ISSN 2301-8984 (Print), 2809-1884 (Online)
Published by Institute of Computer Science

Swarm driven automatic feature selection and classification framework for parkinson voice data

Muhammad Azhar Prabukusumo[1], Hondor Saragih[2], Jonson Manurung[3]
[1,2,3] Informatika, Universitas Pertahanan Republik Indonesia, Bogor, Indonesia

ARTICLE INFO

Article history:

Received Oct 12, 2025
Revised Oct 27, 2025
Accepted Oct 30, 2025

Keywords:

Feature Selection;
Machine Learning;
Parkinson's Disease;
Particle Swarm Optimization;
Voice Analysis.

ABSTRACT

Parkinson's disease (PD) severely impairs motor and vocal functions, and early detection is crucial for effective intervention. Conventional diagnostic procedures remain subjective and time-consuming, highlighting the need for automated, data-driven approaches. This study aims to develop an intelligent and fully automated framework integrating Particle Swarm Optimization (PSO)–based feature selection with ensemble machine learning classifiers for PD detection using voice data. The proposed Swarm-Driven Automatic Feature Selection and Classification Framework (SAFSCF) automates data preprocessing, adaptive feature optimization, and classification within a unified pipeline. The framework was evaluated on the Parkinson's Speech Dataset comprising 743 numerical features. Baseline models achieved accuracies of 0.7738 (Logistic Regression), 0.8651 (Random Forest), and 0.8690 (Gradient Boosting). After PSO optimization, the feature set was reduced by nearly 50% to 382 attributes, achieving a test accuracy of 0.8421 slightly higher than the full-feature model (0.8355). Convergence plots confirmed that PSO effectively minimized the fitness function while maintaining high classification stability. Feature importance analysis revealed that the most discriminative attributes were derived from log energy, Teager Kaiser energy operators (TKEO), MFCCs, Shimmer, and entropy-based features biomarkers known to reflect Parkinsonian speech degradation. These findings demonstrate that the proposed framework enhances computational efficiency and interpretability, offering a reproducible and scalable solution for non-invasive, voice-based PD diagnosis.

Corresponding Author:

Muhammad Azhar Prabukusumo,
Informatika,
Universitas Pertahanan Republik Indonesia,
Kawasan IPSC Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat 16810, Indonesia.
Email: azhar.prabukusumo @idu.ac.id

1. INTRODUCTION

In the era of big data, the exponential growth of digital information has posed significant Parkinson's disease (PD) is a progressive neurodegenerative disorder that severely impairs motor control, speech articulation, and overall quality of life (Kavya et al., 2022; Sheng et al., 2021). Early detection plays a vital role in enabling timely intervention and effective treatment planning, yet conventional diagnostic procedures still rely on clinical assessments that are subjective, time-consuming, and require specialized expertise (Nazir et al., 2025; Shao et al., 2023). With advances in biomedical engineering, voice-based analysis has emerged as a promising non-invasive and cost-effective approach for detecting PD (Gawali, 2024; Quamar et al., 2025). Subtle vocal impairments such as reduced pitch variation and tremor-induced distortions often appear in the early stages of the disease and can be quantitatively captured through acoustic biomarkers (et al., 2025). This development has encouraged the integration of Artificial Intelligence (AI), Machine Learning (ML), and digital signal processing to create automated systems capable of identifying Parkinson's disease through voice data (Dixit et al., 2023; Rabie & Akhloufi, 2025).

