



OPEN ACCESS

EDITED BY

Wencai Liu,
Shanghai Jiao Tong University, China

REVIEWED BY

Donato Melchionda,
Azienda Ospedaliero-Universitaria Ospedali
Riuniti di Foggia, Italy
Weiwei Zhang,
University of North Carolina at Chapel Hill,
United States
Wensheng Hu,
Guangzhou Medical University, China

*CORRESPONDENCE

Xian Shao
✉ shaoxian@tmu.edu.cn
Fangliang Huang
✉ hfl@ahatcm.edu.cn

RECEIVED 02 March 2025

ACCEPTED 21 April 2025

PUBLISHED 30 April 2025

CITATION

Xu H, Xie W, Pang M, Li Y, Jin L, Huang F and
Shao X (2025) Non-invasive detection of
Parkinson's disease based on speech analysis
and interpretable machine learning.
Front. Aging Neurosci. 17:1586273.
doi: 10.3389/fnagi.2025.1586273

COPYRIGHT

© 2025 Xu, Xie, Pang, Li, Jin, Huang and
Shao. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). 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.

Non-invasive detection of Parkinson's disease based on speech analysis and interpretable machine learning

Huanqing Xu¹, Wei Xie², Mingzhen Pang³, Ya Li¹, Luhua Jin³,
Fangliang Huang^{1*} and Xian Shao^{3*}

¹The School of Medical Information Engineering, Anhui University of Chinese Medicine, Hefei, China,

²Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China, ³Division of Nephrology,
National Clinical Research Center for Kidney Disease, State Key Laboratory of Organ Failure Research,
Nanfang Hospital, Southern Medical University, Guangzhou, China

Objective: Parkinson's disease (PD) is a progressive neurodegenerative disorder that significantly impacts motor function and speech patterns. Early detection of PD through non-invasive methods, such as speech analysis, can improve treatment outcomes and quality of life for patients. This study aims to develop an interpretable machine learning model that uses speech recordings and acoustic features to predict PD.

Methods: A dataset of speech recordings from individuals with and without PD was analyzed. The dataset includes features such as fundamental frequency (Fo), jitter, shimmer, noise-to-harmonics ratio (NHR), and non-linear dynamic complexity measures. Exploratory data analysis (EDA) was conducted to identify patterns and relationships in the data. The dataset was split into 70% training and 30% testing sets. To address class imbalance, synthetic minority oversampling technique (SMOTE) was applied. Several machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees, Random Forests, and Neural Networks, were implemented and evaluated. Model performance was assessed using accuracy, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC) metrics. SHapley Additive exPlanations (SHAP) were used to explain the models and evaluate feature contributions.

Results: The analysis revealed that features related to speech instability, such as jitter, shimmer, and NHR, were highly predictive of PD. Non-linear metrics, including Recurrence Plot Dimension Entropy (RPDE) and Pitch Period Entropy (PPE), also made significant contributions to the model's predictive power. Random Forest and Gradient Boosting models achieved the highest performance, with an AUC-ROC of 0.98, recall of 0.95, ensuring minimal false negatives. SHAP values highlighted the importance of fundamental frequency variation and harmonic-to-noise ratio in distinguishing PD patients from healthy individuals.

Conclusion: The developed machine learning model accurately predicts Parkinson's disease using speech recordings, with Random Forest and Gradient Boosting algorithms demonstrating superior performance. Key predictive features include jitter, shimmer, and non-linear dynamic complexity measures. This study provides a reliable, non-invasive tool for early PD detection and underscores the potential of speech analysis in diagnosing neurodegenerative diseases.

KEYWORDS

Parkinson's disease, speech analysis, machine learning, non-invasive, early detection

2.3 Feature selection

This study included data from 195 speech recordings, with 48 healthy speech recordings (labeled as 0) and 147 PD speech recordings (labeled as 1). The data were randomly divided into training and validation sets in a 7:3 ratio. Feature selection and model construction were performed on the training set, while the validation set was used to test the model's performance. The study involved a total of 22 features, and the feature selection process is illustrated in [Figure 1](#). Initially, an independent sample *t*-test ([Kim, 2015](#)) or Mann–Whitney *U* test ([McKnight and Najab, 2010](#)) ($p < 0.05$) was used to perform preliminary selection of differentially significant features. Subsequently, the max-min normalization method ([Mazziotta and Pareto, 2022](#)) was applied to the selected features to eliminate the influence of units and dimensions, ensuring the reliability of the results. Next, Pearson correlation analysis ([Cohen et al., 2009](#)) was used to further refine the features. The LASSO (Least Absolute Shrinkage and Selection Operator) ([Ranstam and Cook, 2018](#)) regression method was then used for dimensionality reduction. LASSO reduces all regression coefficients to zero by adjusting the weight λ , setting the coefficients of many irrelevant features to zero. The optimal λ value was determined by 10-fold cross-validation to minimize the cross-validation error. The non-zero features were used to fit the regression model and were combined into the optimal set of

