2025 International Conference on Advanced Mechatronics and Intelligent Energy Systems

Voice Analysis-Based Parkinson's Disease Detection Method Research

AIPCP25-CF-AMIES2025-00060 | Article

Voice Analysis-Based Parkinson's Disease Detection Method Research

Longhui Ge [1], Anzhe Sun [2] and Wenkai Zhang [3, a)]

[1]*School of Telecommunications and Intelligent Manufacturing, SIAS University, Zhengzhou, China, 451150*
[2]*School of Physical Science and Information Engineering, Liaocheng University, Liaocheng, China, 252000*
[3]*College of Pharmaceutical Information Engineering, Guangdong Pharmaceutical University, Guangdong, China, 527332*

[a)] Corresponding author: *2023401959@stu.lcu.edu.cn*

Abstract. Early and accurate diagnosis of Parkinson's disease (PD) remains challenging due to its insidious onset and reliance on subjective clinical criteria. Given that over 90 % of PD patients exhibit measurable speech impairments, this study develops a non-invasive, low-cost screening method by analyzing vocal features with machine learning. We preprocess a standardized dataset of 188 subjects (107 male, 81 female; ages 33–87), augmenting and normalizing speech recordings, then perform feature selection to reduce 287 acoustic variables to 31 key predictors. Three classification algorithms—Logistic Regression, XGBoost, and Support Vector Machine—are trained on an 80 %/20 % train/validation split. The Logistic Regression model achieves the best balance of accuracy (86 %), precision (healthy: 0.77; PD: 0.89), recall (healthy: 0.71; PD: 0.92), and F1-scores (healthy: 0.74; PD: 0.91), demonstrating robustness against data imbalance. XGBoost exhibits overfitting, and SVM underperforms on healthy sample detection. Our findings confirm that simple, interpretable models can effectively distinguish PD-related dysarthric speech from normal speech, offering a promising tool for large-scale screening. Future work will address environmental noise, dialect variation, and sample diversity to enhance cross-lingual applicability and clinical utility.

INTRODUCTION

Research on Parkinson's disease (PD) can be traced back to the 19th century. Despite advances in neuroscience, genetics, and molecular biology that have elucidated its pathological mechanisms, diagnosis, and treatment, PD remains one of the most prevalent neurodegenerative disorders worldwide. Its pathological hallmark is the degeneration and death of dopaminergic neurons in the substantia nigra of the midbrain [1], leading to a marked reduction in dopamine levels. This deficit manifests as intermittent motor symptoms—such as resting tremor and bradykinesia—that progressively worsen. Notably, approximately 92 % of PD patients exhibit quantifiable speech abnormalities (e.g., reduced loudness and slurred articulation), suggesting potential biomarkers for non-invasive diagnosis. Epidemiological data indicate that over 10 million individuals currently live with PD globally, and this number is projected to rise to 17 million by 2040 [2, 3], underscoring the urgent need for early screening. However, existing diagnostic approaches rely heavily on invasive procedures (e.g., cerebrospinal fluid analysis) and subjective clinical assessment, which are costly, complex, and prone to inter-observer variability [4].

Parkinson's disease is a chronic neurodegenerative disorder that severely diminishes patients' quality of life and imposes a substantial socioeconomic burden. Early diagnosis and timely intervention are critical to slowing disease progression and improving outcomes. Yet clinical diagnosis still depends primarily on patient history and neurological examination, with symptom severity (e.g., bradykinesia, tremor) evaluated subjectively. The absence of widely accepted objective biomarkers results in limited sensitivity and specificity, low early-diagnosis rates, and high misdiagnosis rates. The internationally accepted diagnostic standard is the UK Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria (UK Brain Bank Criteria), which include:

1. Core motor symptom criteria: bradykinesia plus at least one of resting tremor (4–6 Hz), rigidity (cogwheel or lead-pipe), or postural instability (excluding other causes).

2. Supportive criteria: unilateral onset, levodopa responsiveness, progressive course, hyposmia, REM sleep behavior disorder (RBD), etc.

3. Exclusion criteria: cerebellar signs, supranuclear gaze palsy, early severe autonomic failure (suggestive of atypical parkinsonism).

