



OPEN ACCESS

EDITED BY

Corinne A. Jones,
The University of Texas at Austin,
United States

REVIEWED BY

Kristin Teplansky,
The University of Texas at Austin,
United States
Sitanshu Sekhar Sahu,
Birla Institute of Technology, Mesra,
India

*CORRESPONDENCE

Yan Fu
laura_fy@mail.hust.edu.cn

SPECIALTY SECTION

This article was submitted to
Parkinson's Disease and Aging-related
Movement Disorders,
a section of the journal
Frontiers in Aging Neuroscience

RECEIVED 04 September 2022

ACCEPTED 20 October 2022

PUBLISHED 10 November 2022

CITATION

Wang Q, Fu Y, Shao B, Chang L,
Ren K, Chen Z and Ling Y (2022) Early
detection of Parkinson's disease from
multiple signal speech: Based on
Mandarin language dataset.
Front. Aging Neurosci. 14:1036588.
doi: 10.3389/fnagi.2022.1036588

COPYRIGHT

© 2022 Wang, Fu, Shao, Chang, Ren,
Chen and Ling. This is an open-access
article distributed under the terms of
the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution
or reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Early detection of Parkinson's disease from multiple signal speech: Based on Mandarin language dataset

Qiyue Wang¹, Yan Fu^{1,2*}, Baiyu Shao¹, Le Chang²,
Kang Ren^{2,3}, Zhonglue Chen^{2,3} and Yun Ling^{2,3}

¹School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China, ²HUST-GYENNO CNS Intelligent Digital Medicine Technology Center, Wuhan, China, ³Gyenno Science Co., Ltd., Shenzhen, China

Parkinson's disease (PD) is a neurodegenerative disorder that negatively affects millions of people. Early detection is of vital importance. As recent researches showed dysarthria level provides good indicators to the computer-assisted diagnosis and remote monitoring of patients at the early stages. It is the goal of this study to develop an automatic detection method based on newest collected Chinese dataset. Unlike English, no agreement was reached on the main features indicating language disorders due to vocal organ dysfunction. Thus, one of our approaches is to classify the speech phonation and articulation with a machine learning-based feature selection model. Based on a relatively big sample, three feature selection algorithms (LASSO, mRMR, Relief-F) were tested to select the vocal features extracted from speech signals collected in a controlled setting, followed by four classifiers (Naïve Bayes, K-Nearest Neighbor, Logistic Regression and Stochastic Gradient Descent) to detect the disorder. The proposed approach shows an accuracy of 75.76%, sensitivity of 82.44%, specificity of 73.15% and precision of 76.57%, indicating the feasibility and promising future for an automatic and unobtrusive detection on Chinese PD. The comparison among the three selection algorithms reveals that LASSO selector has the best performance regardless types of vocal features. The best detection accuracy is obtained by SGD classifier, while the best resulting sensitivity is obtained by LR classifier. More interestingly, articulation features are more representative and indicative than phonation features among all the selection and classifying algorithms. The most prominent articulation features are F1, F2, DDF1, DDF2, BBE and MFCC.

KEYWORDS

Parkinson's disease, early detection, dysarthria, dysphonia features, machine learning, fully automatic detection model

Introduction

In China, the incidence rate of Parkinson's disease (PD) is 1.7% in the population over 65 years old. There are nearly 3 million patients with Parkinson's disease (PWP) in China, accounting for half of the total number of PWP in the world, and about 100,000 new patients are diagnosed every year. Currently there is no medical cure for PD, but if patients receive timely diagnosis and treatment at the early stage of the disease, early intervention can be applied to delay the disease progress and safeguard daily lives (Singh et al., 2007). Clinical studies have shown that PWP often show some characteristic speech disorders in the early stage (Ho et al., 1999). In 1970, Darley et al. (1969) first studied the pronunciation characteristics of PWP, and found that PWP usually have low volume, increased breath sound, single tone hoarseness and other problems, indicating speech is a useful signal for distinguishing PWP from healthy people (Little et al., 2009; Sapir et al., 2010). Speaking is a highly complex movement that requires the coordination of many nerves and muscles. PWPs commonly have motor deficits, mainly involving the oral, pharyngeal and jaw muscles. Laryngeal and vocal cord tremor, asymmetrical vocal cord closure time, jaw joint dyskinesia and respiratory disorders all lead to voice tremor, unclear speech, slowed speech rate and sunken intonation. At present, there are four main groups of language features used to detect PWP: phonatory, articulatory, prosodic and cognitive-linguistic (Moro-Velazquez et al., 2021).

Phonatory features model abnormal patterns in the vocal fold vibration, whose features were extracted mainly from sustained vowels. Phonation in PWP is characterized by bowing and inadequate closure of vocal folds (Hanson et al., 1984). Articulation deficits in PD patients are mainly associated with reduced amplitude and speed of lip, tongue and jaw movements (Ackermann and Ziegler, 1991), as a result of delayed movements of their tuning organs and a stiff and inflexible tongue, whose features were extracted mainly from running speech (Orozco-Arroyave et al., 2016; Kuruvilla-Dugdale et al., 2020). Prosodic features are paralinguistic, such as pitch variation or the representation of emotions among others (Harel et al., 2004). The cognitive-linguistic analysis examines the vocabulary, phrase construction and the existence of word repetitions (Illes et al., 1988). Among the four types of features, phonation and articulation features are better obtainable, with good constancy to Unified Parkinson's Disease Rating Scale (UPDRS) and thus most applied in speech analysis on PWP (Dromey et al., 1995; Tykalova et al., 2017; Moro-Velazquez et al., 2021).

PD causes abnormal vocal fold vibration, which can be reflected by the presence of noise and other perturbations caused by incomplete closure (Rusz et al., 2011), abnormal phase closure and phase asymmetry or vocal tremor (Perez et al., 1996). Dysfunction measures including noise or frequency and amplitude perturbations are applied to assess the severity of PD in telemonitoring situations (Little et al., 2009; Tsanas

et al., 2010, 2014). The problem is that recordings are done in noisy environment and different equipment draw in different noise, thus affecting dysphonia features being extracted. Studies by Novotny (Novotny et al., 2014) indicate that imprecise consonant articulation can indicate PD-related symptoms. However, it used DDK speech only, which restrict the possible articulatory combinations. Other works employ frequency features, namely Mel Frequency Cepstrum Coefficients (MFCC) and Band Bark Energies (BBE) from running speech, and other features obtained after the segmentation of specific regions, providing good results (Orozco-Arroyave et al., 2016). It is evidenced that the speech of PWP has lower values of relative fundamental frequency, which is the ratio between the fundamental frequency in the cycles of a vowel before or after a voiceless consonant and the typical fundamental frequency during the utterance (Little et al., 2009; Sapir et al., 2010). Other studies perform the tracking of vowel formants during articulation, including onset and offset (Skodda et al., 2012; Bang et al., 2013; Whitfield and Goberman, 2014) and found that as formants reflect the position of the tongue, a reduction of the articulation ranges could subsequently limit the frequency ranges of the formants.

A comparison of PD detection techniques is performed using the acoustic materials extracted from sustained vowels and running speech test, proving that two acoustic materials have better detection performance than utilizing sustained vowels only (Rusz et al., 2013; Bocklet et al. (2013) used phonatory, prosodic and articulatory features jointly, yielding results of 80% of accuracy in PD detection. In any case, all study efforts focus on identify the most representative features for PWP detection but have not reached agreement. The main features for speech sample classification vary across languages (Eyigoz et al., 2020). Different feature extraction methods and different datasets can also obstruct the unification of features (Karan et al., 2020; Zhang et al., 2021). It is one of the main goals for related studies to reduce the number of features by choosing the most relevant for PWP detection.

