**Machine Learning Algorithms for Diagnosis of Parkinson's Disease Based on Voice Characteristics**

**A PROJECT REPORT**

***Submitted by***

**P. Sai Manoj (19113101)**

**K. Sumanth Kumar Reddy (19113105)**

**Under the guidance of**

**Dr. M. Sambath**

**Associate Professor**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

**COMPUTER SCIENCE AND ENGINEERING**



**HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE**

**CHENNAI - 603 103**

MAY 2023



**BONAFIDE CERTIFICATE**

Certified that this project report **Machine Learning Algorithms for Diagnosis of Parkinson's Disease Based on Voice Characteristics** is the bonafide work of **P. Sai Manoj (19113101)** and **K. Sumanth Kumar Reddy (19113105)** who carried out the project work under my supervision during the academic year **2022-2023**.

Dr. J. THANGAKUMAR, Dr. M. SAMBATH,

**HoD,**

Department of CSE,

Hindustan Institute of Technology & Science.

**ASSOCIATE PROFESSOR,**

Department of CSE,

Hindustan Institute of

Technology & Science.

**INTERNAL EXAMINER EXTERNAL EXAMINER**

Name: Designation:

Name:

Designation:

Project Viva - voce conducted on

2

**ACKNOWLEDGEMENT**

First and foremost, we would like to thank **ALMIGHTY** who has provided us the strength to do justice to our work and contribute our best to it.

We wish to express our deep sense of gratitude from the bottom of our heart to our guide **Dr. M. Sambath**, **Associate Professor,** **Department of Computer Science and Engineering,** for his motivating discussions, overwhelming suggestions, ingenious encouragement, invaluable supervision, and exemplary guidance throughout this project work.

We would like to extend our heartfelt gratitude to **Dr. J. Thangakumar, Associate Professor & Head, Department of Computer Science and Engineering** for his valuable suggestions and support in successfully completing the project.

We wish to thank our Project Co-Ordinator and Panel members for keeping our project in the right track. We would like to thank all the teaching, technical and non-technical staff of Department of Computer Science and Engineering for their courteous assistance.

We thank the management of **HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE** for providing us the necessary facilities and support required for the successful completion of the project.

As a final word, we would like to thank each and every individual who have been a source of support and encouragement and helped us to achieve our goal and complete our project work successfully

**ABSTRACT**

Parkinson's disease (PD) is a central nervous system disorder that primarily affects the motor system and progresses over time. Non-motor symptoms typically develop over time and become more common as the condition worsens. Parkinson's disease is also known as spontaneous or intrinsic parkinsonism, hypokinetic stiffness syndrome, paralyzed agitans, and shaking palsy. Tremors, bradykinesia, rigidity, and postural instability are symptoms of PD, a progressive neurological condition that affects the motor system. Deep brain stimulation, medicine, and other therapies can all be used to treat the symptoms of PD, but there is no known cure as of yet. For optimal disease care and the creation of new medicines, Parkinson's disease must be identified as early and accurately as possible.

The goal of this work is to create a model that, using relevant clinical and demographic data, can identify Parkinson's disease. It makes use of a range of machine learning techniques, including ensemble techniques like Gradient Boosting Classifier and Random Forest Classifier, as well as Logistic Regression, the Support Vector Machine (SVM), K-nearest-neighbors (KNN), and others. A significant dataset that contains data on changes in speaking patterns was used to train the model. Using measures for accuracy, recall, precision, and F1-score, the machine learning model's performance is assessed and contrasted with that of the most widely used techniques for Parkinson's disease diagnosis.

**CHAPTER**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **NO.** |  | |
|  | **Acknowledgement** | **i** |
|  | **Dedication** | **ii** |
|  | **Abstract** | **iii** |
|  | **List of Figures** | **iv** |
|  | **List of Abbreviations** | **v** |
| **1** | **INTRODUCTION** | **1** |
|  | **1.1** Overview | **1** |
|  | **1.2** Motivation for the project | **2** |
|  | **1.3** Problem Definition and Scenarios | **2** |
|  | **1.4** Organization of the report | **3** |
|  | **1.5** Summary | **3** |
| **2** | **LITERATURE REVIEW** | **4** |
|  | **2.1** Introduction  **2.2** Literature Review  **2.3** Summary | **4**  **5**  **8** |
|  |  |  |

**TITLE PAGE NO.**

1. **PROJECT DESCRIPTION 9**
   1. Introduction **9**
   2. Objective **9**
   3. Existing System **9**
   4. Shortcomings of Existing System **10**
   5. Proposed System **10**
   6. Benefits of [Proposed System](#_TOC_250019) **11**
2. SYSTEM DESIGN 12
   1. Introduction **12**
   2. Dataset Description **12**
   3. Normalization **13**
   4. Machine Learning Algorithms **14**
   5. Ensemble Method **15**
   6. Evaluation **16**
   7. Evaluation Metrics **16**
   8. [Summary **1**](#_TOC_250017)**7**
3. PROJECT REQUIREMENTS 18
   1. Hardware and Software Specification **18**
   2. [Technologies Used **1**](#_TOC_250013)**8**
4. MODULE DESCRIPTION 20
   1. Introduction **20**
   2. Modules **20**
5. IMPLEMENTATION 23
   1. Introduction **23**
   2. [User Interface](#_TOC_250006) **23**
6. RESULT ANALYSIS 26
   1. Introduction **26**
   2. Results **26**
7. CONCLUSION AND FUTURE WORK 31
   1. Conclusion **31**
   2. [Future Work](#_TOC_250003) **32**
8. INDIVIDUAL TEAM MEMBER’s

**33**

REPORT

* 1. [Individual Objective](#_TOC_250001) **33**
  2. Role of the Team Members **33**
  3. Contribution of Team Members **33**

[REFERENCES](#_TOC_250000) 34

**APPENDIX A : SAMPLE SCREEN**

**APPENDIX B : SAMPLE CODE**

**APPENDIX C : PLAGIARISM REPORT**

**APPENDIX D : FUNDING / PATENT /**

**PUBLICATION DETAILS**

**APPENDIX E : TEAM DETAILS**

|  |  |  |
| --- | --- | --- |
|  | **LIST OF TABLES** |  |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 8.1 | Classification Outcomes Using the Ensemble Technique and All Five Machine Learning Algorithms | 27 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF FIGURES** |  |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| 4.1 | Architecture Diagram | 13 |
| 7.1 | ipynb notebook | 23 |
| 8.1  8.2 | Confusion matrix in the two class classification assignments  Roc curve for the five-machine learning algorithm and ensemble method | 28  29 |

# LIST OF ABBREVIATIONS

|  |  |
| --- | --- |
| AUC  CBS  DTI  HRI  IMU | Area Under the Curve  Corticobasal Syndrome  Diffusion Tensor Imaging  Harmonics-To-Noise Ratio  Inertial Measurement Unit |
| KNN  MFCC  MGSVM  MRI  MSA  MS-SGCN  NLP  PCA  PD | K-Nearest Neighbors  Mel Frequency Cepstral Coefficients  Medium Gaussian Kernel support vector machine  Magnetic Resonance Imaging  Multiple Systems Atrophy  Multi-Scale Sparse Graph Convolutional Network  Natural Language Processing  Principal Component Analysis  Parkinson's disease |
| PSP | Progressive Supranuclear Palsy |
| RBF  ROC | Radial Basis Function  Receiver Operating Characteristic |
| SVM | Support Vector Machine |

**CHAPTER 1**

**INTRODUCTION**

## Overview

Parkinson's disease (PD) is a neurological condition that affects the motor system and progresses over time. It is characterized by tremors, bradykinesia, rigidity, and postural instability. There is presently no cure for PD, despite the serious effects it has on affected people and their family. For optimal disease care and the creation of new medicines, PD must be identified as early and accurately as possible. PD is believed to be caused by a synthesis of inherited and environmental factors, while its precise cause is unknown. The hallmark of Parkinson's disease (PD) is the degeneration of dopamine-producing neurons in the substantia nigra, a part of the brain responsible for controlling movement.

The motor symptoms of PD are caused by a decrease in dopamine levels in the brain as a result of this degradation. PD has traditionally been identified through clinical evaluations, such as the assessment of motor symptoms, medical background, and family history. These approaches, however, are arbitrary and open to inaccuracy. Machine learning algorithms have showed potential as a tool for Parkinson's disease diagnosis in recent years. The goal of this research is to create a machine learning model for diagnosing Parkinson's disease.

