Parkinson's Disease Detection Analysis through Machine Learning Approaches

BY

Annesha Acharjee
ID: 181-15-10718

Md Tanvir Siddiki Rume
ID: 152-15-6090

Arafat Jahan
ID: 181-15-10653

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

Supervised By

Moushumi Zaman Bonny
Assistant Professor
Department of CSE
Daffodil International University

Co Supervised By

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

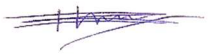
DHAKA, BANGLADESH

JANUARY 2022

APPROVAL

This project titled "**Parkinson's Disease Detection Analysis through Machine Learning Approaches"**, submitted by Annesha Acharjee, Md Tanvir Siddiki Rume and Arafat Jahan to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 02-01-2022.

BOARD OF EXAMINERS

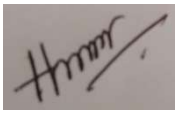


Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Moushumi Zaman Bonny
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Md. Mahfujur Rahman
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Md Arshad Ali
Associate Professor
Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Moushumi Zaman Bonny, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Moushumi Zaman
Bonny

Assistant Professor
Department of CSE
Daffodil International University

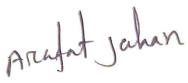
Co-Supervised by:

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by

Annesha Acharjee
ID: 181-15-10718
Department of CSE
Daffodil International University

Md Tanvir Siddiki Rume
ID: 152-15-6090
Department of CSE
Daffodil International University



Arafat Jahan
ID: 181-15-10653
Department of CSE
Daffodil International University

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty God for His divine blessing that makes it possible to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Moushumi Zaman Bonny**, **Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of machine learning, data mining to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice,reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Touhid Bhuiyan,** Head**,** Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

Machine learning and data mining are crucial in health care, as well as medical information and detection. Machine learning approaches are now being utilized to improve awareness of a variety of critical health issues, including diabetes detection, neuron cell tumor diagnosis, COVID 19 identification, and so on. Parkinson's disease is basically a diseases for our senior citizen in Bangladesh. Parkinson's Disease indications often seem progressively and got worst with time. People got effected have trouble walking and communicating with the condition advances. Patients can also psychological and social vagaries, nap problems, hopelessness, reminiscence loss, and weariness. Parkinson's disease can happen in both men and women. Though, men are affected by the illness at a proportion that is around partial of them are women.

In this regarding research we have to get out the accurate ML algorithm to find out the disease with predictable dataset and the model of the following machine learning classifiers.

Therefore, nine ML classifier are secondhand to portion study to use machine learning approaches like as follows, Naive Bayes, Adaptive Boosting, Bagging Classifier, Decision Tree Classifier, Random Forest classifier, XBG Classifier, K Nearest Neighbor Classifier, Support Vector Machine Classifier and Gradient Boosting Classifier are used.

TABLE OF CONTENTS

TABLE LIST

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Machine learning is a popular area of study these days. Various statistics and machine learning techniques are used in various fields. Machine learning may be used in domains such as promotion, health and medical problems, weather prediction and so on. Many disorders in the medical industry may be diagnosed or predicted by machines using machine learning techniques. Chronic illnesses are a major danger to the worldwide health program of the twenty-first century. Parkinson's Disease is a large disease to face Chronic problems that is happening all over the world including Bangladesh as well.

The increased frequency of chronic illnesses, such as Parkinson's Disease has major ramifications for developing country health and economic productivity. The fast rise in mutual hazard factors, particularly among the deprived people, like old diabetes, high blood pressure, hypotension and overweightness, will outcome in even higher and deeper costs that emerging countries are ill-equipped to bear. There is sort of experience to long-lasting illnesses, particularly Parkinson's Disease, owing to the global health group's concentration viral disease a deficiency of awareness. There is an urgent need to focus on it, as well as implementation on comprehensive, active, and preventative long-lasting illness management plans.

There are many clinics which store the dataset of the patients of Parkinson's Diseases patients. Complete refactoring dataset, that will be helpful for taking a decision to predict the disease earlier before it happens. Using machine learning algorithms methods on these data, it is possible to learn many kinds of information and use this information to envisage the sickness. People od Bangladesh much of senior citizens of our country who has been sorrow from Parkinson's disease. In Bangladesh most of the people are not anxious about the illness at all. So, the pretentious people by Parkinson's disease are cumulative rapidly.

Detecting the disease in early stage could be great help for the affected people. As follows, People have to take the essential steps.

Be unaffected Predictive may be used to forecast the illness. Classification, regression, and categorization are some of the approaches that may be employed. By machine learning algorithm we can predict the disease before it has happened with the help of dataset.

We have used In this regarding research we have to get the accurate machine learning approaches to find out the disease with predictable dataset and the model of the following machine learning classifiers.

Therefore, nine machine learning algorithms are work with machine learning tools namely: Naive Bayes, Adaptive Boosting, Bagging Classifier, Decision Tree Classifier, Random Forest classifier, XBG Classifier, K Nearest Neighbor Classifier, Support Vector Classifier and Gradient Boosting Classifier are used.

The primary aims of the research to detected the best accurate result , to warn if the person is at aware to the illness, and to compare the results of multiple machine learning approaches to get the best approach.

1.1 Background

1.1.1 How Parkinson’s Disease infected

Parkinson’s disease is a human brain disorder usually occurs in the brain nerve cell in the area of the brain’s movement control part. The brain cells or neurons could die due to dis Parkinson’s Disease. Neurologists looking forward on researching on this disease for a decade. basically a neurological disorder that disable a person’s stability of holding stuffs.

Parkinson's indications often seem gradually and got worst with period. The anxiety of walking and communicating to the condition advances. They have mental and change in regular life, depression, memory problems, and weariness. Parkinson's disease can distress both males and women. Although the majority of individuals with Parkinson's acquire the illness around the age of 60. About 6 to 12% of persons with Parkinson's Disease have early-onset it happens before 51. The deadly diseases frequently, but not always, inherited, and certain varieties have a genetic component.

Rather than genetic disfunction affected to the aged people. The age limit can be 50 to 60 to get affected by this disease.

1.1.2 Symptoms of Parkinson's Disease

Parkinson's disease grows when neuron in basal ganglia, a area of the brain which governs drive, get damaged or die. Usually, the nerve cells, or neurons, create dopamine, a key brain neuro object. If neuron is hampered produce less dopamine, subsequent in Parkinson's with the walking or focus issues. Doctors still unsure of why the cells are dying frequently.

Symptoms can be seen in Parkinson's Disease effected people be like,

- Stiffness of the limbs and trunk Remorse in the hands, arms, legs, jaw, or head
- Movement is slow.
- Impaired balance and coordination which can lead to falls.
- Depression
- Mood swings
- Difficulties in eating, chewing, and speaking
- Bladder issues or constipation
- Skin problems
- Sleep disturbances.

1.1.3 Diagnosis of Parkinson’s Disease

For a huge variety of illnesses can produce indications comparable to the disease. Parkinson’s Disease is a word used to describe persons who exhibit Parkinson's-like symptoms but are not suffering from it. While many illnesses may first be misdiagnosed as Parkinson's, particular medical testing as well as response to pharmaceutical therapy may help to distinguish them, getting a good result is crucial. There are currently no nongenetic blood or laboratory tests available to diagnose Parkinson's disease. To make a diagnosis a person’s whole life history are needed.

