Detection of Tremor using AI based techniques for early detection of Parkinsons Disease

A project report submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Technology

in

Electronics & Computer Engineering

by

Emad Mohamed (21BLC1543)

Shruti Tiwari (21BLC1179)

School of Electronics Engineering,

Vellore Institute of Technology, Chennai,

Vandalur-Kelambakkam Road,

Chennai - 600127, India.

April 2025

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Declaration

I hereby declare that the report titled ***Detection of Tremor using AI based techniques for early detection of Parkinsons Disease*** submitted by us to the School of Electronics Engineering, Vellore Institute of Technology, Chennai in partial fulfillment of the requirements for the award of **Bachelor of Technology** in **Electronics and Computer Engineering** is a bona-fide record of the work carried out by me under the supervision of ***Dr.Kiruthika V***.
I further declare that the work reported in this report, has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or University.

Sign: _____

Name & Reg. No.: _____

Date: _____

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Electronics Engineering

Certificate

This is to certify that the project report titled ***Detection of Tremor using AI based techniques for early detection of Parkinsons Disease*** submitted by ***Emad Mohamed*** **(21BLC1543) and Shruti Tiwari (21BLC1179)** to Vellore Institute of Technology Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** in **Electronics and Computer Engineering** is a bona-fide work carried out under my supervision. The project report fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Supervisor	**Head of the Department**
Signature:	Signature: .....
Name:	Name: .....
Date:	Date:

Examiner

Signature: .....

Name: .....

Date:

(Seal of the School)

Abstract

Parkinson's Disease (PD) is a progressive neurological condition that gradually impacts motor function, quality of life, and the ability to carry out everyday activities. One of its key symptoms—tremor—often appears on and off, making it difficult to catch in its early stages without continuous tracking. Fortunately, with the rapid advancements in Machine Learning (ML) and Deep Learning (DL), identifying Parkinsonian tremor has become far more effective than relying solely on traditional clinical observations. This research introduces a comprehensive, AI-powered strategy for early tremor detection, with a strong emphasis on feature extraction, optimization techniques, and ensemble-based classification.

The proposed model follows a structured three-stage pipeline: data ingestion, data transformation, and model training. It all begins with collecting raw smartwatch sensor data, which is then carefully preprocessed—this includes handling missing values, applying normalization, and selecting the most meaningful features. By honing in on the most important features and fine-tuning model parameters, the aim is to build a reliable and easy-to-use system for detecting tremors in real time. To make this solution practical and accessible, the trained model is integrated into a web-based application, allowing both users and healthcare providers to upload data and instantly receive feedback on tremor patterns.

In addition to detection, the system goes a step further by including an exercise recommendation module. This feature offers personalized suggestions aimed at managing tremor symptoms and supporting better motor function. Together, these components form a holistic solution—bridging the gap between early detection and ongoing symptom management. The paper walks through every stage of development, from data preparation and feature selection to model evaluation and web deployment. Finally, it discusses the model's performance and how this technology can positively influence the future of PD care and monitoring.

Acknowledgements

We wish to express our sincere thanks and deep sense of gratitude to our project guide, Dr. Kiruthika V, Professor, School of Electronics Engineering, for her consistent encouragement and valuable guidance offered to us in a pleasant manner throughout the course of the project work.

We are extremely grateful to Dr. Ravishankar A, Dean Dr. Reena Monica, Associate Dean (Academics) & Dr. John Sahaya Rani Alex, Associate Dean (Research) of the School of Electronics Engineering, VIT Chennai, for extending the facilities of the School towards our project and for his unstinting support.

We express our thanks to our Head of the Department Dr. Annis Fathima A for her support throughout the course of this project.

We also take this opportunity to thank all the faculty of the School for their support and their wisdom imparted to us throughout the course.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

Contents

List of Figures

Chapter 1

Introduction

The World Health Organization (WHO) categorizes neurological disorders, including Parkinson's Disease (PD), as the leading cause of disability worldwide, affecting more than 10 million people worldwide. PD is also known as a chronic neurodegenerative disorder most defined by its motor symptoms, including tremor, rigidity, and bradykinesia (slowness of movement). Perhaps the most difficult of all of PD to identify is the first tremor, which occurs at irregular intervals and cannot be easily diagnosed unless continuously monitored. The Parkinson's Foundation has indicated that nearly 60,000 new cases of PD are diagnosed each year in the United States alone, and the disease is increasing. PD is a serious risk to the patient and the healthcare system. PD symptoms like tremor and mild loss of movement may be overlooked until the disease has progressed far, and there is little time for early treatment. Delayed diagnosis will result in decreased mobility, cognitive impairment, and decreased quality of life. In addition, if untreated, PD will also result in serious complications like falls, dysphagia, and psychiatric illness like depression and anxiety. Though an incurable, irreversible disease, early diagnosis and proper management can go a long way in slowing its progression and improving patient outcomes.

International health standards stress the need to diagnose PD early so that neuroprotective therapies and lifestyle management can be started, as well as enhance mobility and quality of life. Nevertheless, access to specialist neurological services is usually patchy geographically, especially for those living in rural or under-resourced areas. They often go undiagnosed for years because of lack of awareness, insufficient access to movement disorder specialists, and dependency on subjective clinical exams that do not necessarily identify tremors in the early stages. Such a diagnosis gap emphasizes the immediate need for convenient, objective, and long-term monitoring tools for PD symptoms.

With the fast-paced development of digital health technologies, wearable technology and AI-based apps have become revolutionary tools for monitoring neurological diseases. Current research indicates that the market for digital health is growing exponentially, with newer apps incorporating sensor-based monitoring and machine learning (ML) algorithms for real-time disease monitoring. Yet, most available PD-related apps are only interested in tracking general symptoms and not real-time tremor monitoring and progression tracking. Consequently, there exists a fundamental gap in digital health solutions that are unique to early PD tremor detection and personalized treatment.

In spite of the existence of online tools developed for movement disorder testing, present software implementations cannot integrate high-resolution sensor data with AI-based diagnostic approaches. Present software cannot perform tremor pattern analysis, and as a result, it cannot provide generalized conclusions without the specificity to enable timely diagnosis and informed clinical decisions. This implies the necessity of a novel user-driven application that will integrate wearable sensor data, machine learning-based feature extractions, and predictive modeling for early and accurate tremor detection. This research aims to bridge this gap by identifying the most important characteristics on which PD prediction of Parkinsonian tremor can be made with minimal reliance on subjective opinion and extraneous variables. Creating an app based on such characteristics will improve the accuracy of PD diagnosis and enable early management, particularly in remote areas. A science-driven solution with the use of real-world sensor data and AI-driven analysis will not only improve the quality of diagnosis but also provide patients and clinicians with real-time information, thus leading to better long-term outcomes.

Chapter 2

Literature Survey

2.1 Detecting Parkinsonian Tremor From IMU Data Collected in-the-Wild Using Deep Multiple-Instance Learning

In this study, Parkinson's Disease tremor symptoms have been identified through the analysis of data gathered with Inertial Measurement Units worn by patients in naturalistic settings. Compared to the data gathered in traditional laboratory-controlled environments, this field data tends to be more variable and noisy. In response to these challenges, the researchers employed Deep Multiple-Instance Learning (MIL) as a tactic that takes individual data segments as a whole to be bags or collections of instances, of which only a small proportion may potentially contain the useful signal (the tremor). This tactic improved tremor classification in naturalistic environments because it obviated the task of laborious marking of each unique data point. The researchers noted that MIL-based methods showed a greater capacity to generalize to real-world variability than has been seen in the past with traditional deep learning classifiers. However, these models were stated to require refinement, particularly with respect to the task of model interpretation and the ability to maintain invariant performance across all demographic and clinical subgroups.

2.2 Deep Transfer Learning Based Parkinson's Disease Detection Using Optimized Feature Selection

In the current study, the research team explored the integration of deep transfer learning with an enhanced feature selection process to detect Parkinson's Disease with physiological or movement-related data. Transfer learning was utilized by utilizing pre-trained deep neural networks that are capable of learning general patterns from massive datasets. The models were then fine-tuned for the specific Parkinson's Disease-related classification task. An automatic feature selection process was utilized to enhance the efficiency and focus of the model. This process allowed the selection and utilization of only relevant features or biomarkers, which enhanced the model's accuracy and speed. The proposed approach significantly enhanced diagnosis efficiency while, at the same time, ensuring minimal computational cost. However, the study established that there is a trade-off between performance and transparency of the model, particularly in clinical environments where interpretability is essential for clinical application.

2.3 Wearable-Enabled Algorithms for the Estimation of Parkinson's Symptoms Evaluated in a Continuous Home Monitoring Setting Using Inertial Sensors

This research explored the application of wearable technology for long-term monitoring of Parkinson's Disease symptoms in real-life settings. The patients were observed in their living spaces through inertial sensors, and motion data gathered were processed with machine learning models to quantify tremor amplitude and other motor impairments. The objective was to achieve an invasive and real-time means of symptom progression assessment without recurrent visits to hospitals. The employed models showed promise to estimate symptom severity with satisfactory precision, affirming the possibility that AI-driven wearables might significantly contribute to remote healthcare monitoring. However, the researchers underscored the need for further research that involves larger cohorts of patients to establish the scalability and consistency of the approach.

2.4 Machine Learning in the Parkinson's Disease Smartwatch (PADS) Dataset

The study used the PADS dataset, which contains sensor data gathered using smartwatches from Parkinson's Disease patients. The experiments were performed by the researchers using diverse machine learning approaches, such as deep neural networks and ensemble methods, to detect tremor occurrences and examine patterns of symptoms. Interestingly, ensemble methods were specifically effective because they can leverage the strengths of diverse algorithms, hence providing better accuracy and reduced error rates. The study highlighted the need for careful feature selection and engineering to achieve the best performance since raw sensor data are likely to consist of irrelevant or redundant signals. The study illustrates the capability of wearable technology in constructing scalable diagnostic systems, while highlighting that further enhancement of automatic feature extraction can make the model perform better.

2.5 Motor Assessment of X-Linked Dystonia Parkinsonism via Machine-Learning-Based Analysis of Wearable Sensor Data

This study specifically aimed to classify and monitor motor symptoms in patients with X-linked Dystonia Parkinsonism (XDP), a rare neurodegenerative disorder. The researchers employed wearable sensors to capture motion, which was then processed with machine learning algorithms. The study aimed to find features in motion patterns that are important for distinguishing between symptom types, such as dystonic movements and Parkinsonian tremors. Classifiers based on the features demonstrated high accuracy, thereby suggesting the potential for analytics based on sensors to be applied in clinical evaluation. Nevertheless, the study authors noted that the models used nowadays tend to overlook fine-grained interactions among various motor features. They recommended that future research should target examining feature interaction analysis and potentially incorporating advanced interpretability tools in order to learn how intricate symptoms are represented in sensor data.