Although these criteria standardize clinical assessment, results are heavily influenced by clinician expertise, leading to inter-observer variability. To overcome these limitations, we propose leveraging computer-based data analysis methods for PD diagnosis. Artificial intelligence (AI), particularly machine learning (ML), has shown great promise in neurological diseases, including PD, for early diagnosis, monitoring, treatment optimization, and mechanistic insights. Voice-based PD diagnosis is an emerging non-invasive, low-cost screening approach that analyzes acoustic features—such as reduced loudness, imprecise articulation, and monotonic pitch—and applies ML classifiers to distinguish PD patients from healthy controls. Early voice abnormalities make this technique especially suited for high-risk population screening and disease progression monitoring. While some studies have explored cross-language voice analysis for PD, investigations using Chinese speech datasets remain scarce. Further research is needed to identify language-independent acoustic biomarkers for cross-lingual PD detection.

In this study, we employ a dataset of 188 PD patients (107 males, 81 females; ages 33–87) and healthy controls. After data preprocessing, we load the CSV file into Pandas, split it into 80 % training and 20 % validation sets, perform feature selection and random sampling, and then train and evaluate ML models. Our goal is to accurately classify PD patients versus healthy individuals based on speech features.

The chapters of the paper are arranged as follows:

Chapter 1: Introduction. This chapter first elaborates on the research background and current status of Parkinson's disease, and after introducing the current main diagnostic methods for Parkinson's disease, it briefly summarizes the research methodology related to this topic and the subsequent chapter arrangement of this paper.

Chapter 2: Method Introduction. This chapter provides a brief explanation of there levant technologies and principles involved in this paper, including data preprocessing and recognition analysis, as well as a preliminary introduction to the model used.

Chapter 3: Experiment and Analysis. This chapter introduces a detection algorithm based on voice features of Parkinson's patients, and validates the effectiveness of the method by comparing the performance of the models used, thereby supporting the research conclusions. Finally, a preliminary analysis of the experimental results is conducted.

Chapter 4: Applications and Challenges. This chapter mainly discusses the potential application scenarios of this research and the current problems, shortcomings, and challenges to be addressed.

Chapter 5: Conclusion and Future Prospects. This chapter explains the limitations of the paper and outlines future research directions for the topic.

METHOD INTRODUCTION

Preprocessing

First, establish a complete data processing environment. During the data loading phase, read the Parkinson's disease speech feature dataset. In the data cleaning stage, remove invalid and abnormal data, and use a grouped averaging strategy to handle duplicate entries. By grouping samples with the same ID and calculating their average values, ensure each patient retains only one complete feature record, thereby guaranteeing data quality. Expand the dataset by adjusting pitch, speech rate, and adding noise to alleviate issues like insufficient data or missing audio segments. Finally, perform data normalization to scale feature values to a uniform range, eliminating differences in units and magnitudes, thereby improving model performance [5].

The feature engineering adopted a multi-level processing approach. First, an iterative feature selection mechanism was implemented to identify and remove highly correlated features by calculating correlation coefficients between them. A correlation coefficient threshold of 0.7 was set—when the correlation between two features exceeds this threshold, one feature is retained while the other is removed.

The feature selection stage employed a three-step processing strategy. Step 1: Separate features from labels. Step 2: Use MinMaxScaler to normalize features, mapping all feature values to the 0-1 range. Step 3: Apply the chi-square test to select the most important 30 features. This significantly reduced data dimensionality, decreasing the number of features from 287 to 31, retaining the most predictive feature combinations.

Classification and Recognition

A standard dataset partitioning scheme was adopted, dividing the data into 80% training set and 20% validation set. This ensures the reproducibility of experimental results, guarantees sufficient training data, and reserves adequate validation data for model evaluation.

Implemented a comparative experiment using three different types of classification models. First is the logistic regression model, logistic regression is mainly applied to binary classification problems, making it suitable for the voice signal data studied in this paper. Logistic regression is an important and significant machine learning algorithm:

For a classification problem, a cost function is established, and then an optimization method is used to iteratively solve for the optimal model parameters [6]. The quality of the solved model is then tested and validated. Although logistic regression has "regression" in its name, it is actually a classification method, mainly used for binary classification problems. Logistic regression is primarily used to detect the probability of Parkinson's disease in experimental data.

Logistic Regression function (also known as the Sigmoid function), with the functional form as shown Figure.1.

