EARLY PARKINSON DISEASE DETECTION WITH FEATURE EXTRACTION USING MACHINE LEARNING

BY

Nafi Bin Monsoor Alvi
ID: 203-15-3868
AND

Md.Al Emon
ID: 203-15-3896

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By
Mr. Shahadat Hossain
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised By

Mr.Shah Md. Tanvir Siddiquee
Lecturer (Assistant Professor)
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled "Early Parkinson Disease Detection With Feature Extraction Using Machine Learning", submitted by Nafi Bin Monsoor Alvi and Md.Al Emon to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13th July, 2024.

BOARD OF EXAMINERS

14-07-2024

Dr. Md. Fokhray Hossain(MFH) — **Board Chairman**
Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Shah Md Tnavir Siddiquee (SMTS) — **Internal Examiner 1**
Assistant Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.07.24

Mr. Md Umaid Hasan(MUH) — **Internal Examiner 2**
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13/7/24

Dr. Shamim H Ripon(DSR) — **External Examiner**
Professor
Department of CSE
East West University

DECLARATION

We hereby declare that, this project has been done by **Nafi Bin Monsoor Alvi(203-15-3868),Md Al Emon(203-15-3896)** under the supervision of **Shahadat Hossain, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree.

Supervised by:

10.07.24

Mr. Shahadat Hossain

Lecturer(Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

Mr. Shah Md. Tanvir Siddiquee
Lecturer(Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Alvi 10.07.24

Nafi Bin Monsoor Alvi
ID: 203-15-3868
Department of CSE
Daffodil International University

Emon 10.07.24

Md. Al Emon
ID:203-15-3896
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Shahadat Hossain, Lecturer(Senior Scale)**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Field name*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Dr. Sheak Rashed Haider Noori** , for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Parkinson's disease (PD) is a neurodegenerative disorder impacting millions worldwide. This study aims to leverage machine learning algorithms to improve the diagnosis and understanding of PD. Building upon existing research that utilizes classifiers, feature extraction, data partitioning, and feature selection, we explore the potential of various feature selection algorithms in maximizing classification accuracy on publicly available PD datasets. The investigation will compare and contrast the performance of feature selection technique, ultimately identifying the method that yields the highest accuracy in PD classification. This research contributes to the growing body of knowledge surrounding the application of machine learning in PD diagnosis and paves the way for further exploration of specific feature sets and classification models to enhance clinical practice and patient outcomes. Here we will work on different selecting algorithms. Such as:PCA. For the result we will use several popular ML techniques. Including: K-NN, Random Forest, Decision Tree Algorithm,XG Boost,ANN.After our study of parkinson disease classification with the feature selectionwith correlation method we got accuracy of 89% on the Random Forest and the result after applying the PCA classifier the accuracy increased to 92%.

TABLE OF CONTENTS

PROBLEMS (EP) AND COMPLEX ENGINEERING ACTIVITIES (EA) ADDRESSING

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
PD	Parkinson Disease
PCA	Principle Component Analysis

CHAPTER 1
INTRODUCTION

1.1 Overview

Parkinson’s disease (PD)is a progressive neurodegenerative disorder immensely affecting the quality of lives of millions of people worldwide R. Prashantha, Sumantra Dutta Roya,et al. (2015). Parkinson’s disease is a disorder of the central nervous system affecting movement and inducing tremors and stiffness a neurodegenerative disorder affecting dopamine neurons in brain Dr. C K GOMATHY1, Mr. B. DHEERAJ KUMAR REDDY2 et al. (2021).Parkinson disease (PD) is the second most common neurodegenerative disorder after Alzheimer’s disease N. P. Narendra, Björn Schuller et al. (2021).PD is a progressive neuropathological degeneration of the CNS in which the motor functions of the human body are affected Iqra Nissar,Waseem Ahmad Mir et al. (2021).Parkinson Disorder, the second most common disease after Alzheimer’s and a major public health problem (Worldwide- 5M), is a progressive neurological disorder in the central nervous system whose cause is still unknown, it could be due to environmental factors and/or genetic factors which increases with time Surekha Tadse,Muskan Jain et al. (2021).Parkinson’s disease (PD) is a long-termed, neurological disorder that causes a person to lose control over several body functions including speech Shail Raval,Rahil Balar et al. (2020).

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting millions globally. Characterized by tremors, rigidity, and slowness of movement, PD poses significant challenges in diagnosis and management. Traditionally, diagnosis relies on clinical assessment, which can be subjective and prone to misdiagnosis, especially in the early stages. Recent advancements in machine learning (ML) offer promising avenues for improving PD diagnosis and understanding. Studies have demonstrated the potential of ML algorithms to analyze various data sources, including clinical data, voice recordings, and movement patterns, to aid in PD diagnosis. This research builds upon existing work by investigating the effectiveness of various feature selection algorithms in enhancing the accuracy of ML-based PD classification. The research utilizes publicly available PD datasets and employs a range of widely recognized feature selection technique,Principal Component Analysis (PCA). Furthermore, this research explores the performance of diverse classification models such as K-Nearest Neighbors (K-NN), Random Forest, Decision Tree,ANN and XG Boost. By comparing the efficacy of these techniques, the research aims to identify the combination that yields the highest accuracy in PD classification.

This research contributes to the ongoing exploration of ML applications in PD diagnosis. Through a comprehensive analysis of various feature selection algorithms and classification models, the research seeks to advance our understanding of how machine learning can improve the accuracy and efficiency of PD diagnosis, ultimately paving the way for potential improvements in clinical practice and patient outcomes.

1.2 Background and Present State

The first description of Parkinson's disease (PD) dates back to James Parkinson's 1817 work "An Essay on the Shaking Palsy." Since then, a great deal of study has improved our knowledge of this crippling illness. With an average onset age of 60 years, Parkinson's disease (PD) primarily affects the elderly, while early-onset PD can manifest before the age of 50. Usually progressing slowly at first, the disease eventually causes serious physical and cognitive deficits.

1.2.1 Background

In the past, postmortem investigations and clinical observation were the main ways that Parkinson's disease (PD) was understood. The discovery of dopamine's function in the brain in the late 20th century shed light on the pathophysiology of the condition. The brain's substantia nigra produces the neurotransmitter dopamine, which is crucial for controlling movement. The hallmark movement symptoms of Parkinson's disease (PD) are caused by the degeneration and death of dopamine-producing neurons in the brain.

