International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(1s) 2026 pp. 1-9
© Cerebration Science Publishing
https://doi.org/10.70917/ijcisim-2026-1965
Article

A Comparative Study of Acoustic and Clinical Speech Feature Vectors for Parkinson's Disease Detection with a Multilayer Perceptron Classifier

Ashwini D. Bhople[1], **Avinash Kapse** [2], **Pravin A. Kharat**[3]

[1]Padmashri Dr. V. B. Kolte College of Engineering Malkapur, India (Research Scholar, Department of Computer Science and Engineering, Anuradha College of Engineering and technology, Chikhli, India), ashwinibhople@gmail.com
[2]Professor, Department of Computer Science & Engineering, Anuradha College of Engineering & Technology, Chikhli, India, askapse@gmail.com
[3]Professor, Department of Computer Science & Endineering, Padmashri Dr. V. B. Kolte College of Engineering, Malkapur, India, pravinakharat82@gmail.com

Abstract: Parkinson's disease (PD) is an advancing neurodegenerative disease that can have speech loss during preliminary stages of the disease, thus allowing detection of the disease using non-invasive and convenient analysis of the voice. In this paper, a comparative analysis of acoustic and clinical speech feature vectors as a method of detection of Parkinson disease is suggested using a multilayer perceptron (MLP) classifier. It evaluates the five sets of features: the traditional Acoustic-Spectral Speech Features (CASSF), statistically pooled Acoustic-Dynamic Speech Features (SP-ADSF), Hybrid Correlation-Nonlinear Acoustic Features (HCNAF) and the Hybrid Acoustic-Dysphonia-Nonlinear Speech Feature set (HADNSF) as well as the state of the art, widely used clinical benchmark of pathological speech analysis with low computational resources the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS). The experimental results suggest that the MLP classifier can perform optimally when it is trained using SP-ADSF features. The suggested model attains an accuracy of 94.59, a sensitivity of 92.50 and an area under the ROC curve of 0.9799 when using this representation, which reflects strong discriminative abilities of the model to diagnose the Parkinson disease. On the other hand, eGeMAPS and the suggested HADNSF have comparably poor performance, which may be explained by specific features of the sets, the high feature dimensionality, and sensitivity of dysphonia and nonlinear features to processing at the segment level. However, these sets of features are still clinically relevant because their explanations of voice disorders can be read and interpreted and have a pathology-based nature. In general, the findings indicate that performance-based acoustic feature representation and clinically based feature sets are complementary rather competitive. This work highlights the need to have the right balance between discriminative performance and clinical applicability in reliable speech-based systems of Parkinson disease detection.

Keywords: Parkinson's Disease Detection; Speech Signal Analysis; Multilayer Perceptron (MLP); Acoustic Feature Extraction; Voice Biomarkers

1. Introduction

Parkinson's disease (PD) is a neurodegenerative disorder that is progressive and mainly involves the motor system, which is characterized by bradykinesia, rigidity, and tremor (Avazzadeh et al., 2021). Besides these motor disorders, speech impairment is also one of the most prevalent and the initial non-motor symptoms of the disease. Parkinsonian speech, or hypokinetic dysarthria, is characterized by a loss of voice loudness, monotonic speech, poor articulation and an increase in voice variability (Chenausky et al., 2011) (Brabenec et al., 2017). The causes of these changes are a failure of the neuromotor control of the subsystems of respiration, phonation, and articulation. Among the most commonly impaired behavioural outputs in PD is speech because its production requires the coordination of numerous motor activities, so even minimal changes in the functioning of the nervous system cause acoustic differences to be identified (Sapir et al., 2008) (Martin et al., 2018).

Although significant progress has been achieved in the scientific knowledge of PD, conventional clinical diagnosis and measurement can majorly be associated with subjective rating scales, including the Unified Parkinson's Disease Rating Scale (UPDRS) (Van Hilten et al., 1994). Those scales require the assessment of experts and may be time-consuming. Though these tests are clinically important, they cannot identify common and subtle speech impairments and cannot be used in longitudinal studies or in the assessment of intervention influence because of clinical conditions of conducting such tests(Regnault et al., 2019). In this respect, speech-based analysis has provided a potential, noninvasive, cost-efficient, and objective method for detecting and evaluating PD. The further development of digital signal analysis and machine learning has contributed to the extraction of discriminative acoustic features of the voice characteristics of patients with PD.

