Classifying Parkinson's Disease from Patient's Acoustic Features Using Deep Learning

A Thesis Submitted

In Partial Fulfillment of the Requirements for the degree of

MASTERS OF TECHNOLOGY
IN
Data Science

by

Tarakashar Das

2K22/DSC/20

Under the supervision of

Ms. Shweta Meena

Assistant Professor, Department of Software Engineering

Delhi Technological University

Department of Software Engineering
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

May, 2024

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENTS

I wish to express my sincerest gratitude to Ms. Shweta Meena for his continuous guidance and mentorship that she provided me during the thesis. She showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this thesis as well as its industrial relevance. She was always ready to help me and clear my doubts regarding any hurdles in this work. Without her constant support and motivation, this thesis would not have been successful.

Signature of Candidates

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Tarakashar Das, Roll No: 2K22/DSC/20 hereby certify that the work is being presented in the thesis entitled Classifying Parkinson's Disease from Patient's Acoustic Features Using Deep Learning in partial fulfillment of the requirements for the award of the Degree of Masters of Technology, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from August, 2024 to May, 2024 under the supervision of Shweta Meena.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made b the candidate is correct to the best of our knowledge.

29/5/2024

Signature of Supervisor

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE BY THE SUPERVISOR

Certified that Tarakashar Das (Roll no. 2K22/DSC/20) has carried out their search work presented in this thesis entitled "Classifying Parkinson's Disease from Patient's Acoustic Features Using Deep Learning" for the award of Masters of Technology from Department of Software Engineering, Delhi Technological University, Dekhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not from the basis for the award of any other degree to the candidate or to anybody else from this or any other University.

Signature 29/5/2024

Shweta Meena

Assistant Professor

Department of Software Engineering

Delhi Technological University

Date: Delhi

Place:

ABSTRACT

Parkinson's Disease (PD) refers to a neurological disorder that is caused due to damage to a part of the brain called the substantia nigra; besides this, the diagnosis of PD is so much more costly and lengthy process. Hence, a cost-effective and efficient system will be helpful for PD patients. Nowadays, the advancement of the algorithms of artificial intelligence is an opportunity to develop an efficient diagnosing system. Till now, no permanent cure has been established; however, early diagnosis of PD can be helpful to lead a better life. According to medical science, speech problem is one of the essential symptoms of PD. Therefore, 22 vocal features of the UCI machine learning repository speech dataset are investigated for diagnosing PD patients. Moreover, Artificial Neural Network (ANN), Ensemble Learning (EL) and Machine Learning (ML) are utilized for PD classification. As we need a faster classification system, feature selection techniques have great importance in developing a better system; we investigate excluding and including feature selection. Low Variance Filter, Analysis of Variance, Principle Component Analysis and Linear Discriminant Analysis, feature selection techniques, are incorporated to select the vital feature. As a result, our proposed ANN model achieves 100% accuracy, F1 score of 100%, recall of 100%, precision of 100%, Kappa Score of 1, and AUC of 1. The efficiency of our proposed on different datasets is demonstrated by comparing with recent research works. At last, an accuracy and classification time trade-off is established to determine the best classifier.

Contents

List of Tables

List of Figures

List of Symbols

TP	True Positive
FP	False Positive
TN	True Negative
TN	False Negative
P_0	The total samples.
P_e	The sum of correct and incorrect samples
Rec	Recall
$Prec$	Precision
Acc	Accuracy

Chapter 1

INTRODUCTION

Neurological disorder refers to the problem of the nerve system. Parkinson's Disease (PD) is one of them, which affects the nerve system. Hence, the efficiency of the body parts controlled by the nervous system is decreased gradually [CLI]. PD is different from others because it is caused due to the death of the brain cell. A chemical messenger of our brain known as dopamine is produced from that types of cell. The symptoms of PD started to be visible, when the level of dopamine is decreased. Initially, the symptoms increase slowly [LLC61]. At an early stage, 10 symptoms are visible. The 10 early symptoms are soft or low voice (speech problems), constipation, tremors, trouble sleeping, small handwriting, loss of smell, masked face, trouble moving or walking, dizziness, and stooping. Fig. 1.1 depicted the important symptoms of PD. Although many symptoms can be developed due to PD, soft or low voice symptoms are the most common early symptoms. Approximately, 75% of people affected by

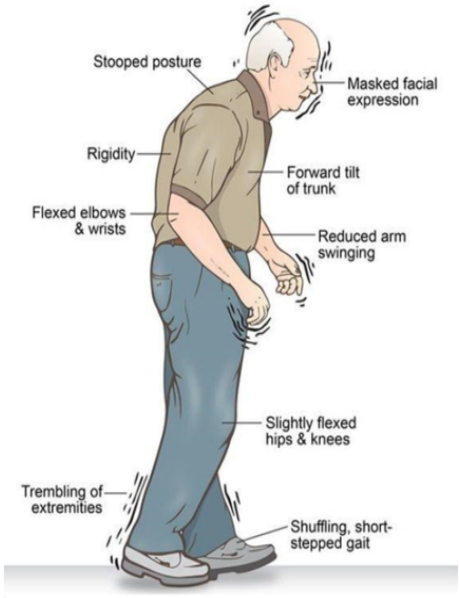

Figure 1.1: Important symptoms of Parkinson disease.

PD experience changes in their voice during the illness [LL61]. The symptoms, as mentioned earlier, are increased gradually according to the increment of loss rate of

brain cells located in substanstia nigra depicted at Fig. 1.2 . The recent belief is that genetic changes and some environmental factors cause the loss of nerve cells. Which decreases the amount of dopamine in the brain [Gov48]. Young and adults are rarely

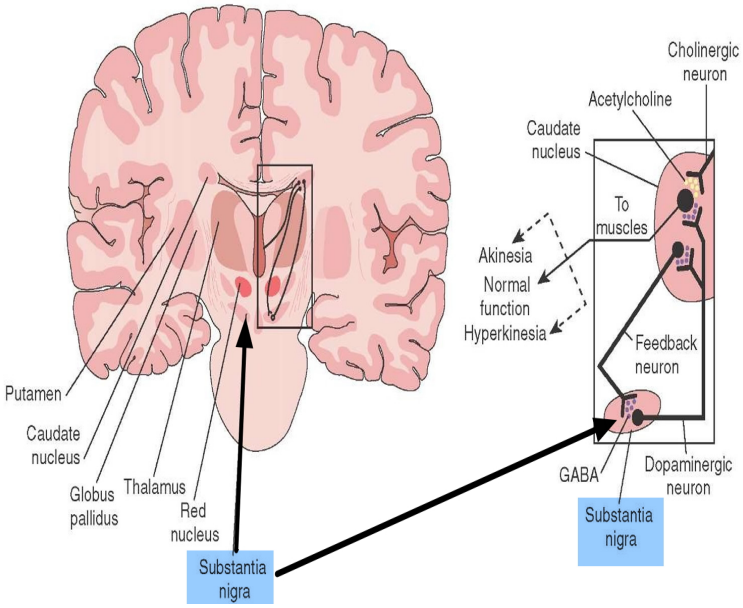

Figure 1.2: Illustration of Substantia Nigra.

affected by PD. People aged around 60 or greater than 60 are at risk of PD. Besides this, men are affected mainly by PD [CLI]. The death rate due to PD is increasing significantly than other neurodegenerative disease. In the past 25 years, the prevalence of PD has doubled [Org]. Almost one million people in the U.S. are struggling with PD, which will rise to 1.2 million by 2030. Hence, it is considered the second most common neurodegenerative disease. Additionally, around 10 million people worldwide suffer from PD, and 90 thousand U.S. people are diagnosed with PD yearly. There are two ways to get rid of this disease. The first one is medications, which cost $2,500 per year. On the other hand, therapeutic surgery costs up to $100,000 per person [LL61]. This massive amount of diagnosing cost and difficulties in diagnosing PD motivates the development of artificial intelligence-based diagnosis tools. Nowadays, different machine learning-based algorithms are used for the diagnosis of different types of disease, such as heart problem detection [KRK$^+$21, RARS21], diabetes prediction [GRK18, CTCA22, KSB$^+$21, AAM21], breast cancer detection [WLW$^+$19, SSR$^+$21], kidney disease detection [NA21, AAJH$^+$18] and so on.

Recently, researchers have applied machine learning and ensemble learning techniques for detecting Parkinson's disease. Some of them are using handwritten patterns of Parkinson's patients and applying machine learning, ensemble learning and deep learning techniques [KCÖ18, XP20, CAH$^+$20, PPR$^+$18, AIPV19, NNT$^+$21]. Besides this, some other researchers use patient questionnaires [DMH$^+$23, PR18] for detecting PD. The accuracy of these works, as mentioned earlier, is between 88%- 99.51%: 88% in [KCÖ18], 89.4% in [XP20], 93.3% in [CAH$^+$20], around 95% in [PPR$^+$18] and

99.51% in [PR18]. Besides, deep neural networks were applied along with machine learning and ensemble learning o diagnose PD because it has a vast hyperparameter tuning scope.However, classifying PD using handwritten patterns and questionnaires is very time-consuming and costly because of the different types of sensors and other equipment. Therefore, for classifying PD, we choose the PD-affected patients' voice data from the UCI machine learning repository Parkinson DataSet [Dat08]. The dataset's data is collected from the acoustic measurements of 31 individuals, 23 out of them are PD patients. From each acoustic measurement, 23 features are extracted from each voice recording, and 195 rows represent the individual's voice recording. Reducing classification time with high accuracy is crucial for a practical task. Different feature selection techniques are available for the establishment of trade-off between classification time and accuracy. To reduce the dimension of the data various feature selection technique is used.

