IDENTIFICATION OF PARKINSON'S DISEASE USING MACHINE LEARNING TECHNIQUE

Project submitted to the

SRM University – AP, Andhra Pradesh

for the partial fulfillment of the requirements to award the degree of

Bachelor of Technology

In

Computer Science and Engineering

School of Engineering and Sciences

Submitted by

Sai Dheeraj Jammu(AP21110011244)
Vishnu Vardhan Vemula(AP21110011011)
Navadeep Kalisetty(AP21110011002)
Hemanth Challa(AP21110010982)

Under the Guidance of

(M.Mallikarjuna Rao)

SRM University–AP

Neerukonda, Mangalagiri, Guntur

Andhra Pradesh – 522 240

Dec, 2024

Certificate

Date: 2-Dec-24

This is to certify that the work present in this Project entitled "**Identification of Parkinson's Disease using machine Learning Technique**" has been carried out by **Sai Dheeraj Jammu** under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

Supervisor

Dr. M Mallikarjuna Rao

Assistant Professor, SRM

University AP.

Certificate

Date: 2-Dec-24

This is to certify that the work present in this Project entitled "**Identification of Parkinson's Disease using machine Learning Technique**" has been carried out by **Vishnu Vardhan Vemula** under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

Supervisor

Dr. M Mallikarjuna Rao

Assistant Professor, SRM

University AP.

Certificate

Date: 2-Dec-24

This is to certify that the work present in this Project entitled "**Identification of Parkinson's Disease using machine Learning Technique**" has been carried out by **Navadeep kalisetty** under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

Supervisor

Dr. M Mallikarjuna Rao

Assistant Professor, SRM

University AP.

Certificate

Date: 2-Dec-24

This is to certify that the work present in this Project entitled "**Identification of Parkinson's Disease using machine Learning Technique**" has been carried out by **Hemanth Challa** under my/our supervision. The work is genuine, original, and suitable for submission to the SRM University – AP for the award of Bachelor of Technology/Master of Technology in **School of Engineering and Sciences**.

Supervisor

Dr. M Mallikarjuna Rao

Assistant Professor, SRM

University AP.

Acknowledgements

We express our heartfelt gratitude to all individuals and organizations whose support and contributions have been invaluable in the pursuit of this research endeavor aimed at leveraging machine learning techniques for Parkinson's disease.

Expressing gratitude for their collaboration, expertise, and tireless efforts in data collection, analysis, and model development, which significantly enriched the project. Prof M.Mallikarjuna Rao and SRM University AP for their expert guidance, invaluable insights into Parkinson's disease pathology, and their continuous support throughout this research. Our deepest thanks to the dedicated team members and collaborators who contributed their expertise and effort, enabling the development and validation of machine learning models for Parkinson's disease analysis. We acknowledge SRM University AP for providing the necessary infrastructure, resources, and conducive environment essential for conducting this research. We acknowledge the pioneering work of researchers in the field of Parkinson's disease, whose insights and discoveries laid the groundwork for our investigations.

Table of Contents

Abstract

Parkinson's disease is characterized by a decrease in dopamine production as a result of brain cell degeneration. Dopamine is in charge of communication between corridors. The part of the brain related with body control and ignorance of body movements. As a result, the complaint presents as a diapason of movement illnesses as well as non-motor characteristics. It is now demonstrated that non-motor symptoms can appear many times prior to the start of motor signs. as a result, timely and precise Opinion is critical in deciding whether to halt or slow the process. The complaint was stopped in its tracks. In this context, ensemble Boosting algorithms, for example, are examples of machine literacy (ML) algorithms contribute significantly to the early identification of Parkinson's disease early stage. Four boosting methods are investigated in this work. and implemented in the UCI Parkinson's Disease dataset. Following a thorough examination In terms of simulation, the ML models produced satisfactory results. of several performance factors such as delicacy, perfection,F1- Score, AUC, Youden, particularity, and mistake rate are all measures of recall.

Abbreviations

SVM	Support Vector Machine
PD	Parkinson's Disease
KNN	K – Nearest Neighbour
ML	Machine Learning
RFE	Recursive Feature Elimination
SMOTE	Synthetic Minority Over Sampling
MDVP	Multidimensional voice program analysis
NHR	Noise/harmonic ratio
HNR	Harmonic/noise
XGBoost	Extreme Gradient Boosting

List of Tables

List of Figures

1. Introduction

Parkinson’s disease is a sluggish neurodegenerative brain disease. Parkinson’s disease is a slow-progressing neurodegenerative brain illness. Dopamine-producing brain cells are found in specific areas of the human brain. The ”substantia nigra” is the name given to the region of the brain where these cells are concentrated. A substance called dopamine helps the substantia nigra and other brain areas that regulate movement in the body communicate with each other. People with dopamine are able to move in a fluid and melodic manner. The inability to produce adequate dopamine results in the motor symptoms of Parkinson’s disease (PD) when 60–80% of the cells that make dopamine are destroyed [3].

The lower brain stem, olfactory pathways, and enteric-nervous system experience the first signs of Parkinson’s disease. [1]From these areas, Parkinson’s disease (PD) extends to the substantia nigra and the brain shell, two higher brain regions. It is believed that the disease starts several years before the initial symptoms, which include constipation, sleep difficulties, loss or decrease of smell, and slowing of movement. Also, voice problems affect 90% of people with Parkinson’s disease. In order to stop the disease’s progression, The researchers are seeking for the techniques to identify the non motor signs as soon as they appear during this illness.