The literature includes Numerous studies have been done on speech characteristics which can be applied in detection of PD. They are traditional acoustic and spectral features like Mel-Frequency Cepstral Coefficients (MFCCs), energy-related features, and prosodic features that are characteristic of articulatory and phonatory qualities of the spoken language (Daoudi et al., 2022) (Liu et al., 2023). Besides, a collection of acoustic measurements, which have been shown to be correlated with the symptoms of Parkinsonian speech, comprise dysphonia-related features, including jitter, shimmer, and harmonic-to-noise ratio, indicative of vocal fold instability and voice breathiness(Chiaramonte & Bonfiglio, 2020). In addition, nonlinear and dynamic features of speech contents have been suggested to explain the anomalies that are caused by the phonation and/or neuromotor control. To examine pathological speech in an objective manner, the extended eGeMAPS has since been put forward as a clinically-oriented feature set used to capture voice quality, pitch variability, and formant-related features and is often used as a baseline on the analysis of Parkinson's disease voice(Eyben et al., 2016)).

The current research has some limitations even though the field has improved. Mostly, the current methodologies have concentrated on a single form of speech feature without giving a full comparison between the performance-based acoustic representations and their clinical-driven counterparts. Moreover, the relationship between the classifiers and features is often not taken into account, because the unique properties of features might affect different machine learning models in a certain manner. Also, even though standardized sets of features like eGeMAPS do add to the clinical interpretability and robustness, data-driven acoustic representations tend to yield better discriminative accuracy in controlled experimental settings. As such, it is imperative to have a detailed analysis that incorporates the classification performance and clinical relevance to help understand the strengths and limitations of the various speech feature representations in detecting the presence of PD.

Based on these observations, the present research paper is a comparative analysis of the performance of five acoustic and clinical speech feature vectors when used with multilayer perceptron (MLP) classifier to detect Parkinson disease. The feature representations being studied are: Conventional Acoustic-Spectral Speech Features (CASSF), Statistically Pooled Acoustic-Dynamic Speech Features (SP-ADSF), Hybrid Correlation-Nonlinear Acoustic Features (HCNAF), the newly proposed Hybrid Acoustic-Dysphonia-Nonlinear features (HADNSF) as well as the eGeMAPS feature set which is clinically verified. The study will propose an explanation of the trade-off between the discriminative performance and the clinical relevance of speech-based Parkinson's disease recognition by assessing such feature vectors in a common experimental context.

The subsequent sections of this paper is organised as follows. Section II describes the speech dataset and methods of extracting features applied. In Section III we then provide our methodology in classifying and our evaluation protocol. Section IV examines the empirical findings and comparative analysis. Section V then conclusively closes the paper and expresses directions in future work.

2. Dataset And Preprocessing

The dataset used in this paper is the MDVR-KCL dataset, a free speech corpus of Parkinson disease speech that is available on Zenodo (Jaeger et al., 2019). The specific dataset was gathered in September 2017 at King's College London (KCL) Hospital and has voice samples of 37 volunteers, among whom 16 had PD and 21 were healthy control (HC) volunteers. The speech recordings were based on a read-paragraph task to make the recordings similar across speakers. All recordings were annotated with the help of clinical experts on the basis of the known clinical rating scales, including the Hoehn and Yahr (H&Y) scale, UPDRS II Part 5, or UPDRS III Part 18 which depict the severity of the disease and motor deficits (Marinus, 2004).

The voice recordings were carried out within a clinical examination room in the setting that very much resembles the realistic conditions of the phone call and the participants were placed with a smartphone close to the mouths. The signal of speech was recorded through a Motorola Moto G4 smartphone and saved in an uncompressed

16-bit 44.1 kHz waveform (WAV) source. Since the recordings have comparatively clean signals, with the microphone being 10 centimeters away as the speaker in a controlled setting, they can be used as an invaluable tool in acoustic or pathological sound analysis.