Our contributions are:

- Investigate the potentiality of various ensemble methods and machine learning approaches in classifying PD. Where we find out the best hyperparameter by using hyperparameter tuning.
- The potentiality of artificial neural network has been investigated for classifying PD [LTAJ21].
- Investigate the potentiality of others hyper-parameters of ANN [Kar20].
- Implementation of hyper-parameter tuning on ANN to improve the performance [MPB19].
- Investigate the importance of various voice features in diagnosing PD using different feature selection techniques.
- Improving the overall accuracy of the work [Kar20, LTAJ21, MPB19].
- Investigate the chance of time complexity reduction by classification time and accuracy trade-off establishment. However, in performance measurement, accuracy is the crucial factor.

Chapter 2

LITERATURE REVIEW

Ensemble learning, machine learning (ML), and neural networks significantly impact computer-based disease classification and diagnosis in health care for accurate and faster decisions. Some of our related works are described here in 3 categories: (a) Patients Voice Features, (b) Patients Handwritten Drawings, (c) Patients EEG Based, and (d) Patients Questionnaires.

2.1 Benchmark Works Based on Patients Voice Properties

The work in [Kar20] proposed a machine learning-based method on the UCI machine learning repository dataset created by Max Little from the University of Oxford [Dat08]. They applied Support Vector Machine (SVM),Artificial Neural Network (ANN), and Classification and Regression Trees (CART). Besides this, they also performed different techniques for feature selection like Feature Importance (FI) and Recursive Feature Elimination (RFE). Finally, they got 93.84% for SVM with RFE feature selection. The study in [WLZG18] classified Parkinson from the UCI dataset [UCI09], and they used KNN, logistic regression (LR), RF, and deep multi-layer perceptron (DMLP). The results of this task showed that the DMLP has better potential with an accuracy of 80 percent for PD classification. Polat et al. [PN20] presented a unique methodology for patients' voice features based on PD classification. They investigated three distinct ML classifiers on the UCI machine learning repository database [UCI19]: SVM with Gaussian kernel, LR, and weighted KNN. As a result, the weighted KNN outperforms with 89.46 percent accuracy. In this study, Ali et al. [AZZL19] proposed a hybrid methodology for classifying PD on different kinds of vocal data. They collected data from the Neurology Department at Istanbul University [SIS+13] and investigated the potentiality of NN with the Genetic algorithm (GA) hyper-parameter tuning method. To reduce the dimension of the Dataset, they used Linear Discriminant Analysis (LDA) and achieved 95% accuracy. Lamba et al. [LTAJ21] classified Parkinson's disease from the UCI dataset [Dat08]. They investigated three ML algorithms, such as naive bayes (NB), k-nearest neighbor, and random forest (RF), and 3 techniques for feature selection: extra tree, genetic algorithm and mutual information gain. As a result, the RF classifier with a genetic algorithm showed a better performance with 95.58 percent. In this work,

Almeida et al. [ARFC$^+$19] investigated the importance of patients' voice and ML methods in PD diagnosis. Hence, they implemented KNN, optimal power flow, SVM, and MLP on the Dataset collected from smartphone (SP) and acoustic cardioid (AC), achieving the best accuracy of 94.55 percent for KNN. Mathur et al. [MPB19] presented an ML-based system on various acoustic features from the UCI dataset [Dat08]. They applied KNN, SVM, and ANN to this Dataset. Finally, they got the best accuracy of 91.28% for AdaBoost with KNN. Table 2.1 presents all the acoustic feature based tasks mentioned earlier.

Table 2.1: Summary of some relevant task on PD classification using acoustic features

References	Name of Dataset	Number of Data	Feature Selection	Best model	Accuracy
[Kar20]	UCI machine learning repository [Dat08]	195	Yes	SVM	93.84%
[LTAJ21]	UCI machine learning repository [Dat08]	195	Yes	RF	95.58%
[MPB19]	UCI machine learning repository [Dat08]	195	No	AdaBoost	91.28%
[WLZG18]	Parkinsons Telemonitoring dataset of UCI [UCI09]	5875	No	DMLP	80%
[PN20]	UCI machine learning repository [UCI19]	240	No	KNN	89.46%
[AZZL19]	Department of Neurology at Istanbul University [SIS$^+$13]	1208	Yes	LDA-NN-GA	95%
[ARFC$^+$19]	Selfmade dataset	N/A	No	KNN	94.55%

2.2 Related Works based on Handwritten Drawing

The authors in [KCÖ18] proposed a CNN-based model for finding the underlying pattern of different drawings from the Wacom Cintiq 12WX graphics handwriting dataset and achieved an accuracy of 88%. Chakraborty et al. [CAH$^+$20] proposed a methodology based on CNN to distinguish between healthy and PD patients. The proposed CNN model was applied to the Kaggle's data repository dataset. Finally, they obtained 93.3% percent accuracy. Pereira et al. [PPR$^+$18] utilized a CNN model to find the hidden pattern of handwritten drawings images from the HandPD dataset. The CNN model achieved around 95% accuracy in distinguishing healthy and PD patients. The work in [XP20] suggested a method based on ensemble learning, using RF and the PCA method together on the NewHandPD dataset. As a result, the model shows 89.4 percent accuracy. All the above mentioned handwritten drawing based work is presented briefly in Table 2.2.

Table 2.2: Brief description of some relevant task on PD classification using handwritten drawing

References	Name of Dataset	Feature Selection	Number of Data	Best Model	Accuracy
[KCÖ18]	UCI Dataset 2018 [SSG$^+$19]	No	72	CNN	88%
[XP20]	NewHandPD [PDCSM19]	Yes	594	RF	89.4%
[CAH$^+$20]	UCI Dataset 2018 [SSG$^+$19]	No	55	CNN	93.3%
[PPR$^+$18]	HandPD dataset	No	92	CNN	95%

2.3 Benchmark Works using EEG

Author in [dOdSA+20] proposed an EEG-based PD diagnosis methodology. Their goal is to classify into three groups. For classification purposes, they performed morphological extreme learning machine, Bayes net, multilayer perceptron (MLP), naive Bayes, SVM, J48, extreme learning machine, random tree, random forest, and morphological extreme learning machine. As a result, a random forest with 50 trees outperforms with an accuracy of 99%. Aljalal et al. [AAA+22] researched the potentiality of EEG signals in diagnosing PD. Hence, they used two public datasets. The first is from the University of San Diego, California [EEG20], and the second is from the study conducted by the University of New Mexico [EEG17]. They used linear discriminant analysis, RF, KNN, Quadratic discriminant analysis and SVM for classification purposes, achieving the highest accuracy of 99% for KNN. The work in [OHR+20] EEG signals are used to diagnose PD. The EEG signals were collected from the Hospital Universiti Kebangsaan Malaysia Ethics Committee, and applied 13-layer CNN on the signals. Finally, they achieved 88.25 percent accuracy, 91.77 percent specificity, and 84.71 percent sensitivity.

All the works related to EEG are described in Table 2.3.

Table 2.3: Brief description of related works on Parkinson's disease using EEG signals

References	Name of Dataset	Feature Selection	Number of Data	Best Model	Accuracy
[dOdSA+20]	N/A	N/A	Yes	RF	99%
[AAA+22]	SanDiego dataset [EEG20] and UNM dataset [EEG17]	31, 54	No	KNN	99%
[OHR+20]	Hospital Universsiti Kebangsaan Malaysia Ethics Committee	40	No	CNN	88.25%

2.4 Benchmark Works using Questionnaire

Tarakashar et al. [DMH+23] proposed an artificial neural network-based system on the PPMI dataset [MJL+11]. They applied several feature selection techniques along with some traditional machine learning and ensemble learning methods. As a result, they achieved 99.51% accuracy for the ANN model. Prashanth et al. [PR18] applied various ML techniques like LR, RF, SVM, and boosted trees to build predictive models that could classify healthy normal and PD patients. They used the PPMI [MJL+11] dataset and achieved the accuracy of >95%, as shown in Table 2.4.

Table 2.4: Summary of patients questionnaire based work on PD

References	Name of Dataset	Feature Selection	Number of Data	Best Model	Accuracy
[DMH+23]	PPMI [MJL+11]	5704	Yes	ANN	99.51%
[PR18]	PPMI [MJL+11]	5704	Yes	SVM	>95%

Chapter 3

METHODOLOGY

Firstly, the dataset is imported from Google Drive, and exploratory data analysis (EDA) is employed for better understanding of the insights of the data. Then, a missing value handling operation is applied to fill in the missing values if they exist. After that, data scaling is implemented to scale the whole dataset into a specific range. As our dataset is imbalanced, hence oversampling method is used to balance the dataset. Following the dataset balancing step, we applied 2 methods. For the first one, we took all the features together. For the second one, to reduce the model's dimensionality and time complexity various feature selection methods were employed. Afterward, we split the dataset into two parts, trained our model on the training dataset, and analogously performed the random search hyper-tuning method to find out the best parameter set. Finally, the model performance is evaluated using f1-score, recall, accuracy, etc. Fig. 3.1 depicted the workflow of our suggested framework.