While there has been significant work done in PD tremor detection and monitoring, there are some areas that are still lacking. Firstly, the majority of the studies did not explore the inter-relations between predictive features in depth, which could potentially

yield more information regarding the evolution of PD. For example, while Detecting Parkinsonian Tremor From IMU Data Collected in-the-Wild utilized in-the-wild sensor data, it did not focus on the interactions between various sensor-derived features.

Second, while some of the research studies did incorporate feature selection, none of them did an extensive analysis with an array of selection algorithms ranging from simple statistical methods to complex deep learning approaches. Secondly, studies such as the Wearable-Enabled Algorithms for the Estimation of Parkinson's Symptoms and the Machine Learning in the Parkinson's Disease Smartwatch (PADS) Dataset were performed on small datasets, which poses an issue with the generality of the models to larger patient groups. To address these shortcomings, more robust feature selection techniques and studies of interrelationships with diverse datasets are necessary.

To bridge this gap, the present work is interested in the analysis of the interactions among different features seen from wearable sensors and the comparison of different feature selection algorithms on a larger dataset. By analyzing the interaction among different movement-related features and the careful selection of the most significant ones, the present work aims to develop a model for Parkinsonian tremor detection that is more accurate, robust, and transferable using the minimum possible required features.

Chapter 3

Methodology

The proposed model for predicting Parkinson's Disease (PD) and its variations has been developed as a comprehensive pipeline incorporating multiple advanced techniques to enhance classification accuracy. Our approach involves three distinct classification models: a multi-class model that differentiates between healthy individuals, PD, and Differential Disease (DD); a binary classifier distinguishing between healthy and unhealthy individuals; and a specialized model for classifying PD versus DD.
The proposed hierarchical framework (Fig. 1) integrates multi-modal data for classification. The methodology used to extract the best and optimized performing model (Fig. 2) is given below with a detailed breakdown of each component, enriched with technical and clinical insights.

3.1 Data Collection

The data collection for this study is the Parkinson's Disease Smartwatch (PADS) dataset, which contains clinical assessments of a heterogeneous sample of Parkinson's Disease (PD) patients, individuals who present with similar movement disorders, and healthy control participants. These assessments were collected through a smart-device-based configuration that consisted of two wrist-worn smartwatches and a smartphone. The two smartwatches simultaneously collected sensor data from 11 interactive movement tasks specifically developed by expert neurologists for the detection of subtle motor impairments.

5,159 measurement steps were collected from 469 participants. The data set includes raw acceleration and rotation signals from smartwatch sensors, together with detailed movement step information, demographic variables, medical histories, and Parkinson's

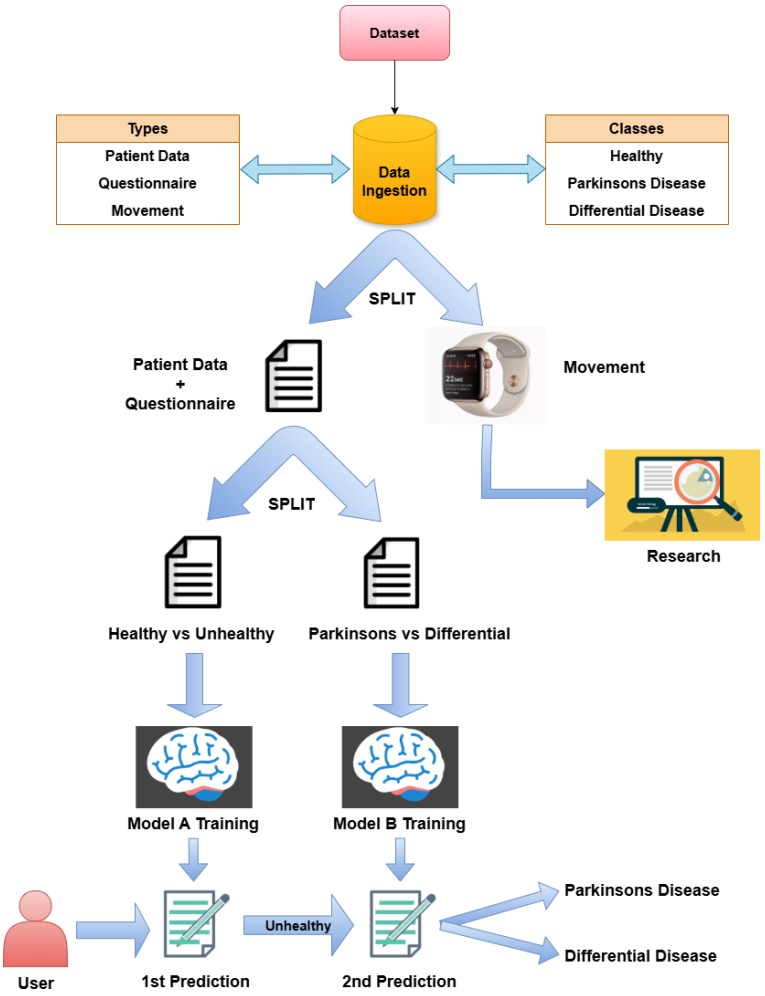


FIGURE 3.1: Flow-chart of the project

Disease-specific non-motor symptoms. Data were collected in two phases: (1) self-completion of an electronic questionnaire, which supplied demographic and medical information, including 30 binary answers for Parkinson's Disease-specific non-motor symptoms as delineated by the Parkinson's Disease Non-Motor Symptoms Questionnaire of the International Parkinson and Movement Disorder Society, and (2) a movement-based assessment using a central smartphone application. For training and testing of the model, an 80-20 train-test split was used to maintain an even data distribution. Table 1 has the 30 questionnaires, and Table 2 has movement data tasks. Diagnoses were validated by board-certified neurologists, and all data were stringently validated for accuracy. Cases of uncertain class assignment were fixed, invalid records were deleted, and personal data were pseudonymized for privacy.

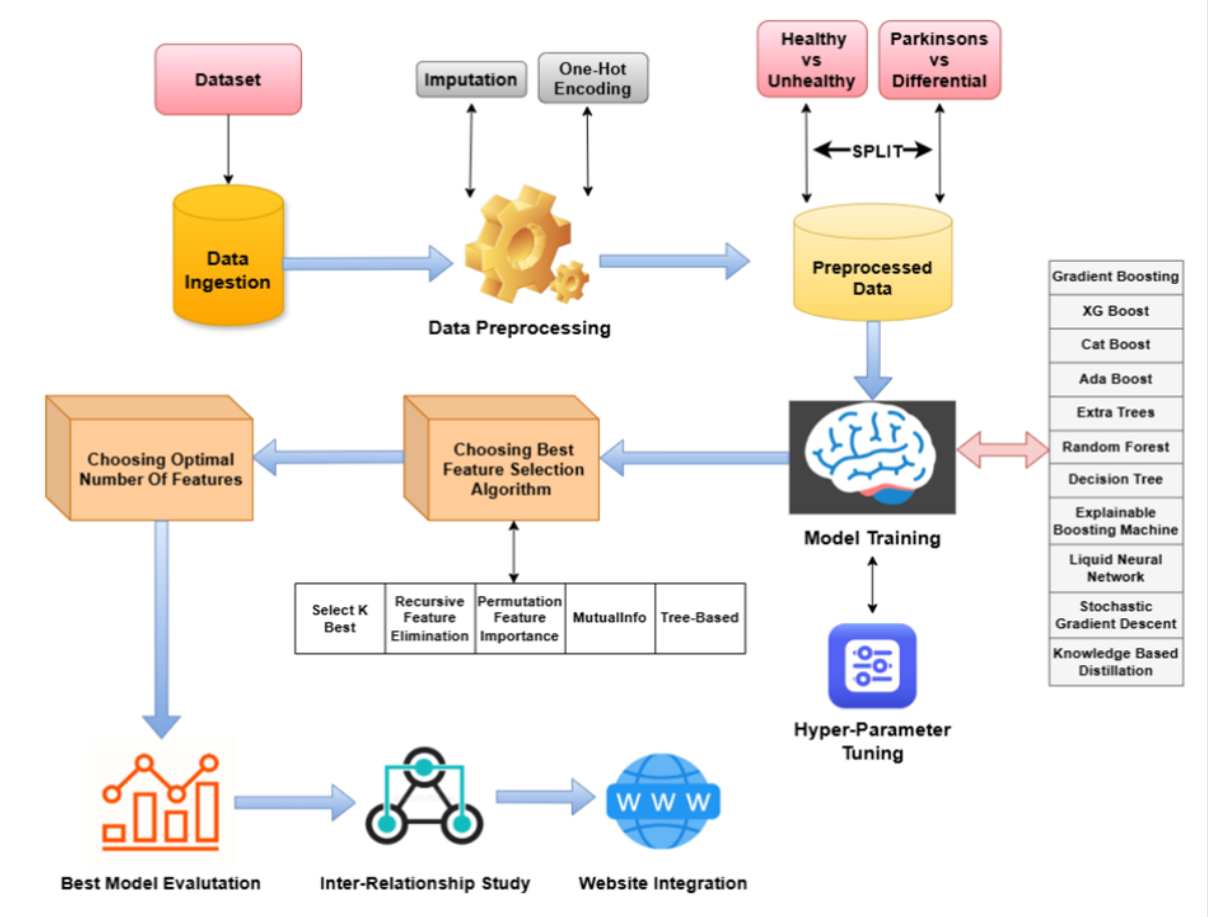


FIGURE 3.2: Proposed Model

3.2 Data Preprocessing

Data preprocessing is a crucial step in ensuring the dataset is properly structured for model training by addressing inconsistencies, normalizing features, and selecting the most relevant attributes. Our dataset consisted of both questionnaire-based features and movement-based sensor data, requiring tailored preprocessing strategies for each type.

