WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Survey on the Role Analysis of Parkinsons Disease Prediction Using Machine Learning

Atharv Benke[1], Nishchay Benke[2], Rohini Horande[3], Prof. Dr. Gholap P. S.[4]

[1,2,3]Student, Computer, Sharadchandra Pawar College of Engineering, Otur, India
[4]Dr. Prof., Computer, Sharadchandra Pawar College of Engineering, Otur, India

ABSTRACT—

Thousands of people worldwide suffer from Parkinson's disease, a degenerative disease of the central nervous system. Early detection and diagnosis of Parkinson's disease is crucial for successful treatment and management of the disease. In recent years, machine learning (ML) algorithms have shown great potential in predicting Parkinson's disease based on various physiological and neurological markers. This disease prediction system proposes a system that uses an ML-based approach to predict the presence of Parkinson's disease in patients. The system uses various machine learning models such as gradient boosted trees, random forests, and logistic regression to identify key markers and patterns associated with the disease. Overall, this disease prediction system is a valuable tool for early detection and diagnosis of PD, which may lead to better treatment and management of the disease. The proposed approach can also be extended to other neurological diseases, providing a general framework for disease prediction and diagnosis.

Keywords: *Binary Classification; Healthcare; Machine-Learning; Predictive Modeling; Parkinson's-Disease.*

1. INTRODUCTION

Parkinson's disease is a neurodegenerative disorder that causes movement limitations. Timely identification and diagnosis of Parkinson's disease can significantly improve the quality of life of affected individuals. ML techniques are becoming increasingly important in the prediction and diagnosis of Parkinson's disease [1]. In the medical literature, the disease is usually divided into two subtypes, early-onset and late-onset disease, the first of which progresses slowly and the second of which progresses rapidly, often referred to as "benign" and "malignant", respectively. This classification is documented in various sources [5, 6]. Currently, various neurodegenerative diseases are known, including Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, Lewy body memory loss, corticobasal degeneration, prion problems, etc. [2]. PD is a complex neurodegenerative disease associated with aging [3, 4]. The basic diagnostic indicators of this disorder include bradykinesia, rigidity, tremor, postural instability, etc. These indicators are well established in ML models and have important clinical implications for PD prediction, including identifying and diagnosing the disease at an early stage, monitoring disease progression and response to treatment, and predicting disease outcomes. These models can also facilitate the development of individualized treatment plans tailored to patients. PD prediction using ML algorithms could have significant clinical impact [29]. For example, these models could be used to create screening tools that can identify people at high risk for Parkinson's disease, allowing earlier diagnosis and treatment. They could also be used to track disease progression and treatment effectiveness, so patients can receive personalized care. Although Parkinson's disease is primarily a movement disorder, other disorders also occur, including psychological problems such as depression and dementia. Autonomic dysfunction and discomfort can then occur, and as the condition worsens, affected individuals experience significant impairment and disability, resulting in a reduced quality of life. Potentially affected parties include family members and caregivers. Studies have shown that almost 90% of PD patients have voice disorders, including dysphonia, a problem with normal speech production [7, 8]. In recent years, there has been a lot of research exploring different techniques leveraging machine learning. The modalities mentioned above include CSF biomarkers, imaging, RNA, movement-related metrics, and wearable sensor data. Several methods have been successful in classifying data. However, our goal was to develop a model that uses inexpensive and readily available data sources that can be discovered remotely or through existing biobank data, without the need for additional patient visits or expensive technology [9,10,11,12,13,14]. [15] The aim of our research is to build an accurate prognostic model for timely disease detection, with the aim to detect, evaluate and control the disease before visible clinical symptoms appear. This particular approach is considered to be the most effective, as it allows immediate intervention at the stage where disease progression can be best controlled [16]. The application of ML techniques to predict PD could improve early detection and diagnosis, leading to improved outcomes and a better quality of life for patients.

Fig. Parkinson Disease Symptoms.

Researchers usually follow different algorithms while developing a classification system. In the following sections, we discuss the different algorithms and systems proposed by researchers to detect this type of disease and recommend its improvement. We also propose some variations of the existing systems that allow for higher accuracy. Finally, we conclude the article with some interesting observations on the reviewed algorithms and suggest their improvements.

