Parkinson's Disease Prediction using Machine Learning and Data Analytics

Made by: Vaibhav

Email: vaibhavraopalri@gmail.com

April, 2021

DECLARATION

I, ‘Vaibhav’, student of ‘Bachelor of Engineering in Big Data Analytics’, session:2019-2023, Department of Computer Science and Engineering, Apex Institute of Technology, Chandigarh University, Punjab, hereby declare that the work presented in this Project Work entitled ‘Parkinson’s Disease Prediction using Machine Learning and Data Analytics’ is the outcome of my own bona fide work and is correct to the best of our knowledge and this work has been undertaken taking care of Engineering Ethics. It contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Vaibhav

Date: 17/4/21

Place: Chandigarh, Punjab

Table of Contents

ABSTRACT

Parkinson's disease is a disorder that affects small regions in the brain that control movement, posture and balance. It is a complex disease that has many different symptoms, so that not everyone with the condition suffers from the same problems. Parkinson's disease is named after the British doctor who wrote the first book about the disease, in 1817, that made it an easily recognized entity. Parkinson called it, "The Shaking palsy," or "paralysis agitans."

Parkinson's may affect anyone at any time. Well known identities diagnosed with the condition include Muhammad Ali, Michael J Fox, Janet Reno, Billy Graham, Bob Hoskins and the late Pope John Paul II and Donald Chipp

History of Parkinson's disease:

Although components of possible Parkinson's disease can be found in very early documents, James Parkinson wrote the first clear medical description in 1817. In the mid-1800s, Jean-Martin Charcot was particularly influential in refining and expanding this early description, as well as disseminating information about Parkinson's disease on a global scale.

Early medicines of Parkinson's illness depended on exact perception, and anticholinergic medications were utilized as ahead of schedule as the nineteenth century.

LIST OF FIGURES/IMAGES

LIST OF TABLES

1. INTRODUCTION

Parkinson's disease is a disorder that affects small regions in the brain that control movement, posture and balance

PD is common. It affects about 500,000- one million Americans, or about 1% of people over the age of 60. After Alzheimer's disease, it is the second most common progressive, neurological disorder in the US. In the state of Rhode Island alone, with a population of only one million people, there are an estimated 1-2,000 people with PD. Although there is a large amount of research on PD, we still don't know what causes it. And we even have some trouble diagnosing it at times.

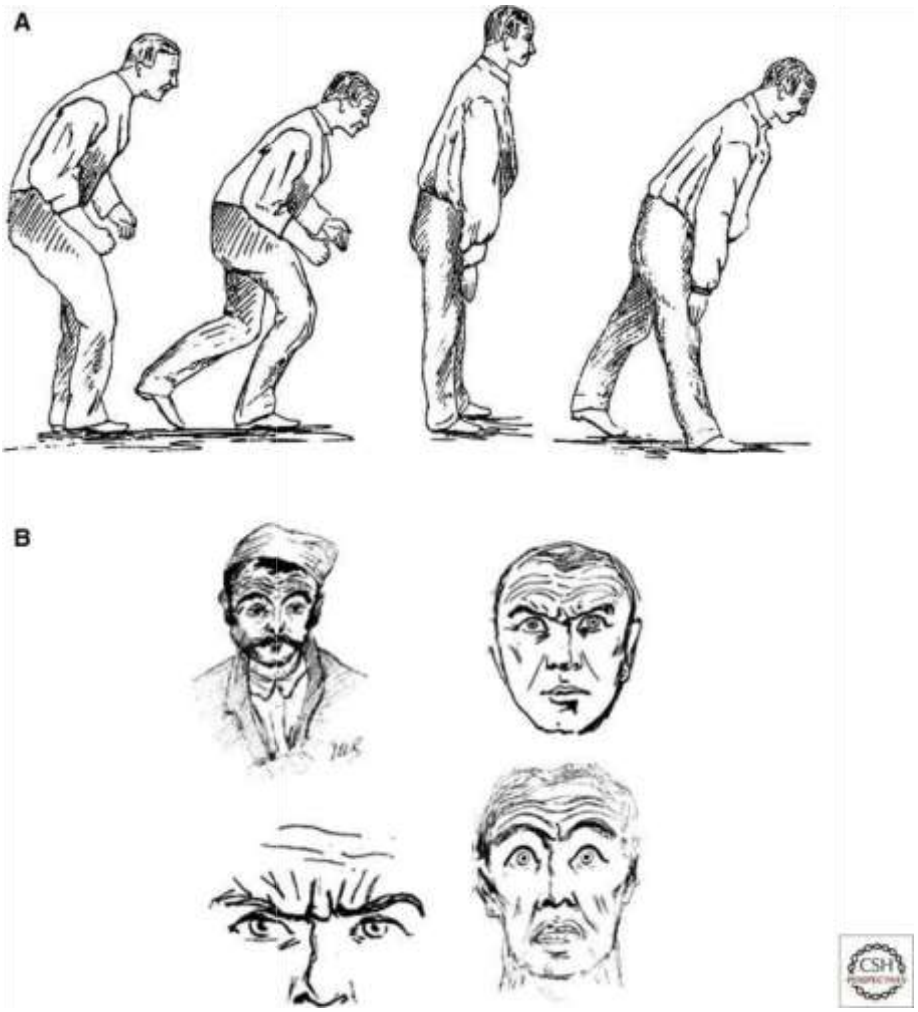


Figure 1

1.1 Problem Definition

Parkinson Disease is a Disease that cannot be detected through any kind of dedicated medical test. The revelation of dopaminergic shortages in Parkinson's sickness and the manufactured pathway of dopamine prompted the main human preliminaries of levodopa Further generally significant anatomical, biochemical, and physiological investigations distinguished extra pharmacological and neurosurgical focuses for Parkinson's infection and permit present day clinicians to offer a variety of treatments pointed toward improving capacity in this still serious sickness.

Since a vast majority of population suffers from this disease and faces consequences, hence forth it is very essential to find ways to predict as to what extent a person is suffering from Parkinson's disease.

Figure 2

1.2 Project Overview/Specifications

The aim of this project is to predict the presence of Parkinson's disease using machine learning algorithms and data analysis. The end product of this project is an application that creates an interface for a user to enter/input a compatible Parkinson disease dataset and predict as to what patients in that data set are suffering from Parkinson's disease.

PD is a neurological disease. It affects certain brain cells that help in controlling the movement and coordination of muscles in the body. It is associated with degeneration of the basal ganglia of the brain and a deficiency of the neurotransmitter dopamine. Dopamine is a hormone and neurotransmitter, a chemical that is generated by brain cell. It is used to send signals to other brain cells to control the muscle activity. PD causes, degeneration of dopamine in the brain cell which is unable to control the movement and activity of muscles.

This project is only compatible with the UCI Machine Learning repository.

1.3 Dataset Information:

Data Set Characteristics: Multivariate	Number of Instances: 197
Area: Life	Attribute Characteristics: Real
Number of Attributes: 23	Date Donated: 2008-06-26
Associated Tasks: Classification	Missing Values? N/A

This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals ("name" column).
The data is in ASCII CSV format. The rows of the CSV file contain an instance corresponding to one voice recording. There are around six recordings per patient; the name of the patient is identified in the first column.

1.4 Hardware Specifications

The minimum Hardware specifications required for this project are:

1. Processor:

   a. Intel i3 processor or above

   b. Any other processor with similar minimum specifications.