The most obvious symptoms, which are motion (motor) related, are tremor, the sensation of brad stiffness, and shuffling/stooped gait. Dysautonomia, neuropsychiatric problems (changes with mood, cognition, behavior, or thought), sensory symptoms (especially a diminished sense of smell), and sleep troubles are examples of non-motor symptoms. Before suffering motor symptoms, people may first experience non-motor symptoms like constipation, anosmia, and REM Behavior Disorder. In general, memory loss, mental illness, orthostasis, and more severe falls develop later. The model will integrate Gradient Boosting Classifier and Random Forest Classifier using algorithms like Logistic Regression, SVM, KNN, and ensemble approaches after being trained on a relevant dataset that contains information on changes in speaking patterns. Accuracy, recall, precision, and F1-score metrics will be used to assess the performance of the machine learning model. The initiative has the potential to increase the speed and accuracy of Parkinson's disease diagnoses as well as offer useful information for the creation of new, more efficient treatments.

# Motivation for project

The scope of a project on detecting Parkinson's disease using machine learning typically involves using data from various sources, such as speech and movement patterns, to train a model that can accurately diagnose the disease.

The motivation for this project is to develop a non-invasive and efficient method for early diagnosis of PD, which can lead to better patient outcomes and a better understanding of the disease.

# Problem Definition and Scenarios

PD is the most frequent kind of parkinsonism, often known as idiopathic parkinsonism because it has no known aetiology. Because of the accumulation of the misfolded protein the protein alpha- in the brain and its spread throughout the brain, PD is categorized as a synucleinopathy and more specifically as an alpha-synucleinopathy (synucleinopathy). Similar movement symptoms may be present in several Parkinson-plus syndromes, but there may be a variety of additional symptoms. Many of them also have synucleinopathies. In Lewy body dementia, motor symptoms appear before cognitive decline and hallucinations. Instead, multiple system atrophy (MSA) can have a Parkinsonian, cerebellar, or autonomic predominance and has an early onset of autonomic dysfunction (like orthostasis).

Several Parkinson-plus illnesses involve tau rather than alpha-synuclein. Corticobasal syndrome (CBS) and progressive supranuclear paralysis (PSP) are two instances. PSP has been connected to frontotemporal dementia symptoms and is characterized by stiffness, early falls, bulbar symptoms, and vertical gaze impairment. The signs of CBS include dystonia, alien limb, myoclonic jerks, and asymmetric parkinsonism. These onset times and associated symptoms can aid in separating some movement abnormalities from indeterminate Parkinson disease.

* 1. **Organization of the report**

The overall report revolves around the objective of Machine Learning Algorithms for Diagnosis of Parkinson's Disease Based on Voice Characteristics.

First chapter deals with introduction of Machine Learning Algorithms for Diagnosis of Parkinson's Disease Based on Voice Characteristics. In that we have included overview, motivation, objectives, and scope. Second chapter deals with literature review. In that we include details of every literature survey of we collected. Third chapter deals with project description. In this we can see the objective of the project work existing system with disadvantages, proposed system, and its advantages. Fourth chapter deals with system design. This includes system architecture. Fifth module deals with project requirements. In this chapter we will see the Hardware and Software Specification and technologies which are used. Sixth module deals with module description. In this chapter we will see in detail about each module and its functions. Seventh chapter deals with implementation. In this chapter we will implement the project working process. Eighth chapter deals with the results obtained during the implementation process of the project. Nineth chapter deals with conclusion and future work of the project. Section 10 presents the individual team members report.

* 1. **Summary**

Based on voice features, a proposed system for diagnosing Parkinson's disease (PD) uses machine learning techniques to identify the condition, including Logistic Regression, SVM, KNN, and ensemble approaches. The accuracy, recall, precision, and F1-score will be used to assess the machine learning model's performance. The initiative has the potential to expedite and enhance Parkinson's disease diagnosis and offer useful information for the creation of fresh, more efficient treatments.

# CHAPTER 2

# LITERATURE REVIEW

## Introduction

## The brain controls the vibration of the vocal folds in the larynx, which produces the voice. Degeneration of dopamine-producing neurons in the brain can cause changes in the voice in Parkinson's disease, such as decreased loudness, pitch fluctuation, and articulatory accuracy. Acoustic analysis, which includes extracting elements from the speech signal such as pitch, jitter, shimmer, and formants, can be used to assess these changes objectively. Machine learning algorithms may then be trained on these features to categorize people as having Parkinson's disease or being healthy.

## Numerous research has looked into using machine learning algorithms to diagnose Parkinson's disease based on speech characteristics. Tsanas et al. (2012), for example, employed a support vector machine (SVM) classifier to differentiate between PD and healthy persons using 16 variables retrieved from sustained vowel phonation. Using a leave-one-out cross-validation strategy, they attained an accuracy of 97.6%. Little et al. (2016) employed a logistic regression model to predict Parkinson's disease status based on 27 variables derived from a range of speech activities, including sustained vowels, reading, and monologues. Using a 10-fold cross-validation technique, they attained a 90% accuracy. Numerous studies have looked into how to diagnose Parkinson's disease using speech characteristics. Tsanas and additional for instance, to identify between individuals with Parkinson's disease and healthy controls, a study published in 2012 employed a support vector machine (SVM) classifier and 16 characteristics extracted from sustained vowel phonation. By using a leave-one-out cross-validation technique, they were able to attain an accuracy of 97.6%. Little & co. (2016) used a computed relapse model to predict Parkinson's infection status based on 27 parameters obtained from a variety of discourse activities, such as supported vowels, reading, and conversations. By using a 10-fold cross-validation approach, they were able to attain an accuracy of 90%.

## Literature Review

Andrea Sabo et al. [1] advised evaluating the concomitant validity of two gait measurement techniques in Parkinson's disease patients. A Zeno instrumented walkway, which analyses gait characteristics such stride length, stride time, and cadence, was utilised in the first method. The second technique was video-based gait analysis, which includes gathering gait features from video recordings of people walking. The study discovered a significant connection between the video-based gait analysis and the gait parameters recorded using the Zeno instrumented walkway, indicating that both techniques are viable and reliable ways to quantify gait in people with Parkinson's disease. The study's findings have significant significance for assessing gait in people with Parkinson's disease because both approaches can yield useful data for the condition's diagnosis and treatment.

Rui Guo et al. [2] manifest offered research that suggests a novel machine learning-based technique for evaluating gait in Parkinson's disease patients. The technique makes use of a deep learning system known as a multi-scale sparse graph convolutional network (MS-SGCN), which can evaluate gait data at various sizes and extract key properties. According to the study, the MS-SGCN performed better than conventional machine learning algorithms in terms of determining Parkinsonian gait. Additionally, the MS-SGCN was able to spot modest gait abnormalities that are a sign of Parkinson's disease development, which is helpful for tracking the condition and informing clinical judgement. The study's results suggest that the MS-SGCN has the potential to be a useful tool for assessing gait in people with Parkinson's disease since it can give more precise and in-depth information about these people's gait patterns.

Robbin Romijnders et al. [3] proposed a survey that examined elderly folks and those with Parkinson's disease in order to test the precision of gait event detection utilising inertial measurement unit (IMU) sensors. IMUs are tiny sensors that may be fastened to the body and used to monitor acceleration and movement. In both older persons and those with Parkinson's disease, the study demonstrated that IMU-based gait event recognition was reliable for identifying gait events including heel strikes and toe-offs during curved walking and turning. The study also discovered that Parkinson's disease-related alterations in gait patterns, such shorter steps and more step variability, could be precisely detected by the IMU sensors.

Kimberley-Dale Ng et al. [4] recommended using computer vision techniques to evaluate older persons with dementia's mobility and fall risk. The study discovered that utilising computer vision techniques, gait metrics including speed, step length, and stride duration could be precisely quantified. The study also discovered that these gait characteristics were related to fall risk in dementia-affected older persons, with those who had slower gait speeds, shorter step lengths, and longer stride times being more susceptible to falling. The results of the study suggest that methods based on computer vision can be effective for assessing mobility and fall risk in elderly dementia patients. Computer vision algorithms can provide objective and non-invasive gait measures to track changes in mobility over time and identify people who are more likely to fall.

T.Sathiya et al. [5] provided a dataset of Parkinson's disease patients and healthy controls, using the information to train the Random Forest Classifier algorithm. Based on the patient's symptoms, the algorithm was able to diagnose Parkinson's disease with excellent accuracy and sensitivity. The study's findings point to the Random Forest Classifier algorithm as a potential diagnostic tool for Parkinson's disease. Parkinson's disease may be quickly and painlessly diagnosed using machine learning approaches, which can enhance patient outcomes and lessen the strain on healthcare systems. The study also emphasises the opportunity for machine learning methods to be applied in the creation of breakthrough diagnostic tools for various illnesses.

