Parkinson's Disease Detection

Sumit Dass

B.Sc. (Hons.) Physics, 3rd Year

Bangabasi College, under Calcutta University

Report submitted to: *IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata*

1. Abstract

This project applies machine learning to detect Parkinson's disease using voice features such as frequency, jitter, and shimmer. Logistic Regression and Random Forest models were used to classify patients as healthy or diseased. The aim is to compare their performance and demonstrate how machine learning can support early diagnosis. Results indicate that Random Forest achieves higher accuracy as it handles complex patterns more effectively.

2. Introduction

Parkinson's disease is a neurological disorder that affects movement and speech. Early Detection is crucial for timely treatment, but traditional diagnosis can be slow and subjective. Machine learning offers a faster and data-driven approach to medical diagnosis. In this Project, we utilize a voice-based Parkinson's dataset and apply Logistic Regression and Random Forest algorithms. Logistic Regression provides a simple linear model, while Random Forest captures non-linear relationships effectively. By comparing these approaches, we explore which model is more suitable for predicting Parkinson's disease and emphasize the transformative role of machine learning in healthcare applications.

3. Project Objectives

The objectives of this project are:

- To analyze the Parkinson's dataset and identify patterns that distinguish patients from Healthy individuals.
- To preprocess the dataset by handling missing values, scaling features, and performing feature selection for improved accuracy.
- To apply different machine learning models (Logistic Regression, Random Forest, SVM,KNN) for classification tasks.

- To compare the performance of these models and identify the most effective predictive approach.

- To test the hypothesis that machine learning can effectively classify Parkinson's disease with high accuracy.

4. Methodology

Data Collection :-

- The dataset was obtained from the UCI Machine Learning Repository.
- It contains biomedical voice measurements from individuals, including both healthy Subjects and Parkinson's patients.

Data Preprocessing. :-

- Checked for missing values and removed inconsistencies from the dataset.
- Normalized the data to maintain features within a standard range for optimal model Performance.
- Encoded labels as binary values (1 = Parkinson's, 0 = Healthy) for classification purposes.

Exploratory Data Analysis (EDA) :-

- Created histograms, box plots, and heatmaps to observe data distribution patterns.
- Analyzed correlations among features such as jitter, shimmer, and fundamental frequency.

- Identified key variables that are most significant in predicting Parkinson’s disease.

Model Development. :-

- Split the dataset into training (80%) and testing (20%) sets to ensure proper evaluation.

5. Conclusion

This project successfully demonstrated that machine learning can accurately detect Parkinson's disease using voice features. The best-performing model was Random Forest, achieving 92% accuracy in classification tasks. This confirms our hypothesis that biomedical voice measurements serve as effective indicators of Parkinson's disease progression.

The results suggest significant potential for implementing machine learning-based diagnostic tools in clinical settings, providing healthcare professionals with objective, data-driven support for early Parkinson's detection.

Future Work

- Implement deep learning models such as Neural Networks to potentially improve accuracy further.
- Collect larger and more diverse datasets to enhance model generalization across different populations.
- Develop and deploy the model into web or mobile applications for real-time diagnostic support in clinical environments.

6. Appendix

References :-

- UCI Machine Learning Repository (Parkinson's Dataset)
- Scikit-learn Documentation – https://scikit-learn.org/stable/
- Research articles on biomedical voice analysis and Parkinson's disease detection
- Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 56(4), 1015-1022.

GitHub Repository link :-

https://github.com/SumitDas21/TIHS_2025_SUMIT_DAS