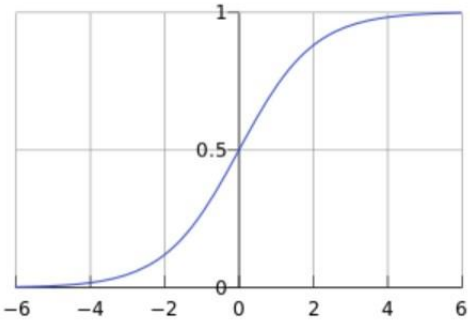

FIGURE 1. *Logistic regression*

In this experiment, y is a qualitative variable, where y=1 is defined as having Parkinson's disease, and y=0 indicates the opposite. There are the calculation process.

$$g(z) = \frac{1}{1+e^{-z}} \tag{1}$$

$$\pi(x) = \frac{1}{1+e^{-w^T x}} \tag{2}$$

$$p_1 = P(y=1|x) = \frac{1}{1+e^{-w^T x}} \tag{3}$$

$$p_0 = P(y=0|x) = 1 - P(y=1|x) = \frac{e^{-w^T x}}{1+e^{-w^T x}} \tag{4}$$

XGBoost classifier: Decision trees are the fundamental components of the XGBoost classifier. Each subsequent decision tree is built based on the prediction bias of the previous one, and splits are made sequentially. The voice features of Parkinson's disease patients are input into each decision tree of XGBoost. Each corresponding node in the decision trees has a predictive weigh t_w. By summing up "the predictive weights in each decision tree", the final prediction result is obtained—whichever is larger determines the final outcome. The formula for the n-th tree is defined as follows:

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^t) + \Omega(f_t) \tag{5}$$

According to the definition of this classifier, the objective function is transformed as:

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{t-1} + f_t(\boldsymbol{x_i})) + \Omega(f_t) \tag{6}$$

Support Vector Machine model (SVM) model: The Support Vector Machine is a classification model typically used for small to medium-scale binary and multi-class classification problems. It separates samples of different categories by finding the optimal hyperplane in the feature space.

EXPERIMENTAL ANALYSIS

Dataset Introduction

Datasets play a crucial role in studying the impact of artificial intelligence on Parkinson's disease. In PD detection research, there are primarily three types of datasets that provide patients' voice samples.

UC Irvine Parkinson's Disease Detection Dataset: The UC Irvine Parkinson's Disease Detection Dataset is a standard benchmark dataset for machine learning and deep learning research. Provided by the University of California, Irvine (UCI), it is part of the UCI database. The dataset contains relevant data from 188 patients with Parkinson's disease, including 107 males and 81 females, aged between 33 and 87 years. This data is primarily used for detecting and classifying Parkinson's disease, holding significant clinical importance as early and accurate diagnosis is crucial for timely intervention and personalized treatment. Variables in the dataset encompass various features such as time-frequency characteristics, Mel Frequency Cepstral Coefficients (MFCCs), wavelet transform features, vocal fold features, and TWQT features. These features help researchers better understand the physiological state of Parkinson's patients and develop more precise diagnostic models.

Experimental Environment and Preparation

Since Parkinson's disease primarily affects patients' vocal cord vibrations and speech prosody, voice feature extraction is a critical step in detecting Parkinson's disease. Patient speech contains complex information, including sound frequency, syllable intensity, speech rate, etc. The extraction of voice features determines the accuracy of experimental detection, so the experiment has relatively high requirements for hardware equipment. Sound is generated by the vibration of objects creating sound waves, and a microphone is a device that captures these sound waves and converts them into analog voltage signal outputs. The analog signal is amplified by a preamplifier, high-frequency components are removed to avoid aliasing, then sampled at a 44.1kHz sampling rate, and finally quantized to encode the data into PCM format.

Evaluation Criteria

The logistic regression model achieved an accuracy of 82.98% on the training set and 80.12% on the validation set, with only a 2.86% difference between the training and validation sets, indicating excellent model stability. In terms of specific prediction performance, the precision for healthy samples reached 0.77, recall was 0.71, and the F1 score was 0.74. The predictions for diseased samples were even more outstanding, with precision at 0.89, recall as high as 0.92, and an F1 score of 0.91. The overall accuracy reached 0.86.Thismodel maintained a good balance in predicting both types of samples, with particularly outstanding performance in identifying diseased samples. The support numbers show that the validation set contained 14 healthy samples and 37 diseased samples, and the model maintained stable performance despite this imbalanced data distribution. The confusion matrix for the logistic regression model is shown Figure.2.