These days, machine learning (ML) is widely utilized to diagnose medical diseases due to its high accuracy and ease of implementation. In the literature, ML has also been used to treat Parkinson’s disease. examined the research on feature selection (FS) to be applied to machine learning (ML) in brain surgery. During Parkinson’s disease brain surgery, an Machine Learning -based technique is employed to identify the precise region to be operated on. The research conducted after PD is diagnosed are the main topic of this publication. estimated the cognitive effects of Parkinson’s disease (PD) using machine learning techniques [10]. an ML application predicted the tremor level of PD patients. ML also carried out the PD stage prediction. However, the majority of the research has concentrated on the early detection of Parkinson’s disease (PD) by making use of machine learning model, which attempts to forecast PD based on motion data collected from subjects’ upper limbs. Both PD patients and healthy individuals who were used as experimental subjects were required to wear devices in their upper limbs and complete

1.1 Literature Overview

In the Nissar et al [11]. study, the researchers used a variety of machine learning methods to try and identify Parkinson's disease. They selected relevant features for their model by employing two methods: minimum Redundancy Maximum Relevance (mRMR) and recursive feature elimination (RFE). A variety of ML- models, including logistic regression, naive bayes, K-nearest neighbors (KNN), random forest classifier, decision tree classifier, support vector machine (SVM), multilayer perceptron (MLP), and XGBoost, were used to assess the ability to identify Parkinson's disease. According to the findings, the XGBoost model that used the mRMR feature selection strategy outperformed all other cutting-edge techniques in the identification of Parkinson's disease, achieving the highest accuracy across all feature subsets.

Ashour et al. [4] introduced a new process for selecting features in two stages that meant to identify patients with Parkinson's disease (PD) who exhibit voice loss. They deployed a Support Vector Machine (SVM) classifier to compute a collection of weighted hybrid chosen features for this purpose. The suggested method relies on a type of SVM called n-cubic kernel, which showed better ability in identifying voice loss. in persons with Parkinson's disease. Polat introduced a method for identifying imbalanced class distributions in a dataset related to Parkinson's Disease (PD). They combined the use of SMOTE (Synthetic Minority Over-sampling Technique) and random forest classifiers. By building a random forest classifier with balanced data, they were able to achieve high accuracy in PD detection. Additionally, Jain and colleagues used deep neural networks along with SMOTE oversampling to address the prediction problem, and their approach yielded the best performance in detecting Parkinson's Disease [12].

Suvita [14] presented binary Rao PD classification systems that leverage Rao algorithms, reducing the requirement for parameter adjustment. They used the hyperbolic tangent V-shaped function to change continuous Rao techniques into binary form and optimized the 'k' value for KNN using binary Rao. These approaches were examined utilizing PD datasets, and their effectiveness was determined by the Friedman rank test. In the evaluation, the proposed binary Rao algorithms were also tested contrary to state-of-the-art algorithms.

F. Cavallo et al. [5] and their team conducted a study to predict Parkinson's disease by analyzing motion data from people's upper limbs. They implanted a device in both Parkinson's disease patients and healthy individuals, had them perform various activities, and collected spatiotemporal and frequency data. Afterward, they used various supervised learning techniques to classify and predict Parkinson's disease based on the collected data.