2 .Operating system specs:

   a. A 64-bit version of Microsoft Windows 8 or above (Windows 10 preferable)

3. RAM

   b. Minimum RAM requirement : 2 GB RAM

   Recommended RAM: 8 GB RAM

4. Memory SPACE:

a. 2.5 GB hard disk space, SSD recommended

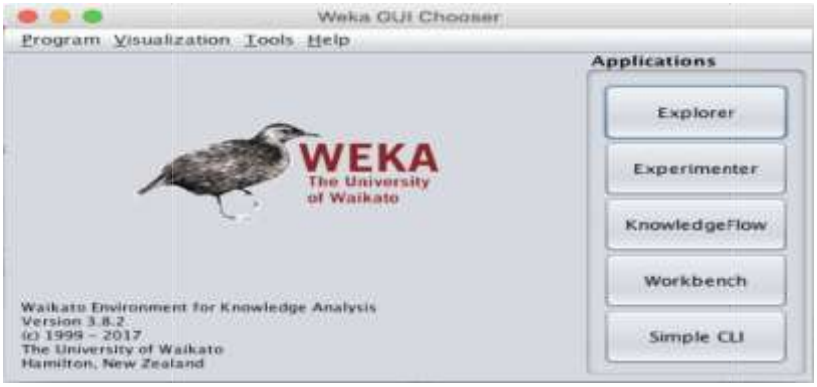
1.5 Software Specifications:

1) Computer having Python with version 3.8.9 or above.

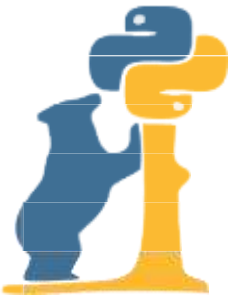
2) Python IDE environment e.g. (Spyder/pyCharm/Jupiter Lab etc.)

3) Pip must be in working condition for installing different libraries.

4) WEKA must installed in the System.



5) System must have following libraries: pandas, scipy,numpy, matplotlib, seaborn, seaborntensorflow, Tkinter, weka-python wrapper class.



2. LITERATURE SURVEY

All of the existing systems only comprise of research work with no application based interface compatibility.

A lot of research has been done to predict Parkinson's disease in a patient, but less work has been reported to predict its severity. These works have used various machine learning techniques.

According to the paper written by Udaya Kumar, Magesh kumar, by using data compression technique such as Discrete Cosine Transform (DCT) number of instances could be minimized, after data compression, attribute selection was done using several WEKA build in methods such as ChiSquared, GainRatio, Infogain after identifying the important attributes, they evaluated attributes one by one using stepwise regression. Based on the selected attributes they processed in WEKA by using cost sensitive classifier with various algorithms the classified result showed on an average 70%. . It can further be improved using ADABOOST.

Any of the existing systems does not comprise of any type of interactive platform with the user.

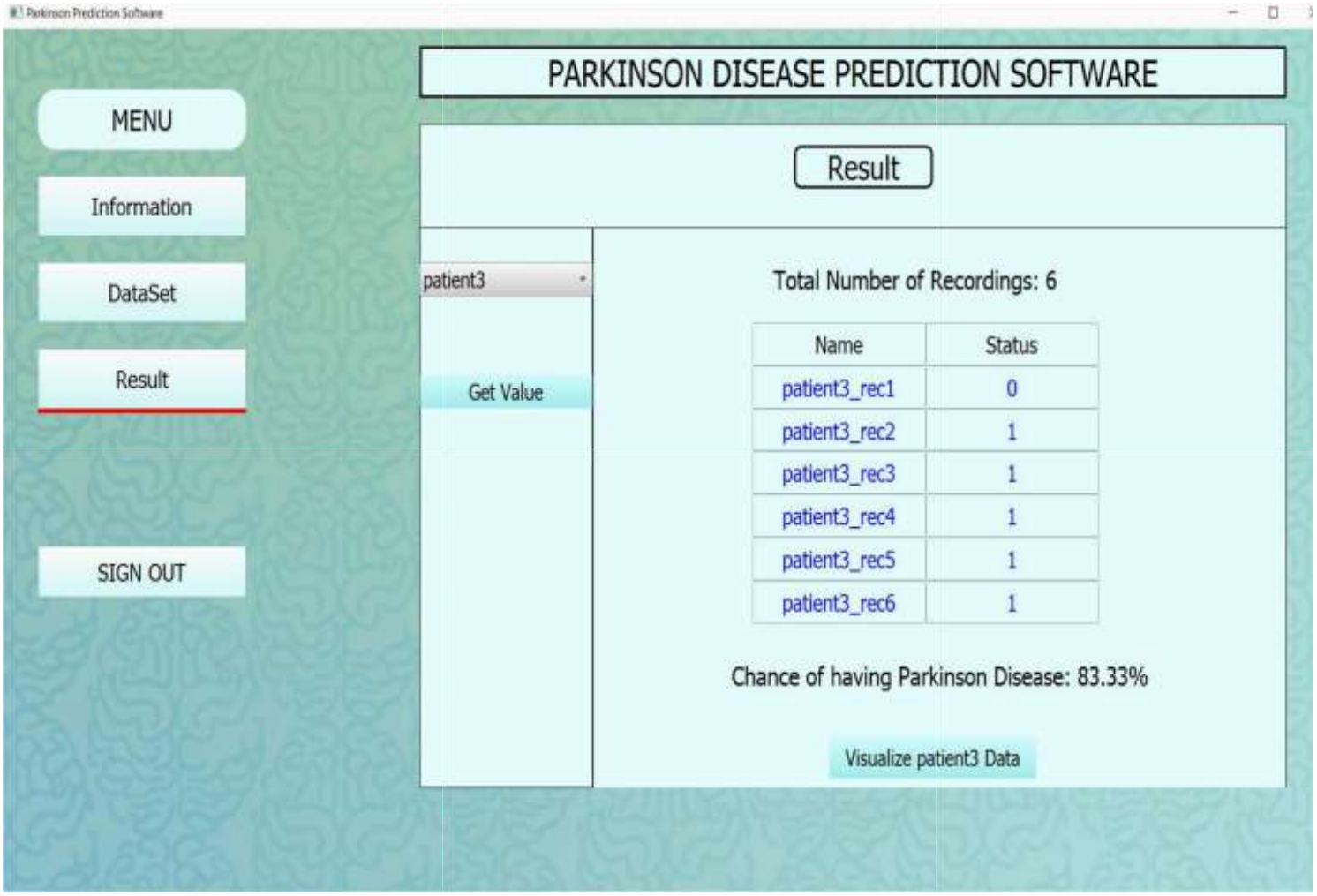
2.1 Existing System

2.2 Proposed System & Innovation

The proposal is to create an application based interface that visualizes the prediction of patients in the UCI Dataset. There are many algorithms to predict whether a person is having Parkinson's disease or not. Here, we will be using the concept of decision trees to classify between different attributes available and choose the appropriate attributes. To improve the efficiency of this algorithm, we will also use AdaBoost which will boost the algorithm's efficiency and create a model with its help.

2.2.1 Newness of Model

The newness of this model exists in the fact that the prediction model which is used for predicting the values of status of Parkinson of the testing dataset uses ADABOOST CLASSIFIER which is a modern day classifier with improved



modifications and it is more acceptable to a greater percentage of cross validation methods.

It also exists in the fact that the project focuses on a GUI based depiction that can be understood by anyone, as compared to other existing models which only focus on a CLI based approach which is very hard to understand.

