Early Parkinson's Detection via Aggregated Acoustic Features and Ensemble Learning

Mohammed F. Zamil [1], Dunia H. Hameed [2], Zainab M. Alameen[2]
[1]Department of Bioinformatics, College of Biomedical Informatics, University of Information Technology and Communications (UoITC), Baghdad, Iraq
[2]Department of Intelligent Medical Systems, College of Biomedical Informatics, University of Information Technology and Communications (UoITC), Baghdad, Iraq

Article Info

Article history:

Received May 1, 2026
Revised June 1, 2026
Accepted June 27, 2026

Keywords:

Aggregation Feature
Acoustic Features
Ensemble learning
Machine Learning (ML)
Parkinson Disease (PD)

ABSTRACT

Parkinson's disease (PD) is progressive neurological disorder. The early detection is critical for symptoms reduction. Though early diagnosis is a challenging task even for trained doctors, as studies reported around 25% of Parkinson patients were mistakenly diagnosed in their early stages. So, this study proposes an ensemble machine learning model that trained on 240 voice samples. The 240 audio samples were a mix between sick and healthy 80 persons. The subject-level splitting with Group K-Fold across validation was the first approach, while the other was an advanced feature aggregation. The performance of the models with the stacked of ensemble of Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP) is greatly improved by the aggregation which obtained the highest level of accuracy (88.75%) and F1-score (88.88%). This aggregation method increased robustness by lowering intra-subject variability, while ensemble learning used complementary model to improve classification. The results showed that the meta classifier's complementary capabilities enable it to exploit various decision boundaries, resulting in enhanced generalization and higher diagnostic performance in detecting PD.

Corresponding Author:

Mohammed F. Zamil
Department of Bioinformatics, College of Biomedical Informatics, University of Information Technology and Communications (UoITC), Baghdad, Iraq
Email: mfadhil@uoitc.edu.iq

1. INTRODUCTION

Humans suffer from some neurodegenerative diseases that result from gradual damage to nerve tissue in the Central Nervous System (CNS) and Peripheral Nervous Systems (PNS), causing disability and potentially threatening patients' lives. The primary cause of all these diseases can be the accumulation of damaged or altered proteins. This pathological effect is the result of insufficient or ineffective protein quality control systems that aim to eliminate damaged or unnecessary proteins. Alzheimer's disease, Parkinson's Disease (PD), Huntington's disease, and amyotrophic lateral sclerosis have been classified as neurodegenerative diseases. One of the most important neurodegenerative diseases is Parkinson's disease, which is caused by the loss of dopaminergic neurons and the deposition of Lewy bodies in the substantia nigra. It is a common neurodegenerative disease, and its incidence rate increases annually. It has been shown that certain genes contribute significantly to their development [1][2]. PD is considered an important topic as it affects the quality of life; it includes several symptoms, such as cognitive decline, especially mild cognitive impairment [3][4]. It is the second most popular neurological disease that affects persons aged 65 years or greater by 2%-3%, and it is estimated to affect approximately 10 million people worldwide, with about 3% among people over 80 years old [5], so the biggest influencing factors are the age of the individual and possible associations with various air pollutants. The main features of PD are slowness of movement with tremor or stiffness at rest, which are defined by the International Parkinson and Movement Disorder Society (MDS) [6]-

[7]. People with Parkinson's disease experience several symptoms, some of which are related to movement, such as slowness of movement, tremors, and rigidity. Others are non-motor symptoms, including cognitive impairment, sleep disturbances, and pain [8][9]. Early detection of PD is necessary to aid doctors in disease diagnosis and prescribe timely treatment. Currently, there are no validated biomarkers for effective early diagnosis of PD. Therefore, there is an urgent need for the healthcare sector to utilize Artificial Intelligence (AI) technologies to diagnose Parkinson's disease earlier and more accurately. Many methods to detect PD earlier and all approaches use machine learning methods which are effective in diagnosis, like the using of handwriting analysis which can be used to specify the symptoms of this disease, several types of handwriting like spiral and wave drawings used as input for the systems of machine learning methods [10][11]. Another approach depends on sequence of sentences to help in the assessment of writing development over time [12][13]. The wide range of studies explored the outcomes of AI with clinical applications; one of these papers found how AI improved diagnosis by treatment planning, medical image analysis, and synthetic data generation; it also improved medical education, healthcare marketing, and financial workflows [14]. So, studies proved that AI could have a big role in Parkinson treatment planning, monitoring and diagnosing especially in its early stages [15].