2. Methodology

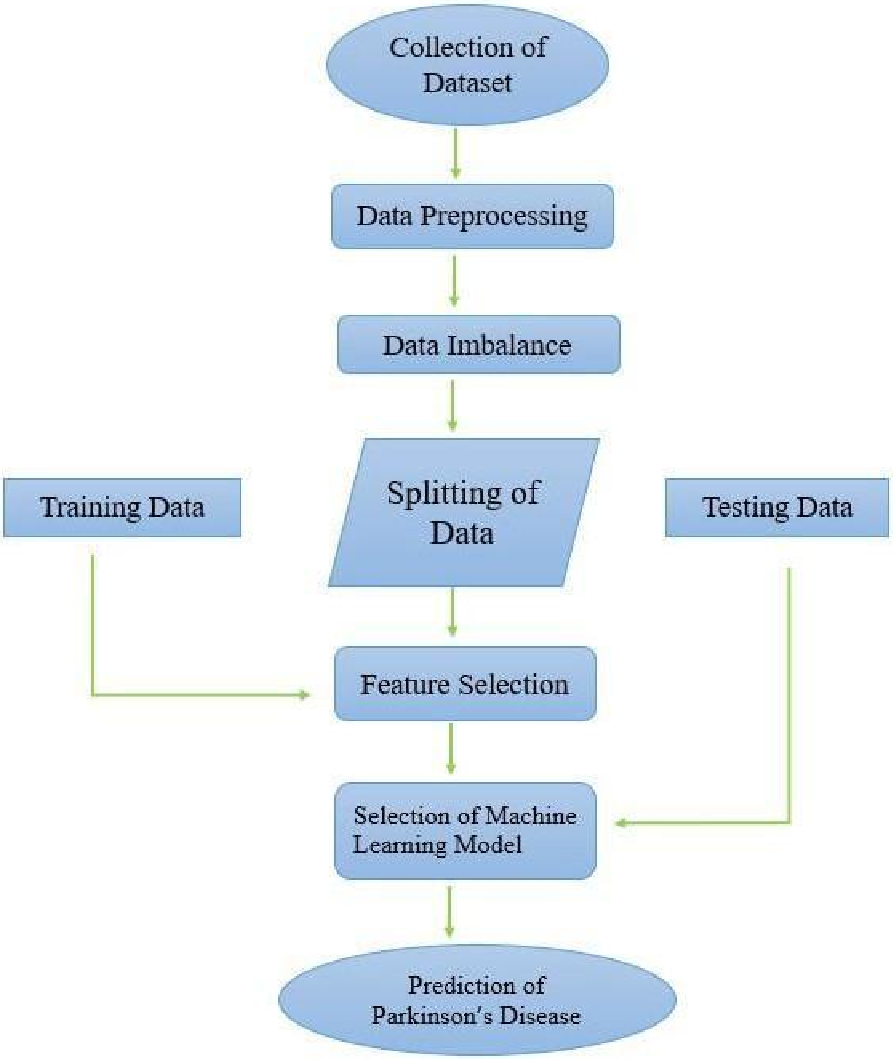


Fig 1 Proposed Methodology

2.1 Dataset Collection

The dataset which is sourced from the UCI Machine Learning Repository, comprises information from 195 individuals, encompassing 48 patients were infected with Parkinson's disease and 147 healthy individuals. The data collection involved recording the participants as they sustained the phonation of the vowel sound "/a/" using a microphone. Each row in the dataset refer to a vocal recordings, while each column refers to distinct vocal measures. The column labeled ”status” differentiates between healthy individuals (marked as 0) and individuals with Parkinson’s disease (marked as 1), indicating that the ultimate goal of the dataset is to distinguish between these two groups based on vocal characteristics

This dataset includes vocal records of each individual with Parkinson’s disease (PD) tracked during their medical treatments. Each patient’s profile comprises 24 clinical features, including several vocal basic frequencies such as MDVP: Fo, MDVP: Fhi, MDVP: Flo, and measures of frequency variation like Jitter and MDVP: RAP. Additionally, it contains amplitude variation measures such MDVP: shimmer, shimmer (dB), shimmer: APQ, and shimmer DDA. The dataset also features ratios of noise to tonal components (NHR, HNR), nonlinear dynamical complexity measures (RPDE, D2), signal fractal scaling exponent (DFA), nonlinear measures of basic frequency variation (spread 1, spread 2, PPE), and the health status of subjects (0 for healthy, 1 for Parkinson’s Disease “.

Feature Description

Sn.no	Attribute	Description
1	MDVP:Fo(Hz)	Pitch period fundamental frequency
2	MDVP:Fhi(Hz)	Upper fundamental frequency limit or maximum voice modulation threshold.
3	MDVP:Flo (Hz)	Minimum or lower limit vocal fundamental frequency
4	MDVP:Jitter(%) , MDVP:Jitter(Abs), MDVP:RAP MDVP:PPQ Jitter:	MDVP is an important measure of the frequency of vibrations in vocals that folds from pitch period to pitch mark at the start of the next cycle DDP MDVP is an important measure of the frequency Of vibrations in vocals that folds from pitch period to pitch mark at the start of the next cycle.
5	Jitter and Shimmer	The absolute difference between the frequencies of each cycle only when the average is normalized.
6	NHR , HNR	Two measures of the ratio of noise to the tonal components in the voice
7	RDPE	Recursive Period density entropy quantifies the signal until it is periodic
8	D2	Using a fractal object, the correlation dimension is utilized to determine dyshonia in speech.
9	DFA	Deternded Fluctuation Analysis –The signal’s fractal

		scaling exponent
10	spread1,spread2	Three nonlinear measures of fundamental Frequency variation.
11	PPE	On a logarithmic scale, pitch period entropy is used to measure unexpected fluctuations in speech.
12	Status	status with ’0’ indicates the healthy person where as status with ’1’ indicates the Parkinsons disease

2.2 Data Preprocessing

It is the process of processing raw data and making the data acceptable for the selected machine Learning model, It is considered as crucial stage for creating a model. A real world data may contain data noise, missing values , which is not suitable for a machine learning model. Data Preprocessing mainly comprises of data cleaning, data integration, data transformation, data reduction, data normalization. It is necessary to arrange the given data in a numerical form in order to prevent any problems at later stages In this Step we will Convert all text values into numerical form is the solution to this problem. Later Scaling is performed Scaling is a technique that can convert data values into shorter ranges. Rescaling and Standardization can be used for the process of scaling the data.

2.3 Data Imbalance

Data imbalance occurs when the distribution of classes that make up a substantial part of the data set are known as majority classes, whereas those that make up a lower proportion are known as minority classes. This imbalance can pose challenges in various fields, such as machine learning, where algorithms may struggle to accurately predict underrepresented classes. Classes that represent a significant portion of the dataset are referred to as majority classes. These classes typically have a larger number of instances or samples which are compared with various categories within dataset. On the other hand, minority classes constitute a smaller proportion of the overall dataset, often having fewer instances or samples compared to the majority classes.

In a scenario of data imbalance, models trained on such data may exhibit biased behavior, as they tend toward the majority class, which has more instances. Since the model may not have enough examples to understand patterns and generate accurate predictions for these classes, this bias can lead to poor performance in predicting minority classes.

Addressing data imbalance is crucial for improving model performance. Traditional strategies increasing the representation of the smaller group, decreasing the prevalence of the larger group. These techniques aim to create a more balanced dataset, Enabling the system to enhance its learning capabilities for the patterns associated with all classes. Here for this dataset we have followed SMOTE Technique.