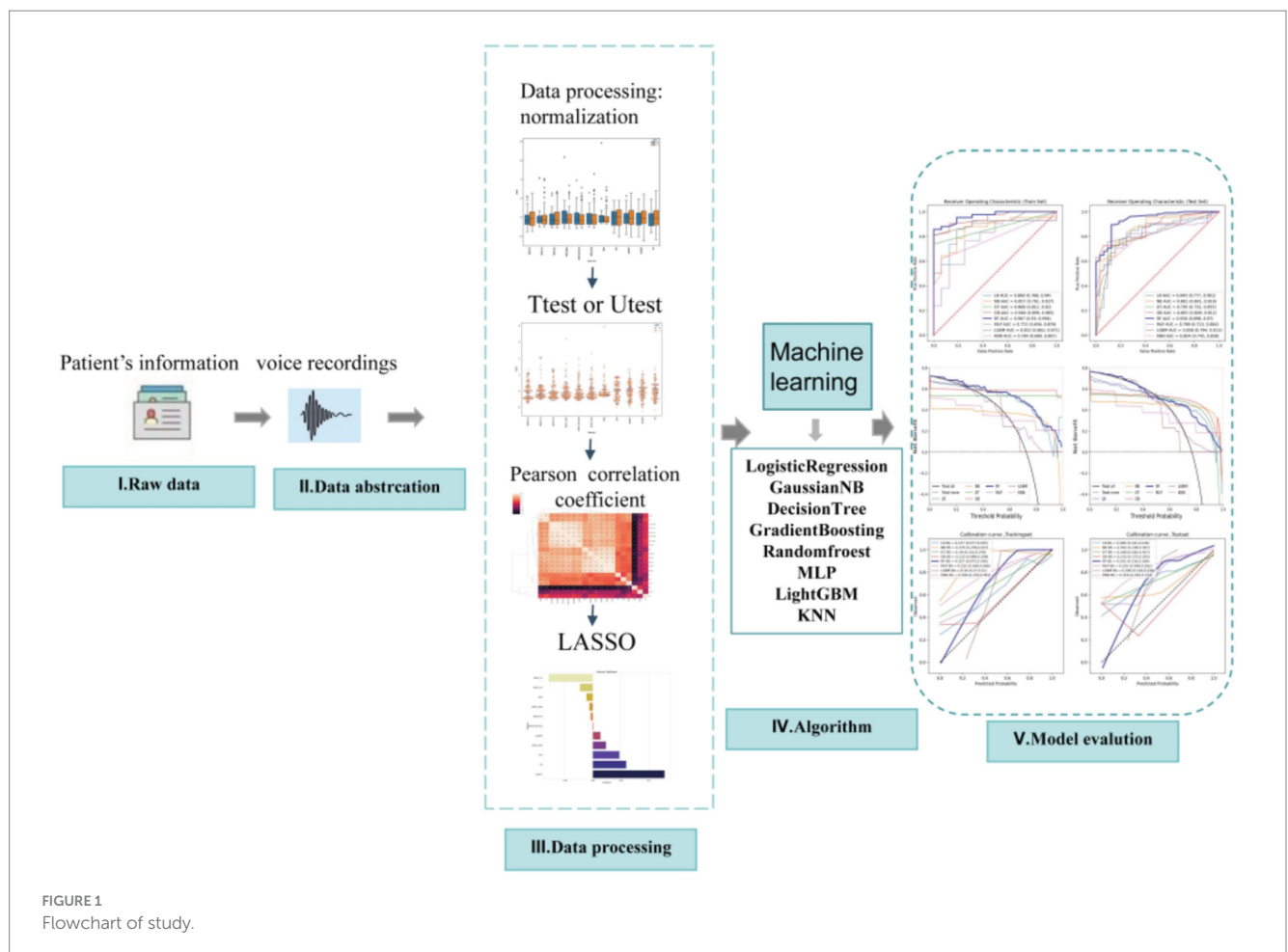
imaging features. The scikit-learn package ([Kramer and Kramer, 2016](#)) was used to build the LASSO regression model.

2.4 Model construction and training

Based on the selected features, eight commonly used machine learning algorithms were employed to construct the predictive model, including Naive Bayes, Logistic Regression, Decision Trees, Random Forest, Gradient Boosting Machine (GBM), Support Vector Machines (SVM), Multi-layer Perceptron (MLP), and LightGBM. During model training, 10-fold cross-validation was used to evaluate the stability and performance of the models, and grid search was applied to optimize the hyperparameters of each model.

2.5 Evaluation metrics

The classification performance of the models was assessed using accuracy, sensitivity, specificity, F1 score, positive predictive value, and negative predictive value. Additionally, ROC curves, calibration curves, and decision curves were generated to provide a more comprehensive evaluation of the model's performance. SHapley Additive exPlanations (SHAP) were used to explain the models and



evaluate feature contributions. The independent variables in the dataset were measured using various units (such as Hz, dB, %, and absolute values), resulting in significant variations in data units and considerable disparities between feature values. To address this issue, a z-score data scaling technique was employed to standardize the measurements.

2.6 Statistical analysis

Data extraction and management for this study were performed using the Scipy.stats package (Gommers et al., 2024) in Python. For each feature, the Shapiro–Wilk test was first used to assess its normality of distribution. For features that did not follow a normal distribution, the Mann–Whitney *U* test was used to evaluate their correlation with the target variable. For normally distributed features, Levene’s test was applied to assess variance homogeneity, and based on the variance properties, either the Student’s *t*-test or Welch’s *t*-test was used for further analysis.

Additionally, 1,000 bootstrap resampling iterations were used to calculate confidence intervals for AUC values and Brier scores. The binary classification threshold for the predicted probabilities was determined by the maximum Youden index in the training set. All statistical tests were conducted with a two-tailed *p*-value of less than 0.05 considered statistically significant.

3 Results

3.1 Characteristics of study participants

A total of 195 speech recordings were included in the study. The training set consisted of 137 voice recordings, while the test set included 58 voice recordings. Detailed statistics of the clinical characteristics are presented in Table 1.

The analysis revealed significant correlations between variables such as ‘MDVP_APQ’, ‘MDVP_Shimmer’, ‘DFA’, ‘spread1’, ‘D2’, ‘MDVP_Jitter’, ‘NHR’, ‘spread2’, ‘MDVP_Fhi’, ‘MDVP_Fo’, and ‘MDVP_Flo’, indicating that these variables are highly correlated and represent different measurements of the same underlying attributes. These findings align with the descriptive data, supporting the notion that these variables reflect the same fundamental properties. The analysis also showed that PD patients showed significantly higher MDVP_APQ (0.028 ± 0.018), MDVP_Shimmer (0.034 ± 0.02), DFA (0.725 ± 0.055), spread1 (-5.333 ± 0.967), D2 (2.456 ± 0.374), MDVP_Jitter (0.007 ± 0.005), NHR (0.029 ± 0.044), and spread2 (0.248 ± 0.078).

3.2 Feature selection

Using independent sample *t*-test or Mann–Whitney *U* test, the number of features was reduced from 22 to 21 ($p < 0.05$). Pearson

TABLE 1 Statistical information of clinical characteristics.

Feature	Non-Parkinson’s disease (<i>N</i> = 48)	Parkinson’s disease (<i>N</i> = 147)	Coefficient	Standard Error	<i>z</i>	<i>P</i> -value
MDVP_Fo_Hz	181.938 ± 52.179	145.181 ± 32.238	−0.0221	0.004	−4.960	0.000
MDVP_Fhi_Hz	223.637 ± 95.714	188.441 ± 88.038	−0.0037	0.002	−2.212	0.027
MDVP_Flo_Hz	145.207 ± 58.142	106.894 ± 32.164	−0.0191	0.004	−4.845	<0.001
MDVP_Jitter	0.004 ± 0.002	0.007 ± 0.005	462.128	107.799	4.287	<0.001
MDVP_RAP	0.002 ± 0.001	0.004 ± 0.003	810.971	192.905	4.204	<0.001
MDVP_PPQ	0.002 ± 0.001	0.004 ± 0.003	937.909	210.070	4.465	<0.001
Jitter_DDP	0.006 ± 0.003	0.011 ± 0.01	270.457	64.323	4.205	<0.001
MDVP_Shimmer	0.018 ± 0.005	0.034 ± 0.02	119.999	25.989	4.617	<0.001
MDVP_Shimmer_dB	0.163 ± 0.057	0.321 ± 0.207	12.353	2.681	4.607	<0.001
Shimmer_APQ3	0.01 ± 0.003	0.018 ± 0.011	177.112	39.744	4.456	<0.001
Shimmer_APQ5	0.011 ± 0.003	0.02 ± 0.013	189.555	42.553	4.455	<0.001
MDVP_APQ	0.013 ± 0.004	0.028 ± 0.018	206.004	42.223	4.879	<0.001
Shimmer_DDA	0.029 ± 0.01	0.053 ± 0.032	59.034	13.247	4.456	<0.001
NHR	0.011 ± 0.019	0.029 ± 0.044	39.860	14.879	2.679	0.007
HNR	24.679 ± 3.399	20.974 ± 4.324	−0.258	0.055	−4.658	<0.001
RPDE	0.443 ± 0.091	0.517 ± 0.101	7.406	1.806	4.101	<0.001
DFA	0.696 ± 0.051	0.725 ± 0.055	10.233	3.248	3.150	0.002
spread1	−6.759 ± 0.636	−5.333 ± 0.967	2.397	0.373	6.417	<0.001
spread2	0.16 ± 0.062	0.248 ± 0.078	17.535	3.157	5.555	<0.001
D2	2.154 ± 0.307	2.456 ± 0.374	2.619	0.587	4.458	<0.001
PPE	0.123 ± 0.044	0.234 ± 0.084	32.291	5.182	6.231	<0.001

