International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Patient Level Intelligent Diagnosis of Parkinson Disease Using Structured Audio Digital Features and Machine Learning

Haoyu Hu[1], Zhang Qing[2], Rongrong Cheng[3], Wu Ling*
Anhui University of Finance and Economics
No. 962 Caoshan Road, Bengbu City, Anhui Province, China 233030
*Corresponding author: Wu Ling

Abstract**—*Parkinson disease is a progressive neurodegenerative disorder in which speech impairment often appears before severe functional decline. This study develops a patient level intelligent diagnosis framework based on structured audio digital features extracted from sustained vowel phonation. The dataset contains 756 recordings collected from 188 patients with Parkinson disease and 64 healthy controls, with three repeated phonations per subject. To reduce subject leakage and align the prediction target with practical screening scenarios, repeated recordings were aggregated at the patient level, yielding 252 modeling units and 753 predictors after excluding the subject identifier and class label. The feature space integrates baseline perturbation descriptors, intensity parameters, formant frequencies, bandwidth parameters, vocal fold indicators, Mel frequency cepstral coefficients, wavelet domain features, and tunable Q factor wavelet transform descriptors. Support vector machine, multilayer perceptron, random forest, and XGBoost were compared under the same evaluation protocol. Random forest achieved the best overall thresholded performance, with an accuracy of 0.8333, recall of 0.9681, F1 score of 0.8966, and AUC of 0.8373. The findings indicate that structured speech features can support lightweight and interpretable auxiliary screening for Parkinson disease.*

Keywords**—** *Audio digital features: intelligent diagnosis: machine learning: Parkinson disease: speech analysis: TQWT.*

I. INTRODUCTION

Parkinson disease is a progressive neurodegenerative disorder with substantial motor and non-motor burden, and timely recognition is important for treatment planning, follow-up management, and remote care [1].

Speech provides an attractive signal source for Parkinson disease screening because it is inexpensive, non-invasive, and suitable for repeated acquisition outside specialist settings. Prior telediagnosis studies have shown that dysphonia measurements can support remote discrimination between Parkinson disease and healthy controls [2], [4], [6].

The traditional acoustic-machine-learning route combines perturbation measures, nonlinear recurrence descriptors, cepstral statistics, and time-frequency features. Classical work highlighted the value of nonlinear voice descriptors such as RPDE, DFA, and PPE [3], while more recent studies extended this line using structured vocal features and tunable Q-factor wavelet transform descriptors [5].

However, three methodological problems remain common. First, repeated recordings are often split at the record level, which may inflate performance through subject leakage. Second, high-dimensional structured acoustic features introduce substantial redundancy relative to the available sample size. Third, many studies emphasize headline accuracy without sufficient analysis of threshold-independent discrimination and feature contribution.

To address these issues, this paper develops a patient-level intelligent diagnosis model based on structured audio-digital information. The main contributions are fourfold: patient-level aggregation to reduce subject leakage, multi-block acoustic feature integration, a unified preprocessing and model comparison pipeline, and feature-level interpretation of discriminative acoustic patterns.

II. MATERIALS AND METHODS

A. Data Source and Sample Construction

The data were collected by the Department of Neurology, Istanbul University Cerrahpasa Faculty of Medicine. The cohort contains 188 patients with Parkinson disease and 64 healthy controls. The Parkinson disease group includes 107 males and 81 females aged 33 to 87 years, with a mean age of 65.1 ± 10.9 years. The healthy control group includes 23 males and 41 females aged 41 to 82 years, with a mean age of 61.1 ± 8.9 years. Voice acquisition was performed at 44.1 kHz after clinical examination, and each participant produced the sustained vowel a three times. The dataset and feature grouping correspond to the structured speech benchmark introduced in the TQWT comparison study [5].

The file contains 756 recordings. After excluding the subject identifier from the predictor space and aggregating the three repeated recordings of each subject by feature-wise averaging, the final patient-level matrix contains 252 subjects and 753 predictors. Because repeated phonations from the same person share speaker-specific information, record-level random splitting would create leakage between training and test folds. Patient-level aggregation therefore provides a more conservative and clinically meaningful estimate of generalization.