Before the feature extraction process, all the audio recordings were converted to mono and resampled to a standard sampling rate of 16 kHz as a standard. First, a selective long-silence removal method was used in order to minimize the presence of non-informative parts of speech and retain the original properties of speech. This was followed by computing the short-time Root Mean Square (RMS) energy per frame, and frames with an energy level lower than a pre-set threshold were labeled as silence(Tan & Lindberg, 2010). The frame time was about 30 ms. Instead of deleting all silent frames, silence segments longer than 500 ms were deleted. Morphological smoothing was used on the silence mask in order to prevent the fragmentation of speech regions by the silence mask.

After eliminating silence, all of the processed speech records were divided into ten equal, non-overlapping segments of equal length. Since the dataset contains 37 recordings by the participants, this segmentation method produced 370 speech sections. It requires a segment-based analysis, which has the benefits of providing the effective volume of training samples and the representation of the speaker-specific-specific speech dynamics with short-period dynamics. The class designation of every segment is determined by the associated speaker (PD or HC). Subject-wise data partitioning was used to make sure that all segments of a single subject were retained in isolation, either during training or testing, to prevent leakage of data and ensure the evaluation proceeded without subject dependence.

The preprocessing pipeline, shown in Fig. 1 involves the conversion of audio signals to mono, resampling to 16 kHz, optionally owing out long silent passages and dividing each recording into ten equal length excerpts. It is interesting to notice that no data augmentation methods were used. These preprocessing stages provide the basis of a uniform and repeatable platform on which later feature extraction and classification is to be applied.

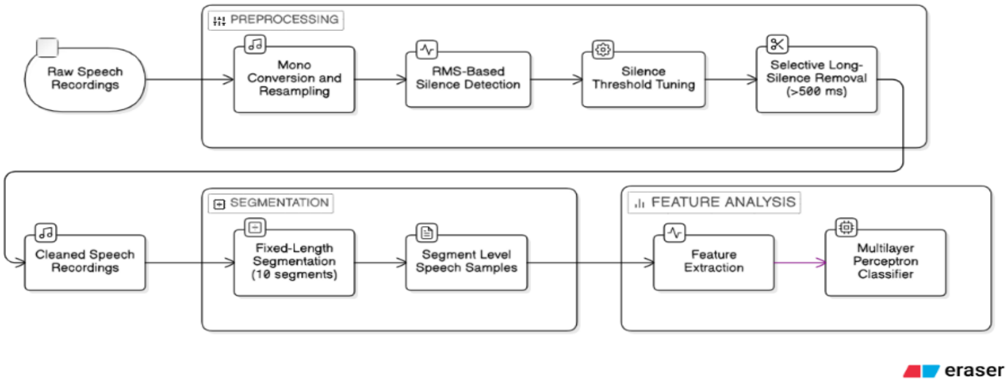

Fig. 1. Block diagram of the speech processing pipeline including preprocessing, segmentation, feature extraction, and classification.

3. Feature Extraction

This section gives a detailed account of the process used to extract speech features that were used in this study. To identify different features of Parkinsonian speech, we used five complementary feature vectors per speech segment. The representations of these features are created in a manner that involves a wide range of acoustical spectral, dynamic, prosodic, clinical (dysphonia-related), correlation-based, and nonlinear attributes of speech signals. Although methods varied widely to produce these features, they were all determined at the segment level and statistical pooled, giving rise to representations of equal dimension, regardless of the properties of the separate samples. This consistency helps them to be directly fed into the multilayer perceptron (MLP) classifier. The following discussion covers the feature vectors related to the extracted features and dimensionality.

The first, Conventional Acoustic Speech Features (CASSF), is the standard baseline acoustic description of speech. Time-domain descriptors include Zero-Crossing rate and RMS energy, whereas spectral descriptors include spectral centroid, spectral bandwidth and spectral roll-off. Also, cepstral coefficients (MFCC 1-13) and harmonic chroma feature (Chroma 1-12) are added. The CASSF includes 30 features and can be used as a reference point to test more sophisticated sets of features (Garg et al., 2025).

