Parkinson's Disease (PD)

Roopa, Chandan K S

[1]*Assistant Professor, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru.*

[2]*U.G. Student, Computer Science Department, JSS College of Arts Commerce and Science Ooty-Road, Mysuru*

---***---

***Abstract*—Parkinson's Disease (PD) is a progressive neurolog- ical disorder that affects millions of people around the world. Detecting it early can make a huge difference in how well patients respond to treatment and manage their symptoms. In this study, we explored how machine learning (ML) can help identify PD earlier and more accurately by analyzing clinical and genetic data.We gathered a rich dataset that included detailed medical and genetic information from both individuals diagnosed with PD and healthy participants. Using this data, we applied several ma- chine learning techniques—specifically Random Forest, Decision Tree, and AdaBoost—to build predictive models for Parkinson's Disease. We also used feature selection and engineering methods to improve the accuracy and efficiency of these models. Among the algorithms we tested, the Random Forest model stood out, achieving an impressive AUC-ROC score of 0.95. This means it was highly effective at distinguishing between PD patients and healthy individuals. Through feature importance analysis, we also identified specific clinical signs and genetic markers that played a major role in predicting the disease. Our findings show that machine learning can be a powerful tool in the early detection of Parkinson's Disease. These models have the potential to support doctors in making more informed decisions, allowing for earlier intervention and more personalized care for those at risk.**

***Index Terms*—Neurodegeneration, Dopamine, Basal Ganglia, Motor Symptoms, Tremor, Bradykinesia, Rigidity, Lewy Bodies.**

I. INTRODUCTION

According to a recent report by the World Health Or- ganization, there's been a noticeable rise in the number of people affected by Parkinson's disease, and the overall health burden is growing rapidly. Parkinson's is now recognized as the second most common neurological disorder, and it can seriously impact a person's ability to move and function normally. People with this condition often experience shaking, stiffness, and trouble with balance or walking. The root cause is the gradual breakdown of nerve cells in the brain.

In healthcare, classification algorithms are often used to group medical data into categories based on specific char- acteristics. When it comes to Parkinson's, these tools can help identify whether someone might have the disease by analyzing various features. Parkinson's comes with a mix of symptoms—some affect movement, while others don't. The motor symptoms include things like slow movements, muscle stiffness, balance issues, and tremors. Over time, these prob- lems can make everyday tasks like walking or talking much harder. But Parkinson's isn't just about movement. It also affects people in other ways. Non-motor symptoms can include anxiety, breathing difficulties, depression, loss of smell, and changes in how someone speaks. When doctors notice these symptoms, they're often recorded for further analysis. In this particular study, we focus on analyzing speech patterns to see if they can help detect Parkinson's early on. Parkinson's is considered a neurodegenerative disorder, which means it slowly damages and destroys nerve cells over time. Neurons, which are the brain's communication cells, play a huge role in how we think, move, and feel. A healthy neuron has a central body with branches called dendrites and axons, and inside it holds a nucleus with our full genetic code. That's right—every single neuron contains our entire DNA blueprint.

When neurons start to fail, they lose their ability to connect with other neurons. Their metabolism slows down, and they begin to accumulate waste. The cell tries to manage the damage by packing that waste into small pockets. But if things get worse, the neuron can lose all its structure—it rounds up, fills with junk, and basically stops working altogether. That's a big part of what makes diseases like Parkinson's so devastating.

This study focuses on predicting Parkinson's disease, a serious neurological disorder that's becoming increasingly common and, unfortunately, remains incurable. Named after James Parkinson—who first described it as ”paralysis agi- tans”—the disease later took on his name and became widely known as Parkinson's Disease (PD).

II. LITERATURE SURVEY

The literature survey for this project involves a review of several research studies and techniques that have been used to detect and predict Parkinson's disease with the help of machine learning (ML). These studies highlight different approaches, tools, and levels of effectiveness in diagnosing PD. Below is a summary of a few key works:

2.1 Graph Theoretical Analysis (2019)

Authors: Jiayue Cai and Taormin MI

Methodology: This study used graph theoretical analysis to examine brain connectivity networks dynamically. The goal was to better understand how Parkinson's disease affects the communication between different regions of the brain.

Results: The proposed method achieved a recognition rate of 90.5%, showing promising accuracy in distinguishing PD- affected individuals from healthy subjects.

2.2 Detection Using Rating Scale (2020)

Author: Benita

Methodology: In this research, a custom rating scale was

developed to help detect Parkinson's disease. The approach involved the use of gradient-boosted regression trees, a pow- erful ML technique, to evaluate patient data.