Nongenetic blood tests available to diagnose Parkinson's disease. A diagnosis is made based on a person's medical history and a neurological examination. Another distinguishing trait of the disease is improvement following treatment.

1.1.4 Parkinson's Disease Test

There is no suggested or conclusive lab or imaging test to test Parkinson's disease. Though, the authority thatt authorized the DaTscan imaging scan in 2011. Doctors may now obtain precise images of the brain's dopamine system using this approach. A DaTscan includes the administration of a tiny quantity of a radioactive medication and the use of a machine known as a single-photon emission computed tomography scanner, which is comparable to an MRI. The chemical attaches to dopamine transmitters in the brain, revealing the location of dopaminergic neurons. Dopaminergic neurons are the brain's source of dopamine; dopamine deficiency causes Parkinson's disease. The findings of a DaTscan cannot demonstrate this disease at time.

1.2 Research Motivation

The disease is a human mental sickness in the brain usually occurs in brain nerve cell in the area of the brain's movement control part. The brain cells or neurons could die due to dis Parkinson's Disease. Neurologists has been researching on this disease for a decade.

Our Motivation is to detect the disease as early as possible with machine learning algorithms. We have arranged serval machine learning algorithm model to calculate the best accurate algorithm to get the most accurate algorithm for detecting Parkinson's Disease in the earliest stage as possible.

1.3 Problem Description

The dataset is used here collected from UCI machine learning library dataset for Parkinson's Disease. Choosing the optimal predictive model in terms of accuracy, precision, recall, f-measure, and learning rate. Models created by employing several classifiers for machine learning and assessing their results Preparing and accurate enough.

1.4 Research Objective

- The research's key goals that anticipate the illness using a machine learning technique, to warn if a being is at danger on the disease, and to compare the results of differnet machine learning algorithms to decide which ML technique delivers the greatest presentation.
- To use the dataset to consider train the data and get the prediction result.
- To get a better statistical machine-learning algorithm to get the best accurate result with percentage.

1.5 Research Organization

The paper structure has given as given:

- The related works are defined in chapter 2
- The methodology is described in chapter 3.
- The findings are discussed in chapter 4.
- The conclusion and future work are discussed in chapter 5.

CHAPTER 2
RELATED WORK

Pahuja et. al. has said in their work that According to their research, 90 % of persons with Parkinson's disease have voice impairment. Numerous experiments have been conducted to automate the diagnosis of Parkinson's disease using speech datasets. On a speech database, the True positive, true negative, total organization accuracy, and geometric of ANN, KNN, and SVM classifiers. A similar argument is taking place over the Wisconsin Breast Cancer Database and the Pima Indian Diabetes Dataset. Artificial Neural Networks with the Levenberg Marquardt equation have been discovered. algorithm have the maximum percent for speech datasets. With machine learning it will be much more helpful to the doctors around the world [1].

Ali et. al. has presented in their paper that In their study, we looked into the challenge of detecting PD using stored dataset. The data under examination was extremely unbalanced. We empirically proved the bias in machine learning models due by unbalanced data. n They perform poorly with data because the approaches toward the dominant class in the data. As a consequence, we discovered a high rate of sensitivity but a low rate of precision for the PD detector since the ill group was in the plurality and the normal class would be in the minority. We used the random under sampling strategy to reduce the bias in the built models. Following the optimization, it has been a great work for machine learning work [2].

Prashanth et. al. has described in their paper that Conclusion Early detection of Parkinson's disease is critical so that neuroprotective and early management techniques can be implemented. In this study, we used preclinical indicators of non-motor To distinguish between early PD and healthy normal, imaging procedure has been used here were used. We noticed that using SVM was beneficial. resulted in an almost flawless classification. The following are the study's contributions: We extend Kang in which they employed CSF measures and found a limited diagnostic usefulness. We employ additional relevant and significant aspects pertaining to non-motor in our work [3].

Senturk et. al. has presented in their paper, an FS scheme to the early identification The Parkinson's disease model was created by analyzing on physically fit people. P6rimary aim was to improve the model's performance and accuracy while minimizing the compute cost of the classification task. The results of the approaches were evaluated with and without FS, and the amazing effect of FS was proven. The findings integrating FS techniques with classification methods is quite effective. Especially once dealing with speech signs. If the process can diagnosis the diseases earlier by the physician that it can be cured [4].

Parisi et. al. has described in their work that, thorough comparison of the proposed hybrid model's classification performance was found to be the most accurate, in contrast towards the MLP, the LSVM, the SVM-MLP applied in available commercially apps, and the ML-based classification methods from picked published studies. As just a consequence, the in the early diagnosis of Parkinson's disease patients in a clinical setting, improving the accuracy and reliability of existing automated speech recognition systems. processing systems. in discrimination. The machine learning classifiers will help the diagnosis [5].

Mandal et. al. has presented in their paper that; our key answers are classifier is used with to improve the classification of classifiers by predictive analysis We proved experimentally the proper ML techniques and methodologies that might be applied to improve the existing state-of-the-art of PD treatment method for patients. This collection's inferential analysis from experimental results is a watershed point in the treatment of Arthritis. and others have done considerable work on the PD categorization. These investigations' greatest reported accuracies are 91.4 percent, 92.9 percent, and 96.93 percent, respectively. This research looked at the study. They have used serval machine learning approaches to get more efficient result [6].

Wang et. al. has showed in their paper that, Early identification of Parkinson's disease is critical for gaining a better knowledge of the illness's etiology, initiating therapeutic measures, and finding effective therapies. Based on premotor parameters like presented a deep learning algorithm to automatically identify between normal persons and PD patients. The suggested deep learning model demonstrated strong detection capability, with an accuracy of 96.45 percent. This is mostly owing to the deep learning model's favorable capabilities in learning linear and nonlinear features from PD data without the requirement for hand-crafted feature extraction. The outcomes demonstrate that deep learning can do a lot of stuffs and more efficiently results shows that the detection has a superior study to for predicting the diseases with the deep learning studies [7].

Salvatore et. al. has described in their study that, The effective and reliable test of MRI markers to diagnose Parkinson's disease has recently gained a lot of interest. So far, many biomarkers have been shown to be more sensitive to Parkinson's test of diseases then PSP. Morphological abnormalities in the brainstem and cerebellar peduncles, for example, have been demonstrated to be good clinical markers. Conventional MRI with manual morphometric measurement, on the other hand, is the only proven MRI-based assessment used in clinical practice in Parkinson's disease [8].

Wroge et. al. has presented in their study that, Disease diagnosis and prediction are now achievable utilizing noninvasive speech biomarkers as features in automated machine learning frameworks. Our study compares the efficacy of several machine learning classifiers in illness detection with noisy and high-dimensional data. Clinical level accuracy is feasible with careful feature selection. These findings are interesting because they might pave the way for new ways to monitor patient health and neurological illnesses using speech data. Because of the models' remarkable accuracy with these brief audio snippets, there is reason to assume that denser feature sets comprising spoken word, video, and graphics will be developed in the future [9].