3.1 Dataset Details

The dataset used in this study is collected from the UCI machine learning repository [Dat08]. The UCI machine learning repository is a big database the machine learning expert uses. Our dataset contains 195 observations, each with 23 features described in Table 3.1, collected from 31 people. Out of 31 people, 23 are with PD, and the rest of the people are healthy. The dataset's feature characteristics are real, meaning the real number represents all the features. Each column represents a specific voice measure, and each row denotes the individual voice recording. All 23 features are related to the voice signal, which defines the characteristics of the voice signal. Out of 195 observations, 147 are PD patients' data and 48 observations from healthy people. The whole dataset was splited into two parts, and the biggest portion of 80% data is used as a training dataset, and 20% of the data for testing. Our target is to classify the PD based on the 22 features. Because the status column defines the target variable, where 0 means healthy person and 1 means PD patients. Table 3.2 describes a short description of the features.

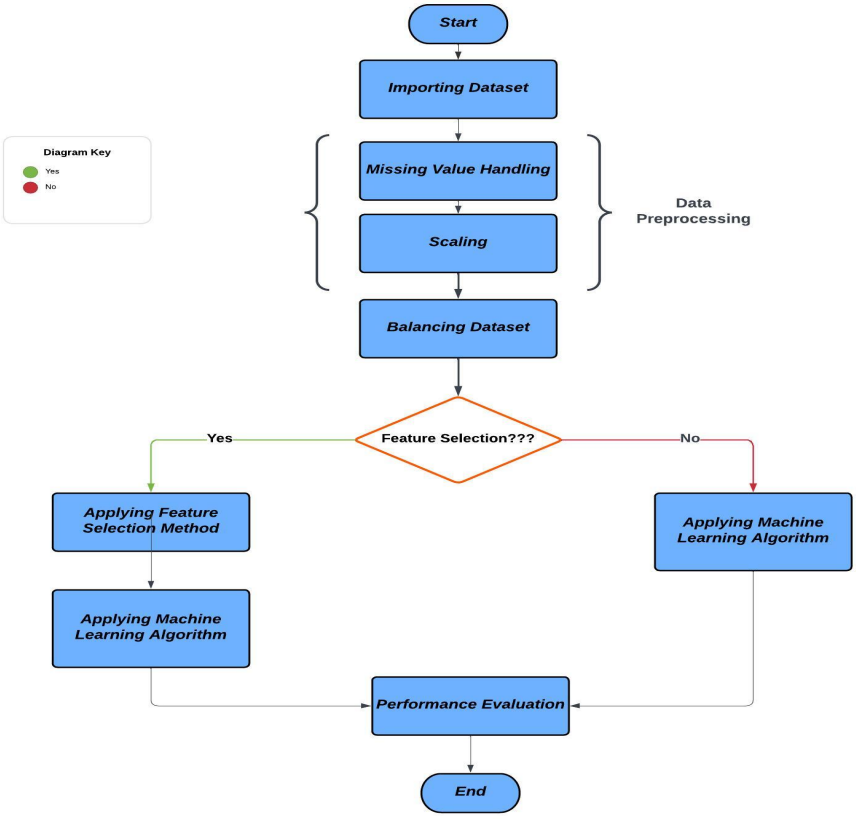

Figure 3.1: The proposed framework for PD classification.

3.2 Exploratory Data Analysis(EDA)

To fine-tune the used dataset, it is an important step. Because it helps a researcher discover the data's insights and behavior, sometimes, the dataset may include outliers and missing values, which misguides the model and leads to a bad prediction. We can find out and handle the missing values and outliers by applying EDA before the model training. Different types of Data visualization packages, NumPy, Pandas, and Statistical Methods, are used in this study for EDA.

3.2.1 Statistical analysis

It is a process that includes collecting and analyzing massive amounts of data. When we conduct research, we often have to go through a huge amount of data. Analyzing these large amounts of data within a short time is difficult. Nevertheless, we can do it in a short time by applying different types of statistical parameters like mean, standard deviation(std), quartile values, minimum number(min), maximum number(max), etc. In this research work, we applied descriptive statistical analysis, which gives us a clear description of the data. Table 3.3 illustrates a short statistical description of the features

Table 3.1: Description of used dataset

Details	Source of Information
Source of dataset	UCI machine learning repository [Dat08]
Type of disease	Parkinson's Disease (Neurodegenerative)
Observations	195
Features	23
Representation of target class	Binary(1-PD, 0-Healthy)

of our dataset.

3.2.2 Missing Value Handling

Data included in our dataset are collected from real patient records. Therefore, there are many chances of having missing values. Data can be missing for many technical issues and other reasons. Missing data is a barrier for many statistical analysis methods and machine-learning models. Hence, we checked the missing values in our dataset, and as a result, we got false. That means our dataset has all the values.

3.3 Data Preprocessing

3.3.1 Feature scaling

Collected data from real patients may be in different units, and the distance between the values of the features may be large. Due to this big distance between feature values, the gradient descent process takes more time, increasing the training time. So, to avoid this problem, we employed a standardization technique to convert the real values of the features in the range of -1 to 1 [Gee]. The equation used for accomplishing this is given in Eq. 3.1.

$$a_i = \frac{a_i - Mean(a_i)}{SD(a_i)} \tag{3.1}$$

here, the Mean(a_i) indicates the mean of a, SD(a_i) defines the standard deviation of a and a_i represents the example of feature a.

Table 3.2: Brief description of the available features in dataset

Feature No.	Name of Features	Description
1	MDVP:Fo(Hz)	Average fundamental frequency of vocal
2	MVDP:Fhi(Hz)	Maximum fundamental frequency of vocal
3	MVDP:Flo(Hz)	Minimum fundamental frequency of vocal
4	MVDP:Jitter(%)	Fundamental frequency measurement
5	MVDP:(Abs)	Fundamental frequency measurement
6	MVDP:RAP	Fundamental frequency measurement
7	MVDP:PPQ	Fundamental frequency measurement
8	Jitter:DDP	Fundamental frequency measurement
9	MDVP:Shimmer	Variation in amplitude measurement
10	MDVP:Shimmer(dB)	Variation in amplitude measurement
11	Shimmer:APQ3	Variation in amplitude measurement
12	Shimmer:APQ5	Variation in amplitude measurement
13	MDVP:APQ	Variation in amplitude measurement
14	Shimmer:DDA	Variation in amplitude measurement
15	NHR	The ratio of noise to tonal components in the voice measurement
16	HNR	The ratio of noise to tonal components in the voice measurement
17	status	The health status of the subject (one) - Parkinson's, (zero) - healthy
18	RPDE	Nonlinear dynamical complexity measures
19	D2	Nonlinear dynamical complexity measures
20	DFA	Signal fractal scaling exponent
21	spread1	Nonlinear measures of fundamental frequency variation
22	spread2	Nonlinear measures of fundamental frequency variation
23	PPE	Nonlinear measures of fundamental frequency variation

3.4 Balancing Dataset

Balancing dataset is a method used to balance the percentages of the data belonging to different classes. Imbalanced dataset may lead to bad model performance because it makes the model biased to the majority class. Our dataset consists of 147 data from PD patients and 48 from a healthy person. Hence, our dataset is imbalanced, and we have to balance our dataset before model training to avoid a biased model. As we have a small amount of data, the Synthetic Minority Oversampling Technique (SMOTE) is performed for balancing the dataset by generating new data from the minority class. After applying SMOTE the number of data of the minority class increased from 48 to 147.

3.5 Feature Selection (FS)

When a dataset contains so many features to represent individual data, and all the features do not impact the classification, choosing a subset of features is an essential

Table 3.3: Descriptive analysis of the features in dataset

Feature No.	Feature Name	Mean	Standard deviation	Minimum	Maximum
1	MDVP:Fo(Hz)	154.23	41.39	88.33	260.10
2	MVDP:Fhi(Hz)	197.10	91.49	102.14	592.03
3	MVDP:Flo(Hz)	116.32	43.52	65.48	239.17
4	MVDP:Jitter(%)	0.01	0	0	0.03
5	MVDP:(Abs)	0	0	0	0
6	MVDP:RAP	0	0	0	0.02
7	MVDP:PPQ	0	0	0	0.02
8	Jitter:DDP	0.01	0.01	0	0.06
9	MDVP:Shimmer	0.03	0.02	0.01	0.12
10	MDVP:Shimmer(dB)	0.28	0.19	0.08	1.30
11	Shimmer:APQ3	0.02	0.01	0	0.06
12	Shimmer:APQ5	0.02	0.01	0.01	0.08
13	MDVP:APQ	0.02	0.02	0.01	0.14
14	Shimmer:DDA	0.05	0.03	0.01	0.17
15	NHR	0.02	0.04	0	0.31
16	HNR	21.89	4.43	8.44	33.50
17	RPDE	0.5	0.10	0.26	0.69
18	D2	0.72	0.06	0.57	0.83
19	DFA	-5.68	1.09	-7.96	-2.43
20	spread1	0.23	0.08	0.01	0.45
21	spread2	2.38	0.38	1.42	3.67
22	PPE	0.21	0.09	0.04	0.53

step before training the model. Because unnecessary features may lead to bad prediction (overfitting) and high computational time [Gee18], many methods can be used to reduce the dimensionality. In our work, we apply 4 feature selection (FS) techniques: Analysis of variance (ANOVA), Low variance filter (LVF), and Linear discriminant analysis (LDA), and Principal component analysis (PCA) . In this part, the FS mentioned above methods are described.