To date, the majority of studies examining the key characteristics of hypokinetic dysarthria and their relationship to speech intelligibility have been conducted with speakers of English. However, extending this research to languages other than English is important for both theoretical and clinical reasons. Because acoustic cues that strongly influence intelligibility in PD may vary cross-linguistically, which is vital in assessment and treatment planning (Hsu et al., 2017). At present, the speech signal diagnosis for PD patients in China is still in its infancy. Hsu et al. (2017) made a comparison between Mandarin speakers and English speakers on key features of hypokinetic dysarthria. In 2011, Zhang et al. (2011) verified the feasibility of PD Chinese speech detection. Based on their study (Zhang, 2017; Haq et al., 2019) concentrated on phonetic measurements like vocal perturbation and nonlinear measurements to classify PWP. However, they only focused on the vowel pronunciation of PD patients. Although other

studies (Su and Chuang, 2015; Fang et al., 2020; Li et al., 2020) filled in this gap through collecting speech samples from vowel pronunciation and running speech test jointly, most of them focus only on time-variant features like MFCC, etc. So this study also intends to figure out if other proposed features recognized can support accurate classification in Chinese speech. Thus it is a worthwhile approach to implement the detection by integrating the automatic feature selection method in case that the system can identify the best representative features by comparing the best detection results based on the newly collected dataset.

Meanwhile existing studies have proposed many machine learning methods to automatically detect PWP although most are based on the manually selected features. Hazan et al. (2012) chose three F1-F2-based acoustic metrics, Formant Centralization Ratio (FAR), Vowel Articulation Index (VAI) and F2i/F2u (the second formant of vowel *i* divided by the second formant of vowel *u*), using the Support Vectors Machine (SVM) with a radial basis function (RBF) kernel for the classification. The best accuracy reached 94%. Gullapalli and Mittal (2022) used various classifiers like Logistic Regression, SVM, KNN, CNN, Deep Neural Network, Boosting, Bagging, Random Forest, and illustrate a comparison on their accuracies, based on MFCC, JTFA, MDVP and TQTW as main features. To date, feature selection has been successfully used in medical applications, where it cannot only reduce dimensionality and but also help us understand the causes of a disease better (Remeseiro and Bolon-Canedo, 2019). Some studies also applied machine learning methods to feature selection. Lamba et al. (2022) tested several combinations of three feature selection approaches (mutual information gain, extra tree, and genetic algorithm) and three classification algorithms (Naive Bayes, KNN, and Random Forest). The combination of genetic algorithm and Random Forest classifier has shown the best performance with 95.58% accuracy. Solana-Lavalle et al. (2020) used Wrappers feature subset selection with four classifiers (KNN, Multi-layer perceptron, SVM, and Random Forest), obtaining the highest accuracy of 94.7% with SVM. The fully automatic model mentioned above performs well in corpora such as English and German, but it is still unknown whether this kind of model can be well applied to the detection of Chinese PD.

In this study, data are collected from two types of speech tasks (namely sustain vowel sound and running speech test), and a completely automatic model is proposed. Multiple speech signals are extracted and are fed into the hybrid combinations of three feature selectors and four classifiers, detecting PWP automatically. The final detection results are used to compare the performance of both selection algorithms and classifying algorithms. The manuscript is organized as follows. Section “Materials and methods” elaborates the dataset and discusses the automatic PD detection model, with classifier validation methods and evaluation metrics. In section “Results,” experimental results are given in details. Section “Discussion”

makes a discussion about the results. Some concluding remarks are given in the section “Conclusion.”

Materials and methods

Automatic detection model

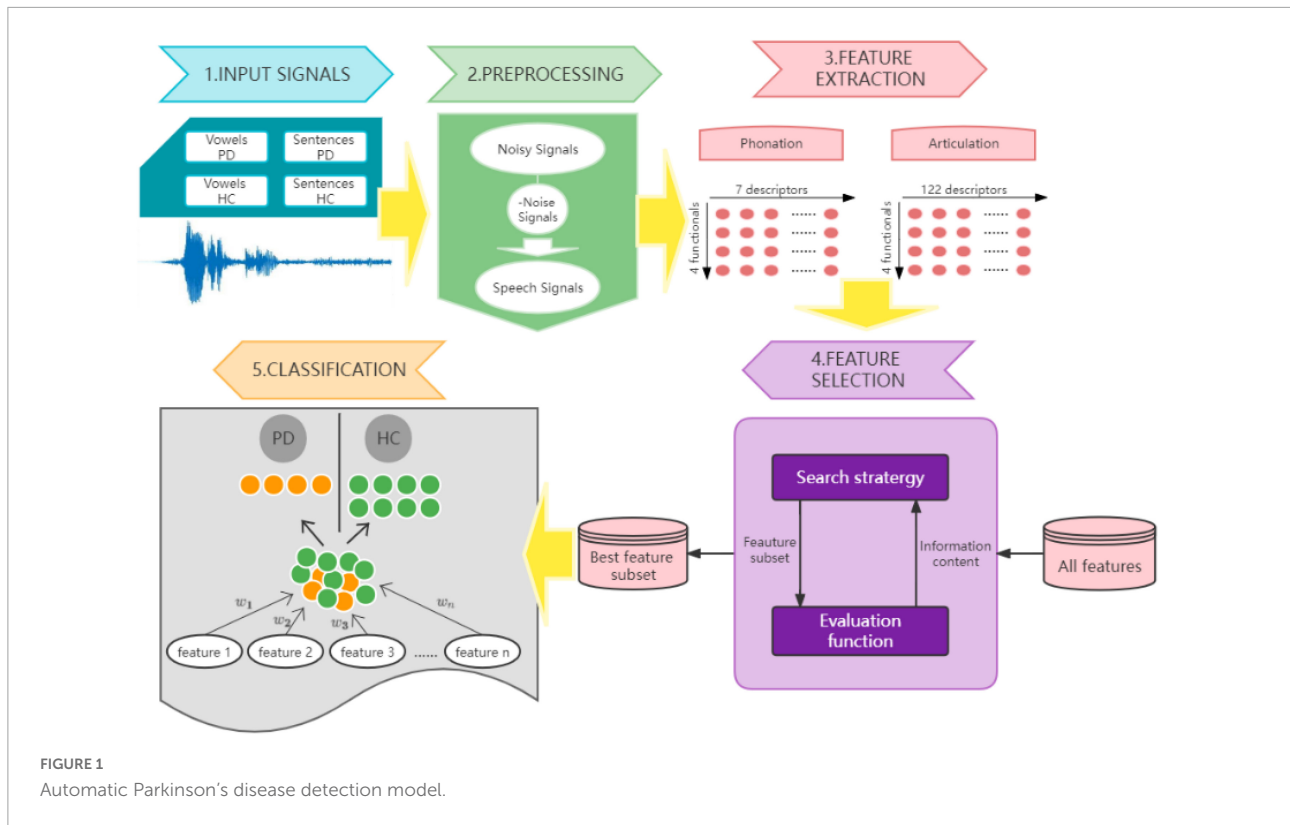
In this study, an automatic detection model is proposed, including feature extraction, feature selection and tester classification, as shown in Figure 1.

When the data are collected, noise reduction is firstly conducted. The noise-reduced speech signal is applied to feature extraction processing, and two types of static vector features are extracted, respectively. The feature extraction is run on the open-source algorithm on GitHub.¹ For feature selection module, the study chose three algorithms to filter the extracted features to compare the best selection algorithm that can integrated in the proposed model. For feature classifying, four algorithms are tested to determine the best to be embedded in the model.