Shallu Sehgal et al. [6] explained the fundamental ideas and procedures of optimal grass hopper algorithms, as well as examples of their uses in diverse industries. analysing the outcomes and constraints of earlier research that employed enhanced grass hopper algorithms for the diagnosis of Parkinson's disease. evaluating the performance of the improved grass hopper algorithm in comparison to other techniques, such as conventional clinical evaluations and other machine learning algorithms, for the diagnosis of Parkinson's disease. The accuracy, sensitivity, and non-invasiveness of the improved grass hopper algorithm were discussed, along with some of its weaknesses and restrictions on how effectively it may be applied to other illnesses or populations.

Armando de Jesús Plasencia Salgueiro et al. [7] formulated using a smartphone app to track people with Parkinson's disease's gait and gauge their adherence to treatment is a possible study subject. To assess the gait data and categorise the user's state, the app employs deep reinforcement learning algorithms. During the smartphone-based gait assessment, data on the subject's walking gait is gathered using the smartphone's accelerometer and gyroscope. The state of the individual is then determined by analyzing this data using deep reinforcement learning algorithms. In order to train the model, the algorithms employ a reward system in which successful results (like a steady gait) are rewarded and unsuccessful outcomes (like a shuffling gait) are penalized. As a result, the algorithm can grow and develop over time.

Chih-Chien Tsai et al. [8] using magnetic resonance imaging (MRI), the non-invasive imaging method known as diffusion tensor imaging (DTI) measures the diffusion of water molecules in the brain. White matter pathways in the brain, which are crucial for motor function, may be examined using DTI to learn more about their organization and structural integrity. Machine learning is a subset of artificial intelligence that makes predictions or categorizes data using algorithms. Machine learning algorithms are trained on DTI data from people with different forms of Parkinsonism in the case of DTI for the differential diagnosis of Parkinsonism in order to uncover patterns or characteristics that discriminate between the various illnesses. Several machines learning methods, including SVM and artificial neural networks, can be used to interpret the DTI data. The machine learning algorithms may identify fresh data from individuals with unknown diagnoses after being trained on a collection of labelled data (i.e., data with known diagnoses).

Ashena Gorgan Mohammadi et al. [9] an artificial neural network that can learn to encode and decode data is called an autoencoder. Autoencoders are trained on the vocal traits of people with Parkinson's disease and healthy people in the event of Parkinson's disease diagnosis in order to find patterns or qualities that set the two groups apart. Pitch, amplitude, and frequency modulation are some of the numerous approaches that may be used to study vocal features. These vocal traits may be represented by autoencoders as a group of latent variables, which can subsequently be utilized to provide a diagnosis.

Vikas Mittal et al. [10] devised that the fundamental frequency (pitch), jitter, shimmer, harmonics-to-noise ratio (HNR), and formant frequencies have all been used to categories Parkinson's illness. Signal processing methods are utilized to extract these properties from voice signals, which are then employed as input features by machine learning algorithms. With the aid of machine learning techniques like Weight the K-NN algorithm (Nearest Neighbour, W K-NN), Logistic Regression, and Medium Gaussian Kernel Support Vector Machines (MGSVM), Parkinson's disease has been categorized using audio data. These algorithms are used to categories new, unknown speech signals as either Parkinson's disease or healthy after being trained on a labelled dataset of speech signals from Parkinson's disease patients and healthy persons.

**2.3 Summary**

Identifying Parkinson's disease applying machine learning methods including Logistic Regression, SVM, KNN, and ensemble approaches, a suggested system based on voice characteristics can identify PD. Using methods like Logistic Regression, SVM, KNN, and ensemble approaches, the model will combine Gradient Boosting Classifier and Random Forest Classifier after being trained on a pertinent dataset that comprises data on changes in speaking patterns. To evaluate the effectiveness of the machine learning model, accuracy, precision, recall, and F1-score measures will be employed.

# CHAPTER 3

# PROJECT DESCRIPTION

## 3.1 Introduction

After being trained on a pertinent dataset containing data on changes in speaking patterns, the model will merge Gradient Boosting Classifier and Random Forest Classifier using methods like Logistic Regression, SVM, KNN, and ensemble approaches. Accuracy, precision, recall, and F1-score will be used to assess the machine learning model's performance. The project has the potential to speed up and improve the diagnosis of Parkinson's disease and provide helpful data for the development of new, more effective treatments. The goal of this research is to use machine learning to reliably identify Parkinson's disease by training a model utilizing data from numerous sources, such as speech and movement patterns.

# 3.2 Objective

# The objective of using machine learning algorithms like Support Vector Machine (SVM), Logistic Regression, and Naive Bayes for detecting Parkinson's disease is to build a model that can accurately diagnose the disease based on various vocal features.

# Training the machine learning algorithms using the selected features and a labeled dataset of patients diagnosed with Parkinson's disease and healthy individuals.

# Assessing the effectiveness of the trained models by the use of metrics like F1 score, recall, accuracy, and precision.

# To integrate the model according to few selected features which we are used to differentiate and identifying the disease.

# 3.3 Existing System

The Parkinson's Foundation's Parkinson's Voice Initiative is a project that uses machine learning algorithms to analyze voice recordings and identify Parkinson's disease. Patients can record their voice using the system's Speak Out app and send the recordings to researchers for examination. The technique has a high degree of accuracy in identifying Parkinson's disease.

Alzheimer's disease in the study paper Diagnosis using Voice, a machine learning approach for Parkinson's disease diagnosis based on voice traits is suggested. The algorithm employs a collection of speech recordings from Parkinson's sufferers and healthy controls, classifying the recordings using characteristics including pitch, jitter, and shimmer. The technique proved successful in identifying Parkinson's disease with high accuracy.

Detection of Parkinson's disease Support Vector Machine with Mel Frequency Cepstral Coefficients: Another study suggests a machine learning technique for Parkinson's disease diagnosis based on speech features. Mel Frequency Cepstral Coefficients (MFCC) feature extraction is used by the system to extract features from voice recordings, and a Support Vector Machine (SVM) classifier is used to categorize the recordings as Parkinson's or healthy. The technique proved successful in identifying Parkinson's disease with high accuracy.

# 3.4 Shortcomings of Existing System

For machine learning algorithms to learn and create reliable predictions, a lot of high-quality data is needed. The algorithm's accuracy may be impacted by the amount of data provided in the instance of Parkinson's disease diagnosis based on voice features. Individuals with PD may have a wide range of symptoms, which may also have an impact on their voice. When faced with such variation, machine learning algorithms may find it difficult to detect the condition effectively. Age, gender, and the existence of other medical disorders are a few more variables that may have an impact on voice characteristics. To produce reliable predictions, machine learning algorithms may need to take these elements into consideration.

# 3.5 Proposed System

# Logistic Regression, SVM, KNN are supervised learning algorithms and ensemble approaches such as Gradient Boosting Classifier and Random Forest Classifier are used.

# A group of machine learning approaches known as ensemble methods combine several models to provide predictions. Specifically, we used two different algorithms, SVM and Random Forest Classifier, and combined them using the voting classifier.

# Two pre-processing techniques are employed, they are standard scalar and principal component analysis (PCA), to pre-process the data prior to training our machine learning model.

# The data will be divided into training and testing sets, and to perform cross-validation to avoid overfitting.

# To build this model, a labeled dataset of patients with and without Parkinson's disease is used. It will train the algorithms on this dataset, and compare their performance in terms of metrics such as accuracy, recall, precision, F1-score, etc.

# 3.6 Benefits of Proposed System

In advance of the appearance of physical symptoms, machine learning algorithms can identify early Parkinson's disease indicators. Better treatment outcomes and earlier action may result from this early detection.

Based on vocal features, Parkinson's disease can be diagnosed non-invasively and painlessly, which may be more tolerable for patients than other diagnostic techniques. Patients in isolated or underserved places may find machine learning algorithms to be more readily available if they are employed with portable devices.

Algorithms for machine learning can be modified to meet the demands of each patient individually and can be changed in response to fresh information.