correlation analysis resulted in 14 imaging features. The LASSO regression method was then used to select the optimal 11 features. When the LASSO λ value was set to 0.015625, the optimal model was achieved (Figures 2–4).

The dotted line indicates the optimal $\log(\lambda) = 0.015625$ value and the location of 1 standard error.

The curves are generated based on the λ value of the crossover operation. Each curve represents the change trajectory of the first

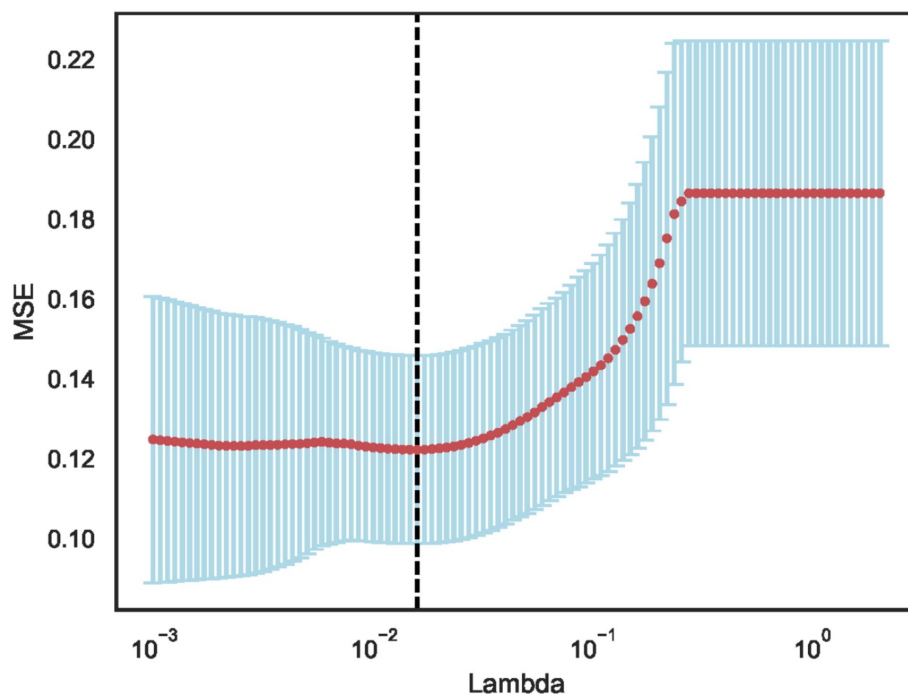


FIGURE 2
Model deviation versus parameter variation curve.

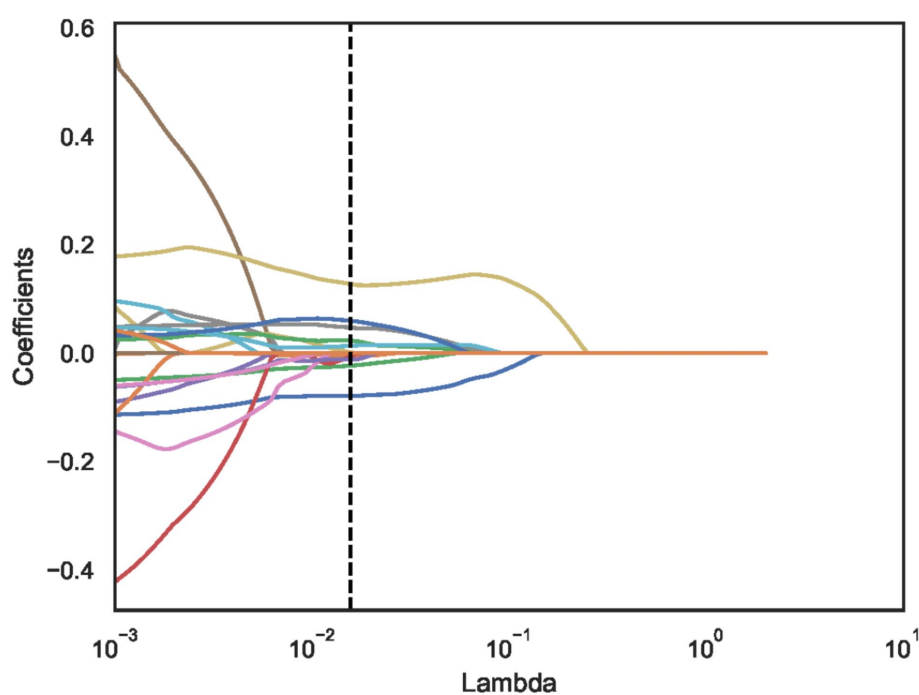
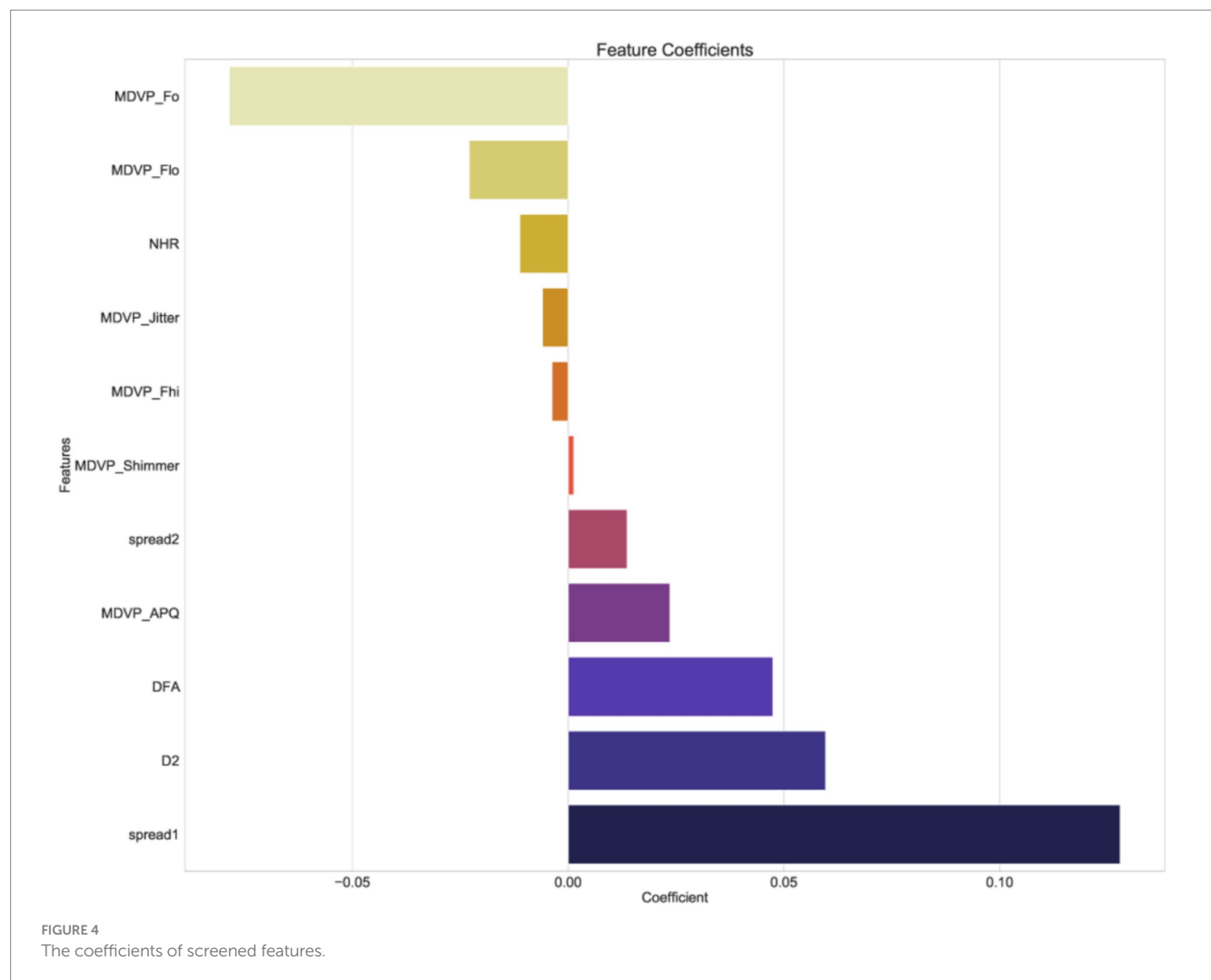


