PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING

PROJECT REPORT

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB

NAME:HEMANT VG

SUBJECT:INT 247

REGNO:11910284

SECTION:KM055

ROLL NO:A-16

DECLARATION

I the solemnly declare that the project report titled Parkinson's Detection Using Machine Learning. is based on my own work carried out during my study under the supervision of

Dr. Sagar Pande I assert the statements made, and conclusions drawn are an outcome of my research work. I further certify that The project report is original and it has been done by me

And I declare that I have followed the Guidelines provided to Me.

Name: HEMANT VG

Reg no: 11910284

Acknowledgement

I would like to express my sincere thanks and gratitude to my **Dr. Sagar Pande** for letting me work on this project. I am very grateful to him for letting me an opportunity to individually perform and complete this project on time.

Name: Hemant VG

Reg NO:11910284

INDEX

ABSTRACT:

The Parkinson's disease is a neurological disorder. It Predominantly produces neurons which affect the specific area a brain. And the symptoms of Parkinson's disease will slowly develop over years and this disease will make our brain dead. but the symptoms are different from other people because diversity. The symptoms are Tremor, slowness of movements, shivering, Bradykinesia, Limb rigidity, Vocal Impairment, and balance problems. There is no permanent cure for this disease. Some affected people can feel better by taking medicines. And for some patients will undergo surgery to improve from the common symptoms.

The Parkinson's Disease is an most common disease in the world. There are two types of disease Communicable and non-communicable diseases. Parkinson disease is a communicable disease which can be spread from person to person. And the symptoms will slowly increases over years. And finally it may lead to dead. It is not cure able still now. To overcome from the symptoms surgery has to be taken.

In this project I have used the several machine learning models to predict the different accuracy i.e. Support vector machines (SVM), Random Forest classifier, K-nearest neighbors and XGBoost. To predict the accuracy of the models.

Keywords: Parkinson's disease, SVM, KNN, Random Forest, XGBoost

INTRODUCTION:

Parkinson's Disease is a brain neurological disorder. This leads to an reduction called dopamine in the brain. And many different symptoms are associated with Parkinson's disease and the most common symptoms that the affected patients will have slowness in body movements, muscle hardness. That the Parkinson's disease cannot be cured at the advanced stage as this disease will slowly increase day by day. But with the help of medications and surgery may be used to control the symptoms.

Parkinson's disease occurs when neurons in an area of the brain called the substantia nigra become impaired or die. These cells normally produce dopamine, a chemical called neurotransmitter that helps the cells of the brain communicate, transmits signals and messages between areas in the brain. When these nerve cells become impaired or die, they produce less dopamine. Dopamine is especially important for the operation of another area of the brain called the basal ganglia. This area of the brain is responsible for organizing the brain's commands for body movement. The loss of dopamine causes the movement symptoms seen in people with Parkinson's disease.

People with Parkinson's disease also lose another neurotransmitter called norepinephrine. This chemical is needed for proper functioning of the sympathetic nervous system. This system controls some of the body's autonomic functions such as digestion, heart rate, blood pressure and breathing. Loss of norepinephrine causes some of the non-movement-related symptoms of Parkinson's disease.

PARKINSON'S DISEASE SYMPTOMS:

Parkinson's disease symptoms will vary from person to person. And it is not noticed at the early stages. This symptom often starts from one side of the body and then it turns to affect from the other side of the body.

The most common Parkinson's disease symptoms are:

- **Tremor:** Shaking begins in your hands and arms. It can also occur in your jaw or foot. In the early stages of the disease, usually only one side of your body or one limb is affected. As the disease progresses, tremor may become more wide spread. It worsens with stress. Tremor often disappears during sleep and when your arm or leg is being moved.
- **Slowness of movement (bradykinesia):** This is the slowing down of movement and is caused by your brain's slowness in transmitting the necessary instructions to the appropriate parts of the body. This symptom is unpredictable and can be quickly disabling. One moment you may be moving easily, the next you may need help moving at all and finishing tasks such as getting dressed, bathing or getting out of a chair. You may even drag your feet as you walk.
- **Rigid muscles/stiff limbs:** Rigidity is the inability of your muscles to relax normally. This rigidity is caused by uncontrolled tensing of your muscles and results in you not being able to move about freely. You may experience aches or pains in the affected muscles and your range of motion may be limited.
- **Unsteady walk and balance and coordination problems:** You may develop a forward lean that makes you more likely to fall when bumped. You may take short shuffling steps, have difficulty starting to walk and difficulty stopping and not swing your arms naturally as you walk. You may feel like your feet are stuck to the floor when trying to take a step.
- **Muscle twisting, cramps:** You may experience a painful cramp in your foot or curled and clenched toes. Dystonia can occur in other body parts.
- Decreased facial expressions: You may not smile or blink as often as the disease worsens; your face lacks expression.
- Speech and vocal changes: Speech may be quick, become slurred or be soft in tone. You may hesitate before speaking. The pitch of your voice may become unchanged monotone.
- The affected person will not have an good body posture.
- Depression and anxiety.
- Chewing and swallowing problems, drooling.
- Urinary problems.
- Sleeping Disturbances including disrupted sleep, acting out your dreams, and restless leg syndrome.
- Loss of smell, as well as skin problems.
- Pain, lack of interest, fatigue, change in motion and weight, vision changes.
- Low blood pressure and urinary problems

Fig1: Parkinson's disease symptoms.

PARKINSON'S DISEASE DIFFERENT STAGES

Parkinson's disease mostly attacks the aged person above age 60. And for the aged people the stage of the disease will be in advanced stage. About more men are attacked to Parkinson's disease than women. The Parkinson's disease can be treated for the people below age of 50. As it is the early stage of the Parkinson's disease.

The Parkinson's disease has different stages that symptoms affect the person slowly from early stage to advanced stage.

Early stage:

In Parkinson's disease it usually starts from early stage then, mild stage and typically occur slowly and do not interfere with daily activities. Sometimes early symptoms are not easy to detect, or you may think early symptoms are simply normal signs of aging. You may have fatigue or a general sense of uneasiness. You may feel a slight tremor or have difficulty standing.

Often, any of the family member will notice some of the awkwardness that the affected person will do. They may notice things like body stiffness or lack of normal movement no arm swing when walking, slow or small handwriting, lack of expression in your face, or difficulty getting out of a chair.

Mid stage

In this stage the Symptoms will start getting worse. Tremor, muscle stiffness and movement problems may now affect both sides of the body. Balance problems and falls are becoming more common. You may still be fully independent but daily tasks of everyday living, such as bathing and dressing, are becoming more difficult to do and take longer to complete.

Mid-late stage

In this Mid late stage, it is very difficult to diagnosis of Parkinson's disease. If the affected person age is more than 60 it is difficult to cure. And they will start feeling the symptoms like Standing and walking are becoming more difficult and may require assistance with a walker. You may need full time help to continue to live at home.

Advanced stage

In this Advanced stage the person affected from Parkinson's disease will undergo fulltime hospitalized And the person cannot stand walk and now the

person requires a wheelchair to get around or are bedridden. You may experience hallucinations or delusions. You now require full-time nursing care.

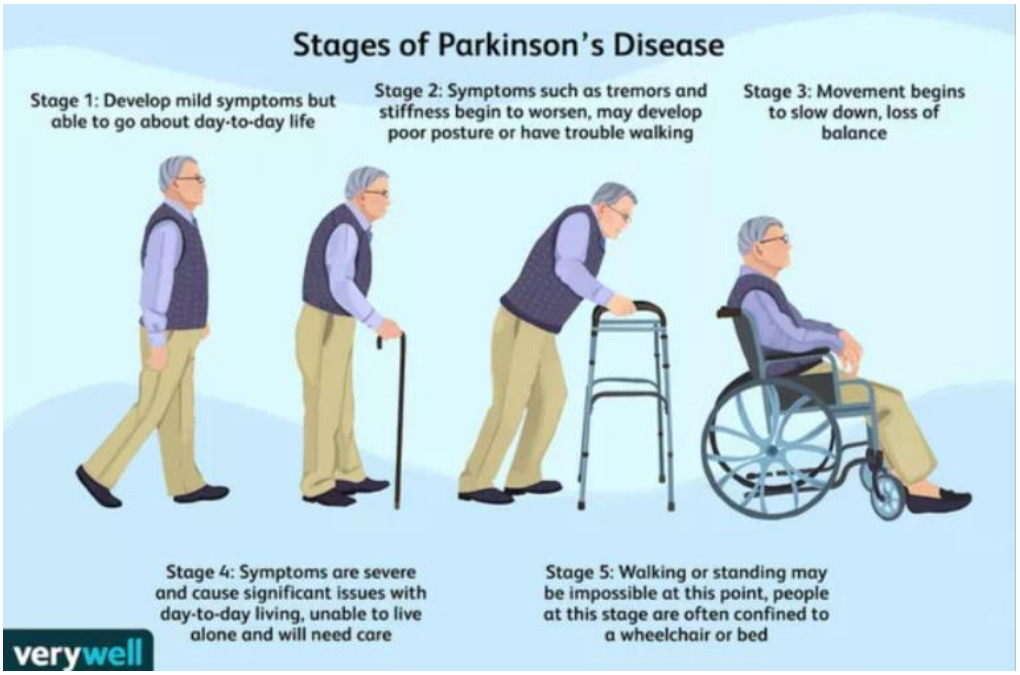


Fig2: Different stages of Parkinson's disease

Problem Statement:

Parkinson's disease is a progressive nervous system disorder that affects movement leading to shaking, stiffness, and difficulty with walking, balance, and coordination. Parkinson's symptoms usually being gradually and get worse over time.

Diagnosis of Parkinson's Disease commonly demands a neurological record of patient with observations of motor skills in numerous conditions. It gets more difficult for the clinician at the early stages during the diagnosis when motor effects are not yet severe. A patient needs to revisit the clinic often to track the progress of the disease over time.

The productive screening process does not demand a medical visit and can be more helpful. People who are having Parkinson's Disease show distinctive voice characteristics, therefore voice recordings are a beneficial tool for the diagnosis. Implementation of machine learning algorithms on the speech dataset for accurate diagnosis of the disease would be a productive screening step before visiting the doctor.

OBJECTIVE:

The main aim of this project is to predict weather the person is affected with the Parkinson's Disease are not using machine learning. In This project the different machine learning models has been used to predict accurately.

The following objective of the model are.

- Developing the model to predict the patient affected with the Parkinson's disease or not.
- Predicting the accuracy of each machine learning models
- Analysing the dataset and splitting the dataset into training and testing for the prediction of the Parkinson's disease
- Different Classification model is used for the Prediction of Parkinson's Disease.

LITREATURE REVIEW:

This portion of the paper describes the background of the project with the explanation of overview of Parkinson's disease, overview of machine learning, and the related works.

PARKINSION'S DISEASE:

As nowadays We all know that each person has very less nutrition in our body. Some People doesn't prefer a good food or lack of water content in our body may leads to a disease.

Parkinson's Disease is one of the most common diseases in the world. There are two types of diseases they are communicable and non-communicable diseases.

The Parkinson's Disease is a communicable disease as it spread from people to people. And it affects the main region of the brain and cell does not work properly and it may lead to brain dead also.

The Parkinson's disease symptoms will slowly start and affect our body. Still now this disease is not cure able at the advanced stage.

As I have said before it will mostly affect the people age above 60.

And the symptoms can be cured with some medications and surgery.

Related Work:

Pereira et al (2017) reviewed that the illness of the disease is characterized that this is an neurological disorder which affects the neurons of the brain. Parkinson's disease easily affects the person with average age in many countries. According to Parkinson's diseases foundation (2015), about 10 million people worldwide have PD, one million of them in the United States (Parkinson's Disease Foundation, 2015). The website of the Parkinson's Disease Society stated that one individual in every 500 British people has this disease, and it is expected that this number will increase 3-fold in the next 50 years. Normally this illness becomes worse over time and mostly affects people between 50-70 years old. PD was first described by James Parkinson, a British physician, in 1817 and there is still no treatment for PD (Lones et al., 2014; Pereira et al., 2015).

PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING

Parkinson disease has different types of symptoms which will affect our body slowly over years and years. As this disease will start from mild stage with normal symptoms like cough, cold and throat infection. When the Parkinson's disease stages get improved to movement disorder, blood pressure, and slowly affect the vocal part of the body at this stage the person has to take some medications, and then if the person is in advanced stage the person need to hospitalize with personal guidance of doctor and nurse as the person could not stand. And person need to go with the surgery to get a relief from the symptoms.

The people who are having Parkinson's Disease mostly 90% of them have a speech impairment, only 3% to 4% of PD patient receives speech therapy and also only one of the most important factor for PD is age, the patient of PD are most of them are aged between 45-60, (Levine et al 2003). The speech of PD patient have change in the frequency specter in their voice because they loss the control of the limb, which decrease the frequency of the audio. So, the low frequency region gives important data to differentiate the speech impairments in PD. Unified Parkinson disease rating scale (UPDRS) is used to find the severity of the PD by help of clinical expertise and experience (Dobson et al 2008).

