MACHINE LEARNING APPROACH FOR PREDICTING PARKINSON'S DISEASE AT EARLY STAGES

1S.Goutham Nayak, 2B.Prabhas, 3G. Sai Nikhil ,4T.Rakesh, 5 Dr K Madan Mohan
1,2,3,4 students, 5 Associate Professor
Department of Information Technology
Guru Nank Institute of Technology

ABSTRACT:

Parkinson's Disease (PD) is a progressive neurological disorder that often causes speech impairments such as dysphonia, making voice analysis a useful non-invasive diagnostic tool. This study proposes a hybrid machine learning framework using a Voting Classifier that combines K-Nearest Neighbors (KNN), Random Forest, and XGBoost to improve detection accuracy and robustness. The model applies feature normalization and dimensionality reduction to enhance performance. By aggregating predictions from multiple classifiers, the system achieves more reliable and interpretable results, supporting early and effective Parkinson's disease detection.

Furthermore, the proposed system addresses common challenges such as class imbalance and feature variability by leveraging ensemble learning techniques. The integration of multiple models ensures better generalization and stability across different datasets. This approach not only improves prediction performance but also provides a scalable and practical solution for real-world clinical applications, assisting healthcare professionals in timely diagnosis and decision-making.

Keywords:

- Parkinson's Disease
- Voice Analysis
- Machine Learning
- Voting Classifier
- Ensemble Learning
- Early Detection

1. INTRODUCTION

Parkinson's Disease results from degeneration of dopamine-producing neurons, causing tremors, stiffness, slowed movements, and impaired speech. Dysphonia is an early indicator, detectable through changes in pitch, loudness, and articulation. Traditional diagnosis is subjective and time-consuming. Machine learning enables automated, non-invasive detection using speech features. This study proposes an ensemble Voting Classifier combining KNN, Random Forest, and XGBoost to improve accuracy and reliability.

1.1 SCOPE OF THE PROJECT

The scope of this project focuses on the early detection of Parkinson's Disease using speech-based features and machine learning techniques. The system is designed to analyze vocal characteristics extracted from recorded speech samples to classify individuals as Parkinson's patients or healthy subjects. It utilizes a Voting Classifier combining KNN, Random Forest, and XGBoost algorithms to enhance prediction performance. The project emphasizes non-invasive, cost-effective, and data-driven diagnosis. This work can be extended to real-time speech monitoring systems, mobile health applications, and clinical decision support tools. Additionally, the model can be adapted to larger and more diverse datasets for improved accuracy. The proposed framework has potential applications in healthcare screening and telemedicine platforms.

1.2 OBJECTIVE

The primary objective of this project is to develop an accurate and reliable machine learning system for the early detection of Parkinson's Disease using speech data. One objective is to extract and analyze relevant vocal features that indicate speech impairments associated with Parkinson's disease. Another objective is to implement the K-Nearest Neighbors algorithm to capture similarity-based patterns in voice data. The project also aims to utilize Random Forest and XGBoost classifiers to improve robustness and handle complex feature relationships. An important objective is to integrate these classifiers using a Voting Classifier to achieve higher accuracy compared to individual models. Feature normalization and preprocessing are performed to enhance model performance. The system aims to reduce misclassification and improve early diagnosis. Another objective is to ensure model interpretability for medical relevance. The project also seeks to provide a scalable and efficient solution suitable for real-world healthcare applications.

2. Existing System & Limitations

Traditional ML models (SVM, Decision Trees, Naive Bayes, Logistic Regression) have been used for PD detection. **Limitations:**

- Poor handling of class imbalance.
- Overfitting on small datasets.
- Assumption of feature independence.
- Limited interpretability.

3. Literature Survey

Several studies have explored machine learning and deep learning approaches for early detection of Parkinson's Disease (PD) using speech signals:

