Deep phenotyping methods for early detection of Parkinson's disease. *J Biomed Inform* 2020 Apr;104:103362. [doi: [10.1016/j.jbi.2019.103362](https://doi.org/10.1016/j.jbi.2019.103362)] [Medline: [31866434](https://pubmed.ncbi.nlm.nih.gov/31866434/)]
29. Gunduz H. Deep learning-based Parkinson's disease classification using vocal feature sets. *IEEE Access* 2019;7:115540-115551 [FREE Full text] [doi: [10.1109/ACCESS.2019.2936564](https://doi.org/10.1109/ACCESS.2019.2936564)]

Abbreviations

ANN: artificial neural network
AVEC: Audio-Visual Emotion recognition Challenge
CERNET: China Education and Research Network
COVID-19: coronavirus disease 2019
CPU: central processing unit
GeMAPS: Geneva Minimalistic Acoustic Parameter Set
HNR: harmonics-to-noise ratio
LASSO: least absolute shrinkage and selection operator
MAE: mean absolute error
MEX: MATLAB executable
MSE: mean square error
PD: Parkinson disease
RMSE: root mean square error
SVM: support vector machine
SVR: support vector regression
UCI: University of California Irvine
UPDRS: Unified Parkinson Disease Rating Scale

Edited by G Eysenbach; submitted 12.03.20; peer-reviewed by J Lei, C Yu; comments to author 20.04.20; revised version received 06.07.20; accepted 26.07.20; published 16.09.20

Please cite as:

Zhang L, Qu Y, Jin B, Jing L, Gao Z, Liang Z

An Intelligent Mobile-Enabled System for Diagnosing Parkinson Disease: Development and Validation of a Speech Impairment Detection System

JMIR Med Inform 2020;8(9):e18689

URL: <http://medinform.jmir.org/2020/9/e18689/>

doi: [10.2196/18689](https://doi.org/10.2196/18689)

PMID: [32936086](https://pubmed.ncbi.nlm.nih.gov/32936086/)

©Liang Zhang, Yue Qu, Bo Jin, Lu Jing, Zhan Gao, Zhanhua Liang. Originally published in JMIR Medical Informatics (<http://medinform.jmir.org/>), 16.09.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction

in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <http://medinform.jmir.org/>, as well as this copyright and license information must be included.