Figure 3

2.3 SYSTEM DESIGN

2.3.1 Info

The Parkinson database used in this study is taken from the University of California at Irvine (UCI) machine learning repository .The dataset was created by Max Little of the University of Oxford, in collaboration with the National Centre for Voice and Speech,

Denver, Colorado, who recorded the speech signals. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD). Each column in the table is a particular voice measure, and each row corresponds one of 195 voice recording from these individuals. The main aim of the data is to discriminate healthy people from those with PD, according to —status column which is set to 0 for healthy and 1 for PD.

2.3.2 Identification and Finalization of Input Parameters

The attribute information is as follows:

Attribute	Type	Description
Name	Class	ASCII subject name and recording number
MDVP:Fo(Hz)	Numerical	Average vocal fundamental frequency
MDVP:Fhi(Hz)	Numerical	Maximum vocal fundamental frequency
MDVP:Flo(Hz)	Numerical	Minimum vocal fundamental frequency
MDVP:Jitter(%) MDVP:Jitter(Abs) MDVP:RAP MDVP:PPQ Jitter:DDP	Numerical	Several measures of variation in fundamental frequency

MDVP:Shimmer MDVP:Shimmer(dB) Shimmer:APQ3 Shimmer:APQ5 MDVP:APQ Shimmer:DDA	Numerical	Several measures of variation in amplitude
NHR HNR	Numerical	Two measures of ratio of noise to tonal components in the voice
RPDE D2	Numerical	Two nonlinear dynamical complexity measures
DFA	Numerical	Signal fractal scaling exponent
spread1 spread2 PPE	Numerical	Three nonlinear measures of fundamental frequency variation
Status	Numerical	Health status of the subject (one) - Parkinson's, (zero) - healthy

2.3.3 Critical Values in Decision Tree:

1. Critical Value for PPE: 0.104

   >0.104 □ check MDVP_Jitter

   <=0.104 □ Healthy

2. Critical value for MDVP_Jitter: 0.02

   >0.02 □ check DFA

   <=0.02 □ 50% chance of Parkinson.

3. Critical Value for DFA: 0.689

   >0.689 □ Parkinson

   <=0.689 □ check MDVP_Fo_Hz

4. Critical Value for MDVP_Fo_Hz: 117.878

   >117.878 □50% chance of Parkinson.

   <=117.878 □check Shimmer APQ3

5. Critical Value for Shimmer APQ3: 0.008

   >0.008 □ Healthy

   <=0.008 □ check NHR

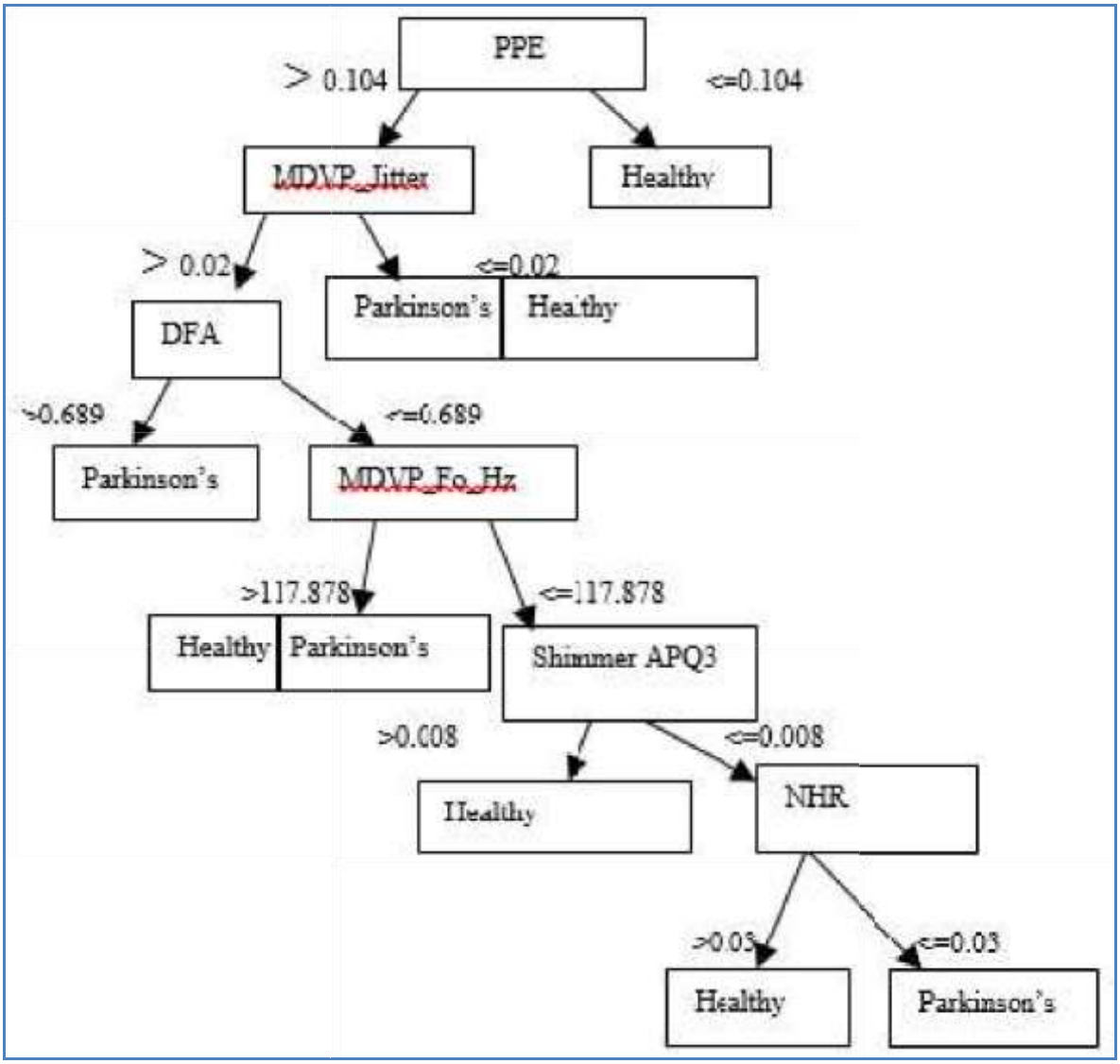
6. Critical Value for NHR: 0.03

   >0.03 □ Healthy

   <=0.03 □Parkinson

2.3.4 ADABOOST CLASSIFIER:

Boosting refers to a general and provably effective method of producing a very accurate prediction rule by combining rough and moderately inaccurate rules of thumb. Boosting has its roots in a theoretical framework for studying machine learning called the —PAC learning model. The AdaBoost algorithm, introduced in 1995 by Freund and Schapire , which solved many of the


PPE
> 0.104
<=0.104
MDVP_Jitter
Healthy
> 0.02
<=0.02
Parkinson's
Healthy
DFA
>0.689
<=0.689
Parkinson's
MDVP_Fo_Hz
>117.878
<=117.878
Healthy
Parkinson's
Shimmer APQ3
>0.008
<=0.008
Healthy
NHR
>0.03
<=0.03
Healthy
Parkinson's

practical difficulties of the earlier boosting algorithms (which came up with the first provable polynomial-time in 1989).

If we want to predict which person has Parkinson disease or not based on the symptoms, we can get a good prediction using Ada-Boost classifier. The algorithm takes as input ($x1$,$y1$),….(xm ,ym) where each xi belongs to some domain or instance space X and each level yi in some level set Y.

In most cases, we assume Y = {-1, +1}. AdaBoost calls a given weak or base learning algorithm repeatedly in a series of rounds t = 1… T.