Database

Clinical practice has shown that sustained vowels (Tsanas et al., 2014) and running speech (Ackermann and Ziegler, 1991) are good materials for detection. This study uses two corpora for testing analysis: sustained vowels and tongue twisters. Tongue twister is regarded as challenging to pronounce because of meeting problems of using correctly the mouth and tongue. It could be assumed that dysarthria would manifest especially during trying to pronounce tongue twister by PD due to the deterioration of articulators (Vilda et al., 2017). Data were collected at National Research Center of Geriatric Diseases, Tongji Hospital under the supervision of SLP professionals. 100 Mandarin-speaking people were recorded. 50 are diagnosed as PD (25 female, 25 male) and 50 healthy people (25 female, 25 male). The 50 PD patients have an average age of 63.57 ± 11.31 years (mean \pm SD) and a mean disease duration of 6.08 ± 3.17 years. According to Hoehn and Yahr (HY) staging scale, all patients were in stage 1–3 (1–1.5 as early stage and 2–3 as middle stage). None of them has a history of language or speech disorders. Their mean motor score according to part III of MDS_UPDRS was 36.43 ± 17.39 . Each person (50 PWP, 50 HC) was recorded three times for each of four speech samples, and a total of 1,200 speech signals were collected for the dataset. The first two recording tasks are sustained vowels (“aaa. . .” and “eee. . .” in Chinese Pinyin), from which the 6-s stable segments are extracted; the other two are short sentences (“si shi si zhi shi shi zi” and “yi zhi da hua wan kou zhe yi zhi da hua ha

¹ <https://github.com/jcvasquezc/DisVoice>



ma”). The voice recordings have been obtained in a realistic environment using a Rode NT-USB microphone 10 cm away from the mouth. And the sampling rate of data was 96 kHz. Data has been stored in a WAVE (.wav) file format. Sustained vowels were used in the phonatory analysis, while running speech test (sentence 1, sentence 2) was added for articulation analysis. Spectral subtraction (SS) (Boll, 1979) is used to clean up the noisy speech signal.

Feature selection algorithm

Three algorithms, Least Absolute Shrinkage and Selection Operator (LASSO), minimum-Redundancy-Maximum-Relevance (mRMR), and Relief-F are applied for automatic feature selection. Each of them is applicable to the selection of high-dimensional data sets and has a wide range of applications in many fields. The number of features is set on the random rule. The most representative features can be determined upon comparing the accuracy of the final classification results.

Least Absolute Shrinkage and Selection Operator

The LASSO (Fonti and Belitser, 2017), mainly for feature selection of high-dimensional data allows the coefficients of features to be compressed even to zero. Capable of making up for the deficiencies of least squares and stepwise regression for local optimal estimation, it can effectively solve the problem of

multicollinearity existing among the features. The LASSO is a particular case of the penalized least squares regression with L_1 -penalty function.

The LASSO estimate can be defined by:

$$\hat{\beta}^{lasso} = arg \min \left\{ \frac{1}{2} \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \tag{1}$$

LASSO transforms each and every coefficient by a constant component λ , truncating at zero. Hence it is a forward-looking variable selection method for regression. It decreases the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. LASSO improves both prediction accuracy and model interpretability by combining the good qualities of ridge regression and subset selection. If there is high correlation in the group of predictors, LASSO chooses only one among them and shrinks the others to zero. It reduces the variability of the estimates by shrinking the some of the coefficients exactly to zero producing easily interpretable models (Muthukrishnan and Rohini, 2016).

Minimum-Redundancy-Maximum-Relevance

Minimum-Redundancy-Maximum-Relevance (mRMR) algorithm (Solana-Lavalle et al., 2020) is a typical feature selection algorithm based on spatial search. It extracts features

of maximum relevance to the target variable while ensuring minimum redundancy between each other. In this algorithm, both redundancy and correlation are used as the metric of mutual information. The steps involved are:

1. Calculate the mutual information of each special x_i with category C :

$$I(x_i; C) \tag{2}$$

2. The average of the mutual information between all features and the category C is calculated to obtain an approximation of D . A subset S of features containing m features is drawn so that the value of D calculated using the features within S is maximized:

$$\max_D(S; C) D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; C) \tag{3}$$

3. Eliminate the redundancy between the selected m features:

$$\min R(S) R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \tag{4}$$

4. Calculate set S of features with maximum-relevance-minimum-redundancy:

$$\text{mRMR} = \max \left[\frac{1}{|S|} \sum_{x_i} I(x_i; C) - \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \right] \tag{5}$$

Relief-F

Relief-F (Park and Kwon, 2007), as the more effective filter-style feature evaluation algorithm is proposed for regression problems where the target attributes are continuous values. Relief algorithm assigns each feature weights, subsequently updated. Features with higher correlation with labels are given higher weights, and vice versa. The steps involved are:

1. Let the training data set be D , the number of samples sampled be m , the feature weight threshold be δ , and the number of nearest samples be k , the feature weights of each characteristic of the output be T .

2. Set all feature weights to zero and make T the empty set.

3. For $i = 1, 2, \dots, m$: (a) Select a random sample R from D ; (b) finds k nearest-neighbor samples $H_j (j = 1, 2, \dots, k)$ of R from the sample set of the same category, and k nearest-neighbor samples $M_j(C)$ of R from the sample set of different categories.

4. For $A = 1$ to N . All features do:

$$W(A) = W(A) - \sum_{j=1}^k \frac{\text{diff}(A, R, H_j)}{mk} + \sum_{\text{Class}(R)} \frac{\frac{p(C)}{1-p(\text{Class}(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C))}{mk} \tag{6}$$

Classifier

Four classifying algorithms, Naive Bayes, K-Nearest Neighbor (KNN), Logistic Regression and Stochastic Gradient Descent are trained by the dataset to explore the best for the whole model.

Naïve Bayes

The input space vector $\mathcal{X} \subseteq R^n$ is the set of n -dimensional vectors and the output space vector is the set of class labels $\mathcal{Y} = \{c_1, c_2, \dots, c_k\}$. The input is the feature vector $x \in \mathcal{X}$ and the output is the class label $y \in \mathcal{Y}$. X is a random vector defined on the input space \mathcal{X} , and Y is a random variable defined on the output space \mathcal{Y} . $P(X, Y)$ is the joint probability distribution of X and Y .

In the Naïve Bayes, for a given input x , the posterior probability distribution $P(Y = c_k | X = x)$ is calculated by the learned model, and the class with the highest posterior probability is used as the class output of x . The posterior probability calculation is performed according to Bayes' theorem. Finally, by the substitution calculation of the formulas, the Naive Bayesian classifier can be expressed as:

$$y = \arg \max P(Y = c_k) \prod_j P(X^{(j)} = x^{(j)} | Y = c_k) \tag{7}$$

K-Nearest Neighbor

K-Nearest Neighbor (KNN) assumes a given training dataset in which the strength classes have been determined. New instances are predicted based on the categories of their k -nearest neighboring training instances, e.g., by majority voting. KNN does not have an explicit learning process, but uses the training dataset to partition the feature vector space and serve as a "model" for its classification. The core idea of the algorithm is that a sample belongs to a class if most of its k -nearest samples belong to that class. And the measurement of distance generally adopts the Euclidean distance:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \tag{8}$$

Logistic Regression

The LR algorithm is a typical and mature classification algorithm, which performs well especially in binary classification problems. Since speech data have many features, each of which has certain level of influence on the final classification result and needs to be linearly weighted. The output of LR is not the exact category, but a probability, and if the result is closer to 0 or 1, the higher the confidence of the classification result is higher. Weighting of each feature can be adjusted by the classification

result during the training process, making the classification result more accurate.

Regression routine steps are as follows:

(1) Find the prediction function.

$$h_{\theta}(x) = g\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}} \quad (9)$$

The value of $h_{\theta}(x)$ indicates the probability that the result will take 1. For input x , the probability that the classification results in category 1 and category 0, respectively, are:

$$P(y = 1 | x; \theta) = h_{\theta}(x) \quad (10)$$

$$P(y = 0 | x; \theta) = 1-h_{\theta}(x) \quad (11)$$

(2) Find the loss function.