3.2.1 Patient Data and Questionnaire Data Preprocessing

Unlike many real-world datasets, our questionnaire dataset had missing values, which were imputed using the median for numerical values and the mode for categorical values. We identified that both age and age_at_ diagnosis were highly correlated and affected model performance, so we retained only age_at_diagnosis. Additionally, outliers were detected in height and weight, and appropriate corrections were applied to ensure consistency. To address class imbalance, we applied the Synthetic Minority Over-sampling

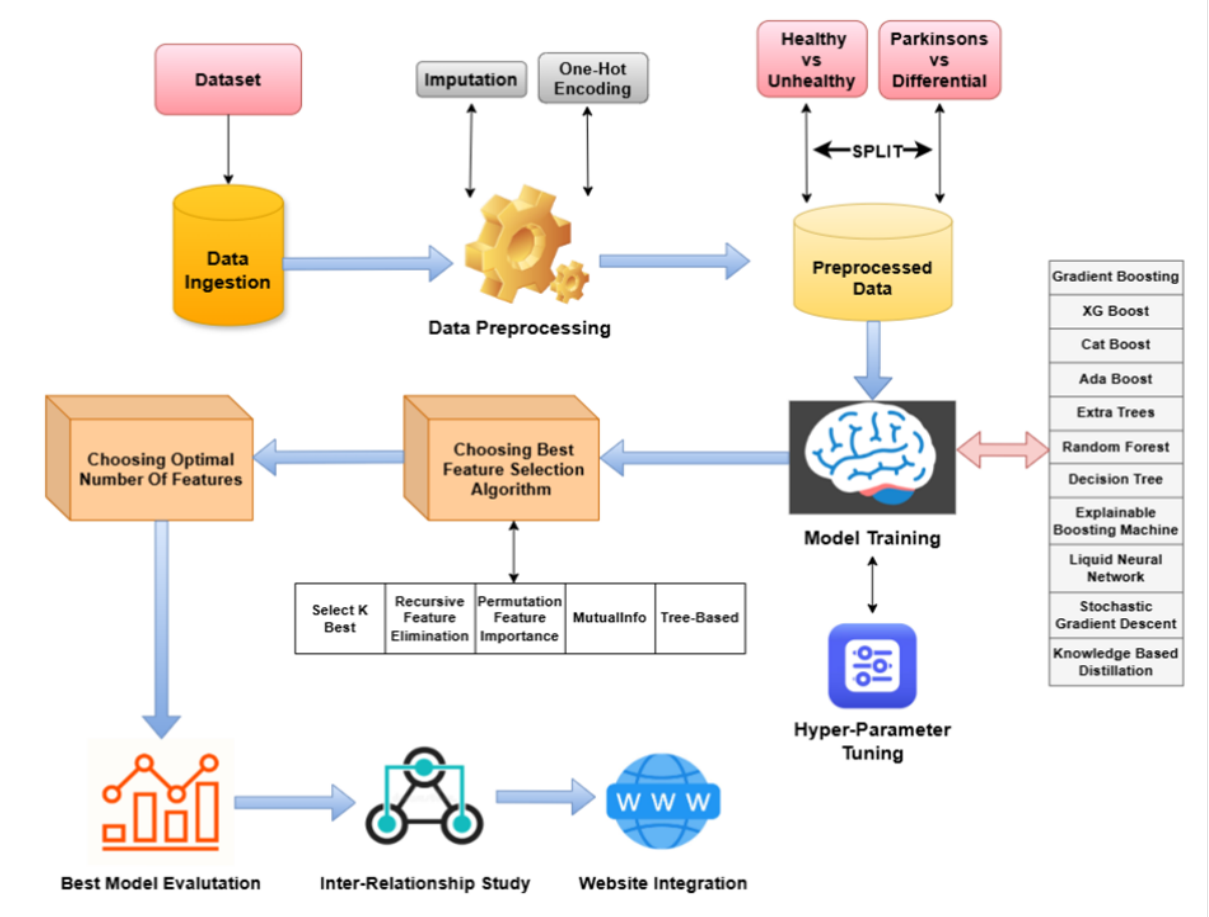


FIGURE 3.3: Methodology

Technique (SMOTE) to prevent bias toward the majority class. Categorical values were encoded using One-Hot Encoding, and numerical features were standardized using a consistent scaling approach. Feature selection played a vital role in optimizing model efficiency. We applied multiple feature selection techniques, including: - Permutation Feature Importance (PFI) to assess the contribution of each feature. - Recursive Feature Elimination (RFE) to iteratively remove less significant features. - Select K Best to retain the most influential variables. - Mutual Information to identify non-linear dependencies among features. Furthermore, excessive Exploratory Data Analysis (EDA) was conducted to identify trends, correlations, and redundancies. This helped refine our preprocessing strategy to improve classification accuracy. For classification tasks, we structured our dataset into two main prediction problems: 1. Healthy vs. Non-Healthy : Parkinson's Disease (PD) and Differential Diagnosis (DD) were combined into a single class (1), while healthy individuals remained as class 0. 2. Parkinson's vs. Differential Diagnosis : Healthy individuals were excluded, and the DD class (previously labeled as 2) was relabeled as 0, while PD remained as 1.

3.2.2 Movement Data Preprocessing

The dataset included time-series motion sensor readings from 469 patients performing 11 predefined tasks with both left and right hands. Each task was recorded using accelerometers and gyroscopes across three axes (X, Y, Z) over a 10-second duration at a sampling rate of 25 Hz, leading to 1024 time points per task. To transform the raw time-series data into structured formats suitable for machine learning, we did high-dimensional feature representation of 66 featurses.
Standard deviation was computed for all 1024 readings per sensor signal per task. Since each patient had 11 tasks per hand, this resulted in 22 tasks in total, with six sensor features per task (Accelerometer X, Y, Z and Gyroscope X, Y, Z). The final dataset contained 66 features per patient (6 features $\times$ 11 tasks $\times$ 2 hands), with the class label appended.

Each of these datasets was split into two separate classification tasks:

1. **Healthy vs Non-Healthy**: Patients labeled as PD and DD were merged into a single class 1, while healthy individuals remained as class 0.
2. **Parkinson's vs Differential Diagnosis**: Healthy individuals were removed, and the DD class was relabeled as 0, while PD remained as 1.

This resulted in a total of six datasets (66-column, 22-column, and 11-column feature sets for both classification tasks). These structured datasets allowed us to systematically evaluate different feature extraction strategies and classification models, ensuring optimal performance and interpretability.

3.3 Feature Selection and Model Optimization

The classification phase involved testing various machine learning algorithms, including Random Forest, Decision Tree, Gradient Boosting, XGBoost, CatBoost, AdaBoost, and Extra Trees Classifier, along with ensemble methods combining these models. Initially, all models were trained using the complete set of 39 features from the preprocessed dataset. However, the results were suboptimal, with lower accuracy and a higher number of false negatives, particularly in distinguishing between Parkinson's Disease (PD) and Differential Diseases (DD).

To improve model performance, we conducted extensive feature selection and data balancing before retraining the models. Feature selection focused on identifying the most

influential variables, and class imbalance was addressed through oversampling techniques such as SMOTE. Instead of selecting the single best-performing model based on accuracy, we employed a Voting Classifier (soft voting), which combines the predictions of all models. This approach leverages the strengths of multiple classifiers, leading to more stable and reliable predictions. The ultimate goal was to maximize classification accuracy while minimizing false negatives, as misclassification of PD or DD can have significant clinical implications.

3.3.1 Feature Set Selection and Optimization

Given the complexity of the dataset, multiple feature sets were tested for different classification tasks:

1. **Healthy vs. Unhealthy Classification:**

   From the questionnaire section, we started with 39 features, including demographic details (height, weight, gender, medical history, etc.), and applied feature selection techniques to reduce this to 14 key features. These selected features had the highest predictive power in differentiating between healthy individuals and those with any movement disorder (PD or DD).

2. **PD vs. DD Classification:**

   Since differentiating between PD and other movement disorders is clinically challenging, an additional 15 features were selected for this specific task. These features were drawn from both questionnaire data and sensor-based movement assessments, ensuring that both motor and non-motor symptoms were considered.

3. **Movement Task Selection:**

   The dataset included 11 interactive movement tasks, each designed to highlight specific movement impairments. However, only two tasks provided the maximum diagnostic value, as they consistently differentiated between PD and DD with high accuracy. These tasks were prioritized for model training.

To determine the best feature subset, five different feature selection techniques were applied to both the questionnaire-based and movement-based features. The selection process focused on finding features that maximized accuracy, precision, recall, and F1-score while minimizing false negatives. For the top two feature selection methods, we varied the number of selected features from 1 to 39 and evaluated performance metrics at each stage.

This iterative approach helped identify the minimum number of features required to achieve optimal model performance, ensuring a balance between accuracy and computational efficiency.

3.3.2 Feature Selection Algorithms

The study employed a combination of five feature selection techniques to systematically identify the most relevant predictors for each classification task:

1. **Select K Best:**

   This statistical method ranks features based on their correlation with the target variable and selects the top K features that contribute the most to classification.

2. **Permutation Feature Importance (PFI):**

   A model-agnostic approach that evaluates feature importance by randomly shuffling each feature and measuring its impact on model performance. Higher importance scores indicate stronger influence on predictions.

3. **Recursive Feature Elimination (RFE):**

   An iterative technique that systematically removes the least important features, refining the model until only the most significant predictors remain. RFE was particularly useful for reducing redundancy in questionnaire-based features.

4. **Tree-Based Feature Importance:**

   Feature importance scores were extracted from tree-based models (Random Forest, Gradient Boosting), which assess how much each feature contributes to reducing impurity (e.g., Gini impurity or entropy). This method helped in ranking movement-related features.

5. **Mutual Information:**

   Measures how much information about the target variable is provided by each feature. Features with high mutual information values were prioritized for final model training.

After applying these techniques, a refined set of 14 features for Healthy vs. Unhealthy classification and 15 features for PD vs. DD classification was selected. These features were used in all subsequent model training steps.

3.3.3 Addressing Class Imbalance with SMOTE

One of the significant challenges in this study was severe class imbalance in the dataset. The distribution of classes was as follows:

- Parkinson's Disease (PD): 58.8% of patients
- Differential Diseases (DD): 24.3% of patients
- Healthy Controls: 16.8% of patients

This imbalance posed a risk of biased model predictions, where classifiers would favor the majority class (PD) while underperforming on the minority classes (Healthy and DD). To counter this, we applied Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the underrepresented classes by interpolating between existing data points. This helped in:

- Balancing the dataset, ensuring models did not disproportionately favor the PD class.
- Reducing bias in predictions, particularly for DD and Healthy classifications.
- Improving recall and reducing false negatives, crucial for diagnostic accuracy.

SMOTE was applied at multiple stages, including both the training phase of individual models and during ensemble learning, to maintain a well-balanced dataset across all classification tasks.

3.4 Hyperparameter Tuning

To further refine model performance, GridSearchCV was used to perform hyperparameter tuning on all classifiers. This process involved:

- Testing a range of hyperparameter values for each model.
- Evaluating both hard and soft voting strategies in ensemble models.
- Selecting the optimal configuration that maximized accuracy and minimized false negatives.

The final model selection was based on a combination of accuracy, precision, recall, F1-score, and false negative rate. The best-performing model was saved for future use, ensuring that it could serve as a reliable and efficient tool for differentiating between Healthy, PD, and DD classifications.

3.5 Performance Metrics

In this study, the performance of the fine-tuned machine learning models was evaluated using several important metrics: accuracy, precision, recall, F1-score, and false negatives. Among these, minimizing false negatives was especially crucial, given the serious health risks posed by undiagnosed cases of Parkinson's disease.