Recent studies have investigated various machine learning and deep learning approaches for Parkinson's disease (PD) detection using voice-based data. (Govindu & Palwe (2022) compared Logistic Regression, Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), and Random Forest classifiers on the MDVP speech dataset, achieving the highest accuracy of 91.83% using Random Forest; however, their reliance on Principal Component Analysis (PCA) for dimensionality reduction limited the model's adaptability in identifying nonlinear and disease-specific vocal features. (Karabayir et al. (2020) further explored Gradient Boosting techniques such as LightGBM and XGBoost on an extended PD voice dataset containing 44 replicated acoustic features, reaching an AUC of 0.951 with only seven selected attributes. Despite showing strong predictive performance, their study was constrained by a small dataset and lacked an adaptive mechanism for feature optimization, reducing its generalization capability across heterogeneous voice samples. Meanwhile, (Iyer et al. (2023) implemented a transfer learning-based deep learning approach using the Inception V3 Convolutional Neural Network (CNN) on spectrogram representations of telephonic voice recordings, achieving an impressive AUC of 0.97. Nonetheless, this model required high computational resources and offered limited interpretability regarding which acoustic biomarkers most contributed to classification. Collectively, these studies demonstrate that although traditional machine learning and deep learning models have achieved notable performance in PD voice classification, they still suffer from limitations in feature redundancy, adaptability, and interpretability. Therefore, a methodological advancement is required an intelligent, swarm-driven, and fully automated framework capable of integrting feature selection and classification within a unified pipeline to enhance efficiency, interpretability, and diagnostic reliability for Parkinson's disease detection.

Traditional feature selection techniques, such as filter, wrapper, and embedded methods, have been widely used to reduce data redundancy and enhance classifier performance (Bashir et al., 2022; Chen et al., 2020). However, these methods are typically static, rely on fixed assumptions, and require manual parameter tuning, which limits their adaptability when handling complex and nonlinear biomedical data (Dhanka et al., 2025). To address these limitations, metaheuristic optimization algorithms inspired by natural processes have been increasingly adopted for feature selection (Khurma et al., 2022). Among them, swarm intelligence methods, particularly Particle Swarm Optimization (PSO), have attracted significant attention due to their capability to explore large search spaces effectively and adaptively identify optimal feature subsets (Gad, 2022).

PSO has demonstrated remarkable performance in biomedical applications because of its strong convergence properties, robustness against local minima, and suitability for high-dimensional data (Al-Shammary et al., 2022). Several studies have shown that PSO-based feature selection can significantly improve classification accuracy in various medical domains such as leukima, brain cancer and breast cancer prediction (Nagra et al., 2024; Sowan et al., 2023). Nevertheless, most of these implementations treat PSO as a separate optimization stage rather than as an integrated component within a unified and adaptive framework. As a result, the full potential of PSO for automated biomedical data analysis remains underutilized (Zaini et al., 2023).

The absence of integration across data preprocessing, feature selection, and classification represents a key gap in the existing literature. Many studies focus primarily on achieving high accuracy but pay little attention to automation, reproducibility, and scalability, which are essential for clinical deployment. Furthermore, the lack of self-adaptive mechanisms and unified processing pipelines makes it difficult to generalize findings across different datasets. Therefore, there is a clear need for an intelligent, adaptive, and fully automated system that combines preprocessing, feature optimization, and classification into a cohesive and clinically applicable framework for PD detection.

In response to these challenges, this study proposes a Swarm-Driven Automatic Feature Selection and Classification Framework for Parkinson's disease voice data. The framework integrates automated data preprocessing, PSO-based adaptive feature selection, and ensemble classification using models such as Random Forest, Logistic Regression, and Gradient Boosting. Unlike conventional approaches, the proposed system embeds PSO as a dynamic and self-adaptive component within the entire classification pipeline. This integration allows the framework to autonomously adjust parameters, identify optimal features, and align them with suitable classifiers, thereby minimizing human intervention and improving reproducibility. Through this approach, the research aims to contribute both methodologically and practically to the development of intelligent, scalable, and clinically meaningful diagnostic systems for early detection of Parkinson's disease.

2. RESEARCH METHOD

Research Framework Overview

This study adopts a computational framework that integrates automated data preprocessing, swarm intelligence-based feature selection, and ensemble classification to improve the accuracy and efficiency of Parkinson's disease (PD) voice data analysis. The proposed method, called the Swarm-Driven Automatic Feature Selection and Classification Framework (SAFSCF), is designed as a fully automated modular pipeline. It eliminates the need for manual tuning and domain-specific intervention, enabling reproducible and scalable model deployment.