Genetic studies have linked multiple gene mutations, including those in the SNCA, LRRK2, and PARK2 genes, to Parkinson's disease (PD). Notwithstanding the fact that most instances are sporadic and lack a strong familial connection, these findings have highlighted the significance of genetic variables in the illness. Additionally linked to environmental issues include pesticide and heavy metal exposure.

1.2.2 Present State

Although Parkinson's disease (PD) is still incurable today, advances in technology and medical research have made it easier to control the disease's symptoms. The mainstay of treatment is pharmacotherapy, which includes drugs like levodopa, dopamine agonists, and MAO-B inhibitors that work by mimicking or replenishing dopamine in the brain. These drugs have the potential to greatly reduce motor symptomatology, but they frequently have negative side effects and lose their effectiveness with time.

For severe Parkinson's disease (PD), deep brain stimulation (DBS) has become a popular surgical therapy option. Electrodes are inserted into particular brain regions during this treatment in order to modify neural activity and reduce symptoms. DBS has been demonstrated to enhance patients' quality of life when they are no longer responding appropriately to medicine.

1.3 Problem Statement

Despite advancements in diagnosis, Parkinson's disease (PD) remains challenging to identify, particularly in its early stages. Traditional methods, primarily relying on clinical assessments, can be subjective and susceptible to misdiagnosis. Therefore, there is a critical need for more objective and accurate methods to improve PD diagnosis. This research addresses this need by investigating the potential of optimizing machine learning algorithms for enhanced PD classification accuracy. Specifically, we will focus on the effectiveness of feature selection techniques and algorithms in maximizing the accuracy of machine learning models for PD diagnosis on publicly available datasets.

1.4 Objectives

Evaluate the effectiveness of different feature selection algorithms in maximizing the accuracy of machine learning models for PD classification. This involves the performance of technique like PCA on publicly available PD datasets. Compare the performance of diverse classification models such as K-NN,Random Forest,Decision Tree,XG Boost and ANN. Identify the feature selection algorithm and classification model that yields the highest accuracy in PD classification on the selected dataset. Contribute to the growing body of knowledge surrounding the application of machine learning in PD diagnosis and pave the way for further exploration of specific feature sets and advanced classification models.

1.5 Scope and Limitations

1.5.1 Data Collection

Make use of Parkinson's disease-related publically accessible datasets, making sure they include pertinent clinical and demographic data.

1.5.2 Feature Selection:

Examine how well the method, Principal Component Analysis (PCA) improve model performance.

1.5.3 Algorithm Optimization:

Try experimenting with various machine learning methods, such as Random Forests,Decision Tree,XG Boost,KNN and ANN.

1.5.4 Data Limitations:

The size, diversity, and completeness of the publicly accessible datasets that the study uses may be constrained. This can have an impact on how broadly applicable the results are.

1.5.5 Feature Quality:

The efficacy of machine learning models could be restricted by the caliber and pertinence of characteristics present in the datasets. It's possible that some crucial elements are absent or poorly depicted.

1.5.6 Algorithm Constraints:

Although the study may be limited by computational resources, perhaps restricting the complexity and scale of models that may be evaluated, a variety of machine learning algorithms will be tested.

1.5.7 Generalization:

The models that have been constructed in this study might not be very applicable to all populations or clinical settings because different locations and demographics have varied illness presentations and diagnostic standards.

1.6 Report Organization

This research explores the potential of machine learning algorithms to improve the diagnosis and management of Parkinson's disease (PD), a progressive neurodegenerative disorder characterized by tremors, rigidity, and slow movement. It investigates the effectiveness of feature selection algorithms in enhancing ML-based PD classification accuracy. The research uses publicly available PD datasets and features like PCA with Synthetic Minority Over-sampling Technique applied to address data imbalances. It compares various classification models to identify the most accurate combination for PD classification. This research explores machine learning applications in Parkinson's disease diagnosis, analyzing feature selection algorithms and classification models to improve accuracy, efficiency, and patient outcomes.

1.7 Summary

Parkinson's disease (PD) diagnosis currently faces limitations due to the subjectivity of clinical assessments. This research explores the potential of optimizing the accuracy of machine learning models for PD classification using various feature selection techniques. By leveraging publicly available PD datasets, the research investigates the effectiveness of feature selection algorithms like PCA in maximizing classification accuracy. Furthermore, the research compares the performance of diverse classification models such as K-NN, Random Forest, Decision Tree,XG Boost and ANN. Ultimately, it aims to identify the feature selection algorithm and classification model that yields the highest accuracy in PD classification on the chosen dataset. This research contributes to the growing knowledge of machine learning applications in PD diagnosis. It paves the way for further exploration of specific feature sets and advanced models, potentially leading to improved clinical practice and patient outcomes.

Key Points:

Focus: Optimizing machine learning accuracy for PD classification using feature selection.

Data: Publicly available PD datasets (with potential data balancing).

Feature Selection: Evaluation of techniques like PCA.

Classification Models: Comparison of K-NN, Random Forest, Decision Tree, XG Boost and ANN.

Outcome: Identifying the best combination for PD classification accuracy.

Impact: Contributing to knowledge on machine learning for PD diagnosis and paving the way for future advancements.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

In previous scientific literature, we see that most of the authors are using different popular ML techniques. Such as: KNN, Random Forest, Decision Tree, logistic regression, Forest PA, Binary Tree, Genetic Algorithms and others. Some of the author done the research just applied ML techniques and found accuracy. And some of them apply other techniques as classifier algorithms, feature extraction data partitioning, feature selection algorithms. Major Author's use voice datasets as their dataset for the research. They apply ML techniques into them, and few do the feature extraction or other techniques to find the best accuracy. Most of the cases the SVM technique performs the better accuracy without the feature extraction and if the author use classifier to the datasets.

2.2 Related Works

According to our study about related works, most of the authors did their research with the raw dataset. They don't give priority to extracting features and other algorithms to find the best feature that can bring great accuracy. In our findings we will use these features to find algorithms and techniques and to find the best value.

Vikas Mittal and co-authors explored PCA Classifier, KNN, Logistic Regression, Medium Gaussian Kernel, and SVM. They achieved an accuracy of 90.3%, indicating the effectiveness of dimensionality reduction techniques like PCA in enhancing classifier performance.