In another hand, researchers study the audio information/features based diagnostic method to identify PD patient's using their voice only. Despite the high potential benefit, there are no public datasets that is enough in size and high quality so the researchers can build high performance models in diagnosing PD patients using their voice recordings yet [16]. However, researchers deployed AI, ML and data preprocessing techniques to overcome such limitations. These techniques provide non-traditional methods in extracting more distinguishable and effective voice features to detect illness cases. For this reason, a lot of research has adopted AI techniques with many diagnostic methods relied on sound. So, Varela-Arellano et al. [17] addressed an early PD identification by utilizing ML algorithms which were applied to acoustic variables in speech recordings. Voice abnormality is one of the first signs of this disease; the researchers used the PD Vocal dataset with Multiple Audio Recordings to train and evaluate multiple machine learning models; including Random Forest (RF), boosting, and bagging. The Boosting classifier exceeded the others by demonstrating the use of voice-based signals in early Parkinson's disease recognition. Another exploration focused on the role of feature selection to increase classifier performance; therefore, it chose Support Vector Machine (SVM) as the most effective model, and offered a low-cost alternative to traditional advanced diagnostic procedures [18].

Das et al. [19] introduced SEFRON, a Spiking Neural Network model trained on the Parkinson Dataset with Replicated Acoustic Features, SEFRON achieved an average accuracy of 91.94%, outperforming MLP, RBF, RNN, and LSTM models. Nevertheless, the study treated the replicated recordings as independent samples and did not apply subject-level validation, which means that results are not reliable due to data leakage between training and testing sets. Audio features can be enough for remote medical diagnosis and telemedicine, the study that was produced by Govindu and Palwe [20] explored various classifiers, and it concluded that K-Nearest Neighbors (KNN) and RF are the most effective. The strategy focused on audio as a non-invasive biomarker and providing a practical and effective option for early detection, especially for people with mobility difficulties. Because of some data are random, another study by Thakur et al.[21] tried various types of ML algorithms with this disease diagnosis, like SVM, RFC, DTC, and Extra Trees Classifier. Then, it compared these models, but it discovered that the Extra Trees Classifier is the best when combined with entropy-based feature selection, with higher performance accuracy. Also, researchers have studied the enhancement that could be achieved by using a combination of ML algorithms and models such as KNN, SVM, and neural networks. The ensemble machine learning techniques in general contribute to PD diagnosing. Besides, all studies that used speech data needed further feature engineering and data pre-processing such as feature selection, data balancing, and model tuning; the study combined classifiers to increase accuracy.

By focusing solely on voice, a rapid, inexpensive, and non-invasive biomarker, a new study aimed to replace traditional diagnostic techniques for providing early detection and opening the way to the broader applications in the diagnosis of neurological disorders [22]. Again, the study was by Du et al.[23] utilized the strong points of MFCS, SVM, and HKNN to produce a creative machine learning approach called LCCB when it condensed on class boundaries and then measured the hyper distance to limit the labels. So, this new technique produced a powerful interpretation to address the hard decision boundaries very well. The results revealed that LCCB outperforms several advanced models, particularly when enhanced with MFCS preprocessing. One of the notable studies demonstrated that combining a wide range of machine learning models can achieve high performance PD classifier by integrating multiple data modalities. The employed models included LR, RF, AdaBoost, SVM, MLP and K-Nearest Neighbors. These models were trained using various data types, including demographic, genomic, and clinical measures, to improve the early prediction performance [24]. Similarly, another different study used a combination between a deep and machine learning techniques for Parkinson's disease classification; it used the diversity of data sources; including audio recordings, MRI, EHRs, and CT scans. It utilized Wolf Search Optimization (WSO) to find the best feature subsets in the data, which are subsequently tuned with a sparse auto encoder to improve model performance.