3.5.1 Analysis of Variance (ANOVA)

Analysis of Variance is a statistical technique to study the significance of more than two samples [Ano]. Means of the samples to measure the differences and variability between the samples and within the samples. ANOVA test gives the results in F statistic (also known as F-ratio), which is used to determine the significance of the samples or features. As our used dataset has more than two samples, we employed an ANOVA test, and based on the F-statistic, we selected the significant features for model training. Here, we set the mean of the f statistic value of all the features as the significance threshold and select those features for model training with the F statistic value greater than the threshold value. The F statistics value of all features are depicted

in Fig. 3.2 and from this figure, it is clear that the features named MDVP: Fo(HZ), Shimmer: APQ5, spread1, PPE, MDVP: Jitter(Abs), MDVP: Shimmer, spread2, and MDVP: APQ. have more F statistic values than threshold values; hence, those features are selected as significant for the models of this work.

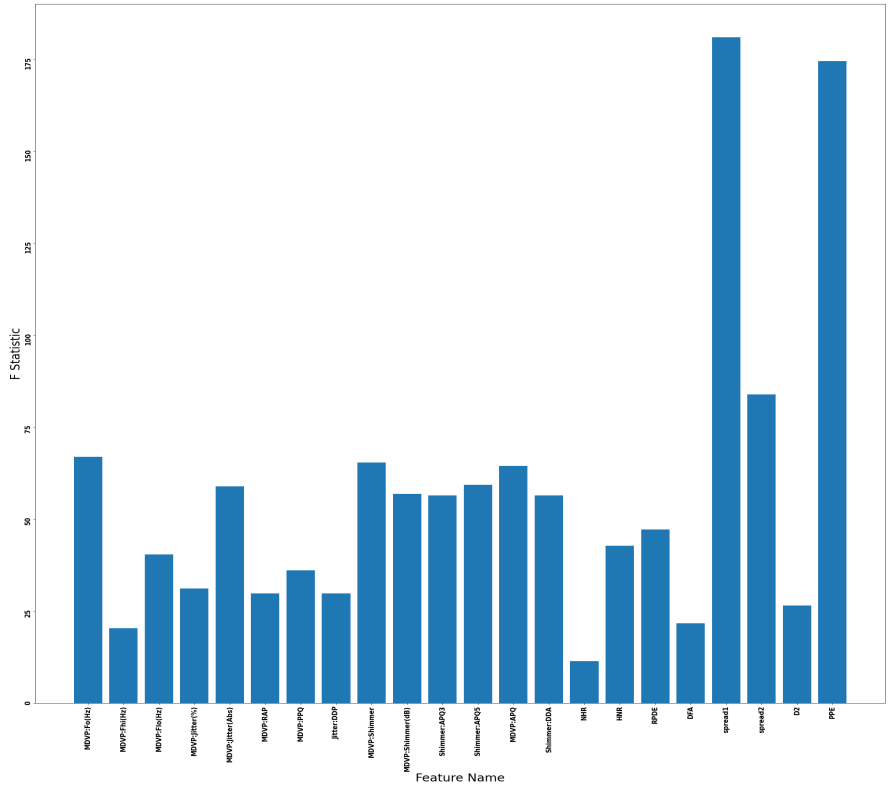

Figure 3.2: F statistic value of all features for ANOVA test.

3.5.2 Low Variance Filter (LVF)

Our dataset contains 22 features, but it is optional that all of those have sufficient variance to help a model in prediction. The features with low variance have a low impact on the model prediction. Hence, we can reduce the dimensionality by choosing those features for model training, which has sufficient variability in data. This technique is called the low variance filter method [LVF]. In our study, we set the mean of the variance (0.8777) of 22 features as threshold and separate 9 features, which have variance greater than the threshold value. Variance scores of all features are depicted in Fig. 3.3, and it is clear from the bar plot that most of the features have variance scores less than 0.8.

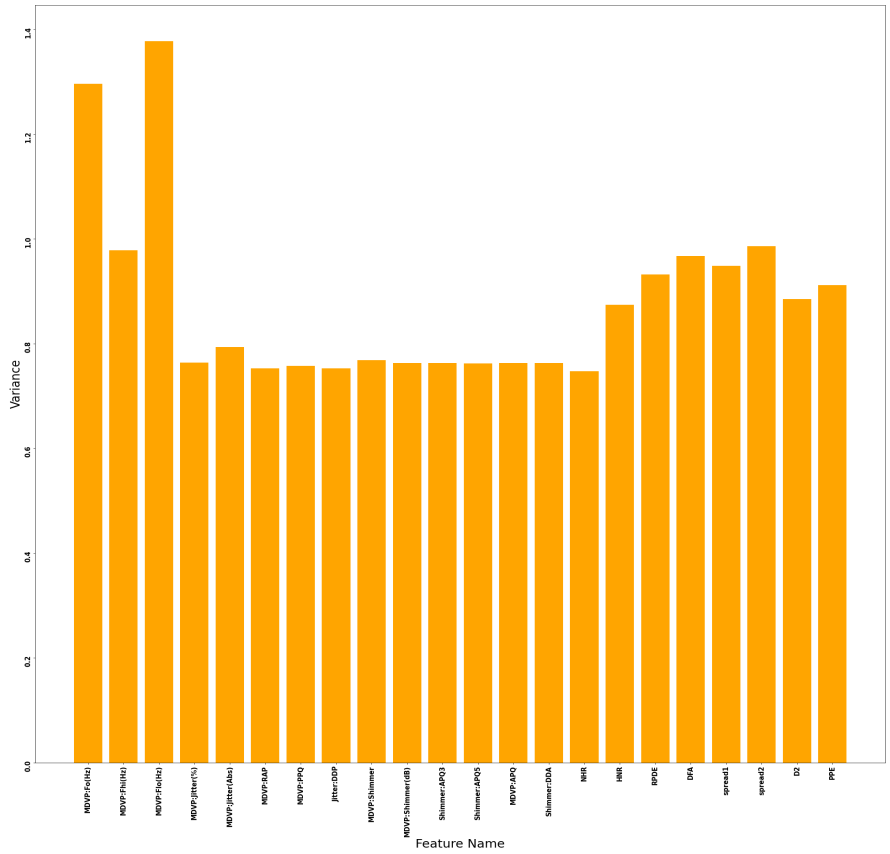

Figure 3.3: Variance score of all features for LVF.

3.5.3 Principal Component Analysis (PCA)

The mostly used dimensionality reduction method is principal component analysis (PCA). It not only reduces the dimension of the data but also enhances its interpretability. Besides this, it converts correlated variables into uncorrelated variables using orthogonal transformation. This orthogonal transformation generates orthogonal axes, called principle components, which capture the maximum variance of the data. The first principle component captures the most variability, followed by the second principle component [PCA]. In our study, we choose the first 12 components, which best represent the used dataset. In Fig. 3.4, 1 represents PD and 0 represents Healthy.

3.5.4 Linear Discriminant Analysis (LDA)

The technique that maximizes the separation between the classes by projecting the data onto a lower-dimensional space is called linear discriminant analysis (LDA). It is applied to find out the linear combination of features so that the classes in the dataset can be separated easily [LD]. There are two assumptions for LDA. The first assumes

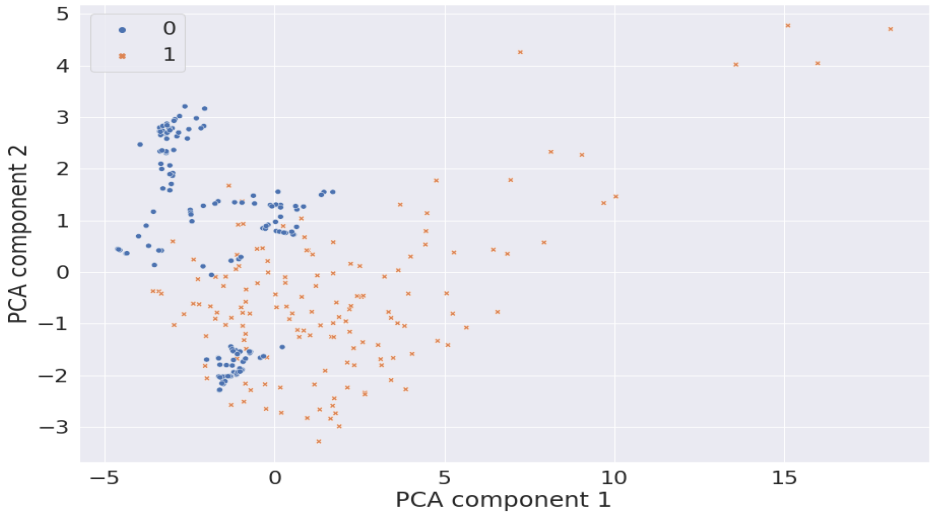

Figure 3.4: Visualization of the data using first two principle components.

that the data has a Gaussian distribution, and the second is that the data is linearly separable. Our dataset contains 2 classes, so we used only 1 linear discriminant in our study.

