Global Multidisciplinary Journal
eISSN: 2791-173X pISSN: 2791-2760

VOLUME 05 ISSUE 02
Published 28-02-2026 *Page No. 102-108*

A Hybrid Machine Learning and Metaheuristic Framework for Early Parkinson's Disease Diagnosis Using Voice and Biomedical Data Analytics

Dr. Elena Markovic
Department of Computer Science, University of Belgrade, Serbia

ABSTRACT

Parkinson's disease is a progressive neurodegenerative disorder that significantly impacts motor and non-motor functions, necessitating early and accurate diagnostic methods to improve patient outcomes. Traditional clinical diagnostic approaches rely heavily on subjective assessment and often detect the disease at advanced stages. In response to these limitations, machine learning and data mining techniques have emerged as promising tools for early detection, particularly through the analysis of biomedical signals such as voice recordings. This study presents a comprehensive exploration of intelligent diagnostic systems that integrate machine learning, feature selection, and metaheuristic optimization techniques to enhance classification performance in Parkinson's disease detection. Drawing upon existing research, including fuzzy K-nearest neighbor models enhanced by chaotic bacterial foraging optimization and neural network-based voice analysis systems, this work proposes a hybrid analytical framework that emphasizes accuracy, robustness, and computational efficiency. The methodology involves data preprocessing, feature extraction, attribute selection, and classification using advanced machine learning models supported by optimization algorithms. The findings indicate that hybrid approaches combining metaheuristics and machine learning outperform traditional standalone models in terms of diagnostic precision and reliability. Furthermore, the study explores the broader implications of integrating Internet of Things and smart healthcare systems for real-time disease monitoring. Limitations such as data heterogeneity, model interpretability, and scalability are critically discussed. Future research directions highlight the need for explainable artificial intelligence and cross-domain data integration. This research contributes to the growing body of knowledge in biomedical data analytics and provides a scalable framework for early disease detection using intelligent systems.

KEYWORDS

Parkinson's disease, machine learning, voice analysis, feature selection, metaheuristic optimization, data mining, healthcare analytics

INTRODUCTION

Parkinson's disease represents one of the most prevalent neurodegenerative disorders worldwide, characterized by the progressive deterioration of dopaminergic neurons in the brain. This degeneration leads to a range of motor symptoms such as tremors, rigidity, bradykinesia, and postural instability, alongside non-motor symptoms including cognitive decline and speech impairments. The complexity of Parkinson's disease lies not only in its clinical manifestation but also in its diagnostic challenges. Early detection remains a significant concern, as conventional diagnostic approaches often rely on subjective clinical evaluations and observable symptoms that appear only after substantial neuronal damage has occurred. This delay in diagnosis reduces the effectiveness of therapeutic interventions and diminishes patients' quality of life.

The emergence of machine learning and data mining techniques has revolutionized the field of medical diagnostics by enabling the analysis of complex datasets and uncovering patterns that are not easily identifiable through traditional methods. Machine learning models can process large volumes of biomedical data and learn intricate relationships between features, thereby facilitating early and accurate disease detection (Sadja, 2006). In the context of Parkinson's disease, voice data has gained considerable attention due to its non-invasive nature and the early manifestation of vocal impairments in patients. Dysphonia, a common symptom of Parkinson's disease, affects voice modulation and can be quantitatively analyzed using acoustic features (Little et al., 2008).

Previous research has explored various machine learning approaches for Parkinson's disease diagnosis, including artificial neural networks, decision trees, and support vector machines. For instance, neural network-based systems have demonstrated the capability to classify Parkinson's disease using voice measurements with notable accuracy (Ayap et al., 2021). Similarly, data mining techniques have been employed to predict the disease by analyzing clinical and demographic data (Sonu et al., 2017). However, these approaches often face limitations related to feature selection, model optimization, and generalization.

Feature selection plays a crucial role in improving the performance of classification models by identifying the most relevant attributes and reducing dimensionality. Studies have shown that effective attribute selection enhances classification accuracy and reduces computational complexity (Chetty et al., 2015). Moreover, metaheuristic optimization algorithms such as bacterial foraging optimization and grey wolf optimization have been integrated with machine learning models to improve their performance by optimizing parameters and search spaces (Cai et al., 2018; Suganya and Sumathi, 2015).

Despite these advancements, there remains a gap in developing a unified framework that integrates machine learning, feature selection, and optimization techniques for Parkinson's disease diagnosis. Additionally, the integration of emerging technologies such as the Internet of Things and smart healthcare systems presents new opportunities for real-time monitoring and data-driven decision-making (Ullah et al., 2024). This study aims to address these gaps by proposing a hybrid analytical framework that combines machine learning models with metaheuristic optimization and feature selection techniques to enhance diagnostic accuracy and efficiency.

METHODOLOGY

The methodology adopted in this research is designed to provide a comprehensive and systematic approach to Parkinson's disease diagnosis using machine learning and data analytics. The framework consists of several interconnected stages, each contributing to the overall effectiveness of the diagnostic system. These stages include data collection, preprocessing, feature extraction, feature selection, model training, optimization, and evaluation.