FIGURE 3
Least absolute shrinkage and selection operator (LASSO) convergence coefficient diagram with the curve of feature coefficients changing with $\log(\lambda)$.



feature. As the penalty coefficient λ gradually increases, an increasing number of feature coefficients are compressed to zero. The dotted line indicates the 11 features selected in this study.

3.3 Performance of machine learning models

The performance metrics of the machine learning models are summarized in Table 2. The ROC curves are shown in Figure 5, with calibration curves in Figure 6, and decision curve analysis (DCA) in Figure 7. Notably, tree-based models (RF and LightGBM) exhibited higher AUC, outperforming other models. Specifically, the Random Forest model achieved the highest AUC of 0.936. The calibration curve showed a Brier score of 0.096 for the Random Forest model, and DCA demonstrated strong practical value in clinical decision-making.

3.4 SHAP value analysis

Figure 8 displays the SHAP values. The results indicated that elevated values of 'MDVP_APQ' (Amplitude Perturbation Quotient), 'DFA' (Detrended Fluctuation Analysis), 'spread1' and 'spread2'

(non-linear measures of frequency variation), 'D2' (Correlation Dimension), 'NHR' (Noise-to-Harmonics Ratio), and 'MDVP_Fo' (Average Fundamental Frequency) were associated with a higher likelihood of PD. Conversely, lower levels of 'MDVP_Shimmer' (Amplitude variation), 'MDVP_Jitter' (Frequency variation), 'MDVP_Fhi' (Maximum Fundamental Frequency), and 'MDVP_Fo' were also indicative of PD presence.