3.6 Hyper-parameter Tuning using Randomized Search Algorithm

In machine learning, hyper-parameter tuning is a common part of finding out the best combination of parameters. Randomized Search (RS) selects the most optimized parameter sets by choosing the parameter randomly. It is one of the most efficient methods of hyperparameter selection. Hence, RS is used in this work to find out the ideal parameters of classifiers for improving the model's accuracy. Without finding the best parameter a classification model cannot perform well [RS].

3.7 Classification

Various classification models with best hyper-parameters are described in this section.

3.7.1 Logistic Regression (LR)

Logistic Regression is a basic machine learning algorithm based on probabilistic predictions. It produces binary output, which helps to solve binary classification problems. The sigmoid function is used, followed by the linear regression equation [KZ01], where the equation used in LR is mentioned in the Eq.3.2.

$$p = \frac{1}{1 + e^{-x}} \tag{3.2}$$

Using Randomized Search, max_iter = 365, C = 0.62, random_state = 1, and solver = 'liblinear' are found as the best hyper-parameter for LR.

3.7.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised machine learning techniques mostly used for classification purposes. Besides this, it is used to solve regression problems. It creates the best line, which divides the n-dimensional plane into different classes using support vectors [ad]. This best line is called a hyperplane. Different kernels separate linear and non-linear data points, making them stronger classifiers than others. Hence, three different kernels, polynomial, linear, and RBF, are used to examine the potential of the SVM classifier. The best hyper-parameter combination of c = 1, gamma = 1, and kernel = 'rbf' is found using hyper-parameter tuning.

3.7.3 Decision Tree (DT)

A tree-based supervised ML algorithm which creates a shape like tree using the features of the data. Every root node of the tree is connected with the leaf nodes, which defines the different class labels, and the internal nodes are used as the decision criteria. The whole dataset is split using criteria like Gini Index and Information Gain. These criteria help the DT algorithm improve classification accuracy by finding out the best split [GKA21]. The most important advantage of DT is that it visualizes the selected features in a tree shape and robust to outliers and . We used hyperparameter tuning to find the best hyperparameters, such as criteria = entropy, max_features = 4, max_depth = 6, and min_sample_leaf = 1.

3.7.4 Random Forest (RF)

Random Forest is an ensemble learning technique that follows bagging approach. It creates many decision trees on different sample data and generates output. After generating output from all the decision trees, a majority vote is taken to generate the final output for the classification problem [Sai]. It is the most used algorithm because it can solve the overfitting problem and robust to outliers. Besides this, it can perform better most of the time without hyperparameter tuning.

3.7.5 K-Nearest Neighbour (KNN)

A distance-based supervised ML algorithm. A general assumption is made that a comparison can be made between new and existing cases. This comparison is made by using the Euclidean distance formula. It is also known as a lazy learner. Because no training phase is needed for this algorithm, it memorizes the existing data. When a new case is taken for classification, the algorithm first takes several nearest neighbors using

the Euclidean distance formula. It makes a majority vote on the selected neighbors. After that, the new case is sorted into the category according to the majority of votes [Sha]. The main advantages of this algorithm are that it is simple to use and can quickly classify new instances into suitable categories. In this study, 1 neighbor is selected using hyper-parameter tuning.Eq.3.3 is used to calculate the distance in KNN :

$$\sqrt{\sum_{j=1}^{n}(x_j - y_j)^2} \tag{3.3}$$

3.7.6 Adaptive Boosting (ADA)

A ensemble learning technique that repeats the training process of weak learners to convert them into a strong learners is known as Adaptive Boosting. Weak learners are sorted together and combined into a strong classifier to increase the classifier's performance. In the internal process of ADA, an increment of the weight of unclassified points happens, and the weight of classified points decreases [Kur]. A strong learner is created in this work by using 20 weak learners, called stumps. Besides this, other used hyperparameters of this work are algorithm= 'SAMME,' base_estimator = 'DecisionTreeClassifier,' mas_feature= 'auto,' random_state= 11 and learning_rate= 0.98.

3.7.7 Extreme Gradient Boosting (XGB)

It is an ensemble method that generates trees sequentially one by one. The second generated tree will solve the error that occurred by the first one [xgb]. This process will continue until a good performance is achieved from the final model. Learning from weak learners' weight helps the algorithm to build a strong learner. This algorithm is well known for its high accuracy and efficiency. Besides this, large datasets can be handled by this algorithm easily. To determine the optimum hyper-parameters of the XGB classifier hyper-parameter tuning is used and an outstanding performance is achieved for 900 decision trees, and each tree has a maximum depth of 2.

3.7.8 Artificial Neural Network (ANN)

A deep learning method based on the human brain's mechanism. Here, many neurons like human brains work together to make the computers like human brains in decision making [jpo20]. An ANN consists of several hidden layers, with 1 input layer at the beginning of the network and 1 output layer at the end. These layers are made of one kind of element called a neuron, whose activity is the same as the human brain's neurons, and several edges connect all the neurons. Besides this, various activation functions like sigmoid, relu, softmax are applied in output and all hidden layers. ANN performs well for regression and classification tasks [tpe22]. An ANN model consist of 1 input layer, 3 hidden layers, and 1 output layer is used in this work. As we are classifying whether a person has PD or not, the sigmoid activation function is used in output layer for binary output. The whole ANN function works like a black box; hence,

parameter optimization can result in good accuracy. This work applied hyperparameter tuning to find the best combination of hyperparameters by changing the combination number of hidden layers, number of neurons in each layer, activation functions in each layer, and two loss functions. Mean Absolute Error (MAE) and Mean Squared Erro (MSE) is employed as loss function. As a result, MSE outperformed others. Additionally, we applied three optimizers named Stochastic Gradient Descent (SGD), Adam, and Root Mean Squared Propagation. The Adam optimizer perform better than other optimizer. In addition, each hidden layer is added one after another. When we added the third hidden layer after the second hidden layer, it gave an outstanding result. So, to improve the result, we added the hidden layer 4, and the model's performance decreased. Hence, we delete the hidden layer 4 from the model and build the final model with 3 hidden layers, MSE as loss function and Adam as optimizer. Finally, according to the performance metrics like Accuracy (Acc.) and F1_score (F1), the 39th experiment shows the best performance. Table 3.4 shows hyperparameter details of the final ANN model.

Table 3.4: Optimized Hyper-parameters of ANN model

Name of parameter	Values	Name of paramete	Values
Input layer neurons	22	Number of epochs	500
First hidden layer neurons	30	Model optimizer	Adam
First hidden layer activation	Sigmoid	batch_size	32
Second hidden layer neurons	40	Model loss function	MSE
Second hidden layer activation	Sigmoid	Output layer neuron	1
Third hidden layer neurons	10	Activation in output layer	Sigmoid
Third hidden layer activation	Sigmoid		

Chapter 4

RESULTS and DISCUSSION

4.1 Requirements of Hardware

Experiments in this work were conducted on an 11th Gen Intel Corei5 along with Windows 11 operating system and 16 Gb RAM configuration.

4.2 Performance Evaluation Metrices

To investigate the performance of the trained model, some performance indicators like recall/sensitivity, precision, accuracy, F1_score, Cohen's Kappa Score (Kappa Score), and AUC are employed in this study. Equations of these performance indicators are given in Eq. (4.2) - (4.3)

$$Recall(Rec.) = \frac{TP}{FN + TP} \tag{4.1}$$

$$Precision(Prec.) = \frac{TP}{FP + TP} \tag{4.2}$$

$$Cohen'sKappaScore = \frac{P_0 - P_e}{1 - P_e} \tag{4.3}$$

$$F1 - score = 2 \times \frac{Rec * prec}{Rec + prec} \tag{4.4}$$

$$Accuracy(Acc.) = \frac{TN + TP}{FP + TN + FN + TP} \tag{4.5}$$

FN, FP, TP, and TN stand false negative, false positive, true positive, and true negative respectively. Additionally, P_e denotes the sum of correct and incorrect samples, and the total samples are represented as P_0.

4.3 Performance of classifiers without using Feature Selection

For the first method of this work, all the features are considered important for this research work. Hence, all the features were included from the used dataset [Dat08], and

various results were obtained from different classifiers. For considering all features, the F1_score for ANN, ADA, XGB, KNN, RF, DT, SVM, and LR is 100%, 92.59%, 96.29%, 92.31%, 92.59%, 92.59%, 88.46% and 80.85%, respectively and the accuracy is 100%, 93.22%, 96.61%, 93.22%, 93.22%, 93.22%, 89.83%, and 84.74%, respectively as represented in Table 4.1.