The Cost-function and J-function are as follows, and they are derived based on the maximum likelihood estimation.

$$Cost\left(h_{\theta}(x), y\right) = \begin{cases} -\log\left(h_{\theta}(x)\right) & y = 1 \\ -\log\left(1-h_{\theta}(x)\right) & y = 0 \end{cases} \quad (12)$$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m Cost\left(h_{\theta}\left(x_i\right), y_i\right) = -\frac{1}{m} \left[\sum_{i=1}^m \left(y_i \log h_{\theta}\left(x_i\right) + \left(1-y_i\right) \log \left(1-h_{\theta}\left(x_i\right)\right) \right) \right] \quad (13)$$

(3) Minimize the loss function and find the regression parameter θ .

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m \left(h_{\theta}\left(x_i\right) - y_i \right) x_i^j \quad (14)$$

Stochastic Gradient Decent

An arbitrary hyperplane w_0, b_0 chosen and then the objective function is continuously minimized using Stochastic Gradient Descent. Assuming that the set of misclassified points M is fixed, the gradient of the loss function $L(w, b)$ is as follows:

$$\nabla_w L(w, b) = - \sum_{x_i \in M} y_i x_i \quad (15)$$

$$\nabla_b L(w, b) = - \sum_{x_i \in M} y_i \quad (16)$$

Select a random misclassification point (x_i, y_i) and update w, b :

$$w \leftarrow w + \eta x_i y_i \quad (17)$$

$$b \leftarrow b + \eta y_i \quad (18)$$

where $\eta(0 < \eta \leq 1)$ denotes the step size, also known as the learning rate in statistics. The loss function $L(w, b)$ can be reduced by iterations until it is 0, which means that the point is correctly classified.

Performance metrics

There are four results for the detection: TRUE POSITIVE (TP) if a PD patient is correctly identified and otherwise FALSE NEGATIVE (FN), TRUE NEGATIVES (TN) if healthy subjects correctly diagnosed and otherwise FALSE POSITIVES (FP).

Accuracy, sensitivity, specificity, precision, false alarm rate, Matthew correlation coefficient, F1 score and the receiver operating curve (ROC) are used to make statistical analysis on the results.

Accuracy represents the percentage that the classification is correct.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (19)$$

Sensitivity or recall is the probability that the outcome of diagnosing PD is positive given that the subjects have PD.

$$Sensitivity = \frac{TP}{TP + FN} \quad (20)$$

Specificity represents the proportion that the outcome of PD is negative given that the subject is healthy.

$$Specificity = \frac{TN}{TN + FP} \quad (21)$$

Precision is the probability that the outcome of diagnosing PD is true.

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

The false alarm rate (FAR) is the probability that the outcome of diagnosing PD is false.

$$FAR = \frac{FP}{TP + FP} \quad (23)$$

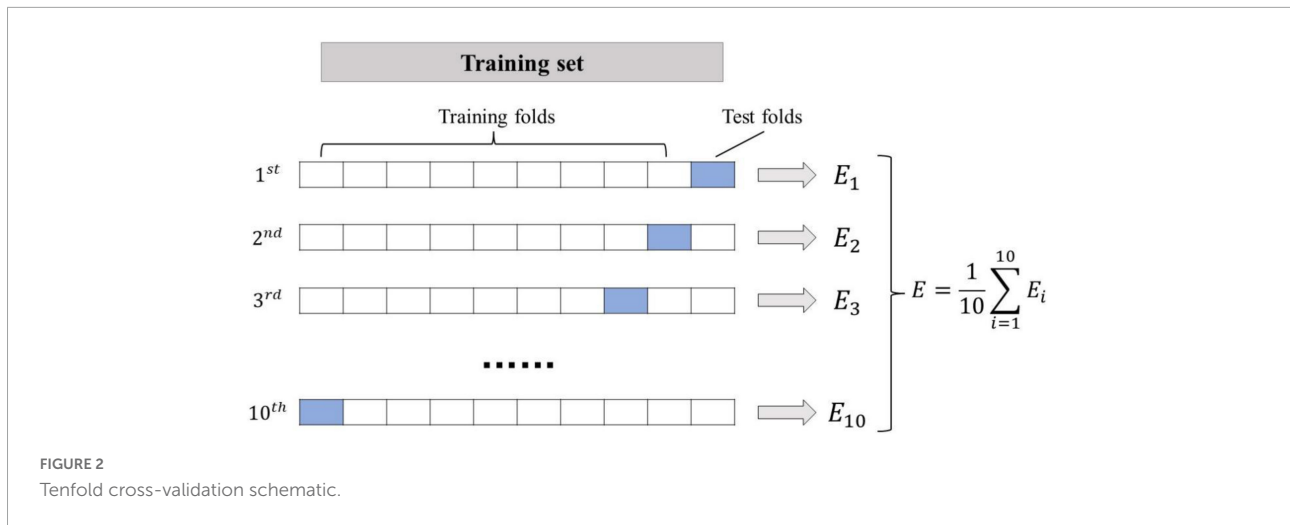
The Matthews correlation coefficient (MCC) is a correlation coefficient between the observed and predicted binary classifications. It returns a value between -1 and +1. When $MCC = 1$, it means that machine learning system perfectly predict the category of the object; When the value is 0, it indicates that the predicted result is worse than the random prediction result; When $MCC = -1$, it illustrates that the predicted classification is completely inconsistent with the actual classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (24)$$

The F_1 score is the harmonic mean of precision and recall.

$$F_1 = 2 \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \quad (25)$$

The receiver operating characteristic (ROC) curve is the plot of the Sensitivity against the false positive rate ($FPR = 1 - Specificity$) in a binary classifier when its threshold is varied.



Training and test set

Feature vectors, from PD or HC, are stored into two sets: C_1 for PD patients, C_2 for HC. Each set is separated into ten fragments, $C_1 = \{C_{1,1}, C_{1,2}, \dots, C_{1,10}\}$ and $C_2 = \{C_{2,1}, C_{2,2}, \dots, C_{2,10}\}$. A fragment $C_{1,i}$ (from C_1) and a corresponding fragment $C_{2,i}$ (from C_2) are randomly combined into C_i . The result of these random mixings is tenfold $\{C_1, C_2, \dots, C_{10}\}$, where each fold contains instances from PD and HC. Among the ten folds, one is picked for testing of a classifier and the other nine are left for training. (Solana-Lavalle et al., 2020). To be specific, in this study, the speech data with 5 PD and 5 HC are used as the test set, and the speech data with 45 PD and 45 HC are used as the training set. The tenfold cross-validation schematic is shown in Figure 2.

Results

Table 1 presents the description of the two main types of features extracted from the speech signal and the corresponding number of features within each group.

TABLE 1 Features extracted from speech signals.

Feature set	Description	Total features
Phonation	Jitter, Shimmer, APQ, PPQ, DF0, DDF0, LogE.	28
Articulation	BBE_on (22 levels), BBE_off (22 levels); MFCC_on (12 levels), MFCC_off (12 levels); DMFCC_on (12 levels), DMFCC_off (12 levels); DDMFCC_on (12 levels), DDMFCC_off (12 levels); F1 (1 level), DF1 (1 level), DDF1 (1 level); F2 (1 level), DF2 (1 level), DDF2 (1 level).	488

Twenty-eight phonation features were extracted. There are seven descriptors, each of which has four values: mean, standard deviation, skewness, and kurtosis. The seven descriptors are Jitter, Shimmer, Pitch Perturbation Quotient (PPQ), Amplitude Perturbation Quotient (APQ), First Derivative of the Fundamental Frequency (DF0), Second Derivative of the Fundamental Frequency (DDF0), and Logarithmic Energy (LogE).

TABLE 2 Features obtained by using three features selection algorithms (std stands for standard deviation).