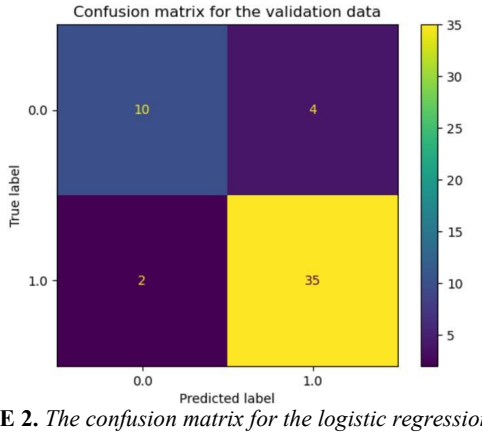

FIGURE 2. *The confusion matrix for the logistic regression model*.

The XGBoost model performed exceptionally well on the training set, with accuracy approaching 100%, but its accuracy dropped to 78.57% on the validation set, showing a performance decline of over 20%. This significant difference clearly indicates severe overfitting issues. The model's high complexity and poor interpretability of prediction results are notable drawbacks in medical diagnostic scenarios.

The support vector machine (SVM) model achieved an accuracy of 71.02% on the training set and 75.87% on the validation set, showing the unusual case where validation performance was higher than training performance. It performed poorly in predicting healthy samples, with precision as low as 0.41, and although recall reached 0.64, the F1 score was only 0.50. The overall accuracy was 0.65, the lowest among the three models. The model exhibited clear imbalance in predicting both classes, particularly with insufficient performance in identifying healthy samples. SVM's confusion matrix as shown Figure.3.

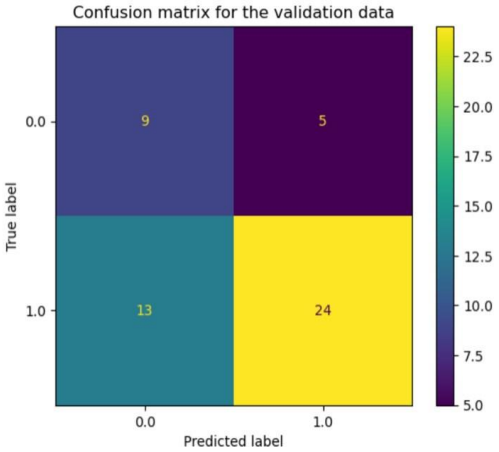

FIGURE 3. *The SVM's confusion matrix*.

Experimental Results Analysis

Through an in-depth comparison of the three models, the logistic regression model demonstrated the best overall performance. It not only achieved the highest prediction accuracy but also outperformed the other two models in terms of stability and class balance. The success of the logistic regression model indicates that for this type of medical diagnostic task, model stability and generalization ability are more important than model complexity. In detecting Parkinson's disease patients' speech data, it provides interpretable prediction results, which holds special value in medical applications.

APPLICATION CHALLENGES

Data Noise Issues

Background noise can significantly reduce the detection accuracy of the model. In real-world application scenarios, common noise sources include wind and rain sounds, traffic noise, microphone background noise, and human interference noise. These noises can distort the spectral features of speech signals, thereby degrading the model's performance and accuracy [5].

Individual Differences Issue

Currently, the aggregated data only includes voice data from Parkinson's patients abroad, which may lead to inaccurate detection for domestic patients [6]. Additionally, Chinese has unique tonal and pronunciation characteristics, with significant dialectal variations across different regions, requiring a large sample collection to supplement the dataset [7,8]. Furthermore, there are subtle differences in diagnostic criteria between domestic and international standards, resulting in the model's inability to adapt to the detection needs of domestic Parkinson's patients.

Small Sample Size Issue

The global prevalence of Parkinson's disease is approximately 0.3%, with limited sample sizes and difficulties in acquisition. Early-stage disease samples are particularly scarce, and samples of different subtypes are unevenly distributed, lacking long-term follow-up data. This leads to insufficient and less accurate datasets [9,10].