# CHAPTER 4

# SYSTEM DESIGN

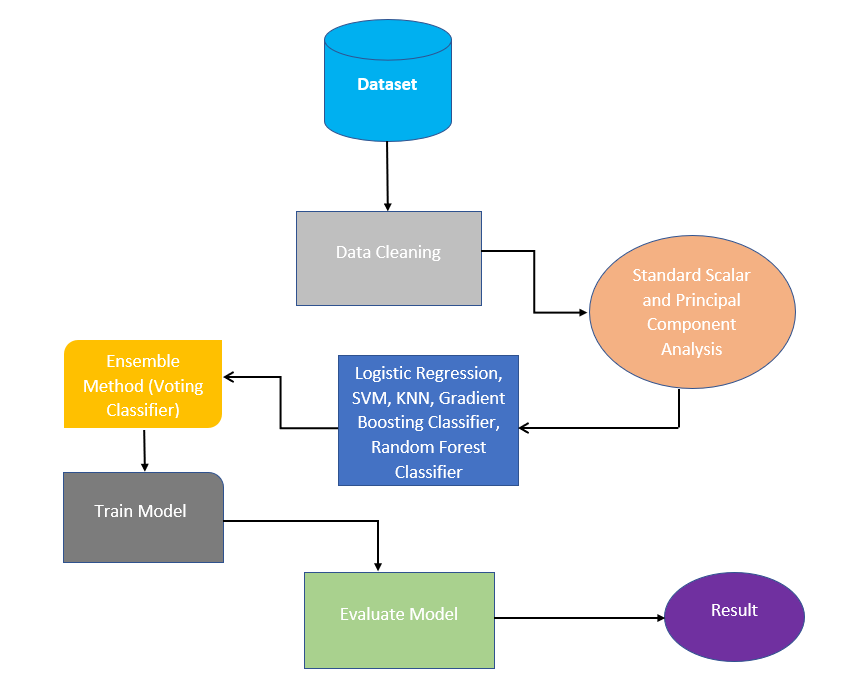
**4.1 Introduction**

The system's components and linkages are shown in high-level detail on the architectural diagram. It would show the various levels and elements of the system, including Ensemble approaches, Machine Learning algorithms, and Pre-processing techniques. Prior to training our machine learning model, the data is pre-processed using two techniques: conventional scalar and PCA. Application of machine learning techniques such as Logistic Regression, SVM, and KNN to the diagnosis of Parkinson's disease. The Gradient Boosting Classifier and Random Forest Classifier will be combined in the model using techniques like Logistic Regression, SVM, KNN, and ensemble approaches.

To evaluate the effectiveness of our machine learning model, we used grid search and cross-validation using 5-fold validation. Cross-validation is the process of separating the data into various folds, training the model on one-fold, and then evaluating it on the other folds. Grid search is a technique that thoroughly examines all possible combinations of a set of hyperparameters to determine which one produces the best results.

**4.2 Dataset description**

The needed dataset for the proposed system is gathered from multiple sources and databases. The dataset that is being gathered is in.csv file format. Parkinson's disease patients provided the voice feature dataset that was used in this research. The dataset was collected from a Kaggle platform. The dataset includes 756 instances with 23 features such as shimmer, jitter, and harmonics-to-noise ratio (HNR). The dataset was preprocessed using standard scalar normalization and principal component analysis. There were no missing values or outliers present in the dataset.



**Figure 4.1: Architecture Diagram**

**4.3 Normalization**

As shown in the above figure 4.1, employed two pre-processing techniques, standard scalar and principal component analysis (PCA), to pre-process the data prior to training our machine learning model. The method known as "standard scalar" entails scaling the characteristics to have a mean of zero and a variance of one. This keeps the scales of all the features uniform and prevents the model from being dominated by features with higher values. Additionally, utilized PCA to make the data's dimensions smaller. PCA is a technique that involves transforming the original features into a new set of features that are linearly uncorrelated and capture the most important variation in the data. This is done to reduce the number of features and to remove any redundant or irrelevant features that may negatively affect the performance of the model.

Specifically, applied standard scalar to each feature of the dataset to scale the values and ensure that all features are on a similar scale. Then applied PCA to the scaled data to reduce the dimensionality of the data. By reducing the number of features, we were able to simplify the model and improve its performance. Together, standard scalar and PCA allowed to pre-process the data in a way that improved the performance of our machine learning model. These techniques are commonly used in machine learning to normalize the data and reduce its dimensionality, which can lead to better performance and more efficient computation.

**4.4 Machine learning algorithms**

Binary classification challenges are handled by the statistical machine learning method known as logistic regression. It operates by simulating the likelihood of a binary result given any number of input variables. The key components of logistic regression include the input variables, output probability, and parameters which are learned during training. The algorithm is particularly useful when the output variable is binary and the input variables are continuous or categorical. In our study, logistic regression was used to predict the presence or absence of Parkinson's disease based on vocal feature data. We used L1 regularization to select important features and tuned the hyperparameters using cross-validation.

A machine learning approach called K-Nearest Neighbours (KNN) is employed for grouping jobs. It works by finding the KNN of a new input sample and assigning it to the class that the majority of the K neighbors belong to. The input specimens, output classes, K value, and proximity metric used to determine sample similarity are the main elements of the KNN algorithm. The algorithm is particularly useful when the decision boundary is nonlinear or when the dataset has a small number of features. In the study, KNN was used to predict the presence or absence of Parkinson's disease based on vocal feature data. We used 5-fold cross-validation to select the optimal K value and used Euclidean distance as the distance metric.

The Support Vector Machine (SVM) algorithm is a machine learning algorithm used for classification tasks. It works by finding the optimal hyperplane that separates the input data into two classes with maximum margin. The key components of the SVM algorithm include the input samples, output classes, kernel function used to map the input data into a high-dimensional feature space, and the hyperparameters used to tune the model. The algorithm is particularly useful when the dataset has a clear margin of separation and when the input data is nonlinearly separable. In the study, SVM was used to predict the presence or absence of Parkinson's disease based on vocal feature data. We used a Radial Basis Function (RBF) kernel and tuned the hyperparameters using grid search.

The Gradient Boosting Classifier algorithm is a popular machine learning algorithm used for classification tasks. It works by iteratively adding new weak learners to the model to improve its accuracy over time. The algorithm uses a loss function to measure the error of each prediction and determines the weights assigned to each weak learner based on their performance. The key components of the Gradient Boosting Classifier algorithm include the weak learner, loss function, and the boosting process used to improve the model's accuracy. The algorithm is particularly useful when dealing with complex, nonlinear relationships between input variables. In our study, Gradient Boosting Classifier was used to predict the presence or absence of Parkinson's disease based on vocal feature data. We used a combination of hyperparameter tuning and early stopping to prevent overfitting of the model.

A well-liked machine learning algorithm for classification tasks is Random Forest. It functions by building numerous decision trees from different randomly selected portions of the data and combining the outcomes to arrive at a final forecast. The algorithm uses bagging and random feature selection to improve the model's accuracy and reduce overfitting. The key components of the Random Forest algorithm include the decision tree, bagging, and random feature selection. The algorithm is particularly useful when dealing with complex, nonlinear relationships between input variables. In our study, Random Forest was used to predict the presence or absence of Parkinson's disease based on vocal feature data. We used a combination of hyperparameter tuning and cross-validation to prevent overfitting of the model.

**4.5 Ensemble method**

In this paper, explored ensemble methods to improve the performance of our machine learning model. Ensemble methods are a class of machine learning techniques that involve combining multiple models to make predictions. The premise behind ensemble approaches is that by merging multiple models, we may increase the generalizability of the model and enhance its overall performance by lowering the likelihood of overfitting. Specifically, we used two different algorithms, SVM and Random Forest Classifier, and combined them using the voting classifier. The voting classifier is a simple ensemble method that aggregates the predictions of multiple models by taking the majority vote. In our case, we combined the predictions of the SVM and Random Forest Classifier to make the final prediction. By combining these two algorithms using the voting classifier, we were able to improve the overall performance of our model. The ensemble method allowed us to reduce the risk of overfitting and increase the generalizability of the model, leading to better performance on unseen data.

**4.6 Evaluation**

Used grid search and cross-validation with 5-fold validation in this study to evaluate the effectiveness of our machine learning model. The process of partitioning the data into different folds, training the model on one-fold, then testing it on the other folds is known as cross-validation. To measure the generalization performance of the model, this procedure is performed for each fold, and the performance indicators are averaged across all folds. In addition, we used grid search to fine-tune the hyperparameters of our model. Grid search is a method that exhaustively searches through a predefined set of hyperparameters to find the optimal combination that yields the best performance. We specified a range of values for each hyperparameter of our model, and grid search evaluated the performance of the model for all possible combinations of hyperparameters within the specified range.

The cross-validation and grid search worked together to evaluate our model's performance on various subsets of data and to pinpoint the most effective hyperparameters. With the help of this method, we were able to accurately measure the effectiveness of the model and enhance it for the needs of our particular dataset.