Feature subset	Features		
	LASSO	mRMR	Relief-F
Phonation features	Kurtosis_Jitter, kurtosis_Shimmer, kurtosis_DF0, kurtosis_PPQ, kurtosis_APQ. (5)	Kurtosis_Jitter, mean_Jitter, mean_Shimmer, std_PPQ, skewness_Jitter, std_APQ, skewness_Shimmer. (7)	Std_LogE, skewness_LogE, mean_DF0, kurtosis_Shimmer, mean_LogE, mean_DDF0, skewness_PPQ, skewness_DDF0, kurtosis_APQ, skewness_APQ, skewness_DF0, kurtosis_PPQ. (12)
Articulation features	Mean_MFCCon_3, mean_MFCCon_4, mean_MFCCoff_3, mean_F1, mean_F2, std_F1, std_DDF1, std_F2, std_DDF2, mean_MFCCon_6. (10)	Mean_F2, mean_BBEon_2, mean_BBEoff_1, mean_DDMFCCoff_1, std_F1, std_DDF1, std_DDF2, mean_F1, mean_BBEon_3, std_F2. (10)	Mean_MFCCon_3, mean_MFCCoff_3, mean_BBEoff_6, mean_MFCCon_4, mean_MFCCoff_4, mean_BBEoff_7, mean_F1, mean_MFCCoff_7, mean_F2, mean_MFCCon_7. (10)

Four hundred and eighty-eight articulation features were extracted. There are 122 descriptors, each of which has 4 values: mean, standard deviation, skewness and kurtosis. The 122 descriptors are segmented into 14 categories: Bark Band Energies in onset transitions (BBE_on), Bark Band Energies in offset transitions (BBE_off), Mel Frequency Cepstral Coefficients in onset transitions (MFCC_on), Mel Frequency Cepstral Coefficients in offset transitions (MFCC_off), First derivative of the MFCCs in onset transitions (DMFCC_on), First derivative of the MFCCs in offset transitions (DMFCC_off), Second derivative of the MFCCs in onset transitions (DDMFCC_on), Second derivative of the MFCCs in offset transitions (DDMFCC_off), First Formant Frequency (F1), First Derivative of F1 (DF1), Second Derivative of F1 (DDF1), Second Formant Frequency (F2), First Derivative of F2 (DF2) and Second Derivative of F2 (DDF2). Each category has a different number of levels, as shown in **Table 1**.

Table 2 presents a description of the selected features. For phonation, each algorithm screens 5, 7, and 12 features for classification each time; for articulation, each algorithm screens 10, 20, 30, and 40 features for classification each time. The best performing feature set and its size corresponding to each algorithm are listed in **Table 2**.

Tables 3–5 show the PD detection performance of four different classifiers (Naïve Bayes, KNN, Logistic Regression,

Stochastic Gradient Descent) when they are tested with a phonation feature set selected by three selection algorithms (LASSO, mRMR, Relief-F). The best results are highlighted in boldface. The best performance metric values for phonation-based PD detection are Accuracy = 0.4941, Sensitivity = 0.7058, Specificity = 0.8070, precision = 0.5100, FAR = 0.4900, MCC = 0.0413, F_1 score = 0.5449. Most these best results appear when the LASSO algorithm is used to select the features. The best-performing classifier is different across different features.

Tables 6–8 show the PD detection performance of the four classifiers, when they are tested with an articulation feature set automatically selected. The best results are also highlighted in boldface. The best performance metric values for articulation-based PD detection are Accuracy = 0.7576, Sensitivity = 0.8244, Specificity = 0.7315, precision = 0.7657, FAR = 0.2343, MCC = 0.5100, F_1 score = 0.7901. All these best results also appear when the LASSO algorithm is used to select the features.

Discussion

For all three selection algorithms, the most prominent features are F1, F2, DDF, BBE and MFCC, which are all from articulation features. F1, F2, DDF1 and DDF2 can represent

TABLE 3 PD-detection performance metrics for four different classifiers by using phonation features selected by Least Absolute Shrinkage and Selection Operator selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
KNN	0.4898	0.7058	0.7058	0.4344	0.5656	0.0413	0.5378
LR	0.3726	0.5167	0.3777	0.3680	0.6320	-0.0837	0.4298
SGD	0.3391	0.4900	0.3400	0.3382	0.6618	-0.1810	0.4002
NB	0.4941	0.1483	0.8070	0.1531	0.8469	-0.0330	0.1507

The best results are highlighted in bold.

TABLE 4 Parkinson's disease-detection performance metrics for four different classifiers by using phonation features selected by minimum-Redundancy-Maximum-Relevance selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
KNN	0.4475	0.6501	0.6501	0.3832	0.6168	-0.0133	0.4822
LR	0.3184	0.3989	0.3400	0.2906	0.7094	-0.2865	0.3362
SGD	0.3558	0.6544	0.1547	0.1350	0.8650	-0.2110	0.2238
NB	0.3141	0.3628	0.4000	0.2281	0.7719	-0.2290	0.2801

The best results are highlighted in bold.

TABLE 5 Parkinson's disease-detection performance metrics for four different classifiers by using phonation features selected by Relief-F selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F_1 score
KNN	0.4354	0.6497	0.6497	0.3697	0.6303	-0.0486	0.4712
LR	0.4351	0.5711	0.4201	0.4471	0.5529	-0.0108	0.5015
SGD	0.4143	0.4067	0.5374	0.3105	0.6895	-0.0337	0.3522
NB	0.4399	0.4128	0.5222	0.3593	0.6407	-0.0708	0.3842

The best results are highlighted in bold.

TABLE 6 PD-detection performance metrics for four different classifiers by using articulation features selected by Least Absolute Shrinkage and Selection Operator selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F ₁ score
KNN	0.5017	0.5851	0.5851	0.4556	0.5444	0.1110	0.5123
LR	0.7453	0.8244	0.6885	0.7586	0.2414	0.5100	0.7901
SGD	0.7576	0.7717	0.7315	0.7657	0.2343	0.5041	0.7687
NB	0.3981	0.5267	0.4080	0.3896	0.6140	-0.0418	0.4479

The best results are highlighted in bold.

TABLE 7 PD-detection performance metrics for four different classifiers by using articulation features selected by minimum-Redundancy-Maximum-Relevance selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F ₁ score
KNN	0.4605	0.6648	0.6648	0.3987	0.6013	-0.0395	0.4985
LR	0.5063	0.6156	0.4584	0.5416	0.4584	0.0627	0.5763
SGD	0.5734	0.5706	0.5988	0.5557	0.4443	0.1688	0.5630
NB	0.3266	0.4644	0.3333	0.3192	0.6808	-0.2448	0.3783

The best results are highlighted in bold.

TABLE 8 PD-detection performance metrics for four different classifiers by using articulation features selected by Relief-F selection algorithm.

	Accuracy	Sensitivity	Specificity	Precision	FAR	MCC	F ₁ score
KNN	0.5482	0.6383	0.6383	0.5076	0.4924	0.1295	0.5655
LR	0.6994	0.7839	0.6450	0.7160	0.2840	0.4277	0.7484
SGD	0.6743	0.7217	0.6326	0.6919	0.3081	0.3590	0.7065
NB	0.5534	0.5361	0.6276	0.5020	0.4980	0.1457	0.5185

The best results are highlighted in bold.

resonances in the vocal tract (Pah et al., 2022) and the capability of the speaker to hold the tongue in a certain position (Ladefoged and Harshman, 1979). BBE and MFCC are common dynamic signals. It was found that oral rotation can be represented by the dynamic characteristics of speech signals (like BBE and MFCC). Although the oral rotation rate of PWP did not decrease significantly, there was a balance among speed, intensity and accuracy. Besides, MFCCs were also computed as a smooth representation of the voice spectrum that considers the human auditory perception. The features mentioned above may mainly reflect the pitch, speed and intelligibility of the tester's speech (Moro-Velazquez et al., 2021), echoing UPDRS, which can reflect the five levels of speech status from 0 to 4 in the scale (Zhang et al., 2017).