The Statistically Pooled Acoustic-Dynamic Speech Features (SP-ADSF) is a new type of feature vector based on the time dynamics of speech obtained by means of statistical pooling. This vector uses the static MFCC features and the first (Δ) and second ($\Delta\Delta$) derivatives, which are described using the mean and standard deviation statistics. Also, the chromatic features, pitch (F0), root mean square (RMS) energy, and the Zero Crossing Rate are represented by means of statistical pooling (Yaman et al., 2019). The SP-ADSF feature vector contains 108 features, which put an accent on dynamic articulation and prosodic variability.

The third feature vector, which is known as the Hybrid Correlation Nonlinear Acoustic Features (HCNAF), is created to describe how features relate to each other and the nonlinear dynamics of speech. It consists of MFCC-pitch coupling, energy-pitch correlations and pairwise inter-MFCC correlations as linear features. The nonlinear properties are Hurst exponent, sample entropy, permutation entropy and detrended fluctuation analysis (Sun et al., 2025). Also, quantification analysis, which measures recurrence, is included. HCNAF comprises 99 features that are intended to represent the complicated dependence that occurs due to impaired neuromotor control. Hybrid Acoustic–Dysphonia-Nonlinear Speech Features (HADNSF) is the fourth representation vector, which has been proposed in the given research as a hybrid representation combining acoustic, dysphonia-associated, and nonlinear features. It contains MFCCs and temporal derivatives of these, energy and temporal, spectral, chroma, pathologically interesting dysphonia-related, and a small number of nonlinear features calculated with Praat. HADNSF is a set of 118 features and it focuses on nonlinear speech behaviour and clinical interpretability. The fifth feature set is the eGeMAPS (Eyben et al., 2016) which was identified as a clinically validated gold standard in pathological speech analysis and comprises of the following features: pitch-related features, formant and spectral features, voice quality measures, and temporal energy measures, making a total of 88 measures. eGeMAPS was calculated on each of the features with the openSMILE toolkit and used as a reference point of comparison (Eyben et al., 2013).

Table I: Summary of Feature Vectors Used in This Study

Feature Set	Domain Coverage	Description	Feature Count
CASSF	Acoustic–Spectral	Baseline acoustic features capturing basic time-domain and frequency-domain characteristics of speech	30
SP-ADSF	Acoustic–Dynamic	MFCC-based static and dynamic features with statistical pooling (mean and standard deviation) to model spectral and temporal speech dynamics	108
HCNAF	Correlation–Nonlinear	Correlation-based MFCC interactions combined with nonlinear dynamical speech descriptors	99
HADNSF	Acoustic–Dysphonia–Nonlinear	Integration of acoustic features with clinically motivated dysphonia and nonlinear speech measures	118
eGeMAPS	Clinical voice features	Clinically validated standardized feature set for pathological speech analysis	88

The five feature vectors that we have used in our study are outlined in Table I, with the domain coverage, size, and the role taken by each of these parameters in our research framework. It should also be pointed out that we did not perform any feature selection or dimensionality reduction since our main goal is to directly compare the discriminative ability of different feature representations.

4. Classification Methodology

This paper used a Multilayer Perceptron (MLP) classifier as the main detecting classifier of the Parkinson disease through speech features. MLPs have found wide applications in the modeling of nonlinear decision boundaries in pattern recognition and biomedical signal analysis, since they have the ability to model complex feature interactions (Gori & Scarselli, 1998). MLPs provide more representational powers than linear classifiers and also have a reasonable cost of computation(Collobert and Bengio, 2004).

The five feature vectors described in Section III were used to model each segment of the speech. The sets of features were tested separately to have fair comparison between the representations of different dimensional sizes. Before classification, the features with static values were filtered to avoid numerical instability and to avoid features of the inputs that did not add importance to the learning process.

All the features were normalized by the z-score normalization technique which normalizes each feature to a mean of zero and a variance of one. Machine learning pipelines also include a known preprocessing step called standardization which is known to increase numerical stability and convergence. Notably, the normalization parameters were calculated only using the training data in each cross-validation fold and then used in the test samples to avoid information leakage (Demircioğlu, 2024).