B. Feature System

The workbook divides the feature space into several interpretable blocks. Baseline features include perturbation and nonlinear descriptors such as PPE, DFA, RPDE, pulse counts, jitter, shimmer, and harmonicity-related measures. Intensity parameters summarize vocal energy control. Formant frequency and bandwidth features characterize vocal-tract resonance behavior. Vocal-fold descriptors capture glottal and harmonic structure. Mel frequency cepstral coefficient statistics represent spectral envelopes together with first-order and second-order temporal dynamics. Wavelet and tunable Q-factor wavelet transform blocks provide multiscale energy, entropy, and Teager-Kaiser energy operator based nonlinear descriptors [3], [5], [6].

TABLE I. Dataset Summary and Modeling Unit

Item	Value
Institution	Istanbul Univ. Cerrahpasa
Task	Sustained vowel a, 3 repeats
Sampling rate	44.1 kHz
Records	756
Modeling units	252
PD subjects	188
Healthy controls	64
Predictors excl. ID and class	753
Missing values	0

Fig. 1. Overall research pipeline from raw recordings to patient-level model evaluation.

This design preserves both physiological interpretability and broad coverage of perturbation, source, resonance, cepstral, and nonlinear time-frequency information. In total, the feature system is intentionally overcomplete, making dimensionality control essential for stable patient-level modeling.

C. Data Preprocessing, Dimensionality Reduction, and Leakage Control

The preprocessing pipeline consisted of numerical parsing, missing-value inspection, model-specific scaling, and dimensionality reduction. After patient-level aggregation, no missing values remained in the matrix, but median imputation was still kept inside each model pipeline as a conservative safeguard. Standardization was applied for SVM and MLP because these classifiers are sensitive to heterogeneous feature scales and covariance structure.

Principal component analysis was used for the SVM and MLP branches to suppress redundancy and stabilize optimization in the high-dimensional setting. Thirty principal components were retained, explaining 76.26% of the standardized variance; the first 10 components explained 56.38%. Aggregation was completed before any cross-validation split, ensuring that each fold contained unique subjects rather than repeated recordings from the same speaker.

D. Model Construction and Evaluation Metrics

Four classifiers were selected to represent distinct modeling assumptions for structured acoustic diagnosis. SVM served as a nonlinear baseline in a compressed feature space; MLP captured nonlinear interactions among latent factors; random forest offered robust ensemble learning on tabular acoustic descriptors; and XGBoost represented gradient-boosting behavior on the same task. All models were evaluated under the same patient-level stratified three-fold cross-validation protocol.

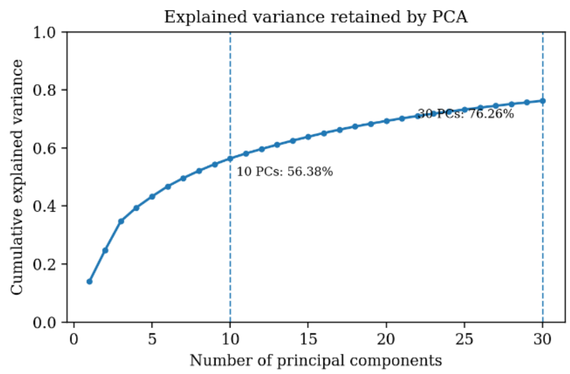

Fig. 2. Cumulative variance retained by PCA. The first 10 principal components explain 56.38%, and the first 30 explain 76.26%.

Accuracy, precision, recall, F1 score, and AUC were used as evaluation metrics. In Parkinson disease auxiliary screening, recall is particularly important because false negatives correspond to missed patients, while F1 score balances detection ability and positive predictive reliability. AUC assesses ranking quality independently of a fixed threshold and is therefore informative under class imbalance.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

$$\text{F1 score} = 2 \times \text{Precision} \times \text{Recall} / (\text{Precision} + \text{Recall}) \quad (4)$$

III. Results

At the patient level, the final matrix contains 188 Parkinson disease cases and 64 healthy controls, corresponding to a class ratio of approximately 2.94:1. The aggregated cohort includes 130 male subjects and 122 female subjects. No missing values remained after aggregation.