For motor symptoms, four main signs are considered as cardinal symptoms: rest tremor, rigidity, bradykinesia, and sometimes postural instability. About 70% of PD patients have a resting tremor which is between 3-5 HZ and it characterized as asymmetrical tremor. The second sign of PD is a feeling of resistance during joints' movements, and it is called cogwheel rigidity (Samii et al., 2004). In other words, it is the converse of smooth movements (Khan Academy, 2015).

Slowing down the movement is the third sign, called bradykinesia; it enlarges with simple movements like handwriting. The fourth symptom is postural instability and this one does not happen in the early stage of PD, in particular for younger patients and it is related to balance, which makes the patient unstable on their feet and may lead to falls (Samii et al., 2004; Khan Academy, 2015).

Some of the non-motor symptoms of PD, like hyposmia, rapid eye movement (REM), sleep behaviour disorder, constipation, and depression may emerge before any motor symptoms by years (Meireles et al., 2012). Many patients also exhibit cognitive dysfunction, and these range from what is called mild cognitive impairment (PD-MCI) to PD dementia (PDD) (Litvan et al., 2012). In several cases, PD-MCI emerges in the early stage of the disease, while

PDD tends to occur after 20 years of having the PD. PD-MCI is defined as thinking and memory problems abnormal to what is expected with normal ageing, but without preventing the patient from carrying out daily routine activities. Moreover, PD-MCI diagnosis is important because it could be a transition to PDD (Meireles et al., 2012). PDD symptoms include impaired short-term memory, executive dysfunction, attention impairment, visual-spatial deficit, behavioural or neuropsychiatric symptoms like psychotic symptoms (hallucinations), changes in personality and mood, anxiety and apathy (Meireles et al., 2012). Figure 1 shows how the symptoms develop over time until reaching the dementia stage. 14 | Page Figure 1 PD symptoms development over time (Meireles et al., 2012)

fietzek et al (2020) given the high dataset size requirements are met through a supervised data collection approach by which we were able to generate informative annotations in one-minute intervals. To our knowledge, collecting expert annotations on a oneminute basis has not been reported to date at such a large scale. Abós et al (2017) described that data characteristic without any priori model. we used XGBoost algorithm for classification, XGBoost algorithm benefit from constant learning or retraining, they don't guarantee optimized classification/regression. However, when trained and maintained, XGBoost learning method have great potential than Logistic regression in solving real world problems. The prior report of using XGBoost technique to diagnose Parkinson's disease are determines according to their cognitive status.

In this Parkinson's disease I have used four machine learning models for predicting the accuracy. XGBoost provided an accuracy of 82 % for the classification of the dataset and the Support vector machine (SVM) provides the accuracy 84% for classification of the dataset. and the Random forest provides the accuracy of 82% for the classification of the dataset. And the KNN provides the accuracy of 79%. The predicted Parkinson's disease has been formulated and compares the accuracy of the model with the train and test datasets. Here the SVM provides the best accuracy then other models.

Mohammad et al (2014) performed a comparative analysis to detect Parkinson Disease using various classifiers like Support vector Machine (SVM), Random Tree (RT), feed forward back-propagation Artificial Neural Network (FBNN) classifiers are utilized in this system. Geetha et al (2011) presented a comparison was made between the classifiers to differentiate between PD and Healthy persons and the study has the dataset contains 195 voice samples and consist of both male and female. The dataset has 23 PD patient and healthy, by comparing all the classifiers, FBNN classifier has achieved 97.37% accuracy.

Max A. et alia (2009) [2] presented a Support Vector Machine (SVM) algorithm to classify between people having Parkinson's disease and people who aren't suffering with the help of dysphonia detection. Yadav, G et alia (2009) [3] have presented a classification and Support Vector Machine Classifier (SVC) to distinguish between people having Parkinson's disease & those who are not. This provides an 76% accuracy, for the classification of the dataset.

Ramani et al (2011) discussed a system to classify PD and Non-PD patient was proposed by utilized Binary Logistic Regression, Linear Discriminant Analysis LDA, Random tree and SVM. The dataset used in this system are from UCI repository of PD, the training dataset consist of 195 samples with 21 features, here the LDA and random tree achieved an accuracy greater than 90%. Resul et al (2010) used various classification models to identify PD. Classification techniques were implemented and analyzed, they are neural network, regression and decision tree. for classification various evaluation methods were used, the performance of the classifiers were evaluated from the results, only Neural network classifier yield the good result among other, here the input dataset was randomly inserted into train and test dataset.

Paul et al (2019) have used a machine learning techniques for predicating student dropout using data mining. In this model decision tree was used to predict the dropout in student and they obtained an accuracy with 97.69% and the prediction was done by using various parapets, which are considered for every student. Mallikarjuna et al (2020) presented the feedback-based approach comparison of the normality and abnormality with the back propagation approach. In the training phase, the extracted feature sequence of a normal walking and abnormal walking, the three classes A, B, C, D normal, Parkinson gait, Hemiplegic gait, Neuropathic gait data sets compared with the normal data set.

Yahia A. et alia (2014) [8] presented a comparison between naive Bayes Classifier and KNN algorithm using Parkinson's voice dataset with sound recordings of people having Parkinson's disease and healthy people. The accuracy achieved by the KNN classifier and Naive Bayes algorithm is 80% and 93.3% respectively.

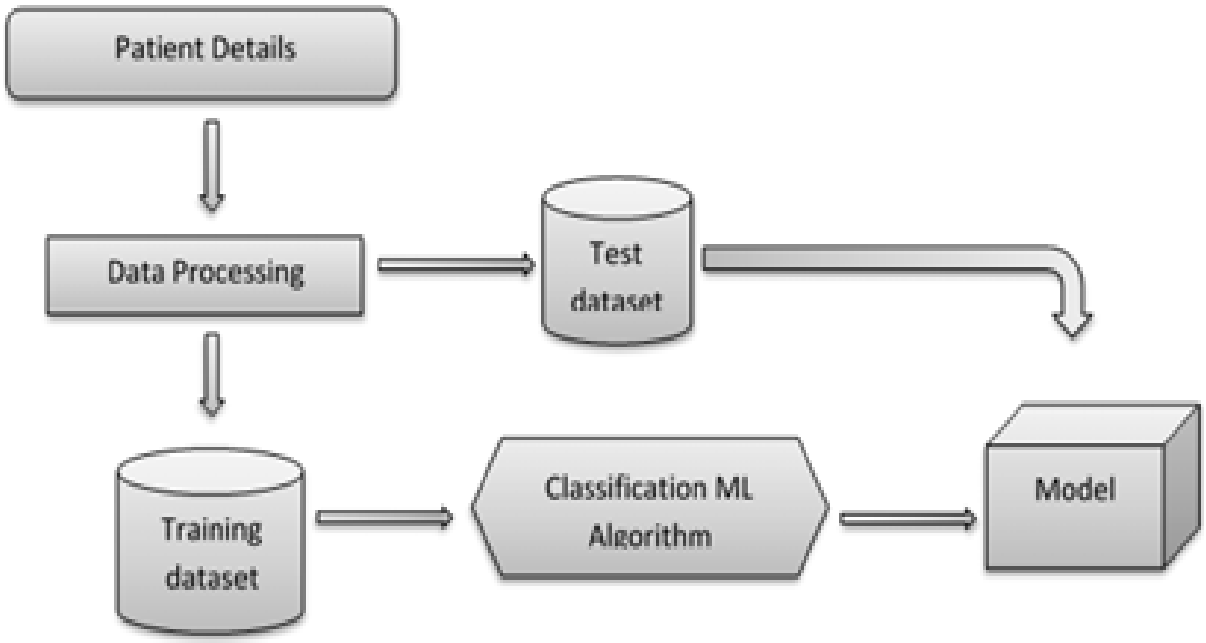
OVERVIEW OF PROJECT:

Project Statement:

In this project the Parkinson's disease detection using machine learning is used to predict weather the person is affected by the Parkinson's disease or not. Using different machine learning models such as

- Support vector machine (SVM)
- Random forest
- K Nearest Neighbor (KNN)
- XGBoost.

Data flow diagram of the model:



Steps involved in implementation of the project:

Step 1: Collecting the datasets of the person affected from Parkinson's disease and not affected.

Step 2: Getting the features and labels from the Collected data sets. And initializing the minmax and scale the feature to normalize.

Step 3: Splitting the collected data set into training and testing.

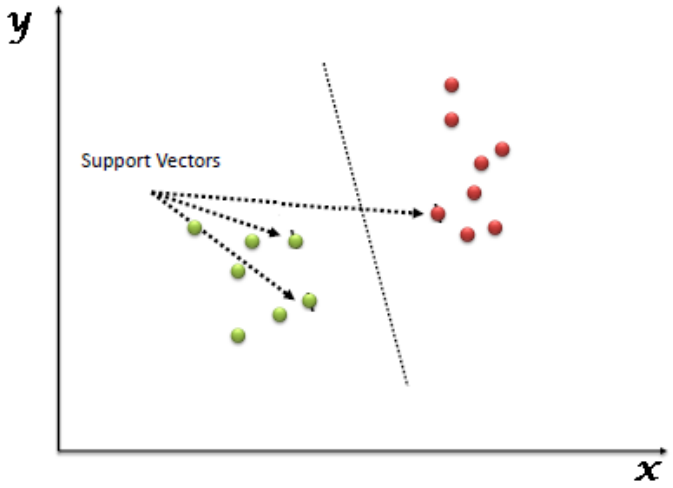
Step 4: Now initializing the different ML classifier model to predict the accuracy of the model.

Step 5: Finally, The model building is performed to predict the person Has Parkinson's disease or Stays Healthy.

MACHINE LEARNING MODELS

SUPPORT VECTOR MACHIES:

Support Vector Machine" (SVM) is a supervised Machine learning Algorithm that can be used for both classification challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well

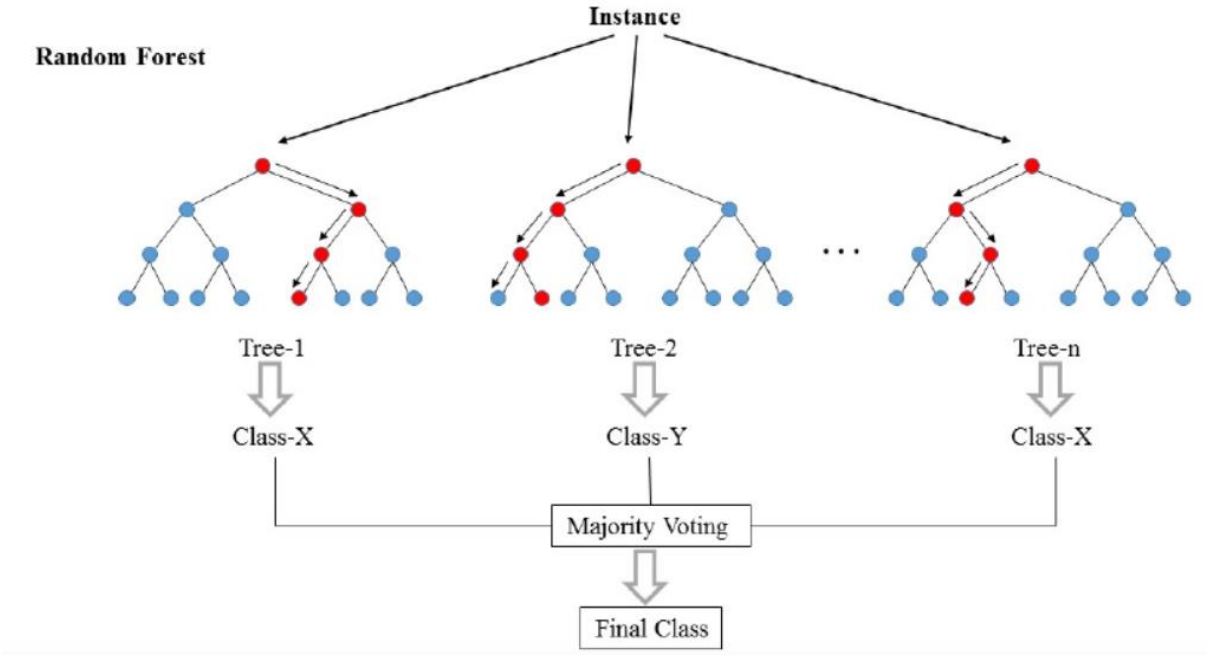


In SVM, we plot each data item in the dataset in an N-dimensional space, where N is the number of features/attributes in the data. Next, find the optimal hyperplane to separate the data. So, by this, you must have understood that inherently, SVM can only perform binary classification to choose between two classes. However, there are various techniques to use for multi-class problems. Support Vector Machine for Multi-Class Problems To perform SVM on multi-class problems, we can create a binary classifier for each class of the data. The two results of each classifier will be.

The data point belongs to that class OR. The data point does not belong to that class.

RANDOM FOREST:

Random forest is a supervised learning algorithm. It can be used for both Classification and Regression problems. It uses an ensemble learning method known as 'bagging' (Bootstrap Aggregation) which is a process of combining multiple classifiers to solve a complex problem.



Random forest creates various random subsets of the given dataset and passes them to different numbers of decision trees and takes the prediction from each tree. Based on the majority votes of prediction, random forest takes the average to predict the output and also to increase the accuracy and overall result. As there are greater numbers of trees in random forests, it prevents the problem of overfitting.