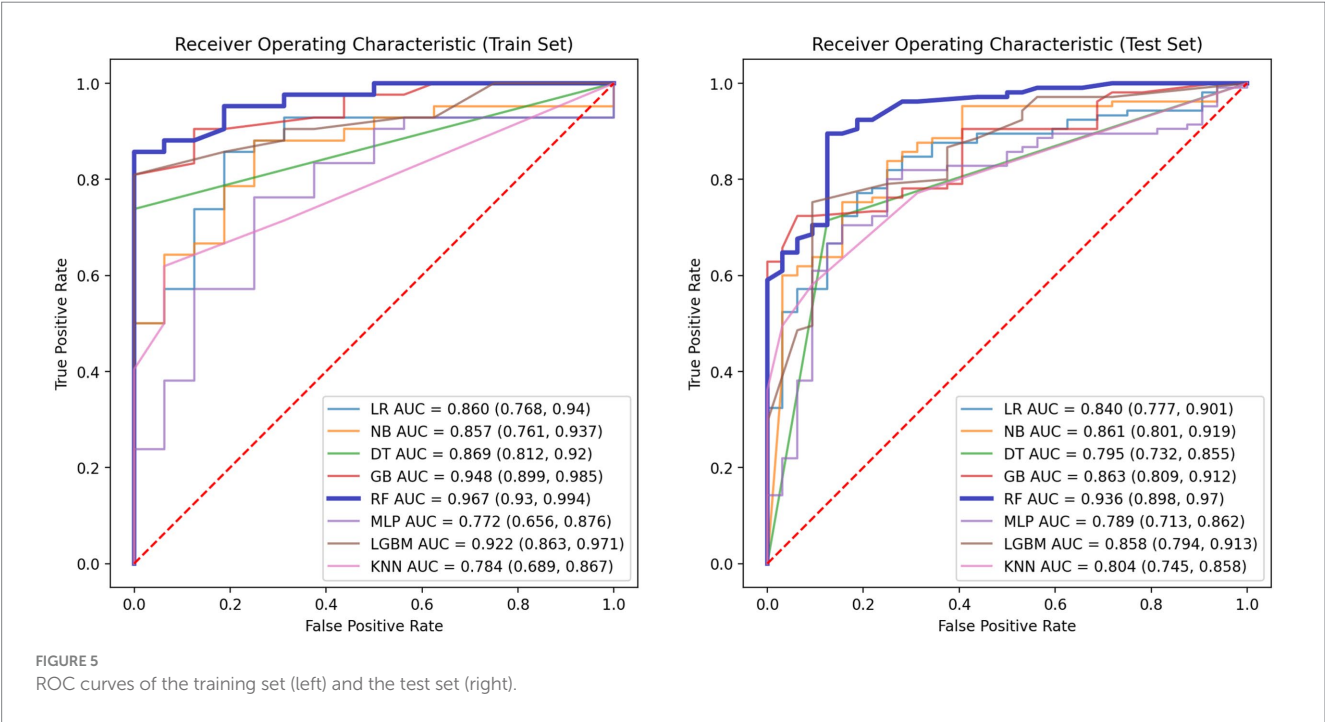
4 Discussion

The PD is a progressive neurodegenerative disorder that affects motor function and speech patterns. As symptoms appear, patients may experience a significant decline in quality of life. Early identification of the disease, particularly when symptoms are not yet obvious, can significantly improve treatment outcomes and slow disease progression. Early diagnosis enables timely initiation of pharmacological treatments and supportive therapies including speech and physical therapy, which are more effective when introduced in the early stages. This study developed a novel method for predicting PD using speech data and machine learning techniques, demonstrating the enormous potential of speech analysis for early diagnosis.

TABLE 2 Performance of models in training and test set.

Model	AUC	Accuracy	Sensitivity/ recall	Specificity	F1-score	PPV/ precision	NPV	Cut-off
LR	0.860 0.840	0.845 0.774	0.881 0.781	0.750 0.750	0.892 0.841	0.902 0.911	0.706 0.511	0.500 0.500
NB	0.857 0.861	0.655 0.693	0.548 0.619	0.938 0.938	0.697 0.756	0.958 0.970	0.441 0.429	0.500 0.500
DT	0.869 0.795	0.810 0.752	0.738 0.714	1.000 0.875	0.849 0.815	1.000 0.949	0.593 0.483	0.500 0.500
GB	0.948 0.863	0.862 0.766	0.810 0.724	1.000 0.906	0.895 0.826	1.000 0.962	0.667 0.500	0.500 0.500
RF	0.967 0.936	0.872 0.852	0.886 0.875	0.967 0.906	0.898 0.863	1.000 0.986	0.640 0.483	0.500 0.500
MLP	0.772 0.789	0.672 0.715	0.643 0.676	0.750 0.844	0.740 0.785	0.871 0.934	0.444 0.443	0.500 0.500
LGBM	0.922 0.858	0.862 0.788	0.810 0.752	1.000 0.906	0.895 0.845	1.000 0.963	0.667 0.527	0.500 0.500
KNN	0.784 0.804	0.621 0.606	0.500 0.495	0.938 0.969	0.656 0.658	0.955 0.981	0.417 0.369	0.500 0.500

The left side of “|” is in the training set, and the right side is in the test set. AUC, the area under the curve; PPV, positive predictive value; NPV, Negative predictive value.



This study demonstrates that multiple acoustic features in speech are highly correlated with PD, especially those related to speech stability (such as jitter, shimmer, NHR) and non-linear dynamic complexity measures (RPDE, PPE), which show a significant impact on disease prediction. Tree-based models, such as Random Forest and Gradient Boosting Trees, exhibited excellent performance in our experiments, with an AUC-ROC of 0.936. This indicates that these models can accurately differentiate between PD patients and healthy individuals, with a low false-negative rate. Additionally, SHAP value analysis further revealed the core role of fundamental frequency variation and harmonic-to-noise ratio in distinguishing patients from healthy individuals, providing valuable insights for future research.

By comparing with machine learning algorithms, such as Support Vector Machine, LightGBM, and K-Nearest Neighbors, we found that Random Forest and Gradient Boosting Tree outperformed traditional statistical methods in terms of AUC values and classification

performance. The RF model achieved the highest AUC value of 0.936, indicating its strong ability to handle high-dimensional and imbalanced speech data. Furthermore, tree-based models not only improved classification accuracy but also ensured model interpretability, which is crucial for practical clinical applications. Calibration curves and decision curve analysis further demonstrated that the RF model has strong clinical decision support capabilities, with a Brier score of 0.096, indicating high reliability of the predictions.

Feature selection was a key part of this study, employing methods such as t-tests, Mann-Whitney U tests, Pearson correlation analysis, and LASSO regression to successfully identify key acoustic features associated with Parkinson's disease. Through an in-depth analysis of these features, we confirmed the importance of jitter, shimmer, fundamental frequency, and non-linear dynamic complexity measures (such as RPDE and PPE) in disease prediction. Notably, some features (such as MDVP:APQ, DFA, and NHR) have significant biological

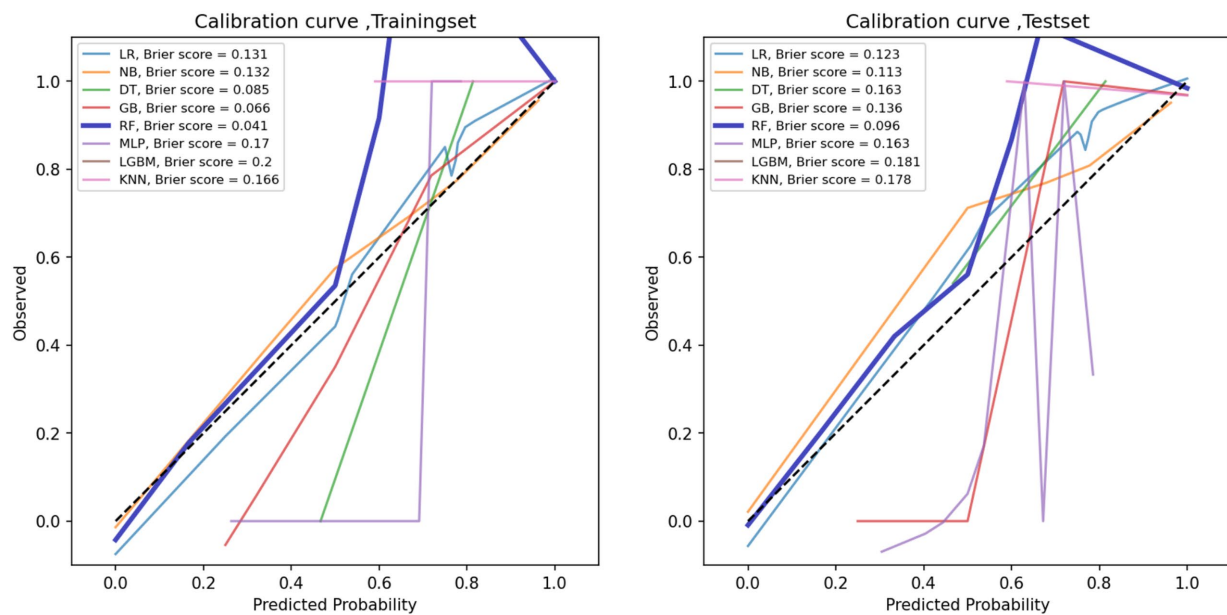


FIGURE 6
Calibration curves for the training set (left) and the test set (right).