It is quite noticeable that articulation-type features are generally more representative than phonation analysis in this study. The reason may be that, the signals employed in phonatory approaches (sustained vowels) are much simpler than those used for articulatory analyses (running speech), including less variability and a smaller amount of kinetic information. Moreover, running speech contains vowels and sonorant segments and therefore methodologies using connected speech can indirectly characterize certain phonatory aspects (Moro-Velazquez et al., 2021). In addition, articulation relates to more

voice organs than phonation features (Hanson et al., 1984; Ackermann and Ziegler, 1991). Phonation features like Jitter and Shimmer are used as significant influential factors in classifiers, with good performance of the results (Orozco-Arroyave et al., 2014; Mekyska et al., 2015; Moro-Velazquez et al., 2021). But in the present study, all phonation feature subsets provide relatively low classification accuracy. Further studies based on more Chinese dataset are expected to explore the reasons.

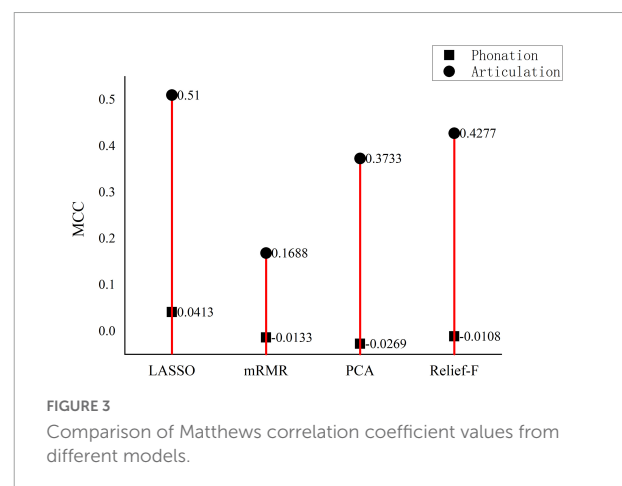
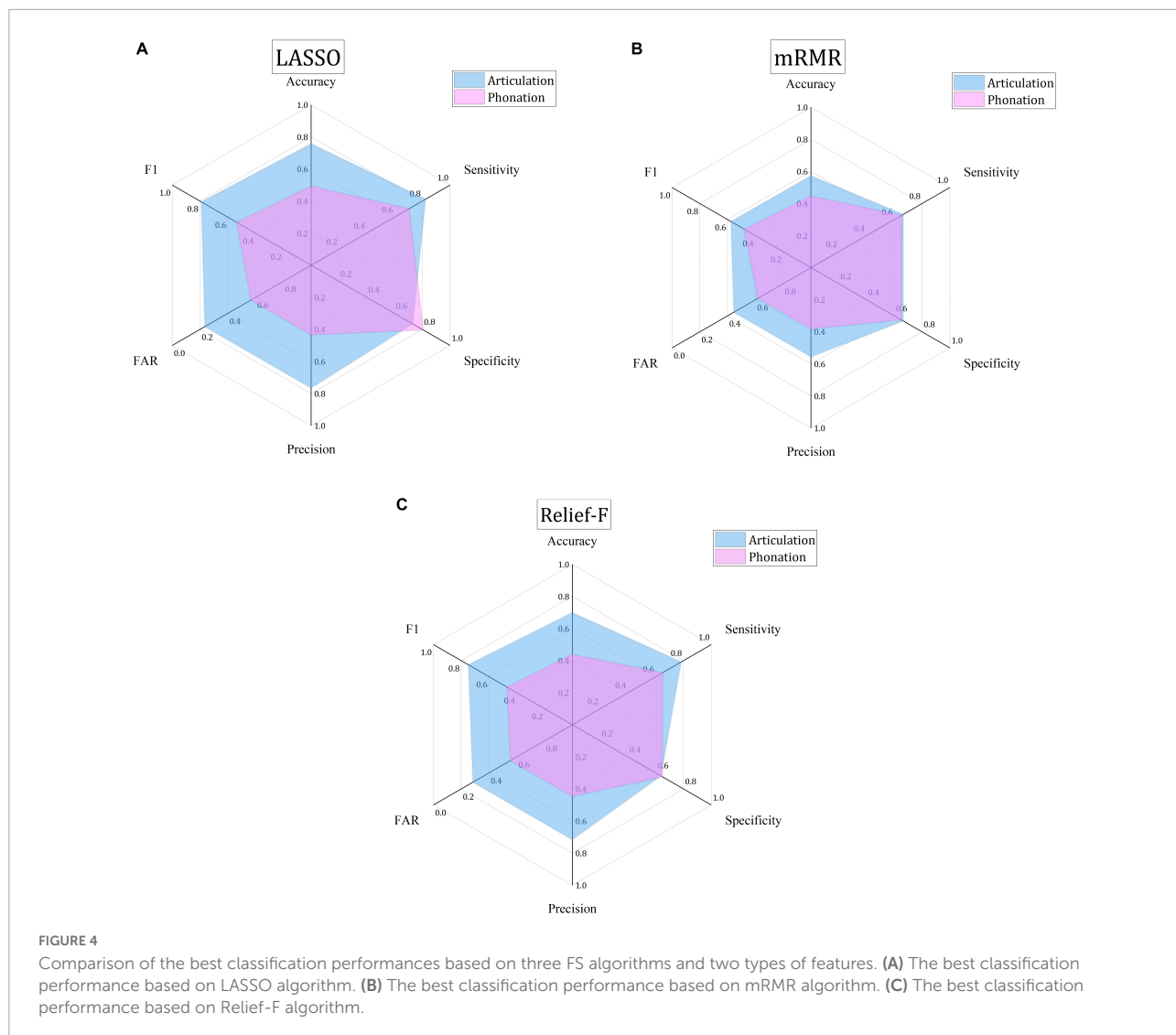


FIGURE 3 Comparison of Matthews correlation coefficient values from different models.



As **Figure 3** shows, the MCC values for the detection with the phonation features are almost all negative small numbers while the MCC for the articulation features are mostly positive big numbers. Therefore, it is recommended that articulation features can be the primary detection feature set applied. Further studies can be made to identify the most representative articulation features when more Chinese language materials are used to test the proposed detection model. Six positive performance values are further compared as shown in **Figure 4**. The articulation (indicated by blue) and phonation (indicated by pink) are put together to show that the articulation feature is better than the phonation feature, indicating articulation can better reflect the phonetic features of Chinese PWP. This finding is quite consistent with the conclusion reached by **Vásquez-Correa et al. (2018)**. It can be inferred that the proposed model can automatically generate good indicators for the following

automatic speech character classification, replacing the manual feature selection.

Meanwhile, it was found that the MCC values of the final detection results were not satisfactory enough. There are several possible reasons: Firstly, the speech data used in this study were from patients at HY1-3 stages, with relatively low degree of dysarthria, and pronunciation defects were not quite obvious. Secondly, the classification features are automatically selected by the model, and the number of input features is large, which may result in overfitting and cause some errors. Thirdly, different from English language, each hieroglyph in Chinese has an individual meaning and corresponds to one syllable, which means each syllable conveys an idea, and the combination of syllables will be different according to different contextual meanings. Chinese speakers normally require more time to think before speaking, causing some pauses, not due to PD (**Pavlovskaya and Hao, 2020**). Chinese speakers breathe

less regularly than English speakers when speaking, which in turn may lead to a misjudgment that Chinese test subjects have unstable vocalizations.

Among the three feature selection algorithms, LASSO performs the best. The main reason may lie in that the parameter estimation of LASSO algorithm with good continuity is suitable for the selection model of high-dimensional data, which is the main characters of the collected signal. Among the four classification algorithms, Logistic Regression performs the best. To observe the final performance results, LASSO & LR and LASSO & SGD are the best combinations of feature selection and classifier with the accuracies of 0.7453 and 0.7576, respectively. Interestingly, for English corpus, the commonly used Wrappers feature subset selection and the classifying techniques like KNN, SVM, MLP and Random Forest do not perform well for Chinese corpus in this study. More speech materials should be collected to train the detection model, and comparing the results with the performance of those algorithms in case.