Challa et. al. has described in their paper that, the analysis of Parkinson's Disease not as forthright as everyone thinks, which means that a single test, such as a blood test or an ECG, cannot identify whether or not a person has the disease. Doctors review a patient's medical history before doing a complete neurological examination. They identify at least two cardinal symptoms in the participants and then subject has the disease. Because this conclusive test for Parkinson's disease, there is a high likelihood of misdiagnosis. Such a circumstance, we may support the doctor by offering a machine learning model. ML techniques such as boosted classification trees, and Bayes Ne are used to create the prediction models. Based on these important characteristics, Bayes Net and multilayer perceptron were developed. It has been seen that the performance has improved. It is shown that Boosted Logistic Regression produces better results. These findings urge us to experiment with alternative ensemble learning strategies. The current study incorporates many machine learning methods that are not employed in. This research is critical in doing a comparative comparison of various machine learning algorithms [10].

Nilashi et. al. has showed in their study that, using machine learning approaches, a hybrid solution for This article established. ISVR was used to predict Clustering and data dimensionality reduction were accomplished using SOM and NIPALS, respectively. Several investigational evaluations were performed obtained from UCI to test the suggested technique. The results showed that combining SOM, NIPALS, and ISVR methods is successful in predicting. The findings also showed that the suggested technique may greatly minimize calculation time. The approach the approach described in this paper was tested using UCI public datasets with Parkinson's disease diagnosis. Furthermore, in comparison. Furthermore, in the case of large healthcare data, which might be complicated datasets with unique features, future research must examine this problem in the creation of new approaches so to rid of these challenges and take use of big data. Additionally, because large hospital data contains spectral, assorted data [11].

Almeida et. al. has presented in their study that they developed a method for detecting the disease utilizing consolidated feature set extraction and ML approaches. On a speech dataset, we used some approaches were used, and the results were examined in terms of ERR and Accuracy. Statistical analysis of ERR data using N-way ANOVA and the Nemenyi test found no statistically significant validating the hypothesis that PD may be detected using a smart-phone. Thesis and Unvoiced portion audio produced much better outcomes than Phonation (P) audio. The most effective individual feature set and classifier of machine learning and data mining instead [12].

Lahmiri et. al. has described in their work that use a number of ML algorithms to identify between healthy people and Parkinson's disease Dyphonia values were used to classify patients. The findings of linear discriminant analysis (LDA), k nearest neighbors (k-NN), nave Bayes (NB), logistic regression (RT), radial basis function networks (RBFNN), support vector machine (SVM), and Distance measure classifier were evaluated (MDC). When compared to another classifiers, the SVM classification classifier performs the best.er outperforms them all. We improved the RBFNN and SVM parameters for our investigation. Future research might examine feasibility to increase the presentation of the SVM and other machine learning approches [13].

Moshkova et. al. has presented in their study that, In this study, the PD and control groups' hand movement activity was recorded. A 3D Leap Motion sensor was used to collect data. Plots showing motion amplitudes as a function of frame number were created.. 25 kinematic characteristics were generated for each movement signal based on three essential motion parameters: speed, amplitude, and frequency. For the training of four classifiers, several feature vector combinations were employed. The findings the identification of Parkinson's disease was acquired. [14].

Rovini et al. has showed in their study that, the goal of this study was to see if acquisitions performed using sensors worn on the feet while doing for important motor parameters that could be used to get the best of persons with high result. The motion data were collected using a sensor individuals completed an experimental procedure consisting of, which is often used to assess Parkinson's disease patients during neurological exams. Data was collected and sent to a PC, where it was filtered and analyzed using suitable algorithms to produce 23 attributes per limb that might reflect patients' characteristics [15].

CHAPTER 3
METHODOLOGY

Here has been described the methodology part with different parts. That as follow,

3.1 Data Description

In this paper, we use 195 Parkinson's disease datasets took at the UCI machine learning repository. While working with following dataset, there are 24 attributes. After training the data we will able to get the best accurate machine learning approaches. The used attributes description has given below,

1. Name: The name of the patients.
2. MDVP:Fo(Hz) : MDVP-Fo of the patient.
3. MDVP:Fhi(Hz) : MDVP:Fhi of the patient
4. MDVP:Flo(Hz) : MDVP:Flo of the patient.
5. MDVP:Jitter(%) : MDVP:Jitter of the patient.
6. MDVP:Jitter(Abs) : MDVP:Jitter of the patient.
7. MDVP:RAP : MDVP:RAP of the patient.
8. MDVP:PPQ : MDVP:PPQ of the patient.
9. Jitter:DDP : Jitter:DDP of the patient.
10. MDVP:Shimmer : MDVP:Shimmer of the patient.
11. MDVP:Shimmer(dB) : MDVP:Shimmer of the patient.
12. Shimmer:APQ3 : Shimmer:APQ3 of the patient.
13. Shimmer:APQ5 : Shimmer:APQ5 of the patient.
14. MDVP:APQ : MDVP:APQ of the patient.
15. Shimmer:DDA : Shimmer:DDA of the patient.

16. NHR : NHR of the patient.
17. HNR : HNR of the paitent.
18. Status : Status of the patient.
19. RPDE : RPDE of the patient.
20. DFA : DFA of the patient.
21. Spread1: Spread1 of the patient.
22. Spread2 : Spread2 of the patient.
23. D2 : D2 of the patient.
24. PPE: PPE of the patient.

3.2 Algorithm description

We have used different machine learning approaches for this research. The description has given below,

3.2.1 Naive Bayes Classifier:

Naive Bayes classifiers are a collection of Bayes' Theorem-based classification techniques. This is not a single algorithm, but an algorithm family in which all of them share a common notion.
It is used to forecast continuous data such as illness rate efficiency. The bayes theorem is a efficient way to predict Parkinson's Disease from given datasets.

3.2.2 Adaboost Classifier:

The Adaptive Boosting classifier is referred to as AdaBoost. Using this strategy, a group of weak classifiers is combined to form a powerful one. The equation describes the use of the AdaBoost method for classification in this case.

It used to predict the best way to survive the best figure. By using it the dataset is more frequent to train.

3.2.3 Bagging Classifier:

The bagging classifier is an ensembled meta-estimator. With each base classifier, it fits random subsets of the original dataset. The forecasts are then aggregated to generate a final prediction. It is often used to minimize the variance of a black-box estimator.
After training, predictions for unknown samples x' may be created by integrating the predictions on x' from all of the various regression trees.

3.2.4 Decision Tree Classifier:

Decision Tree categorization employs tree-like structures. The criteria are represented by the root nodes, and the class label is represented by the child nodes. The consequences of the circumstances are represented by the branches of the root nodes.

Here Decision tree classifier has played an important role to predict the Parkinson's Disease.

3.2.5 Random Forest Classifier:

It is an ensemble tree-based learning method. The Random Forest Classifier is a collection of decision trees chosen at random from a subset of the training data. It uses votes from the several decision trees to determine the final class.

When employing the Random Forest Algorithm to solve regression issues, the mean squared error (MSE) is used.

3.2.6 XGB Classifier:

Because all input variables are numeric and the goal is a binary classification issue, the XGBoost Classifier model provides an ideal data set. It is an example of Gradient Boosting in action.