**4.7 Evaluation Metrics**

The machine learning model's performance is assessed using a variety of criteria. We may assess the model's precision, recall, accuracy, and overall performance using these criteria. Additionally, used receiver operating characteristic (ROC) curves and confusion matrices, also known as error matrices, to illustrate the classification performance of the suggested machine learning methods. A confusion matrix makes it simple to spot misclassification trends because each row (or column) represents expected outcomes and each column (or row) represents real classes. Plotting sensitivity against (1 - specificity) at different threshold values results in the creation of a ROC curve. Additionally, the model's efficacy was assessed using the area under the ROC curve (AUC). Better categorization outcomes are indicated by higher AUC values. Using the values in the confusion matrix, we can calculate metrics such as sensitivity (TP / (TP + FN)) and specificity (TN / (TN + FP)) to further evaluate the performance of the model. It provides a clear and concise summary of the model's predictions and can be used to calculate various metrics to assess the model's accuracy.

**4.8 Summary**

In summary, the architecture design figure 4.1 would provide a high-level overview of the several levels and elements involved in the Parkinson's Disease Diagnosis Based on Voice Characteristics. The effectiveness of our machine learning model is measured using a variety of evaluation measures. These measures gave us a way to gauge the model's overall performance as well as its precision, recall, and accuracy. Additionally, employed receiver operating characteristic (ROC) curves and confusion matrices, also known as error matrices, to show how well the suggested machine learning techniques performed at classifying data.

# CHAPTER 5

# PROJECT REQUIREMENTS

# 5.1 HARDWARE AND SOFTWARE SPECIFICATIONS

# Hardware Requirements:

# 1.19Hz processor speed

# RAM 512 MB

# HDD/SSD:50GB

# Graphic Card 4gb

# Software Requirements

# Operating System: Windows >7 or Mac or Linux

# Dependencies: Python 3.6

# IDE: Google Colab or Spyder or Jupyter Notebook

# 5.2 Technologies Used

# Scikit-learn: Sklearn is a well-known open-source machine learning package for Python. For tasks including classification, regression, clustering, and dimensionality reduction, a wide range of machine learning tools and techniques are available.

# Numpy: A Python package called NumPy is used for data analysis and scientific computing. It provides a powerful N-dimensional array object in addition to tools for working with and using arrays, including discrete Fourier transforms, elementary linear algebra, elementary statistic processes, sorting, choosing, I/O, and many other activities.

# Pandas: Pandas is a Python library for analyzing and manipulating data. On top of the NumPy library, it includes data analysis capabilities as well as high-performance, user-friendly data structures.

# Matplotlib: A Python package used for data visualization is called Matplotlib. It offers a comprehensive selection of top-notch 2D and 3D charting tools and functions that let you make a variety of plots, charts, and graphs to visualize your data.

**Seaborn**: On top of Matplotlib, the Python package Seaborn adds more capabilities for data visualization. It offers a more complex interface for producing statistical visualizations that are both more appealing and educational than those made solely using Matplotlib.

# CHAPTER 6

# MODULE DESCRIPTION

# 6.1 Introduction

# The phrase “modular description of a Parkinson's Disease Diagnosis Based on Voice Characteristics" refers to an entire process for creating a system that utilizes the potential of machine learning algorithms. The system will train and test the data set that contains patient data for people with Parkinson's disease using machine learning algorithms. Overall, a strives to offer a strong and adaptable option for medical treatment that seeks to identify Parkinson's disease.

In this project we are proposing six modules:

# Dataset Description

# Data Cleaning

# Machine learning algorithms

# Ensemble method

# Evaluation Methods

# Evaluation Metrics

# 6.2 Modules

# Dataset Description: Patients with Parkinson's disease donated the voice feature dataset that was utilized in this investigation. The Kaggle platform is where the dataset was gathered. The dataset has 756 cases and 23 characteristics, including shimmer, jitter, and harmonics-to-noise ratio (HNR).

# Data Cleaning: Two pre-processing techniques are employed, they are standard scalar and principal component analysis (PCA), to pre-process the data prior to training our machine learning model. Standard scalar is a commonly used technique that involves scaling the features to have zero mean and unit variance.

# To make the data less dimensional, PCA is utilized. As part of the PCA technique, the original features are converted into an additional set of characteristics that are linearly independent and capture the key variables in the data.

# Machine learning algorithms: This The use of machine learning methods such as Random Forest Classifier, Gradient Boosting Classifier, KNN, SVM, and Logistic Regression.

# The best tools for preparing classification and regression concerns are logistic regression, KNN, and SVM. The two most popular machine learning algorithms are Random Forest and Gradient Boosting.

# Both models are ensembles, which means they combine a number of weak learners to produce a powerful learner.

# While using the same weak learner, Random Forest and Gradient Boosting are very distinct algorithms.

# Ensemble method: A group of machine learning approaches known as ensemble methods combine several models to provide predictions. Specifically, used two different algorithms, SVM and Random Forest Classifier, and combined them using the voting classifier.

# The voting classifier is a simple ensemble method that aggregates the predictions of multiple models by taking the majority vote.

# In our case, we combined the predictions of the SVM and Random Forest Classifier to make the final prediction.

# Evaluation Methods: This Two common evaluation methods, cross-validation with 5-fold validation and grid search are applied. Together, cross-validation and grid search allows to evaluate how well our machine-learning model is performing on multiple subsets of the data and to identify the best hyperparameters for the model.

# This approach enables to obtain reliable estimates of the model's performance and to optimize its performance for our specific dataset.

# Evaluation Metrics: This Several evaluation metrics are used which allow us to measure the accuracy, recall, precision, and overall performance of the model. Additionally, we used receiver operating characteristic (ROC) curves and confusion matrices, also known as error matrices, to illustrate the classification performance of the suggested machine learning methods.

# Additionally, the area under the ROC curve was used to gauge how effective the model was (AUC). Higher AUC values are a sign of better categorization results.

# CHAPTER:7

# IMPLEMENTATION

# 7.1 Introduction

# Google Colab can be used to implement this project. The Foundation. The Colaboratory, often known as "Colab," is an outcome of Google Research. Colab excels in three areas: machine learning, teaching, and data analysis. Anybody may create and execute arbitrary Python code using the browser. This is important because it enables the training of large ML and deep learning algorithms without the requirement of sophisticated hardware or a fast internet connection. Google Colab is the ideal solution for anyone passionate about deep learning and data analysis because it supports both GPU and TPU instances and overcomes the processing constraints of local workstations. Because a Colab notebook can be viewed remotely from any computer via a browser, it is also appropriate for commercial use.

# 7.2 Creating. ipynb notebook in colab

# Use your selected browser to navigate to colab.research.google.com and log in using your Google account. Click on a new notebook to start a new runtime instance.

# 

# Figure 7.1: ipynb notebook

# You can change the notebook's name by clicking on "NEW NOTEBOOK" in the lower-right corner as shown in the above figure 7.1. The cell execution block is where you put your code. To run the cell, enter while holding down shift. It is possible to use a cell's declared variable as a global variable in other cells. The environment will print a variable's value on the final line of code if it is mentioned directly there.

# Now, in order to improve the outcomes of this project, import all the necessary packages.

# The CSV file can be imported into Google Colab in a number of different ways.

# Data loading from a local drive

# To upload a document from the local drive, enter the following codes in the cell and run it.

# from google.colab import files

# uploaded = files.upload()

# Click "choose files" to select to download the CSV file from your computer's hard drive.

# Through Github

# The simplest way to add a CSV file to Colab is with this method. To do this, navigate to the dataset in the repository on GitHub and choose "View Raw". Copy the raw dataset's URL and pass it as an argument to the read\_csv() function.

# url = 'copied\_raw\_github\_link'

# df = pd.read\_csv(url)

# Using Google Drive

# Installing the drive

# This approach is really straightforward and clear.

# Establish a folder in Google Drive.

# To this folder, upload the CSV file.

# The code below should be entered into your Colab Notebook:

from google.colab

import drive

# new\_data = pd.read\_csv("/content/drive/MyDrive/ML\_Parkinson\_son\_dataset.csv")

# print(new\_data.head())

# Use machine learning algorithms in the code to acquire precise results.

# CHAPTER: 8

# RESULT ANALYSIS

# 8.1 Introduction

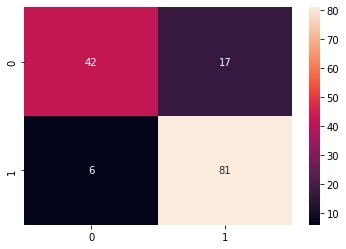
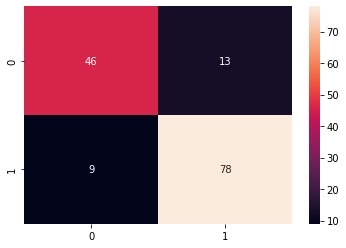
# Parkinson's illness a new approach called "Diagnosis Based on Voice Characteristics" promises to accurately identify PD in patients. This system uses Machine Learning Algorithms to analyze, train, and test the given data collection. Some pre-processing methods, such as PCA and Standard Scaler, are employed to obtain better performance. Confusion matrix will be used to calculate the values of numerous evaluation indicators. Overall, the findings and discussion section provide a complete analysis of Parkinson's Disease Diagnosis Based on Voice Characteristics and insight into its potential impact on the medical industry.