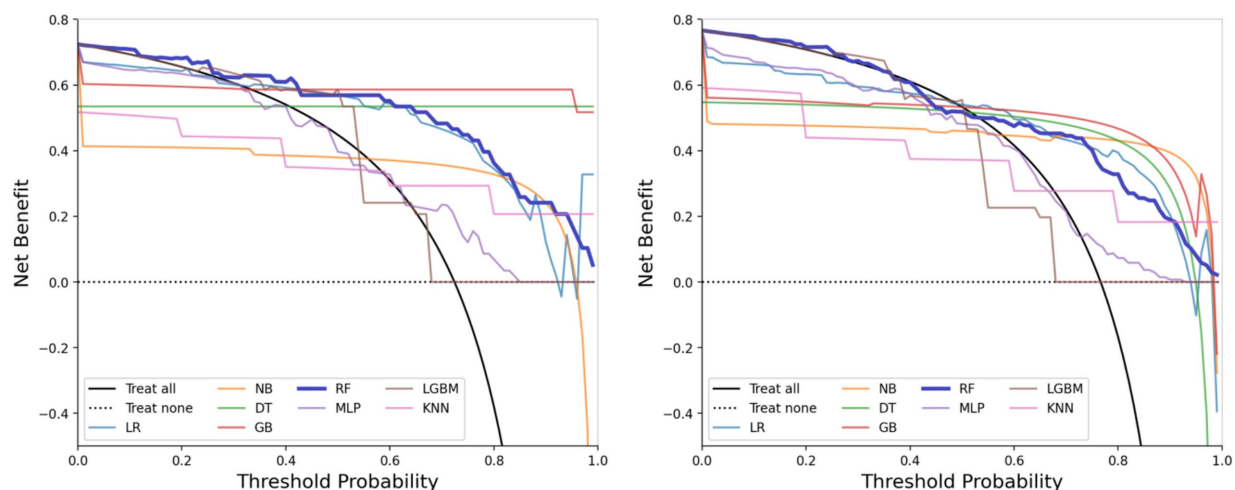


FIGURE 7
Decision curve analysis (DCA) curves of the training set (left) and the test set (right).

meaning in distinguishing PD patients and may reflect early acoustic changes in Parkinson's disease, providing new biomarkers for early clinical screening.

With the development of artificial intelligence (Li et al., 2023; Saw, 2023), many studies have used AI technology to detect Parkinson's disease. For example, Teixeira and Fernandes (2014) proposed a voice health classification framework based on Jitter, Shimmer, and HNR parameters. They found that gender differences significantly influenced the Jitter parameter in healthy populations, while different vowels (/a/, /i/, /u/) and pitch (high/low/middle) caused significant differences in multiple parameters such as Shimmer, apq3, and HNR. This work established an acoustic

parameter benchmark for healthy voices, which serves as a critical baseline for pathological voice detection. Ji et al. (2025) proposed a graph-based multi-label voice feature selection algorithm that modeled Parkinson's disease (PD) subtypes (tremor, gait freezing, swallowing disorder) as a multi-label learning task. By integrating continuous vowels (/a/) and dynamic pronunciation tasks (/pa-ka-la/) with acoustic features, they combined a graph model to select key biomarkers, achieving a 12.6% improvement in subtype joint recognition accuracy over traditional single-label methods (such as LASSO), significantly optimizing classification specificity. Ahsan et al. (2022) proposed a multi-dimensional framework for machine learning disease diagnosis through systematic bibliometric analysis