The framework operates through a series of interconnected stages, starting with delimiter detection and data loading, followed by dynamic header and target identification, feature and target separation, and data cleaning and normalization. Baseline model evaluation is performed using ensemble classifiers, and feature optimization is conducted via Particle Swarm Optimization (PSO). The selected features are then used for model training and validation. Finally, the framework includes performance comparison, visualization, result storage, and interpretability analysis. Each stage functions autonomously, forming an adaptive loop that reduces redundancy, enhances feature interpretability, and strengthens diagnostic reliability in PD classification.

Dataset Description

The dataset used in this study is the Parkinson's Speech Dataset, obtained from the UCI Irvine Machine Learning Repository (Parkinson's Disease Classification, UCI Machine Learning Repository). It contains 756 instances from 188 Parkinson's patients (107 men, 81 women, aged 33–87, mean ± SD: 65.1 ± 10.9) and 64 healthy controls (23 men, 41 women, aged 41–82, mean ± SD: 61.1 ± 8.9) collected at the Department of Neurology, Cerrahpaşa Faculty of Medicine, Istanbul University. Subjects performed three sustained phonations of the vowel /a/ recorded at 44.1 kHz. The dataset includes 754 features extracted using various speech signal processing techniques such as Time-Frequency features, Mel-Frequency Cepstral Coefficients (MFCCs), Wavelet Transform-based features, Vocal Fold features, and Time-Weighted Quadratic Transform (TWQT) features, capturing vocal instability and tremor-related irregularities for Parkinson's disease assessment and classification.

Table 1 Dataset

id	gender	PPE	…	tqwt_kurtosisValue_dec_35	tqwt_kurtosisValue_dec_36	class
0	1	0.85247	…	30.004	189.405	1
0	1	0.76686	…	63.431	45.178	1
0	1	0.85083	…	31.495	47.666	1
1	0	0.41121	…	6.265	40.603	1
1	0	0.3279	…	50.559	61.164	1
1	0	0.5078	…	35.046	3.225	1
2	1	0.76095	…	13.83	77.693	1
2	1	0.83671	…	180.927	50.448	1
2	1	0.80826	…	36.216	3.843	1
3	0	0.85302	…	58.807	387.211	1
3	0	0.80657	…	28.233	26.381	1
3	0	0.82653	…	33.948	269.617	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	0	0.81304	…	26.217	31.527	0

Data Preprocessing

Due to the heterogeneous and inconsistent format of biomedical data, preprocessing is fully automated. The framework first detects the most appropriate delimiter (from ;, ,, or tab) to maximize the number of valid columns. Headers and target columns are then identified by searching for keywords such as "class" or "target." Non-numeric and identifier columns, such as IDs, are removed, and string-based numeric values are converted to floating-point numbers, accommodating both comma and period decimal separators. Missing values are imputed using the mean, and features are standardized using StandardScaler to ensure uniform distribution. Columns that are entirely missing or non-numeric are discarded, resulting in a preprocessed feature matrix (X_scaled) that is numerically stable and ready for PSO-based feature selection.

Baseline Model Establishment

To establish a benchmark prior to feature optimization, three widely used classifiers Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB) are applied to the complete feature set. Logistic Regression is a parametric statistical model that predicts the probability of a binary outcome using the logistic function:

$$P(y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{D} \beta_i x_i)}} \quad (1)$$

where $X = (x_1, x_2, \dots, x_D)$represents the feature vector, $\beta_i$are the model coefficients, and $D$is the total number of features (Bhuyan et al., 2021). Random Forest, on the other hand, is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes for classification tasks (Salman et al., 2024). Gradient Boosting sequentially builds an ensemble of weak learners, typically decision trees, by fitting each new tree to the residual errors of the previous ensemble to minimize a specified loss function (Kunapuli, 2023).