2.3.1 Synthetic Minority Oversampling Technique :

SMOTE, which stands for ”Synthetic Minority Over-sampling Technique,” is an algorithm used in the field of machine learning and data mining in order to clarify or solve the class imbalance problem tasks [13]. This class imbalance usually arises when one class has much fewer instances when compared to another class of the same binary classification problem.

SMOTE algorithm works as follows:

1. Select a minority class instance (i.e., an instance from the class with less data).
2. Identifying the nearest ’k’ neighbors among the minority class instances. The k-value is a user-defined parameter.
3. One of the k-nearest neighbors is to be choosen randomly.
4. Create a new example by combining features from a chosen instance and the randomly chosen neighbor. This is done by selecting a random value between 0 and 1 and using it to weigh the contributions of the two instances to create the synthetic sample.
5. Repeat these steps upto the desired level of oversampling is reached.

The SMOTE algorithm produces a dataset where the classes are more evenly distributed, which can help machine learning models perform better on classification tasks. SMOTE is widely used in various domains, especially when dealing with imbalanced datasets “Fig. 2”.

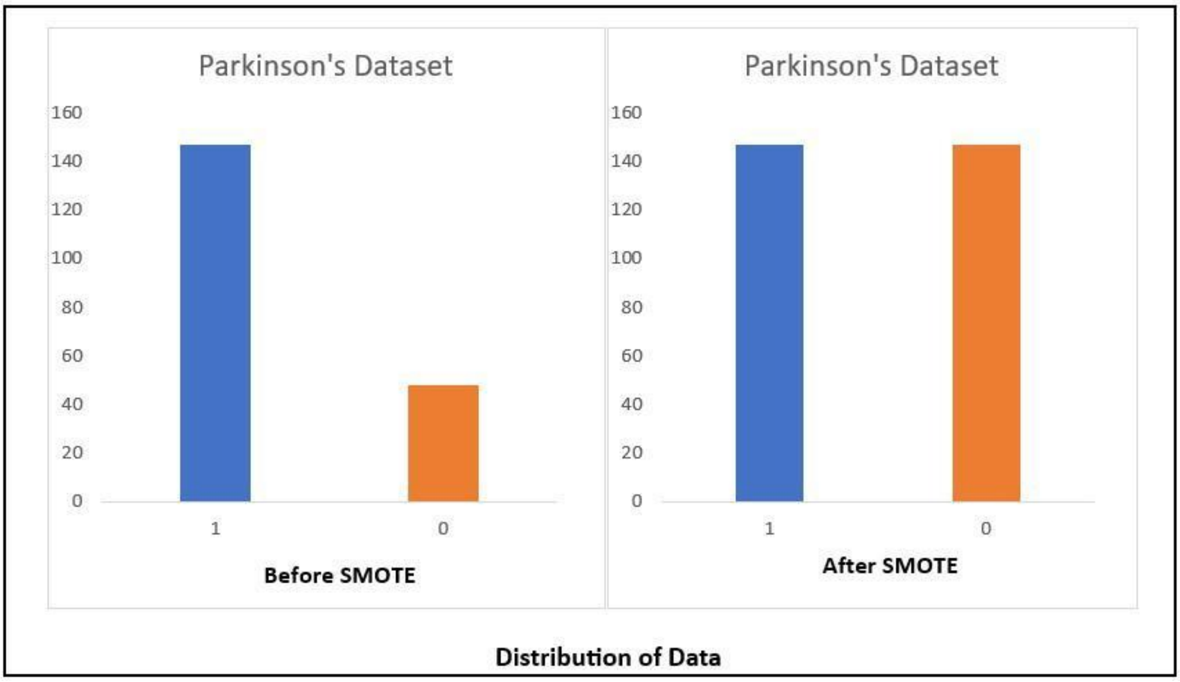


Figure 2 Distribution of Data

2.4 Feature Selection

In the process of developing a machine learning model only a handful of attributes from the dataset prove to be valuable for the model. Feature selection is the process of determining which of the model's key characteristics should be included. A feature is an element that affects or is helpful for an issue. Feature selection is a technique for lowering the model's input variable by using only relevant data in order to reduce overfitting in the model. We need Feature selection to help the model to perform well by selection important or relevant features.

Recursive feature elimination (RFE) [6] Feature selection helps decrease the number of input variables in a model by focusing solely on pertinent data, thereby mitigating overfitting. It is a systematic and automated method for identifying the most important features from a provided dataset. RFE works by recursively fitting a model to the data, evaluating the importance of each element and removing the least significant elements at each iteration. This process is continued till the required number of features or a specific performance criterion is reached.

In simpler terms, RFE begins with all features in the dataset, repeatedly trains the model while iteratively removing the least significant features, and terminates when the predefined condition is met. The result is a subset one of the most informative and relative features for a particular given task, which can improve model efficiency, reduce overfitting, and improve interpretability.

2.5 Data Splitting

In this phase, the dataset gets split into a training set and a testing set, usually done to prevent overfitting. For instance, in the case of the Parkinson's disease dataset, 80% is allocated for training and 20% for testing. The training set is utilized to build models, while the test set is utilized to assesses their performance.