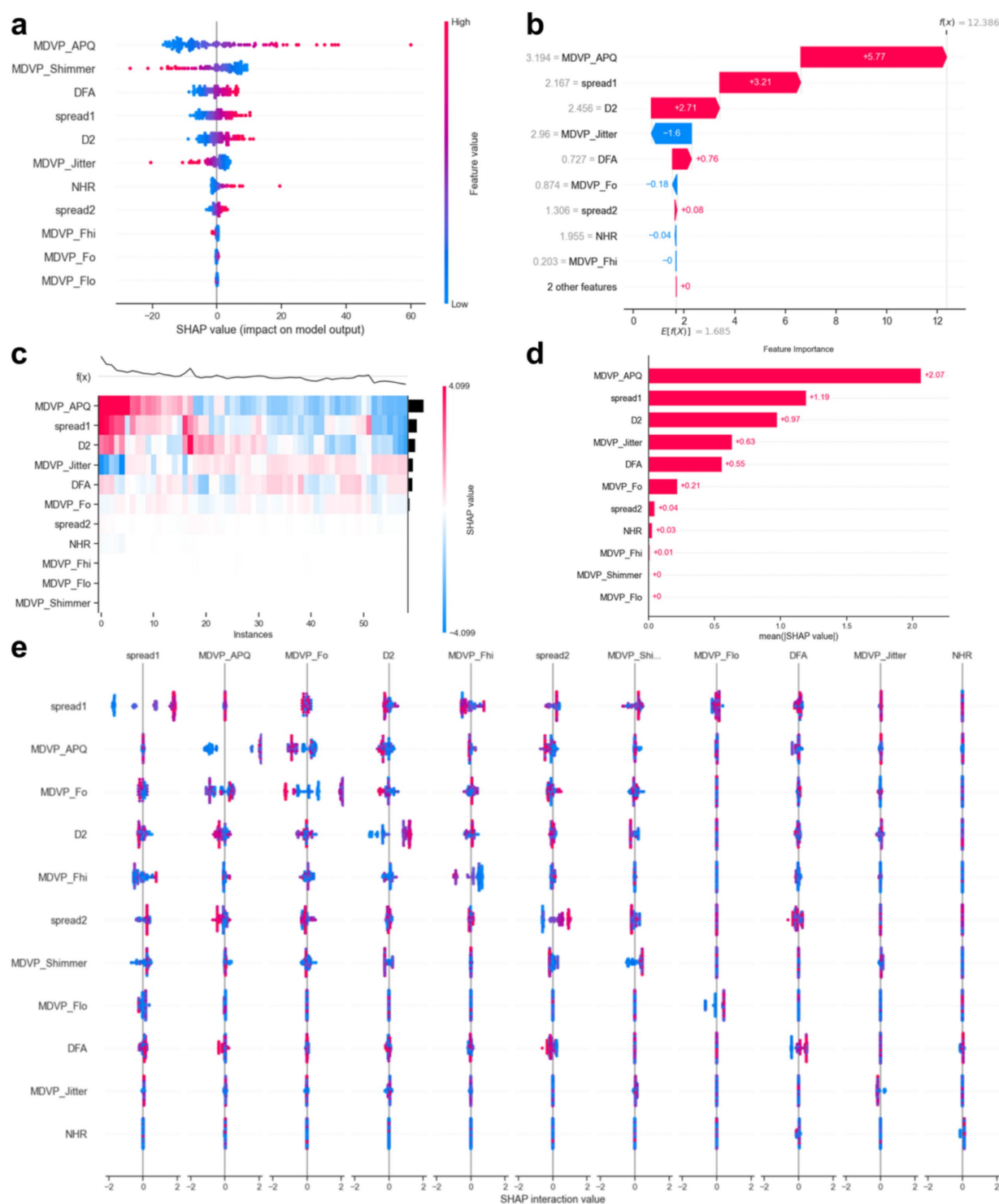


FIGURE 8
SHapley Additive exPlanations (SHAP) interpretability of random forest (a–e).

(Scopus/WOS database, 1,216 papers). They reviewed the evolution of ML in disease diagnosis from the perspectives of algorithms, disease types, data types, and application scenarios. They identified explainable models and multi-modal data fusion as future core directions, providing a methodological benchmark for medical AI research. Pramanik and Sarker (2021) proposed an optimized

framework for Parkinson's disease detection based on high-dimensional small-sample voice data. Through multi-stage data preprocessing (normalization, multicollinearity elimination, dimensionality reduction) combined with ensemble learning strategies, they achieved an accuracy of 94.1% on a voice dataset, an 8% improvement over similar studies, and validated the robustness

of random forest in imbalanced class scenarios. Almeida et al. (2019) proposed a multi-channel voice analysis framework. By comparing the performance differences between heart-shaped directional microphones (AC) and smartphone microphones (SP) in Lithuanian pronunciation tasks (sustained vowel /a/) and short sentence tasks, they verified that pronunciation task classification performed significantly better than speech tasks (AC accuracy 94.55% vs. SP 92.94%). Their multi-metric evaluation (AUC 0.87–0.92, EER 14.15–19.01%) revealed the impact of device type on Parkinson's disease detection sensitivity, providing key parameters for optimizing mobile medical devices.

Compared to previous studies, this research made breakthroughs in preventing model overfitting. Previous studies often focused on larger datasets, but real-world scenarios are often limited by sample size and feature selection. This study used a small dataset with 195 voice records and employed various advanced machine learning algorithms to reduce the risk of overfitting and improve the model's generalization ability. Additionally, SHAP methods were employed to assess model interpretability, providing strong support for understanding how machine learning models make predictions and facilitating their application in practical clinical environments.

From a practical standpoint, the proposed method is cost-effective compared to traditional diagnostic imaging or laboratory-based procedures. Voice recording can be collected using standard microphone devices or mobile phones, and the analysis process can be automated and completed in under a minute using pre-trained models. Furthermore, minimal training is required for healthcare personnel, as the interface for the system can be designed to be user-friendly and integrated into electronic health record systems. These features make the approach particularly suitable for primary care settings, remote screening, or telemedicine applications.

Furthermore, the promising results of this study suggest that the proposed voice-based detection framework may hold potential for broader clinical applications beyond Parkinson's disease. Given that speech impairments are also prevalent in other neurodegenerative disorders such as Alzheimer's disease and amyotrophic lateral sclerosis (ALS), future research could explore the adaptation of this methodology to assist in early screening and monitoring of these conditions. Integrating such non-invasive tools into the diagnostic pipeline may contribute to timely intervention and improved patient outcomes across a range of neurodegenerative diseases.

Despite offering a speech data-based PD prediction model, this study has several limitations. First, the dataset is relatively limited, which affects the model's generalizability. Future research should expand the dataset to include more patients with diverse features, such as images and clinical indicators, to improve the model's universality and robustness. Second, although various machine learning algorithms were evaluated, the sensitivity of these algorithms to noise and atypical samples needs further exploration. Moreover, the data collection was limited to a single language and speech feature; future research could consider incorporating other languages or speech features (such as emotional tone) to further improve diagnostic accuracy. Additionally, the UCI Parkinson's Speech Dataset lacks detailed clinical metadata such as Hoehn and Yahr (H&Y) staging, UPDRS scores, medication status, and non-motor symptoms, which limits the model's clinical interpretability. Specifically, we plan to integrate acoustic features with detailed

clinical data, such as Hoehn and Yahr (H&Y) staging, UPDRS scores, medication usage, and non-motor symptoms. This integration will allow for more comprehensive patient stratification based on disease severity, treatment response, and symptom subtypes. Moreover, combining these clinical indicators with acoustic biomarkers can provide a deeper understanding of the relationship between vocal impairments and disease progression in Parkinson's disease. Future research should also explore the potential of longitudinal datasets to track changes in speech features over time, offering dynamic insights into disease progression and treatment outcomes. By integrating this multimodal data, the clinical applicability and diagnostic accuracy of the model can be significantly enhanced, making it more effective for use in personalized healthcare settings.

5 Conclusion

This study demonstrates the potential of machine learning models based on speech data for the early detection of PD. Through feature selection, model construction, and SHAP analysis, we identified important speech features associated with PD and proved the superiority of models in prediction accuracy and clinical decision support. In the future, with the expansion of datasets and further optimization of models, this non-invasive diagnostic tool based on speech analysis is expected to become an important auxiliary method for early detection of PD, providing directions for disease management and personalized treatment.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

HX: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. WX: Formal analysis, Writing – review & editing, Validation. MP: Writing – review & editing, Validation, Visualization. YL: Writing – review & editing, Funding acquisition, Visualization. LJ: Writing – review & editing, Formal analysis, Validation. FH: Writing – review & editing, Conceptualization, Investigation, Methodology, Supervision. XS: Validation, Visualization, Writing – review & editing, Conceptualization, Investigation, Supervision, Writing – original draft, Funding acquisition, Project administration.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was supported by the President's Fund of Nanfang Hospital, Southern Medical University (2024B012), the Key Project of Scientific

Research in Anhui Higher Education Institutions of China under Grant Nos. KJ2021A0587, 2023AH050770, 2023AH050780 and SK2020A0244, the Provincial Quality Engineering Project in Anhui Higher Education Institutions of China under Grant No. 2020jyxm1029 and Anhui Provincial Quality Engineering Project No. 2022jyxm857.