Multi-Layer Perceptron (MLP) classifier was written with the help of scikit-learn library, was fitted with two hidden layers, of 128 and 64 neurons, respectively. The activation function used in the hidden layers was the Rectified Linear Unit (ReLU) to add nonlinearity, whereas binary classification at the output layer was carried out by the Sigmoid activation function (Ramachandran et al., 2017). Model training was performed using the Adam optimization algorithm, which enables the model to fit the data by finding the right model weights and biases and uses an adaptive learning rate strategy to reach efficient convergence in a wide range of deep neural networks. The number of training iterations was set to a maximum of 500 and a fixed random seed was used to ensure reproducibility.

To perform an independent evaluation and thus reduce optimistic bias, we used a cross-validation strategy, namely GroupKFold (5 fold). In speech-based disease detection, it is important that the samples are produced by the same speaker and used in testing since it avoids subject-level leakage between training and testing (Jeancolas et al., 2017). All folds were expected to make predictions and posterior probabilities and these were added together to form performance measures on the whole dataset.

The classification performance was measured using standard classification measures, such as accuracy, specificity, F1-score, precision, recall (sensitivity) and the area under the receiver operating characteristic curve (ROC-AUC). These are normally used as a measure of medical decision-support systems, both in general terms and in terms of specific classes(Medic et al., 2019).

Along with the MLP classifier, a Random Forest (RF) model was also used to determine the importance of the features and compare them by using ROC evaluation. Random Forests are the most common ensemble tree-based models that are used to estimate the relevance of features because these models are both stable and interpretable (Alhams et al., 2024). However, the presented results of all primary classification are based on the MLP classifier.

5. Experimental Setup And Evaluation Metrics

The main aim of this experimental research is to compare the levels of performance of CASSF, SP-ADSF, HCNAF, HADNSF, and eGeMAPS in the process of identifying a case of Parkinson disease in a consistent and standardized classification system. In order to establish a fair and reproducible comparison, evaluation of all feature sets was done using a similar classifier configuration, preprocessing strategy and validation criterion. Following preprocessing and segmentation, 370 speech units were obtained out of 37subjects. Subject-wise evaluation approach was employed in order to reduce the impact of Optimism bias through level of segment correlation. Specifically, a five-fold GroupKFold cross-validation process with subject IDs as grouping variables was performed. This ensured that each piece of a subject was allocated to either training or test set in each fold, thereby allowing a speaker-independent assessment. It used z-score normalization to perform feature normalization in every fold of cross-validation. The normalization parameters were estimated using the training data and these were then applied to the test data to avoid information leaking. The Multilayer Perceptron (MLP) outlined in Section IV was trained using four folds and tested using the rest of the fold. This was repeated until all of the folds had been tested. The overall performance metrics were then calculated using the predictions and posterior probabilities of all folds in order to obtain overall performance indicators of the complete dataset. The rate of accurate classification of the samples was used to determine the accuracy. Precision is a metric of the accuracy of positive (PD) prediction, whereas recall (sensitivity) is a metric of the effectiveness of the classifier in detecting all the relevant cases. Specificity is an evaluation of the ability of the classifier to identify healthy control subjects. The F1 -score is a balance measure of performance by a classification algorithm and should be used to measure performance when class imbalance occurs. Along with these two measures, the ROC-AUC was also used to evaluate the discriminative ability of the classifier at different decision levels. ROC curves have been created to demonstrate improved performance levels whereas precision-recall curves have been applied in other situations. Class-specific patterns of prediction were explained using the confusion matrices.

6. Results

The findings of the experiment under MLP classifier are reported in this section and were implemented with the fivefold subject-wise GroupKFold cross-validation. All of the results are given at the segment level and averaged within folds, ensuring a speaker independent analysis. The results of the MLP classifier in classifying the five analyzed feature representations including CASSF, SP-ADSF, HCNAF, HADNSF and eGeMAPS are tabulated in Table II. The metrics are used to evaluate the performance as the accuracy, precision, recall (sensitivity), specificity, F1-score and ROC-AUC.