# 8.1 Results

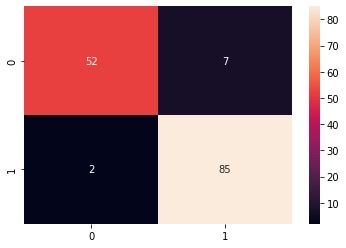
# Five machine learning algorithms and the ensemble technique were used. Their categorization outcomes for the two classification tasks in the 5-fold cross-validation procedure is displayed in Table 8.1. The SVM and voting classifier did the best and achieved a 0.93 accuracy in the two-class classification task, followed by the other classifiers that are the highest precision was obtained by Random Forest (0.89), followed by KNN (0.84), logistic regression (0.84), and gradient boosting (0.75).

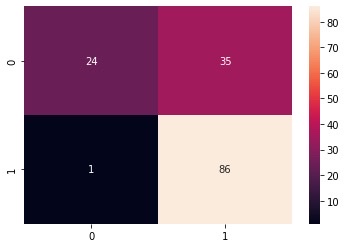
**TABLE 8.1** Classification Outcomes Using the Ensemble Technique and All Five Machine Learning Algorithms

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **Logistic Regression** | 0.84 | 0.85 | 0.84 | 0.84 |
| **KNN** | 0.84 | 0.85 | 0.84 | 0.85 |
| **SVM** | 0.94 | 0.94 | 0.94 | 0.94 |
| **Gradient Boosting Classifier** | 0.75 | 0.81 | 0.75 | 0.72 |
| **Random Forest Classifier** | 0.90 | 0.90 | 0.90 | 0.90 |
| **Voting Classifier** | 0.94 | 0.94 | 0.94 | 0.94 |

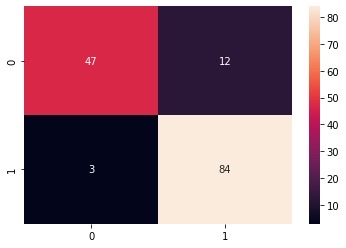
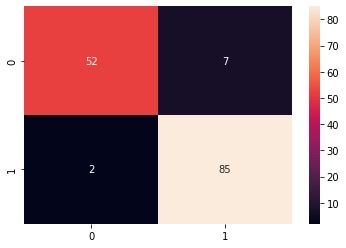
****

|  |  |
| --- | --- |
| **(a)** | **(b)** |

****

****

|  |  |
| --- | --- |
| **(c)** | **(d)** |

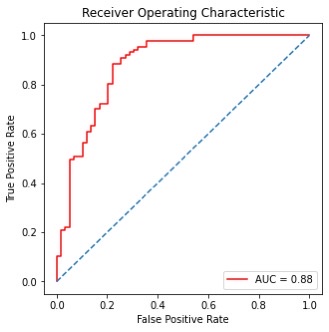
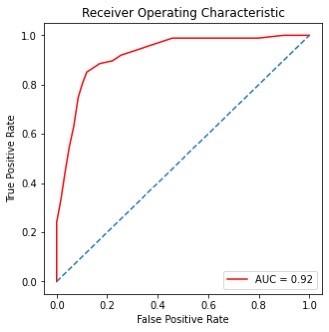
****

|  |  |
| --- | --- |
| **(e)** | **(f)** |

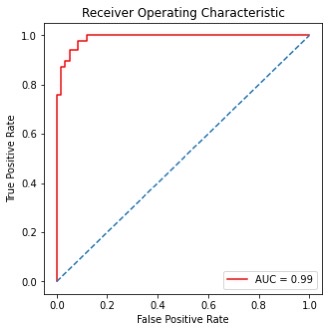
**Figure 8.1:** Confusion matrix in the two class classification assignments.

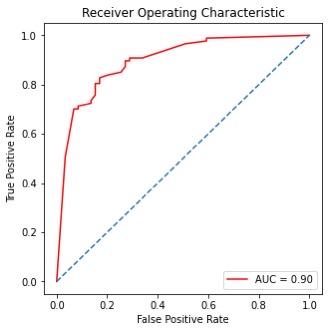
There are five machine learning algorithms and ensemble method. (a) For the two classes, logistic regression. (b) For the two classes, K - Nearest neighbor. For the two classes, support vector machine (c). (d) For the two classes, gradient boosting. (e) For the two classes, Random Forest (f) Voting classifier, and the rate of accurate classification increases as the diagonal cells' colour becomes darker as seen in the earlier figure.

A binary classification model's effectiveness is graphically depicted by the Receiver Operating Characteristic (ROC) curve. The graph shows how sensitivity and specificity are traded off for various classification criteria. A popular metric for summarising the overall efficacy of the model is the area under the curve (AUC). An AUC of 1.0 would indicate a flawless model, whereas an AUC of 0.5 would indicate a model with no value for prediction.

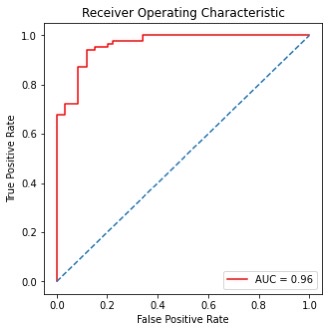
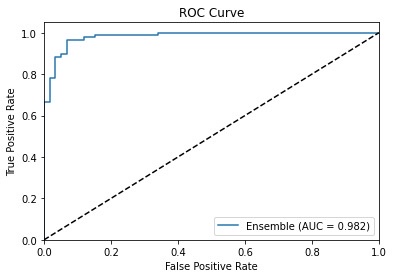
****

|  |  |
| --- | --- |
| **(a)** | **(b)** |

****

****

|  |  |
| --- | --- |
| **(c)** | **(d)** |

****

|  |  |
| --- | --- |
| **(e)** | **(f)** |

# Figure 8.2: Roc curve for the five-machine learning algorithm and ensemble method (a) Logistic regression. (b) K - Nearest Neighbour. (c) Support vector Machine. (d) Gradient Boosting and (e) Random Forest and (f) Voting classifier. The true-positive rate is shown on the Y-axis, and the false-positive rate is shown on the X-axis.

# The fig [8.2] (a)shows an ROC curve for a logistic regression model predicting the presence of Parkinson disease based on disease status. The AUC for this model is 0.88. The fig [8.2] (b)shows an ROC curve for a KNN model and AUC for this model is 0.92. The fig [8.2] (c)shows an ROC curve for a SVM model and AUC for this model is 0.99. The fig [8.2] (d)shows an ROC curve for a gradient boosting model and AUC for this model is 0.90. The fig [8.2] (e)shows an ROC curve for a random forest classifier and AUC for this model is 0.96. The fig [8.2] (f)shows an ROC curve for a voting classifier and AUC for this model is 0.98 indicating that it SVM and voting classifier performs well in distinguishing between cases and non-cases.

# CHAPTER 9

# CONCLUSION AND FUTURE WORK

# 9.1 Introduction

# The diagnosis and treatment of PD have benefited greatly from the advances made in the field of machine learning in many facets of healthcare. A neurodegenerative condition called Parkinson's disease (PD) is characterised by a steady decline in motor abilities. For effective intervention and management of Parkinson's disease, an early and precise diagnosis is essential. Researchers have recently looked into the prospect of using machine learning algorithms to analyse voice characteristics as a non-invasive and economical way to diagnose Parkinson's disease (PD). The goal of this project is to research and assess several machine learning algorithms for Parkinson's disease diagnosis based on voice data.

# A more thorough and reliable diagnosis of Parkinson's disease might be achieved by combining voice analysis with other diagnostic modalities, such as imaging methods or clinical evaluations. Real-time voice monitoring through integration with wearable technology and cellphones could provide continuous assessment and early motor function degradation identification. It would also be advantageous to investigate how interpretable machine learning models are for the diagnosis of Parkinson's disease. Understanding the specific voice patterns or variables that influence the algorithms' classification choices may offer important new perspectives on the underlying physiological changes related to Parkinson's disease.