CONCLUSIONS

The logistic regression model achieved an accuracy of 82.98% on the training set and 80.12% on the validation set, demonstrating excellent model stability. In terms of specific prediction performance, it achieved a precision of 0.77, recall of 0.71, and F1-score of 0.74 for healthy samples. The predictions for diseased samples were even more outstanding, with a precision of 0.89, recall of 0.92, and F1-score of 0.91. The overall accuracy reached 0.86. This model maintained a good balance in predicting both types of samples, with particularly strong performance in identifying diseased samples. The support numbers showed that the validation set contained 14 healthy samples and 37 diseased samples, yet the model maintained stable performance despite this imbalanced data distribution.

The model performed exceptionally well on the training set, with accuracy nearing 100%, but its accuracy dropped to 78.57% on the validation set. This significant discrepancy clearly indicates severe overfitting issues. The model's high complexity and poor interpretability of prediction results are notable drawbacks in medical diagnostic scenarios.

The support vector machine model achieved an accuracy of 71.02% on the training set and 75.87% on the validation set, showing the unusual case where validation performance surpassed training performance. It performed poorly in predicting healthy samples, with a precision of only 0.41. Although the recall rate reached 0.64, the F1 score was merely 0.50. The overall accuracy was 0.65. The model exhibited significant imbalance in predicting both classes, particularly in identifying healthy samples.

Despite facing multiple challenges, voice analysis-based Parkinson's disease detection remains promising. With advancements in noise reduction technology, deep learning models such as DNN, CNN, and RNN can be optimized to significantly reduce the impact of noise on speech recognition systems. At the same time, collecting a large number of voice samples from Chinese Parkinson's patients and incorporating dialect recognition modules will make the system more tailored to the Chinese Parkinson's patient population. In the future, wearable devices could be developed to enable sustainable detection and early warning, and by integrating expert knowledge bases, patients could receive scientific and personalized treatment recommendations.

CONTRIBUTION

All the authors contributed equally and their names were listed in alphabetical order.

REFERENCES

1. Parkinson's Disease and Movement Disorders Group, Neurology Branch, Chinese Medical Association; Parkinson's Disease and Movement Disorders Group, Neurology Branch, Chinese Medical Doctor Association, "Chinese Guidelines for the Treatment of Parkinson's Disease (4th Edition)," Chinese Journal of Neurology, vol. 53, no. 12, pp. 1-12 (2020).
2. B. R. Bloem, M. S. Okun, and C. Klein, "New Evidence of the Parkinson's Disease Pandemic," Journal of Parkinson's Disease, vol. 13, no. 4, pp. 1-12 (2023).
3. A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Automated Detection of Speech Disorders in Parkinson's Disease Using Deep Learning," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 5, pp. 2155-2165 (2022).
4. A. Rana, A. S. Rawat, A. Bijarvan, and H. Bahuguna, "Application of Multilayer (Perceptron) Artificial Neural Networks in Diagnostic Systems: A Systematic Review," in Proceedings of the 2018 International Conference on Research in Intelligent and Computing Engineering (RICE), San Salvador, El Salvador, August 22-24, 2018, pp. 1-6.
5. M. Gao, J. Sun, Q. Li, et al., "Towards trustworthy image super-resolution via symmetrical and recursive artificial neural network," Image and Vision Comput. 105519 (2025).
6. R. Wang, J. Zhu, S. Wang, T. Wang, J. Huang, and X. Zhu, "Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking," Int. J. Multimed. Inf. Retr. 13, 39 (2024).
7. F. Wang, M. Ju, X. Zhu, Q. Zhu, H. Wang, C. Qian, and R. Wang, "A Geometric algebra-enhanced network for skin lesion detection with diagnostic prior," J. Supercomput. 81, 1-24 (2025).
8. R. S. D. T. Paz, A. L. Imbiriba, and L. C. Correa, "Ankle dorsiflexion range of motion is decreased in Parkinson's disease patients with freezing of gait and high l-dopa equivalent daily dose," Journal of Bodywork & Movement Therapies, vol. 25, pp. 42868-874 (2025).
9. D. Hemmerling, M. Kaczmarska, B. Krawczyk, et al., "Gait analysis in mixed reality for Parkinson's disease assessment," Biomedical Signal Processing and Control, vol. 61, pp. 107659-107659 (2025).
10. M. Mozafar, D. Z. Manshadi, Z. Molaei, et al., "Microstructural patterns in substantia nigra subregions are associated with depression and olfactory impairments in Parkinson's disease," Clinical Neurology and Neurosurgery, vol. 251, pp. 108817-108817 (2025).