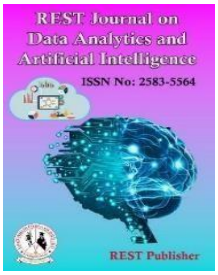
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/62



Predicting Parkinson's Disease using Machine Learning

Ch Lahari, V Leela Manikanta Gopal, N Pavan Kumar, B Kanaka Swamy
Anurag University, Hyderabad, Telangana, India
*Corresponding Author Email: 21eg106b49@anurag.edu.in

Abstract. *Parkinson's disease (PD) is a progressive neurodegenerative disease which affects motor functions of the body, including tremors, rigidity and bradykinesia. Early detection of PD allows effective interventions to be initiated in time to improve the quality of life of patients, but is not easy due to disease symptoms being similar to other disorders. In this work, we propose a machine learning based diagnostic system which attempts to leverage voice measurements along with clinical characteristics in order to improve early diagnosis of PD. We used several machine learning models such as Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN) and Logistic Regression. The combination of genetic risk score along with clinical assessment resulted in better performance, penalized Logistic Regression and XGBoost. As compared to traditional approaches, a noninvasive, telemedicine supported and scaled out diagnostic solution is proposed.*

Keywords: *Parkinson's Disease, Early Detection, Machine Learning, Random Forest, Logistic Regression, XGBoost, Clinical Features, Predictive Modeling.*

1. INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disease of motor function, characterized by tremor, rigidity, and bradykinesia. As the second most common neurological disorder, early diagnosis is still difficult due to lack of definitive biomarkers [1]. Machine learning (ML) is an emerging approach for the prediction of PD, incorporating the use of voice pathology, clinical data and biomarkers from patients. A number of methods including Random Forest (RF), Support Vector Machine (SVM), k-Nearest Neighbors (KNN) and Neural Networks have been explored for early detection of PD and obtained excellent accuracy rates [2]. Quantum computing techniques have also been used to improve computational efficiency in the diagnosis of PD [3]. This study explores the effectiveness of advanced ML models in predictive PD using feature-basedclassification and progression analysis. Through the integration of clinical and speech-based data sets we propose a novel approach that would improve the accuracy and early detection rate and thus could contribute to improving patient care [4], [5].

2. BACKGROUND

1. **Overview of Parkinson's disease:** Parkinson's disease (PD) is a "progressive neurodegenerative disease", which affects millions of people worldwide, in particular people older than 50 years of age. The disease results in 'motor impairments' (tremors, muscle rigidity, bradykinesia), and there is a 'progressive' progression over time with no known cure. The disease is not contagious and therefore the most effective method of early diagnosis is 'clinical evaluation and imaging techniques', because clinical evaluation and imaging techniques are often 'subjective and costly' [1].
2. **Role of Machine Learning:** Machine Learning (ML) has emerged as a promising approach to early diagnosis of PD by selecting biological, clinical and movement-based PD samples from various sources. Numerous studies use support vector machine (SVM), random forest (RF), k-Nearest

Neighbors (KNN) and XGBoost to improve prediction.

3. **Integrating Clinical and Genetic Data**: But challenging challenges remain such as model generalization, dataset variability and reducing false positives for many of these models. Large quality datasets are required to avoid bias and over fitting for many models. Also an effort is needed towards explainable AI techniques to make models interpretable for clinical applications.
4. **Challenges in Prediction:** We still have problems in this respect, including guaranteeing the generalization of models to a large population, dealing with variability in datasets, and controlling false positives. Many models require a large dataset with high quality labelled data or the model cannot be explained correctly for clinical applications.
5. **Future Directions**: Future research is focusing on hybrid ML models, deep learning techniques, and telemedicine applications to enhance PD prediction. The goal is to develop scalable, accessible, and accurate diagnostic tools for early intervention and personalized treatment planning [5].

3. LITERATURE REVIEW

Numerous studies have explored the application of machine learning techniques towards prediction of Parkinson's disease (PD), from traditional statistical models to computationally sophisticated deep learning and quantum computing techniques, which significantly enhance both diagnostic accuracy and speed. A paper (2002), Machine Learning Methods for Predicting Parkinson's Disease Progression, evaluated the predictive ability of Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) approaches to PD classification. The paper conducted clinical, genetic and biomarker features and found that ML approaches significantly improve classification accuracy, but encountered challenges in feature selection and data imbalance, impacting the overall performance of the model [1]. To address these problems, a paper published in 2024 entitled "Machine Learning Approach for Parkinson's Disease Prediction using Quantum Computing Techniques" proposes an ensemble learning method employing Artificial Neural Networks (ANN), Decision Trees, Nave Bayes and KNN that leverages quantum computing as part of the learning mechanism to increase the processing efficiency and classification accuracy. The paper reporting ensemble learning methods proved to be more efficient than conventional ML methods but experienced computational complexity and scalability issues when used in real-world clinical settings [2]. At the same time, recent advances have focused on the integration of multi-modal data sources like voice patterns, handwriting analysis and gait to enhance robustness of PD detection models. Such integration has been shown to have improved predictive performance, especially in early stage diagnosis where conventional clinical measures are still not applicable. The application of deep learning approaches such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to extract high level representations from medical imaging and biosensor data also continues. Although both methods offer high degree of accuracy they are computationally expensive and thus not applicable in real time applications.