# 9.2 Conclusion

In conclusion, our research aimed to explore the effectiveness of various machine learning algorithms in predicting Parkinson's disease using vocal features. Through our experimentation with five different algorithms, we found that SVM and Voting Classifier yielded the highest accuracy rates, with an overall accuracy of 93%, respectively. Our use of ensemble methods further improved the accuracy of our models. These findings have important implications for the early detection and diagnosis of Parkinson's disease, which can be critical for successful treatment outcomes. By using vocal features and machine learning algorithms, we have demonstrated the potential for accurate and efficient diagnosis of Parkinson's disease. However, there are limitations to our study. Our dataset was limited to vocal features, and further research is needed to determine the effectiveness of these algorithms with larger and more diverse datasets. Future research may also examine the use of additional factors to diagnose Parkinson's disease more accurately, such as data on movement or brain imaging. Overall, our research sheds important information on the possible application of algorithms based on machine learning in Parkinson's disease detection and emphasises the significance of ongoing research in this field to enhance results of diagnosis and therapy.

**9.3 Future work**

Continuous data on speech qualities and other physiological indicators can be collected by wearable technology, such as smartwatches and fitness trackers. The development of machine learning algorithms that can examine this data and find early indications of Parkinson's disease may be the main goal of future research. It may be possible to create mobile applications that evaluate voice traits using machine learning algorithms to offer a quick and easy way to diagnose Parkinson's disease. A mobile app that allows users to record their speech could diagnose them and offer treatments. Technologies for natural language processing (NLP) could be used to examine speech's content and organization in addition to its voice qualities. This may offer a more thorough evaluation of the patient's condition and increase the precision of the diagnosis.

**CHAPTER 10.**

**INDIVILDUAL TEAM MEMBER’S REPORT**

**10.1 OBJECTIEVE OF INDIVIDUAL MEMBERS**

1. **P. SAI MANOJ (19113101)**

To design a model architecture and to develop the algorithm.

1. **K. SUMANTH KUMAR REDDY (19113105)**

To maintain an updated data set samples for getting more accuracy and testing.

**10.2 ROLE OF TEAM MEMBERS**

1. **P. SAI MANOJ (19113101)**

Identification of the issue, review of the literature, creation of the module description, and work on the shortcomings of the current system.

1. **K. SUMANTH KUMAR REDDY (19113105)**

Error detection, algorithm selection that is appropriate, testing, and evaluation.

**10.3** **CONTRIBUTION OF TEAM MEMBERS**

1. **P. SAI MANOJ (19113101)**

* Exploration of academic publications in order to identify existent systems.
* Model implementation.
* The planning of Module description.
* Cooperation in project presentation, paper, and report.

1. **K. SUMANTH KUMAR REDDY (19113105)**

* Understanding and selecting algorithms.
* Model implementation.
* Paper evaluation and correction.
* Participation in project presentation, paper, and report.

**REFERENCES**

[1] Sabo, A., Gorodetsky, C., Fasano, A., Iaboni, A., & Taati, B. (2022). Concurrent Validity of Zeno Instrumented Walkway and Video-Based Gait Features in Adults With Parkinson’s Disease. IEEE Journal of Translational Engineering in Health and Medicine, 10, 1-11.

[2] Guo, R., Shao, X., Zhang, C., & Qian, X. (2021). Multi-scale sparse graph convolutional network for the assessment of Parkinsonian gait. IEEE Transactions on Multimedia, 24, 1583-1594.

[3] Romijnders, R., Warmerdam, E., Hansen, C., Welzel, J., Schmidt, G., & Maetzler, W. (2021). Validation of IMU-based gait event detection during curved walking and turning in older adults and Parkinson’s Disease patients. Journal of neuroengineering and rehabilitation, 18(1), 28.

[4] Ng, K. D., Mehdizadeh, S., Iaboni, A., Mansfield, A., Flint, A., & Taati, B. (2020). Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia. IEEE journal of translational engineering in health and medicine, 8, 1-9.

[5] Sehgal, S., Agarwal, M., Gupta, D., Sundaram, S., & Bashambu, A. (2020). Optimized grass hopper algorithm for diagnosis of Parkinson’s disease. SN Applied Sciences, 2, 1-18.

[6] Salgueiro, A. D. J. P., Shichkina, Y., García, A. G., & Rodríguez, L. G. (2021). Parkinson’s disease classification and medication adherence monitoring using smartphone-based gait assessment and deep reinforcement learning algorithm. Procedia Computer Science, 186, 546-554.

[7] Mohammadi, A. G., Mehralian, P., Naseri, A., & Sajedi, H. (2021). Parkinson's disease diagnosis: The effect of autoencoders on extracting features from vocal characteristics. Array, 11, 100079.

[8] Belić, M., Bobić, V., Badža, M., Šolaja, N., Đurić-Jovičić, M., & Kostić, V. S. (2019). Artificial intelligence for assisting diagnostics and assessment of Parkinson’s disease—A review. Clinical neurology and neurosurgery, 184, 105442.

[9] Gunduz, H. (2021). An efficient dimensionality reduction method using filter-based feature selection and variational autoencoders on Parkinson's disease classification. Biomedical Signal Processing and Control, 66, 102452.

[10] Mohammed, M. A., Elhoseny, M., Abdulkareem, K. H., Mostafa, S. A., & Maashi, M. S. (2021). A multi-agent feature selection and hybrid classification model for Parkinson's disease diagnosis. ACM Transactions on Multimidia Computing Communications and Applications, 17(2s), 1-22.

[11] Haq, A. U., Li, J. P., Memon, M. H., Malik, A., Ahmad, T., Ali, A., ... & Shahid, M. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson’s disease using voice recordings. IEEE access, 7, 37718-37734.

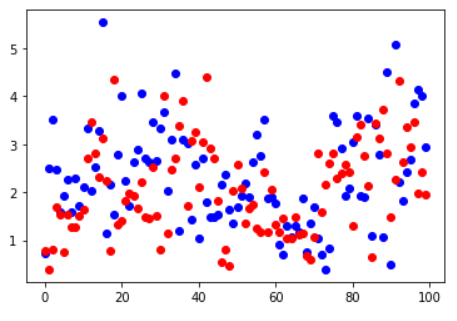
[12] Sathiya, T., Reenadevi, R., & Sathiyabhama, B. (2021). Random Forest Classifier based detection of Parkinson’s disease. Annals of the Romanian Society for Cell Biology, 2980-2987.

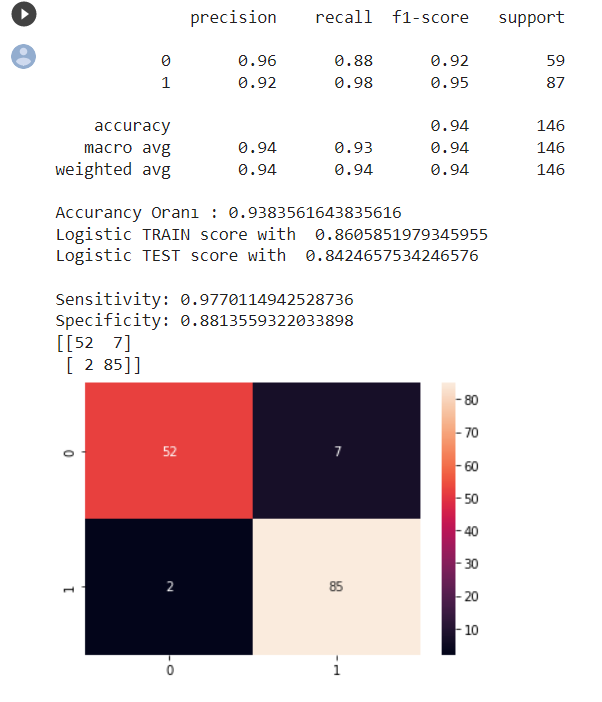
[13] Masud, M., Singh, P., Gaba, G. S., Kaur, A., Alroobaea, R., Alrashoud, M., & Alqahtani, S. A. (2021). CROWD: crow search and deep learning-based feature extractor for classification of Parkinson’s disease. ACM Transactions on Internet Technology (TOIT), 21(3), 1-18.

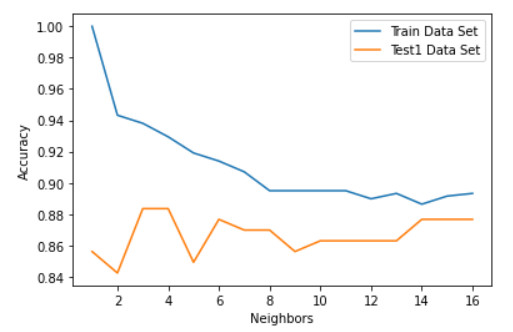
[14] Tsai, C. C., Chen, Y. L., Lu, C. S., Cheng, J. S., Weng, Y. H., Lin, S. H., ... & Wang, J. J. (2022). Diffusion tensor imaging for the differential diagnosis of Parkinsonism by machine learning. Biomedical Journal.