Accuracy gave a general idea of how well the model performed overall. Precision helped us understand how many of the predicted positive cases were actually correct, while recall measured the model's ability to catch all true positive cases. Since medical datasets often have imbalanced class distributions, the F1-score—being a harmonic mean of precision and recall—offered a more balanced view of performance.

False negatives, in particular, carried significant weight in this context, as each missed diagnosis could result in delayed treatment and worsening of the condition. By considering all these metrics together, the evaluation provided a more complete and realistic picture of how effective the model was in detecting Parkinson's disease.

1. Accuracy= (TP+TN)/(Total Number of Data Points)
2. Precision= TP/(TP+FP)
3. Recall= TP/(TP+FN)
4. F1-Score= (2*Precision*Recall)/(Precision+Recall)

3.6 Development of Digital Application

This project ended with the deployment of a Flask-based web application for Parkinson's disease prediction. The application uses an interactive user interface in which users are able to enter pertinent medical data, which is fed into an optimized machine learning model. The predictions are displayed in real-time, signifying the proper integration of front-end input forms and back-end predictive analytics. After the user has entered information, the application uses the optimal machine learning model to determine the presence or absence of Parkinson's disease.

Chapter 4

Results

4.1 Data Sources and Preprocessing Strategies

The dataset comprises three primary data sources:

- Questionnaire-based features (30 attributes)
- Movement-based sensor tasks (11 attributes)
- Demographic attributes (9 attributes)

Each data type required a tailored preprocessing approach to ensure data integrity and model compatibility.

4.1.1 Preprocessing Steps

To enhance the quality and reliability of the dataset, several key preprocessing steps were implemented:

1. Handling Missing Values and Duplicates Duplicate entries and null values were identified and removed to eliminate redundancy and maintain consistency. Given the diverse nature of the data sources, specific imputation strategies were applied:
   (a) Questionnaire Data: Since the questionnaire consists primarily of categorical (Yes/No) responses, missing values were imputed using the mode (most frequently occurring response) to preserve the dominant representation. For numerical values within the questionnaire, missing entries were replaced with the mean to maintain distribution integrity.

(b) Demographic Attributes: These attributes, including factors such as age, height, and weight, were handled based on their data type. Missing numerical values were imputed using the mean to retain statistical properties, while categorical variables were imputed using the mode, consistent with the questionnaire preprocessing approach.

(c) Movement-Based Sensor Data: Missing values were handled based on task-specific patterns and sensor reliability. Given the structured nature of movement tasks, statistical techniques were applied to maintain temporal consistency and ensure meaningful data representation.

2. Feature Encoding and Standardization To ensure compatibility with machine learning models:

   (a) Categorical features were converted into numerical representations using one-hot encoding. Its distribution is shown in Figure 4.1

   (b) Numerical attributes were standardized to ensure uniformity and prevent scale imbalances.

3. Outlier Detection and Handling Outliers in numerical variables were identified and addressed using appropriate thresholding techniques to prevent skewed predictions.

By implementing these preprocessing strategies, the dataset was refined to improve model efficiency and interpretability. This ensured that each data source—questionnaires, movement tasks, and demographic features—was optimally prepared for classification and predictive modeling.

For the classification tasks, we employed multiple machine learning models, including Random Forest, Decision Tree, Gradient Boosting, XGBoost Classifier, CatBoost Classifier, AdaBoost Classifier, and Extra Trees Classifier. Additionally, we utilized ensemble learning techniques, including a Voting Classifier (soft voting), which combined the predictions of multiple models to enhance overall accuracy. The models were trained separately for movement-based and questionnaire-based data, ensuring a comprehensive evaluation of feature importance across different data types.

The accuracies attained by each model for Healthy vs Unhealthy classification is shown in Figure 4.2 and for PD vs DD is shown in Figure 4.3.

For the Healthy vs. Unhealthy classification, the movement-based model, using the Stretch Hold & Cross Arms task, achieved 84% accuracy with zero false negatives, ensuring high sensitivity in detecting unhealthy cases. The questionnaire-based model, utilizing 14 selected features, performed even better, achieving 92% accuracy, demonstrating the effectiveness of symptom-based diagnosis.

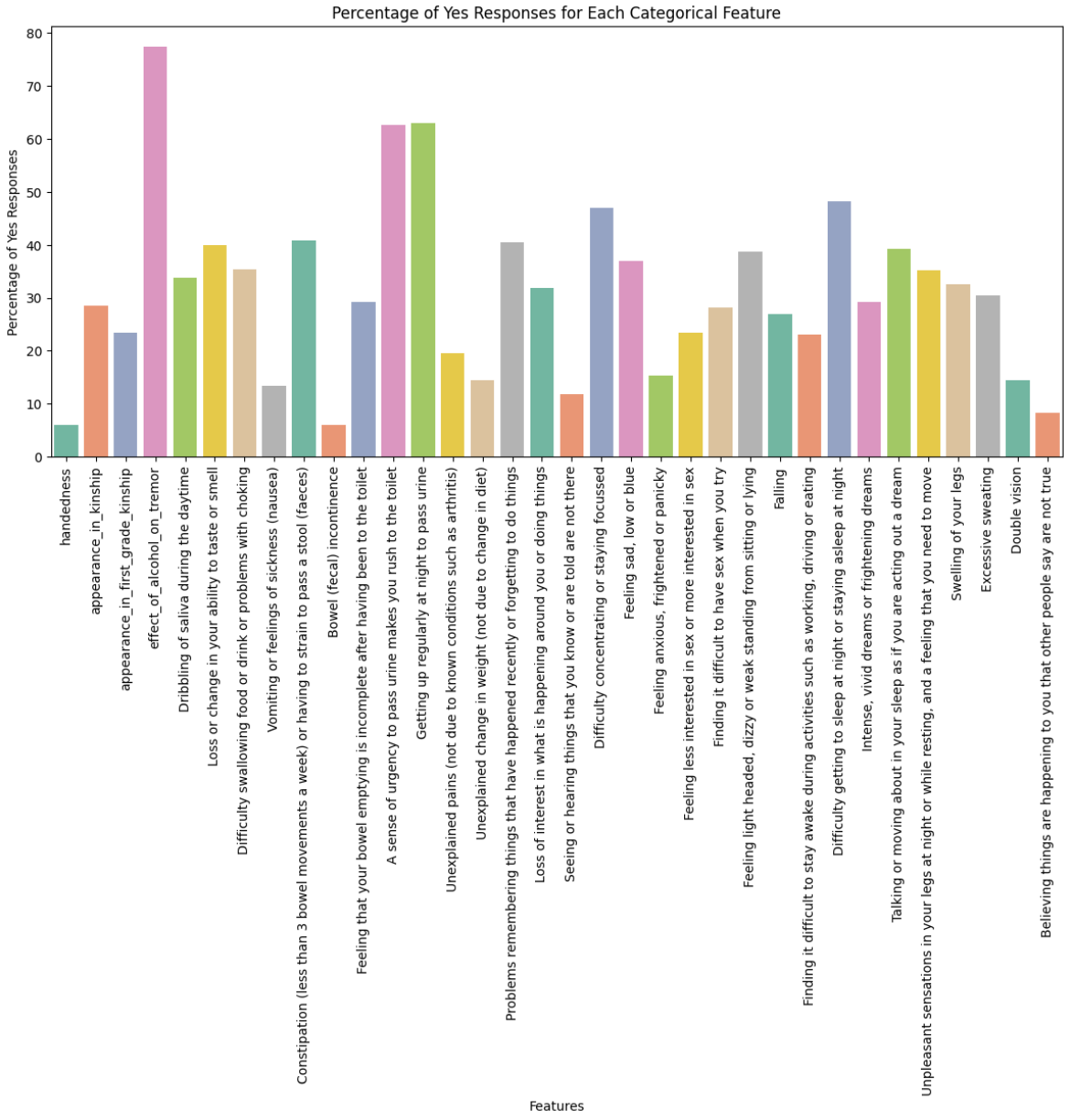


FIGURE 4.1: Categorical Features Distribution

For Parkinson's vs. Differential Diagnosis (PD vs. DD) classification, the movement-based model, trained on Drink Glass & Hold Weight tasks, attained 82% accuracy, leveraging movement-based impairments to distinguish between conditions. The questionnaire-based model, utilizing 29 features, achieved 85% accuracy, confirming that patient-reported symptoms and medical history play a significant role in differential diagnosis.

In the multi-class classification task, distinguishing between Healthy Controls (HC), Parkinson's Disease (PD), and Differential Diagnosis (DD), the movement-based model using all features achieved 81% accuracy, while the questionnaire-based model using all features reached 78% accuracy. These results highlight the complexity of differentiating

Classifier Name	Accuracy	Precision	Recall	F1 Score	Confusion Matrix
Ensemble	87.23	91	93.42	92.2	(11, 7), (5, 71)
Extra Trees	85.1	90.8	90.8	90.8	(11, 7), (7, 69)
CatBoost	87.23	91	93.42	92.2	(11, 7), (5, 71)
Random Forest	87.23	90	94.7	92.3	(10, 8), (4, 72)
Gradient	89.36	91.25	96	93.6	(11, 7), (3, 73)
XGB	87.23	91	93.42	92.2	(11, 7), (5, 71)
AdaBoost	90.43	91.36	97.37	94.26	(11, 7), (2, 74)
Decision Tree	83	88.5	90.8	90	(9, 9), (7, 69)

FIGURE 4.2: Healthy vs Unhealthy Accuracies

between multiple conditions but also demonstrate the effectiveness of leveraging both movement and questionnaire data for comprehensive classification.

To further enhance predictive performance, we developed an ensemble model by combining the outputs of multiple classifiers, improving overall robustness. The soft voting classifier was particularly effective in aggregating predictions, allowing for a balanced and optimized decision-making process. These results underscore the importance of a multi-modal approach, integrating diverse data sources and classification techniques to achieve high accuracy in predicting Parkinson's Disease and its variations.

To enhance our model's performance, hyperparameter tuning was employed on all the machine learning algorithms using GridSearchCV, testing both hard and soft voting strategies.The goal was to build a model with minimal features while ensuring high accuracy and reducing false negatives. To achieve this, the five feature selection algorithms were evaluated using two different features sets—11 features and 12 features. Figure 4 illustrates the accuracy of each algorithm with the number of false negatives associated with their respective accuracy.

Based on the results, the two feature selection algorithms that consistently demonstrated superior performance were Permutation Feature Importance (PFI) and Select K Best. Both methods yielded high accuracy while effectively minimizing false negatives, making them the most reliable choices compared to other algorithms.