The algorithm will maintain a distribution or set of weights over the training set. The weight of this distribution on training example i on round t is denoted by Dt

(i). At first stage all weights set equally, but on each round, the weights of misclassified examples are increased so that the weak learners is forced to focus on hard examples in the training set.

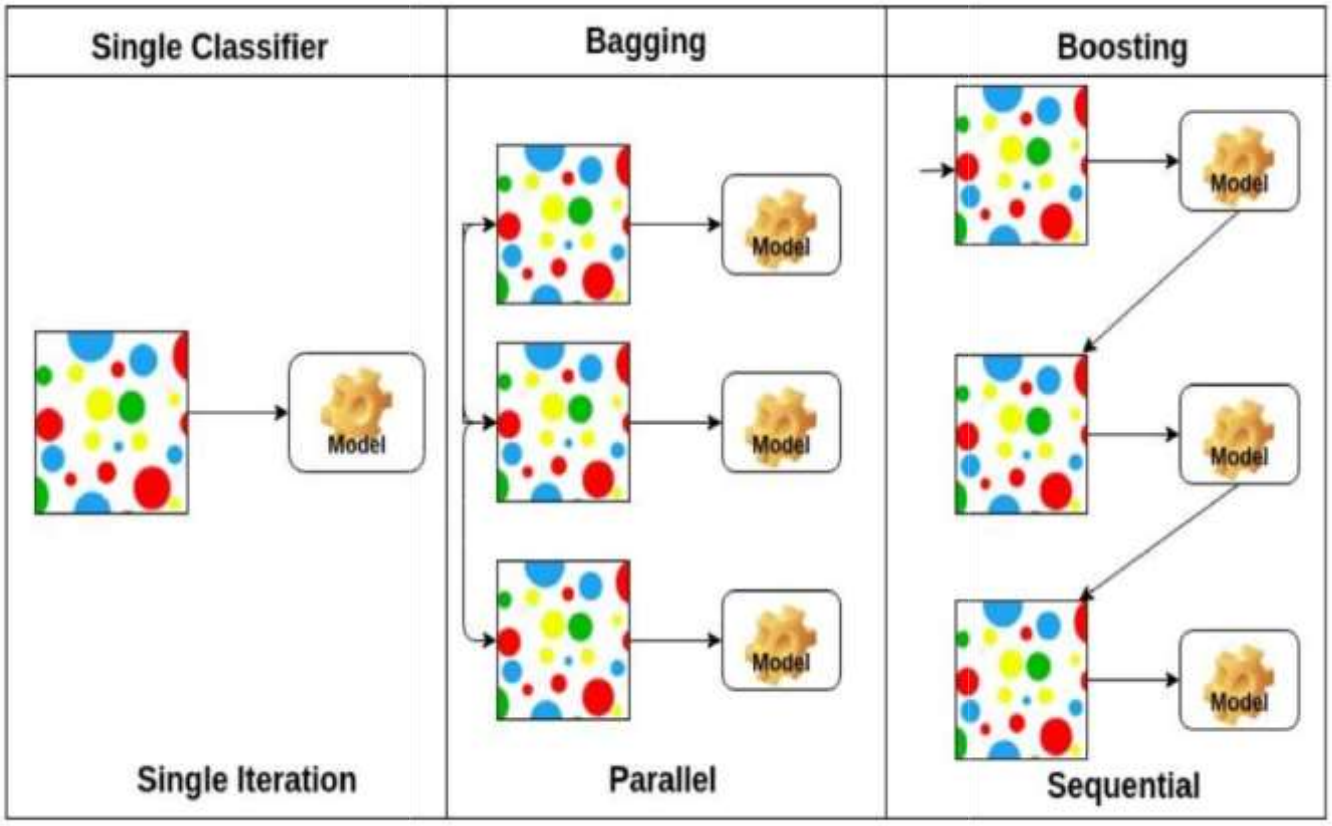


Figure 4

The weak learner's job is to find out the weak hypothesis

h_t : X -> {-1, +1} appropriate for the distribution D_t.

The goodness of the weak hypothesis is measured by its error:

Here the error is measured with respect to the distribution D_t , on which the weak learner was trained.

Given: $(x_1 , y_1),\ldots\ldots, (x_m , y_m)$ where $x_i \in X , y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$
For t … T:
-Train weak learners using distribution D_t
-Get weak hypothesis h_t: X-> {-1, +1} with error
$\epsilon_t = \Pr_{i\sim D_t}[h_t(x_i) \neq y_i]$
-Choose $\alpha_t = ½ \ln (\frac{1-e_t}{e_t})$
-Update:

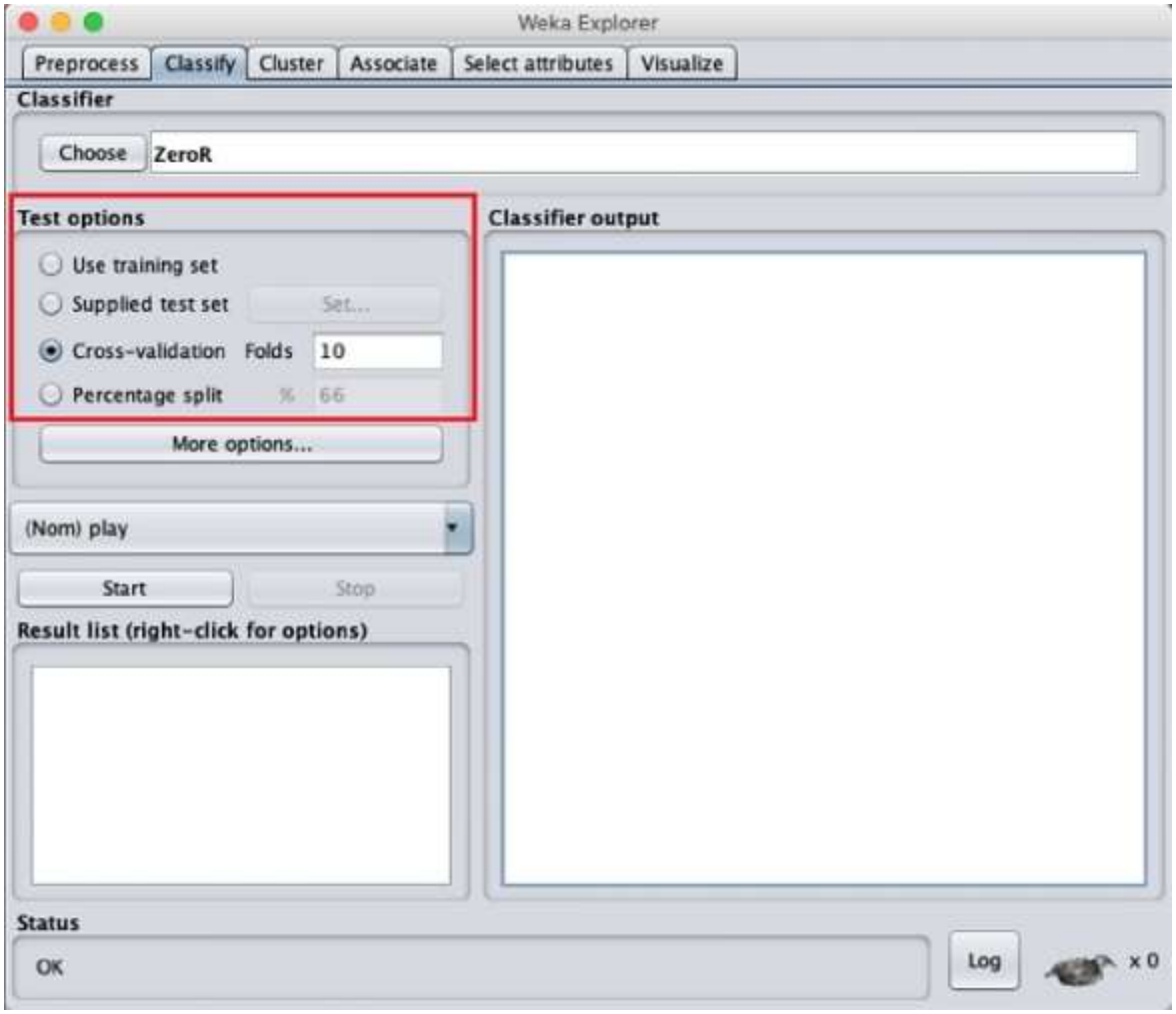
$$D_{t+1} = \frac{D_t(i)}{Z_t} \begin{cases} e^{-at} & if\ h_t(x_i) = y_i \\ e^{at} & if\ h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