To ensure fair and unbiased evaluation of these classifiers, a Stratified K-Fold Cross-Validation with $k = 3$is employed, which divides the dataset into three subsets while preserving the proportion of each class within each fold. Each fold serves as a test set once while the remaining folds are used for training, and the process is repeated three times. The mean cross-validation accuracy is computed as:

$$\text{baseline\_acc} = \frac{1}{K} \sum_{j=1}^{K} \frac{\text{Number of correctly classified instances in fold } j}{\text{Total instances in fold } j} \quad (2)$$

where $K = 3$. This baseline accuracy provides a quantitative reference for evaluating the effectiveness of subsequent feature selection using PSO. By comparing the performance of classifiers trained on the full feature set with those trained on the PSO-selected subset, the contribution of the feature selection process to both predictive accuracy and dimensionality reduction can be assessed.

Swarm-Based Feature Selection using PSO

The Particle Swarm Optimization (PSO) algorithm is used to select an optimal subset of features and is implemented as a binary discrete PSO. Each particle is represented by a binary vector, where 1 indicates a selected feature and 0 indicates an unselected feature (Shami et al., 2022). The dimensionality of each particle corresponds to the total number of features after preprocessing, and a population of 100 particles is initialized randomly. The velocity matrix is initially set to zero and updated iteratively.

The fitness function balances classification accuracy and subset compactness, and is defined as (Aldossary, 2025):

$$\text{Fitness} = \alpha(1 - \text{Accuracy}) + \beta \frac{k}{D} \quad (3)$$

where: Accuracy= mean cross-validation accuracy of Logistic Regression, k= number of selected features in the current particle, D= total number of available features, α, β= weight coefficients (set to 0.9 and 0.1, respectively)

The goal is to minimize the fitness value, so higher accuracy and smaller feature subsets produce better solutions.

The velocity and position of each particle are updated as follows:

$$v_i = wv_i + c_1 r_1 (pbest_i - x_i) + c_2 r_2 (gbest - x_i)$$
$$x_i = \frac{1}{1 + e^{-v_i}} \quad (4)$$

where: v_i= velocity of particle i, x_i= position of particle i(interpreted as the probability of selecting a feature), $pbest_i$= personal best position of particle i, $gbest$= global best position among all particles, $w = 0.72$= inertia weight, $c_1 = c_2 = 1.49$= cognitive and social learning coefficients, r_1, r_2= random numbers uniformly distributed in [0,1]
Binary feature selection is determined using a threshold of 0.5 applied to x_i. The algorithm iterates for 10 epochs, progressively refining both personal and global best positions to converge on an optimal feature subset.

Classification and Evaluation

After the PSO algorithm converges, the optimal feature subset, referred to as selected_features, is extracted. Two Random Forest classifiers are subsequently trained: one using all available features and another using only the features selected by PSO. An 80:20 stratified train-test split is applied to ensure that the class distribution remains balanced in both training and testing sets. The performance of each model is evaluated in terms of classification accuracy, number of features used, and execution time. Comparing the baseline model trained on all features with the optimized model trained on the PSO-selected subset provides a quantitative measure of the contribution of PSO to both accuracy improvement and dimensionality reduction.

Visualization and Result Analysis

The convergence behavior and feature selection process are visualized to facilitate interpretability. An accuracy versus iteration plot illustrates the progressive improvement in model performance as PSO iterates, while the feature selection history summarizes the fitness value, accuracy, and number of selected features across all iterations. All outputs, including the PSO history, the final selected features, and result plots, are automatically saved to the designated output directory, ensuring reproducibility and transparency of the analysis.