Fig. 4.1, illustrated that the ANN performs better than other classifiers. ANN has a big hyper-parameter tuning scope; hence, it can explore the insights of the data better than other classifiers. As a result, ANN can learn from the data well and performs better than other classifiers. In addition, we achieved 100% precision, 100% recall, 1 AUC, and 1 Kappa Score from the ANN model.

Table 4.1: Classifier's performance without using feature selection

Model	Acc.	Prec.	Recall	F1-score	AUC	Kappa Score	Classification time (mean)
SVM	89.83%	92%	85%	88.46%	0.8946	0.79	0.00002(Sec)
DT	93.22%	92.59%	92.59%	92.59%	0.9317	0.8634	0.000008(Sec)
LR	84.74%	95%	70.37%	80.85%	0.8362	0.69	**0.0000063(Sec)**
KNN	93.22%	96.00%	88.88%	92.31%	0.9288	0 .8626	0.0000619(Sec)
RF	93.22%	92.59%	92.59%	92.59%	0.9317	0.8634	0.000035(Sec)
ADA	93.22%	92.59%	92.59%	92.59%	0.9317	0.8634	0.0003(Sec)
XGB	96.61%	96.29%	96.29%	96.29%	0.9658	0.8634	0.000036(Sec)
ANN	**100%**	**100%**	**100%**	**100%**	**1**	**1**	0.0008(Sec)

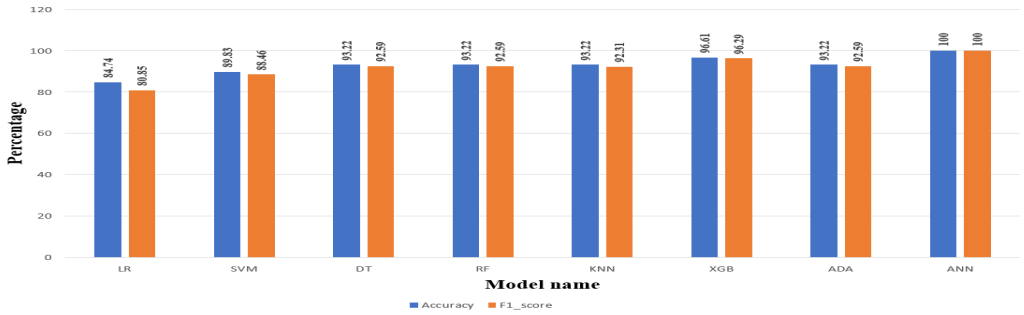

Figure 4.1: Comparison of performance between different classifiers for all features.

4.4 Performance of classifiers using Feature Selection

In this section, we went through the second method of this study, including FS strategies for reducing the model complexity. Before model training, the ANOVA, LDA, LVF, and PCA FS methods were applied. The ANN model was not considered here because ANN has the characteristics of improving performance by selecting the features automatically. First of all, the ANOVA test was performed and obtained the best performance of 93.22% accuracy, 92.59% F1-score, 92.59% precision, Kappa Score of 0.8634, AUC of 0.9317 and 92.59% recall for the AdaBoost model as described in Table 4.2. After that, the LDA feature selection technique is applied, and the KNN model provides a good

performance of 79.66% accuracy, 77.77% precision, 77.77% recall, 77.77% F1-score, a Kappa Score of 0.5903 and AUC of 0.7951 as illustrated in Table 4.3. Then PCA is applied for dimensionality reduction, and for the first 12 transformed PCA components, the SVM model outperformed with the best precision of 93.1%, accuracy of 96.61%, recall of 100%, AUC of 96.87%, F1_score of 96.42%, and 0.9321 of Kappa score shown in 4.4. Last but not least, the LVF feature selection technique was performed, and the SVM classifier outperformed with a F1_score of 96.42%, accuracy of 96.16%, precision of 93.1%, Kappa score of 0.9321, recall of 100%, and AUC of 0.9687 described in table 4.5. After analyzing all the results of Table 4.2, Table 4.3, Table4.4 and Table 4.5, the SVM model using the LVF and PCA feature selection method shows same performance when considering performance measurement indicator, but the SVM model using PCA method takes less time (0.00002 (Sec)) for classification than the SVM model using LVF method. Hence, we determined the SVM model using PCA as the best after using the feature selection method. Fig. 4.2 illustrates the performance of the models used in second method.

Table 4.2: Classifier's performance using Analysis of Variance test.

Model	Acc.	Prec.	Recall	F1-score	AUC	Kappa Score	Classification time (mean)
SVM	93.19%	96%	88.88%	92.31%	0.9288	0.8626	0.00004(Sec)
DT	86.44%	88%	81.48%	84.61%	0.8605	0.7252	0.00001(Sec)
LR	84.74%	90.9%	74.07%	81.63%	0.8391	0.6882	**0.000008(Sec)**
RF	91.53%	92.31%	88.88%	90.56%	0.9132	0.8287	0.00003(Sec)
KNN	91.53%	95.83%	95.18%	90.19%	0.9103	0.8277	0.00004(Sec)
XGB	93.22%	92.59%	92.59%	92.59%	0.9317	0.8634	0.00003(Sec)
ADA	**93.22%**	**92.59%**	**92.59%**	**92.59%**	**0.9317**	**0.8634**	0.0003(Sec)

Table 4.3: Classifier's performance using Linear Discriminant Analysis

Model	Acc.	Prec.	Recall	F1-score	AUC	Kappa Score	Classification time (mean)
SVM	79.66%	80%	74.07%	76.92%	0.7922	0.5878	0.00001(Sec)
DT	79.66%	94.11%	59.26%	72.72%	0.7807	0.5781	**0.000005(Sec)**
LR	79.66%	82.61%	70.37%	76%	0.7893	0.5855	**0.000005(Sec)**
RF	76.27%	78.26%	66.66%	72%	0.7552	0.5164	0.0003(Sec)
KNN	**79.66%**	**77.77%**	**77.77%**	**77.77%**	**0.7951**	**0.5903**	0.00004(Sec)
XGB	79.66%	74.07%	76.29%	79.22%	0.7922	0.5879	0.00007(Sec)
ADA	79.66%	66.66%	75%	78.65%	0.7865	0.583	0.000007(Sec)

Table 4.4: Classifier's performance using Principle Components Analysis

Model	Acc.	Prec.	Recall	F1-score	AUC	Kappa Score	Classification time (mean)
SVM	**96.61%**	**93.1%**	**100%**	**96.42%**	**0.9687**	**0.9321**	0.00002(Sec)
DT	91.52%	92.31%	88.88%	90.56%	0.9131	0.8287	0.00001(Sec)
LR	81.35%	78.57%	81.48%	80%	0.8136	0.6255	0.000009(Sec)
RF	91.52%	92.31%	88.88%	90.56%	0.9131	0.8287	0.00014(Sec)
KNN	93.22%	100%	85.18%	92%	0.9259	0.8618	0.0001(Sec)
XGB	91.52%	92.3%	88.88%	90.56%	0.9132	0.8288	0.000032(Sec)
ADA	91.52%	92.31%	88.88%	90.56%	0.9132	0.8287	**0.000001(Sec)**

Table 4.5: Classifier's performance using Low Variance Filter.

Model	Acc.	Prec.	Recall	F1-score	AUC	Kappa Score	Classification time (mean)
SVM	**96.61%**	**93.1%**	**100%**	**96.42%**	**0.9687**	**0.9321**	0.000033(Sec)
DT	89.83%	82%	85.19%	88.46%	0.8946	0.7939	0.000019(Sec)
LR	83.05%	86.95%	74.07%	80%	0.8235	0.6546	**0.0000088(Sec)**
RF	94.91%	96.15%	92.59%	94.33%	0.9473	0.8973	0.00053(Sec)
KNN	91.52%	100%	81.48%	89.79%	0.9074	0.8267	0.000043(Sec)
XGB	95%	96.15%	92.59%	94.33%	0.9473	0.8973	0.00003(Sec)
ADA	95%	92.86%	96.29%	94.54%	0.9502	0.8973	0.000011(Sec)

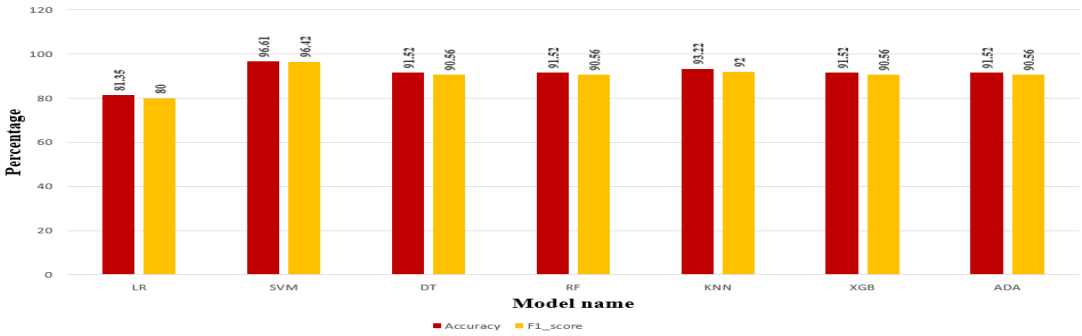

Figure 4.2: Comparison of performance between different classifiers with best feature selection method PCA.