Where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution.
Output of the final hypothesis:
$H(x) = \text{sign}(\sum_{t=1}^{T} \alpha_t h_t(x))$

2.3.5 USE OF WEKA AND PANDAS FOR DATA CLEANING AND ANALYSIS:

We use Weka so as to have a visual depiction of the data and to make a model of the dataset using ADABOOST CLASSIFER using the Pandas Library in Python.


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) play
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

Pandas:


Pandas

```
from scipy.io import arff
import pandas as pd

data = arff.loadarff("E:/Minor Project/parkin167temp.arff")
df = pd.DataFrame(data[0])
print(df)
```

```
In [1]: runfile('E:/Minor Project/Model/model.py', wdir='E:/Minor Project/Model')
                  name  MDVP:Fo(Hz)  MDVP:Fhi(Hz)  ...    spread2        D2       PPE
0    b'phon_R01_S01_1'      119.992       157.302  ...  0.266482  2.301442  0.284654
1    b'phon_R01_S01_2'      122.400       148.650  ...  0.335590  2.486855  0.368674
2    b'phon_R01_S01_3'      116.682       131.111  ...  0.311173  2.342259  0.332634
3    b'phon_R01_S01_4'      116.676       137.871  ...  0.334147  2.405554  0.368975
4    b'phon_R01_S01_5'      116.014       141.781  ...  0.234513  2.332180  0.410335
..                 ...          ...           ...  ...       ...       ...       ...
161  b'phon_R01_S34_4'      202.805       231.508  ...  0.213353  2.470746  0.189032
162  b'phon_R01_S34_5'      202.544       241.350  ...  0.220617  2.576563  0.159777
163  b'phon_R01_S34_6'      223.361       263.872  ...  0.345238  2.840556  0.232861
164  b'phon_R01_S35_1'      169.774       191.759  ...  0.414758  3.413649  0.457533
165  b'phon_R01_S35_2'      183.520       216.814  ...  0.355736  3.142364  0.336085

[166 rows x 24 columns]
```

Figure 5

2.3.6 CREATION OF INTERFACE USING PYTHON

After having a prediction model, we then use python to create an application based interface where we can see the results of the patients in the dataset and we visualize their analysis

Figure 6

3. PROBLEM FORMULATION

As of now, there are plenty of research works that are going on Parkinson's Disease(PD) which is just on the borderline of being the second most common disease in the world and the number of cases only seem to grow day by day. This situation leads us to create and build a prediction system for Parkinson's disease.

Parkinson's Symptoms:

The presentation of symptoms varies greatly between individuals diagnosed and no two people will be affected in the same way. The provisional medical diagnosis is based on symptoms because there is no definitive medical test or radiological procedure which diagnoses Parkinson's. The diagnostic criterion is composed of four cardinal symptoms which are:

- Tremor
- Bradykinesia
- Muscle rigidity
- Postural instability
- Festination of speech

Tremor is related to an imbalance of neurotransmitters, dopamine and acetylcholine, for this reason, tremor may be the least responsive symptom to dopamine replacement therapy. A classic tremor presentation of Parkinson's involves the thumb and first finger and is referred to as 'pill rolling'.

Bradykinesia affects all activities of daily living, walking, talking, swallowing and speaking. In the eyes and face it presents as a decreased blink rate and a lack of facial expression. The term used to describe slowness of thought experienced by people with Parkinson's is bradyphrenia.

Muscle rigidity is commonly present in the wrist, shoulder and neck. It may also manifest as a slightly flexed elbow on the affected side. Early reports of a painful shoulder are associated with increased muscle rigidity and tone. The pain

associated with Parkinson's is often underestimated and reported, and is usually associated with muscle rigidity.

Postural instability and gait disturbances often develop later in the progression of the condition. If a loss of postural reflexes and resulting falls occur early, it is not suggestive of typical Parkinson's. In early Parkinson's, the posture may show a slight flexion of the neck or trunk with a slight lean to one side. Gait changes include reduced arm swing (unilateral) and shortened stride height and length which may lead to shuffling.

There are some non- motor symptoms as well, along with the motor symptoms

- Fatigue
- Swallowing changes
- Depression
- Anosmia
- Constipation
- Micrographia
- Anxiety
- Sweating

Parkinson Disease can have a profound effect on speech and voice. Although symptoms vary widely from patient to patient, the speech symptoms most commonly demonstrated by patients with PD are reduced vocal loudness, monopitch, disruptions of voice quality, and abnormally fast rate speech.

This cluster of speech symptoms is often termed Hypokinetic Dysarthria. The most common symptom of Hypokinetic Dysarthria is Hypophonia, or reduced vocal loudness. Patients demonstrating this symptom may be unaware of the volume at which they are speaking and may be require frequent requests to speak louder.

The symptoms can be very evident and is usually mild at the beginning and then get more complex and the functionality lost varies on several conditions. The list of signs and symptoms mentioned in various sources for Hypokinetic Dysarthria includes the 7 symptoms listed below:

- Hoarse voice
- Breath voice
- Coarse voice
- Tremulous voice
- Single pitched voice
- Monotone voice
- Sudden pitch changes

Though the health care providers could diagnose the presence of Parkinson's disease based on the symptoms by the physical examination, the assess ability of the symptoms becomes difficult more particularly in case of elderly people. As the illness progresses the signs like tremor, walking problem and speech variations becomes clearer. The main point is that the diagnosis must concentrate on ruling out the other ailments that share the similar symptoms.

The signs that need to be looked for are:

- Slow opening and inadequate closing of the vocal folds
- Slows down voluntary movements

4. REPRESENTATION OF THE MODEL

An appropriate pictorial representation of how the framework of the model is embedded is essential for a good understanding of the project.