While developing a prediction model, the classifier collaborates with artificial neural networks and decision tree-based algorithms.

3.2.7 K Nearest Neighbor (KNN):

The KNN technique is a sort of supervised machine learning approach to prediction of classification and regression. The KNN classifier is a powerful method for evaluating any prediction. Here's the proof:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$$

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

3.2.8 Gradient Boosting Classifier: Gradient enhancement is a greedy technique that can swiftly override a training data. Gradient Boosting algorithm can be used as not only continuous prediction model but also used for a categorical prediction model.

Here an example of Parkinson's Disease effected patients,

dizziness	anxiety	nausea	Age
False	True	False	59
True	False	True	65
Ture	True	False	70

Gradient Boosting Classifier Table

3.2.9 Support Vector Algorithm:

The "Support Vector Machine" is a ML algorithm to solve particular problem itself. . however, mostly used in randomized problems. Per aged data can recognized different points for the value that could be featured from SVM for the result of a given position.

3.2 Proposed Model

This section illustrates the execution processes. The research was carried out using Python and the Scikit-learn module.

3.2.1 Input Data

In this research total 195 patient data collected to predict an analysis and find out performance. This dataset collected from UCI machine learning repository

3.2.2 Data Prepossessing

In this collection, a total of 24 attributes are utilized, 24 of which are results and findings and one of which is a response variable. As a result, use a mapping to transform nominal properties to quantitative variable. The nominal value in this information is function. In this dataset, the insignificant value contains name MDVP:Fo(Hz) MDVP:Fhi(Hz) MDVP:Flo(Hz) MDVP:Jitter(%) MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP MDVP:Shimmer MDVP:Shimmer(dB) Shimmer:APQ3 Shimmer:APQ5 MDVP:APQ Shimmer:DDA NHR HNR Status RPDE DFA spread1 spread2 D2 PPEthis are the attribute Using a mapping function, we turn this to a numeric number.

And now our data collection is brimming with numerical values. Furthermore, the dataset is using for 80 percent in training and 20 percent in testing method.

3.2.3 Feature Extraction

In the following section we have regenerated the correlation that has shown in metrics are shown using target attribute to feature attribute.

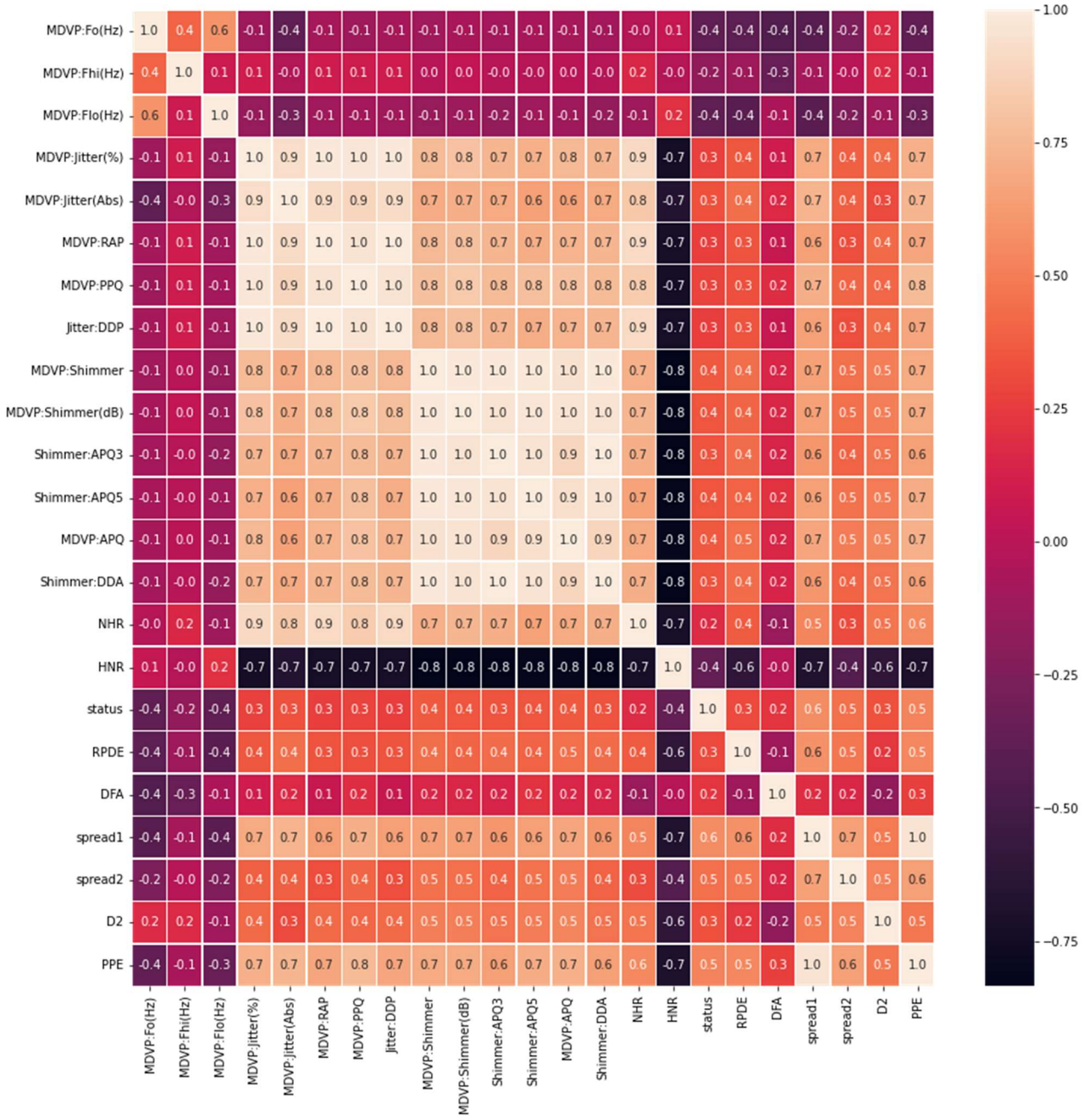


Figure 1: correlation metrics are shown using target attribute to feature attribute

3.2.4 Machine Learning Classifier

We have used In this regarding research we have to calculate the most accurate machine learning algorithm to find out the disease with predictable dataset and the model of the following machine learning classifiers.

We have used In this regarding research we have to get the accurate machine learning approaches to find out the disease with predictable dataset and the model of the following machine learning classifiers.

Therefore, nine machine learning classifier are used to measure analysis using machine learning tools namely: Naive Bayes, Adaptive Boosting, Bagging Classifier, Decision Tree Classifier, Random Forest classifier, XBG Classifier, K Nearest Neighbor Classifier, Support Vector Classifier and Gradient Boosting Classifier are used.

.

3.2.5 Cross-Validation

Cross validation is a approach for verifying ML algorithm that entails training multiple ML algorithms methods on accessible input data sets and evaluating them on the comparable data. Cross-validation, such as filling to generalize a sequence, can be used to detect overfitting. For a specific predictive modeling issue is a common approach in applied ML since it is easy to understand and use. and produces ability estimates with lower bias than other techniques. Because our data set is limited.

K fold is using for validation here with sub k.