The CASSF proved to be accurate at 0.7324 and F1-score 0.6667 with only the conventional acoustic-spectral features used which implies that it has limited discriminative ability. Conversely, the HCNAF demonstrated more favorable results as it obtained the accuracy of 0.7811 and the F1-score of 0.7429. On the other hand, the performance of the HADNSF was relatively poor with a 0.6838 accuracy and F1-score of 0.6164. The custom eGeMAPS that was motivated by clinical reasons achieved moderate performance in general, and its accuracy and ROC-AUC were 0.7770 and 0.8148, respectively.

Table II: MLP Classification Performance

Feature Set	Accuracy	Precision	Recall	F1-score	Specificity	ROC–AUC
CASSF	0.7324	0.7226	0.6188	0.6667	0.8190	0.7861
HCNAF	0.7811	0.7548	0.7312	0.7429	0.8190	0.8082
HADNSF	0.6838	0.6483	0.5875	0.6164	0.7571	0.7444
eGeMAPS	0.7770	0.7541	0.7188	0.7360	0.8214	0.8148
SP-ADSF	0.9459	0.9487	0.9250	0.9367	0.9619	0.9799

Among the considered sets of features, SP-ADSF showed the best performance with the highest values of a number of metrics: accuracy (0.9459), precision (0.9487), recall (0.9250), F1-score (0.9367), specificity (0.9619), and ROC-AUC (0.9799). Evaluation on the confusion matrix shows that SP-ADSF has a high true positive rate on the detection of the Parkinson disease, with a reasonable false positive rate. Additional evaluation based on ROC and precision-recall curves supports the fact that SP-ADSF is better across threshold independent. The importance of features analysis performed with the help of Random Forest classifier on the offered SP-ADSF feature and demonstrated in Figure 2 suggests that one of the major factors of interest is provided by prosodic and acoustic-dynamic features. Fundamental frequency, MFCC delta statistics, chroma feature set, RMS energy and zero-crossing rate were found to be significant features with mean and standard deviation. On the whole, the results indicate that statistically pooled acoustic and dynamic characteristics are more useful in distinguishing the presence of Parkinson’s disease in the context of our experimental study.