4.1 Phase diagram

Shown below is a flow chart depiction of how the things take place in the project model. :

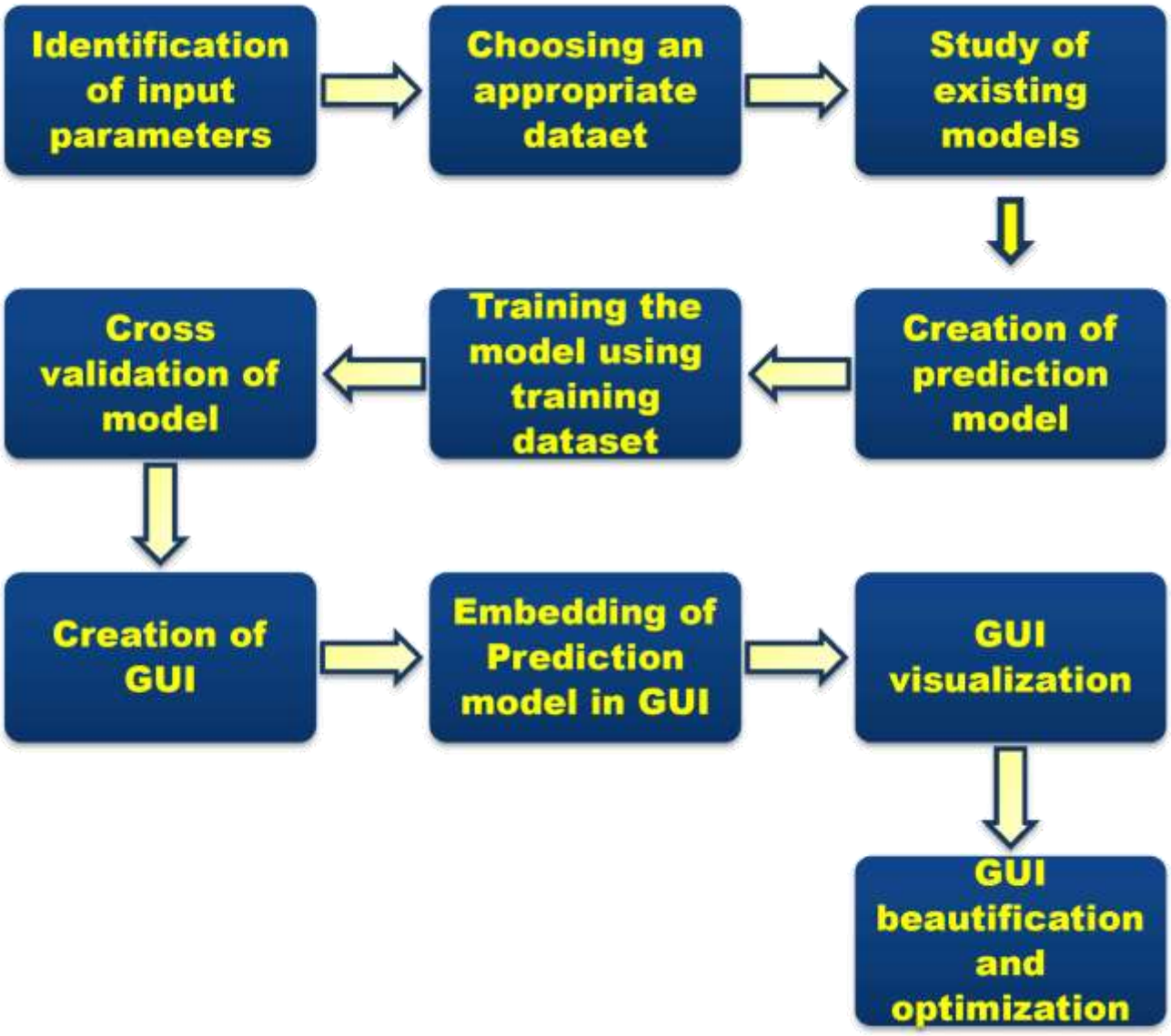


Figure 7

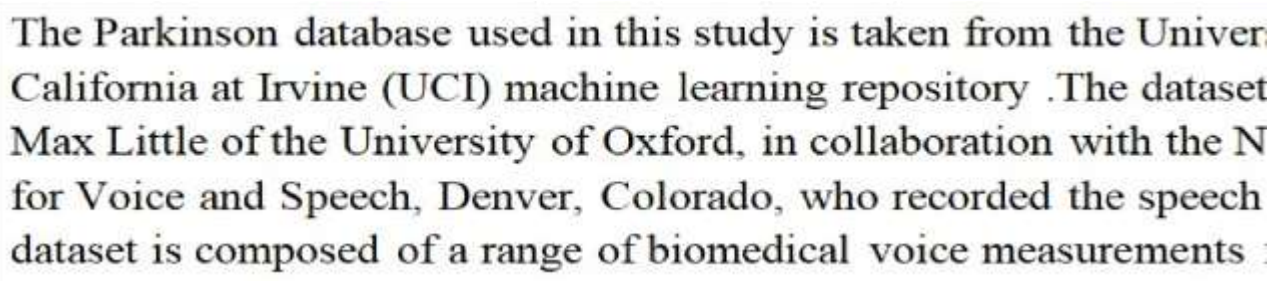
5. METHODOLOGY

Methodology describes the structure and framework of how the model building takes place.

5.1 Framework and Modelling

The following methodology will be followed to achieve the objectives defined for proposed research work:

1. Detailed study of the UCI Machine Repository's Parkinson dataset will be done.
2. Identification of final parameters to be used for prediction will be done.
3. Installation of Python IDLE'S and required libraries will be done.
4. AdaBoost Classifier will be used along with weka for making a model.
5. The UCI dataset will be bifurcated into training set and testing set.
6. A model will be created in Python which will use the AdaBoost classifier to make predictions. The model will be trained using our training data set.
7. A python based application GUI interface will be made to display patient's data and visualize it.
8. The trained model will be deployed within the GUI code and it will now be used to predict Parkinson status for the testing dataset.
9. The GUI code is improved in terms of presentability, beautification and optimization.


The Parkinson database used in this study is taken from the Univer
California at Irvine (UCI) machine learning repository .The dataset
Max Little of the University of Oxford, in collaboration with the N
for Voice and Speech, Denver, Colorado, who recorded the speech
dataset is composed of a range of biomedical voice measurements

measure, and each row corresponds one of 195 voice recording from these individuals (lnamel column). The main aim of the data is to discriminate healthy people from those with PD, according to —statusl column which is set to 0 for healthy and 1 for PD