The holdout solution is now available with the test as one of the k subsets and the remaining k-1 subsets forming a training set. each time. For maximal model efficacy, the error estimate is aggregated across all k trials.

3.2.6 Evaluation Metrics

The evaluation metrices is to compute the study's outcomes, we must use precision whereas TP indicates True Positive, FP means False Positive, TN is True Negative, and FN defines False Negative.

Precision: Predictive Precision or
The generates of predictive values as follows,
TP / TP + FP,

Recall: The Recall values generates,
TP / TP + FN

F **Measure**: The F-1 score values generates follows,
2 * TP / 2 * TP + FP + FN

Accuracy: The accuracy values generate follows,
TP + TN / TP + TN + FP + FN

3.2.7 Best Model

The precision of the given dataset is also assessed in order to calculate the accuracy of different sorts of algorithms., with the most accurate attained to utilizing a (KNN) K Nearest Neighbor Classifier.

This machine learning algorithm provides the best accurate result after training the given dataset.

3.2.8 Confusion Metrics

This machine learning algorithms will be produced model for predicting the disease. We will find the best possible accuracy among them.

A.TP stands for true positive. That is, we expected the TP FP value will make perfect prediction for the model.

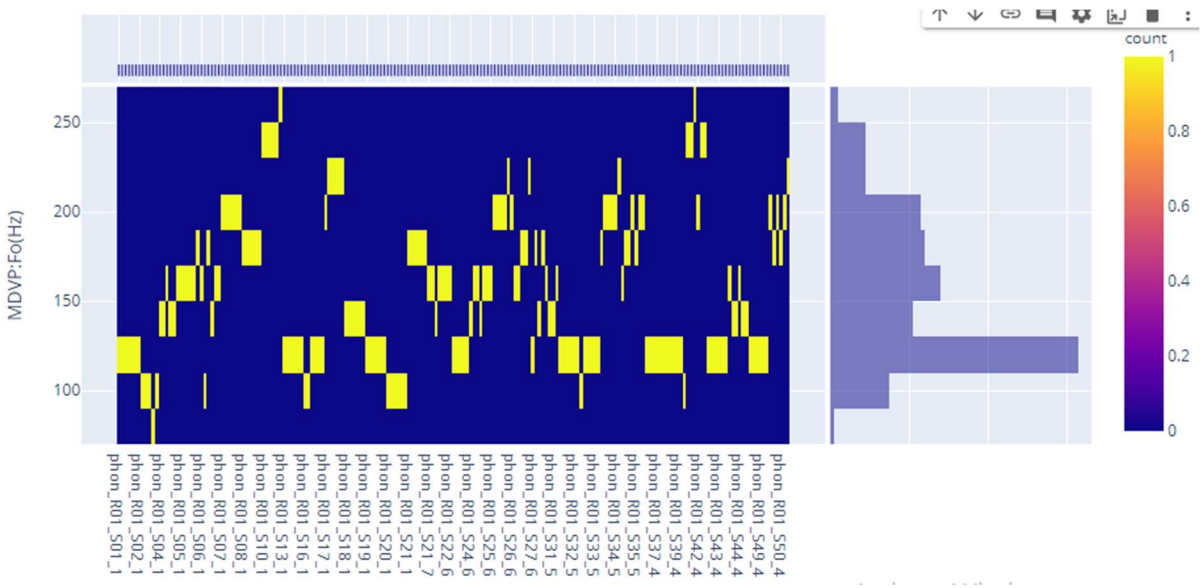
B. FP stands for incorrect. we expected that when a label was incorrect.

C. FN denotes false negatives. If there is no predicted.

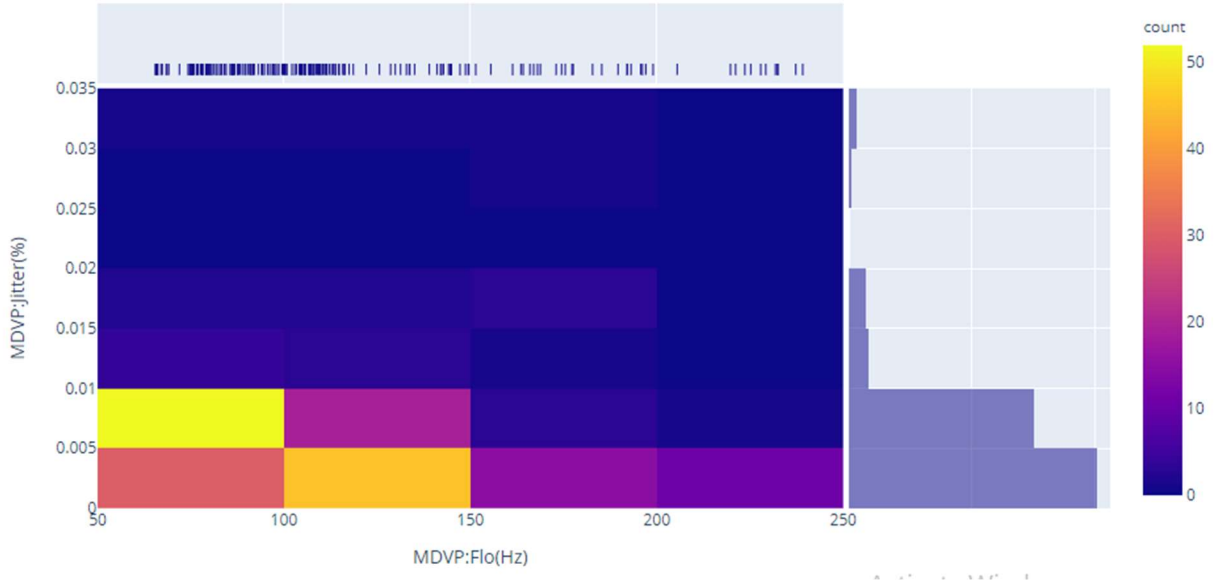
D. TN stands for true negative, which means that if another label correctly predicts.