Enes Celik and colleagues employed Logistic Regression, SVM, Extra Trees, Gradient Boosting, Random Forest, Naive Bayes, KNN, and ANN. Their results indicated that SVM achieved the best performance with an accuracy of 91%, showcasing the algorithm's robustness in various settings.

Zayrit Soumaya and colleagues focused on Genetic Algorithm and SVM, achieving an accuracy of 91.18%. This study further supports the effectiveness of combining evolutionary algorithms with traditional classifiers to improve diagnostic accuracy.

In their study, Zehra Karapinar Senturk and co-authors applied Binary Tree, SVM, and Artificial Neural Network (ANN) algorithms. They found that SVM provided the best performance with an accuracy of 93.84%, demonstrating the algorithm's robustness in handling non-linear data for PD diagnosis.

Surekha Tadse and colleagues explored multiple algorithms including Decision Tree, Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). Among these, the Decision Tree algorithm achieved the highest accuracy of 94.87%, underscoring its effectiveness in classifying PD patients based on clinical data.

2.3 Comparison between existing works

According to our literature review we accomplish the following information. Author used ML techniques in the detection of the PD. The result is satisfactory, but they don't apply feature extraction techniques to their research. A few authors use the feature extraction method. We are going to use feature extraction methods with PCA technique and try to find the best features that can help us to lead to our goal. The authors who apply the other techniques and feature extraction methods are working better than other techniques. We are going to use various techniques to our datasets and then we will apply ML techniques. We will apply different techniques to different ML methods. Best feature extraction methos its most important to our works. We will apply the best features extraction method to different algorithms to do the best.

Table 2.1 Comparative analysis with previous work

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm
1.	Surekha Tadse et al. [1]	Decision Tree,Logistic Regression,K-Nearest Neighbours,Suport vector machine	Decissoin tree 94.87%
2.	Zehra Karapinar Senturk et al [2]	Binary tree,Svm, Ann	SVM=93.84%
3.	Rohit Lamba et al. [3]	Naïve Bayes,KNN,Random forest Classification, Geneteic algorithm	Genetic algorithm and Random Forest=95.58%
4.	Arvind Kumar Tiwari. [4]	bagging,boosting,random forest, rotation forest,random subspace,svm,multipalyer perception,decission tree.	Random forest=95.9%
5.	Shail Raval et al. [5]	Logistic regression,SVM,KNN,DT,SGD,RF,AB,HV	RF AND HV-97.5%
6.	N. P. Narendra et al. [6]	pipeline system,end to end system	68.56%
7.	Vikas Mittal. et al. [7]	PCA classifier,k-nn,logostic regression,medium gaussian kernal,support vector machine	90.3%
8.	Gunjan Pahuja et al. [9]	Multilayer Perceptron, Support Vector Machine and K-nearest neighbor,ANN	ANN=96.89%
9.	Moumita Pramanik et al. [10]	Forest PA,decision tree,Decision forest comparison	Decision tree=95%
10.	Enes Celik et al. [11]	Logistic regression, Support Vector Machine, Extra Trees, Gradient Boosting and Random Forest,Naive Bayes, K-Nearest	Support Vector Machine=91%

		Neighbour, Artificial Neural Network	
11.	Dr. C K Gomathy et al[12]	XGBoost,Decision tree, Navie Bayes	XGBoost=0.948
12.	Zayrit Soumaya et al[13]	Genetic algorithm,SVM	Genetic algorithm and SVM=91.18%

The Table 2.1 shows the most usages algorithms to the related works and also give a overview of the accuracy.

2.4 Open Issues

There are issues regarding the generalizability of study findings to larger groups due to the prevalence of studies that concentrate on certain demographic or clinical datasets. To provide reliable diagnostic tools that are relevant everywhere, future study should examine how algorithms work across a range of racial and cultural backgrounds, age groups, and geographic locations.

While they frequently achieve great accuracy, complex machine learning models like ensemble methods and neural networks are difficult to grasp. Comprehending the cognitive process that underlies these models is essential for gaining therapeutic acceptance and comprehension of disease causes.

Efficient integration of machine learning models into current healthcare procedures is essential for their successful application in clinical settings. In order to promote acceptance among healthcare professionals, it is imperative to tackle concerns pertaining to data accessibility, real-time processing, and user interface design.

Large-scale patient data is a prerequisite for machine learning models, which raises questions regarding data security, privacy, and moral application. Maintaining patient trust and regulatory compliance requires ensuring compliance with laws like GDPR and HIPAA as well as clear data management procedures.

2.5 Summary

Previous scientific literature primarily uses popular machine learning techniques like KNN, Random Forest, Decision Tree, logistic regression, Forest PA, Binary Tree, and Genetic Algorithms. Some researchers apply these techniques alone, while others use classifier algorithms, feature extraction, and feature selection algorithms. SVM technique often performs better with feature extraction.

The present literature review delves into the latest developments in the application of machine learning algorithms for the diagnosis of Parkinson's disease (PD), emphasizing diverse approaches and results from numerous studies.

For their potential to evaluate clinical data and increase diagnostic accuracy, machine learning approaches including Decision Trees, Support Vector Machines (SVM), Random Forests, Genetic Algorithms, and Artificial Neural Networks (ANN) have been thoroughly studied. The analyzed studies show a considerable variation in algorithm performance, with stated accuracies varying from over 90% to 97.5% .

Important discoveries include how well ensemble techniques, such as Random Forest and Genetic Algorithms, handle complicated datasets and maximize feature selection. SVM and ANN have also demonstrated strong performance in managing non-linear data patterns, which increases its usefulness in the diagnosis of Parkinson's disease.

CHAPTER 3
Research Methodology

3.1 Overview

"Early parkinson disease detection with feature extraction using machine learning" is a research project that focuses on applying cutting-edge AI methods to the field of medical diagnostics, with a particular focus on PD detection. Parkinson's disease (PD) is a progressive illness that affects the neurological system and the body parts that are nerve-controlled. The study explores the difficulties posed by conventional PD detection techniques and aims to solve them by combining the best machine learning techniques with the most optimal feature selection.