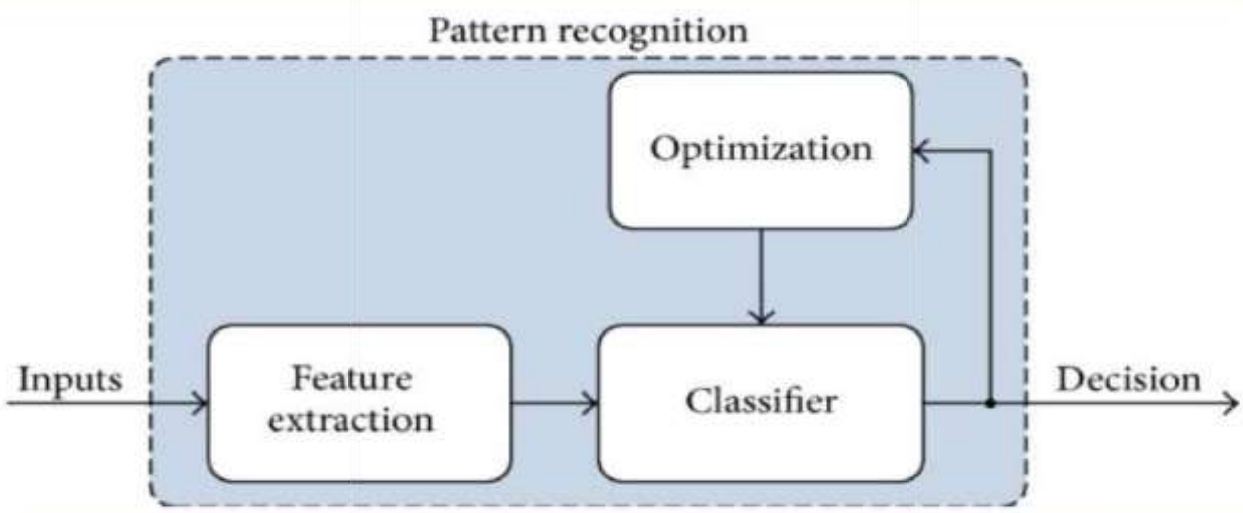


Figure 1

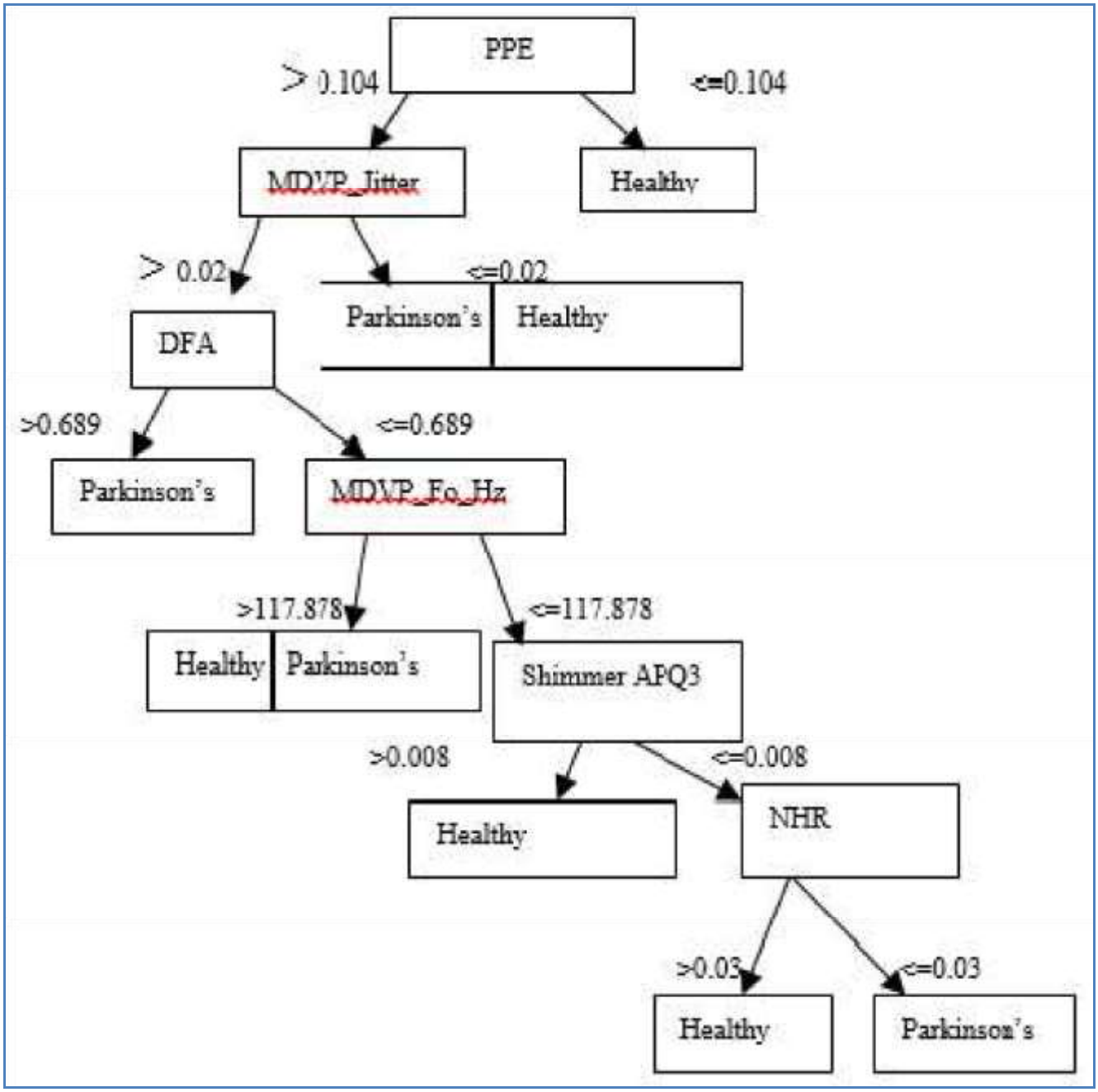
The block diagram of the pattern diagnosis concept.

5.2 Attribute Tree Hi erarchy :

There is some kind of deviation in voice measurement - it splits Jitter, shimmer and Noise Harmonic Ratio (NHR). Jitter and Shimmer is a variation of amplitude and frequency that reflect some specific irregularities in voice. In Fig. 2, the first node Pitch Period Entropy (PPE) follows two decisions. If the measured PPE value is <=0.104, then the people are consider as healthy.

If it is greater than > 0.104, it undergoes next decision MDVP_Jitter. In the second decision if the value is <=0.02, then the people are healthy with 50% probability and if it is >0.02, it goes to next condition Detrented Fluctuation Analysis (DFA).

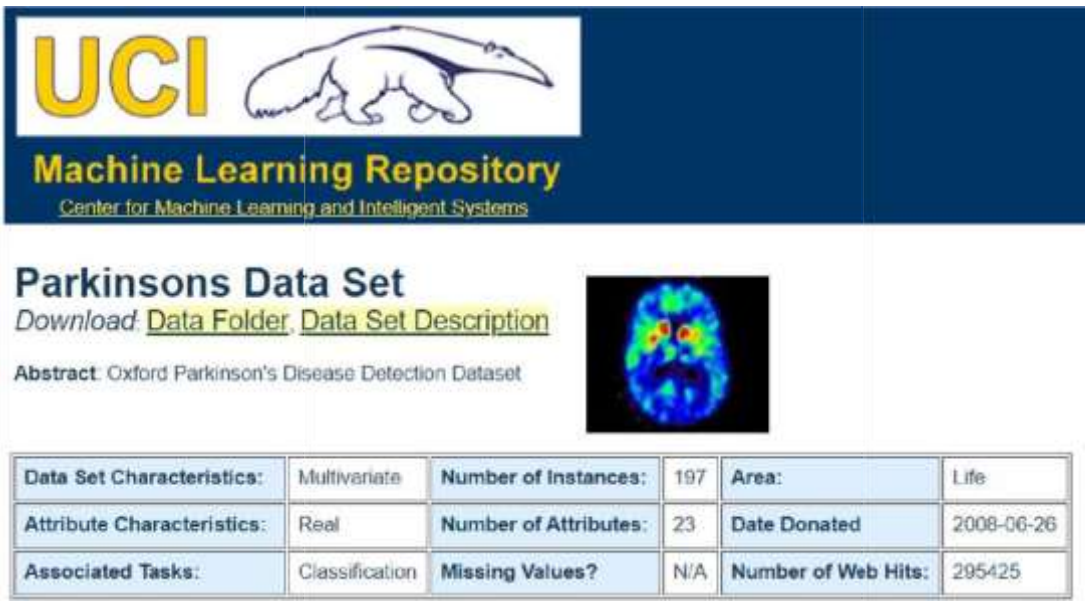
If DFA value is >0.682, then the people are affected by PD. If it is the procedure is repeated in other nodes RPDE, shimmer APQ3 ar



6. TENTATIVE PLAN FOR PROPOSED WORK

1. Collecting the Dataset:

   This step requires selecting an appropriate dataset to work upon having enough instances for a successful prediction model.