- **PD-Net (2023):** Islam et al. proposed a hybrid deep neural network combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) with a multi-head attention mechanism. Using mel spectrogram and MFCC features, the model achieved **99% accuracy**, demonstrating the effectiveness of spectral feature fusion [1].
- **Ali et al. (2023):** This study applied filter feature selection and genetic algorithms with ensemble learning. Decision Tree and Random Forest achieved **100% accuracy** on one dataset, highlighting the importance of feature selection and ensemble methods in PD detection [2].
- **Bukhari & Ogudo (2024):** An AdaBoost-based classifier trained on the UCI voice dataset achieved strong performance, with **AUC = 0.99** and high precision/recall scores. Cross-validation confirmed robustness across iterations [3].
- **Shyamala & Navamani (2024):** Introduced the Interpretable Feature Ranking XGBoost (IFRX) model, which integrates explainable AI techniques such as SHAP for feature importance ranking. This approach improved interpretability while maintaining high accuracy [4].
- **Rabie & Akhloufi (2024):** Conducted a comprehensive review of ML and DL techniques for PD detection, covering audio, gait, and medical imaging datasets. The study emphasized challenges such as data diversity, privacy, and the need for explainable AI to ensure clinical trust [5].

4. Proposed System

The proposed system is designed to detect Parkinson's Disease at early stages using speech data. Vocal features such as pitch, jitter, shimmer, and harmonic-to-noise ratio are extracted and preprocessed through normalization to ensure consistency. To improve accuracy and robustness, three classifiers—K-Nearest Neighbors (KNN), Random Forest, and XGBoost—are integrated into a Voting Classifier. KNN captures similarity patterns, Random Forest reduces overfitting, and XGBoost handles complex feature interactions and class imbalance. By aggregating their predictions, the ensemble model delivers reliable, interpretable results, making it suitable for clinical use and scalable for applications such as mobile health and telemedicine.

5. METHODOLOGIES

5.1 MODULES NAME:

Modules Name:

- Accumulating Resources
- Examining Insights
- Normalizing Data
- Activating the Model
- Tuning Hyperparameters
- Performance Evaluation
- Calculating Probabilities

5.2. Module Explanation

- **Accumulating Resources:** Collect speech datasets of Parkinson's patients and healthy individuals.
- **Examining Insights:** Extract key vocal features such as jitter, shimmer, and pitch variations.
- **Normalizing Data:** Standardize features to ensure fair comparisons and improve accuracy.
- **Activating the Model:** Apply the Voting Classifier combining KNN, Random Forest, and XGBoost.
- **Tuning Hyperparameters:** Optimize parameters like k-value, tree depth, and learning rate for best performance.
- **Performance Evaluation:** Assess accuracy using metrics such as precision, recall, and F1-score.

- **Calculating Probabilities:** Provide probability scores for predictions to enhance interpretability.