3. RESULTS AND DISCUSSIONS

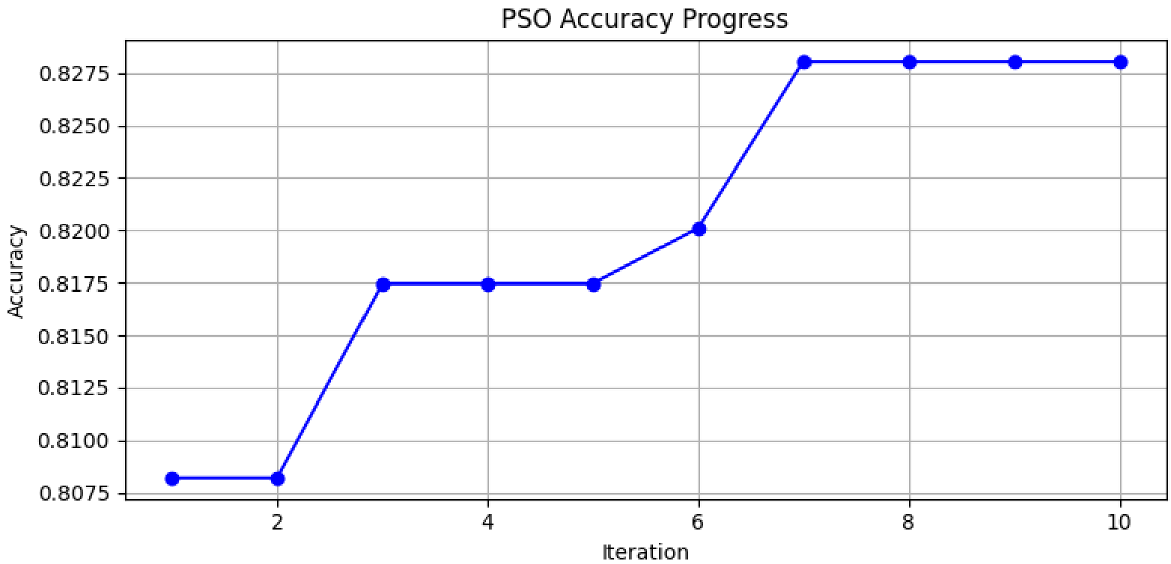

Figure 1 PSO Accuracy Progress

The Particle Swarm Optimization (PSO) feature selection framework was applied to the preprocessed dataset containing 743 numerical voice features, with *class* as the target variable. Baseline classifiers Logistic Regression, Random Forest, and Gradient Boosting achieved mean accuracies of 0.7738, 0.8651, and 0.8690, respectively. After PSO optimization, the dimensionality was reduced to 382 features while maintaining high predictive accuracy. As shown in Figure 1 (PSO Accuracy Progress), the classification accuracy increased steadily over iterations, reaching a stable value around 0.8280 after the 7th iteration, indicating strong convergence behavior.