2.6 Machine Learning Model

Supervised machine learning is a branch of computer science where algorithms learn patterns from the labeled data to make predictions or decisions. It's a key concept in data analysis and predictive modeling. [2] The process involves training a model using a dataset with input-output pairs, allowing the model to understand relationships between inputs and corresponding outputs. There are two main types: regression, used for predicting continuous values, and classification, for categorizing data into predefined classes.

2.6.1 Random Forest Classifier

The random forest classifier comes under supervised learning machine learning algorithm which can be used for both classification as well as regression in Machine Learning [8].The concept behindthe Random forest classifier is ensemble learning, in which numerous classifiers are used to tackle a particularly complexproblem.A number of trees based on different selections of the provided dataset and also subset of the given dataset and averages them. Here we have applied Random forest classifier on our Parkinson's dataset and we got results as follows

Accuracy	Precision	F1-score	Recall
94.91%	1	0.947	0.9

Table 1

2.6.2 K – Nearest Neighbour

K- nearest neighbours comes under supervised machine learning techniques that sorts data by identifying its closest neighboring data points to classify it. KNN is more suitable for the dataset that contains numerical data and for small amount of features.

KNN doesn't make any assumptions about underpinning data distribution so it is known as non-parametric algorithm. KNN is instance-based because it makes predictions based on the instances (data points) in the training set. Here K is called hyper parameter and it plays a vital role in KNN. The choice of k is very important and depends on the nature of the provided data. A lesser values of k increases the model sensitivity to noise, whereas a larger k value makes it less sensitive but might smooth out the decision boundaries. Here we have applied K-Neighbours classifier on our Parkinson's dataset and we got results as follows

Accuracy	Precision	F1-score	Recall
86.44%	0.8928	0.862	0.8334

Table 2

2.6.3 Navie Bayes Classifier

Navies Bayes comes under supervised machine learning technique used for the classification of data. It model the input data for a given class. Navie Bayes approach works on the basis on the assumption that all the features in the given input data are independent to each other. In classification, it assists in figuring out the probability of a particular class given some observed features. It works as follows

1) Given a set of features and a class label, Naive Bayes measures the probability of each class for a given instance.

2) The class with highest probability is selected based on the predicted class for the instance.

Accuracy	Precision	F1-score	Recall
83.05%	0.95	0.807	0.7

Table 3

2.6.4 Ensemble Method

Here in ensemble method we have combined a group of classifiers using voting classifier method. In this method we have combined Random Forest classifier , Gaussian Naive Bayes classifier and K-Neighbour so that we get the better machine learning model to predict the output. The results are as follows :

Accuracy	Precision	F1-score	Recall
89.83%	0.961	0.892	0.83

Table 4

2.6.5 Extreme Gradient(XGBoost) Boosting

XGBoost, refers to Extreme Gradient Boosting, which is a famous machine learning algorithm used for both classification as well as regression tasks [9]. XGBoost incorporates several enhancements over traditional gradient boosting methods, Typically, the method begins with a basic model, such as a shallow tree. In order to fix the errors in the joined ensemble, additional trees are added repeatedly. Regarding the ensemble predictions, each tree is fitted to the loss function's negative gradient. To prevent overfitting, regularization terms are used to regulate the complexity of the individual trees. The final outcome will be found by summing up the predictions of all the trees in the ensemble. Here we have applied XGBoost classifier on our Parkinson's dataset and we got results as follows :