Data dimensionality can be reduced using a statistical method called principal component analysis (PCA). The original variables are changed into a new set known as the principle components, which are orthogonal to one another. The degree of variance that each of these primary components extracts from the data determines their ranking. The greatest variance is captured by the first principal component, then by the second, and so forth. PCA aids in data simplification by lowering the number of dimensions without sacrificing important information. This facilitates the visualization and interpretation of complicated datasets. In domains including finance, bioinformatics, and machine learning, PCA is extensively utilized. Additionally, it aids in feature extraction and noise reduction. The data covariance matrix's eigenvalue decomposition is a step in the process.

The main focus of the study is on how different machine learning techniques, including XGBoost, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Artificial Neural Networks (ANN) are applied to identify Parkinson's disease. The study's comparative analysis section explores the distinctions and overlaps between these various classification strategies. Classification entails determining whether Parkinson's disease is present based on particular biomarkers and trends seen in medical records. By looking at various approaches, the study hopes to give a thorough grasp of their advantages and disadvantages, guiding the creation of more reliable and efficient diagnostic instruments.

Primarily, the study focuses on the nexus of artificial intelligence, medical diagnosis, and neurodegenerative disease identification, with a particular focus on enhancing the diagnosis of Parkinson's disease by means of sophisticated machine learning methodologies. The results of this study are intended to benefit not just the medical community's theoretical understanding but also its practical application, with the ultimate goal of enhancing patient outcomes and diagnostic precision.

3.2 System Design

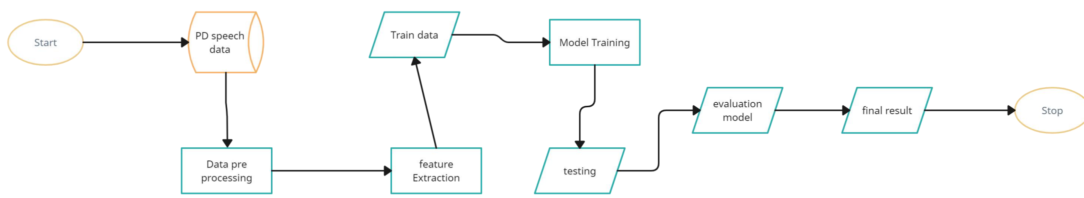

Figure 3.1: Proposed Methodology

Figure 3.1 expose the proposed methodology of our Algorithms.

3.3 Hardware/Software Requirements

There are particular prerequisites for the implementation of the research on "early parkinson disease detection with feature extraction using machine learning" in order to guarantee the study's effective completion. These implementation requirements cover aspects like data gathering and model evaluation in addition to hardware and software components.

3.3.1 Hardware Requirements:

High-performance computing resources: Having access to a computing infrastructure with enough memory and processing power to execute the intricate tasks associated with data analysis and machine learning.

GPU, or graphics processing unit: In order to drastically reduce the time needed for model creation, a GPU is necessary for speeding up the training of machine learning techniques (ML techniques including XC Boost, Artificial Neural Networks (ANN), Decision Tree, XGBoost, K-Nearest Neighbors (KNN), and Random Forest).

3.3.2 Software Requirements:

Machine learning frameworks: the creation, training, and assessment of machine learning models using well-known frameworks like PyTorch or TensorFlow.
The Python programming language offers a flexible and extensively utilized framework for executing deep learning algorithms and carrying out data analysis.

3.3.3 Data Collection:

Diverse and representative dataset: To guarantee the stability and generalizability of the created models, a thorough dataset of medical voice cases comprising both normal and Parkinson's disease cases was gathered.
Annotated dataset: The availability of data with precise labels identifying the presence or absence of Parkinson's disease for the purpose of training and validating the models.

3.3.4 Model Training and Evaluation:

PCA**(**Principal Component Analysis**):** The method known as Principal Component Analysis (PCA) is used to decrease the dimensionality of data while preserving the majority of its variation. The original variables are converted into a new collection of uncorrelated variables known as principle components, which are arranged according to how much variance they can account for. PCA aids in classification jobs by streamlining the dataset, facilitating analysis and visualization without sacrificing important information. When working with high-dimensional data, this technique is especially helpful as it improves the effectiveness and performance of several machine learning algorithms. In general, PCA is an effective approach for preparing data in classification issues.

SVM**:** An effective supervised learning approach for regression and classification problems is called Support Vector Machine (SVM). It operates by determining which hyperplane in the feature space best divides the various classes. By increasing the margin between the closest data points for each class—referred to as support vectors—the ideal hyperplane can be found. Because SVM can handle non-linear data using a variety of kernel functions (linear, polynomial, radial basis function, etc.), it is especially useful in high-dimensional fields and adaptable. SVM is a popular option in many applications, including text classification, picture recognition, and bioinformatics, because of its high accuracy and robustness.

KNN:A straightforward and user-friendly supervised learning technique for classification and regression problems is K-Nearest Neighbors (KNN). It categorizes data points in the feature space according to the majority class of its k-nearest neighbors. KNN works well in many real-world applications and is efficient for small datasets despite its simplicity; nevertheless, it can be computationally demanding for large datasets.

XG Boost**:**Based on the gradient boosting framework, XGBoost (Extreme Gradient Boosting) is a machine learning technique that is incredibly effective and potent. By constructing decision trees iteratively—each tree correcting the mistakes of the preceding one—and optimizing for speed and accuracy using sophisticated methods like regularization and parallel processing, it improves model performance. XGBoost's scalability and capacity to manage big datasets with intricate patterns make it a popular choice for both real-world and competitive applications.

Decision Tree:A Decision Tree is a versatile machine learning algorithm used for classification and regression tasks. It works by recursively splitting the data into subsets based on the value of input features, creating a tree-like model of decisions. Each internal node represents a feature, each branch represents a decision rule, and each leaf node represents an outcome, making decision trees easy to interpret and understand**.**

Random Forest:Random Forest is an ensemble learning method used for classification and regression tasks. It builds multiple decision trees during training and merges their results to improve the model's accuracy and robustness. By averaging the predictions of many trees, Random Forest reduces overfitting and enhances generalization, making it a powerful and reliable algorithm for a wide range of applications.

ANN:Ann, short for artificial neural networks, is a fundamental concept in machine learning, mimicking the human brain's interconnected neurons to process complex data. Through layers of interconnected nodes, Ann models can learn patterns and make predictions, driving advancements in fields from computer vision to natural language processing.