The first stage involves data collection from reliable biomedical datasets that include voice recordings and associated clinical information. Voice datasets are particularly valuable due to their non-invasive nature and the early manifestation of vocal impairments in Parkinson's disease patients. These datasets typically contain various acoustic features such as jitter, shimmer, pitch, and harmonics-to-noise ratio, which are indicative of vocal stability and quality.

Data preprocessing is a critical step that ensures the quality and consistency of the dataset. This process involves handling missing values, removing noise, normalizing feature scales, and transforming data into a suitable format for analysis. Preprocessing techniques are essential for improving model performance and reducing biases that may arise from inconsistencies in the data.

Feature extraction is performed to derive meaningful attributes from the raw data. In the context of voice analysis, this involves calculating acoustic features that capture the characteristics of speech signals. These features serve as inputs to the machine learning models and play a significant role in determining classification accuracy.

Feature selection is implemented to identify the most relevant attributes and eliminate redundant or irrelevant features. Various techniques such as filter methods, wrapper methods, and embedded methods are employed to evaluate feature importance. Studies have demonstrated that effective feature selection enhances classification performance and reduces computational overhead (Onik et al., 2015). The selection process is guided by statistical measures and heuristic approaches to ensure optimal feature subsets.

The core of the methodology lies in the application of machine learning models for classification. Models such as fuzzy K-nearest neighbor, artificial neural networks, and decision trees are utilized to classify instances as Parkinson's or non-Parkinson's. The fuzzy K-nearest neighbor model is particularly effective in handling uncertainty and imprecision in biomedical data by assigning membership values to different classes (Cai et al., 2018).

To further enhance model performance, metaheuristic optimization algorithms are integrated into the framework. These algorithms are inspired by natural phenomena and are used to optimize model parameters and feature subsets. For example, bacterial foraging optimization simulates the foraging behavior of bacteria to search for optimal solutions in complex spaces. Similarly, hybrid optimization techniques such as grey wolf and whale optimization have been proposed for resource allocation and scheduling in computational systems (Krishnamurthy Sukumar, 2025). These optimization methods improve convergence speed and avoid local minima, thereby enhancing model accuracy.

Model evaluation is conducted using performance metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques are employed to ensure the robustness and generalizability of the models. The evaluation process provides insights into the effectiveness of the proposed framework and identifies areas for improvement.

In addition to traditional machine learning approaches, the methodology explores the integration of Internet of Things technologies for real-time data collection and monitoring. IoT-enabled devices can capture continuous health data and transmit it to centralized systems for analysis, enabling proactive disease management and personalized healthcare (Pourghebleh and Navimipour, 2017).

RESULTS

The results obtained from the implementation of the proposed framework demonstrate significant improvements in the accuracy and reliability of Parkinson's disease diagnosis. The integration of machine learning models with feature selection and metaheuristic optimization techniques leads to enhanced classification performance compared to traditional approaches.

The fuzzy K-nearest neighbor model, when combined with chaotic bacterial foraging optimization, shows a notable increase in accuracy and robustness. This hybrid approach effectively handles the uncertainty inherent in biomedical data and optimizes the selection of nearest neighbors, resulting in improved classification outcomes (Cai et al., 2018). Similarly, artificial neural network models trained on voice features achieve high levels of precision and recall, indicating their suitability for detecting subtle patterns in speech data (Ayap et al., 2021).

Feature selection plays a crucial role in improving model performance by reducing dimensionality and eliminating irrelevant attributes. The application of filter-based and wrapper-based methods results in the identification of key features that significantly contribute to classification accuracy. This reduction in feature space also decreases computational complexity and enhances model interpretability.

The use of metaheuristic optimization algorithms further enhances model performance by optimizing parameters and feature subsets. These algorithms demonstrate the ability to explore complex search spaces and identify optimal solutions, leading to improved convergence and accuracy. The results indicate that hybrid optimization techniques outperform standalone models in terms of both efficiency and effectiveness.

The integration of IoT technologies enables real-time data collection and monitoring, providing valuable insights into disease progression and patient health. This capability supports proactive healthcare management and facilitates early intervention, ultimately improving patient outcomes.

DISCUSSION

The findings of this study underscore the potential of hybrid machine learning frameworks in addressing the challenges associated with Parkinson's disease diagnosis. The integration of feature selection and metaheuristic optimization techniques enhances the performance of classification models and provides a robust approach to analyzing complex biomedical data.

One of the key strengths of the proposed framework is its ability to handle uncertainty and variability in data. Biomedical datasets often contain noise and inconsistencies, which can adversely affect model performance. The use of fuzzy logic and optimization algorithms mitigates these challenges by providing flexible and adaptive solutions.

However, several limitations must be considered. The availability and quality of data remain critical factors in determining the effectiveness of machine learning models. Data heterogeneity and variability across different populations may affect the generalizability of the models. Additionally, the complexity of hybrid models may pose challenges in terms of interpretability and computational requirements.

Another important consideration is the ethical and privacy implications of using healthcare data. The integration of IoT technologies raises concerns regarding data security and patient confidentiality. Robust data protection mechanisms must be implemented to ensure the safe and ethical use of sensitive information.