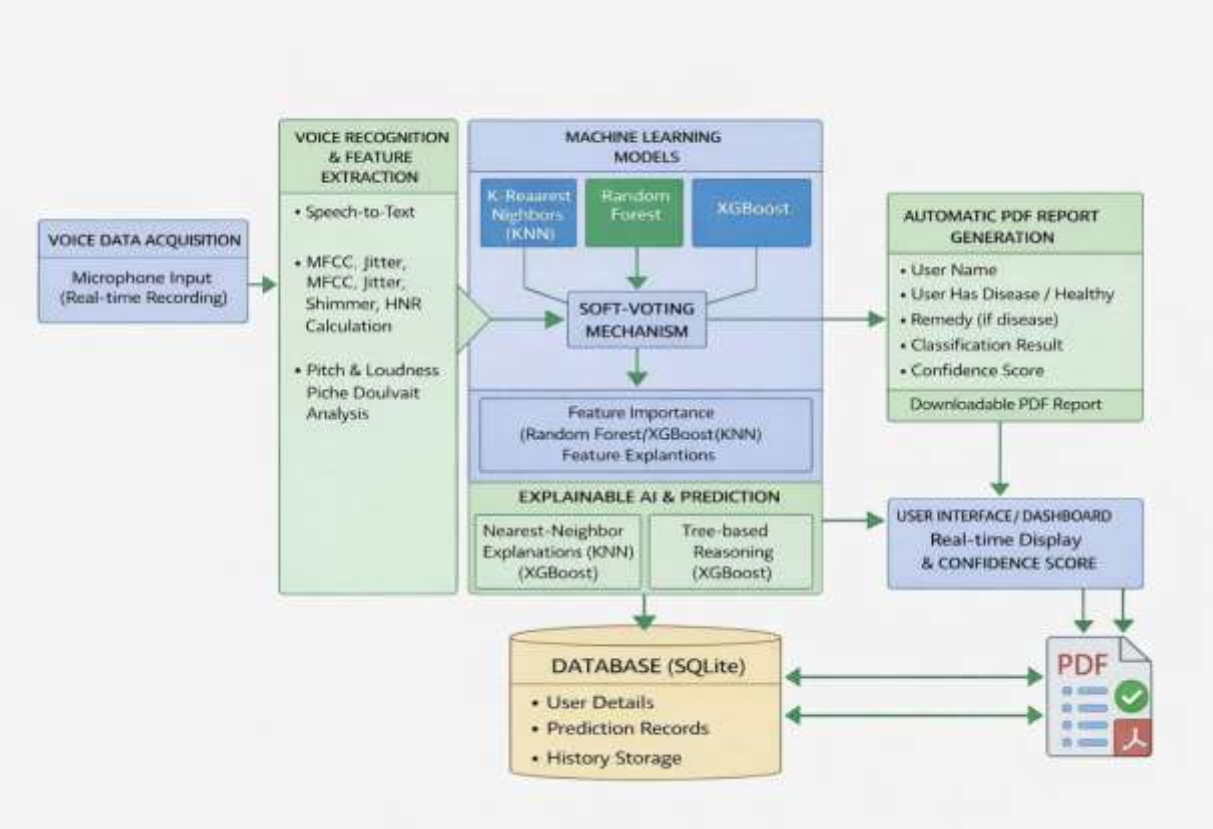
6. Algorithm Used

The proposed system integrates three machine learning algorithms—**K-Nearest Neighbors (KNN), Random Forest, and XGBoost**—to improve detection accuracy and reliability. KNN classifies patients based on similarity patterns in vocal features, Random Forest reduces overfitting through ensemble decision trees, and XGBoost provides high predictive power by handling complex feature interactions and class imbalance. Their outputs are combined using a **Voting Classifier**, which aggregates predictions to deliver robust, consistent, and interpretable results for early Parkinson's Disease detection.

7. Requirements Engineering

- **Hardware Requirements:** Dual Core processor, 4GB RAM, 500GB HDD.
- **Software Requirements:** Windows 10, Python (Spyder IDE).
- **Functional Requirements:** Speech feature extraction, classification using Voting Classifier, probability output for interpretability.
- **Non-Functional Requirements:**
    - *Usability:* Automated process with minimal user intervention.
    - *Reliability:* Stable performance using Python platform.
    - *Performance:* Fast response time with optimized ML algorithms.
    - *Supportability:* Cross-platform compatibility and scalability.

8. SYSTEM ARCHITECTURE:



9. DEVELOPMENT TOOLS

9.1 Python

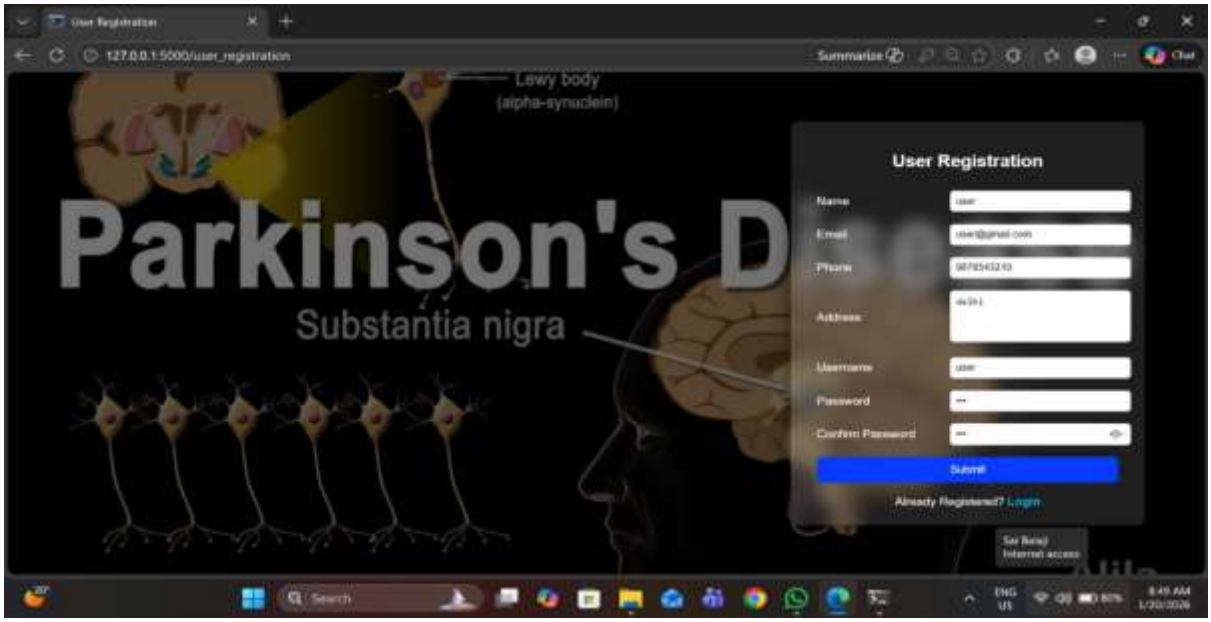
Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