3.4 Project Management and Financial Analysis

● Project Objectives:

A. To find the disease in the early stage.

B. To give effective solution for patients of Parkinson disease

C. To provide a platform to detect if the patient is actually suffering by this disease.

● Project Timeline:

A. Phase 1: Research and Project Planning (2-4 weeks)

B. Phase 2: Dataset collecting (6-8 weeks)

C. Phase 3: Dataset filtering (8-12 weeks)

D. Phase 4: Data pre-processing (10-14 weeks)

E. Phase 5: Train data (4-6 weeks)

F. Phase 6: Model training (2-4 weeks)

G. Phase 7: Evaluation Model (4-6 weeks)

● Resource Planning:

A. Equipment and Tools:

● Environment for Development and Testing

● Superb Computing Systems for Progress Design of ML Algorithms

● Version Control System.

B. Software and Technologies:

● Need Jupiter notebook or Google Co-laboratory.

● Back-End Technologies (e.g., Node.js, Django, Flask)

● Database Management System

C. Data and Content:

● Product Data: Gathered by online source or sourced through medical databases/APIs

● User Documentation: Prepared by our own team

D. Training and Skill Development:

● Ensure that we have skills in Machine Learning, DSA , database management and backend development.

● Provide additional training on specific technologies and tools as needed.

● Communication Plan:

A. Visiting Hospitals:

Purpose: For Update dataset .

Participants: Team members.

Frequency: Bi-monthly or as specified in the project plan.

B. Change Control Meetings:

Purpose: Discuss for any changes to the project scope or others with supervisor.

Participants: Supervisor,and team members.

Frequency: As needed to full fill our research.

Table 3.1 Total Cost

SN	Components	Cost (BDT)
01.	Visiting Stakeholders	500-1000
02.	Data Collection and Processing	500-1000
03.	Documentation and Report Writing	500-1000
04.	Contingency	1000-1500
	Total Cost	2500-4500

Table 3.1 gives the overview of our final cost of this research.

3.5 Summary

The literature review suggests that SVM method is commonly used for accuracy, but it doesn't include classifiers, feature extractors, or data partitioning. To achieve best accuracy, feature extraction and suitable ML techniques will be applied to the dataset. And a proposed design figure mention in the paper.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

In this study, the implementation process begins with data visualization, followed by thorough checks for null and missing values. Principal Component Analysis (PCA) is then applied across various dimensions to optimize classifier performance. Subsequently, a model is loaded and trained using a diverse set of machine learning algorithms including Random Forest, XG Boost, Decision Trees, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). This comprehensive approach aims to explore and optimize predictive capabilities across different methodologies.

4.2 Train Model/Prototype Design

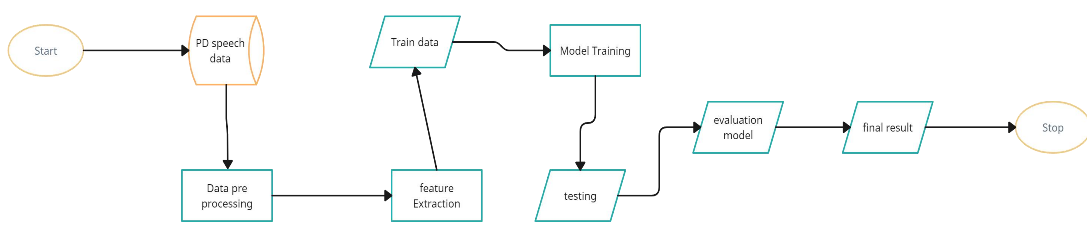

Figure 4.1: Train Model

Figure 4.1 is our applied algorithm to train our model.

4.3 System Testing:

In the testing phase, a Python script accepts an array of numerical features as input to predict the presence of Parkinson's disease using a pre-trained machine learning model. The script utilizes the model to make predictions based on the input data, assessing whether the individual is likely affected by Parkinson's or not, thereby validating the model's predictive capabilities in a practical application scenario.

4.4 Summary

The implementation overview included initial steps of data visualization, handling null values, applying PCA for dimensionality reduction, and training multiple machine learning algorithms. The proposed methodology focused on leveraging these steps to build and optimize a predictive model for Parkinson's disease detection. A Python code snippet was developed for testing purposes, where an array of numerical features is inputted to predict the Parkinson's disease based on the trained model's predictions. This approach aims to validate the model's effectiveness and practical application in diagnosing the disease.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

In this study, PCA was applied to the dataset with varying values of PCA (e.g.100%, 90%, 85%, 80% etc.) to optimize feature selection and improve model performance. The models were then retrained with these reduced datasets, showing varying results. Notably, when PCA was set to 85% the Random Forest algorithm achieved the highest accuracy of 92%. This demonstrates the effectiveness of PCA in enhancing the model's predictive accuracy by selecting the most relevant features.

5.2 Experiment/ Simulation result

This below graphical representation of the correlation matrix showing the correlation coefficients between variables. Each cell in the grid represents the correlation between two variables and is color-coded according to the value of the correlation coefficient indicated by the color bar on the right side of the heatmap. The colors range from red (1.00) indicating a perfect positive correlation, through white (0), indicating no correlation, to blue (-1.00) indicating a perfect negative correlation.

The heatmap includes labeled rows and columns with various abbreviations such as MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), HNR, RPDE, DFA, spread1, spread2, D2, PPE, among others. These abbreviations likely represent different parameters or features related to voice or sound analysis given their naming convention.