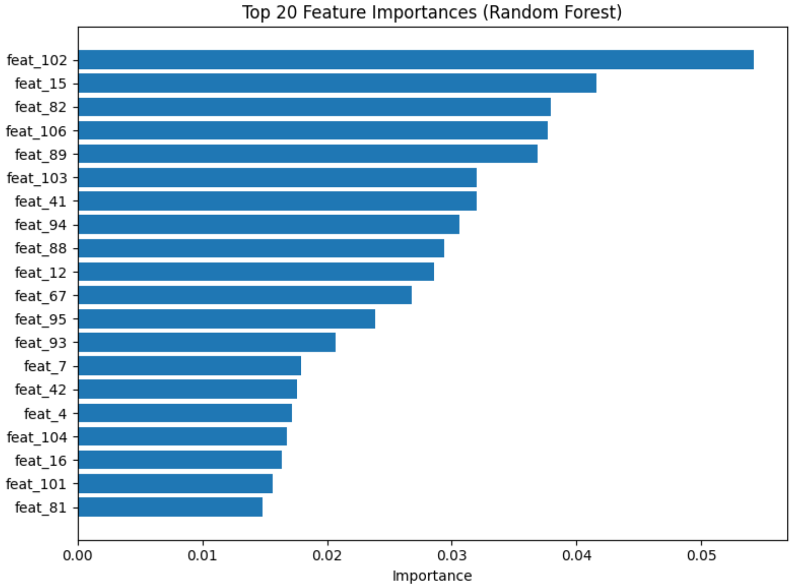

Fig 2: Feature Importance Profile for MLP-Based Parkinson's Disease Detection

7. Discussion And Conclusion

We performed a comparative study of five sets of features that described the characteristics of speech to identify Parkinson disease (PD) in a single Multilayer Perceptron (MLP) classification system. The experimental results indicate the importance of feature design towards improving the performance of the classification in detecting Parkinson disease using speech-based systems. Statistically Pooled Acoustic -Dynamic Speech Features (SP-ADSF) system proved to be more effective in numerous test measures. The findings suggest that the statistical pooling of MFCC-based spectral population, their derivatives or double derivatives, is an effective model of phonation variability due to Parkinsonian motor impairments. By comparison, the clinically-motivated feature sets (eGeMAPS) and hybrid representations (HADNSF) had seen a somewhat worse performance in the presented task, which shows that the effectiveness depends on the data, which is more effective in the given task when analyzing the data segments and in low-resource settings. The analysis involving feature importance displayed that prosodic features, dynamic MFCC statistics, and harmonic parameters are important in making the classification decisions. The results are consistent with the typical aspects of Parkinsonian speech including a lack of pitch variability and changed spectral effects. Although features that can be clinically interpreted are still useful in pathological speech analysis, our findings indicate that performance-based solutions of acoustic-dynamic features can perform better in a combination with neural network classifier. Finally, this paper has shown that statistically pooled acoustic dynamic characteristics are a good and consistent representation of the possibility of identifying a person with Parkinsonism disease by studying their speech. Future work will involve validating our framework on a larger dataset, subject-level fusion of decisions, and a combination of clinically interpretable features and performance-driven feature representation to enhance both robustness and clinical applicability.

References

1. Alhams, A., Abdelhadi, A., Badri, Y., Sassi, S., & Renno, J. (2024). Enhanced Bearing Fault Diagnosis Through Trees Ensemble Method and Feature Importance Analysis. *Journal of Vibration Engineering & Technologies*, *12*(S1), 109–125. https://doi.org/10.1007/s42417-024-01405-0
2. Avazzadeh, S., Baena, J. M., Keighron, C., Feller-Sanchez, Y., & Quinlan, L. R. (2021). Modelling Parkinson's Disease: iPSCs towards Better Understanding of Human Pathology. *Brain Sciences*, *11*(3), 373. https://doi.org/10.3390/brainsci11030373
3. Brabenec, L., Mekyska, J., Galaz, Z., & Rektorova, I. (2017). Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. *Journal of Neural Transmission*, *124*(3), 303–334. https://doi.org/10.1007/s00702-017-1676-0
4. Chenausky, K., Macauslan, J., & Goldhor, R. (2011). Acoustic Analysis of PD Speech. *Parkinson's Disease*, *2011*(1), 1–13. https://doi.org/10.4061/2011/435232