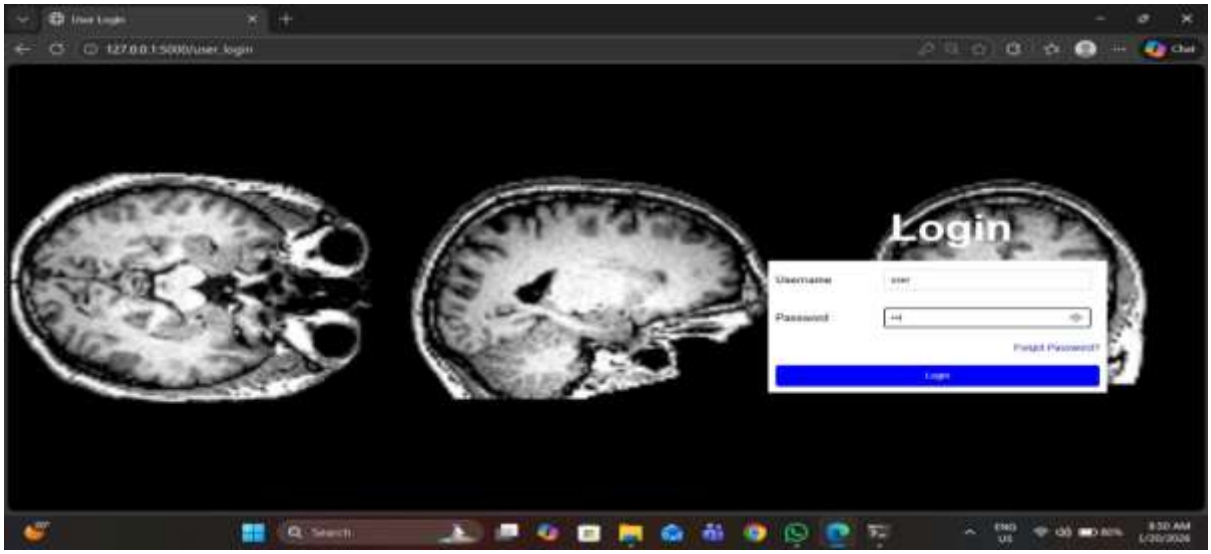
9.2 History of Python

- Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.
- Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.