TABLE 1. Literature Review

Year	Title of the Paper	Algorithms	Drawbacks
2023	Machine Learning Methods for Predicting Parkinson's Disease Progression	SVM, KNN, Random Forest, Logistic Regression	Feature Selection Challenges, High Dimensional Data
2024	Machine Learning Approach for Parkinson's Disease Prediction through Quantum Computing Techniques	Artificial Neural Networks, Decision Trees, Naïve Bayes, KNN	Computational Complexity, Scalability Issues

4. FINDINGS AND LIMITATIONS

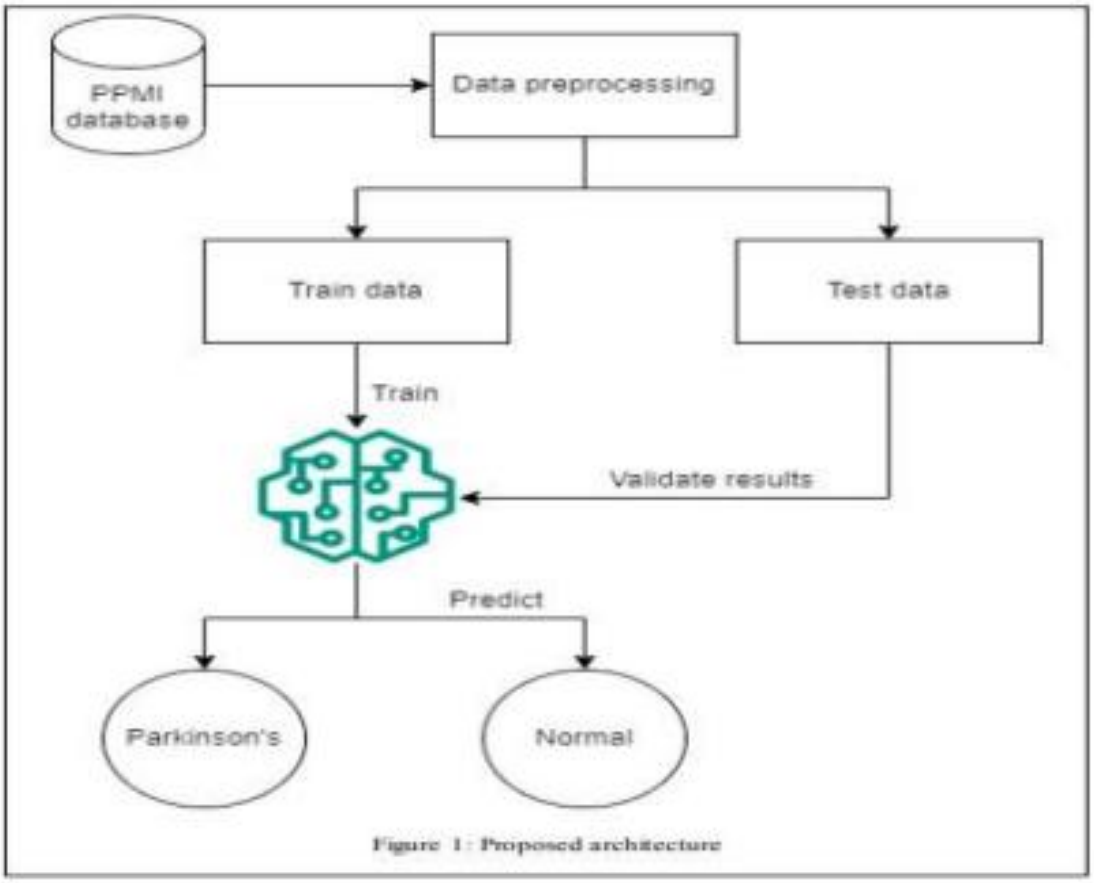


FIGURE 1. Proposed System Architecture

Artificial intelligence techniques have been used to successfully develop modern predictive models of Parkinson's Disease (PD) that achieve high diagnostic accuracy and high early detection. Unfortunately, while many improvements in detecting and preventing PD have been achieved over the past ten years, challenges remain with regard to their practical implementation in clinical settings. These include data variability, computational resource requirements, model interpretability, and deployment limitations in real time. One of the most notable problems with existing ML-based prediction models for PD is their performance across different patient populations. Machine learning algorithms are well designed to generalize well across different datasets, but they often suffer from these limitations when faced with high intra-patient variability, varying disease progression rates, and heterogeneous biomarker data. Clinical symptoms and genetic components contributing to PD vary significantly per individual, therefore it is difficult for models trained on specific datasets to perform consistently across different data sources, and have a limited supply of high quality, labeled data. This means the model could be biased when representing data across different demographics. Turns out experiments such as transfer learning, domain adaptation, and federated learning can be leveraged to further increase generalizability of the models, while also increasing the potential for biases to be avoided if the system is trained on specific datasets. Another significant challenge involves attaining real-time processing while preserving high predictive accuracy. Today, deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) require large computational resources to train and infer from. A number of quantum computing-based approaches have been proposed as a way to improve the accuracy of the processing, but on the other hand, difficulties have still to be solved with respect to hardware scalability as well as to address real-world applications. Many clinical applications rely on using PD detection models on computationally limited resources, either physical edge devices or cloud-based platforms. A number of optimization strategies such as model compression, feature pruning and hardware acceleration have been proposed to improve the

processing efficiency, but at the same time retain the high diagnostic performance. If interpretability and explainability of the machine learning model is unsatisfactory, it will pose a critical barrier in clinical adoption. Since many modern machine learning models, such as XGBoost and deep neural networks, are highly accurate but function as black boxes, it is difficult to understand the reason for the predictions made by them. However, as a result of lack of interpretability the medical community have raised concerns about the trustworthiness of the decisions made by these models. To address the problem, new approaches are being developed such as Explainable AI (XAI), SHAP (Shapley Additive Explanations) and Local Interpretable ModelAgnostic Explanations (LIME). Explainable decision making processes are needed for the diagnosis of PD. Another major limitation is the scalability and adaptability of ML-based prediction models for PD in different healthcare settings. The syndrome manifests as symptoms that steadily deteriorate, and early detection requires extended analysis of longitudinal data from several sources such as clinical assessment, genetic profiling and sensor-based monitoring. Traditional machine learning models are trained on static data and thus cannot adapt to real time patient data. It is being explored in this paper how to use self-supervised learning and reinforcement learning techniques to enable models to learn from continuously evolving patient data and achieve better prediction accuracy.

5. FUTURE SCOPE

Machine learning based prediction models for Parkinson's disease (PD) predictions have improved in accuracy with great progress. New challenges remain, such as using multi-modal datasets, including clinical, genetic and sensor-based data, to improve predictions accuracy. Hybrid learning approaches can be used to bridge the traditional ML and deep learning challenges, while federated learning approaches can allow data to be securely accessed across institutions. Models that are able to update real-time based on the latest patient data can be developed and shared between institutions.

6. CONCLUSION

The system proposes a machine learning-based prediction model for Parkinson's disease using sophisticated models such as Support Vector Machine (SVM), Random Forest (RF) and XGBoost to improve the diagnostic accuracy of patients with the objective of improving disease prognosis, early detection and disease management. Also, feature selection techniques such as PCA and ensemble learning are used to obtain optimal model performance under the limitations of data variability. Although it proved to be effective, the problematic aspects such as heterogeneity in the dataset, computational limitations, and interpretability of PD prediction models remain. Hybrid ML approaches for complex PD prediction could be explored, as well as federated learning to ensure secure sharing of the data in the future, and real-time adaptive models for improving performance of PD prediction.

REFERENCES

[1]. Y. Teletska, V. Trofymenko, O. Vietrov and A. Baiev, "Machine Learning Methods for Predicting Parkinson's Disease Progression," 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 2023, pp. 6-10, doi: 10.1109/ELIT61488.2023.10310787.

[2]. A. K, L. H. R, N. H. K, S. K and Y. C. L, "Machine Learning Approach for Parkinson's disease Prediction through Quantum Computing Techniques," 2024 Second International Conference on Advances in Information Technology (ICAIT), Chikkamagaluru, Karnataka, India, 2024, pp. 1-6, doi: 10.1109/ICAIT61638.2024.10690380.

[3]. T. Bhattacharya, K. T. Thomas and L. Mathew, "Parkinsons Disease Progression Prediction using Advanced Machine Learning Techniques," 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), Greater Noida, India, 2024, pp. 1-5, doi: 10.1109/ICEECT61758.2024.10739044.

[4]. N. Alapati, N. Anusha, P. Joharika, N. J. Jerusha and P. Tanuja, "Prediction of Parkinson's Disease using Machine Learning," 2023 Second International Conference on Electronics and Renewable Systems (ICEARS), Tuticorin, India, 2023, pp. 1357-1361, doi: 10.1109/ICEARS56392.2023.10085443.

[5]. S. S, A. S, G. V. V. Rao, P. V, K. Mohanraj and R. Azhagumurugan, "Parkinson's Disease Prediction Using Machine Learning Algorithm," 2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS), Chennai, India, 2022, pp. 1-5, doi: 10.1109/ICPECTS56089.2022.10047447.