Accuracy	Precision	F1-score	Recall
98.93%	1	0.983	0.966

Table 5

3. Discussion

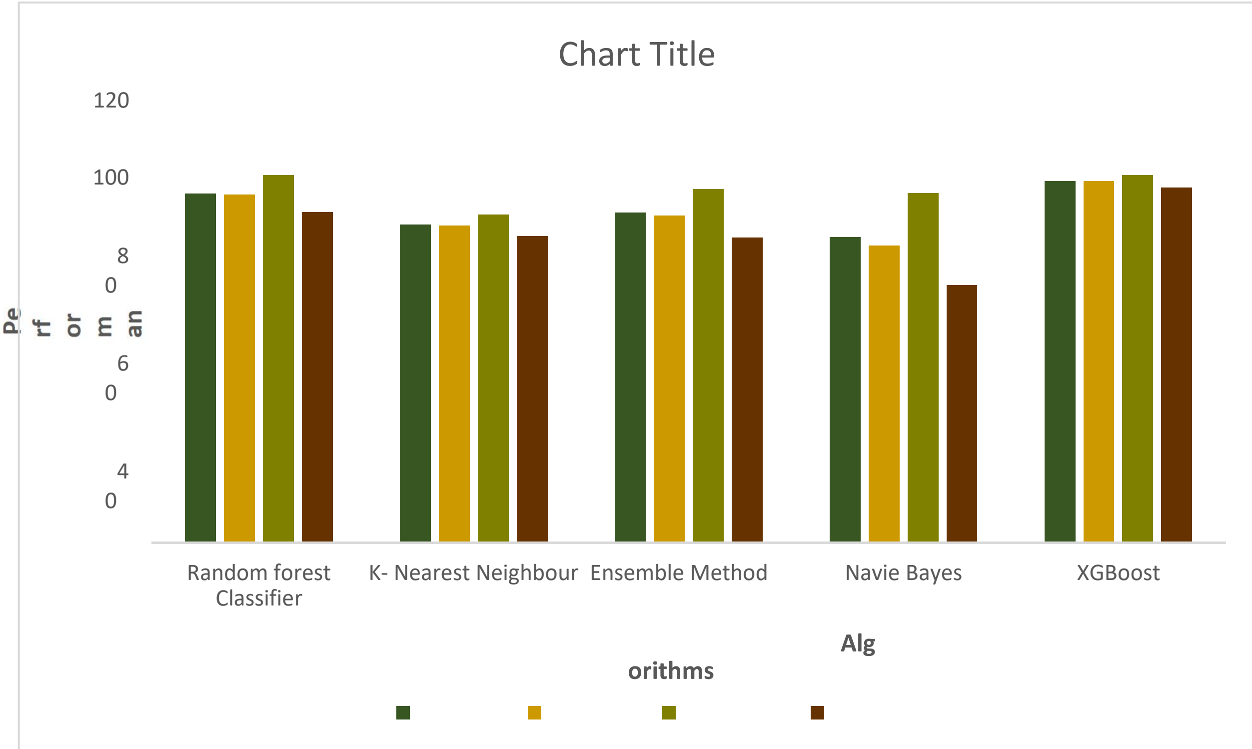


Figure 3. Bar graph of Performances

Table 6. [Performance of ALL Models]

Algorithms	Accuracy	f1 - score	Precession	Recall
Random forest Classifier	94.91	94.7	100	90
K- Nearest Neighbour	86.44	86.2	89.28	83.34
Ensemble Method	89.83	89	96.15	83
Navie Bayes	83.05	80.7	95	70
XGBoost	98.3	98.3	100	96.6

Table 6 All the above data in table are in %

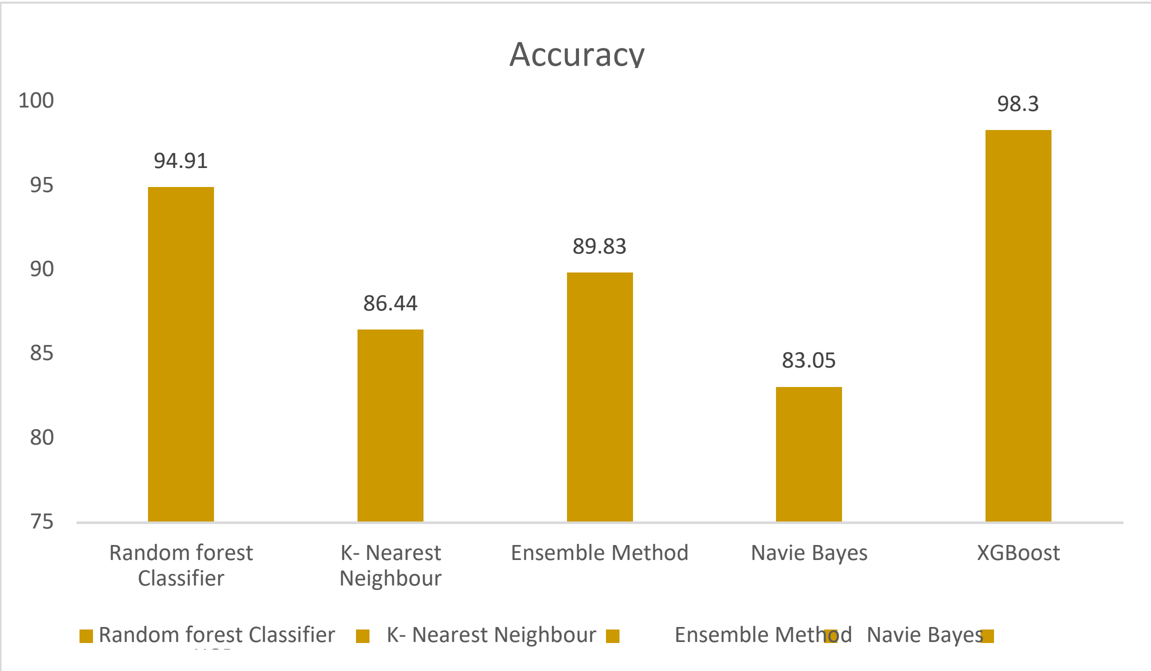


Fig 4 Accuracy of Models

4. Concluding Remarks

In this work a machine learning model was created to detect Parkinson's disease early by using speech signals from both individuals with Parkinson's and those who are healthy. In this research work the data set contains data imbalance so to balance the data in the dataset we have used synthetic minority oversampling technique(smote) . Late we have choose Recursive feature elimination that contains support Vector Machine (svm) to classifying the features to obtain the most important features so that the model performance gets enhanced. After selecting the features now we have performed different classifiers like K-nearest Neighbour classifier, decision tree classifier along with that we have tried an ensemble methods in which we have combine K-Neighbour classifier, Naïve Bayes Classifier, decision tree classifier and we got an accuracy of 86.44% . Later on we have tried with a boosting method i.e we have used Extreme Gradient Boosting (XGBoost) and we got an accuracy of 98.3% which gives better accuracy than ensemble method and other classifiers.