Classifier Name	Accuracy	Precision	Recall	F1 Score	Confusion Matrix
Ensemble	84.6	83.9	96.3	89.67	(14, 10), (2, 52)
Extra Trees	86	85.25	96.3	90.4	(15, 9), (2, 52)
CatBoost	82	80.3	98.15	88.3	(11, 13), (1, 53)
Random Forest	83.33	81.54	98.15	89	(12, 12), (1, 53)
Gradient		80.95	94.4	87.18	(12, 12), (3, 51)
XGB	84.6	83.87	96.3	89.65	(14, 10), (2, 52)
AdaBoost	77	75	100	85.7	(6, 18), (0, 52)
Decision Tree	66.6	81.81	66.6	73.47	(16, 8), (18, 36)

FIGURE 4.3: PD vs DD Accuracies

Following this, we systematically trained models using feature sets ranging from 1 to 39 features, applying both feature selection techniques. The primary selection criterion, as previously discussed, was to identify the most accurate model while using the fewest number of features and maintaining minimal false negatives. This approach ensured a balance between model efficiency and predictive performance.

After extensive experimentation across multiple combinations of features for both classifiers, we identified the best-performing models:

- For Healthy vs. Unhealthy classification, the optimal model was achieved using 14 selected features.
- For Parkinson's Disease vs. Differential Diagnosis (PD vs. DD) classification, the best model utilized 29 selected features.

The top contributing features for these classifications are displayed in Figure 4.4 and Figure 4.5, showcasing the feature importance rankings determined by the PFI algorithm. The features are arranged in descending order of importance, highlighting the key predictors that significantly influenced model performance.

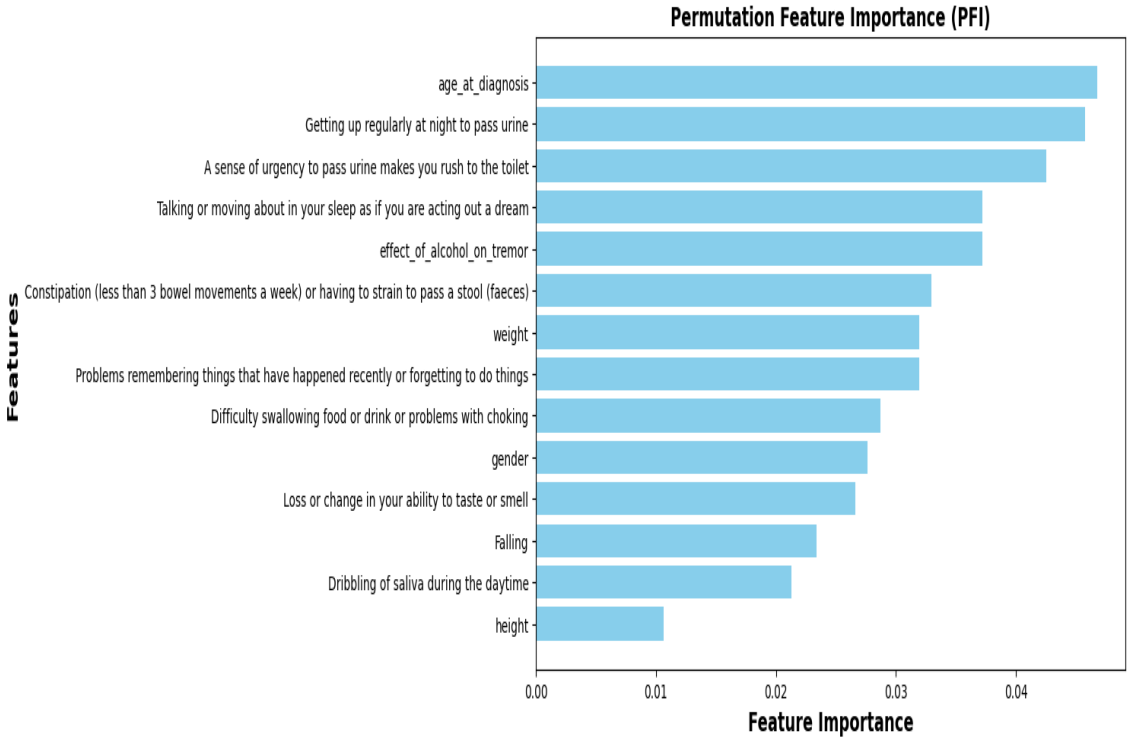


FIGURE 4.4: PFI Healthy vs Unhealthy Features

To further validate model effectiveness, confusion matrices for the two classifiers are presented in Figure 4.6 and Figure 4.7, providing a detailed breakdown of true positives, false positives, true negatives, and false negatives. Additionally, the AUC (Area Under the Curve) curves, shown in Figure 4.8 and Figure 4.9, illustrate the models' discriminative power.

The ROC-AUC scores achieved were as follows:

- Healthy vs. Unhealthy classification: 0.94, indicating excellent model performance with high sensitivity and specificity.
- PD vs. DD classification: 0.86, confirming strong predictive capability, though with slightly increased complexity in distinguishing between Parkinson's Disease and other differential diagnoses.

These findings underscore the effectiveness of feature selection in enhancing classification accuracy while reducing computational overhead. By leveraging PFI and Select K Best, we ensured that only the most informative and relevant features were retained, improving model generalization and robustness. The results further highlight the importance of multi-modal feature integration, combining movement-based assessments and patient-reported symptoms for a comprehensive, data-driven diagnostic approach.

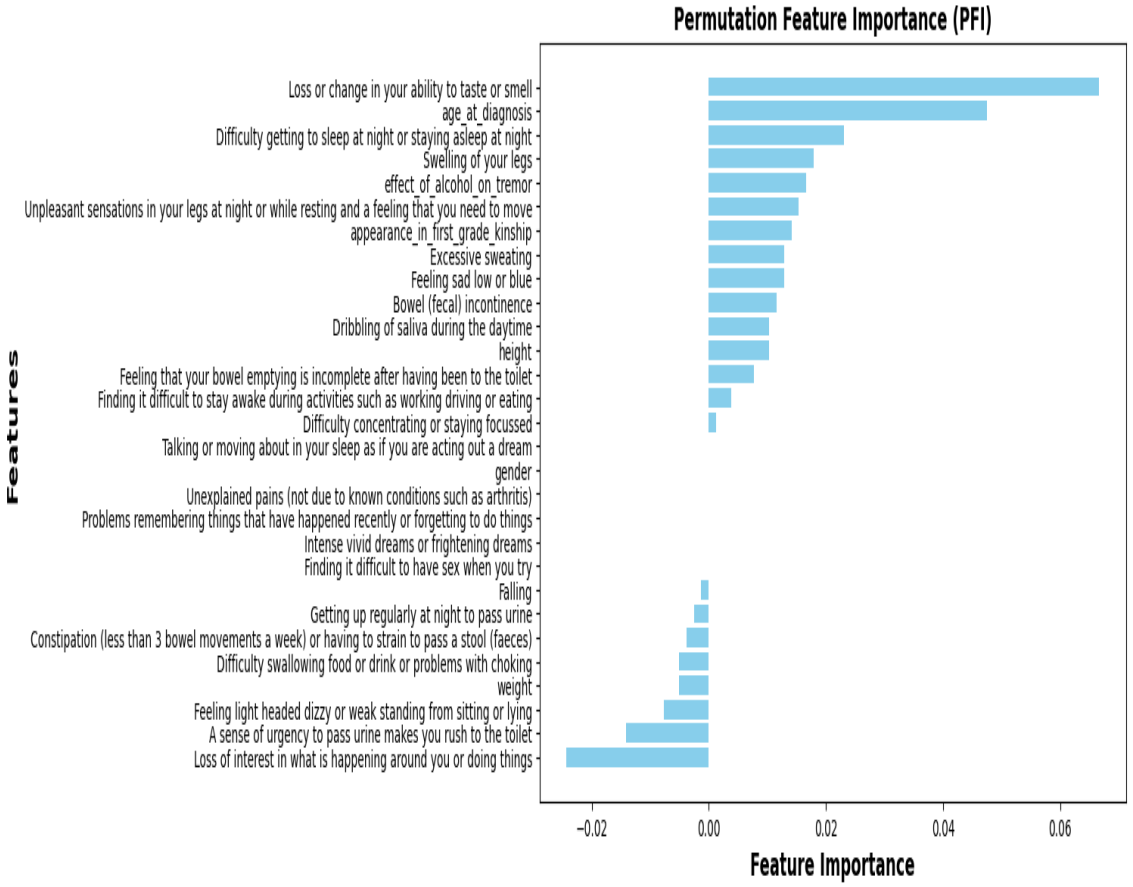


FIGURE 4.5: PFI PD vs DD Features

To improve the efficiency of movement-based classification, we aimed to identify the minimum number of movement tasks required to achieve high classification accuracy while reducing redundancy and effort in data collection. Instead of utilizing all 11 movement tasks or both hands (22 tasks in total), our goal was to determine whether a reduced subset of tasks could yield similar predictive performance.

To accomplish this, we systematically trained and evaluated multiple machine learning models for the two primary classification tasks. We initially trained models using all 11 tasks and observed their performance. Then, we iteratively reduced the number of tasks while monitoring classification accuracy to determine the optimal task subset for each classification problem.

For Healthy vs Unhealthy Classification, our analysis revealed that using only two specific movement tasks with a single hand achieved nearly the same accuracy as using all 11 tasks or both hands (22 tasks total).
The best-performing model for Healthy vs. Unhealthy classification was achieved using the two movement tasks: StretchHold and CrossArms, which resulted in an accuracy of 84%.