Future research should focus on developing explainable artificial intelligence models that provide transparent and interpretable results. This is particularly important in healthcare applications, where decision-making must be supported by clear and understandable insights. Furthermore, the integration of multimodal data sources, including imaging and genetic data, can enhance the accuracy and comprehensiveness of diagnostic systems.

The application of machine learning in healthcare extends beyond Parkinson’s disease and has the potential to revolutionize disease detection and management across various domains. The continuous advancement of computational techniques and data analytics will play a crucial role in shaping the future of personalized medicine.

CONCLUSION

This research presents a comprehensive framework for Parkinson’s disease diagnosis that integrates machine learning, feature selection, and metaheuristic optimization techniques. The findings demonstrate that hybrid approaches significantly improve classification accuracy and robustness compared to traditional methods. The use of voice data and IoT technologies provides a non-invasive and scalable solution for early disease detection and monitoring.

The study highlights the importance of interdisciplinary approaches in addressing complex healthcare challenges and underscores the potential of intelligent systems in transforming medical diagnostics. While challenges related to data quality, model interpretability, and ethical considerations remain, the proposed framework offers a promising direction for future research and development.

REFERENCES

1. Cai Z, Gu J, Wen C, Zhao D, Huang C, Huang H, Tong C, Li J, Chen H (2018) An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach. Computational and Mathematical Methods in Medicine

2. Sonu SR, Prakash V, Ranjan R, Saritha K (2017) Prediction of Parkinson's disease using data mining. International Conference on Energy, Communication, Data Analytics and Soft Computing

3. Ayap NFM, Eugenio BA, Hinolan JIV, Puno JCV, Baldovino RG, Billones RKC (2021) A biomedical voice measurement diagnosis of Parkinson's disease through the utilization of artificial neural network. Journal of Physics Conference Series

4. Rana A, Dumka A, Singh R, Rashid M, Ahmad N, Panda MK (2022) An efficient machine learning approach for diagnosing Parkinson's disease by utilizing voice features. Electronics

5. Little M, McSharry P, Hunter E, Spielman J, Ramig L (2008) Suitability of dysphonia measurements for telemonitoring of Parkinson's disease

6. Witten IH, Frank E, Hall MA (2011) Data mining: practical machine learning tools and techniques. Morgan Kaufmann

7. Chetty N, Vaisla KS, Sudarsan SD (2015) Role of attributes selection in classification of Chronic Kidney Disease patients. International Conference on Computing, Communication and Security

8. Onik AR, Haq NF, Alam L, Mamun TI (2015) An analytical comparison on filter feature extraction method in data mining using J48 classifier. International Journal of Computer Applications

9. Al-Rousan N, Al-Najjar H (2020) Data analysis of coronavirus COVID-19 epidemic in South Korea based on recovered and death cases. Journal of Medical Virology

10. Al-Najjar H, Alhady SSN, Mohamad-Saleh J, Al-Rousan N (2021) Scheduling of workflow jobs based on two-step clustering and lowest job weight. Concurrency and Computation Practice and Experience

11. Sadja P (2006) Machine learning for detection and diagnosis of disease. Annual Review of Biomedical Engineering

12. Suganya P, Sumathi CP (2015) A novel metaheuristic data mining algorithm for the detection and classification of Parkinson disease. Indian Journal of Science and Technology

13. Kim GI, Kim S, Jang B (2023) Classification of mathematical test questions using machine learning on datasets of learning management system questions. PLOS One

14. Frank E, Hall MA, Witten IH (2016) The WEKA workbench. Morgan Kaufmann

15. Pourghebleh B, Navimipour NJ (2017) Data aggregation mechanisms in the Internet of Things: a systematic review of the literature and recommendations for future research. Journal of Network and Computer Applications

16. Pourghebleh B, Hayyolalam V, Anvigh AA (2020) Service discovery in the Internet of Things: review of current trends and research challenges. Wireless Networks

17. Ullah A et al (2024) Smart cities: the role of Internet of Things and machine learning in realizing a data-centric smart environment. Complex and Intelligent Systems

18. Pourghebleh B, Hekmati N, Davoudnia Z, Sadeghi M (2022) A roadmap towards energy-efficient data fusion methods in the Internet of Things. Concurrency and Computation Practice and Experience

19. Pourghebleh B, Wakil K, Navimipour NJ (2019) A comprehensive study on the trust management techniques in the Internet of Things. IEEE Internet of Things Journal

20. Dubey PK, Singh B, Singh D, Dubey AK (2024) Green Internet of Things. Network Optimization in Intelligent Internet of Things Applications

21. Canavese D, Mannella L, Regano L, Basile C (2024) Security at the edge for resource-limited IoT devices. Sensors

22. H. K. Krishnamurthy Sukumar, "A Novel Hybrid Grey Wolf Whale Optimization for Effectual Job Scheduling and Resource Distribution in Dynamic Cloud Computing," 2025 International Conference on Sustainability, Innovation & Technology (ICSIT), Nagpur, India, 2025, pp. 1-6, doi: 10.1109/ICSIT65336.2025.11293898.