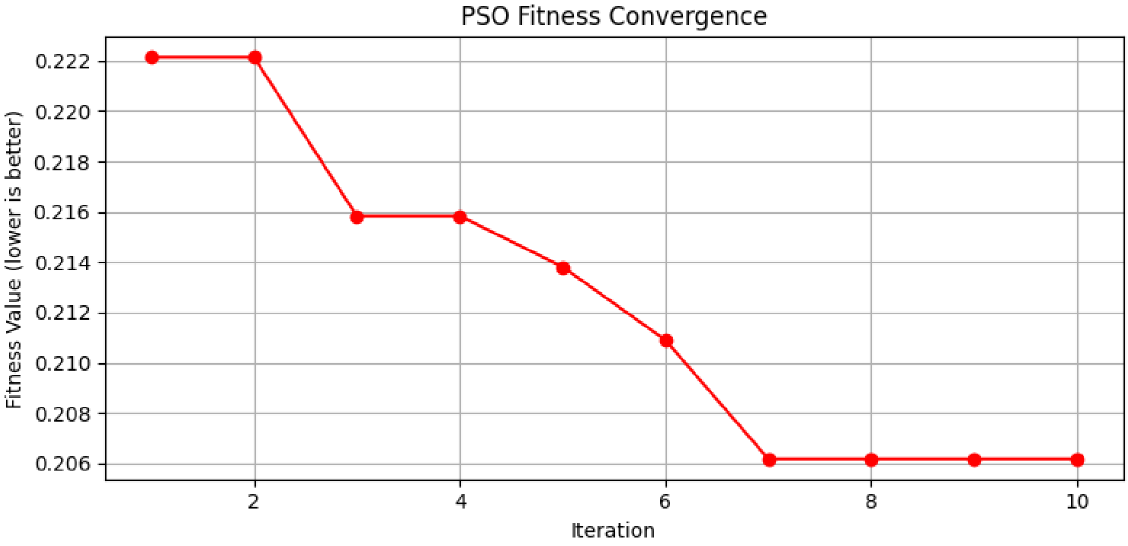

Figure 2 PSO Fitness Convergence

The optimization trend illustrated in Figure 2 (PSO Fitness Convergence) demonstrates a continuous decrease in the fitness value, confirming the swarm's ability to minimize the objective function efficiently. The lowest fitness value of 0.2062 was achieved at iteration 10, corresponding to the optimal feature subset. The final model attained a test accuracy of 0.8421 using the PSO-selected features, slightly outperforming the 0.8355 accuracy from the full feature set, showing that PSO successfully balances dimensionality reduction with model performance.

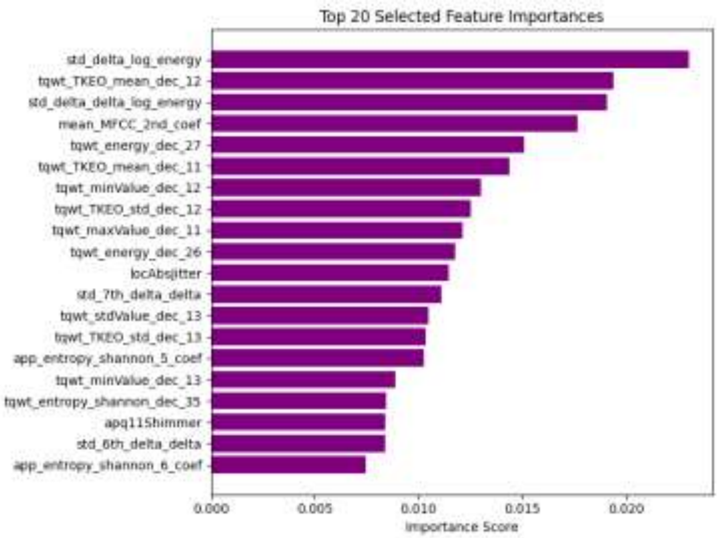

Figure 3 Top 20 Selected Feature Importance

Finally, Figure 3 (Top 20 Selected Feature Importances) reveals that the most influential features predominantly belong to well-established Parkinsonian biomarkers, including log energy derivatives, Teager–Kaiser Energy Operators (TKEO), Mel-Frequency Cepstral Coefficients (MFCCs), Shimmer, and entropy-based measures. These results demonstrate that the PSO algorithm not only enhances computational efficiency but also preserves clinical interpretability by emphasizing physiologically meaningful variables related to Parkinson's disease.

Discusion

The results indicate that the application of Particle Swarm Optimization (PSO) as a feature selection technique substantially improves the efficiency and performance of Parkinson's disease classification. The consistent accuracy improvement observed across iterations (Figure 1) and the convergence pattern in fitness reduction (Figure 2) reflect the robust exploration–exploitation balance inherent in the PSO algorithm. By reducing the feature set from 743 to 382 without sacrificing model performance, PSO effectively eliminates redundant or noisy attributes that may hinder classifier generalization. The enhanced test accuracy of 0.8421, compared to 0.8355 from the full feature set, highlights PSO's capacity to achieve optimal dimensionality reduction while maintaining or even improving predictive capability. This finding corroborates prior research emphasizing the strength of swarm-based metaheuristics in optimizing feature selection for high-dimensional biomedical datasets.

Moreover, the analysis of the top 20 selected features (Figure 3) provides valuable clinical insights into the underlying pathology of Parkinson's disease. The prominence of features such as log energy derivatives, TKEO, MFCCs, Shimmer, and entropy-based measures aligns with established literature identifying these attributes as significant acoustic biomarkers for distinguishing Parkinsonian speech patterns. This outcome underscores the dual advantage of PSO: it not only refines model accuracy but also enhances interpretability by prioritizing physiologically meaningful variables. Such results demonstrate the potential of integrating PSO-driven feature selection with advanced classifiers to develop computationally efficient, clinically interpretable diagnostic tools. These findings contribute to the growing body of evidence supporting the application of evolutionary computation in medical data analysis, offering a promising avenue for precision diagnostics in neurodegenerative disorders.