Random forest searches for the most important feature while splitting a node which helps in building a better model.

XGBOOST:

XGBoost was introduced in the year 2014. XGBoost is a new Machine Learning algorithm designed with speed and performance in mind. XGBoost stands for **eXtreme Gradient Boosting and** is based on decision trees. In this project, we will import the XGBClassifier from the xgboost library this is an implementation of the scikit-learn API for XGBoost classification.

The boosting Technique consists of three simple steps:

- An initial model F0 is defined to predict the target variable y. This model will be associated with a residual (y – F0)
- A new model h1 is fit to the residuals from the previous step
- Now, F0 and h1 are combined to give F1, the boosted version of F0. The mean squared error from F1 will be lower than that from F0:

$$F_1(x) <- F_0(x) + h_1(x)$$

To improve the performance of F1, we could model after the residuals of F1 and create a new model F2:

$$F_2(x) <- F_1(x) + h_2(x)$$

This can be done for *'m'* iterations, until residuals have been minimized as much as possible:

$$F_m(x) <- F_{m-1}(x) + h_m(x)$$

The mean minimized the error here. When MAE (mean absolute error) is the loss function, the median would be used as F0(x) to initialize the model. A unit change in y would cause a unit change in MAE as well.

For MSE, the change observed would be roughly exponential. Instead of fitting hm(x) on the residuals, fitting it on the gradient of loss function, or the step along which loss occurs, would make this process generic and applicable across all loss functions.

Gradient descent helps us minimize any differentiable function. Earlier, the regression tree for hm(x) predicted the mean residual at each terminal node of the tree. In gradient boosting, the average gradient component would be computed.

For each node, there is a factor γ with which hm(x) is multiplied. This accounts for the difference in impact of each branch of the split. Gradient boosting helps in predicting the optimal gradient for the additive model, unlike classical gradient descent techniques which reduce error in the output at each iteration.

The following steps are involved in gradient boosting:

- F0(x) – with which we initialize the boosting algorithm – is to be defined:

$$F_0(x) = argmin_\gamma \sum_{i=1}^{n} L(y_i, \gamma)$$

- The gradient of the loss function is computed iteratively:

$$r_{im} = -\alpha \left[\frac{\partial (L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}, \text{ where } \alpha \text{ is the learning rate}$$

- Each hm(x) is fit on the gradient obtained at each step
- The multiplicative factor γm for each terminal node is derived and the boosted model Fm(x) is defined:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

K-Nearest Neighbors.

K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. The following two properties would define KNN well

The algorithm's learning is:

1. Instance-based learning: Here we do not learn weights from training data to predict output (as in model-based algorithms) but use entire training instances to predict output for unseen data.

2. Lazy Learning: Model is not learned using training data prior and the learning process is postponed to a time when prediction is requested on the new instance.

3. Non -Parametric: In KNN, there is no predefined form of the mapping function.

The implementation of the KNN Algorithm.

1.Load the data

2.Initialize K to your chosen number of neighbors

3. For each example in the data

3.1 Calculate the distance between the query example and the current example from the data.

3.2 Add the distance and the index of the example to an ordered collection

4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. If regression, return the mean of the K labels

8. If classification, return the mode of the K labels

CIRCUIT DESCRIPTION:

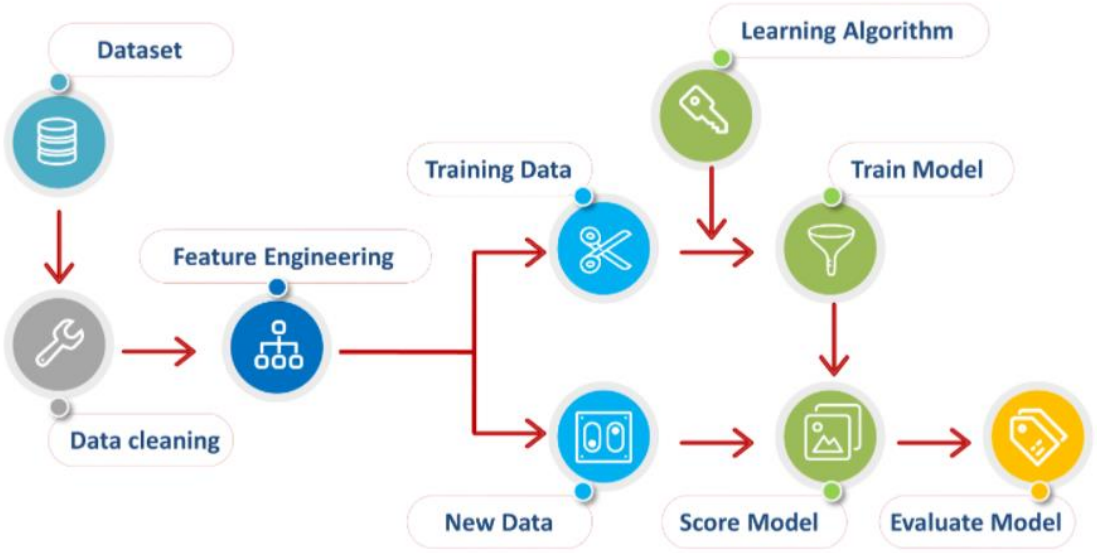
Data-Pre-processing:

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

Implementation:

Here, we will see the structure of the data Pre-processing. As how it works. And How it analyses the performance of the dataset from the machine learning algorithms. Such as XGBoost, SVM, Random Forest classifier, KNN.



Steps involved in data Pre-processing:

Step 1: Importing the required libraries for the model

Step 2: Importing the datasets in the required, model for processing

Step 3: Handling the missing data

Step 4: Encoding the data in categorical

Step 5: Splitting the data into test set and train set

Step 6: Feature scaling.

SOFTWARE DETAIL

The tools used for the Proposed model prediction are.

In this model the mainly utilized is python, and then python libraries for the model then numpy, pandas. Then for the graphical representation matplotlib has been used. And then seaborn is utilized for measuring representation.

Hardware and software requirements:

Hardware	Requirements
RAM	8GB or above
Hard disk	500mb or above
Processor	i3 processor or above
Operating system	Windows 7 or above
Environment Used	Jupiter notebook, Google colab, Spyder, VS code, Pycharm

Data set Description:

The dataset used for the prediction on Parkinson's disease is taken from Kaggle. And the dataset is in the CSV format. In this data set it contains the composed range of biomedical voice measurement of 31 people and 23 with the Parkinson's Disease, Each column in the table is voice measure, and each row corresponds one of 195 voice recordings.

The main aim of the data is to discriminate healthy people from those with PD, according to status. The column which is set to 0 for healthy and 1 for Parkinson's Disease.

DATA SET INFORMATION:

MDVP: Fo(Hz) - Average vocal fundamental frequency
MDVP: Fhi(Hz) - Maximum vocal fundamental frequency
MDVP: Flo(Hz) - Minimum vocal fundamental frequency
MDVP: Jitter(%),

MDVP: Jitter(Abs),

MDVP:RAP,

MDVP:PPQ,

Jitter:DDP - Several
measures of variation in fundamental frequency
MDVP:Shimmer,MDVP:Shimmer(dB),Shimmer:APQ3,Shimmer:APQ5,

MDVP: APQ,Shimmer:DDA - Several measures of variation in amplitude
NHR,HNR - Two measures of ratio of noise to tonal components in the voice
status - Health status of the subject (one) - Parkinson's, (zero) - healthy
RPDE,D2 - Two nonlinear dynamical complexity measures
DFA - Signal fractal scaling exponent
spread1,spread2,PPE - Three nonlinear measures of fundamental frequency variation

RESULT AND CONCLUSION & FUTURE SCOPE:

Result Analysis:

In this project the Parkinson's Disease Detection using Machine Learning Models. The different accuracy of this model has been found using different models of ML.

		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

TP = True Positive: shows the person is not Predicted with positive in Parkinson's disease

FP = False Positive: Shows that the person is not or incorrectly affected to the Parkinson

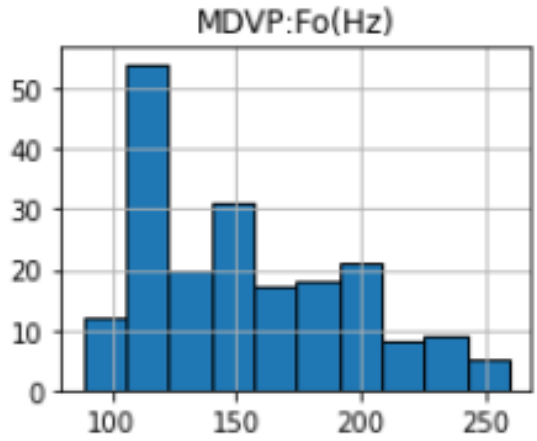
FN = False Negative: Shows that the person stays healthy

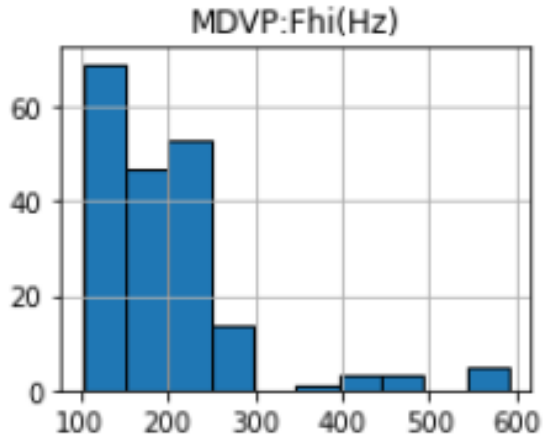
TN = The person is completely affected from Parkinson's disease.

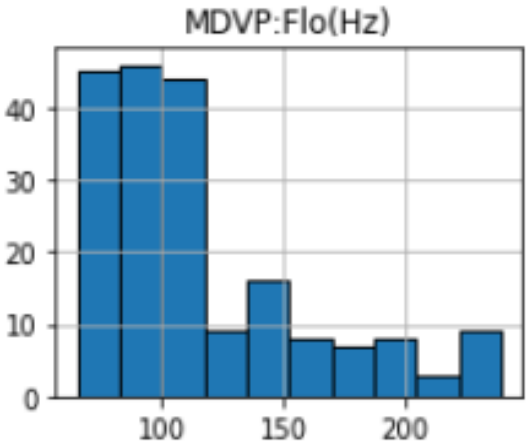
Accuracy Analysis of the different classifiers.

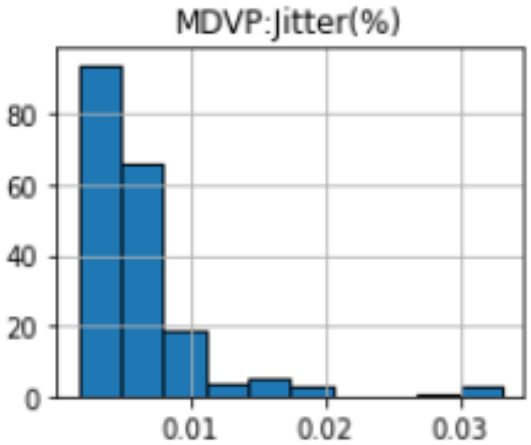
CLASSIFIER	ACCURACY
XGBOOST	82%
Support vector machines	84%
Random Forest	82.5%
K-Nearest Neighbors	79.5%

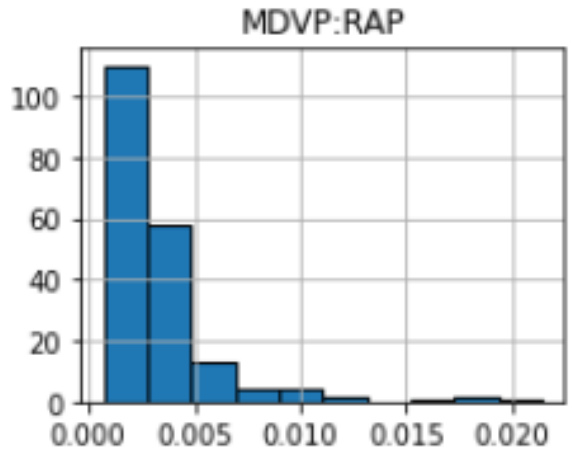
Histogram Representation of the model:

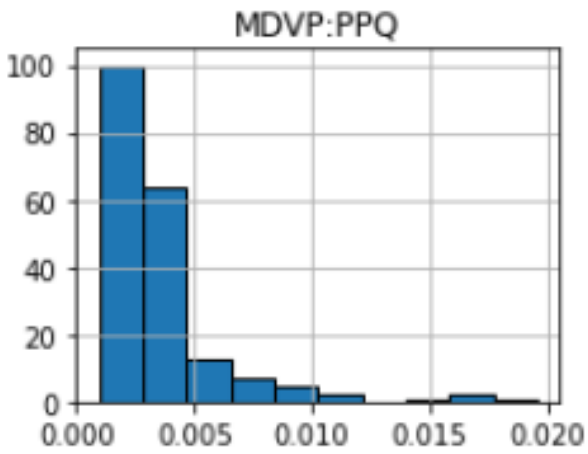


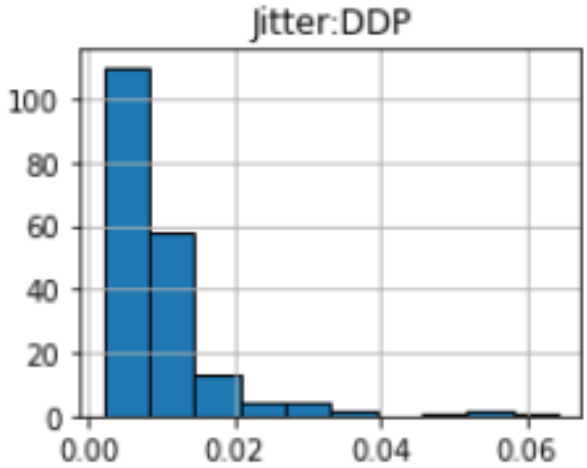


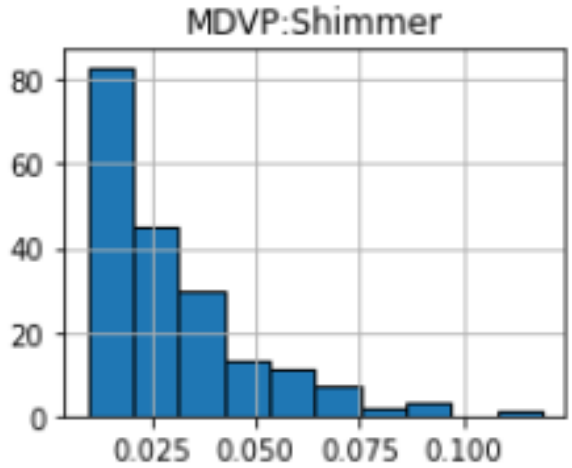


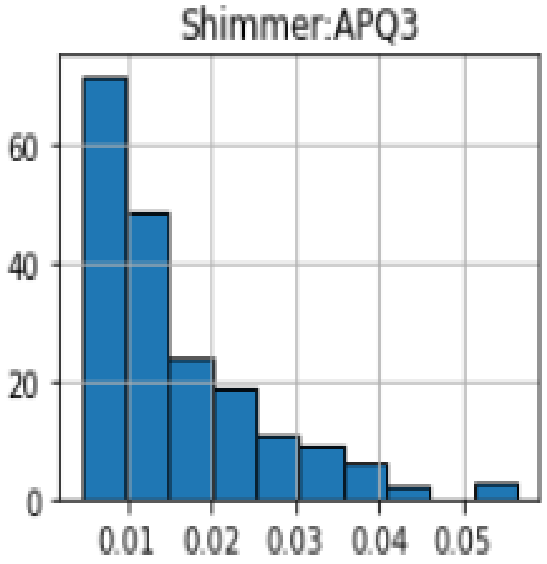


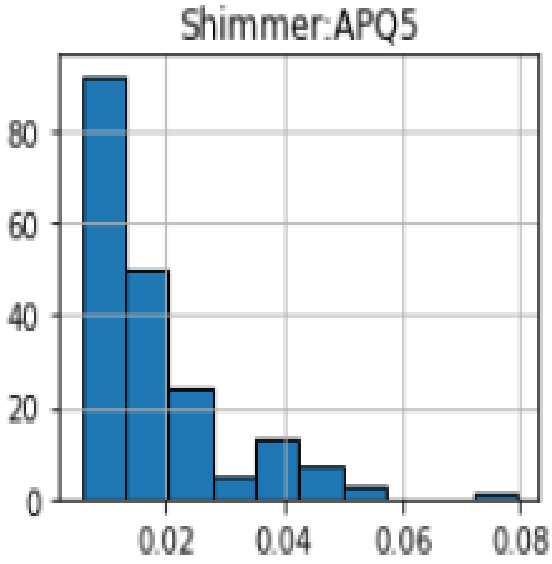

MDVP:RAP
100
80
60
40
20
0
0.000
0.005
0.010
0.015
0.020

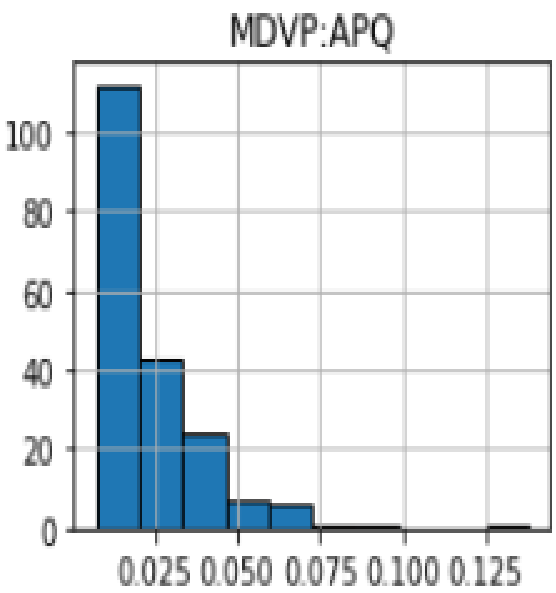

MDVP:PPQ
100
80
60
40
20
0
0.000
0.005
0.010
0.015
0.020

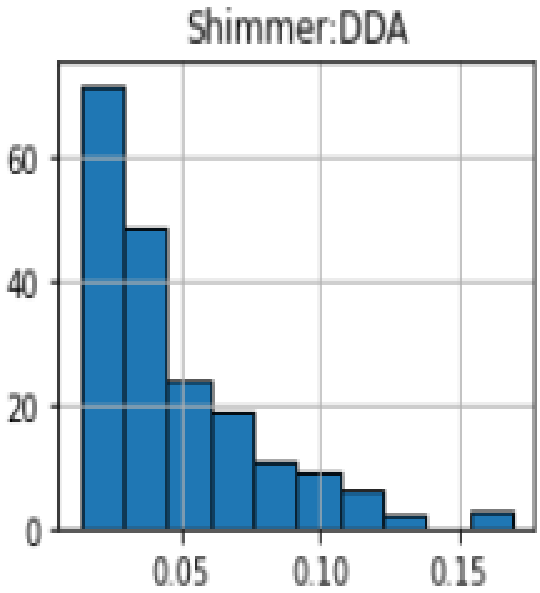

Jitter:DDP
100
80
60
40
20
0
0.00
0.02
0.04
0.06

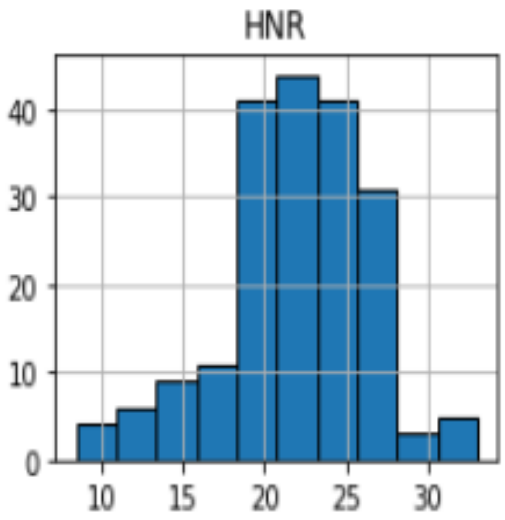

MDVP:Shimmer
80
60
40
20
0
0.025
0.050
0.075
0.100

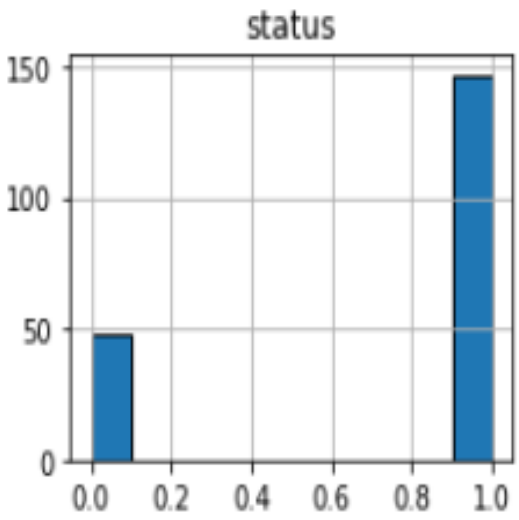

Shimmer:APQ3
60
40
20
0
0.01
0.02
0.03
0.04
0.05

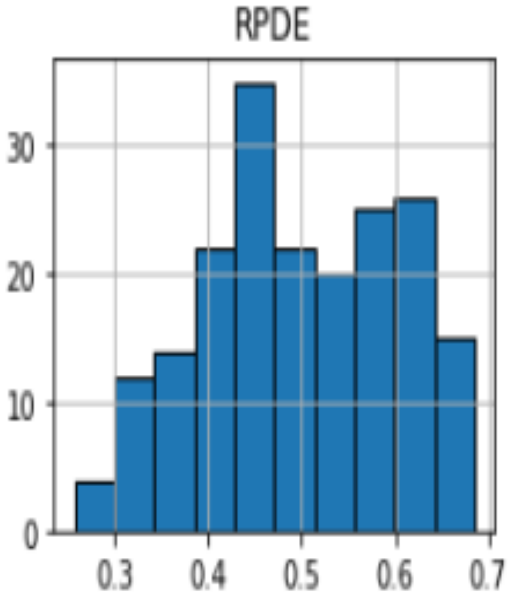

Shimmer:APQ5
80
60
40
20
0
0.02
0.04
0.06
0.08

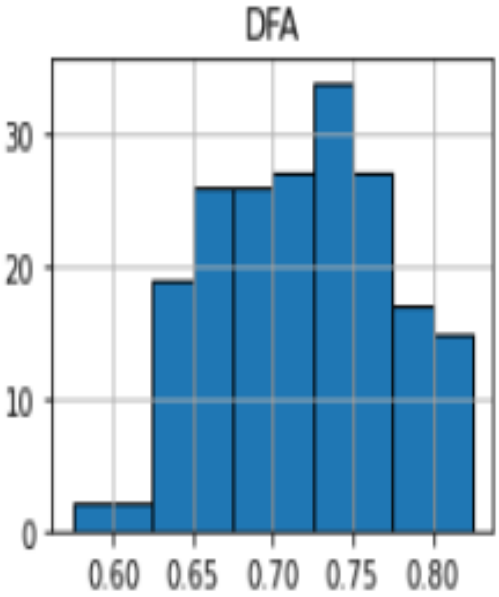