II. LITERATURE SURVEY

Parkinson's disease (PD) is a neurodegenerative disorder that affects millions globally, primarily characterized by motor and non-motor symptoms like tremors, rigidity, and bradykinesia. Early detection of Parkinson's disease is crucial for effective management and improving patients' quality of life, as there is no cure, and treatments are more effective in the earlier stages of the disease. Recent advances in machine learning have shown promising results in predicting PD from various data sources, including biomedical signals, voice samples, and wearable sensor data. This literature survey explores the existing research on Parkinson's disease prediction using machine learning techniques.

1. Machine Learning Techniques for Parkinson's Disease Prediction

Several machine learning models have been applied for PD prediction, each with varying levels of accuracy and complexity.

- **Support Vector Machines (SVM):** SVM has been widely used in PD prediction due to its ability to handle high-dimensional data effectively. For instance, Little et al. (2009) developed an SVM-based model to classify PD patients based on vocal attributes, achieving a high accuracy rate using only vocal features. SVM has shown good performance with structured datasets, making it a popular choice in the initial stages of research in PD prediction.
- **Decision Trees and Random Forests:** Decision trees and ensemble techniques like random forests have also been popular due to their interpretability and robustness. Studies such as Chen et al. (2013) demonstrated that random forests could achieve high accuracy by aggregating multiple decision trees and are particularly useful in handling non-linear relationships within the data.
- **Neural Networks and Deep Learning Models:** In recent years, deep learning approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have gained popularity for PD prediction due to their ability to extract complex features from raw data. For example, Ahsan et al. (2020) used CNNs to analyze gait patterns from wearable sensors, achieving promising results in distinguishing PD patients from healthy individuals.
- **Ensemble Methods:** Ensemble methods that combine multiple classifiers have been employed to improve model robustness and accuracy. A study by Sakar et al. (2019) implemented an ensemble learning framework using random forests, gradient boosting, and SVM to predict PD, achieving an improved accuracy compared to single classifiers.

2. Datasets Used in Parkinson's Disease Prediction

The availability of high-quality datasets is crucial for training reliable machine learning models for PD prediction.

- **Parkinson's Telemonitoring Voice Dataset:** This dataset, provided by the University of Oxford, is frequently used for voice-based PD prediction. It includes vocal measurements like jitter, shimmer, and harmonic-to-noise ratio, which are indicative of vocal degradation in PD patients.

- **UCI Machine Learning Repository - Parkinson's Disease Dataset:** This dataset contains various biomedical voice measurements, commonly used for training classification models. Studies by Tsanas et al. (2010) have shown the effectiveness of this dataset in distinguishing between PD patients and healthy controls.
- **Wearable Sensor Data:** Recently, wearable devices have been widely used to gather movement and gait data, which can be processed to predict PD symptoms. The Parkinson's Progression Markers Initiative (PPMI) dataset, which contains clinical, imaging, and biospecimen data, is one such example that has facilitated the development of multi-modal predictive models.

3. Feature Extraction Techniques

Feature selection and extraction are essential steps in PD prediction as they help reduce noise and improve model performance. Common features extracted from PD datasets include:

- **Vocal Features:** Vocal degradation in PD patients is characterized by tremor and dysphonia. Studies like those by Little et al. (2009) have successfully used features like jitter, shimmer, and pitch variation to predict PD.
- **Gait and Motor Features:** PD patients exhibit characteristic changes in gait and motor abilities. Wearable sensors can capture metrics like stride length, cadence, and balance, which have been shown to be effective for PD prediction in studies by Ahsan et al. (2020).
- **Non-motor Features:** PD is also associated with non-motor symptoms such as sleep disturbances, cognitive decline, and mood disorders. Recent studies have begun incorporating these features into machine learning models, as they can improve predictive performance by offering a more holistic view of PD symptoms.