In general, prior studies confirm using vocal features to build ML models with reasonable classification performance, although still enhancement in working with small size and replicated samples have a potential for further improvement [25]. For thus, this study investigates the effect of building ensemble and individual machine learning models that relied on subject-level aggregation of replicated voice recording features with very limited dataset. To evaluate the improvement, the proposed approach was compared to same models trained on row recording.

The rest of the paper is organized as follows: Section 2 introduces the data considered for this study, section 3 explains the methodology followed in the paper, section 4 shows experiments and results, section 5 elaborates a discussion about findings reported in the study, and finally, section 6 presents the conclusions derived from this work.

2. METHODOLOGY

The study methodology starts first with dataset exploring, preprocessing, splitting and then models development and selection. Finally, details of how experiments are conducted and evaluation metrics used.

2.1 Dataset and Pre-processing Methods

In this paper we aimed to find the best algorithm that can build an acceptable detection model using a small dataset such as Replicated Acoustic Features Parkinson Dataset [26]. Diagnose Parkinson's disease at early stages used acoustic features (voice recordings) were derived from sustained vowel phonations. The new dataset used the Replicated Acoustic Features Parkinson Dataset, which comprised 240 recordings from 80 individuals (40 with Parkinson's, 40 healthy), with three voice samples per subject. The primary goal was to evaluate the impact of different data preparation strategies, raw instance-level modeling where each voice record was treated as single data point versus subject-level where feature aggregation method applied to the three records of the same person. These two data preparations methods applied to see their effectiveness on classification performance of various machine learning models.

2.2 Data Sampling and Preparation

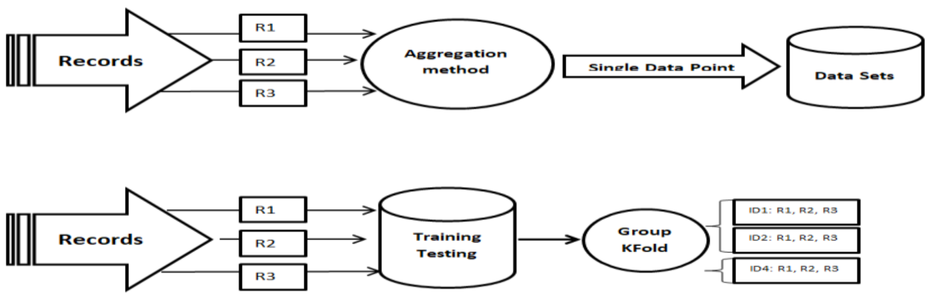

Figure 1. Data Preparation Steps

In the first approach, we used all 240 voice samples as independent instances. However, to avoid data leakage due to repeated samples from the same subject, we implemented subject-level splitting using Group K-Fold cross-validation, where the ID of each subject was used as the grouping factor. This data sampling method ensured that voice samples from the same individual are not used in learning and testing in the same time, as such incidents consider as data leakage which is critical issue that can cause unreliable performance results.

In the second approach, we applied advanced feature aggregation to summarize the three recordings per subject into a single instance. For each acoustic feature, we computed several statistical descriptors including mean, standard deviation, minimum, maximum, median, and Inter Quartile Range (IQR). This transformation yielded a dataset of 80 aggregated instances, each corresponding to a unique subject, thus allowing the models to learn from consolidated, subject-level profiles. Such aggregation methods may result in some data pattern loss, besides reducing data samples from 240 to 80. Generally, we cannot be certain whether it will contribute to enhancing or reducing models' performances until applying the learning process and comparing results.

2.3. Feature Normalization

All features in both approaches were standardized using z-score normalization to ensure a common scale, which is critical for distance-based and regularized models like SVM, KNN, and Multi-Layer Perceptron (MLP). The scale was fitted only on the training folds and applied to the test folds during cross-validation to prevent information leakage.

2.4. Model Selection and Evaluation

In the next stage, we evaluated a diverse set of supervised classification models, like Logistic Regression (LR), SVM, RF, KNN, MLP, and XGBoost. Additionally, we implemented a stacked ensemble classifier that combines the predictions of LR, SVM, RF, and MLP as base learners, with a logistic regression meta-classifier. These models belong to different learning mechanisms which can make comparisons become more effective.