MDVP:APQ
100
80
60
40
20
0
0.025 0.050 0.075 0.100 0.125

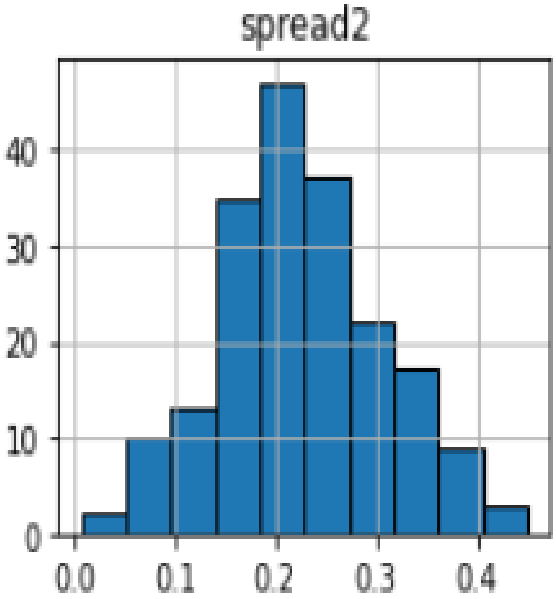

Shimmer:DDA
60
40
20
0
0.05
0.10
0.15

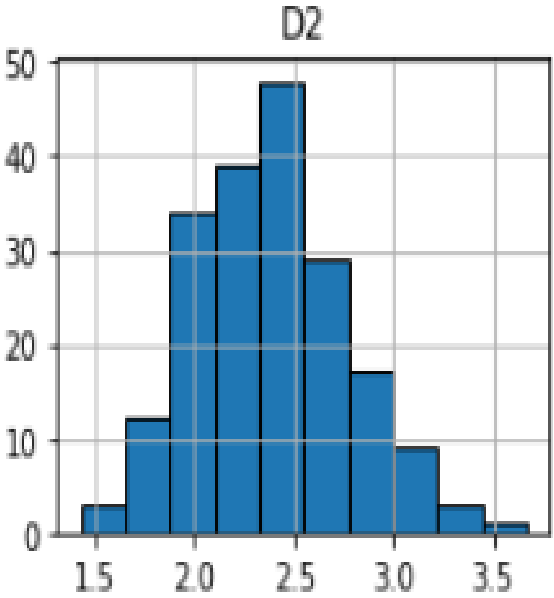

HNR
40
30
20
10
0
10
15
20
25
30

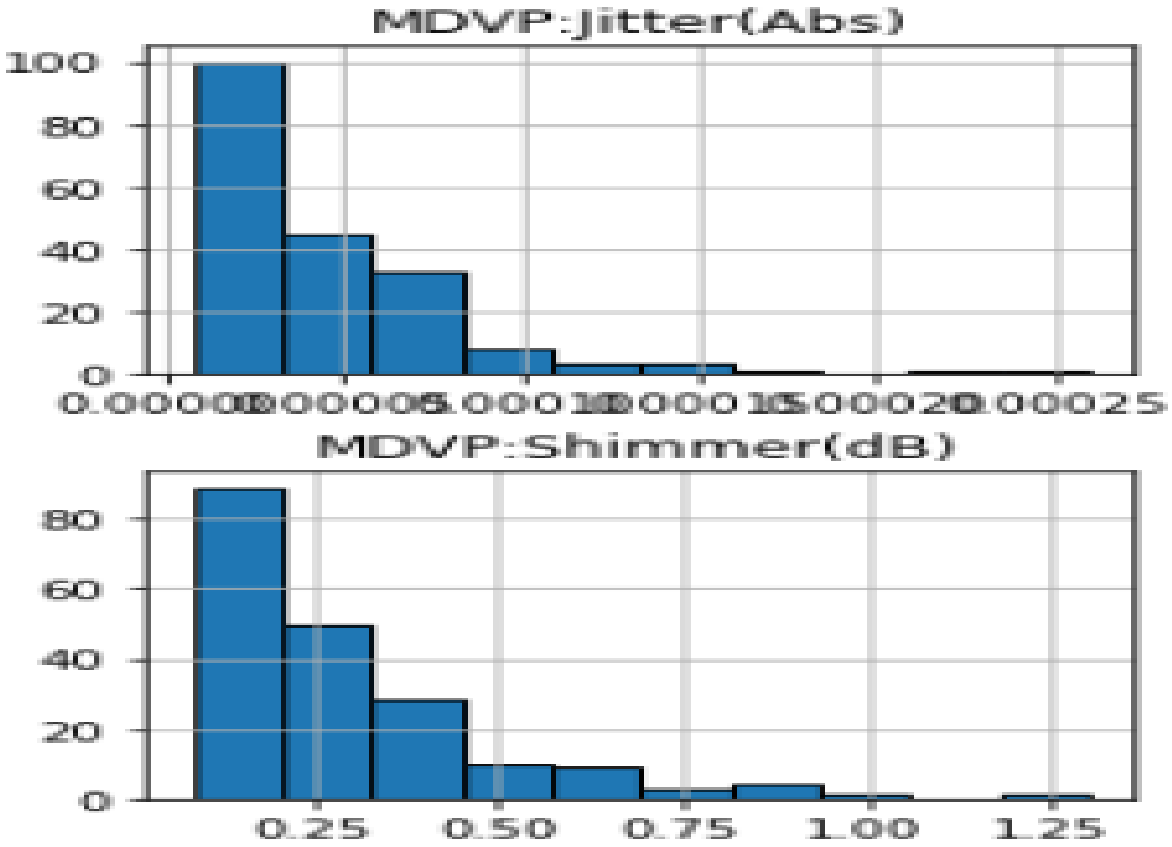

status
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0

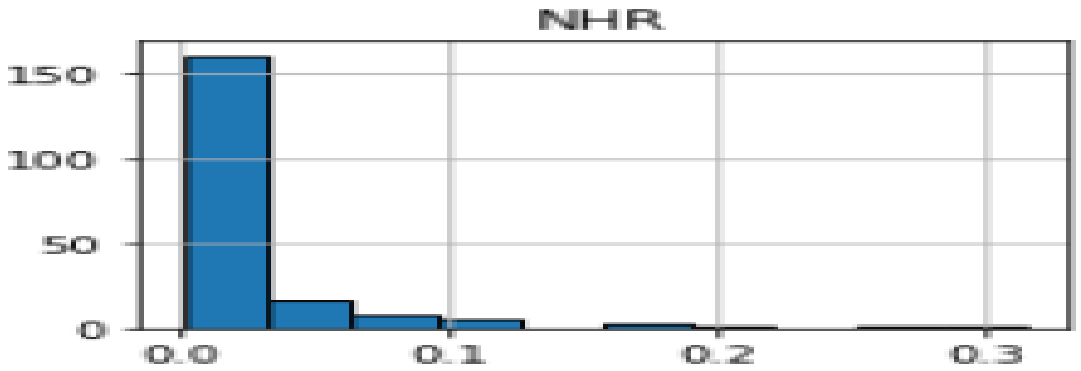

RPDE
30
20
10
0
0.3
0.4
0.5
0.6
0.7

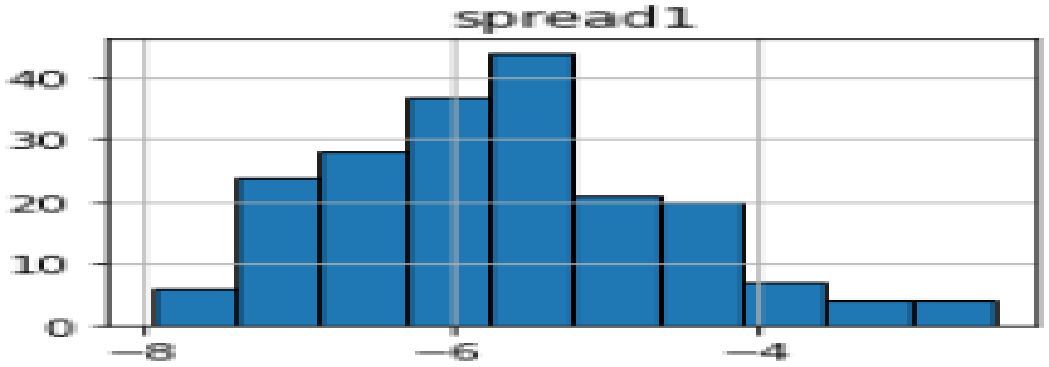

DFA
30
20
10
0
0.60
0.65
0.70
0.75
0.80

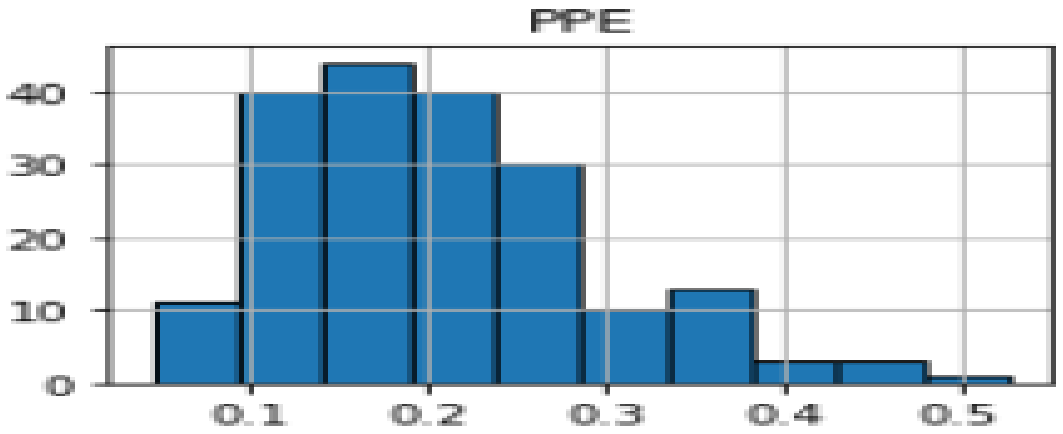

spread2
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4

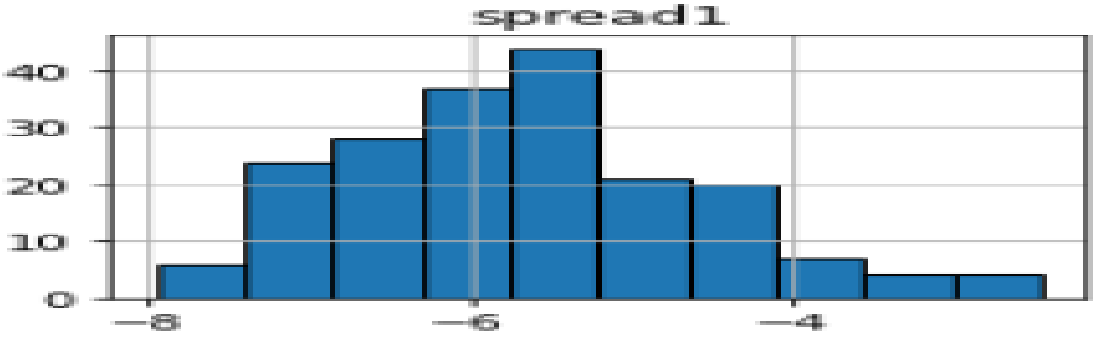

D2
50
40
30
20
10
0
1.5
2.0
2.5
3.0
3.5


MDVP:Jitter(Abs)
100
80
60
40
20
0
0.00025
MDVP:Shimmer(dB)
80
60
40
20
0
0.25
0.50
0.75
1.00
1.25