4.5 Selection of Best Model Based On Performance Measurement Indicator

From results of the 2 methods mentioned earlier (without feature selection and using feature selection), we conclude that the ANN model of our first method carries out the best result with an accuracy of 100%. However, some ensemble and traditional ML models perform well with good accuracy, we determined the ANN model as the best model because the improvement of 0.1% is important for classifying any disease. ANN model has a characteristic that it can find out the hidden pattern of the data more effectively. Hence, the ANN model performed better than other models. On the contrary, from the result of our second method shown in Table 4.2 Table 4.3, Table 4.4, and Table 4.5, the SVM using PCA method for feature selection performed better with an accuracy of 96.61%. Finally, it is visible from the result of both methods that the ANN model of the first method is more robust for PD classification using patients' voice features. It also indicates that although all the features are not equally important, an internal relation exists between them. Hence, no features can not be eliminated. Therefore, the incorporation of all the features is important to classify PD. Besides this, the huge hyper-parameter tuning scope of ANN helps ANN discover the underlying pattern of the data. Fig. 4.3 depicts the confusion matrix of the ANN model.

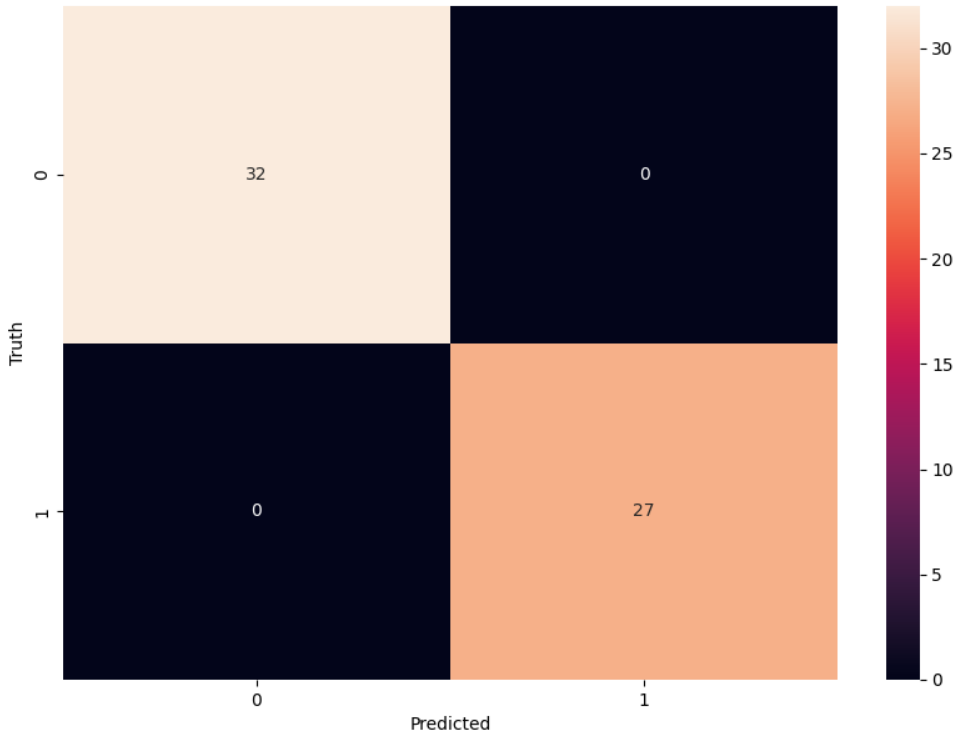

Figure 4.3: Confusion matrix of the ANN model for classifying PD.

4.6 Trade-off Between Classification Time and Accuracy

Trade-off between classification time and accuracy is an important part of the classification task. Because an efficient and accurate classification system is important for real-world disease classification, sometimes both cannot be handled together. Hence, accuracy and classification time trade-off is established here. It is visible from Table 4.1 that the LR model takes 0.0000063 seconds for the classification task with an accuracy of 84.74%. On the contrary, from Table (4.2) - (4.5), the ADA model using the PCA feature selection method takes 0.000001 seconds with 91.52% accuracy for classification, which is less than others. However, the accuracy of the classifiers is more important than the classification time in the case of disease classification because a faster system with low accuracy is considered a low-quality system. Hence, it is determined that the ANN model is the best because of its 100% accuracy described in Table 4.1 (as no external feature selection is needed for ANN).

4.7 Comparison with Other Benchmark Work

The work [Kar20, LTAJ21, MPB19] is conducted on the UCI machine learning dataset [Dat08]. However, they all used traditional machine learning algorithms and some ensemble algorithms. In [Kar20], the SVM model's accuracy of 93.84% is achieved. On the other hand, 95.58% and 91.28% accuracy is achieved, respectively, for [LTAJ21] and [LTAJ21] using RF and the AdaBoost model. We used ANN on the same dataset and some ML and EL algorithms. Besides this, hyper-parameter tuning is used in our work to find the best parameter set of the model. Table 4.6 depicts that, this work's using ANN model performs better than others.

Table 4.6: Comparison between our work and Vocal Dataset based work

References	Best model	Accuracy
[Kar20]	SVM	93.84%
[LTAJ21]	RF	95.58%
[MPB19]	AdaBoost	91.28%
This work	**ANN**	**100**%

Chapter 5

CONCLUSION AND FUTURE SCOPE

Artificial intelligence and data analysis play an important role in the healthcare industry. In this paper, we discussed a method of Parkinson's Disease classification from patients' voice-related data. Till now, no permanent cure has been established, but some medicines and therapeutic surgery are available. However, the existing treatments are an expensive and lengthy process. Hence, early detection of PD may help to lead a better life. Nowadays, different machine learning and deep learning algorithms are used for different disease diagnoses; hence, an advanced system for PD classification can be developed using machine learning and deep learning algorithms. Recent studies on PD says that speech problem is a very common symptom of PD at an early stage. Hence, we investigate the speech data to classify PD.

To achieve better performance, we applied ANN, AdaBoost, XGBoost, RF, DT, KNN, SVM, and LR classifiers on speech data, distinguishing PD patients from healthy cases. Besides this, we employed different feature selection techniques like PCA, LDA, ANNOVA, and LVF to reduce computational complexity. As a result, our suggested ANN model outperformed with an accuracy of 100%, precision of 100%, recall of 100%, 1 of Kappa score, and 1 of AUC for all features. ANN outperformed others because it can find out the underlying pattern of the unobserved data after model training. This result proves that all speech features are important to classify PD as healthy. Hence, feature selection techniques are less important for classifying PD. This scenario also indicates that if we consider some features, then much important information may lost because there is an internal relationship between all features. Although ANN consumes more time in PD classification, the accuracy of ANN is better than others. Hyper-parameter tuning impacts this result because it helps to find the most suitable combination of parameters.

We have a plan to explore brain MRI, EEG, handwritten drawing of PD patients for diagnosing PD etc.

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

PLAGIARISM VERIFICATION

Tittle of the Thesis: Classifying Parkinson's Disease from Patient's Acoustic Features Using Deep Learning

Total Pages: 30 Name of the Scholar: Tarakashar Das

Supervisor: (1) Ms. Shweta Meena

Department of Software Engineering

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: Turnitin Similarity Index: 8% Total Word Count: 7810

I wish to express my sincerest gratitude to Ms. Shweta Meena for his continuous guidance and mentorship that she provided me during the thesis. She showed me the path to achieve my targets by explaining all the tasks to be done and explained to me the importance of this thesis as well as its industrial relevance. She was always ready to help me and clear my doubts regarding any hurdles in this work. Without her constant support and motivation, this thesis would not have been successful.

Date:

Candidates Signature **Signature of Supervisor**

Bibliography

[AAA+22] Majid Aljalal, Saeed A. Aldosari, Khalil AlSharabi, Akram M. Abdurraqeeb, and Fahd A. Alturki. Parkinson & rsquo;s disease detection from resting-state eeg signals using common spatial pattern, entropy, and machine learning techniques. *Diagnostics*, 12(5), 2022.

[AAJH+18] Ahmed J. Aljaaf, Dhiya Al-Jumeily, Hussein M. Haglan, Mohamed Alloghani, Thar Baker, Abir J. Hussain, and Jamila Mustafina. Early prediction of chronic kidney disease using machine learning supported by predictive analytics. In *2018 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–9, 2018.

[AAM21] Nour Abdulhadi and Amjed Al-Mousa. Diabetes detection using machine learning classification methods. In *2021 International Conference on Information Technology (ICIT)*, pages 350–354, 2021.

[ad] *Support Vector Machine Algorithm*.

[AIPV19] Maria Teresa Angelillo, Donato Impedovo, Giuseppe Pirlo, and Gennaro Vessio. Performance-driven handwriting task selection for parkinson's disease classification. In *International Conference of the Italian Association for Artificial Intelligence*, pages 281–293. Springer, 2019.

[Ano] *Analysis of Variance (ANOVA) Explanation, Formula, and Applications*.