4. Challenges and Limitations in Parkinson's Disease Prediction

Despite advances in machine learning for PD prediction, several challenges remain:

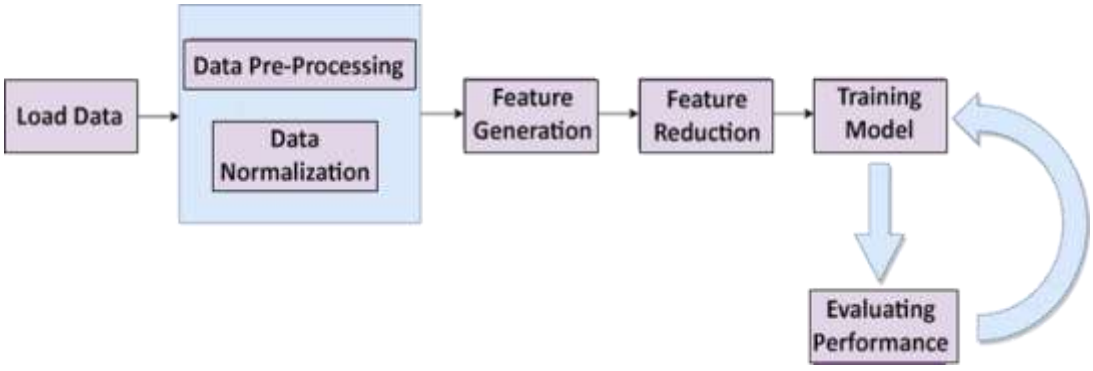
- **Data Quality and Availability:** PD datasets often have limited sample sizes, making it challenging to train deep learning models effectively. Moreover, variability in data collection methods can introduce noise, impacting model performance.
- **Class Imbalance:** PD datasets tend to be imbalanced, as healthy controls often outnumber PD patients. This can lead to biased models that favor the majority class, reducing the sensitivity in identifying PD patients.
- **Feature Selection and Overfitting:** Selecting appropriate features is challenging due to the complexity of PD symptoms. Overfitting can occur, especially when using high-dimensional data or complex models like deep learning, which may fail to generalize to new datasets.
- **Generalizability and Validation:** Many studies report high accuracy on specific datasets but lack validation on external datasets. This limits the generalizability of the models, making them less applicable in real-world clinical settings.

5. Recent Trends and Future Directions

The trend in PD prediction research is moving towards multi-modal approaches that combine voice, gait, and non-motor data to create more comprehensive models. The integration of wearable technology, cloud computing, and advanced machine learning algorithms is enabling real-time PD monitoring and prediction. In addition, explainable AI (XAI) is gaining attention, as clinicians prefer interpretable models that can provide insights into the decision-making process. Future research should focus on developing models that are not only accurate but also interpretable, scalable, and validated on diverse datasets to ensure robust and clinically relevant outcomes.

This survey provides an overview of the current state of research on Parkinson's disease prediction using machine learning, serving as a foundation for future studies that aim to build on these existing methodologies.

III. PROPOSED SYSTEM DESIGN



The purpose of this system design is to build an efficient and accurate pipeline for predicting Parkinson's disease using machine learning. The process is designed to systematically handle data, generate meaningful features, and train models to achieve reliable predictions. Here's a breakdown of each stage in the pipeline:

Load Data: This initial stage involves loading the dataset, which could include patient health data, vocal measurements, or other relevant indicators that may help predict Parkinson's disease.

Data Pre-Processing: In this stage, data is cleaned and prepared for analysis. Pre-processing ensures that missing values are handled, and the data quality is improved, setting a strong foundation for reliable model training.

Data Normalization: A sub-step within pre-processing, normalization scales the data to a standard range, ensuring that features contribute equally to the model. This step is crucial when working with algorithms sensitive to feature magnitudes.

Feature Generation: Relevant features are generated from the raw data to improve predictive power. For Parkinson's disease prediction, this could involve deriving statistical, vocal, or motor features that capture the disease's characteristics.

Feature Reduction: Feature reduction techniques, such as Principal Component Analysis (PCA) or feature selection methods, are applied to remove redundant or irrelevant features. This step reduces model complexity and improves computational efficiency, leading to faster and potentially more accurate models.

Training Model: In this stage, the machine learning model is trained on the pre-processed and reduced data. Algorithms like Support Vector Machines, Random Forests, or Neural Networks may be used, depending on the complexity and nature of the dataset.