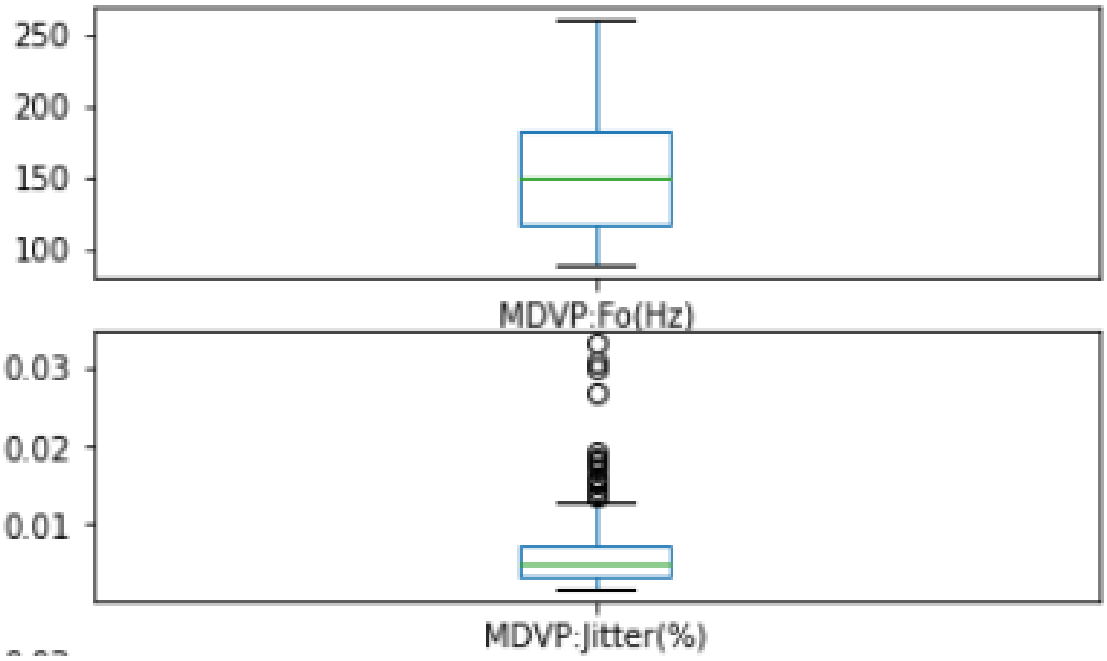

NHR
150
100
50
0
0.0
0.1
0.2
0.3

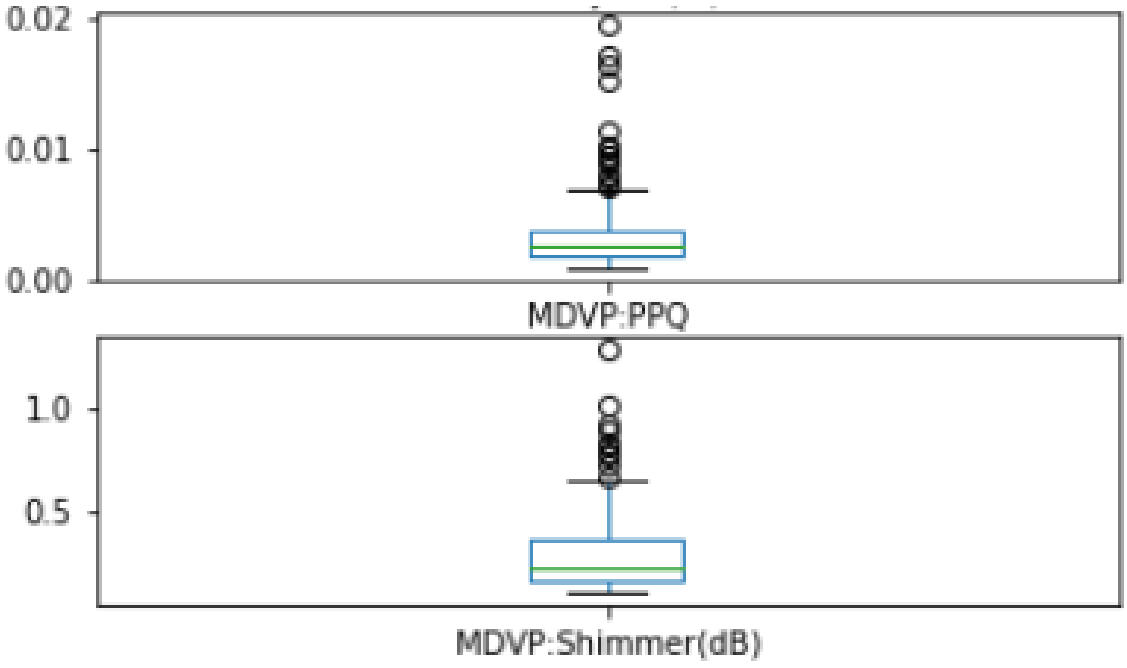

spread1
40
30
20
10
0
−8
−6
−4

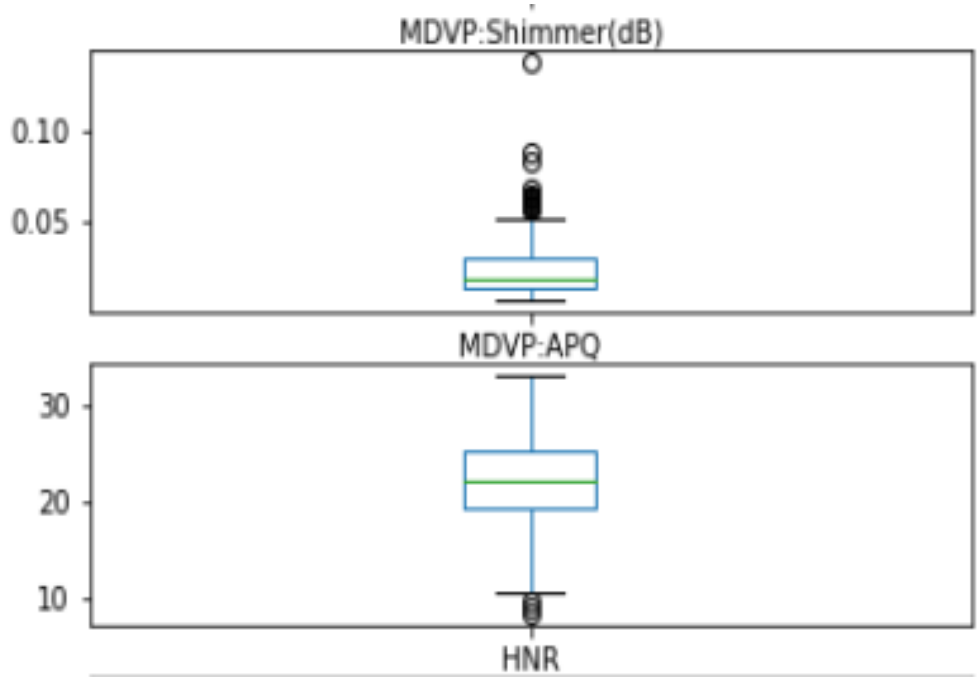

PPE
40
30
20
10
0
0.1
0.2
0.3
0.4
0.5

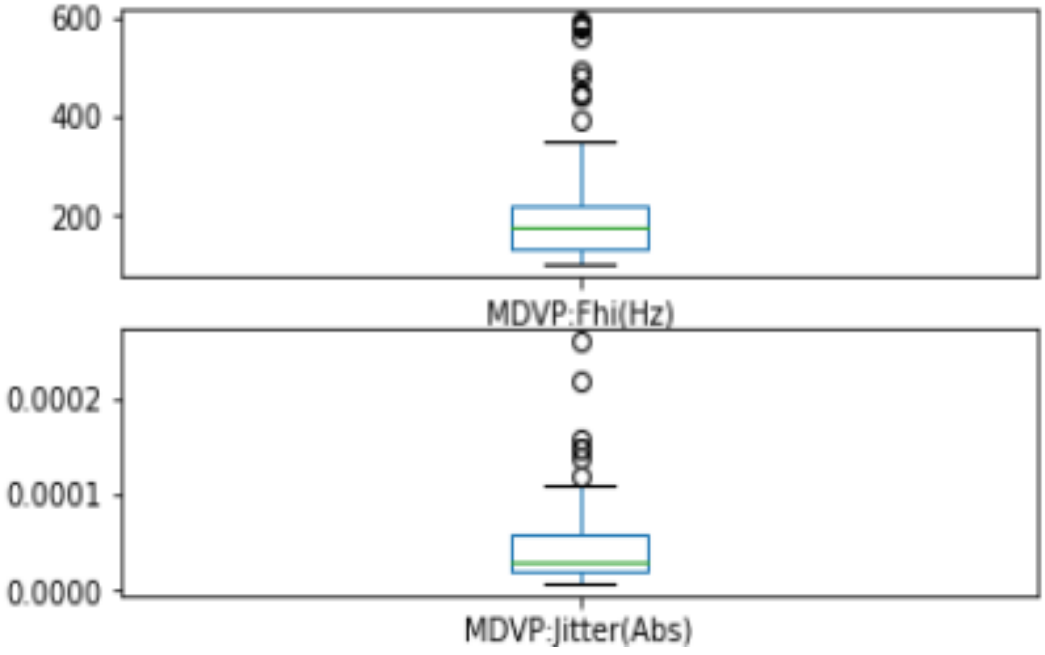

spread1
40
30
20
10
0
−8
−6
−4

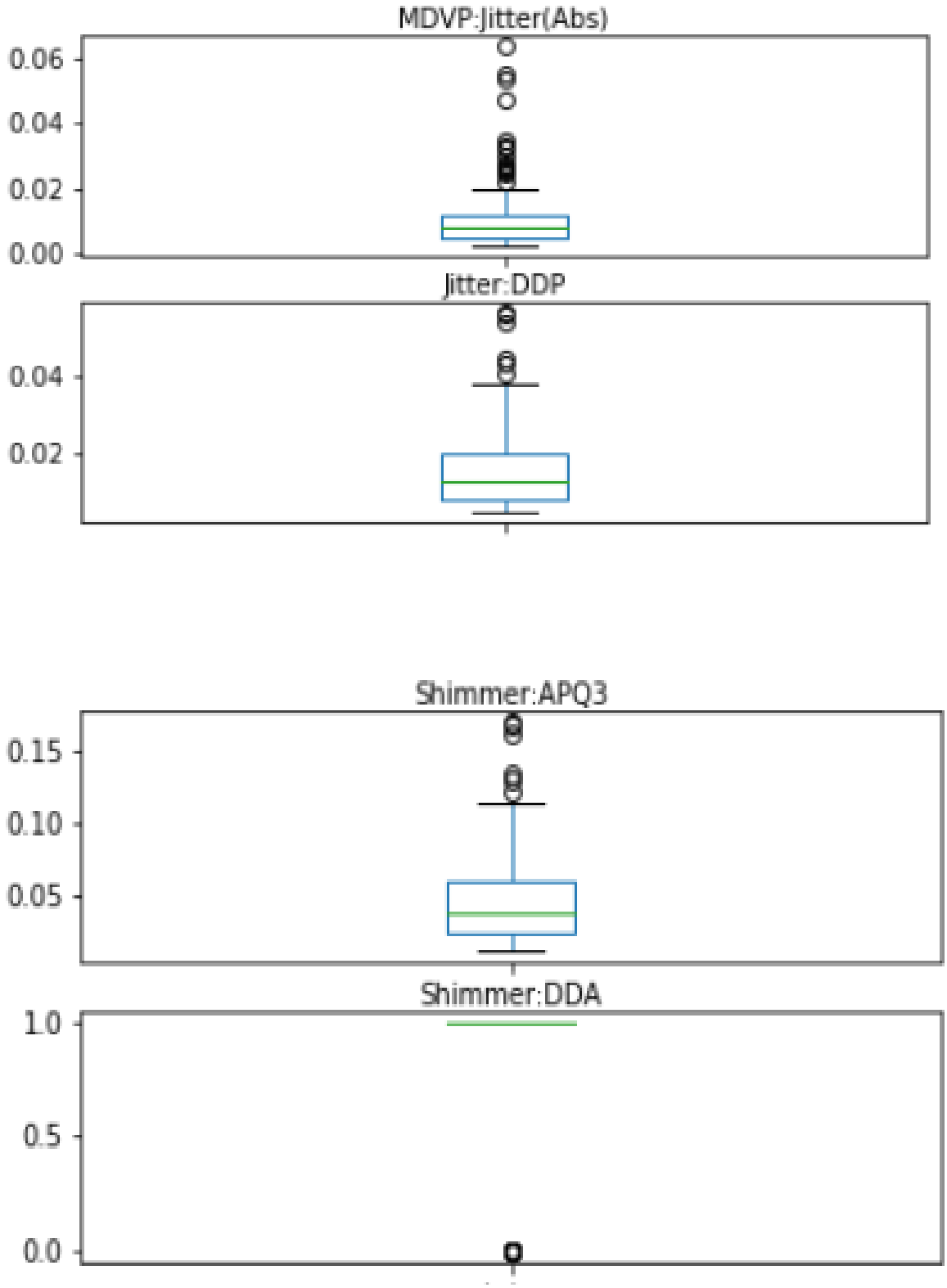
Feature Extraction:

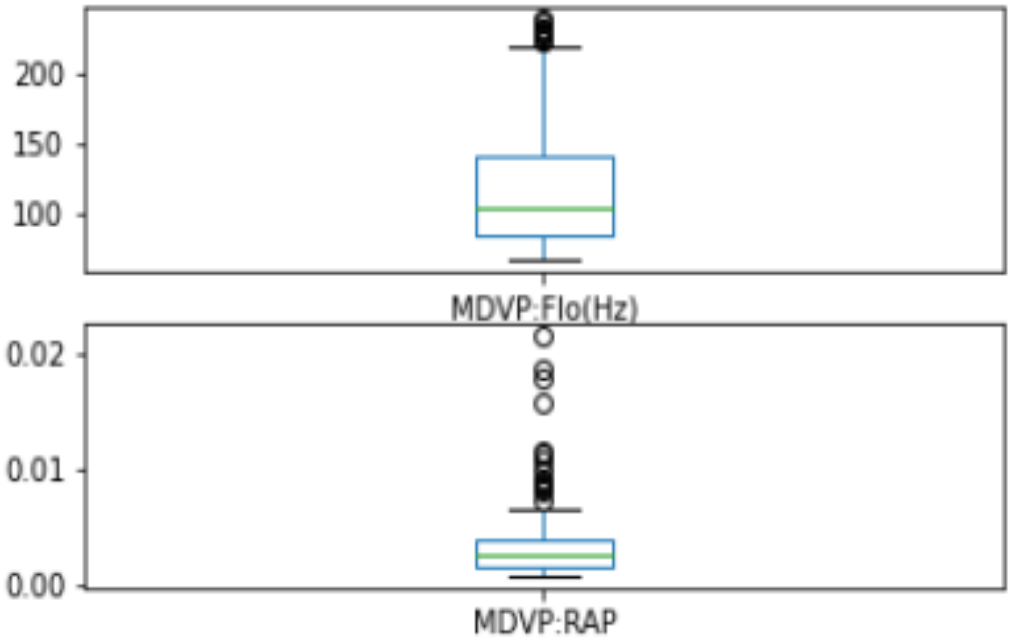


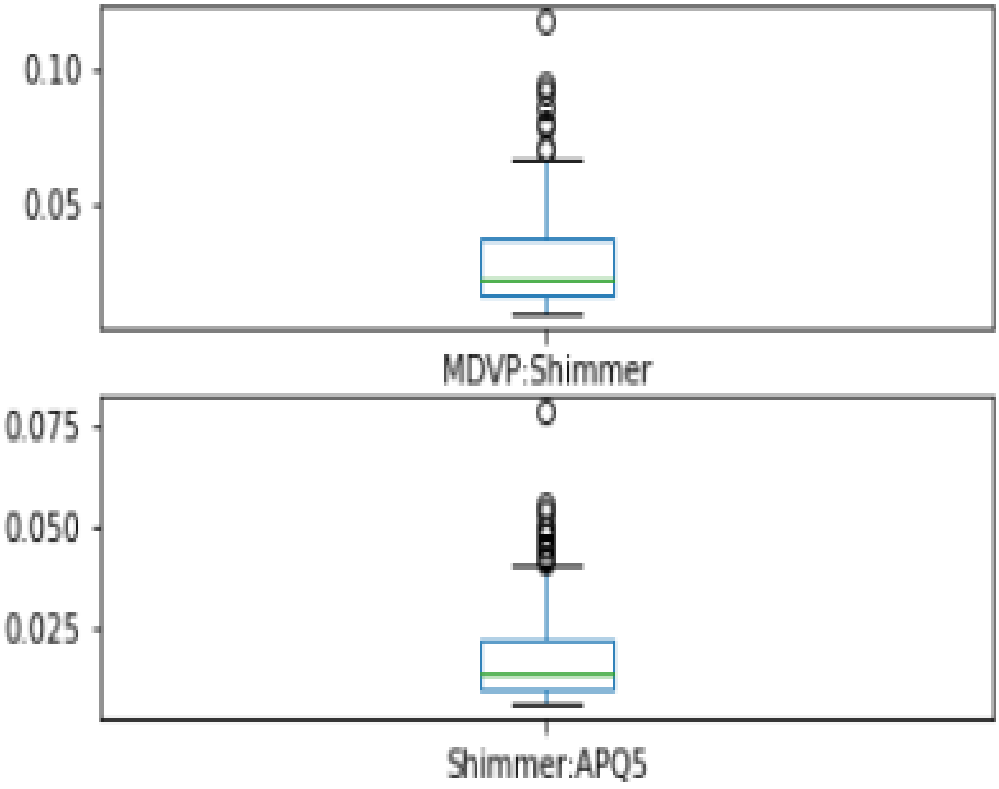


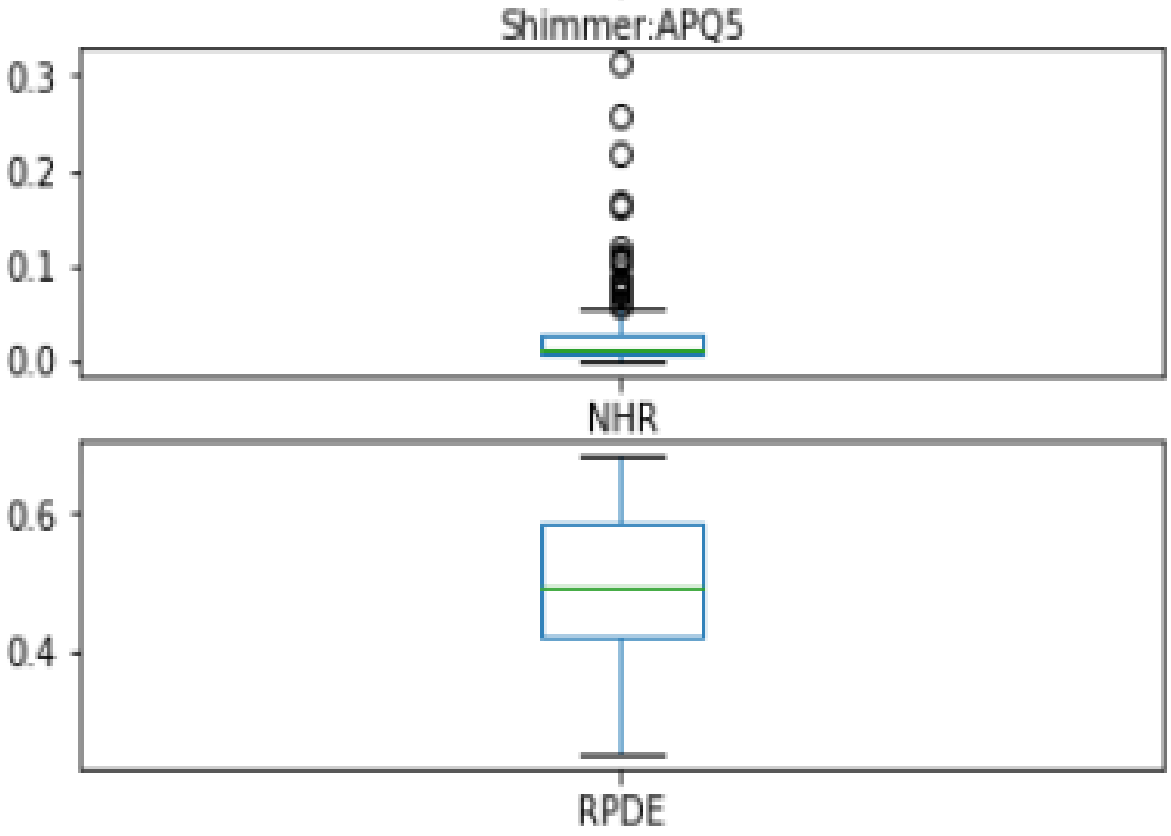

MDVP:Shimmer(dB)
0.10
0.05
MDVP:APQ
30
20
10
HNR

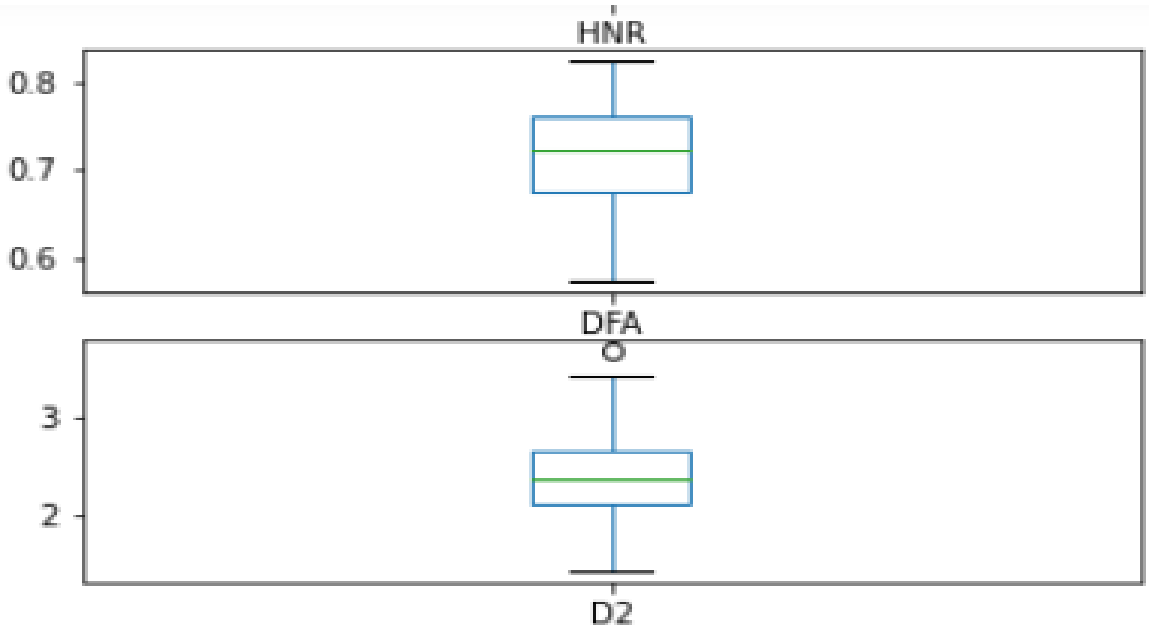

600
400
200
MDVP:Fhi(Hz)
0.0002
0.0001
0.0000
MDVP:Jitter(Abs)

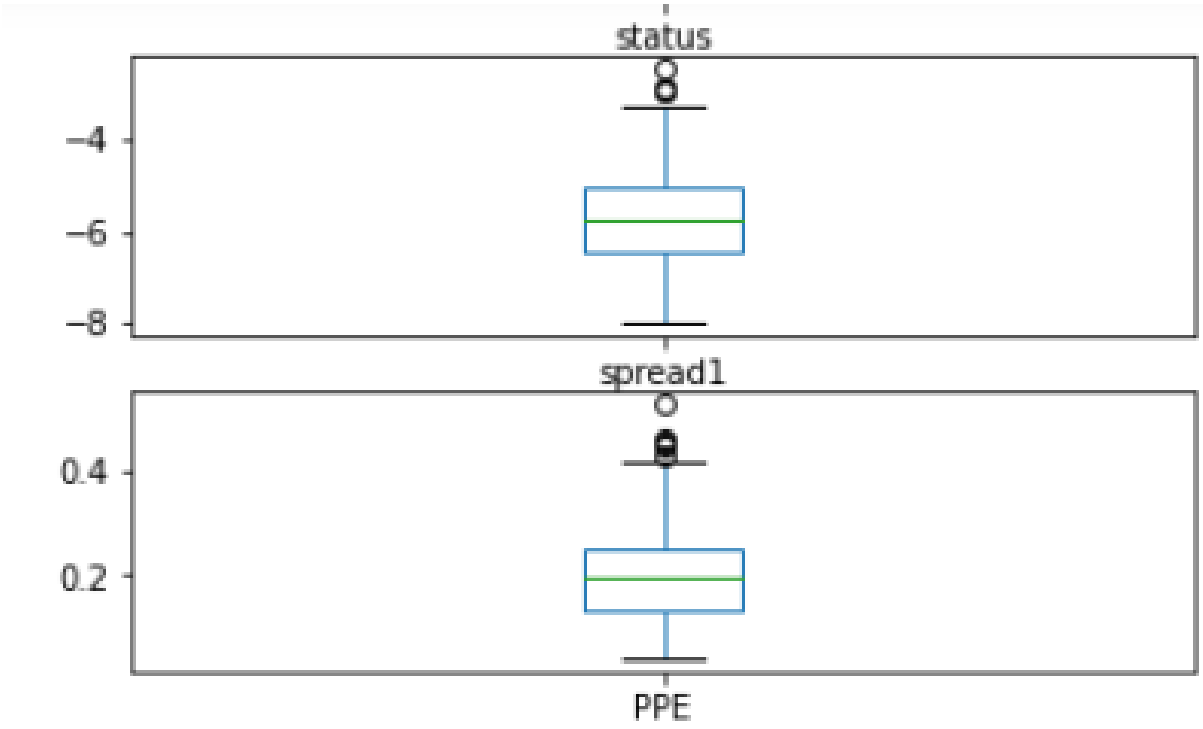

MDVP:Jitter(Abs)
0.06
0.04
0.02
0.00
Jitter:DDP
0.04
0.02
Shimmer:APQ3
0.15
0.10
0.05
Shimmer:DDA
1.0
0.5
0.0