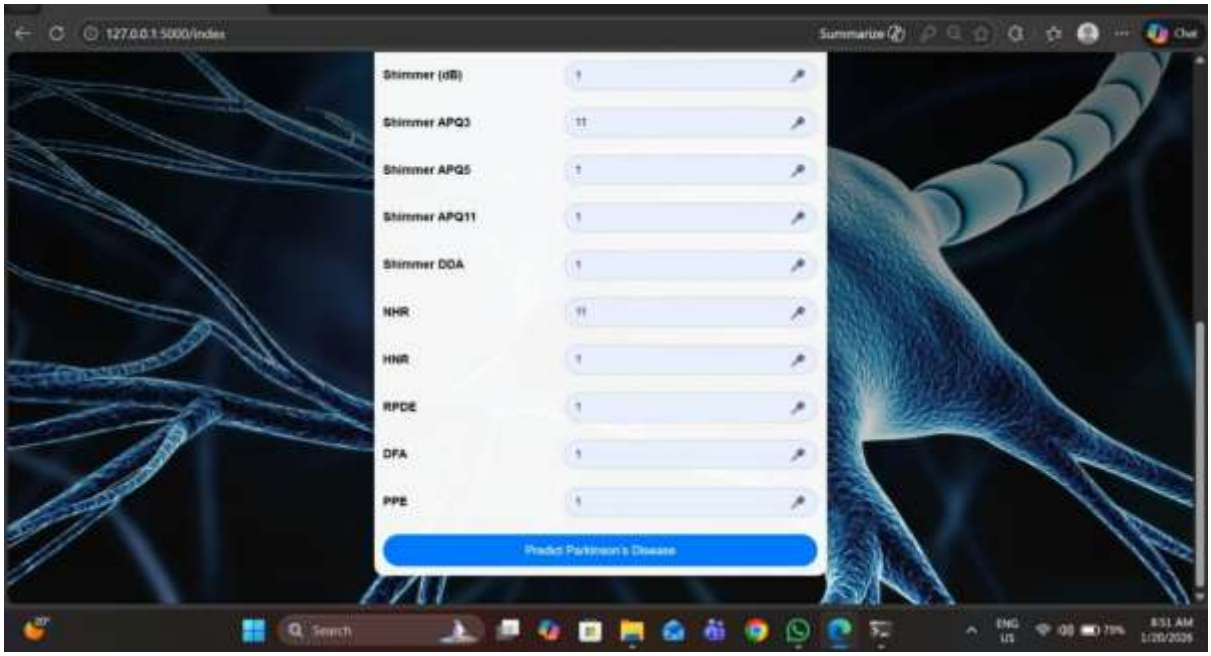
10. RESULTS

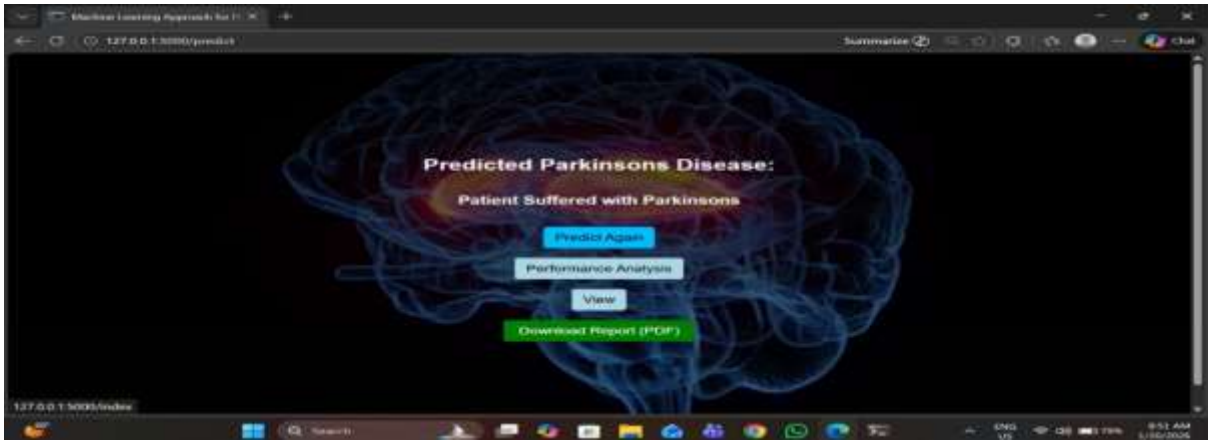
This project is implements like application using python and the Server process is maintained using the SOCKET & SERVERSOCKET and the Design part is played by Cascading Style Sheet.