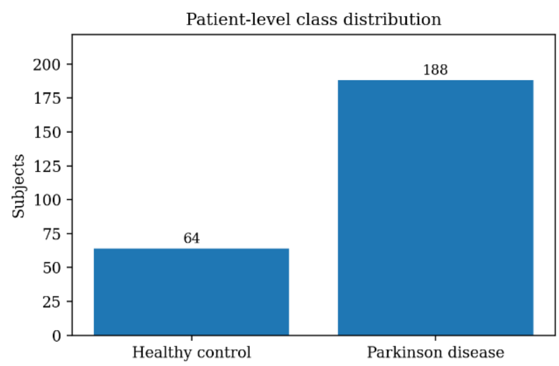

Fig. 3. Patient-level class distribution after subject aggregation.

Representative descriptive statistics show that Parkinson disease subjects tend to have higher DFA and RPDE values but lower mean intensity and lower first formant frequency than healthy controls. This pattern is consistent with the view that Parkinson disease affects both nonlinear control stability and vocal energy organization during sustained phonation.

TABLE II. Descriptive Statistics of Representative Features by Class

Feature	Healthy controls (mean ± SD)	PD (mean ± SD)
PPE	0.767 ± 0.110	0.739 ± 0.123
DFA	0.664 ± 0.062	0.713 ± 0.063
RPDE	0.431 ± 0.122	0.509 ± 0.123
numPulses	372.245 ± 104.946	307.539 ± 82.985
meanIntensity	77.072 ± 2.904	72.950 ± 7.524
f1	716.143 ± 119.236	623.696 ± 123.926

A. Patient-Level Model Performance Comparison

The four classifiers were benchmarked under the same patient-level cross-validation protocol. Random forest achieved the best overall thresholded performance, with the highest accuracy and F1 score. XGBoost obtained the highest AUC, suggesting slightly stronger ranking behavior across thresholds, whereas MLP delivered the most balanced trade-off between sensitivity and specificity.

TABLE III. Patient-Level Model Performance Comparison

Model	Accuracy	Precision	Recall	F1 score	AUC
SVM	0.7659	0.8449	0.8404	0.8427	0.7874
MLP	0.8214	0.8705	0.8936	0.8819	0.8313
RF	0.8333	0.8349	0.9681	0.8966	0.8373
XGBoost	0.7976	0.8216	0.9309	0.8728	0.8408

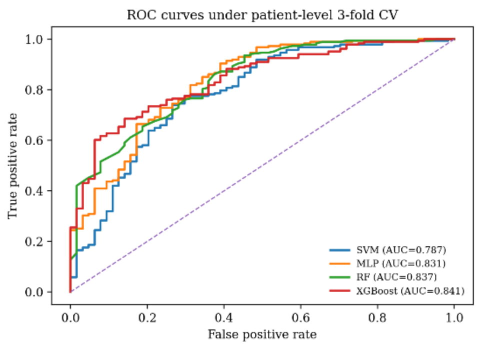

Fig. 4. ROC curves of the four classifiers under patient-level stratified three-fold cross-validation, generated from out-of-fold predicted probabilities.

B. Feature Contribution and Acoustic Interpretation

To interpret the discriminative structure of the acoustic space, a random forest model was refit on the complete aggregated matrix and impurity-based feature importances were extracted. The top-ranked variables were dominated by dynamic Mel frequency cepstral coefficient statistics and tunable Q-factor wavelet transform entropy or nonlinear-energy descriptors. This indicates that Parkinson disease related dysphonia is represented less by static spectral position alone than by instability in spectral dynamics and subband energy organization.

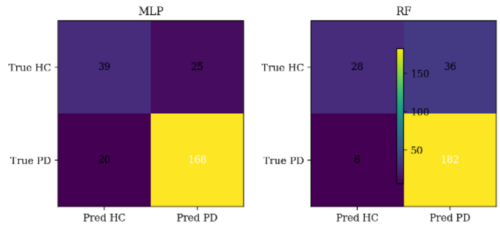

Fig. 5. Confusion matrices of the two strongest thresholded classifiers, highlighting the trade-off between missed PD cases and false alarms.

From an acoustic perspective, delta and delta-delta MFCC statistics represent changes and accelerations in short-term spectral shape and are therefore sensitive to unstable articulatory timing. TQWT entropy and TKEO-related descriptors capture subband irregularity and nonlinear energy bursts, which are consistent with impaired phonatory control in Parkinson disease [3], [5], [6]. The acoustic-meaning labels reported in Table IV should therefore be interpreted as signal-processing summaries rather than as standalone causal claims.