4. CONCLUSION

This study demonstrates that the proposed Swarm-Driven Automatic Feature Selection and Classification Framework effectively integrates Particle Swarm Optimization (PSO) with ensemble learning to identify a compact and clinically meaningful subset of voice features for Parkinson's disease (PD) detection. By reducing dimensionality by nearly 50% from 743 to 382 features while maintaining or slightly improving predictive accuracy (0.8421 compared to 0.8355 using all features), the framework proves its capacity to enhance computational efficiency, scalability, and interpretability in biomedical voice analysis. These findings highlight PSO's potential as an adaptive and data-driven tool for optimizing complex, nonlinear biomedical datasets, particularly when clinical interpretability is essential. However, the framework exhibited relatively lower specificity in classifying healthy subjects and was tested on a single-center dataset, which may limit generalizability. Future research should address these limitations by incorporating larger and more diverse multi center datasets, integrating class-balancing strategies such as Synthetic Minority Over-sampling Technique (SMOTE), and exploring hybrid PSO variants or cost-sensitive ensemble models to further improve sensitivity specificity balance. Ultimately, this work contributes a reproducible and extensible foundation for the development of automated, clinically interpretable, and scalable AI-assisted systems for early Parkinson's disease diagnosis.

REFERENCES

Al-Shammary, D., Albukhnefis, A. L., Alsaeedi, A. H., & Al-Asfoor, M. (2022). Extended particle swarm optimization for feature selection of high-dimensional biomedical data. *Concurrency and Computation: Practice and Experience*, *34*(10), e6776. https://doi.org/10.1002/cpe.6776

Aldossary, M. (2025). Q-MobiGraphNet: Quantum-Inspired Multimodal IoT and UAV Data Fusion for Coastal Vulnerability and Solar Farm Resilience. *Mathematics*, *13*(18), 3051.

Ashok, R. S., & Anil, K. D. (2025). Machine learning-based early detection of Parkinson's disease using handwriting and vocal features. *Research on Engineering Structures and Materials*. https://doi.org/10.17515/resm2025-835ml0422rs

Bashir, S., Khattak, I. U., Khan, A., Khan, F. H., Gani, A., & Shiraz, M. (2022). A Novel Feature Selection Method for Classification of Medical Data Using Filters, Wrappers, and Embedded Approaches. *Complexity*, *2022*(1), 8190814. https://doi.org/10.1155/2022/8190814

Bhuyan, H. K., Chakraborty, C., Pani, S. K., & Ravi, V. (2021). Feature and subfeature selection for classification using correlation coefficient and fuzzy model. *IEEE Transactions on Engineering Management*, *70*(5), 1655–1669.

Chen, C. W., Tsai, Y. H., Chang, F. R., & Lin, W. C. (2020). Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. *Expert Systems*,