Overall, the results prove the feasibility of applying a fully automatic model to Chinese PD detection is feasible, even though the results are not satisfactory when compared with the detection model based on the English corpus (Solana-Lavalle et al., 2020). But the combined performance of LASSO with LR and SGD are both above 0.7, quite convincing to motivate further development on the proposed detection including automatic feature selection and classification when there are no universally accepted representative features for PWP early detection.

Conclusion

The novel contribution of this study is establishing an automatic model with machine learning methods based on Mandarin language dataset, dealing the whole process of PD detection based on speech signals from extraction, selection to classifying automatically. It is possible that the gap-filling in setting up representative feature reservoirs for Chinese language can be accelerated through automatic feature selection model.

The current study only proved its feasibility and future work should be focused on developing robust and accurate methods for the automatic and unobtrusive detection for Chinese PWP, and a dedicated algorithm for feature extraction specific to Chinese speech features.

This study also gives good hints for feature selection and classifier strategy. The most representative feature set of PWP is articulation, from which ten features automatically selected is enough for the following classifier. To improve the accuracy of detection, bigger dataset should be collected to test whether the articulation features are the best representative for Chinese-speaking PD. LASSO performs the best feature selection and LR performs the best classification, while two combinations of

LASSO & LR and LASSO & SGD all performs well. So, in this study, the best model proposed is to filter 10 articulation features with LASSO algorithm and use them in SGD or LR classifier. Further studies can be made to explore the rules of selecting among LASSO & LR or LASSO & SGD.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Data collection and sharing for this project was approved by the Medical Ethics Committee of Tongji Hospital (A053/IEC/2021). Prior to data acquisition, all patients involved gave written informed consent to the study procedures, and to pseudonymized storage of voice recordings and further speech analyses.

Author contributions

QW: methodology, software, formal analysis, investigation, writing – original draft and review and editing, and visualization. YF: conceptualization, validation, formal analysis, writing – original draft, supervision, and project administration. BS: data curation, resources, and investigation. LC: software, formal analysis, and writing – review and editing. KR: study design and writing – review the draft. ZC: study design and data collection. YL: study design. All authors contributed to the article and approved the submitted version.

Funding

This work was partly funded by the National Natural Science Foundation of China (71771098) and partly by Exploration Project of HUST-GYENNO CNS Intelligent Digital Medicine Technology Center.

Acknowledgments

This work was partly supported by the Speech-Language Pathologists (SLP) from Tongji Hospital affiliated to Tongji Medical College of Huazhong University of Science and Technology.

Conflict of interest

Authors KR, ZC, and YL were employed by Gyenno Science Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ackermann, H., and Ziegler, W. (1991). Articulatory deficits in parkinsonian dysarthria: An acoustic analysis. *J. Neurol. Neurosurg. Psychiatry* 54, 1093–1098. doi: 10.1136/jnnp.54.12.1093
- Bang, Y. I., Min, K., Sohn, Y. H., and Cho, S. R. (2013). Acoustic characteristics of vowel sounds in patients with Parkinson disease. *NeuroRehabilitation* 32, 649–654. doi: 10.3233/nre-130887
- Bocklet, T., Steidl, S., Nöth, E., and Skodda, S. (2013). “Automatic evaluation of parkinson’s speech-acoustic, prosodic and voice related cues,” in *Proceedings of the 14th annual conference of the international speech communication association*, (Lyon: ISCA), 1149–1153.
- Boll, S. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* 27, 113–120. doi: 10.1109/tassp.1979.1163209
- Darley, F. L., Aronson, A. E., and Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *J. Speech Hear. Res.* 12, 246–269. doi: 10.1044/jshr.1202.246
- Dromey, C., Ramig, L. O., and Johnson, A. B. (1995). Phonatory and articulatory changes associated with increased vocal intensity in parkinson disease: A case study. *J. Speech Lang. Hear. Res.* 38, 751–764. doi: 10.1044/jshr.3804.751
- Eyigoz, E., Courson, M., Sedeño, L., Rogg, K., Orozco-Arroyave, J. R., Nöth, E., et al. (2020). From discourse to pathology: Automatic identification of Parkinson’s disease patients via morphological measures across three languages. *Cortex* 132, 191–205. doi: 10.1016/j.cortex.2020.08.020
- Fang, H., Gong, C., Zhang, C., Sui, Y., and Li, L. (2020). “Parkinsonian Chinese speech analysis towards automatic classification of Parkinson’s disease,” in *Proceedings of the machine learning for health NeurIPS workshop*, (New York, NY: PMLR), 114–125.
- Fonti, V., and Belitser, E. (2017). Feature selection using lasso. *VU Amst. Res. Pap. Bus. Anal.* 30, 1–25.
- Gullapalli, A. S., and Mittal, V. K. (2022). “Early detection of Parkinson’s disease through speech features and machine learning: A review,” in *ICT with intelligent applications*, eds T. Senjyu, P. N. Mahalle, T. Perumal, and A. Joshi (Singapore: Springer), 203–212. doi: 10.1007/978-981-16-4177-0_22
- Hanson, D. G., Gerratt, B. R., and Ward, P. H. (1984). Cinagraphic observations of laryngeal function in parkinson’s disease. *Laryngoscope* 94, 348–953. doi: 10.1288/00005537-198403000-00011
- Haq, A. U., Li, J. P., Memon, M. H., Khan, J., Malik, A., Ahmad, T., et al. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson’s disease using voice recordings. *IEEE Access* 7, 37718–37734. doi: 10.1109/ACCESS.2019.2906350
- Harel, B., Cannizzaro, M., and Snyder, P. J. (2004). Variability in fundamental frequency during speech in prodromal and incipient Parkinson’s disease: A longitudinal case study. *Brain Cogn.* 56, 24–29. doi: 10.1016/j.bandc.2004.05.002
- Hazan, H., Hilu, D., Manevitz, L., Ramig, L. O., and Sapir, S. (2012). “Early diagnosis of Parkinson’s disease via machine learning on speech data,” in *Proceedings of the 2012 IEEE 27th convention of electrical and electronics engineers in Israel*, (Piscataway, NJ: IEEE), 1–4.
- Ho, A. K., Inseker, R., Marigliani, C., Bradshaw, J. L., and Gates, S. (1999). Speech impairment in a large sample of patients with Parkinson’s disease. *Behav. Neurol.* 11, 131–137. doi: 10.1155/1999/327643
- Hsu, S. C., Jiao, Y., McAuliffe, M. J., Berisha, V., Wu, R. M., and Levy, E. S. (2017). Acoustic and perceptual speech characteristics of native mandarin speakers with Parkinson’s disease. *J. Acoust. Soc. Am.* 141:EL293. doi: 10.1121/1.4978342
- Illes, J., Metter, E. J., Hanson, W. R., and Iritani, S. (1988). Language production in Parkinson’s disease: Acoustic and linguistic considerations. *Brain. Lang.* 33, 146–160. doi: 10.1016/0093-934X(88)90059-4
- Karan, B., Sahu, S. S., and Mahto, K. (2020). Parkinson disease prediction using intrinsic mode function based features from speech signal. *Biocybern. Biomed. Eng.* 40, 249–264. doi: 10.1016/j.bbe.2019.05.005
- Kuruville-Dugdale, M., Salazar, M., Zhang, A., and Mefferd, A. S. (2020). Detection of articulatory deficits in Parkinson’s disease: Can systematic manipulations of phonetic complexity help? *J. Speech Lang. Hear. Res.* 63, 2084–2098. doi: 10.1044/2020_JSLHR-19-00245
- Ladefoged, P., and Harshman, R. (1979). “Formant frequencies and movements of the tongue,” in *Proceedings of the UCLA working papers in phonetics*, (Los Angeles, CA: UCLA), 39–52.
- Lamba, R., Gulati, T., Alharbi, H. F., and Jain, A. (2022). A hybrid system for Parkinson’s disease diagnosis using machine learning techniques. *Int. J. Speech Technol.* 25, 583–593. doi: 10.1007/s10772-021-09837-9
- Li, Y., Tan, M., Fan, H., Li, J., Xu, Z., Bian, R., et al. (2020). Lee Silverman voice therapy can improve the speech of Chinese-speakers with Parkinson’s disease. *Chin. J. Phys. Med. Rehabil.* 42, 245–248. doi: 10.3760/cma.j.issn.0254-1424.2020.03.013
- Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., and Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease. *IEEE Trans. Biomed. Eng.* 56, 1015–1022. doi: 10.1109/tbme.2008.2005954
- Mekyska, J., Galaz, Z., Mzourek, Z., Smekal, Z., Rektorova, I., Eliasova, I., et al. (2015). “Assessing progress of Parkinson’s disease using acoustic analysis of phonation,” in *Proceedings of the 2015 4th international work conference on bioinspired intelligence (IWOB1)*, (Piscataway, NJ: IEEE), 111–118.
- Moro-Velazquez, L., Gomez-Garcia, J. A., Arias-Londoño, J. D., Dehak, N., and Godino-Llorente, J. I. (2021). Advances in Parkinson’s disease detection and assessment using voice and speech: A review of the articulatory and phonatory aspects. *Biomed. Signal Process. Control* 66:102418. doi: 10.1016/j.bspc.2021.102418
- Muthukrishnan, R., and Rohini, R. (2016). “LASSO: A feature selection technique in predictive modeling for machine learning,” in *Proceedings of the 2016 IEEE international conference on advances in computer applications (ICACA)*, (New York, NY: IEEE), 18–20. doi: 10.1109/icaca.2016.7887916
- Novotny, M., Ruzs, J., Ěmejla, R., and Růžička, E. (2014). Automatic evaluation of articulatory disorders in Parkinson’s disease. *IEEE/ACM Trans. Audio Speech Lang. Process.* 22, 1366–1378. doi: 10.1109/TASLP.2014.2329734
- Orozco-Arroyave, J. R., Belalcázar-Bolaños, E. A., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Haderlein, T., and Nöth, E. (2014). “Phonation and articulation analysis of Spanish vowels for automatic detection of Parkinson’s disease,” in *Text, speech and dialogue*, eds P. Sojka, A. Horák, I. Kopeček, and K. Pala (Berlin: Springer), 374–381.
- Orozco-Arroyave, J. R., Hönl, F., Arias-Londoño, J. D., Vargas-Bonilla, J. F., Daqrouq, K., Skodda, S., et al. (2016). Automatic detection of Parkinson’s disease in rousing speech spoken in three different languages. *J. Acoust. Soc. Am.* 139, 481–500. doi: 10.1121/1.4939739
- Pah, N. D., Motin, M. A., and Kumar, D. K. (2022). Phonemes based detection of Parkinson’s disease for telehealth applications. *Sci. Rep.* 12, 1–9. doi: 10.1038/s41598-022-13865-z
- Park, H., and Kwon, H.-C. (2007). “Extended relief algorithms in instance-based feature filtering,” in *Proceedings of the sixth international conference on advanced*