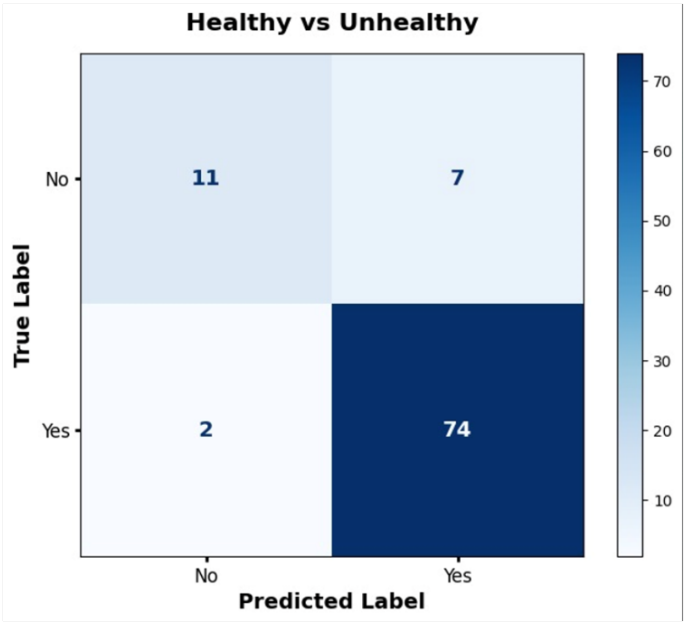


FIGURE 4.6: Confusion Matrix 1

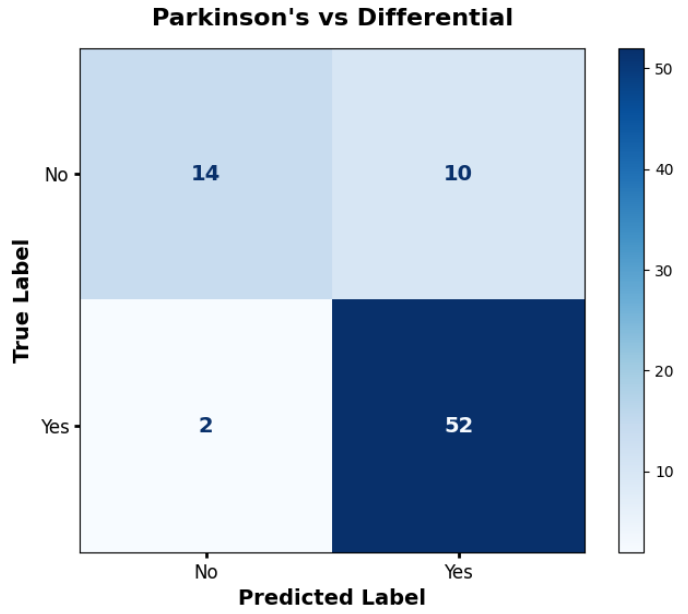


FIGURE 4.7: Confusion Matrix 2

This result suggests that just two carefully selected tasks provide sufficient information to distinguish between healthy and unhealthy individuals, eliminating the need for redundant movement-based assessments.

By reducing the number of required tasks, we significantly simplify the data collection process, making movement-based assessments more practical and efficient for real-world applications. The graph in Figure 4.10 shows the accuracies attained for different numbers of tasks employed.

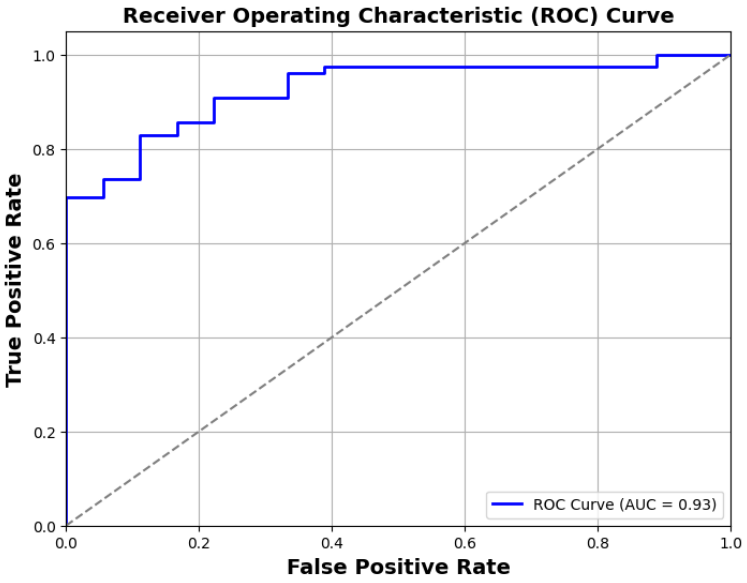


FIGURE 4.8: Healthy vs Unhealthy AUC curve

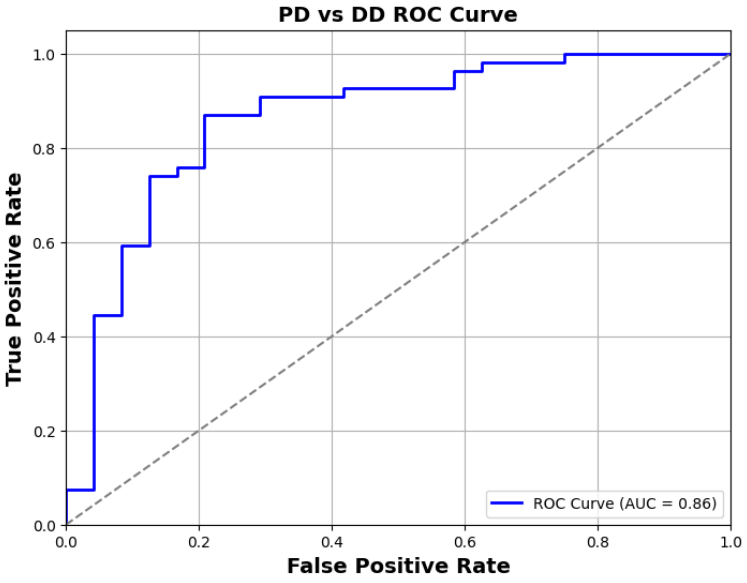


FIGURE 4.9: PD vs DD AUC curve

Similarly, for Parkinson's Disease vs. Differential Diagnosis classification, the best-performing model required only two movement tasks: DrinkGlass and HoldWeight, achieving 77% accuracy.

Notably, the two most important tasks for PD vs. DD classification were different from those identified for Healthy vs. Unhealthy classification, highlighting the distinct movement impairments associated with Parkinson's Disease and its differential diagnoses.

This finding underscores the importance of selecting the right movement assessments

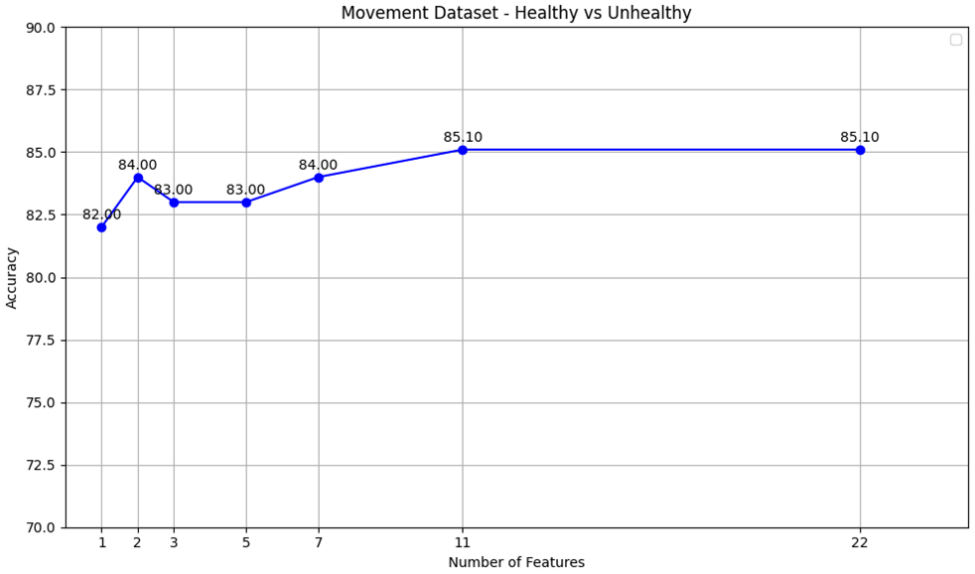


FIGURE 4.10: Number of Tasks vs Accuracy

based on the specific classification goal, rather than relying on an extensive set of generalized tasks. Figure 4.11 shows the accuracies attatined for different numbers of tasks employed.

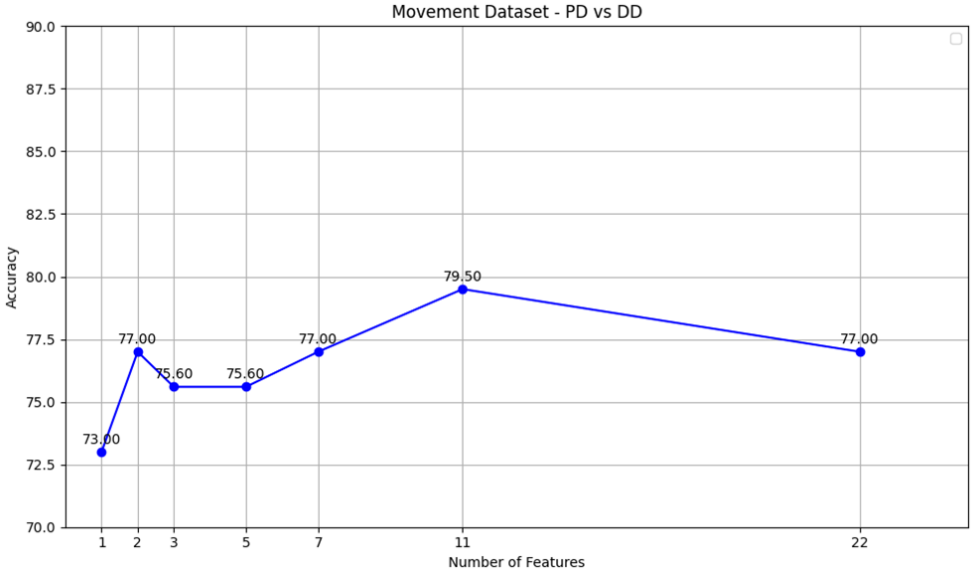


FIGURE 4.11: Number of Tasks vs Accuracy

The feature importance of each task is displayed in Figure 4.12.

To further validate the feature selection process, we analyzed the inter-relationships among movement tasks by plotting a correlation heatmap (Figure 4.13). This allowed us to verify that tasks excluded from the final models were still captured through their

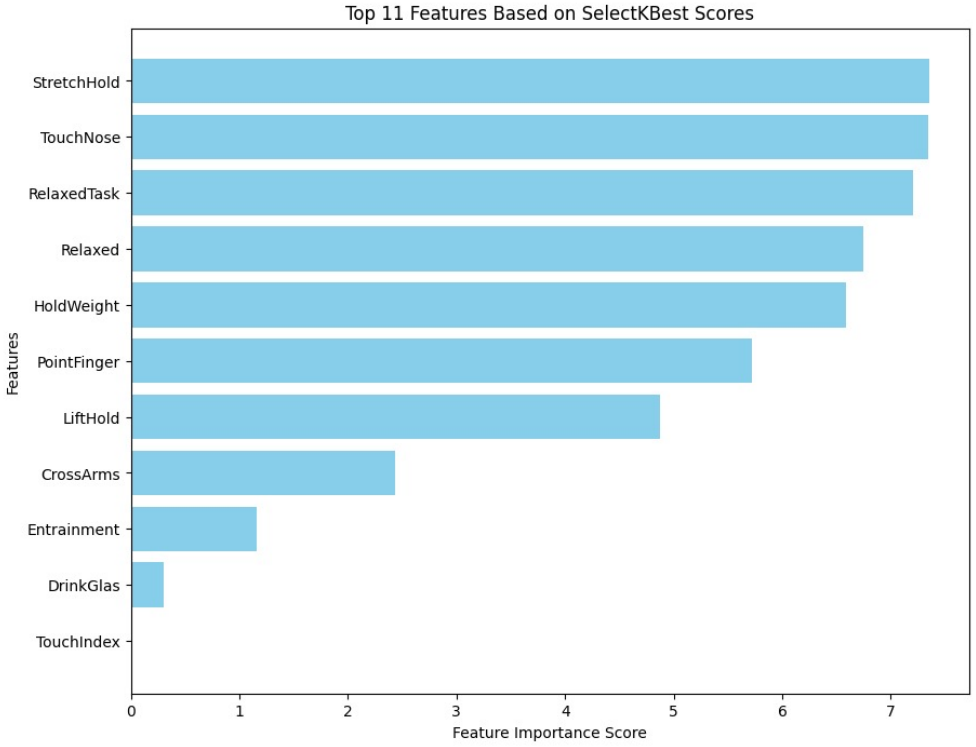


FIGURE 4.12: Tasks' Ranking

correlations with the selected tasks. The correlation values indicate the degree of association between different movement tasks.