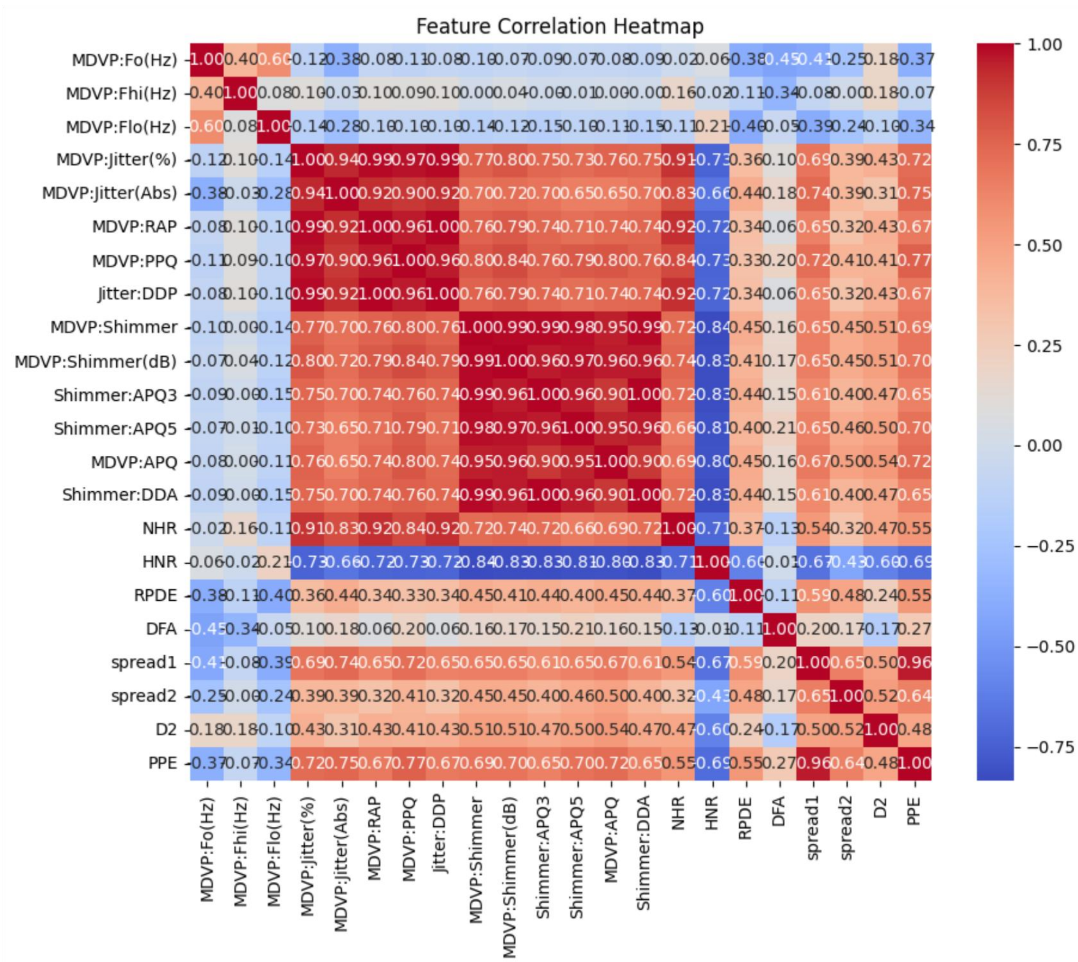

Figure 5.1: Correlation Heat Map of the Features

After analyzing the the heatmap we consider 14 features for our machine learning techniques and train our algorithms.Here is the accuracy results for the algorithms.

Table 5.1 Accuracy Results

Algorithms	**Accuracy(%)**
Decision Tree	87
XG Boost	**89**
KNN	84
Random Forest	84
ANN	79

We can see a accuracy percentage of our train model after correlation.(Table 5.1)

We have taken the XG Boost as our best performing model for the selected features.

After then apply the PCA variance we get a better result before.below table shows the comparison.we try the variance by the component into (100%, 90%, 85%,80%) and so on.when we apply the 85% variance to the attributes we got a better result in the algorithms.

Table 5.2 Simulation Result of machine learning algorithms with PCA

Decision Tree(%)	**XG Boost(%)**	**KNN(%)**	**Random Forest(%)**	**ANN(%)**	**PCA variance(%)**
84	76	84	89	76	100
79	76	84	89	76	90
79	**79**	**84**	**92**	**79**	**85**
79	76	84	89	76	80
76	74	84	82	76	70
76	76	84	84	76	60
79	76	84	82	76	50

This Table 5.2 is a simulation of the machine learning algorithms when we applies the variance of the classifier PCA

5.3 Performance/Comparative Analysis

Two phases of machine learning were conducted in this study: one without a PCA classifier and the other in selected features with correlation. It's interesting to note that using PCA really enhanced the outcomes, especially when using the Random Forest technique. This improvement emphasizes how important PCA is for improving feature selection, which results in more precise and efficient model performance.

Table5.3 Comparative analysis of selected features training algorithms and PCA applied training algorithms accuracy

Algorithms	Selected features(%)	PCA applied features(%)
Random Forest	84	92
KNN	84	84
XG Boost	89	79
Decision Tree	87	79
ANN	79	79

The Table 5.3 shows the final comparison accuracy of the Correlation applied algorithms and PCA applied algorithms.

After applying the classifier(PCA) table shows the the difference of the accuracy variety and Random forest perform the best accuracy with 92%.Other hand XG Boost performance has dropped to 79%.

5.4 Summary

In this Parkinson's disease detection study, two machine learning models were evaluated: one without using the PCA classifier and another with the PCA classifier. The model without PCA yielded poor testing results, indicating suboptimal feature selection and performance. Conversely, the model incorporating PCA significantly improved testing accuracy, particularly enhancing the performance of the Random forest. This highlights the importance of dimensionality reduction techniques like PCA in improving model accuracy and robustness in medical diagnosis applications.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The lives of patients, their families and medical professionals may be significantly impacted by the use of machine learning algorithms in the diagnosis and treatment of Parkinson's disease (PD). This section explores the various ways in which these developments enhance both the well-being of the individual and the group.

Early and accurate diagnosis of PD allows for timely intervention, which can significantly slow disease progression and manage symptoms more effectively.. Patients may have a higher quality of life as a result, retaining their independence and capacity to carry out everyday tasks for extended periods of time.

More accurate diagnostic instruments enable the creation of individualized treatment programs that are catered to the unique requirements of every patient. This can raise overall patient well-being, lessen adverse effects, and increase therapeutic efficacy.

Families are better able to make decisions on care and lifestyle modifications when a diagnosis is apparent. Enhancing communication, fortifying family ties, and creating a nurturing atmosphere for the patient can all result from this.

By giving medical practitioners more dependable diagnostic tools, machine learning algorithms can boost their confidence in making Parkinson's disease diagnoses. This may result in quicker and more accurate diagnosis, which would benefit the patient.

6.2 Impact on society and Environment

The broader society is greatly impacted by the use of Parkinson's disease (PD) in the diagnosis and treatment of Parkinson's disease (PD). These developments have the potential to enhance public health, provide economic gains, and promote increased social acceptance and assistance for those impacted by the illness. This section examines the complex societal effects of machine learning-assisted PD diagnosis.