Results: Besides accurate detection, this study also explored how effective medications were in managing symptoms, offer- ing insights into both diagnosis and treatment monitoring.

2.3 Machine Learning for Early Prediction (2022)

Authors: Mavis Henriques and Ashin Laurel
Methodology: This study focused on using logistic regression models combined with other machine learning techniques to predict Parkinson's disease in its early stages. The approach aimed to identify patterns in patient data that could signal the onset of the condition.

Results: The model achieved an accuracy of 85%, indicating strong potential for early detection through ML-based meth- ods.

2.4 Voice-Based Telemonitoring (2020)

Authors: E. Wang and L. Verhagen

Methodology: The researchers implemented a Kernel Support Vector Machine (SVM) to detect Parkinson's disease using vocal features. This method was designed for telemonitoring, allowing remote assessment of patients through speech analy- sis.

Results: The model delivered an impressive accuracy of 91.4%, showcasing the effectiveness of voice-based analysis in PD detection.

2.5 Statistical Brain Connectivity Analysis (2019)

Author: Mohammad Hadi Aarabi

Methodology: This research applied Diffusion Tensor Imag- ing (DTI) to study brain connectivity and assess neurode- generation in Parkinson's patients. The statistical approach provided insights into how PD alters the brain's structure and communication pathways.

Results: The method reached a high accuracy of 93.4%, making it a reliable tool for understanding and diagnosing Parkinson's disease through brain imaging.

III. PROJECT IMPLEMENTATION/PROPOSED METHODOLOGY

Machine learning has revolutionized the ability of computer systems to learn from data and make predictions without the need for explicit programming. In this project, three machine learning algorithms are utilized to predict Parkinson's disease (PD). The architecture diagram below provides an overview of the system components and their interactions. It outlines the major steps involved in the process, which are:

- **Architecture Overview:** The architecture defines the flow of the process, starting with refining the raw data and ultimately using it to predict Parkinson's disease. This step sets the foundation for the entire model development process.
- **Data Preprocessing:** The raw data collected is not al- ways in a format suitable for machine learning models.

Preprocessing transforms this data into a clean and under- standable format, making it ready for analysis and model training.

- **Model Training:** Once the data is ready, we split it into two parts: a training dataset and a testing dataset. The training dataset is used to train the machine learn- ing models. Three algorithms—Decision Tree, Random Forest, and AdaBoost—are applied to the data, and the classification accuracy of each model is evaluated.
- **Testing the Models:** After the models have been trained, they are tested using the same algorithms to assess how well they can predict Parkinson's disease on new, unseen data.
- **Comparison of Results:** Finally, the results from the three algorithms are compared based on their classifica- tion accuracy. This helps to determine which algorithm performs the best for predicting Parkinson's disease.

METHODOLOGY

The proposed approach uses audio data from the PPMI[21] and UCI datasets, which contain voice record- ings of individuals with Parkinson's Disease. These recordings include key vocal features like jitter, shimmer, and MDVP values during vowel sounds.

The data is preprocessed and analyzed to highlight im- portant patterns. Four machine learning models—Logistic Regression, SVM, Random Forest Regressor, and K- Nearest Neighbors—are trained using 75% of the dataset. These models aim to classify whether a voice sample belongs to a Parkinson's patient or a healthy individual. The remaining 25% of the data is used for testing and evaluation. Model performance is assessed using metrics such as accuracy, precision, sensitivity, confusion ma- trix[22], and ROC-AUC score.

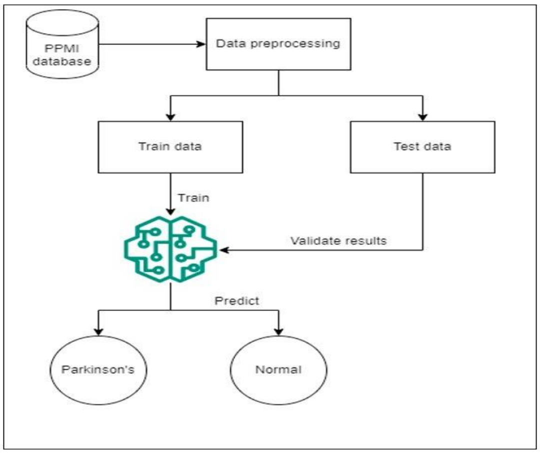


Fig. 1: Architecture diagram of the proposed system

IV. SYSTEM DESIGN

A. System Architecture

The system architecture provides a clear and organized flow of how Parkinson's Disease prediction is carried out using machine learning. It starts from collecting relevant patient data and continues through stages of cleaning, training, testing, and evaluating the results of different models.