37(5), e12553. https://doi.org/10.1111/exsy.12553
Dhanka, S., Sharma, A., Kumar, A., Maini, S., & Vundavilli, H. (2025). Advancements in Hybrid Machine Learning Models for Biomedical Disease Classification Using Integration of Hyperparameter-Tuning and Feature Selection Methodologies: A Comprehensive Review. *Archives of Computational Methods in Engineering*, 1–36. https://doi.org/10.1007/s11831-025-10309-5
Dixit, S., Bohre, K., Singh, Y., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2023). A Comprehensive Review on AI-Enabled Models for Parkinson's Disease Diagnosis. *Electronics (Switzerland)*, *12*(4), 783. https://doi.org/10.3390/electronics12040783
Gad, A. G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering*, *29*(5), 2531–2561. https://doi.org/10.1007/s11831-021-09694-4
Gawali, A. (2024). *Voice Analysis for Disease Screening*. Sant Gadge Baba Amravati University, Amravati.
Govindu, A., & Palwe, S. (2022). Early detection of Parkinson's disease using machine learning. *Procedia Computer Science*, *218*(2022), 249–261. https://doi.org/10.1016/j.procs.2023.01.007
Iyer, A., Kemp, A., Rahmatallah, Y., Pillai, L., Glover, A., Prior, F., Larson-Prior, L., & Virmani, T. (2023). A machine learning method to process voice samples for identification of Parkinson's disease. *Scientific Reports*, *13*(1), 1–9. https://doi.org/10.1038/s41598-023-47568-w
Karabayir, I., Goldman, S. M., Pappu, S., & Akbilgic, O. (2020). Gradient boosting for Parkinson's disease diagnosis from voice recordings. *BMC Medical Informatics and Decision Making*, *20*(1), 228. https://doi.org/10.1186/s12911-020-01250-7
Kavya, S., Viswanathan, P., Perumal, R., & Charan, S. (2022). Impact of communication difficulty on the quality of life in individuals with Parkinson's disease. *Annals of Movement Disorders*, *5*(1), 49–54. https://doi.org/10.4103/AOMD.AOMD_45_21
Khurma, R. A., Aljarah, I., Sharieh, A., Elaziz, M. A., Damaševičius, R., & Krilavičius, T. (2022). A Review of the Modification Strategies of the Nature Inspired Algorithms for Feature Selection Problem. *Mathematics*, *10*(3), 464. https://doi.org/10.3390/math10030464
Kunapuli, G. (2023). *Ensemble Methods for Machine Learning*. Simon and Schuster.
Nagra, A. A., Khan, A. H., Abubakar, M., Faheem, M., Rasool, A., Masood, K., & Hussain, M. (2024). A gene selection algorithm for microarray cancer classification using an improved particle swarm optimization. *Scientific Reports*, *14*(1), 19613. https://doi.org/10.1038/s41598-024-68744-6
Nazir, A., Hussain, A., Singh, M., & Assad, A. (2025). Deep learning in medicine: advancing healthcare with intelligent solutions and the future of holography imaging in early diagnosis. *Multimedia Tools and Applications*, *84*(17), 17677–17740. https://doi.org/10.1007/s11042-024-19694-8
Quamar, D., Ambeth Kumar, V. D., Rizwan, M., Bagdasar, O., & Kadar, M. (2025). Voice-Based Early Diagnosis of Parkinson's Disease Using Spectrogram Features and AI Models. *Bioengineering*, *12*(10), 1052. https://doi.org/10.3390/bioengineering12101052
Rabie, H., & Akhloufi, M. A. (2025). A review of machine learning and deep learning for Parkinson's disease detection. *Discover Artificial Intelligence*, *5*(1), 24. https://doi.org/10.1007/s44163-025-00241-9
Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/bjml/2024/007
Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, *10*, 10031–10061. https://doi.org/10.1109/ACCESS.2022.3142859
Shao, J., Feng, J., Li, J., Liang, S., Li, W., & Wang, C. (2023). Novel tools for early diagnosis and precision treatment based on artificial intelligence. *Chinese Medical Journal Pulmonary and Critical Care Medicine*, *1*(3), 148–160. https://doi.org/10.1016/j.pccm.2023.05.001
Sheng, J., Amankwah-Amoah, J., Khan, Z., & Wang, X. (2021). COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions. *British Journal of Management*, *32*(4), 1164–1183. https://doi.org/10.1111/1467-8551.12441
Sowan, B., Eshtay, M., Dahal, K., Qattous, H., & Zhang, L. (2023). Hybrid PSO feature selection-based association classification approach for breast cancer detection. *Neural Computing and Applications*, *35*(7), 5291–5317. https://doi.org/10.1007/s00521-022-07950-7

Zaini, F. A., Sulaima, M. F., Razak, I. A. W. A., Zulkafli, N. I., & Mokhlis, H. (2023). A Review on the Applications of PSO-Based Algorithm in Demand Side Management: Challenges and Opportunities. *IEEE Access*, *11*, 53373–53400. https://doi.org/10.1109/ACCESS.2023.3278261