TABLE IV. Top Important Features and Their Acoustic Meanings

Feature	Importance	Acoustic meaning
std_8th_delta_delta	0.0255	2nd order MFCC dynamics
std_8th_delta	0.0120	1st order MFCC dynamics
std_delta_delta_log_energy	0.0114	2nd order energy change
std_9th_delta_delta	0.0113	Higher order cepstral acceleration
std_delta_log_energy	0.0107	1st order energy variation
tqwt_entropy_shannon_dec_36	0.0107	High order TQWT complexity
tqwt_TKEO_mean_dec_12	0.0105	Nonlinear subband energy
std_6th_delta	0.0104	Mid order cepstral transitions
tqwt_entropy_log_dec_12	0.0092	Subband log entropy
std_9th_delta	0.0089	Higher order transitions

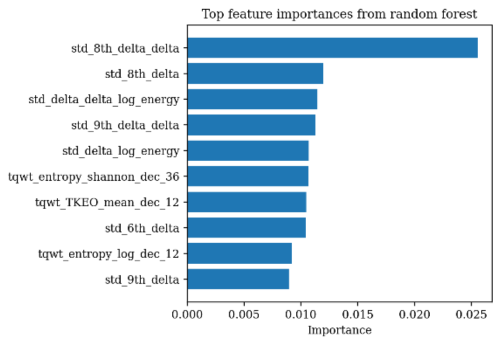

Fig. 6. Top feature importances estimated from the random forest model fitted to the aggregated dataset.

IV. Discussion

The relative behavior of the models is consistent with their structural assumptions. Random forest is robust to nonlinear interactions, scale heterogeneity, and mixed feature blocks, which is advantageous in high-dimensional acoustic diagnosis. XGBoost provides strong ranking behavior across thresholds, whereas MLP benefits from PCA-compressed latent representations. SVM remains competitive but is more constrained by a single kernel boundary in the compressed space.

The patient-level design is a substantive methodological improvement because it better reflects real-world deployment, where diagnosis is made for a person rather than for an isolated recording. This is also closer to the way remote dysphonia screening is framed in prior telediagnosis studies [2], [4], [6]. The proposed framework remains lightweight and interpretable, which is valuable for telemedicine and resource-constrained screening settings.

Several limitations remain. The sample size is modest relative to the dimensionality of the feature space. The healthy-control group is substantially smaller than the Parkinson disease group. Only sustained-vowel phonation was analyzed, and no independent external validation cohort was available. Future work should therefore consider nested cross-validation, multimodal fusion, connected speech tasks, and external clinical validation.

V. Conclusion

This study developed an intelligent patient-level diagnosis framework for Parkinson disease using structured audio-digital features extracted from sustained-vowel recordings. By aggregating repeated phonations at the subject level, the design reduced subject leakage and aligned evaluation with the actual screening target.

The experimental results show that random forest achieved the best overall thresholded classification performance, XGBoost obtained the highest AUC, and dynamic MFCC and TQWT-based descriptors were especially important for distinguishing Parkinson disease from healthy controls. The framework should be regarded as a technically grounded auxiliary screening baseline.

Acknowledgment

This study was supported by the Innovation and Entrepreneurship Program of Anhui University of Finance and Economics (Project No. 202410378133).

References

[1] J. Massano and K. P. Bhatia, “Clinical approach to Parkinson's disease: Features, diagnosis, and principles of management,” *Cold Spring Harbor Perspectives in Medicine*, vol. 2, no. 6, art. a008870, 2012.

[2] C. O. Sakar and O. Kursun, “Telediagnosis of Parkinson's disease using measurements of dysphonia,” *Journal of Medical Systems*, vol. 34, no. 4, pp. 591-599, 2010.

[3] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *BioMedical Engineering OnLine*, vol. 6, art. 23, 2007.

[4] B. Erdogdu Sakar, G. Serbes, and C. O. Sakar, “Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease,” *PLoS ONE*, vol. 12, no. 8, art. e0182428, 2017.

[5] C. O. Sakar, G. Serbes, A. Gunduz, H. C. Tunc, H. Nizam, B. E. Sakar, M. E. Tutuncu, T. Aydin, H. Isenkul, and H. Apaydin, “A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform,” *Applied Soft Computing*, vol. 74, pp. 255-263, 2019.

[6] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, “Suitability of dysphonia measurements for telemonitoring of Parkinson's disease,” *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 4, pp. 1015-1022, 2009.