6.2.1 Public Health Improvement

Algorithms for machine learning make it easier to identify Parkinson's disease (PD) early and accurately, allowing for prompt medical attention. Early detection can greatly decrease the course of a disease, improving outcomes for patients and lessening the overall toll that the illness has on the healthcare system. A population that is healthier and has fewer severe cases requiring expensive medical treatment can result from this preventive strategy

Tracking the course of an illness and the efficacy of treatment can be aided by machine learning's predictive analytics and continuous monitoring. In the long run, this can improve the management of chronic disorders like Parkinson's disease (PD) by resulting in more adaptable and dynamic healthcare practices.

6.2.2 Economic Benefits

Machine learning can save healthcare expenses by improving diagnostic accuracy and lowering the need for multiple tests and misdiagnosis. More focused therapies result from precise and efficient diagnosis, which lowers the cost burden on patients and healthcare systems.

Patients with Parkinson's disease (PD) can prolong their time in the workforce and preserve their functional abilities with early diagnosis and appropriate medication. This may have favorable financial effects by lowering the financial losses brought on by incapacity and early retirement from Parkinson's disease.

6.2.3 Environment

Multiple environmental implications arise from the use of machine learning (ML) in the diagnosis and treatment of Parkinson's disease (PD). Even while enhancing patient outcomes and healthcare efficiency is frequently the main priority, it's important to take into account how these technological improvements effect environmental sustainability. This section examines the multiple environmental effects of machine learning (ML) in healthcare, with a focus on PD diagnosis.

6.3 Ethical Aspects

While there are many advantages to using machine learning in the diagnosis and treatment of Parkinson's disease (PD), there are also significant ethical issues to be addressed. It is imperative to tackle these issues in order to guarantee the conscientious and fair application of these technologies. Key ethical considerations pertaining to patient consent, algorithmic bias, transparency, data privacy, and equitable access are outlined in this section.

Large volumes of data, frequently containing sensitive patient information, are needed for machine learning models. It is crucial to protect the privacy of this data. To safeguard patient privacy, strong data encryption techniques, safe storage options, and stringent access controls must be put in place.

Patient data should be anonymised whenever feasible to further protect privacy. By eliminating or obscuring personal identifiers from datasets, anonymization preserves the usefulness of the research while making it challenging to link specific patient data.

The proper development and application of these technologies depends on addressing the ethical issues surrounding the integration of machine learning into the diagnosis of Parkinson's disease. Important ethical considerations include protecting patient privacy, obtaining informed permission, ensuring fairness, openness, accountability, and enhancing clinical judgment.

6.4 Sustainability Plan

The sustainability plan outlines the necessary tactics and protocols to ensure the environmental, fiscal, and social sustainability of the application of machine learning (ML) to Parkinson's disease (PD) diagnosis and treatment. This technique aims to maximize long-term advantages while minimizing negative impacts.

Use renewable energy sources like solar and wind power in data centers and implement energy-efficient techniques. Reduce energy usage by optimizing server workloads and making use of energy-efficient hardware.

Use machine learning (ML) to cut down on the number of diagnostic tests required, saving money on medical supplies. Use predictive analytics to make sure that tests are only run when absolutely required.

Promote the use of environmentally friendly materials in the creation of ML hardware and diagnostic tools. Prioritize longevity and recyclable materials to reduce your environmental footprint.

6.5 Summary

The use of ML in PD diagnosis has a big social impact. ML-based solutions can enhance patient outcomes and quality of life while lessening the overall strain on healthcare systems by enabling early and accurate detection. Through resource allocation optimization and the reduction of unnecessary testing, these technologies can effectively reduce healthcare expenditures. Patients who receive an early diagnosis are able to retain their functional abilities for longer, which increases economic production and lowers losses related to disability. Furthermore, ML increases public knowledge and comprehension of Parkinson's disease (PD), providing patients and their families with improved information and assistance. Stronger support networks and more proactive health management may result from this, improving social welfare.

The use of machine learning in PD diagnoses improves environmental sustainability. Through the reduction of many diagnostic tests and procedures, these technologies help to limit waste generation and conserve medical resources. Medical equipment lasts longer and produces less electronic waste when it is used optimally. Energy consumption in data centers can be reduced by using energy-efficient algorithms and advances in green computing, even though machine learning models demand a lot of processing power.

A comprehensive sustainability plan ensures that the integration of ML into PD diagnostics remains environmentally, economically, and socially responsible. Key components of this plan include. Implementing green data centers and optimizing ML algorithms to reduce energy consumption.

Reducing redundant testing and optimizing medical equipment usage to conserve resources.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

Parkinson's disease (PD) is a neurological disease that worsens with time and affects millions of people globally. The diagnosis and treatment of Parkinson's disease (PD) remain difficult despite tremendous advances in medical science and practice. The use of machine learning (ML) as a transformative tool to improve the precision and effectiveness of Parkinson's disease (PD) diagnosis has been investigated in this thesis.

By using sophisticated feature selection strategies and classification models, ML algorithms have demonstrated potential to increase the precision of Parkinson's disease diagnosis. ML can identify early indications of Parkinson's disease (PD) that conventional clinical examinations might overlook by examining intricate patterns in medical data.

Early detection of Parkinson's disease (PD) by machine learning (ML) can result in prompt medical therapies that may decrease the disease's progression and enhance patient outcomes. Better symptom management and more successful treatment strategies are made possible by early diagnosis.

Multiple tests and consultations can be avoided with the help of ML-driven diagnostics, which can expedite the diagnosis procedure. This optimization can increase the effectiveness of healthcare delivery, reduce healthcare expenses, and preserve medical resources.

In conclusion,the application of machine learning to the diagnosis and treatment of Parkinson's disease is a noteworthy development in the medical field. ML has the potential to revolutionize the diagnosis and treatment of Parkinson's disease (PD) by increasing diagnostic accuracy, facilitating early identification, and maximizing resource usage. To maximize these technologies' beneficial effects on people and society as we create and improve them, we must remain dedicated to sustainability, moral behavior, and patient empowerment.

7.2 Further Suggested Works

There are still many opportunities for further study and advancement, even after this thesis examined the application of machine learning (ML) in the diagnosis and treatment of Parkinson's disease (PD) and illustrated its possible advantages. By addressing these issues, machine learning (ML) in healthcare can be applied more effectively, efficiently, and morally.

7.2.1 PCA Classification Attributes:

We applied the PCA classifier to out datasets and find the best accuracy on the variance percentage of 85%. Future findings may be conduct to find the 85% exact attributes.