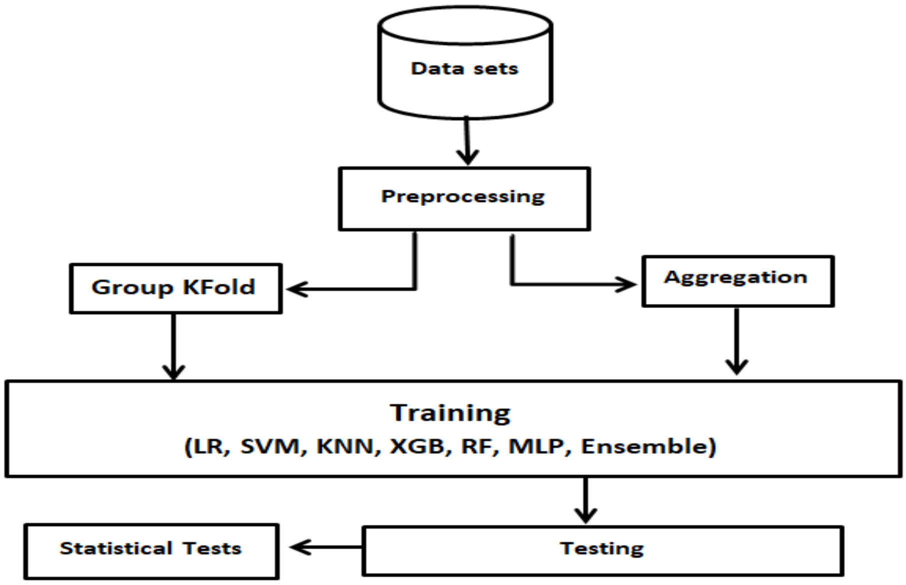

Figure 2. Block Diagram of Proposed Model

Each of the selected models belongs to a different family so each of them can capture different types of patterns and decision boundaries. For example, a linear model such as LR can detect linear patterns that are interpretable relationships. While RF, which is a bagging ensemble method, was fine at handling non-linearity and feature interactions. In another hand, SVM and MLP both could handle complex patterns. The models were aggregated using a stacking ensemble method, in which meta-model learns how to combine them to improve overall predictive performance.

Table 1. Models' hyper-parameters settings.

Model	Hyperparameters
Logistic Regression	solver = saga; penalty = elasticnet; C = 0.01; l1_ratio = 0.1; max_iter = 1000
Random Forest	n_estimators = 50; max_depth = None; min_samples_leaf = 4; min_samples_split = 3;
SVM	kernel = sigmoid; C = 0.1; gamma = 10; probability = True
MLP	hidden_layers = (64, 16); activation = logistic; solver = adam; alpha = 0.0001; learning_rate = constant; max_iter = 300;
Stacking Ensemble	base learners = LR, RF, SVM, MLP; meta-classifier = Logistic Regression

The models in this study were built based on a predefined set of hyperparameters selected to optimize models' performance. To enable the reproducibility of the experiments and allow results replicating, Table 1 summarizes the hyperparameter configurations of all classifiers, including the base learners and the stacking ensemble.

Cross-validation is a widely used technique for evaluating the performance and generalization ability of machine learning models. Instead of relying on a single train/test split, it partitions the data into multiple folds, trains the model on a subset of the folds, and tests it on the remaining fold. Evaluation results are result of 5-fold cross-validation, one difference for non-aggregated data points we used specifically group K fold with subject IDs to avoid data leakage. The proposed models and data preprocessing settings evaluated using accuracy, precision, recall and F1-score to determine the best combination. To determine whether there was a significant difference between the raw and aggregated methods, a paired t-test on all evaluation metrics.

3. RESULTS AND DISCUSSION

Table 2 presents the classification models' evaluation metrics obtained using different datasets preprocessing and sampling methods. Using raw recordings, LR shows the highest performance model according to F1 score (79.81%) and accuracy (79.58%). However, this performance was significantly outperformed when aggregation methods applied 80 rows instead of 240; the stacked classifier achieved 88.88% F1-score, 90.00% recall, and 88.75% accuracy.

Interestingly, all models demonstrated enhanced performance when aggregations techniques were employed. For example, SVM improved from 78.28% to 87.97% F1-score, while KNN improved from 78.80% to 86.19%. This indicates that the aggregation methods helped reduce noise and outlier values in the data and improved its overall quality, and this reflected the overall performance enhancement on all models deployed the enhanced aggregated method.