[ARFC+19] Jefferson S Almeida, Pedro P Rebouças Filho, Tiago Carneiro, Wei Wei, Robertas Damaševičius, Rytis Maskeliūnas, 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.

[AZZL19] Liaqat Ali, Ce Zhu, Zhonghao Zhang, and Yipeng Liu. Automated detection of parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network. *IEEE Journal of Translational Engineering in Health and Medicine*, 7:1–10, 2019.

[CAH+20] Sabyasachi Chakraborty, Satyabrata Aich, Eunyoung Han, Jinse Park, Hee-Cheol Kim, et al. Parkinson's disease detection from spiral and

wave drawings using convolutional neural networks: A multistage classifier approach. In *2020 22nd International Conference on Advanced Communication Technology (ICACT)*, pages 298–303. IEEE, 2020.

[CLI] MAYO CLINIC. *Parkinson's disease*.

[CTCA22] Jyotismita Chaki, S. Thillai Ganesh, S.K Cidham, and S. Ananda Theertan. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: A systematic review. *Journal of King Saud University - Computer and Information Sciences*, 34(6, Part B):3204–3225, 2022.

[Dat08] Parkinsons data set, 2008.

[DMH+23] Tarakashar Das, Sabrina Mobassirin, Minhaz Hossain, Syed Md, Aka Das, Anik Sen, Khaleque Md. Aashiq Kamal, and Kaushik Deb. Patient questionnaires based parkinson's disease classification using artificial neural network. *Annals of Data Science*, 2023.

[dOdSA+20] Ana Paula S. de Oliveira, Maíra Araújo de Santana, Maria Karoline S. Andrade, Juliana Carneiro Gomes, Marcelo C. A. Rodrigues, and Wellington P. dos Santos. Early diagnosis of parkinson's disease using eeg, machine learning and partial directed coherence. *Research on Biomedical Engineering*, 36, 2020.

[EEG17] Unm dataset, 2017.

[EEG20] Sandiego dataset, 2020.

[Gee] *What is Standardization in Machine Learning*.

[Gee18] *Introduction to Dimensionality Reduction*, 2018.

[GKA21] Jinee Goyal, Padmavati Khandnor, and Trilok Chand Aseri. A comparative analysis of machine learning classifiers for dysphonia-based classification of parkinson's disease. *International Journal of Data Science and Analytics*, 11(1):69–83, 2021.

[Gov48] UK Government. *Causes Parkinson's disease*, 1948.

[GRK18] Swapna G., Vinayakumar R., and Soman K.P. Diabetes detection using deep learning algorithms. *ICT Express*, 4(4):243–246, 2018.

[jpo20] *Artificial Neural Network Tutorial*, 2020.

[Kar20] Zehra Karapinar Senturk. Early diagnosis of parkinson's disease using machine learning algorithms. *Medical Hypotheses*, 138:109603, 2020.

[KCÖ18] Pedram Khatamino, İsmail Cantürk, and Lale Özyılmaz. A deep learning-cnn based system for medical diagnosis: an application on parkinson's disease handwriting drawings. In *2018 6th International Conference on Control Engineering & Information Technology (CEIT)*, pages 1–6. IEEE, 2018.

[KRK+21] Likitha KN, Nethravathi R, Nithyashree K, Ritika Kumari, Sridhar N, and Venkateswaran K. Heart disease detection using machine learning technique. In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pages 1738–1743, 2021.

[KSB+21] L.V. Rajani Kumari, P. Shreya, Mehrunnisa Begum, T. Pavan Krishna, and M. Prathibha. Machine learning based diabetes detection. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pages 1–5, 2021.

[Kur] Shruti SureshanVihar Kurama. *Introduction to AdaBoost for Absolute Beginners*.

[KZ01] Gary King and Langche Zeng. Logistic regression in rare events data. *Political analysis*, 9(2):137–163, 2001.

[LD] *Linear Discriminant Analysis (LDA) in Machine Learning*.

[LL61] MultiMedia LL. *Parkinsons Foundation*, 1961.

[LLC61] MultiMedia LLC. *Parkinsons Foundation*, 1961.

[LTAJ21] Rohit Lamba, Gulati Tarun, Hadeel Fahad Alharbi, and Anurag Jain. A hybrid system for parkinson's disease diagnosis using machine learning techniques. *International Journal of Speech Technology*, 2021.

[LVF] *The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes)*.

[MJL+11] Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

[MPB19] Richa Mathur, Vibhakar Pathak, and Devesh Bandil. Parkinson disease prediction using machine learning algorithm. pages 357–363, 2019.

[NA21] G Nandhini and J Aravinth. Chronic kidney disease prediction using machine learning techniques. In *2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)*, pages 227–232, 2021.

[NNT+21] Aleksei Netšunajev, Sven Nõmm, Aaro Toomela, Kadri Medijainen, and Pille Taba. Parkinson's disease diagnostics based on the analysis of digital sentence writing test. *Vietnam Journal of Computer Science*, 8(04):493–512, 2021.

[OHR+20] Shu Lih Oh, Yuki Hagiwara, U. Raghavendra, Rajamanickam Yuvaraj, N. Arunkumar, M. Murugappan, and U. Rajendra Acharya. A deep learning approach for parkinson's disease diagnosis from eeg signals. *Neural Computing and Applications*, 32:389–401, 2020.

[Org] World Health Organization. *Parkinson's disease*.

[PCA] *Principal Component Analysis(PCA)*.

[PDCSM19] Antonio Parziale, Antonio Della Cioppa, Rosa Senatore, and Angelo Marcelli. *A Decision Tree for Automatic Diagnosis of Parkinson's Disease from Offline Drawing Samples: Experiments and Findings*, pages 196–206. 09 2019.

[PN20] Kemal Polat and Majid Nour. Parkinson disease classification using one against all based data sampling with the acoustic features from the speech signals. *Medical Hypotheses*, 140:109678, 2020.

[PPR+18] Clayton R Pereira, Danilo R Pereira, Gustavo H Rosa, Victor HC Albuquerque, Silke AT Weber, Christian Hook, and João P Papa. Handwritten dynamics assessment through convolutional neural networks: An application to parkinson's disease identification. *Artificial intelligence in medicine*, 87:67–77, 2018.

[PR18] R Prashanth and Sumantra Dutta Roy. Early detection of parkinson's disease through patient questionnaire and predictive modelling. *International journal of medical informatics*, 119:75–87, 2018.

[RARS21] M. Snehith Raja, M. Anurag, Ch. Prachetan Reddy, and Nageswara Rao Sirisala. Machine learning based heart disease prediction system. In *2021 International Conference on Computer Communication and Informatics (ICCCI)*, pages 1–5, 2021.

[RS] *Hyperparameter Tuning Using Randomized Search*.

[Sai] Anshul Saini. *An Introduction to Random Forest Algorithm for beginners*.

[Sha] Shivam Sharma. *K-Nearest Neighbour: The Distance-Based Machine Learning Algorithm.*

[SIS+13] Betul Erdogdu Sakar, M Erdem Isenkul, C Okan Sakar, Ahmet Sertbas, Fikret Gurgen, Sakir Delil, Hulya Apaydin, and Olcay Kursun. Collection and analysis of a parkinson speech dataset with multiple types of sound recordings. *IEEE Journal of Biomedical and Health Informatics*, 17(4):828–834, 2013.

[SSG+19] C Okan Sakar, Gorkem Serbes, Aysegul Gunduz, Hunkar C Tunc, Hatice Nizam, Betul Erdogdu Sakar, Melih Tutuncu, Tarkan Aydin, M Erdem Isenkul, and Hulya Apaydin. A comparative analysis of speech signal processing algorithms for parkinson's disease classification and the use of the tunable q-factor wavelet transform. *Applied Soft Computing*, 74:255–263, 2019.

[SSR+21] Humza Sami, Mahnoor Sagheer, Kashif Riaz, Muhammad Qasim Mehmood, and Muhammad Zubair. Machine learning-based approaches for breast cancer detection in microwave imaging. In *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*, pages 72–73, 2021.

[tpe22] *Artificial Neural Network (ANN)*, 2022.

[UCI09] Parkinsons telemonitoring data set, 2009.

[UCI19] Parkinson dataset with replicated acoustic features data set, 2019.

[WLW+19] Zhiqiong Wang, Mo Li, Huaxia Wang, Hanyu Jiang, Yudong Yao, Hao Zhang, and Junchang Xin. Breast cancer detection using extreme learning machine based on feature fusion with cnn deep features. *IEEE Access*, 7:105146–105158, 2019.

[WLZG18] Shaohua Wan, Yan Liang, Yin Zhang, and Mohsen Guizani. Deep multi-layer perceptron classifier for behavior analysis to estimate parkinson's disease severity using smartphones. *IEEE Access*, 6:36825–36833, 2018.

[xgb] *Introduction to XGBoost Algorithm in Machine Learning*.

[XP20] Shoujiang Xu and Zhigeng Pan. A novel ensemble of random forest for assisting diagnosis of parkinson's disease on small handwritten dynamics dataset. *International Journal of Medical Informatics*, 144:104283, 2020.