3.2.9 Histogram

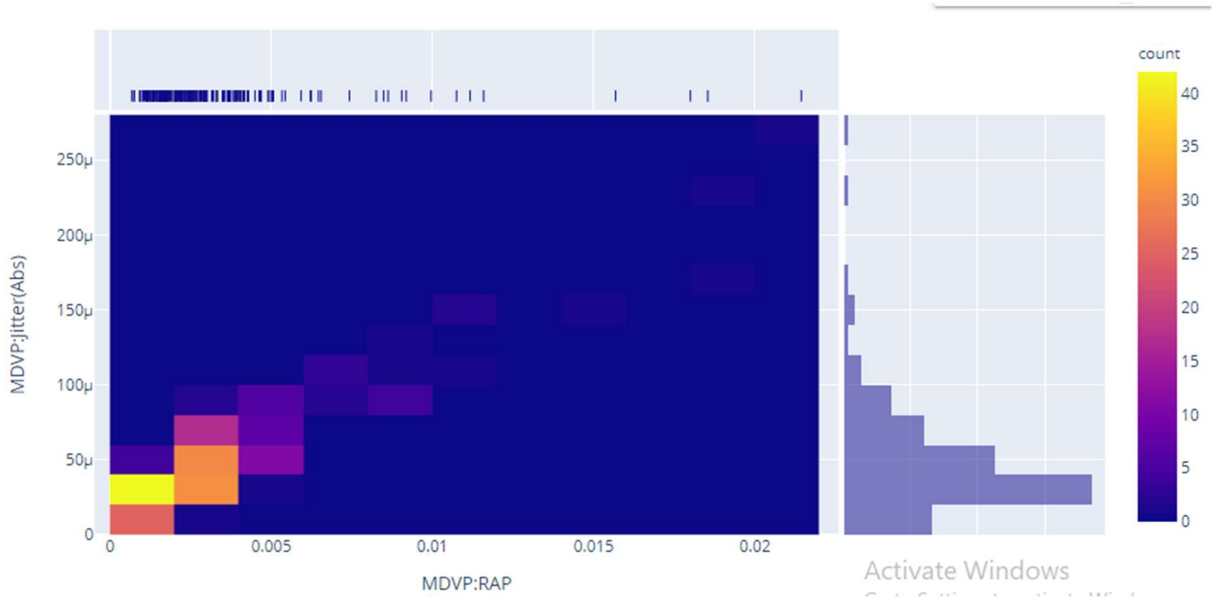
Histogram of x="name", y="MDVP:Fo(Hz)" is given below,



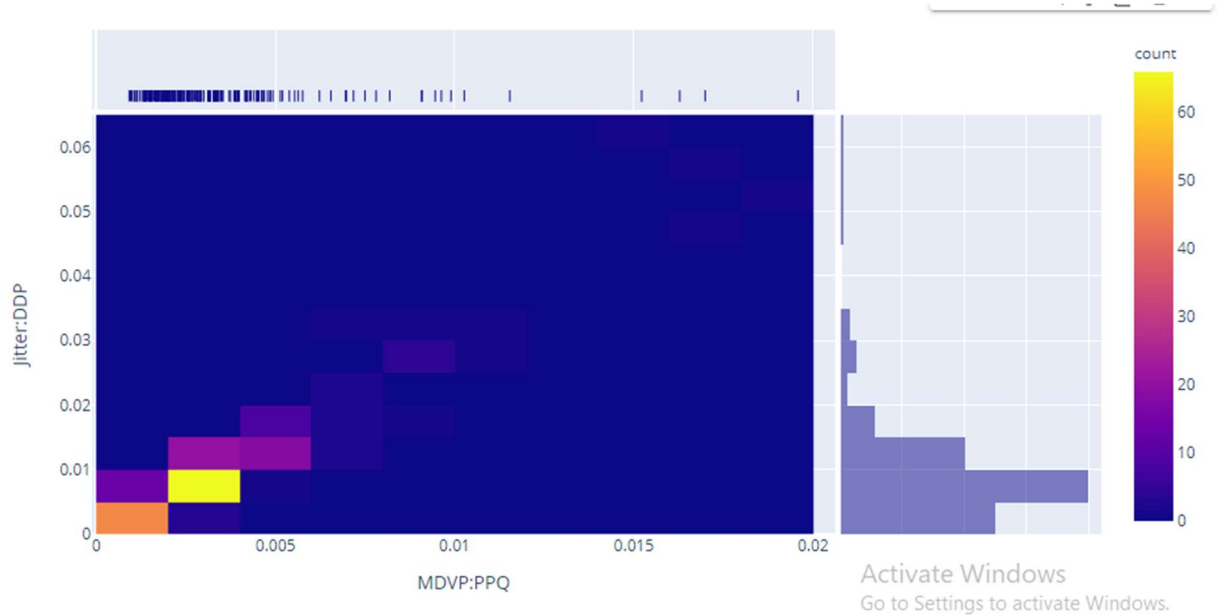
Histogram of x="MDVP:Flo(Hz)", y="MDVP:Jitter(%)" is given below,



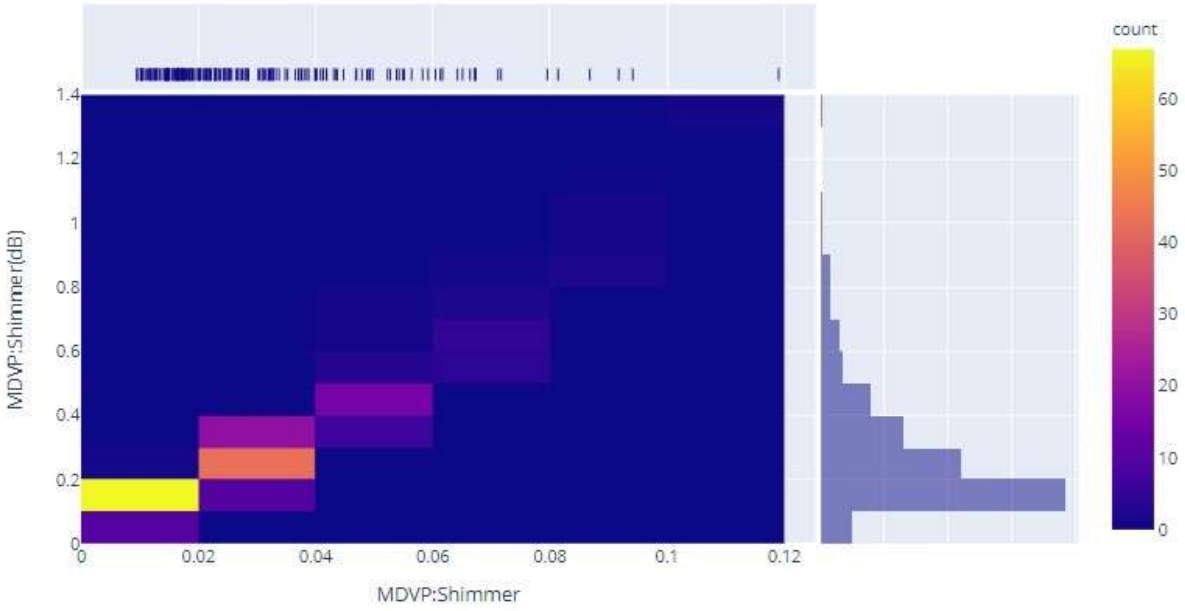
Histogram of x="MDVP:RAP", y="MDVP:Jitter(Abs)"is given below,