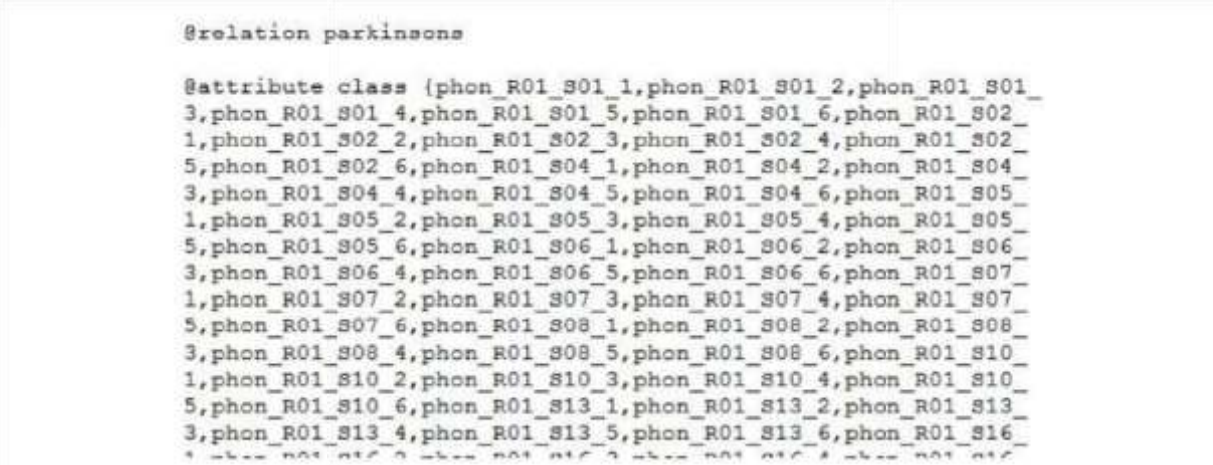
Data Set Characteristics:	Multivariate	**Number of Instances:**	197	**Area:**	Life
Attribute Characteristics:	Real	**Number of Attributes:**	23	**Date Donated**	2008-06-26
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	295425

2. Studying the Dataset:

   In this step, we deeply analyze the dataset that is given and connect it to the practical problem and study it deeply so as to understand the significance of all the attributes.

3. Preprocessing and data cleaning:

   In this process we pre process and clean the data so as to remove any inconsistencies.



```
@relation parkinsons

@attribute class {phon_R01_S01_1,phon_R01_S01_2,phon_R01_S01_
3,phon_R01_S01_4,phon_R01_S01_5,phon_R01_S01_6,phon_R01_S02_
1,phon_R01_S02_2,phon_R01_S02_3,phon_R01_S02_4,phon_R01_S02_
5,phon_R01_S02_6,phon_R01_S04_1,phon_R01_S04_2,phon_R01_S04_
3,phon_R01_S04_4,phon_R01_S04_5,phon_R01_S04_6,phon_R01_S05_
1,phon_R01_S05_2,phon_R01_S05_3,phon_R01_S05_4,phon_R01_S05_
5,phon_R01_S05_6,phon_R01_S06_1,phon_R01_S06_2,phon_R01_S06_
3,phon_R01_S06_4,phon_R01_S06_5,phon_R01_S06_6,phon_R01_S07_
1,phon_R01_S07_2,phon_R01_S07_3,phon_R01_S07_4,phon_R01_S07_
5,phon_R01_S07_6,phon_R01_S08_1,phon_R01_S08_2,phon_R01_S08_
3,phon_R01_S08_4,phon_R01_S08_5,phon_R01_S08_6,phon_R01_S10_
1,phon_R01_S10_2,phon_R01_S10_3,phon_R01_S10_4,phon_R01_S10_
5,phon_R01_S10_6,phon_R01_S13_1,phon_R01_S13_2,phon_R01_S13_
3,phon_R01_S13_4,phon_R01_S13_5,phon_R01_S13_6,phon_R01_S16_
```

4. Creation of a model Using ADABOOST CLASSIFIER FOR PREDICTION

   In this step we use the ADABOOST ML Algorithm to make a model of the given dataset so as to make predictions.

5. Training and Improvement:

   In this step we train the model so as to improve its prediction accuracy.

6. Apply the MODEL:

   Here, we apply the model by using testing data set and see the test results.

7. Creating a python interface for GUI depiction.

   In this step we create a friendly interface where the test results of the cases can be checked and the accuracy can be seen.

8. GUI improvement:

   In the final step we analyze for any improvements that can be done on the interface.

7. OTHER EXISTING APPROACHES

After evaluating the proposed algorithm's performance against the other relevant approaches, it is also essential that its accuracy, sensitivity and specificity are also tested. These are calculated based on the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) measures.

Sensitivity= TP (TP+FN) * 100

Specificity= TN (TN+FP)*100

Accuracy= (TP+TN) (TP+TN+FP+FN)*100

The accuracy, sensitivity and specificity values of classifiers are shown in Table 2.

In the proposed system, the classifier models are chosen in such a way that they have consistency and reliability in their performance, as these models are deployed in the healthcare domain where consistency and accuracy of prediction are of prime factors.

Population (Million)	Decision Tree			SVM			ANN		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
0.5	82.31	85	10	84.15	87	28	86.05	89	3
1.0	86	86	23	85.88	87	18	88	88	10
1.5	84.1	87	18	86.25	89	23	89.87	89	22
2.0	87.5	90	19	89.4	91	22	90.2	93	5
2.5	85.2	92	32	85.99	95	27	88.5	97	4
3.0	89	91	7	90.71	95	7	91	96	7

Classification Accuracy Model

8. FUTURE SCOPE

The future scope of this project lies in its compatibility with various classifiers. In this project, we have used Adaboost M1 classifier which is a high accuracy classifier but a newer yet unstable version Adaboost M2 is also present. New modifications and enhancements can be done in this project with respect to Adaboost classifier.

Advancement in the software which can be done is that the model can be modified to predict the status of Parkinson disease on a real time basis with the help of medical instruments which can collect valid data for 23 attributes.

That conjectures to the fact of availability of instruments which can collect this data and send it for interpretation on a real time basis.

The project can also be enhanced to convert it from a desktop based application to a web based application so that it can work as a distributed system and can be more feasible to a varying audience.

Operating system based conversion can also be done so as to make the software runnable on different operating systems such as Mac OS.

9. REFERENCES

1. UCI dataset: https://archive.ics.uci.edu/ml/machine-learningdatabases/parkinsons/
2. . Parkinson's Brain Disease Prediction Using Big Data Analytics" by N. Shamli and B. Sathiyabhama
   http://www.mecs-press.org/ijitcs/ijitcs-v8-n6/IJITCS-V8-N6-10.pdf

3. Symptom Analysis of Parkinson Disease using SVM-SMO and Ada-Boost Classifiers by Muhtasim Billah https://core.ac.uk/download/pdf/61803688.pdf

4. Parkinson Australia: http://www.parkinsons.org.au/about-ps/about-pd.htm

5... UCI Parkinson repository dataset: https://archive.ics.uci.edu/ml/machinelearning-databases/parkinsons/

6. An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4871978/#:~:text=The%20main%20symptoms%20of%20PD,many%20patients%20and%20their%20families

7. History of Parkinson's disease:

https://link.springer.com/article/10.1007/s00415-006-7002-7

8. Anchana Khemphila and Veera Boonjing. ―Parkinson Disease Classification using Neural Network and Feature selection‖.

9. P. F. Guo, P. Bhattacharya, and N. Kharma, "Advances in detecting Parkinson's disease." In Medical Biometrics. Springer Berlin Heidelberg. pp. 306—314, 2010.