7.2.2 Deep Learning Approaches:

Examine the application of deep learning methods for Parkinson's disease diagnosis, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs). These techniques have the ability to increase the precision of diagnostic models by identifying intricate patterns in time-series.

7.2.3 Ensemble Methods:

Create and assess ensemble learning techniques that boost diagnostic performance by combining many ML algorithms. Ensemble techniques like boosting, bagging, and stacking help improve the accuracy and robustness of models.

7.3 Limitations/ Conflict of Interests

7.3.1 Data Privacy:

There are serious privacy issues when using patient data for machine learning research. It is imperative to guarantee that all the data utilized in this study conforms to applicable data protection laws, including GDPR and HIPAA. Achieving this compliance, though, can be difficult and might restrict the amount of data that is available.

7.3.2 Computational Resources:

Creating and refining sophisticated machine learning models demands substantial processing power. Restrictions on the availability of high-performance computing infrastructure may limit the extent and magnitude of the study.

7.3.3 Compute the applied variance attributes:

We applied PCA classifier into our datasets but we can’t find the exact attributes for the classify datasets.

Reference:

[1] Senturk, Zehra Karapinar. "Early diagnosis of Parkinson’s disease using machine learning algorithms." *Medical hypotheses* 138 (2020): 109603.

[2] Pahuja, Gunjan, and T. N. Nagabhushan. "A comparative study of existing machine learning approaches for Parkinson's disease detection." *IETE Journal of Research* 67.1 (2021): 4-14.

[3] Celik, Enes, and Sevinc Ilhan Omurca. "Improving Parkinson's disease diagnosis with machine learning methods." *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*. Ieee, 2019

[4] Lamba, Rohit, et al. "A hybrid system for Parkinson’s disease diagnosis using machine learning techniques." *International Journal of Speech Technology* (2021): 1-11.

[5] Nissar, Iqra, Waseem Ahmad Mir, and Tawseef Ayoub Shaikh. "Machine learning approaches for detection and diagnosis of Parkinson’s disease-a review." *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. IEEE, 2021.

[6] Prashanth, R., et al. "High-accuracy detection of early Parkinson's disease through multimodal features and machine learning." *International journal of medical informatics* 90 (2016): 13-21

[7] Gomathy, Dr CK, et al. "The Parkinson’s Disease Detection Using Machine Learning Techniques." *International Research Journal of Engineering and Technology (IRJET)* 8.10 (2021): 2395-0056.

[8] Challa, Kamal Nayan Reddy, et al. "An improved approach for prediction of Parkinson's disease using machine learning techniques." *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*. IEEE, 2016.

[9] Pramanik, Moumita, et al. "Machine learning methods with decision forests for Parkinson’s detection." *Applied Sciences* 11.2 (2021): 581.

[10] Raval, Shail, Rahil Balar, and Vibha Patel. "A Comparative Study of Early Detection of Parkinson's Disease using Machine Learning Techniques." *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE, 2020.

[11] Mandal, Indrajit, and N. Sairam. "New machine-learning algorithms for prediction of Parkinson's disease." *International Journal of Systems Science* 45.3 (2014): 647-666

[12] Miljkovic, Dragana, et al. "Machine learning and data mining methods for managing Parkinson’s disease." *Machine learning for health informatics: state-of-the-art and future challenges* (2016): 209-220.

[13] Tadse, Surekha, Muskan Jain, and Pankaj Chandankhede. "Parkinson's Detection Using Machine Learning." *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021.

[13] Tiwari, Arvind Kumar. "Machine learning based approaches for prediction of Parkinson’s disease." *Mach Learn Appl* 3.2 (2016): 33-39.

[14] Soumaya, Zayrit, et al. "The detection of Parkinson disease using the genetic algorithm and SVM classifier." *Applied Acoustics* 171 (2021): 107528.

[15]Little, M., McSharry, P., Hunter, E., Spielman, J., & Ramig, L. (2008). Suitability of dysphonia measurements for telemonitoring of Parkinson’s disease. Nature Precedings, 1-1.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:

EARLY PARKINSON DISEASE DETECTION WITH FEATURE EXTRACTION USING MACHINE LEARNING

Student ID: 203-15-3868,203-15-3896

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **parkinson detection** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**

CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by employing Machine Learning Techniques,Data visualization, and apply feature extraction classifier.The project demonstrates **specialist knowledge (K4)** by conducting Medical science datasets,its symptoms,Data collection methods.	**Page no:** [1-6] **Page no:** [2-9]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing Machine Learning Models.	**Page no:** [13,20,23]
				This project ensures to **K8 (Research Literature)** by synthesizing insights from recent studies,to Parkinson Disease detection using machine	**Page no:** [7-11]

				learning, showcasing a comprehensive understanding of current methodologies.	
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in parkinson disease detection, including limitations of traditional methods. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [7-11]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting PCA classification as the chosen significant solution for enhancing parkinson disease detection amidst multiple potential approaches.	**Page no:** [10]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting medical diagnostics in neurological disorder like parkinson disease, contributing to advancements in public health and healthcare practices which indicates **EP-4**.	**Page no:** [10]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting requirements**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in medical diagnostics which ensures **EP-7**.	**Page no:** [12-16]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in parkinson disease detection through classification and machine learning.	**Page no:** [13-16]
2.	**EA2: Level of interaction**	No		N/A	N/A

3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project contributes to society by improving healthcare through advanced parkinson disease detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.	**Page no:** [26-30]
5.	**EA-5: Familiarity**	Yes		This project expands upon existing research by examining a novel approach in parkinson disease detection machine lerning and classification demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [17-20]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [17-19]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing patient privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible	**Page no:** [26-29]

			knowledge dissemination and societal well-being through the ethical application of advanced healthcare technologies which comply with **CO9**.	
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [12-19, 20-21]

NAfi

ORIGINALITY REPORT

23%	16%	14%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	6%
2	www.mdpi.com Internet Source	1%
3	Surekha Tadse, Muskan Jain, Pankaj Chandankhede. "Parkinson's Detection Using Machine Learning", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021 Publication	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	www.ijnrd.org Internet Source	1%
6	Submitted to University of Technology, Sydney Student Paper	1%
7	"Proceedings of the 2nd International Conference on Big Data, IoT and Machine	<1%