Strongly correlated tasks share redundant information, meaning selecting one implicitly captures the effect of the other.

Example from the heatmap analysis:

- PointFinger and LiftHold (0.72 correlation) → If LiftHold is included, PointFinger's movement characteristics are already embedded in the data.
- StretchHold and RelaxedTask (0.49 correlation) → If StretchHold is included, it partially accounts for RelaxedTask's effect.
- DrinkGlass and HoldWeight (0.55 correlation) → The selected task DrinkGlass retains some information about HoldWeight, making it a reasonable choice for PD vs. DD classification.

This analysis confirms that tasks excluded from the model were adequately represented through their correlations with selected tasks, ensuring no significant loss of predictive power.
And by identifying the two most informative tasks for each classification problem, we

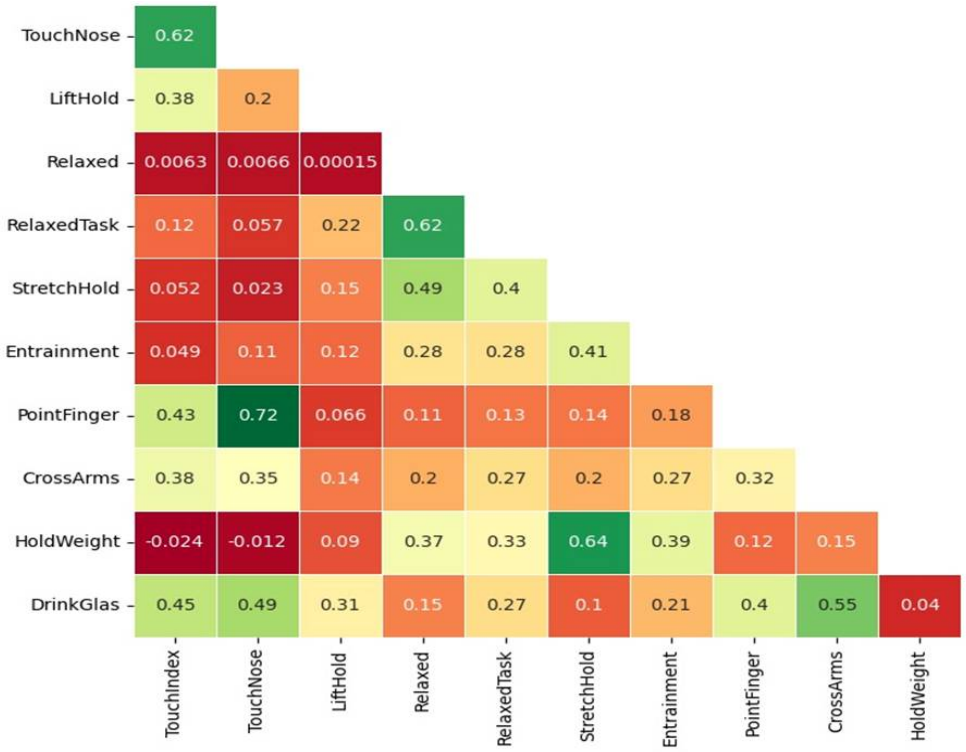


FIGURE 4.13: Inter-Relationship

reduce data collection time, minimize participant fatigue, and streamline diagnostic assessments, making movement-based classification more feasible for clinical and real-world applications.
The results further support the idea that movement impairments are task-specific, emphasizing the necessity of careful task selection in diagnostic models.

After optimizing traditional machine learning models for movement-based classification, we explored the potential of four novel models to assess whether they could further improve classification performance. These models included:

- Stochastic Gradient Descent (SGD) – A widely used optimization technique for large-scale machine learning problems, particularly effective for handling high-dimensional data.
- Explainable Boosting Machine (EBM) – A model designed for interpretability, leveraging generalized additive models (GAMs) with boosting to enhance prediction accuracy while maintaining transparency.
- Liquid Neural Network (LNN) – A biologically inspired neural architecture that dynamically adapts to data, making it particularly effective for time-series and dynamic movement data.

- Knowledge-Based Distillation (KBD) – A transfer learning technique where a simpler student model learns from a complex teacher model, reducing computational overhead while preserving performance.

Each of these models was trained and tested using the optimized feature sets identified in the earlier analysis (i.e., the two most relevant movement tasks per classification problem). The classification accuracies obtained for each model are summarized below in Figure 4.14.

Novel Models	Healthy vs Unhealthy	PD vs DD
Stochastic Gradient Descent	94%	82%
Explainable Boosting Machine	93%	81%
Liquid Neural Network	90%	78%
Knowledge Based Distillation	48%	58%

FIGURE 4.14: Inter-Relationship

Chapter 5

Discussion

5.1 Explanation of the Findings and Their Importance

The study demonstrates the effectiveness of integrating movement-based features and questionnaire data for accurately classifying Healthy vs. Unhealthy individuals and distinguishing Parkinson's Disease (PD) from other differential diagnoses (DD). The results indicate that feature selection plays a crucial role in improving classification performance while minimizing computational complexity.

- Feature Selection and Classification Performance: The Permutation Feature Importance (PFI) and Select K Best methods consistently selected the most relevant features, reducing the number of required inputs while maintaining high accuracy. This highlights the significance of focusing on a smaller, more informative feature set rather than using an exhaustive list of variables.
  - The best feature subset for Healthy vs. Unhealthy classification used 14 features, while for PD vs. DD classification, 29 features were optimal.
  - This selective approach ensured that the models remained computationally efficient while preserving predictive power.
- Movement Task Optimization: The results also highlight the potential to reduce the number of movement tasks needed for classification:
  - Instead of using all 11 movement tasks or both hands (22 tasks total), our analysis found that using only two specific movement tasks per classification problem achieved comparable accuracy.
  - For Healthy vs. Unhealthy classification, the StretchHold and CrossArms tasks alone resulted in 84

  - For PD vs. DD classification, the DrinkGlass and HoldWeight tasks yielded an accuracy of 77
  - These findings suggest that careful task selection can simplify diagnostic assessments, making them more feasible for real-world applications.

- Ensemble and Novel Machine Learning Models

  The soft voting classifier proved to be the most effective, aggregating predictions from multiple models to enhance overall robustness. Additionally, the study explored four novel machine learning models:

  - Stochastic Gradient Descent (SGD) – Efficient for high-dimensional data.
  - Explainable Boosting Machine (EBM) – Provided transparency in predictions.
  - Liquid Neural Network (LNN) – Adapted dynamically to movement-based input.
  - Knowledge-Based Distillation (KBD) – Optimized model performance while reducing computational overhead. These models demonstrated promising performance, highlighting the potential of advanced machine learning techniques in improving movement disorder diagnosis.

- Model Performance Metrics:

  - The ROC-AUC score of 0.94 for Healthy vs. Unhealthy classification confirms excellent model sensitivity and specificity.
  - The ROC-AUC score of 0.86 for PD vs. DD classification suggests strong discriminative power, although distinguishing between PD and other movement disorders remains challenging.

Overall, the results emphasize the importance of combining optimized feature selection, minimal task assessments, and ensemble learning techniques to create an accurate and efficient diagnostic system.

5.2 Limitations of the Current System

While the study presents promising results, certain limitations must be addressed:

1. Dataset Size and Diversity:

   The dataset used may not fully represent the variability of Parkinson's Disease symptoms across different demographic groups.

A larger, more diverse dataset is needed to improve the generalizability of the models.

2. Feature Generalization:

   The selected features were optimized for this specific dataset, but their effectiveness in different clinical settings remains unverified.

   Additional validation on independent datasets is required to confirm the robustness of the feature selection process.

3. Challenges in PD vs. DD Classification:

   The slightly lower ROC-AUC score (0.86) for PD vs. DD classification indicates that distinguishing between Parkinson's Disease and other movement disorders is inherently more complex.

   Overlapping symptoms may lead to misclassification, necessitating additional clinical features (e.g., genetic markers, imaging data).

4. Real-World Implementation Constraints:

   While movement task reduction makes diagnosis more practical, implementing these findings in clinical practice or wearable sensor applications requires further testing.

   The feasibility of using AI-powered diagnostic tools in real-time settings needs additional exploration.

5.3 Potential Improvements and Future Work

To further enhance the accuracy, efficiency, and clinical applicability of the system, several future improvements can be pursued:

1. Expanding the Dataset:

   Increasing the sample size and incorporating data from different populations will help improve model generalization.

   Collaborating with hospitals and research institutions to collect more comprehensive movement and clinical data.

2. Integrating Multi-Modal Data Sources:

   Currently, the system relies on movement-based features and questionnaire responses. Incorporating neuroimaging data, genetic markers, or wearable sensor outputs can improve diagnostic accuracy.

The use of electromyography (EMG) and gait analysis could further refine PD vs. DD classification.

3. Improving PD vs. DD Classification Performance:

   Exploring deep learning models trained on larger datasets to capture subtle movement differences in PD and other disorders.

   Investigating biomechanical models that analyze muscle activation and tremor patterns more effectively.

4. Developing a Real-Time Diagnosis System:

   Implementing the model in mobile applications or wearable devices to allow real-time patient monitoring.

   Testing the feasibility of an AI-assisted diagnostic tool for clinicians to support decision-making.

5. Enhancing Explainability and Interpretability:

   While models like Explainable Boosting Machine (EBM) improve transparency, more work is needed to explain why specific features contribute to classification.

   Developing interpretable AI models that provide insights into movement impairments and suggest personalized treatment recommendations.

By addressing these limitations and improvements, the system can be refined into a clinically viable diagnostic tool, paving the way for early detection and better management of movement disorders.