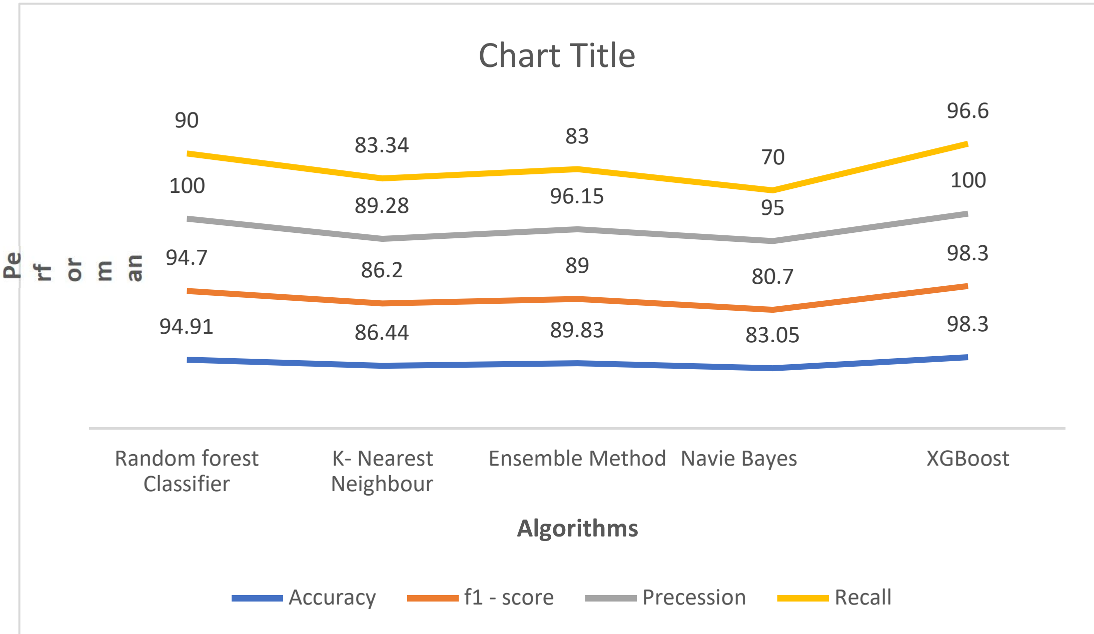
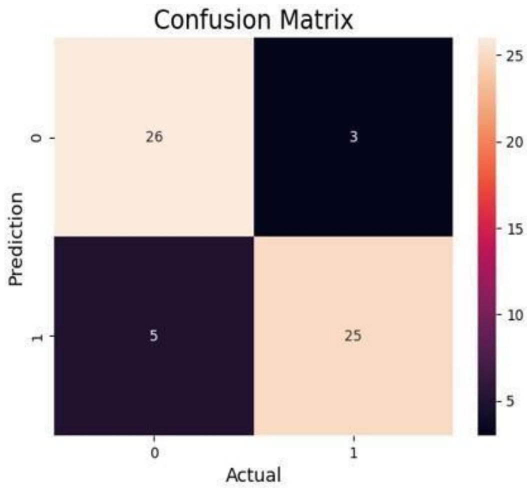
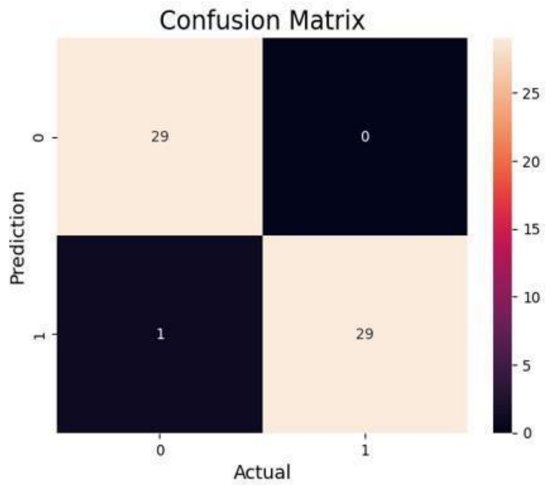


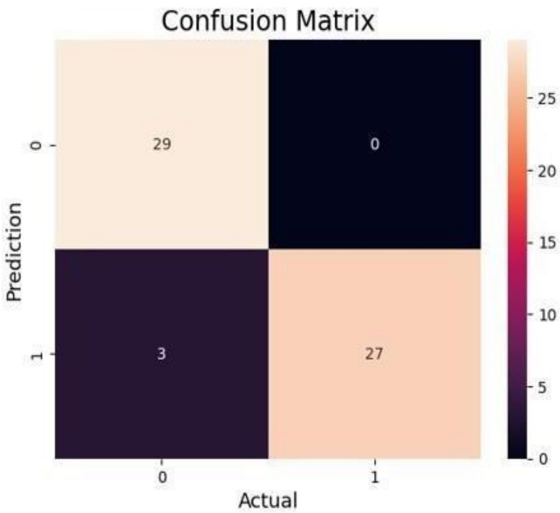
Figure 5 chart-line of Performances



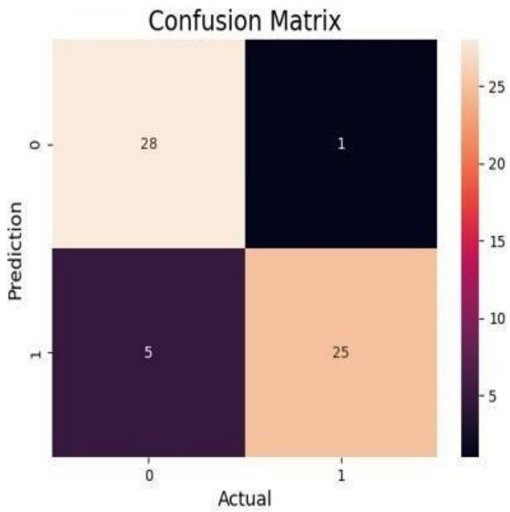
K-Neighbours Algorithm



XGBoost Algorithm



Random Forest Classifier



ensemble Method Classifier

Figure 6 Confusion Matrix of all Methods

5. Future Work

Machine learning has been making significant strides in the field of Parkinson's disease (PD) research, offering promising avenues for future work:

1. **Early Diagnosis and Prediction:** ML algorithms can analyze large datasets, including patient symptoms, genetic information, and imaging data, to identify patterns and markers that could lead to earlier and more accurate diagnosis of PD. Predictive models can also forecast disease progression based on various factors.

