Predictive Modeling for Parkinson's Disease Diagnosis Using Machine Learning

Vansh Yadav
2022559
vansh22559@iiitd.ac.in

Utkarsh Dhilliwal
2022551
utkarsh22551@iiitd.ac.in

Shamik Sinha
2022468
shamik22468@iiitd.ac.in

Vaibhav Singh
2022555
vaibhav22555@iiitd.ac.in

Abstract

This project aims to develop a machine learning model for accurate Parkinson's disease prediction. By analyzing historical medical data specific to Parkinson's disease, the project will employ various ML algorithms, including Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Logistic Regression, among others, to identify key risk factors and enhance early diagnostic capabilities. The model will be trained and tested using publicly available datasets related to Parkinson's disease.

1. Motivation

Parkinson's disease is a serious condition with no cure, making early diagnosis essential. Traditional methods can be limited and subjective, which inspired us to explore machine learning as a tool for early detection. Our goal is to build a model that identifies key risk factors and improves diagnostic accuracy, offering a more reliable and accessible way to predict the disease.

We saw potential in machine learning for non-invasive Parkinson's disease prediction while seeking better methods for early detection and diagnosis of diseases like this.

2. Related Work

DOI: **https://www.nature.com/articles/s41598-024-64004-9**
One study found that Feed-forward Neural Networks (FNN) and K-Nearest Neighbor (KNN) models achieved high performance, with FNN reaching 99.11% accuracy on 195 recordings from 31 patients, indicating strong potential for early PD detection.

DOI: **https://www.sciencedirect.com/science/article/pii/S1877050923000078**
Another study used Support Vector Machines (SVM) and Decision Trees to analyze vocal features, achieving 88% accuracy. It highlighted the importance of feature selection for improving detection accuracy.

DOI: **https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10086231/**
A third study reported 98.31% accuracy with Multi-Layer Perceptron (MLP) and 95% with SVM models on the same dataset, employing techniques like SMOTE and GridSearchCV to enhance performance.

3. Timeline

Our team plans to start the project by end-August. We'll first dive into the basics of machine learning required for the project, focusing on data collection and preparation. By the end of September, we aim to have initial models, like Logistic Regression and K-Nearest Neighbours, up and running, with preliminary testing underway.

In October, we will tackle more advanced techniques, such as Random Forests and SVMs. We'll also prepare for the mid-semester presentation and submit a progress report. After that, our focus will shift to refining the models and applying them to real-world data to ensure accuracy and fairness. By late November, we plan to wrap up the project, finalize the report, and get ready for the end-of-semester presentation where we will report our findings.

4. Individual Tasks

- **Shamik Sinha**: Data Collection, Feature Engineering, RF, KNN, SVM, and Logistic Regression models.
- **Utkarsh Dhilliwal**: Data Collection, Feature Engineering, RF, KNN, SVM, and Logistic Regression models.
- **Vansh Yadav**: Data Collection, Feature Engineering, RF, KNN, SVM, and Logistic Regression models.
- **Vaibhav Singh**: Data Collection, Feature Engineering, RF, KNN, SVM, and Logistic Regression models.

5. Final Outcome

As a team, we're aiming to develop a tool that improves early detection of Parkinson's Disease through accurate diagnostics. We expect to leverage our collective skills to tackle the complexities of the project and produce meaningful results.

Each team member will contribute their expertise to build a robust solution, and together, we hope to advance the field and provide valuable insights for both clinical practice and further research.