Chapter 6

Conclusion

This research introduced a machine learning-driven diagnostic framework for classifying Healthy vs. Unhealthy individuals and differentiating Parkinson's Disease (PD) from other differential diagnoses (DD). By implementing feature selection techniques (Permutation Feature Importance and Select K Best), we reduced the number of input variables while maintaining high classification accuracy. The results demonstrated that only two movement tasks were required for each classification problem, simplifying data collection while preserving diagnostic effectiveness. The soft voting ensemble model exhibited superior performance, achieving a ROC-AUC score of 0.94 for Healthy vs. Unhealthy classification and 0.86 for PD vs. DD classification. Furthermore, the study explored alternative classification models, including Explainable Boosting Machine (EBM) and Liquid Neural Networks (LNN), which showed promising results in terms of interpretability and adaptability. These findings emphasize the potential of machine learning in movement-based diagnostics, allowing for automated, efficient, and accurate assessments of neurodegenerative disorders.

To enhance accessibility and usability, the developed diagnostic system was integrated into a web-based application, allowing users to input relevant movement-based features and obtain predictions instantly. This web implementation ensures that the diagnostic tool is user-friendly, scalable, and accessible from any device with an internet connection. The system's interface includes a prediction form where users enter their movement test results, and upon submission, the classification outcome is displayed. Figure 6.1 illustrates the web-based prediction form for prediction of Healthy vs Unhealthy, while Figure 6.2 presents a sample result screen showing the diagnostic prediction for the same, and if its unhealthy, it is followed by another prediction form for prediciton of Parkinsons disease. And then, Figure 6.3 shows the result if its Parkinsons disease or some other differential disease. This integration facilitates real-world application,

making it easier for both patients and medical professionals to utilize the system without requiring complex installations or technical expertise.

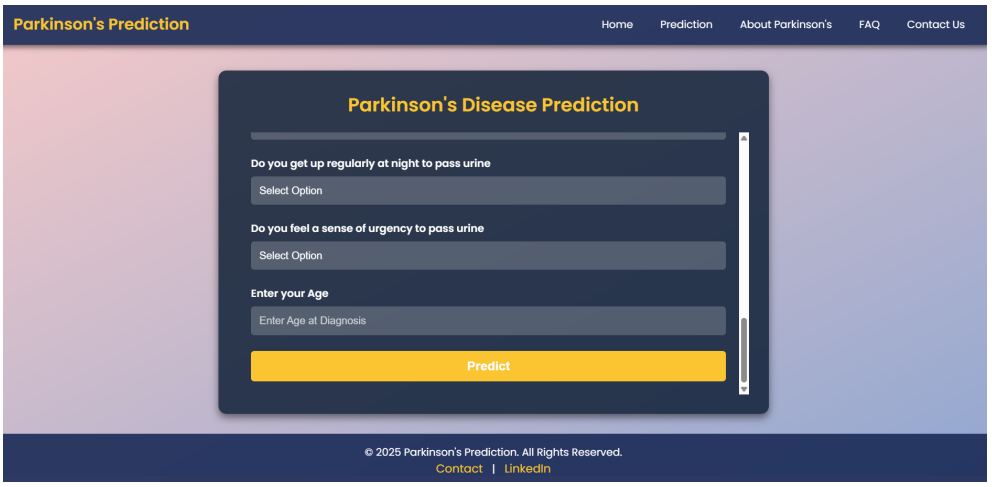


FIGURE 6.1: 1st Prediction Form

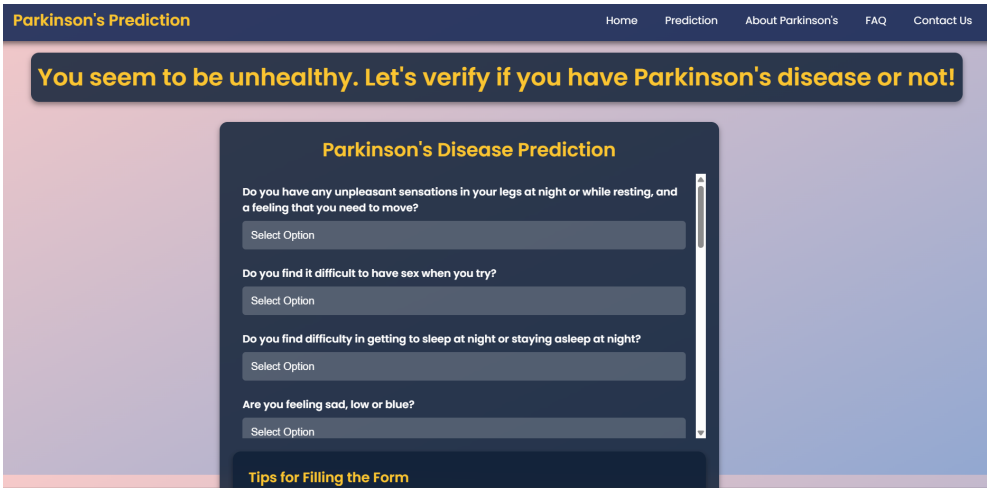


FIGURE 6.2: 2nd Prediction Form

While the results are promising, limitations remain in terms of dataset diversity, feature generalization, and real-world applicability. The dataset used in this study, while effective for model training and validation, may not fully capture the variability of PD symptoms across different demographics, disease stages, and movement patterns. Additionally, the PD vs. DD classification remains challenging due to overlapping symptoms, indicating that further refinement is needed to improve discriminatory power. Another challenge is the reliance on movement-based data alone, which, while valuable, may benefit from integration with multi-modal inputs, such as wearable sensor data, imaging, or genetic information. Future research should explore how combining multiple data sources can enhance diagnostic accuracy and robustness.

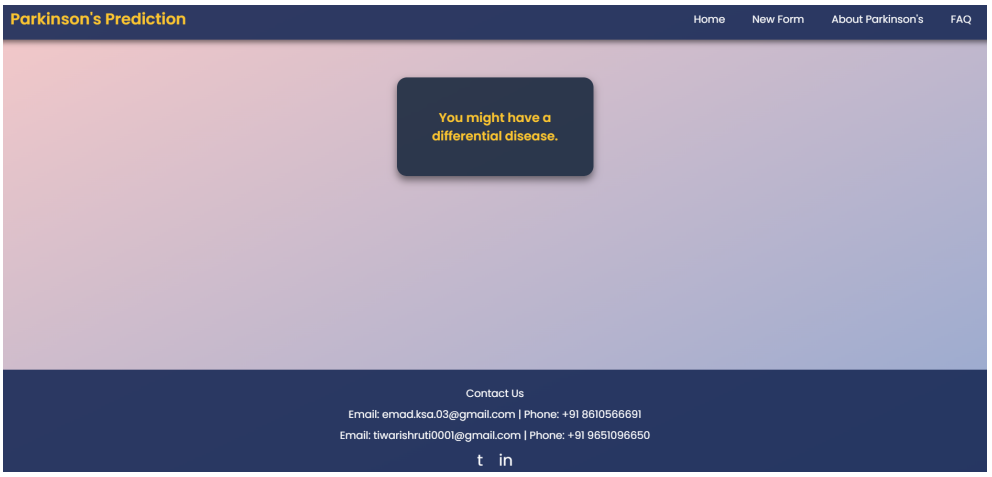


FIGURE 6.3: Final Result

The study lays the foundation for further improvements and real-world deployment of AI-driven movement diagnostics. Future work should focus on expanding the dataset, ensuring it represents a wider population, and validating the model's effectiveness in clinical settings. Additionally, the system could be enhanced by developing a real-time diagnostic tool—potentially integrated into mobile applications or wearable devices—that can assist neurologists in early disease detection and personalized treatment planning. By addressing these aspects, this research contributes to the growing field of AI-powered healthcare solutions, bringing automated, non-invasive, and accessible diagnostics closer to clinical implementation.

Bibliography

[1] Papadopoulos, A., Kyritsis, K., Klingelhoefer, L., Bostanjopoulou, S., Chaudhuri, K. R., and Delopoulos, A. *Detecting Parkinsonian Tremor From IMU Data Collected in-the-Wild Using Deep Multiple-Instance Learning.* IEEE Transactions on Biomedical Engineering, 67(12), 3282-3291, 2020.

[2] Abdullah, S. M., Abbas, T., Bashir, M. H., Khaja, I. A., Ahmad, M., Soliman, N. F., and El-Shafai, W. *Deep Transfer Learning Based Parkinson's Disease Detection Using Optimized Feature Selection.* IEEE Access, 10, 12115-12126, 2022.

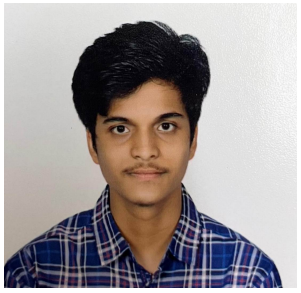
[3] Crowe, C., Sica, M., Kenny, L., O'Flynn, B., Mueller, D. S., Timmons, S., Barton, J., and Tedesco, S. *Wearable-Enabled Algorithms for the Estimation of Parkinson's Symptoms Evaluated in a Continuous Home Monitoring Setting Using Inertial Sensors.* IEEE Journal of Biomedical and Health Informatics, 24(9), 2556-2565, 2020.

[4] Talitckii, A., Anikina, A., Kovalenko, E., Shcherbak, A., Mayora, O., Zimniakova, O., Bril, E., Semenov, M., Dylov, D. V., and Somov, A. *Defining Optimal Exercises for Efficient Detection of Parkinson's Disease Using Machine Learning and Wearable Sensors.* IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29, 2254-2264, 2021.

[5] Parisi, F., Corniani, G., Bonato, P., Balkwill, D., Acuna, P., Go, C., Sharma, N., and Stephen, C. D. *Motor Assessment of X-linked Dystonia Parkinsonism via Machine-Learning-Based Analysis of Wearable Sensor Data.* Scientific Reports, 11(1), 11472, 2021.

[6] Varghese, J., Brenner, A., Fujarski, M., van Alen, C. M., Plagwitz, L., and Warnecke, T. *Machine Learning in the Parkinson's Disease Smartwatch (PADS) Dataset.* IEEE Transactions on Biomedical Engineering, 69(3), 1234-1244, 2022.

Biodata



Name: Emad Mohamed
Mobile No.: +91 8610566691
E-mail: emad.ksa.03@gmail.com
Permanent Address:-
8/78, T-3, Venus Ruby Apts., Ponnappa Street,
Purasaiwalkam, Chennai - 600084

Name: Shruti Tiwari
Mobile No.: +91 9651096650
E-mail: tiwarishruti0001@gmail.com
Permanent Address:-
562/64 Ramgarh Colony, Kanpur Road,
Lucknow - 226012