Table 2. The evaluation metrics by apply the different classification models.

Model	Accuracy	Precision	Recall	F1 Score	Setting
Stacked Classifier	**0.8875**	**0.8933**	**0.9000**	**0.8888**	**Enhanced Aggregated Method**
SVM	0.8750	0.8778	0.9000	0.8797	Enhanced Aggregated Method
KNN	0.8625	0.8788	0.8750	0.8619	Enhanced Aggregated Method
MLP	0.8625	0.9056	0.8250	0.8595	Enhanced Aggregated Method
Logistic Regression	0.8500	0.8605	0.8750	0.8571	Enhanced Aggregated Method
Random Forest	0.8500	0.8760	0.8500	0.8510	Enhanced Aggregated Method
XGBoost	0.8250	0.8566	0.8250	0.8231	Enhanced Aggregated Method
Logistic Regression	**0.7958**	**0.8140**	**0.8000**	**0.7981**	**Raw Recordings + Group K folds**
Random Forest	0.8000	0.8379	0.7667	0.7940	Raw Recordings + **Group K folds**
KNN	0.7958	0.8110	0.7833	0.7880	Raw Recordings + **Group K folds**
SVM	0.7750	0.7765	0.8167	0.7828	Raw Recordings + **Group K folds**
Stacked Classifier	0.7667	0.7942	0.7583	0.7664	Raw Recordings + **Group K folds**
MLP	0.7500	0.7966	0.6917	0.7355	Raw Recordings + **Group K folds**
XGBoost	0.7292	0.7625	0.6833	0.7137	Raw Recordings + **Group K folds**

In this study, various combinations of base learners were explored to optimize overall performance. The highest achieved performance was composed of (LR, RF, SVM and MLP) F1-score (0.8888), the highest recall (0.90), and a strong accuracy (0.8875). The heterogeneity of model mechanisms is an effective factor to enhance performance. Since each model has different capabilities in features extractions which lead to each model complementing the other. The meta-classifier to get overwhelmed view for decisions leading to improved generalization power.

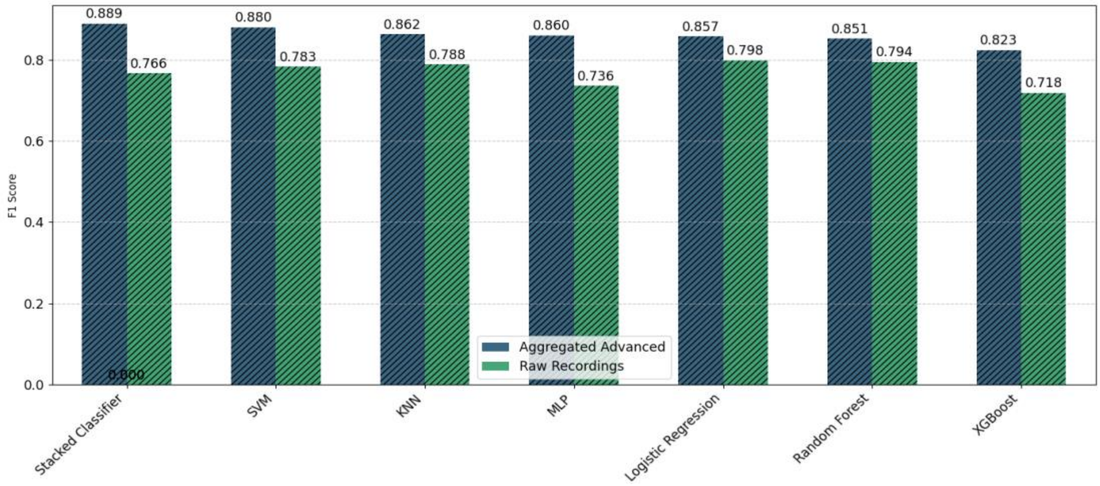

Figure (3): Bar chart demonstrates F1 Score values of two classification methods with many different algorithms

Our stacked model composed of models that have different patterns in their performance as SVM and LR show high recall as they could catch most positive Parkinson patients but their precision lower. In the other hand, MLP and RF have missed more Parkinson cases, but their positive decision is more likely true positive. Combining those two groups of models aimed to produce more balanced models in terms of precision and recall so that it becomes reliable and trusted in medical tasks.