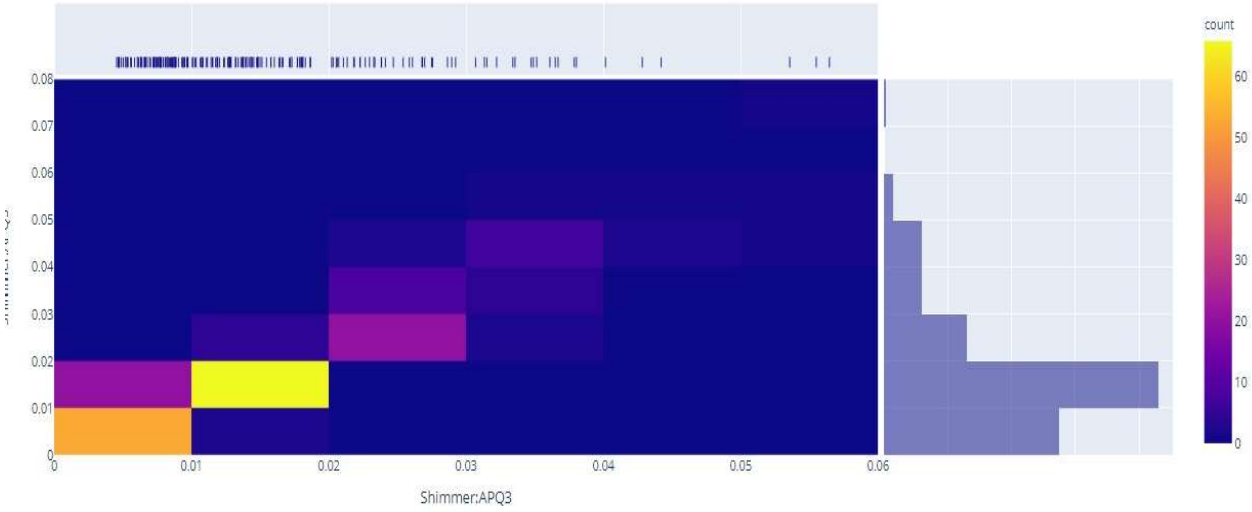
Histogram of x="MDVP:PPQ", y="Jitter:DDP" is given below,



Histogram of x="MDVP:Shimmer", y="MDVP:Shimmer(dB)" is given below,



Histogram of x="Shimmer:APQ3", y="Shimmer:APQ5" is given below,



CHAPTER 4 RESULT
ANALYSIS

4.1 Confusion Metrics Analysis

4.1.1 Confusion Metrics Analysis using Cross Validation

For cross validation, the confusion metrics study is provided in table I, hence, the confusion metrics study is an essential aspect of this research. Anticipated negative is a more crucial feature for achieving excellent accuracy than predicted positive.

We have calculated the predictive positive and predictive negative result here.

In this particular research on Parkinson's Disease prediction, K Nearest Neighbor Classifier (KNN) outperforms all other classifiers in terms of performance with 97% of accuracy.

Table I: Confusion matrix for cross-validation

Models	**Accuracy**	**Predictive Positive**	**Predictive Negative**
XGB Classifier	0.95	5 0	2 32
KNN Classifier	0.97	6 0	1 32
Decision Tree Classifier	0.90	6 3	1 29
Random Forest Classifier	0.92	4 0	3 32
Gradient Boost Classifier	0.90	4 1	3 31
Support Vector Classifier	0.90	3 0	4 32
Bagging Classifier	0.92	5 1	2 31
AdaBoost Classifier	0.90	6 3	1 29
Naive Bayas	0.72	4 8	3 24

4.2 Evaluation Metrics Analysis

4.2.1 Evaluation Metrics Analysis for Cross Validation

Results described in the given Table II using k fold cross validation from different ML classifiers like : Naive Bayes, Adaptive Boosting, Bagging Classifier, Decision Tree Classifier, Random Forest classifier, XBG Classifier, K Nearest Neighbor Classifier, Support Vector Classifier and Gradient Boosting Classifier are used.

We have used in this regarding research we have to get the accurate machine learning approaches to find out the disease with predictable dataset and the model of the following machine learning classifiers. The evaluation matrices and cross validation and True positive and False positive rate as well. In this particular research on Parkinson's Disease prediction, K Nearest Neighbor Classifier (KNN) outperforms all other classifiers in terms of performance with 97% of accuracy

Table II: Evaluation Metrics for Cross Validation

Models	**Accuracy**	**Precision**	**Recall**	**F1-score**	**TP rate**	**TN rate**
XGB Classifier	0.95	1.00 0.94	0.71 1.00	0.83 0.97	1.00	0.94
KNN Classifier	0.97	1.00 0.97	0.86 1.00	0.92 0.98	1.00	0.97
Decision Tree Classifier	0.90	0.67 0.97	0.86 0.91	0.75 0.94	0.66	0.97
Random Forest Classifier	0.92	1.00 0.91	0.57 1.00	0.73 0.96	1.00	0.91
Gradient Boost Classifier	0.90	0.80 0.91	0.57 0.97	0.67 0.94	0.80	0.91
Support Vector Classifier	0.90	1.00 0.89	0.43 1.00	0.60 0.94	1.00	0.89
Bagging Classifier	0.92	0.83 0.94	0.71 0.97	0.77 0.95	0.83	0.94
AdaBoost Classifier	0.90	0.67 0.97	0.86 0.91	0.75 0.94	0.67	0.97
Naive Bayas Classifier	0.72	0.33 0.89	0.57 0.75	0.42 0.81	0.33	0.89

CHAPTER 5
CONCLUSION AND FUTURE WORK

5.1 Conclusion

Parkinson’s Disease has a nerve disorder that usually occur to older people. Age is a definite danger issue for Parkinson's disease. Although the popular of individuals with Parkinson's acquire the illness around the age of 60. Very few have the disease in young age, happen to get aged 50. Early onset Parkinson's disease is frequently inherited, but none for every time, and some types have been related to particular gene alterations. Our target is to predict disease the best period of life span is possible with the help of ML algorithm. We have done the research with help of datasets. After the tarin of different algorithms we have found that on Parkinson’s Disease prediction, K Nearest Neighbor Classifier (KNN) outperforms all other classifiers in the best performance with the accurate result of 97%. With K Nearest Neighbor Classifier (KNN) other machine learning classifiers showed good accuracy as well. But for K Nearest Neighbor Classifier (KNN) the precision, the recall and the f measure values are also having the best accurate prediction and detection result.

5.2 Future Work

For a machine learning and data mining approaches, everything is depended on data. After using more data, the work will have more hidden patterns and dimensions. This dataset had 195 records, which is insufficient for a more accurate result. More information may be added to this dataset in the future, and the greatest diversity of data would be an advantage for subsequent study. For the future this particular research, deep learning will be implemented, as well as image processing or big data analysis will be implemented as well.

REFERENCE

[1] Pahuja, G., & Nagabhushan, T. N. (2021). A comparative study of existing machine learning approaches for Parkinson's disease detection. IETE Journal of Research, 67(1), 4-14.

[2] Ali, L., Zhu, C., Golilarz, N. A., Javeed, A., Zhou, M., & Liu, Y. (2019). Reliable Parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model. Ieee Access, 7, 116480-116489.

[3] Prashanth, R., Roy, S. D., Mandal, P. K., & Ghosh, S. (2016). High-accuracy detection of early Parkinson's disease through multimodal features and machine learning. International journal of medical informatics, 90, 13-21.

[4] Senturk, Z. K. (2020). Early diagnosis of Parkinson's disease using machine learning algorithms. Medical hypotheses, 138, 109603.

[5] Parisi, L., RaviChandran, N., & Manaog, M. L. (2018). Feature-driven machine learning to improve early diagnosis of Parkinson's disease. Expert Systems with Applications, 110, 182-190.

[6] Mandal, I., & Sairam, N. (2014). New machine-learning algorithms for prediction of Parkinson's disease. International Journal of Systems Science, 45(3), 647-666.