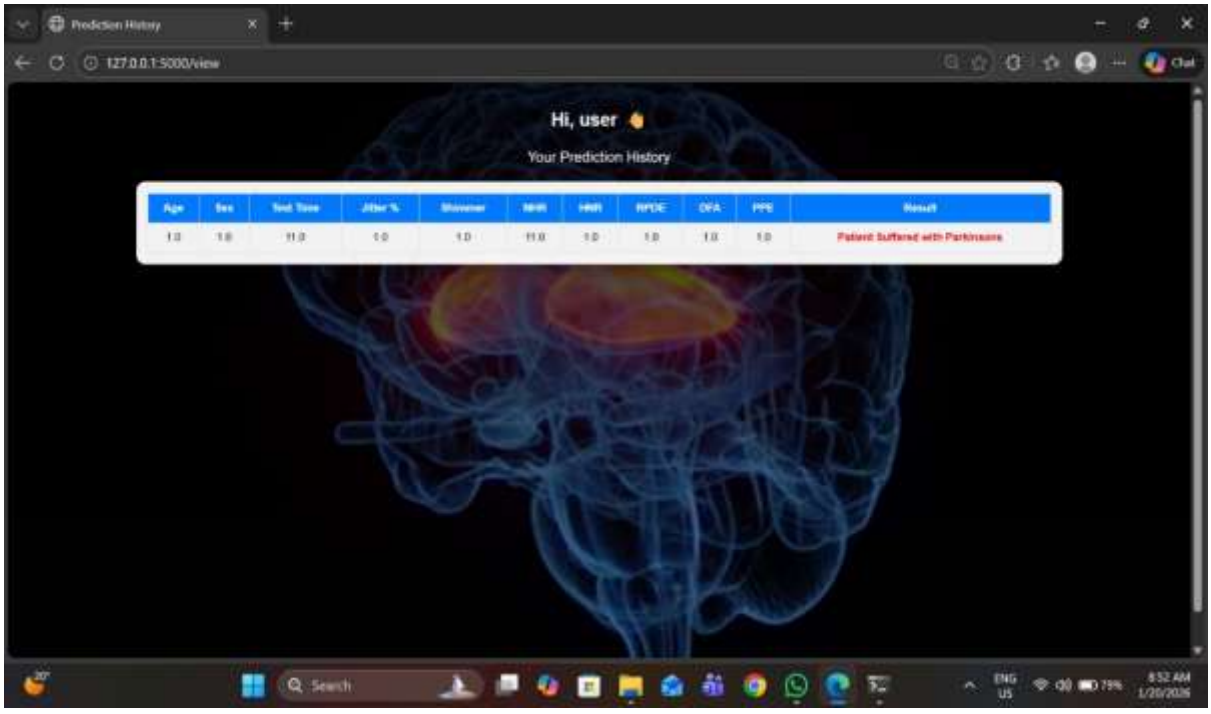

Login
Username
Password
Login

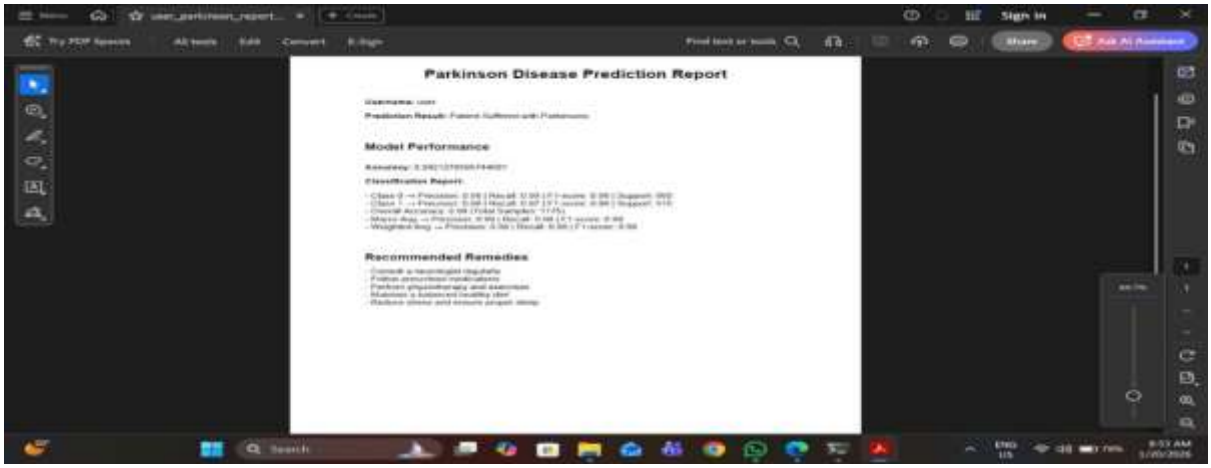

Machine Learning Approach for Predicting Parkinson's Disease
Age
Jitter (%)
Jitter (Abs)
Jitter RAP
Jitter PPQ5
Jitter DDP
Shimmer


Shimmer (dB)
Shimmer APQ3
Shimmer APQ5
Shimmer APQ11
Shimmer DDA
NHR
HNR
RPDE
DFA
PPE
Predict Parkinson's Disease


Predicted Parkinsons Disease:
Patient Suffered with Parkinsons
Predict Again
Performance Analysis
View
Download Report (PDF)


Machine Learning Approach for Predicting Parkinson's Disease at Early Stages


Hi, user
Your Prediction History



11. Future Enhancements

- **Real-time Monitoring:** Extend the system to mobile apps for continuous speech tracking.
- **Telemedicine Integration:** Enable remote diagnosis and support for clinicians.
- **Multilingual Datasets:** Train on diverse languages to improve global applicability.
- **Explainable AI:** Incorporate SHAP/XAI tools for better clinical trust.
- **Scalability:** Adapt framework for larger datasets and cloud deployment.