In addition to stacked model, results (Table 2) show clear enhancements due to preprocessing (aggregations) applied to the dataset. Results show that advanced aggregation (mean, std, min, max, median, and IQR) methods applied to recordings have enabled models to generalize better and capture more informative features associated with PD symptoms. So, in general, the enhancement does not only rely on the ensemble stacked model, but also aggregation of the three patient's records contributes in reducing the sound diffusion, noise and giving clearer stable and coherent acoustic features to better express patient's sound features.

Table 3. Statistical Tests.

Metric	Raw Mean	Aggregated Mean	t-statistic	p-value
Accuracy	0.7688	**0.8531**	8.94	<0.001
Precision	0.7967	**0.8761**	8.51	0.0001
Recall	0.7479	**0.8562**	10.87	<0.001
F1 Score	0.7621	**0.8531**	9.68	<0.001

To prove the enhancement in results, paired t-tests were applied to all metrics across all models. As Table 3 shows, the t-test revealed that the performance differences between raw sound recordings and our enhanced aggregated method were statistically significant ($p < 0.001$) for all metrics. This validates the benefit of subject-level feature samples aggregation in the context of sound datasets. Results show that supervised machine learning models can detect early signs of Parkinson's disease using voice records only biomarkers, specifically when multiple voice recordings preprocessed and aggregated. Such aggregations not only enhance model's performance but also align with real-world clinical protocol as decisions are made per patient, not per recording.

Table 4 shows a comparison between our proposed model with prior relevant studies using same dataset as we used in our study. As the dataset documentation warns that dataset contains three replications belonging to the same patient, so data sampling should avoid separating a patient into training and testing. For thus, we add this note to the comparison table whether the study explicitly mention how data sampling done to avoid data leakage. The proposed method employs subject-aware cross-validation, ensuring that all replications of a given subject remain entirely within either the training or the testing set to prevent any data leakage. Besides, our proposed ensemble model shows high performance comparable to relevant previous studies.

Table 4. Studies Performance Comparison.

Study (Year)	Method	Data leakage prevention	Accuracy	F1-score
Naranjo et al. (2016) [27]	**Subject-based Bayesian approach**	**Yes**	**~75.2%**	**—**
Naranjo et al. / Two-stage (2017) [28]	**Two-stage variable + sample selection + LASSO**	**Yes**	**86.20%**	**—**
Ghaheri et al. (2024) [29]	**SHAP + Hard Voting Ensemble**	**Not stated**	**85.42%**	**84.94%**
Wang/Das et al. / SEFRON (2024) [19]	**Spiking Neural Network**	**No (treated the 240 recordings as independent)**	**88.89%**	**88.5%**
Proposed Study	**Stacking Ensemble**	**Yes (Group K-fold)**	**88.75%**	**88.88%**

4. CONCLUSION

The experimental results revealed the machine learning efficiency for early Parkinson's disease detection by using patient voice recordings. Two approaches were considered: the first utilized all 240 speech samples with subject-aware cross-validation to ensure the confidentiality, while the audio combining across subjects via statistical attributes such as mean, standard deviation, median, minimum, maximum or interquartile range was the second strategy. The outcomes confirmed that the aggregation method achieved the highest accuracy (88.75%) and F1-score (88.88%) with the algorithms Stacked Classifier, SVM, KNN, MLP, Logistic Regression, Random Forest, and XGBoost. The aggregation ways could improve the durability by minimizing the intra-subject variation, while the effectiveness of the ensemble approach could enhance the classification. In conclusion, the complete work must merge the two approaches, feature aggregation, and ensemble strategy for more reliability with the early detection of PD. As future work, we aim to build a multi-modal detection systems that use several and multi-source data types such as voice recordings, handwriting and medical tests.

FUNDING INFORMATION

There is no funding involved.

DATA AVAILABILITY

The data that support the findings of this study are openly available in UCI machine learning repository. UCI Machine Learning Repository. https://doi.org/10.24432/C5701F.