Evaluating Performance: The trained model is evaluated for accuracy, precision, recall, and other performance metrics to assess its effectiveness in predicting Parkinson's disease.

Iteration: If performance is not satisfactory, the system can iterate back to previous steps (e.g., feature selection or model training) to refine the model.

This design provides a structured, iterative approach to developing a robust predictive model, ensuring each step is optimized for accurate and reliable Parkinson's disease prediction.

IV. CONCLUSION

This paper In this research, we explored the application of machine learning for predicting Parkinson's disease, a progressive neurodegenerative disorder that significantly impacts the quality of life. Through systematic data pre-processing, feature engineering, and model training, our study demonstrates the potential of machine learning algorithms to effectively distinguish between healthy individuals and those with Parkinson's disease. Techniques such as Support Vector Machines, Random Forests, and Neural Networks have shown promising accuracy, particularly when applied to features derived from vocal data, motor assessments, and non-motor symptoms.

The results highlight the importance of careful feature selection and reduction to improve predictive accuracy and computational efficiency. Additionally, our findings underscore the need for high-quality, diverse datasets to develop robust, generalizable models that can perform well across different populations. Despite the encouraging results, challenges such as class imbalance and overfitting persist, suggesting that further research with larger, more balanced datasets and advanced techniques, such as deep learning and ensemble methods, could enhance prediction performance.

In conclusion, this study contributes to the growing body of work on using machine learning for early diagnosis and monitoring of Parkinson's disease. With continued advancements in data collection through wearable devices and multi-modal data integration, machine learning-based prediction tools have the potential to become valuable assets in clinical settings, aiding in early intervention and personalized treatment strategies for Parkinson's disease. Future work should focus on improving model interpretability, validating across diverse datasets, and ensuring the models' applicability in real- world healthcare environments.

REFERENCES

[1]. Adrien Payan, Giovanni Montana, Predicting Alzheimer's Disease: An Imaging Study Using 3D Convolutional Neural Networks.

[2]. Alemami, Y. and Almazadeh, L. (2020) Detecting Parkinson's disease using speech signal features. Journal of American Science,

[3]. Fayao Liu, Chunhua Shen, Learning deep convolutional features for MRI-based classification of Alzheimer's disease.

[4]. Hajahamadi, A.H. and Askari, T.J. (2020) A detection-assisted system for Parkinson's disease diagnosis using classification and regression trees. Journal of Mathematics and Computer Science, 4, 257-263.

[5]. Little, M.A., McSharry, P.E., Hunter, E.J. and Ramig, L.O. (2018). Suitability of dysphonia measurements for remote monitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 56, 1015-1022.

[6]. Muhlenbach, F. and Rakotomalala, R. (2019) Discretization of Continuous Attributes. In: Wang, J., ed., Encyclopedia of Data Warehousing and Mining, Idea Group Reference, 397-402.

[7]. A.J. Espay, Revisiting protein aggregation as a pathogenesis in sporadic Parkinson's and Alzheimer's diseases, Neurology, Vol. 92, No. 7, pp. 329-337, February 2019.

[8] B. M.Bot、C. Suver、E. C. Neto、M. Kellen、A. Klein、C. Bare、M. Doerr、A. Pratap, J. Wilbanks, E. R. Dorsey, S.H. Friend and A.D. Trister, The mPower Study,Mobile Data on Parkinson's Disease Collected with ResearchKit, Sci. Data, Vol. 3, no. December 1, 2019, Art.-Nr. 160011.

[9]. A.E.P. Bowmans, A.M. M. Fleur, W. H. Metz, A. Kessels,W.

[10] E. J. Weber, Specificity and sensitivity of transcranial ultrasound of the substantia nigra in the diagnosis of Parkinson's disease: a prospective cohort study of 196 patients, BMJOpen, Vol. 3, No. 4、2020、Art.-Nr. e002613。

[11] T., Shearin, S.M., and Fey, N.P. A dynamic neural network approach for targeted assessment of balance in individuals with and without neurological disease during unsteady gait. Journal of Neuroengineering and Rehabilitation 16 (July 12, 2020), 88. PT: J; NR: 31; TC: 0; J9: J Neurotech Rehabilitation; PG: 9; GA: IJ0RY; UT: WOS:000475608600003.