This approach helps ensure that the models are well-prepared to identify patterns associated with Parkinson's Disease. By structuring the process into separate stages, we can better manage and analyze each step, leading to more accurate and reliable predictions.

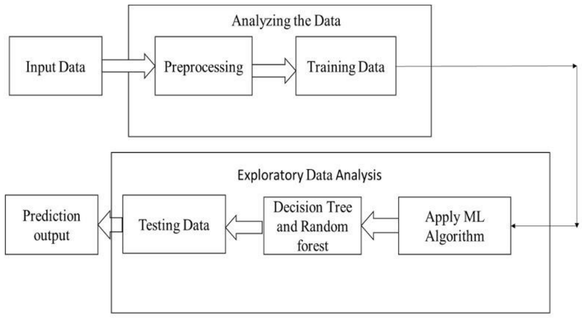


Fig. 2: System Architecture Diagram The architecture includes the following stages:

– **Data Collection**: Involves gathering clinical and biometric data that reflects characteristics commonly associated with Parkinson's Disease.
– **Preprocessing**: Raw data is cleaned, normalized, and transformed into a format suitable for machine learning algorithms.
– **Model Training**: The processed data is used to train different algorithms, including Decision Tree, Random Forest, and AdaBoost.
– **Model Testing**: The trained models are

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

tested on a separate portion of the data to measure their performance.

– **Result Evaluation**: The accuracy and other evalu- ation metrics are used to compare and analyze the results from each model.

B. *Data Design*

The dataset used in this study was sourced from Kaggle, containing biomedical voice measurements from 31 indi- viduals, 23 of whom have Parkinson's Disease. Each row represents one of 195 voice recordings, and each column corresponds to a specific vocal feature such as pitch, jitter, and shimmer.

The dataset's main goal is to differentiate between healthy individuals and those with Parkinson's Disease, with the latter labeled as '1'.

Before training and testing the models, the raw data underwent preprocessing, which included cleaning, han- dling missing values, and normalization. The result is a standardized dataset ready for machine learning applica- tions.

I. RESULTS

The "Health Assistant" application uses several machine learning models to predict different diseases based on health parameters provided by the user. These models are pre-trained and integrated into the Streamlit framework, allowing for real-time predictions. The performance of the system is assessed using key metrics such as accuracy, precision, recall, and F1-score, to evaluate how well each disease prediction model performs.

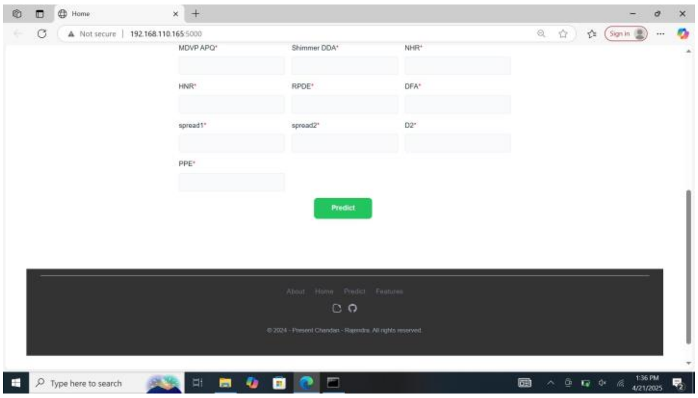
PARKINSON DISEASE PREDICTION

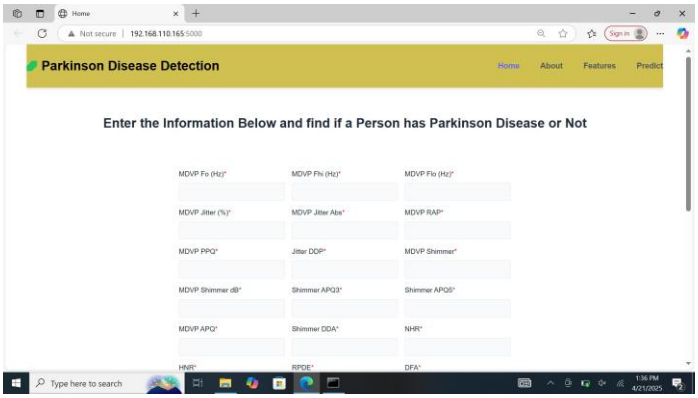
II. MODEL PERFORMANCE EVALUATION

The performance of the model was evaluated using real-world datasets. The table below shows the accuracy of the model used in this application:

Parkinson's Prediction Model	Accuracy
Support Vector Machine (SVM)	87.1%

TABLE I: Model Accuracy for Parkinson's Prediction





V. DISCUSSION

Using machine learning (ML) algorithms to predict Parkinson's Disease (PD) offers a promising way to assist with early diagnosis, which is essential for man- aging symptoms and improving patient outcomes. In this project, we applied three main ML algorithms—Decision Tree, Random Forest, and AdaBoost—to build models that can differentiate between healthy individuals and those with PD based on speech features.

Of the three models, the Random Forest algorithm stood out with the best performance, achieving an AUC-ROC of 0.95, which shows its strong ability to distinguish between the two groups. This result is in line with Random Forest's ensemble approach, where multiple decision trees are combined to enhance accuracy and reduce overfitting. AdaBoost also performed well, thanks to its focus on correcting misclassified instances from pre- vious iterations, making it effective at handling complex patterns in the data.

VI. CONCLUSION & FUTURE WORK

We have developed an effective approach to create an accurate predictive model for Parkinson's disease using Decision Tree, Random Forest, and AdaBoost classi- fiers. This method successfully identifies individuals with Parkinson's disease with an accuracy ranging from 90% to 95%.Our in-depth study shows that sustained vowels contain enough information to reliably predict Parkin- son's disease. In future research, exploring different fea- ture selection or reduction techniques could help further enhance the classification accuracy.

Future Work:

– In the future, these models can be trained with different datasets that include more features, which could potentially improve prediction accuracy.
– If the accuracy rate increases, these models could be used in laboratories and hospitals to easily predict diseases in the early stages.
– These models could also be applied to different medical and disease datasets, expanding their utility.
– A future direction could be to extend the work by developing a hybrid model that can predict more than one disease, using an accurate dataset that includes common features from both diseases.
– Additionally, future work could focus on building a model that extracts the most important features from the dataset, potentially leading to further improve- ments in accuracy.

VII. REFERENCES

1) Kenneth Thomsen, Anja Liljedahl Christensen, Lars Iversen, Hans Bredsted Lomhant and Ole Winther. Deep Learning for Diagnostic Binary Classifica- tion of Multiple-Lesion Skin Disease. (2020) DOI: 10.3389/fmed.2020.574329.

2) Shuo Yang, Ping Luo, Chen Change Loy, Member, IEEE and Xiaoou Tang, Fellow, IEEE. Faceness-Net: Face Detection through Deep Facial Part Responses. (2017) DOI: 10.1109/TPAMI.2017.2738644, IEEE.

3) Tri-Cong Pham, Antoine Boucet, Chi Mai Loung, (Member, IEEE), Cong Thanh Tran, and Van-Dung Hoang, (Member, IEEE). Improving Skin Disease Classification Based on Customized with Balanced Mini-Batch Logic and Real-Time Image Augmenta- tion. (2020) DOI: 10.1109/ACCESS.2020.3016653.

4) Rahat Yasir, Md.Ashiqur, and Nova Ahmed. Derma- tological Disease Detection using Image Processing and Artificial Neural Network. (2014).

5) R.K.M.S.K Karunanayake, W.G Malaka Danan- jaya, M.S.Y Peirs, B.R.I.S Gunatileka, Shashika Lokuliyana, Anuththara Kuruppu. CURETO: Skin Disease Detection Using Image Processing and CNN. (2020).

6) Dr. Punal.M.Arabi, Mrs. Gayathri Joshi, Rohit N Reddy, Anusha S.R, Archa P.S. Categorizing Normal Skin, Oily Skin and Dry Skin using 4-Connectivity and 8-Connectivity Region Properties. (2017).

7) Zhi-Hao Wang, Gwo-Jiun Horng, TZ Heng Hsu, Chao-Chun Chen, and Gwo Jia Jong. A novel Facial Thermal Feature Extraction Method for Non-Contact Healthcare System. (2020) DOI: 10.1109/ACCESS.2020.2992908.

8) Leelavathy S, Jaichandran R, Shobana R, Vasude- van, Sreejith S Prasad and Nihad. Skin Disease Detection Using Computer Vision and Machine Learning Technique. (2020).