5. Chiaramonte, R., & Bonfiglio, M. (2020). Acoustic analysis of voice in Parkinson’s disease: a systematic review of voice disability and meta-analysis of studies. *Revista de Neurología*, *70*(11), 393. https://doi.org/10.33588/rn.7011.2019414
6. Collobert, R., & Bengio, S. (2004). *Links between perceptrons, MLPs and SVMs*. 23. https://doi.org/10.1145/1015330.1015415 Daoudi, K., Das, B., Tykalova, T., Klempir, J., & Rusz, J. (2022). Speech acoustic indices for differential diagnosis between Parkinson’s disease, multiple system atrophy and progressive supranuclear palsy. *NPJ Parkinson’s Disease*, *8*(1). https://doi.org/10.1038/s41531-022-00389-6
7. Demircioğlu, A. (2024). The effect of feature normalization methods in radiomics. *Insights into Imaging*, *15*(1). https://doi.org/10.1186/s13244-023-01575-7
8. Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., Andre, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., & Truong, K. P. (2016). The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing. *IEEE Transactions on Affective Computing*, *7*(2), 190–202. https://doi.org/10.1109/taffc.2015.2457417
9. Eyben, F., Weninger, F., Gross, F., & Schuller, B. (2013). *Recent developments in openSMILE, the munich open-source multimedia feature extractor*. 835–838. https://doi.org/10.1145/2502081.2502224 Garg, A., Aribi, Y., Althobaiti, T., & Bhowmik, T. (2025). An analysis of acoustic features for accented speech classification. *Egyptian Informatics Journal*, *31*, 100743. https://doi.org/10.1016/j.eij.2025.100743 Gori, M., & Scarselli, F. (1998). Are multilayer perceptrons adequate for pattern recognition and verification? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1121–1132. https://doi.org/10.1109/34.730549
10. Jaeger, H., Trivedi, D., & Stadtschnitzer, M. (n.d.). Mobile Device Voice Recordings at King’s College London (MDVR-KCL) from both early and advanced Parkinson’s disease patients and healthy controls [Dataset]. *In Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/https://doi.org/10.5281/zenodo.2867215 Jeancolas, L., Benali, H., Benkelfat, B.-E., Mangone, G., Corvol, J.-C., Vidailhet, M., Lehericy, S., & Petrovska-Delacretaz, D. (2017). *Automatic detection of early stages of Parkinson’s disease through acoustic voice analysis with mel-frequency cepstral coefficients*. *14*, 1–6. https://doi.org/10.1109/atsip.2017.8075567 Liu, Y., Reddy, M. K., Penttila, N., Ihalainen, T., Alku, P., & Rasanen, O. (2023). Automatic Assessment of Parkinson’s Disease Using Speech Representations of Phonation and Articulation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 242–255. https://doi.org/10.1109/taslp.2022.3212829 Marinus, J. (2004). A short scale for the assessment of motor impairments and disabilities in Parkinson’s disease: the SPES/SCOPA. *Journal of Neurology, Neurosurgery & Psychiatry*, *75*(3), 388–395. https://doi.org/10.1136/jnnp.2003.017509
11. Martin, S., Iturrate, I., Millán, J. D. R., Knight, R. T., & Pasley, B. N. (2018). Decoding Inner Speech Using Electrocorticography: Progress and Challenges Toward a Speech Prosthesis. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00422
12. Medic, G., Kosaner Kließ, M., Atallah, L., Weichert, J., Panda, S., Postma, M., & El-Kerdi, A. (2019). Evidence-based Clinical Decision Support Systems for the prediction and detection of three disease states in critical care: A systematic literature review. *F1000Research*, *8*, 1728. https://doi.org/10.12688/f1000research.20498.2 Ramachandran, P., Zoph, B., & Le, Q. (2017). *Searching for Activation Functions*. Cornell University. https://doi.org/10.48550/arxiv.1710.05941
13. Regnault, A., Boroojerdi, B., Meunier, J., Bani, M., Morel, T., & Cano, S. (2019). Does the MDS-UPDRS provide the precision to assess progression in early Parkinson\u2019s disease? Learnings from the Parkinson\u2019s progression marker initiative cohort. *Journal of Neurology*, *266*(8), 1927–1936. https://doi.org/10.1007/s00415-019-09348-3
14. Sapir, S., Ramig, L., & Fox, C. (2008). Speech and swallowing disorders in Parkinson disease. *Current Opinion in Otolaryngology & Head & Neck Surgery*, *16*(3), 205–210. https://doi.org/10.1097/moo.0b013e3282febd3a Sun, L.-C., Tseng, C.-W., Lin, K.-F., & Chen, P.-N. (2025). Hybrid preprocessing and ensemble classification for enhanced detection of Parkinson’s disease using multiple speech signal databases. *Digital Health*, *11*. https://doi.org/10.1177/20552076251352941
15. Tan, Z.-H., & Lindberg, B. (2010). Low-Complexity Variable Frame Rate Analysis for Speech Recognition and Voice Activity Detection. *IEEE Journal of Selected Topics in Signal Processing*, *4*(5), 798–807. https://doi.org/10.1109/jstsp.2010.2057192
16. Van Hilten, J. J., Van Der Zwan, A. D., Zwinderman, A. H., & Roos, R. A. C. (1994). Rating impairment and disability in Parkinson’s disease: evaluation of the Unified Parkinson’s Disease Rating Scale. *Movement Disorders*, *9*(1), 84–88. https://doi.org/10.1002/mds.870090113
17. Yaman, O., Ertam, F., & Tuncer, T. (2019). Automated Parkinson’s disease recognition based on statistical pooling method using acoustic features. *Medical Hypotheses*, *135*, 109483. https://doi.org/10.1016/j.mehy.2019.109483