2. **Treatment Personalization:** ML can assist in tailoring treatment plans by analyzing patient responses to different medications and therapies. This personalized approach aims to optimize treatment effectiveness and minimize adverse effects.

3. **Monitoring Disease Progression:** Wearable devices equipped with sensors can collect continuous data on movement, tremors, gait, and other symptoms associated with PD. Machine learning algorithms can analyze this real-time data to track disease progression, allowing for timely interventions and adjustments to treatment plans.

4. **Remote Patient Monitoring and Support:** ML-powered applications can facilitate remote monitoring of PD patients, providing continuous support and feedback on symptoms. This can improve patient care, allowing for timely interventions and reducing the need for frequent hospital visits.

5. **Ethical Considerations and Bias Mitigation:** As with any medical application of AI, ensuring fairness, transparency, and ethical use of algorithms in PD research is crucial. ML models must be developed and validated on diverse datasets to mitigate biases and ensure their reliability across different demographics.

References

1. **Luis CS Afonso, Gustavo H Rosa, Clayton R Pereira, Silke AT Weber, Christian Hook, Victor Hugo C Albuquerque, and Jo˜ao P Papa**. A recurrence plot-based approach for parkinson's disease identification. Future Generation Computer Systems, 94:282–292, 2019.

2. **John Sahaya Rani Alex** et al. Early detection of parkinson's disease using motor symptoms and machine learning. arXiv preprint arXiv:2304.09245, 2023.

3. **Jefferson S Almeida, Pedro P Rebouc¸as Filho, Tiago Carneiro, Wei Wei, Robertas Damaˇseviˇcius, Rytis Maskeli¯unas, and Victor Hugo C de Albuquerque.** Detecting parkinson's disease with sustained phonation and speech signals using machine learning techniques. Pattern Recognition Letters, 125:55–62, 2019.

4. **Amira S Ashour, Majid Kamal A Nour, Kemal Polat, Yanhui Guo, Wafaa Alsaggaf, and Amira El-Attar.** A novel framework of two successive feature selection levels using weight-based procedure for voice-loss detection in parkinson's disease. Ieee Access, 8:76193–76203, 2020.

5. **Filippo Cavallo, Alessandra Moschetti, Dario Esposito, Carlo Maremmani, and Erika Rovini.** Upper limb motor pre-clinical assessment in parkinson's disease using machine learning. Parkinsonism & related disorders, 63:111–116, 2019.

6. **Prabhleen Kaur Chawla, Meera S Nair, Dattakumar Gajanan Malkhede, Hemprasad Yashwant Patil, Sumit Kumar Jindal, Avinash Chandra, and Mahadev Anant Gawas.** Parkinson's disease classification using nature inspired feature selection and recursive feature elimination. Multimedia Tools and Applications, pages 1–24, 2023.

7. **Sujata Dash.** A systematic review of adaptive machine learning techniques for early detection of parkinson's disease. Artificial Intelligence for Neurological Disorders, pages 361–385, 2023.

8. **J Divya, P Radhakrishnan, G Pavithra, Anandbabu Gopatoti, D Baburao, and R Krishnamoorthy.** Detection of parkinson disease using machine learning. In 2023 International Conference on Inventive Computation Technologies (ICICT), pages 53–57. IEEE, 2023.

9. **Ishu Gupta, Vartika Sharma, Sizman Kaur, and Ashutosh Kumar Singh**. Pca-rf: an efficient parkinson's disease prediction model based on random forest classification. arXiv preprint arXiv:2203.11287, 2022.

10. **Aishwarya Keller and Anukul Pandey**. Smote and enn based xgboost prediction model for parkinson's disease detection. In 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), pages 839– 846. IEEE, 2021.

11. **Jie Mei, Christian Desrosiers, and Johannes Frasnelli.** Machine learning for the diagnosis of parkinson's disease: a review of literature. Frontiers in aging neuroscience, 13:633752, 2021.

12. **Iqra Nissar, Danish Raza Rizvi, Sarfaraz Masood, and Aqib Nazir Mir**. Voice-based detection of parkinson's disease through ensemble machine learning approach: A performance study. EAI Endorsed Transactions on Pervasive Health and Technology, 5(19), 8 2019.

13. **Kemal Polat.** A hybrid approach to parkinson disease classification using speech signal: the combination of smote and random forests. In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), pages 1–3. Ieee, 2019.

14. **Enislay Ramentol, Yail´e Caballero, Rafael Bello, and Francisco Herrera.** Smote-rs b*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using smote and rough sets theory. Knowledge and information systems, 33:245–265, 2012.

15. **Suvita Rani Sharma, Birmohan Singh, and Manpreet Kaur.** Classification of parkinson disease using binary rao optimization algorithms. Expert Systems, 38(4):e12674, 2021.