12. CONCLUSION

This project presents an effective machine learning-based approach for the early detection of Parkinson's Disease using speech signal analysis. Since speech impairment is one of the earliest symptoms of Parkinson's disease, analyzing vocal features provides a non-invasive and cost-effective diagnostic solution. The proposed system utilizes a Voting Classifier that combines K-Nearest Neighbors, Random Forest, and XGBoost algorithms to improve prediction accuracy and robustness. By leveraging the strengths of multiple classifiers, the system overcomes the limitations of individual models and achieves reliable classification results. Proper data preprocessing, feature normalization, and hyperparameter tuning further enhance model performance. Experimental results demonstrate that the ensemble approach provides better accuracy and consistency compared to standalone classifiers. The system offers interpretability and clinical relevance, making it suitable for assisting medical professionals in early diagnosis. Overall, the proposed model contributes to improved healthcare screening and early intervention for Parkinson's Disease. This work highlights the potential of machine learning in supporting intelligent and data-driven medical diagnostic systems.

13. REFERENCES

[l] S. L. Oh, Y. Hagiwara , U. Raghavendra , R. Yuvaraj, N . Arunkumar, M. Murugappan, and U. R. Acharya, "A deep learning approach for Parkinson 's disease diagnosis from EEG signals," Neural Comput. Appl. , vol. 32, no. 15, pp. 10927-10933, Aug. 2020.

[2] C. Loconsole , G. D. Cascarano, A. Brunetti , G. F. Trotta, G. Losavio, V. Bevilacqua , and E. Di Sciascio, "A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis," Pattern Recognit. Lett., vol. 121, pp. 28-36, Apr. 2019 .

[3] b. F. Ertugl-ul, Y. Kaya, R. Tekin, and M. N. Almah, "Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait," Expert Syst. Appl., vol. 56, pp. 156-163, Sep. 2016.

[4] R. Gupta, M. Khari, D. Gupta, and R. G. Crespo, "Fingerprint image enhancement and reconstruction using the orientation and phase reconstruction ," Inf Sci., vol. 530, pp. 201-218, Aug. 2020.

[5] H. M. R. Afzal, S. Luo, M. K. Afzal, G. Chaudhary, M. Khari, and S. A. P. Kumar, "3D face reconstruction from single 2D image using distinctive features," IEEE Access, vol. 8, pp. 180681-180689 , 2020.

[6] R. Raj, P. Rajiv, P. Kumar, M. Khari, E. Verdu , R. G. Crespo, and G. Manogaran , "Feature based video stabilization based on boosted Haar cascade and representative point matching algorithm," Image Vis. Comput., vol. 101, Sep. 2020, Art. no. 103957.

[7] R. Gupta, M. Khari, V. Gupta, E. Verdu, X. Wu , E. Herrera- Viedma, and R. G. Crespo, "Fast single image haze removal method for inhomogeneous environment using variable scattering coefficient ," Comput. Model. Eng. Sci., vol. 123, no. 3, pp. 1175-1192, 2020.

[8] A. Ma, K. K. Lau , and D. Thyagarajan , "Voice changes in Parkinson's disease: What are they telling us?" J. Clin. Neurosci. , vol. 72, pp. 1-7, Feb. 2020.

[9] S. Saravanan, K. Ramkumar, K. Adalarasu , V. Sivanandam , S. R. Kumar, S. Stalin, and R. Amirtharajan , "A systematic review of artificial intelligence (Al) based approaches for the diagnosis of Parkinson's disease," Arch. Comput. Methods Eng., vol. 29, no. 6, pp. 3639-3653, Oct. 2022.

[10] K. A. Shastry, "An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease," H ealthcare Anal ., vol. 3, Nov. 2023, Art. no. 100181.