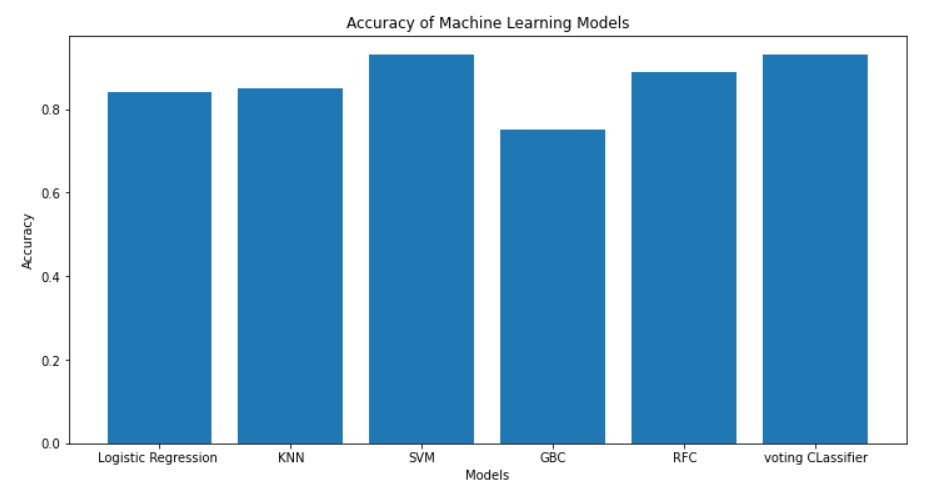
[15] Mittal, V., & Sharma, R. K. (2021). Machine learning approach for classification of Parkinson disease using acoustic features. Journal of Reliable Intelligent Environments, 7(3), 233.

**APPENDIX A: SAMPLE SCREEN**

****

****

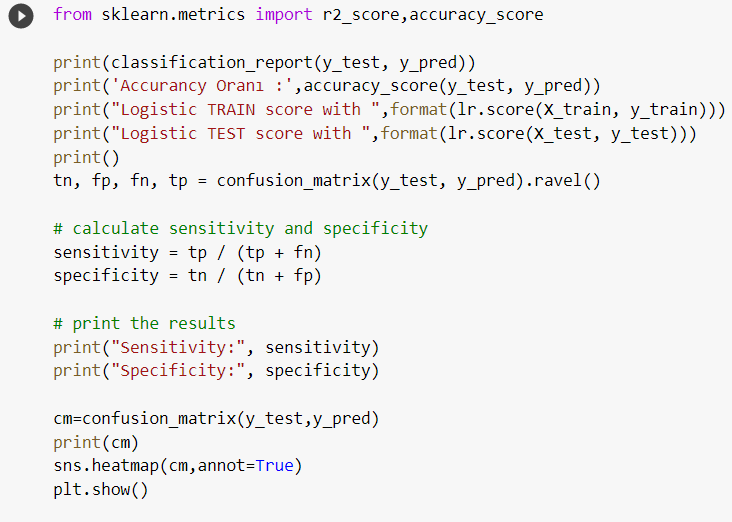
****

****

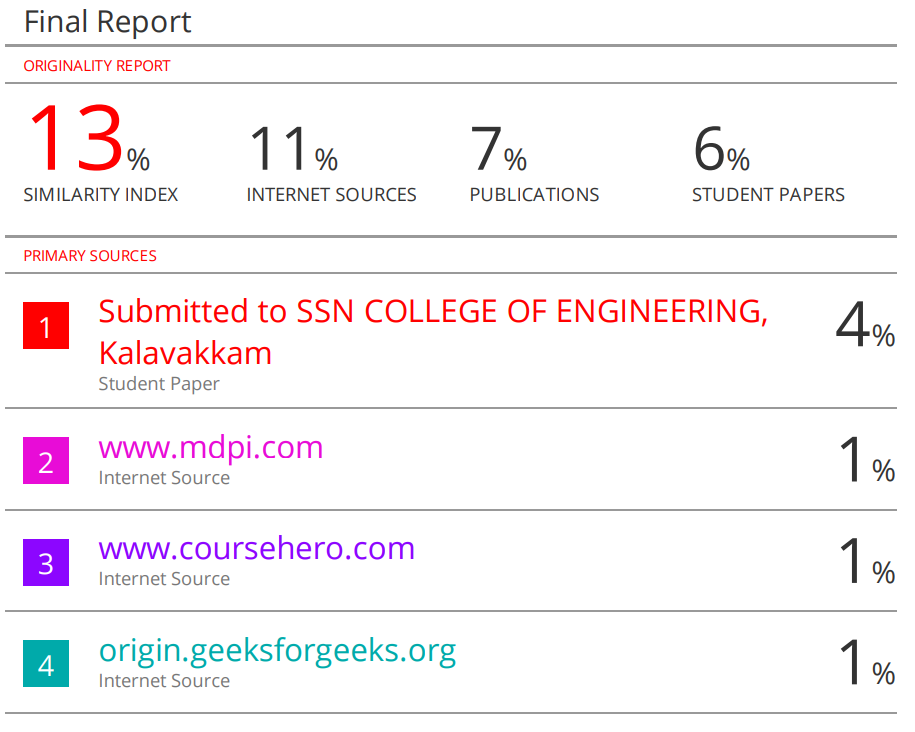
**APPENDIX B : SAMPLE CODE**

****

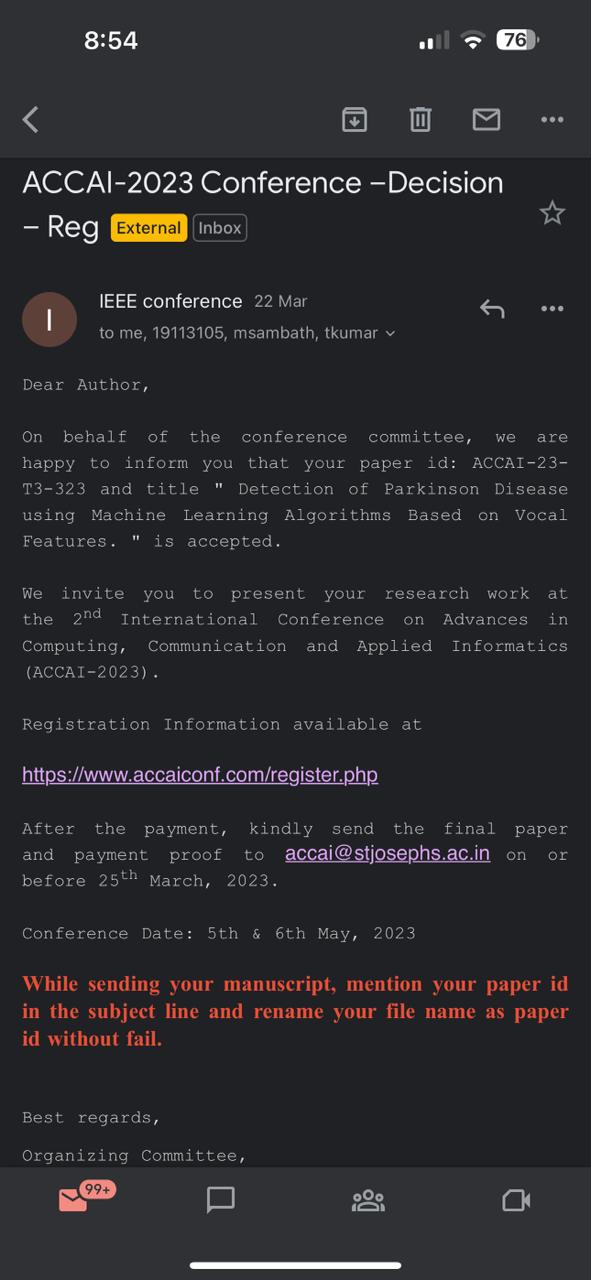
****

****

**APPENDIX C : PLAGIARISM REPORT**

****

**APPENDIX D: FUNDING / PATENT /PUBLICATION DETAILS**



**APPENDIX E : TEAM DETAILS**

**NAME:** P SAI MANOJ

**CLASS:** CSE-8B

**ROLL NUMBER:**19113101

**EMAIL:** 19113101@student.hindustanuniv.ac.in

**NAME:** K SUMANTH KUMAR REDDY

**CLASS:** CSE-8B

**ROLL NUMBER:**19113105

**EMAIL:** 19113105@student.hindustanuniv.ac.in