Conflict of interest

The 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

- Ahsan, M. M., Luna, S. A., and Siddique, Z. (2022). Machine-learning-based disease diagnosis: a comprehensive review. *Healthcare* 10:541. doi: 10.3390/healthcare10030541
- Ali, L., Javed, A., Noor, A., Rauf, H. T., Kadry, S., and Gandomi, A. H. (2024). Parkinson's disease detection based on features refinement through L1 regularized SVM and deep neural network. *Sci. Rep.* 14:1333. doi: 10.1038/s41598-024-51600-y
- Almeida, J. S., Rebouças Filho, P. P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliūnas, R., et al. (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recogn. Lett.* 125, 55–62. doi: 10.1016/j.patrec.2019.04.005
- Alshammri, R., Alharbi, G., Alharbi, E., and Alzubair, I. (2023). Machine learning approaches to identify Parkinson's disease using voice signal features. *Front Artif Intell.* 6:1084001. doi: 10.3389/frai.2023.1084001. PMID: 37056913
- Ciucci, M. R., Grant, L. M., Rajamanickam, E. S., Hilby, B. L., Blue, K. V., Jones, C. A., et al. (2013). Early identification and treatment of communication and swallowing deficits in Parkinson disease. *Semin. Speech Lang* 34, 185–202. doi: 10.1055/s-0033-1358367
- Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., et al. (2009). Pearson correlation coefficient. *Noise Reduct. Speech Proc.* 14, 1–4. doi: 10.1007/978-3-642-00296-0_5
- Gommers, R., Virtanen, P., Haberland, M., Burovski, E., Weckesser, W., Reddy, T. T., et al. "scipy/scipy: SciPy 1.11.4", Zenodo. (2024) [Online]. Available at: <https://zenodo.org/records/10155614> (Accessed July 13, 2024).
- Hossain, M. A., and Amenta, F. (2024). Machine learning-based classification of Parkinson's disease patients using speech biomarkers. *J. Parkinsons Dis.* 14, 95–109. doi: 10.3233/JPD-230002
- Ji, W., Fu, Y., Zheng, H., and Li, Y. (2025). Multi-label speech feature selection for Parkinson's disease subtype recognition using graph model. *Comput. Biol. Med.* 185:109566. doi: 10.1016/j.compbiomed.2024.109566
- Kang, S. H., Lee, H. M., Seo, W. K., Kim, J. H., and Koh, S. B. (2016). The combined effect of REM sleep behavior disorder and hyposmia on cognition and motor phenotype in Parkinson's disease. *J. Neurol. Sci.* 368, 374–378. doi: 10.1016/j.jns.2016.07.057
- Kim, T. K. (2015). T test as a parametric statistic. *Korean J. Anesthesiol.* 68, 540–546. doi: 10.4097/kjae.2015.68.6.540
- Kouli, A., Torsney, K. M., and Kuan, W. L. (2018). Parkinson's disease: etiology, neuropathology, and pathogenesis. *Exon Publicat.* 18, 3–26. doi: 10.15586/codonpublications.parkinsonsdisease.2018.ch1
- Kramer, O., and Kramer, O. (2016). Scikit-learn. *Machine Learn. Evolut. Strat.* 128, 45–53. doi: 10.1007/978-3-319-33383-0_5
- Li, M. P., Liu, W. C., Wu, J. B., et al. (2023). Machine learning for the prediction of postoperative nosocomial pulmonary infection in patients with spinal cord injury. *Eur. Spine J.* 32, 3825–3835. doi: 10.1007/s00586-023-07772-8
- Liang, X. W., Zeng, F. Y., Li, H. M., Li, Y., Lin, Y., Cai, K., et al. (2023). Deep learning based two-dimensional ultrasound for follicle monitoring in infertility patients. *BIO Integr.* 4, 125–131. doi: 10.15212/bioi-2022-0024
- Little, M., McSharry, P., Hunter, E., Spielman, J., and Ramig, L. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *Nat. Prec.* 56, 1015–1022. doi: 10.1038/npre.2008.2298.1
- Little, M., McSharry, P., Roberts, S., et al. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nat. Prec.* 6:23. doi: 10.1038/npre.2007.326.1
- Mazziotta, M., and Pareto, A. (2022). Normalization methods for spatio-temporal analysis of environmental performance: revisiting the min–max method. *Environmetrics* 33:e2730. doi: 10.1002/env.2730
- McKnight, P. E., and Najab, J. (2010). Mann-Whitney U test. The Corsini encyclopedia of psychology, 1. doi: 10.1002/9780470479216.corpsy0524
- Meghraoui, D., Boudraa, B., Merazi-Meksen, T., et al. Parkinson's disease recognition by speech acoustic parameters classification. Modelling and Implementation of Complex Systems: Proceedings of the 4th International Symposium, MISC 2016, Constantine, Algeria. Cham: Springer International Publishing, (2016): 165–173.
- 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 Proc. Control* 66:102418. doi: 10.1016/j.bspc.2021.102418
- Pramanik, A., and Sarker, A. (2021). "Parkinson's disease detection from voice and speech data using machine learning" in Proceedings of international joint conference on advances in computational intelligence. eds. M. S. Uddin and J. C. Bansal (Singapore: Algorithms for Intelligent Systems. Springer).
- Quan, C., Ren, K., Luo, Z., Chen, Z., and Ling, Y. (2022). End-to-end deep learning approach for Parkinson's disease detection from speech signals. *Biocybern. Biomed. Eng.* 42, 556–574. doi: 10.1016/j.bbe.2022.04.002
- Ranstrom, J., and Cook, J. A. (2018). LASSO regression. *J. Brit. Surg.* 105:1348. doi: 10.1002/bjs.10895
- Saw, P. E. (2023). Voice series: interview with professor Dr. Xin Gao, Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences. *BIO Integr.* 4, 141–144. doi: 10.15212/bioi-2023-0011
- Suppa, A., Costantini, G., Ascì, F., di Leo, P., al-Wardat, M. S., di Lazzaro, G., et al. (2022). Voice in Parkinson's disease: a machine learning study. *Front. Neurol.* 13:831428. doi: 10.3389/fneur.2022.831428
- Teixeira, J. P., and Fernandes, P. O. (2014). Jitter, shimmer and HNR classification within gender, tones and vowels in healthy voices. *Proc. Technol.* 16, 1228–1237. doi: 10.1016/j.protcy.2014.10.138
- Ul-Haq, A., Li, J. P., Agbley, B. L. Y., Mawuli, C. B., Ali, Z., Nazir, S., et al. (2022). A survey of deep learning techniques based Parkinson's disease recognition methods employing clinical data. *Expert Syst. Appl.* 208:118045. doi: 10.1016/j.eswa.2022.118045
- Zheng, T., Xinheng, W., Xiaolan, P., et al. (2024). Survey of neurocognitive disorder detection methods based on speech, visual, and virtual reality technologies. *Virtual Real. Intell. Hardw.* 6, 421–472. doi: 10.1016/j.vrih.2024.08.001

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

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.