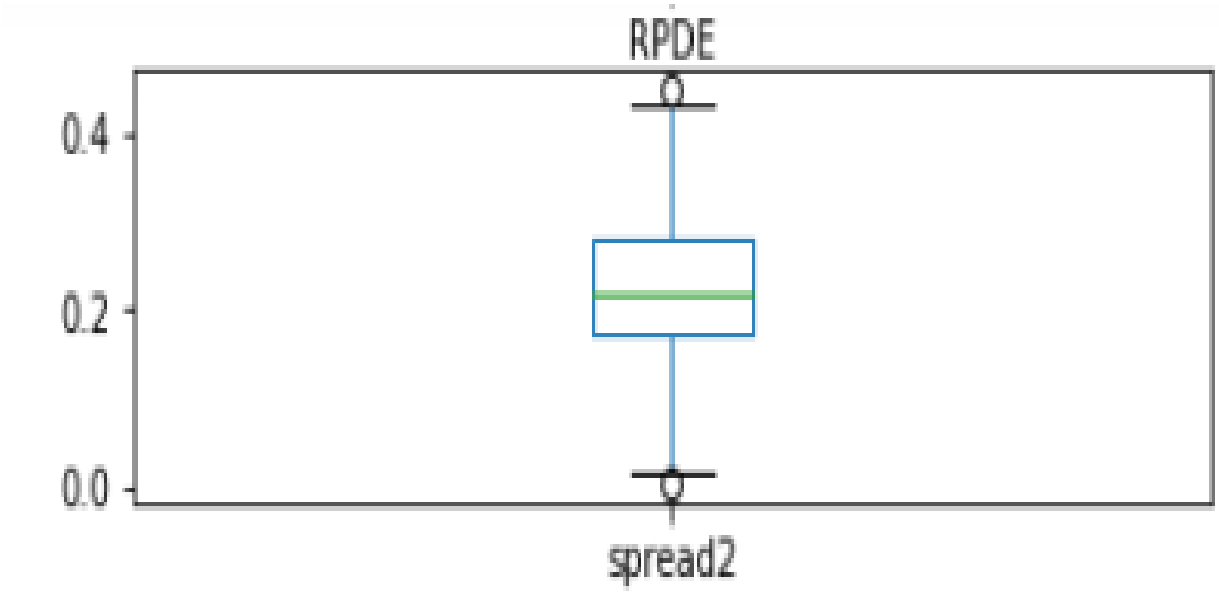

200
150
100
MDVP:Flo(Hz)
0.02
0.01
0.00
MDVP:RAP


0.10
0.05
MDVP:Shimmer
0.075
0.050
0.025
Shimmer:APQ5

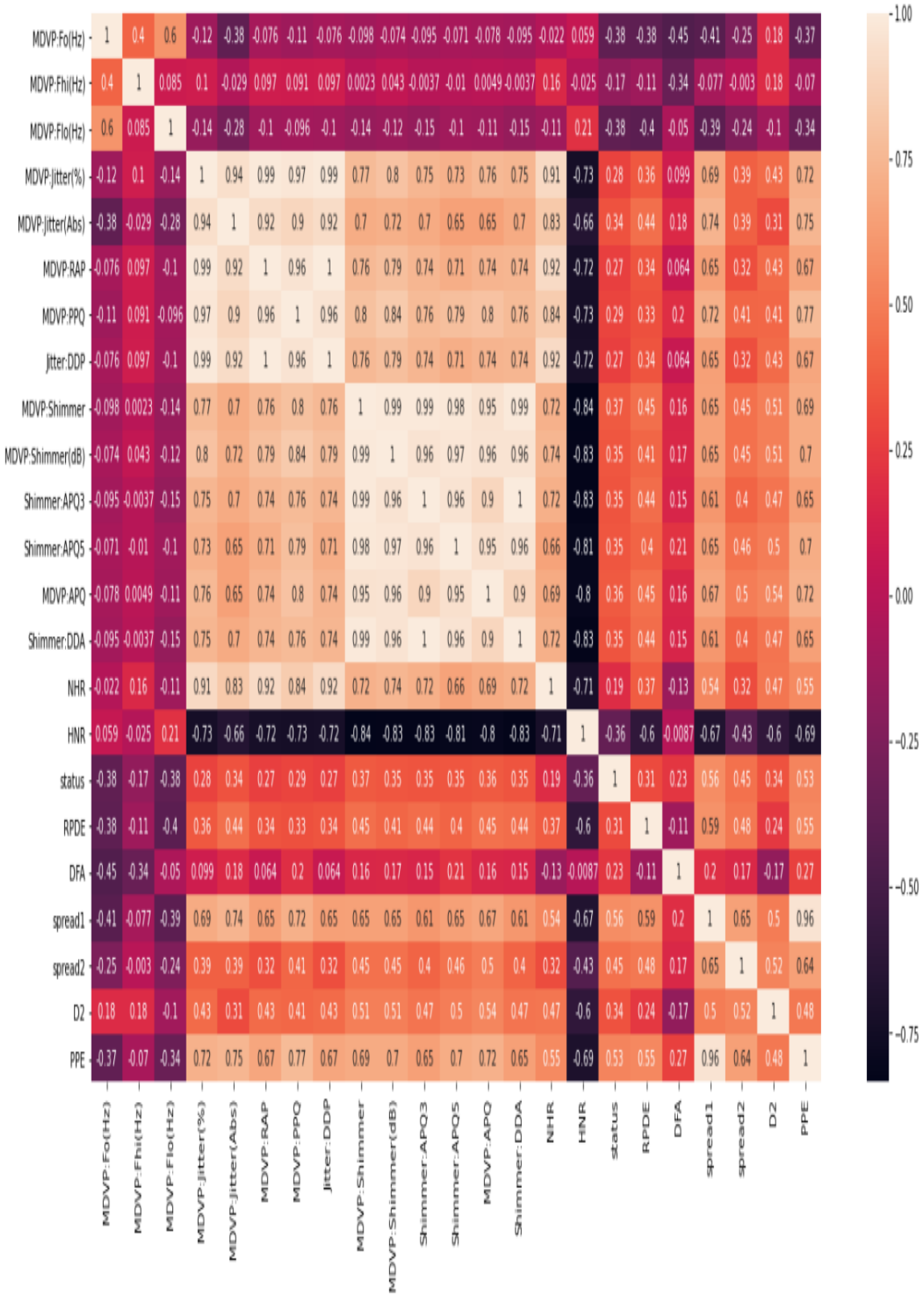

Shimmer:APQ5
0.3
0.2
0.1
0.0
NHR
0.6
0.4
RPDE


HNR
0.8
0.7
0.6
DFA
3
2
D2


status
−4
−6
−8
spread1
0.4
0.2
PPE


RPDE
0.4
0.2
0.0
spread2

Heat Map:


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
PPE
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

PREDICTING THE ACCURACY FROM DIFFERENT MODELS AND HEAT MAP REPRESENTATION.

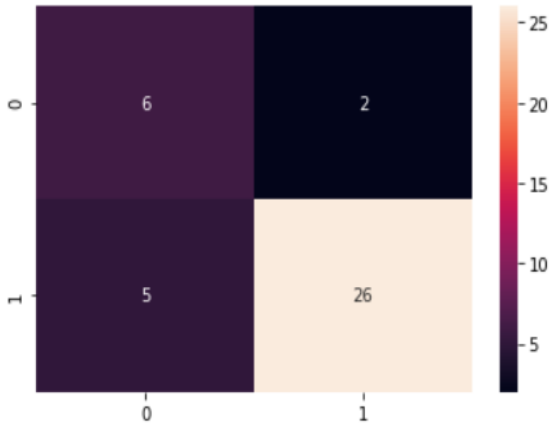
XGBOOST:

```
In [18]: model=XGBClassifier(eval_metric='mlogloss')
         model.fit(x_train,y_train)

         C:\Anaconda\lib\site-packages\xgboost\sklearn.py:1224: UserWarning: The use of label encoder in XGBClassifier is deprecated and
         will be removed in a future release. To remove this warning, do the following: 1) Pass option use_label_encoder=False when cons
         tructing XGBClassifier object; and 2) Encode your labels (y) as integers starting with 0, i.e. 0, 1, 2, ..., [num_class - 1].
           warnings.warn(label_encoder_deprecation_msg, UserWarning)

Out[18]: XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
                       colsample_bynode=1, colsample_bytree=1, enable_categorical=False,
                       eval_metric='mlogloss', gamma=0, gpu_id=-1, importance_type=None,
                       interaction_constraints='', learning_rate=0.300000012,
                       max_delta_step=0, max_depth=6, min_child_weight=1, missing=nan,
                       monotone_constraints='()', n_estimators=100, n_jobs=8,
                       num_parallel_tree=1, predictor='auto', random_state=0,
                       reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1,
                       tree_method='exact', validate_parameters=1, verbosity=None)

In [19]: y_pred=model.predict(x_test)
         print(accuracy_score(y_test, y_pred)*100)

         82.05128205128204
```

ACCURACY Of XGBOOST: 82%

```
In [20]: from sklearn.metrics import confusion_matrix
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)
```

```
Out[20]: <AxesSubplot:>
```



Confusion Matrix:

True Positive : 6(True Healthy), False positive: 2 (True Parkinson)

False Negative : 5(Predicted Healthy) , True Negative:26(Predicted Parkinson)

Support vector Machines: (SVM)

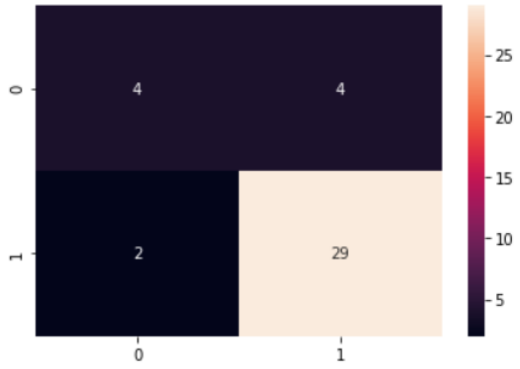
```
In [21]: from sklearn.svm import SVC
         model = SVC(kernel='linear')
         model.fit(x_train, y_train)
         y_pred=model.predict(x_test)
         print(accuracy_score(y_test, y_pred)*100)
```

```
84.61538461538461
```

Accuracy of the SVM: 84%

Confusion Matrix and heatmap:

```
In [22]: from sklearn.metrics import confusion_matrix
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)

Out[22]: <AxesSubplot:>
```



True Positive : 4(True Healthy), False positive: 4(True Parkinson)

False Negative : 2(Predicted Healthy) , True Negative:29(Predicted Parkinson)

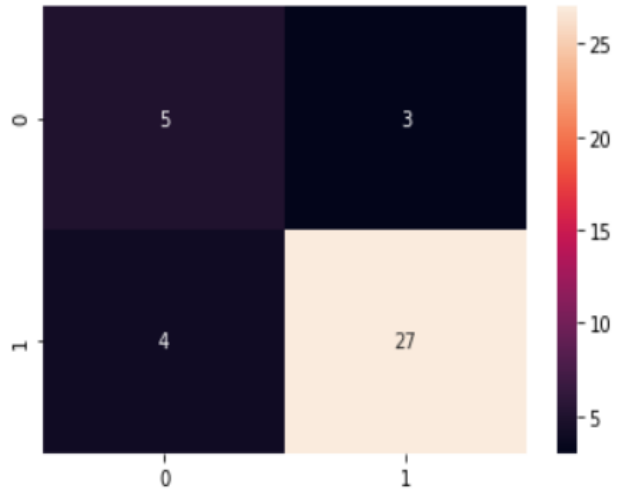
Random Forest:

```
In [23]: from sklearn.ensemble import RandomForestClassifier
         model= RandomForestClassifier(random_state=2)
         model.fit(x_train,y_train)
         y_pred=model.predict(x_test)
         model.score(x_test, y_test)

Out[23]: 0.8205128205128205
```

Accuracy of Random Forest: 82%

Confusion Matrix And heat map:

```
In [24]: from sklearn.metrics import confusion_matrix
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)

Out[24]: <AxesSubplot:>
```



True Positive : 5(True Healthy), False positive: 3(True Parkinson)

False Negative : 4(Predicted Healthy) , True Negative:27(Predicted Parkinson)

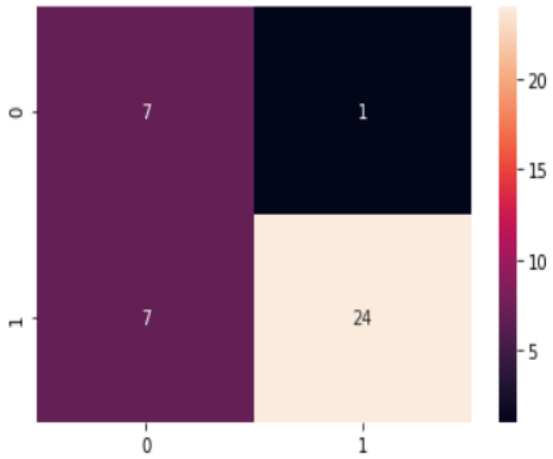
K-Nearest Neighbors:

```
In [25]: from sklearn.neighbors import KNeighborsClassifier
         model = KNeighborsClassifier(n_neighbors=8)
         model.fit(x_train, y_train)

Out[25]: KNeighborsClassifier(n_neighbors=8)

In [26]: model.score(x_test, y_test)

Out[26]: 0.7948717948717948
```

Accuracy of KNN: 79%

Confusion matrix and Heat map:

```
In [27]: from sklearn.metrics import confusion_matrix
         y_pred=model.predict(x_test)
         cm = confusion_matrix(y_test, y_pred)
         sns.heatmap(cm ,annot=True)

Out[27]: <AxesSubplot:>
```



True Positive : 7(True Healthy), False positive: 1 (True Parkinson)

False Negative : 7(Predicted Healthy) , True Negative:24(Predicted Parkinson)

PREDICTION MODEL BUILDING:

- This model building is used to predict the Parkinson's disease by providing the data values.
- Here i have given some input to predict and it predicted as Positive.

```
In [28]: input_data = (222.23600,231.34500,205.49500,0.00266,0.00001,0.00152,0.00144,0.00457,0.01643,0.14500,0.00867,0.01108,0.01200,0.02
         input_data_numpy = np.asarray(input_data)
         input_data_reshape = input_data_numpy.reshape(1,-1)
         std_data = scaler.transform(input_data_reshape)
         prediction = model.predict(std_data)
         print(prediction)

         if (prediction[0] == 0):
           print('The patient has Parkinson')
         else:
           print('The patient does not have Parkinson')

         [0]
         The patient has Parkinson
```

CONCLUSION:

Parkinson's disease affects the CNS of the brain and has yet no treatment unless it's detected early. Late detection leads to no treatment and loss of life. Thus, its early detection is significant. For early detection of the disease, we utilized various machine learning algorithms to detect Parkinson's disease. We checked our Parkinson disease data and found out that Support Vector Machines is the best Algorithm with accuracy of 84% to predict the onset of the disease which will enable early treatment and save a life

References:

1. Shahbakhi, M., Far, D. T., & Tahami, E. (2014). Speech Analysis for Diagnosis of Parkinson's Disease Using Genetic Algorithm and Support Vector Machine. Journal of Biomedical Science and Engineering, 2014

2. Yahia A, Laiali A. (2014). Detection of Parkinson Disease through Voice Signal Features. Journal of American Science 2014; 10 (10), 44-47.

3. Khan Academy. (2015). Movement signs and symptoms of Parkinson's disease | NCLEX-RN | Khan Academy. [online] Available at: https://www.youtube.com/watch?v=4qdD4Ny34cc [Accessed 11 Apr. 2018].

4. Pereira, C.R., Weber, S.A., Hook, C., Rosa, G.H. and Papa, J.P., 2016, October. Deep LearningAided Parkinson's Disease Diagnosis from Handwritten Dynamics. In Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on(pp. 340-346). IEEE.

5.Prechelt, L., 1998. Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Springer, Berlin, Heidelberg.

6. Samii A., J.G. Nutt, B.R. Ransom.Parkinson's disease. Lancet, 363 (2004), pp. 1783-1793

7.Geeks for Geeks- Data Preprocessing

8. Chen, A. H., Lin, C. H., & Cheng, C. H. New approaches to improve the performance of disease classification using nested–random forest and nested–support vector machine classifiers. Cancer, 2 (10509), 102.

9. Das, R. (2010). A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications, 37 (2), 1568-1572

10: Souillard-Mandar, W., Davis, R., Rudin, C., Au, R., Libon, D.J., Swenson, R., Price, C.C., Lamar, M. and Penney, D.L., 2016. Learning classification models of cognitive conditions from subtle behaviors in the digital Clock Drawing Test. Machine learning, 102(3), pp.393-441.

11. C. Lu, B. Deng, Y. Zhu, J. Wang and C. Liu, "Effect of DBS Targeting Striatum on beta Oscillations in Parkinson's Disease," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 502-507. doi: 10.23919/ChiCC.2019.8866389

12. Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding machine learning: From theory to algorithms. Cambridge university press.