REFERENCES

[1] P. Yousefi, S. Ghadirian, M. Mobedi, M. Jafarzadeh, A. Alirezaei, A. Gholami, and A. Tabibzadeh, "Autophagy related genes polymorphisms in Parkinson's disease: A systematic review of literature," *Clinical Parkinsonism & Related Disorders*, vol. 12, p. 100312, 2025, doi: 10.1016/j.prdoa.2025.100312.

[2] K. A. Jellinger, "Mild cognitive impairment in Parkinson's disease: Current view," *Frontiers in Cognition*, vol. 3, Art. no. 1369538, 2024, doi: 10.3389/fcogn.2024.1369538.

[3] F. Zhang, L. Guo, L. Liu, and X. Han, "Diagnostic accuracy of machine learning-assisted MRI for mild cognitive impairment in Parkinson's disease: A systematic review and meta-analysis," *Parkinson's Disease*, 2025, Art. no. 2079341, doi: 10.1155/padi/2079341.

[4] K. A. Silveira, G. B. Souza, C. M. Carpenedo, C. R. M. Rieder, and B. C. Beber, "Associations between quality of life in communication and sociodemographic and clinical aspects of Parkinson's disease," *CoDAS*, vol. 37, no. 5, p. e20240298, 2025, doi: 10.1590/2317-1782/e20240298en.

[5] S. Gao *et al.*, "Early detection of Parkinson's disease through multiplex blood and urine biomarkers prior to clinical diagnosis," *NPJ Parkinson's Disease*, 2025, doi: 10.1038/s41531-025-00888-2.

[6] M. Ianculescu, C. Petean, V. Sandulescu, A. Alexandru, and A.-M. Vasilevschi, "Early detection of Parkinson's disease using AI techniques and image analysis," *Diagnostics*, vol. 14, no. 23, p. 2615, 2024, doi: 10.3390/diagnostics14232615.

[7] K. A. Jellinger, "Pathobiology of cognitive impairment in Parkinson disease: Challenges and outlooks," *International Journal of Molecular Sciences*, vol. 25, no. 1, p. 498, 2023, doi: 10.3390/ijms25010498.

[8] R. Akhmedullin *et al*., "The regional burden of Parkinson's disease in Kazakhstan 2014–2021: Insights from national health data," *Parkinson's Disease*, vol. 2025, Art. no. 4317554, 2025, doi: 10.1155/padi/4317554.

[9] G. Marano *et al*., "Psychiatric and cognitive symptoms of Parkinson's disease: A life's tale," *Revue Neurologique*, vol. 181, no. 4, pp. 265–283, 2025, doi: 10.1016/j.neurol.2024.11.004.

[10] M. Shelk, N. M. Ranjan, and G. Mate, "Detection of Parkinson's disease using machine learning algorithms and handwriting analysis," *Journal of Data Mining and Management*, vol. 8, no. 1, 2022, doi: 10.46610/JoDMM.2023.v08i01.004.

[11] A. Razaq, S. Ramzan, S. Jabbar, M. Iqbal, M. Habib, and U. Raza, "A deep learning framework for early Parkinson's disease detection: Leveraging spiral and wave handwriting tasks with EfficientNetV2-S," *Diagnostics*, vol. 15, p. 2795, 2025, doi: 10.3390/diagnostics15212795.

[12] K. Białek, A. Potulska-Chromik, J. Jakubowski, M. Nojszewska, and A. Kostera-Pruszczyk, "Analysis of handwriting for recognition of Parkinson's disease: Current state and new study," *Electronics*, vol. 13, no. 19, p. 3962, 2024, doi: 10.3390/electronics13193962.

[13] G. Marano *et al*., "Writing the future: Artificial intelligence, handwriting, and early biomarkers for Parkinson's disease diagnosis and monitoring," *Biomedicines*, vol. 13, p. 1764, 2025, doi: 10.3390/biomedicines13071764.

[14] S. S. Bhuyan *et al*., "Generative artificial intelligence use in healthcare: Opportunities for clinical excellence and administrative efficiency," *Journal of Medical Systems*, vol. 49, no. 1, p. 10, 2025, doi: 10.1007/s10916-024-02136-1.

[15] Y. A. Fahim, I. W. Hasani, S. Kabba, and W. M. Ragab, "Artificial intelligence in healthcare and medicine: Clinical applications, therapeutic advances, and future perspectives," *European Journal of Medical Research*, 2025, doi: 10.1186/s40001-025-03196-w.