[7] Wang, W., Lee, J., Harrou, F., & Sun, Y. (2020). Early detection of Parkinson's disease using deep learning and machine learning. IEEE Access, 8, 147635-147646.

[8] Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., ... & Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and Progressive Supranuclear Palsy. Journal of neuroscience methods, 222, 230-237.

[9] Wroge, T. J., Özkanca, Y., Demiroglu, C., Si, D., Atkins, D. C., & Ghomi, R. H. (2018, December). Parkinson's disease diagnosis using machine learning and voice. In 2018 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (pp. 1-7). IEEE.

[10] Challa, K. N. R., Pagolu, V. S., Panda, G., & Majhi, B. (2016, October). An improved approach for prediction of Parkinson's disease using machine learning techniques. In 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1446-1451). IEEE.

[11] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M. (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. Biocybernetics and Biomedical Engineering, 38(1), 1-15.

[12] Almeida, J. S., Rebouças Filho, P. P., Carneiro, T., Wei, W., Damaševičius, R., Maskeliūnas, R., & de Albuquerque, V. H. C. (2019). Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. Pattern Recognition Letters, 125, 55-62.

[13] Lahmiri, S., Dawson, D. A., & Shmuel, A. (2018). Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. Biomedical engineering letters, 8(1), 29-39.

[14] Moshkova, A., Samorodov, A., Voinova, N., Volkov, A., Ivanova, E., & Fedotova, E. (2020, April). Parkinson's disease detection by using machine learning algorithms and hand movement signal from LeapMotion sensor. In 2020 26th Conference of Open Innovations Association (FRUCT) (pp. 321-327). IEEE.

[15] Rovini, E., Maremmani, C., Moschetti, A., Esposito, D., & Cavallo, F. (2018). Comparative motor pre-clinical assessment in Parkinson's disease using supervised machine learning approaches. Annals of biomedical engineering, 46(12), 2057-2068.

APPENDICES

Abbreviation

ML = Machine Learning

ROC = Receiver Operating Characteristic

NB = Naive Bayes

SVC = Support Vector Classifier

AdaBoost = Adaptive Boosting

DTC = Decision Tree Classifier

Appendix: Research Reflections

We had very little expertise with machine learning methods for detection and prediction when we started this project. Our supervisor was really pleasant and accommodating. He provided us with invaluable advice and was quite helpful. We learnt a lot of new knowledge, strategies, and algorithms throughout this study period, and we experimented with a variety of ways. When we first start working with this, we run into a lot of issues, but as time goes on, we get more comfortable with the algorithms.

Finally, we acquired bravery and were encouraged to perform additional study in the future as a result of our investigation.

PLAGIARISM REPORT

Parkinson

ORIGINALITY REPORT

20%	15%	10%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	Submitted to Daffodil International University Student Paper	4%
3	Minhaz Uddin Emon, Al Mahmud Imran, Rakibul Islam, Maria Sultana Keya, Raihana Zannat, Ohidujjaman. "Performance Analysis of Chronic Kidney Disease through Machine Learning Approaches", 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021 Publication	1%
4	healthjade.com Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	Submitted to University of Melbourne Student Paper	1%
7	github.com Internet Source	1%

8	Luca Parisi, Narrendar RaviChandran, Marianne Lyne Manaog. "Feature-driven machine learning to improve early diagnosis of ParKinson's disease", Expert Systems with Applications, 2018 Publication	<1 %
9	coek.info Internet Source	<1 %
10	www.coursehero.com Internet Source	<1 %
11	Salim Lahmiri, Debra Ann Dawson, Amir Shmuel. "Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures", Biomedical Engineering Letters, 2017 Publication	<1 %
12	"Computer Networks and Inventive Communication Technologies", Springer Science and Business Media LLC, 2022 Publication	<1 %
13	Jefferson S. Almeida, Pedro P. Rebouças Filho, Tiago Carneiro, Wei Wei et al. "Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques", Pattern Recognition Letters, 2019 Publication	<1 %

	wood allocation planning", International Journal of Production Economics, 2020 Publication	
22	Submitted to University of Edinburgh Student Paper	<1 %
23	boa.unimib.it Internet Source	<1 %
24	annalsofrscb.ro Internet Source	<1 %
25	apo.ansto.gov.au Internet Source	<1 %
26	hdl.handle.net Internet Source	<1 %
27	hosei.repo.nii.ac.jp Internet Source	<1 %
28	Submitted to Bournemouth University Student Paper	<1 %
29	www.jscdss.com Internet Source	<1 %
30	www.nia.nih.gov Internet Source	<1 %
31	www.tandfonline.com Internet Source	<1 %

14	"Advanced Computing and Intelligent Technologies", Springer Science and Business Media LLC, 2022 Publication	<1 %
15	Submitted to Stourbridge College Student Paper	<1 %
16	inthenews.mayoclinic.org Internet Source	<1 %
17	doctorpenguin.com Internet Source	<1 %
18	"Advances in Computing and Data Sciences", Springer Science and Business Media LLC, 2019 Publication	<1 %
19	Gunjan Pahuja, T. N. Nagabhushan. "A Comparative Study of Existing Machine Learning Approaches for Parkinson's Disease Detection", IETE Journal of Research, 2018 Publication	<1 %
20	www.download.bham.ac.uk Internet Source	<1 %
21	Michael Morin, Jonathan Gaudreault, Edith Brotherton, Frédérik Paradis, Amélie Rolland, Jean Wery, François Laviolette. "Machine learning-based models of sawmills for better	<1 %

32	Mehrbakhsh Nilashi, Othman Ibrahim, Hossein Ahmadi, Leila Shahmoradi, Mohammadreza Farahmand. "A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques", Biocybernetics and Biomedical Engineering, 2018 Publication	<1 %
33	Rohit Lamba, Tarun Gulati, Hadeel Fahad Alharbi, Anurag Jain. "A hybrid system for Parkinson's disease diagnosis using machine learning techniques", International Journal of Speech Technology, 2021 Publication	<1 %
34	movementdisorders.onlinelibrary.wiley.com Internet Source	<1 %
35	thesai.org Internet Source	<1 %
36	"Soft Computing Techniques and Applications", Springer Science and Business Media LLC, 2021 Publication	<1 %
37	Debashree Devi, Saroj K. Biswas, Biswajit Purkayastha. "chapter 5 Early Detection of Parkinson's Disease", IGI Global, 2019 Publication	<1 %

38	Jinee Goyal, Padmavati Khandnor, Trilok Chand Aseri. "Classification, Prediction, and Monitoring of Parkinson's disease using Computer Assisted Technologies: A Comparative Analysis", Engineering Applications of Artificial Intelligence, 2020 Publication	<1 %
39	Yunfeng Wu, Pinnan Chen, Yuchen Yao, Xiaoquan Ye, Yugui Xiao, Lifang Liao, Meihong Wu, Jian Chen. "Dysphonic Voice Pattern Analysis of Patients in Parkinson's Disease Using Minimum Interclass Probability Risk Feature Selection and Bagging Ensemble Learning Methods", Computational and Mathematical Methods in Medicine, 2017 Publication	<1 %
40	"Abstracts", Movement Disorders, 2019 Publication	<1 %
41	"Brain Informatics", Springer Science and Business Media LLC, 2021 Publication	<1 %