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- language processing and web information technology (ALPIT 2007), (New York, NY: IEEE), 123–128. doi: 10.1109/alpit.2007.16
- Pavlovskaya, I. Y., and Hao, L. (2020). *The influence of breathing function in speech on mastering english pronunciation by Chinese students*. Amsterdam: Atlantis Press, 35–42. doi: 10.2991/assehr.k.200205.008
- Perez, K. S., Ramig, L. O., Smith, M. E., and Dromey, C. (1996). The Parkinson larynx: Tremor and videostroboscopic findings. *J. Voice* 10, 354–361. doi: 10.1016/s0892-1997(96)80027-0
- Remeseiro, B., and Bolon-Canedo, V. (2019). A review of feature selection methods in medical applications. *Comput. Biol. Med.* 112:103375. doi: 10.1016/j.combiomed.2019.103375
- Rusz, J., Cmejla, R., Ruzickova, H., and Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *J. Acoust. Soc. Am.* 129, 350–367. doi: 10.1121/1.3514381
- Rusz, J., Cmejla, R., Tykalova, T., Ruzickova, H., Klempir, J., Majerova, V., et al. (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task. *J. Acoust. Soc. Am.* 134, 2171–2181. doi: 10.1121/1.4816541
- Sapir, S., Ramig, L. O., Spielman, J. L., and Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *J. Speech Lang. Hear. Res.* 53, 114–125. doi: 10.1044/1092-4388(2009/08-0184)
- Singh, N., Pillay, V., and Choonara, Y. E. (2007). Advances in the treatment of Parkinson's disease. *Prog. Neurobiol.* 81, 29–44. doi: 10.1016/j.pneurobio.2006.11.009
- Skodda, S., Grönheit, W., and Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS One* 7:e32132. doi: 10.1371/journal.pone.0032132
- Solana-Lavalle, G., Galán-Hernández, J., and Rosas-Romero, R. (2020). Automatic Parkinson disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybern. Biomed. Eng.* 40, 505–516.
- Su, M., and Chuang, K. (2015). "Dynamic feature selection for detecting Parkinson's disease through voice signal," in *Proceedings of the 2015 IEEE MTT-S 2015 international microwave workshop series on RF and wireless technologies for biomedical and healthcare applications (IMWS-BIO)*, (Piscataway, NJ: IEEE), 148–149. doi: 10.1109/IMWS-BIO.2015.7303822
- Tsanas, A., Little, M. A., Fox, C., and Ramig, L. O. (2014). Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 181–190. doi: 10.1109/tnsre.2013.2293575
- Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. (2010). Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. *IEEE Trans. Biomed. Eng.* 57, 884–893. doi: 10.1109/tbme.2009.2036000
- Tykalova, T., Rusz, J., Klempir, J., Cmejla, R., and Ruzicka, E. (2017). Distinct patterns of imprecise consonant articulation among Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Brain Lang.* 165, 1–9. doi: 10.1016/j.bandl.2016.11.005
- Vásquez-Correa, J. C., Orozco-Arroyave, J. R., Bocklet, T., and Nöth, E. (2018). Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease. *J. Commun. Disord.* 76, 21–36. doi: 10.1016/j.jcomdis.2018.08.002
- Vilda, P. G., Mekyska, J., Rodellar, A. G., Alonso, D. P., Biarge, V. R., and Marquina, A. A. (2017). Monitoring parkinson disease from speech articulation kinematics. *Loquens Revista Española De Ciencias Del Habla* 4, 2386–2637. doi: 10.3989/loquens.2017.036
- Whitfield, J. A., and Goberman, A. M. (2014). Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *J. Commun. Disord.* 51, 19–28. doi: 10.1016/j.jcomdis.2014.06.005
- Zhang, J., Xu, W., Zhang, Q., Jin, B., and Wei, X. (2017). "Exploring risk factors and predicting UPDRS score based on Parkinson's speech signals," in *Proceedings of the 2017 IEEE 19th international conference on e-health networking, applications and services (Healthcom)*, (New York, NY: IEEE), 1–6. doi: 10.1109/HealthCom.2017.8210785
- Zhang, T., Hong, W., Chang, F., and Liu, X. (2011). Speech features analysis of Parkinson's disease by vowel class separability. *Chin. J. Biomed. Eng.* 30, 476–480. doi: 10.3969/j.issn.0258-8021.2011.03.026
- Zhang, T., Zhang, Y., Sun, H., and Shan, H. (2021). Parkinson disease detection using energy direction features based on EMD from voice signal. *Biocybern. Biomed. Eng.* 41, 127–141. doi: 10.1016/j.bbe.2020.12.009
- Zhang, Y. N. (2017). Can a smartphone diagnose Parkinson disease? A deep neural network method and telediagnosis system implementation. *Parkinsons Dis.* 2017:6209703. doi: 10.1155/2017/6209703