[16] K. Raut and V. Balpande, "Early detection of Parkinson's disease: Ensemble learning for improved diagnosis," *EPJ Web of Conferences*, vol. 328, p. 01005, 2025, doi: 10.1051/epjconf/202532801005.

[17] Y. O. Varela-Arellano *et al*., "Early diagnosis of Parkinson's disease through machine learning and voice signals," in *Proc. X Latin American Conf. on Biomedical Engineering (CLAIB 2024)*, IFMBE Proc., vol. 121. Springer, 2025, doi: 10.1007/978-3-031-89514-2_23.

[18] Z. Karapinar Senturk, "Early diagnosis of Parkinson's disease using machine learning algorithms," *Medical Hypotheses*, vol. 138, p. 109603, 2020, doi: 10.1016/j.mehy.2020.109603.

[19] Das, P., Nanda, S., Panda, G., Dash, S., Ksibi, A., Alsenan, S., Bouchelligua, W., & Mallik, S. (2024). A robust Parkinson's disease detection model based on time-varying synaptic efficacy function in spiking neural network. BMC Neurology, 24(1). https://doi.org/10.1186/s12883-024-04001-7..

[20] A. Govindu and S. Palwe, "Early detection of Parkinson's disease using machine learning," *Procedia Computer Science*, vol. 218, pp. 249–261, 2022, doi: 10.1016/j.procs.2023.01.007.

[21] K. Thakur, D. S. Kapoor, K. J. Singh, A. Sharma, and J. Malhotra, "Diagnosis of Parkinson's disease using machine learning algorithms," in *Proc. Third Congress on Intelligent Systems (CIS 2022)*, Lecture Notes in Networks and Systems, vol. 608. Springer, 2023, doi: 10.1007/978-981-19-9225-4_16.

[22] S. Srinivasan *et al*., "Detection of Parkinson disease using multiclass machine learning approach," *Scientific Reports*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1038/s41598-024-64004-9.

[23] Q. Du, J. Shen, P. Wen, *et al*., "Parkinson's disease detection by using machine learning method based on local classification on class boundary," *Discover Applied Sciences*, vol. 6, p. 576, 2024, doi: 10.1007/s42452-024-06295-1.

[24] Makarious, M. B., Leonard, H. L., Vitale, D., Iwaki, H., Sargent, L., Dadu, A., Violich, I., Hutchins, E., Saffo, D., Bandres-Ciga, S., Kim, J. J., Song, Y., Maleknia, M., Bookman, M., Nojopranoto, W., Campbell, R. H., Hashemi, S. H., Botia, J. A., Carter, J. F., … Nalls, M. A. (2022). Multi-modality machine learning predicting Parkinson's disease. Npj Parkinson's Disease, 8(1). https://doi.org/10.1038/s41531-022-00288-w.

[25] Y. Xiong and Y. Lu, "Deep feature extraction from the vocal vectors using sparse autoencoders for Parkinson's classification," *IEEE Access*, vol. 8, pp. 27821–27830, 2020, doi: 10.1109/ACCESS.2020.2968177.

[26] Prez, C. (2016). Parkinson Dataset with replicated acoustic features [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5701F.

[27] Naranjo, L., Pérez, C. J., Campos-Roca, Y., & Martín, J. (2016). Addressing voice recording replications for Parkinson's disease detection. Expert Systems with Applications, 46, 286–292. https://doi.org/10.1016/J.ESWA.2015.10.034.

[28] Naranjo, L., Pérez, C. J., Martín, J., & Campos-Roca, Y. (2017). A two-stage variable selection and classification approach for Parkinson's disease detection by using voice recording replications. Computer Methods and Programs in Biomedicine, 142, 147–156. https://doi.org/10.1016/J.CMPB.2017.02.019.

[29] Ghaheri, P., Nasiri, H., Shateri, A., & Homafar, A. (2024). Diagnosis of Parkinson's disease based on voice signals using SHAP and hard voting ensemble method. Computer Methods in Biomechanics and Biomedical Engineering, 27(13), 1858–1874. https://doi.org/10.1080/10255842.2023.2263125.