Institute for Advanced Studies
in Basic Sciences (IASBS)
Gava Zang, Zanjan, Iran

Detecting Parkinson's Disease Using Machine Learning Algorithms

Najmeh Moazzen

Department of Physics, IASBS

Fall 2021

Outline

- **Part I : The Paper**
  - What is PD?
  - Clinical diagnosing PD
  - Importance of ML for diagnosing PD
  - Review of the past ML experiments
  - Methods
  - Results
  - Conclusion

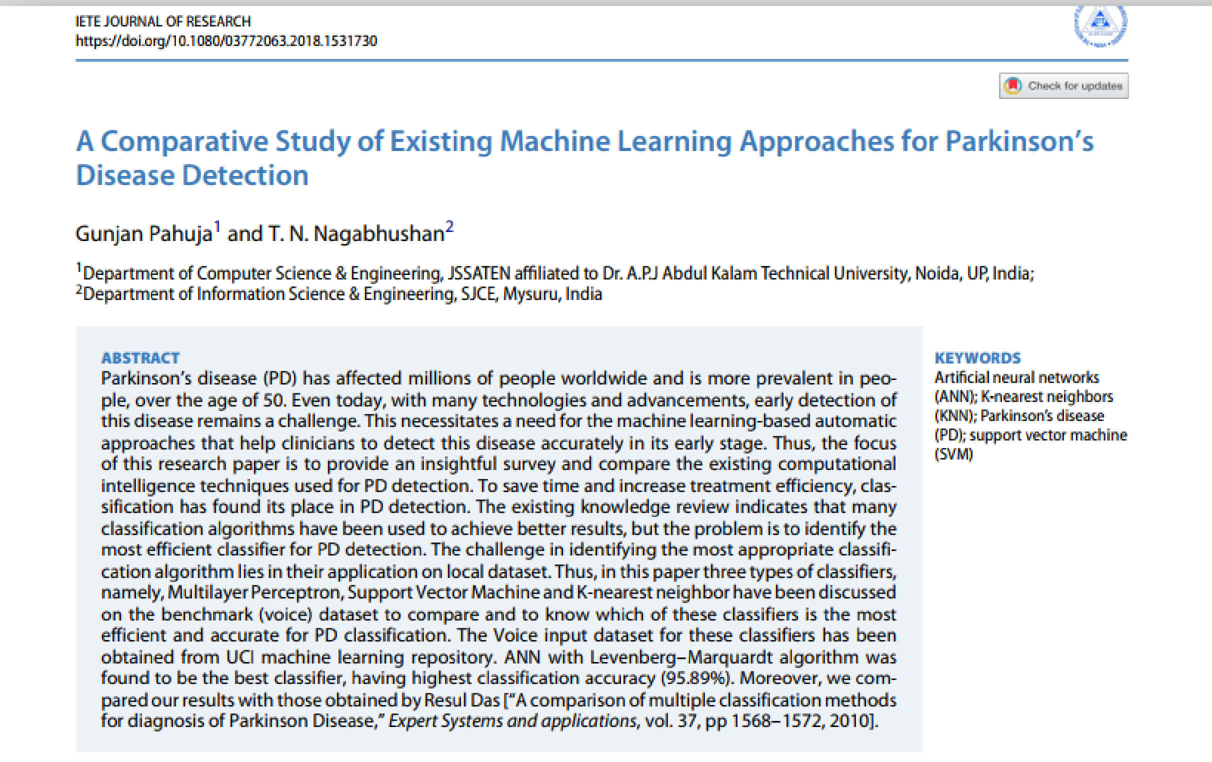
- **Part II : The Project**
  - Our purpose
  - Dataset
  - Models and results
  - Conclusion

Part I

The Paper:

A Comparative Study of Existing Machine Learning Approaches for Parkinson’s Disease Detection

Authors: Gunjan Pahuja and T. N. Nagabhushan



IETE JOURNAL OF RESEARCH
https://doi.org/10.1080/03772063.2018.1531730

Check for updates

A Comparative Study of Existing Machine Learning Approaches for Parkinson’s Disease Detection

Gunjan Pahuja[1] and T. N. Nagabhushan[2]

[1]Department of Computer Science & Engineering, JSSATEN affiliated to Dr. A.P.J Abdul Kalam Technical University, Noida, UP, India;
[2]Department of Information Science & Engineering, SJCE, Mysuru, India

ABSTRACT

Parkinson’s disease (PD) has affected millions of people worldwide and is more prevalent in people, over the age of 50. Even today, with many technologies and advancements, early detection of this disease remains a challenge. This necessitates a need for the machine learning-based automatic approaches that help clinicians to detect this disease accurately in its early stage. Thus, the focus of this research paper is to provide an insightful survey and compare the existing computational intelligence techniques used for PD detection. To save time and increase treatment efficiency, classification has found its place in PD detection. The existing knowledge review indicates that many classification algorithms have been used to achieve better results, but the problem is to identify the most efficient classifier for PD detection. The challenge in identifying the most appropriate classification algorithm lies in their application on local dataset. Thus, in this paper three types of classifiers, namely, Multilayer Perceptron, Support Vector Machine and K-nearest neighbor have been discussed on the benchmark (voice) dataset to compare and to know which of these classifiers is the most efficient and accurate for PD classification. The Voice input dataset for these classifiers has been obtained from UCI machine learning repository. ANN with Levenberg–Marquardt algorithm was found to be the best classifier, having highest classification accuracy (95.89%). Moreover, we compared our results with those obtained by Resul Das [“A comparison of multiple classification methods for diagnosis of Parkinson Disease,” *Expert Systems and applications*, vol. 37, pp 1568–1572, 2010].

KEYWORDS
Artificial neural networks (ANN); K-nearest neighbors (KNN); Parkinson’s disease (PD); support vector machine (SVM)

https://doi.org/10.1080/03772063.2018.1531730

What is PD?

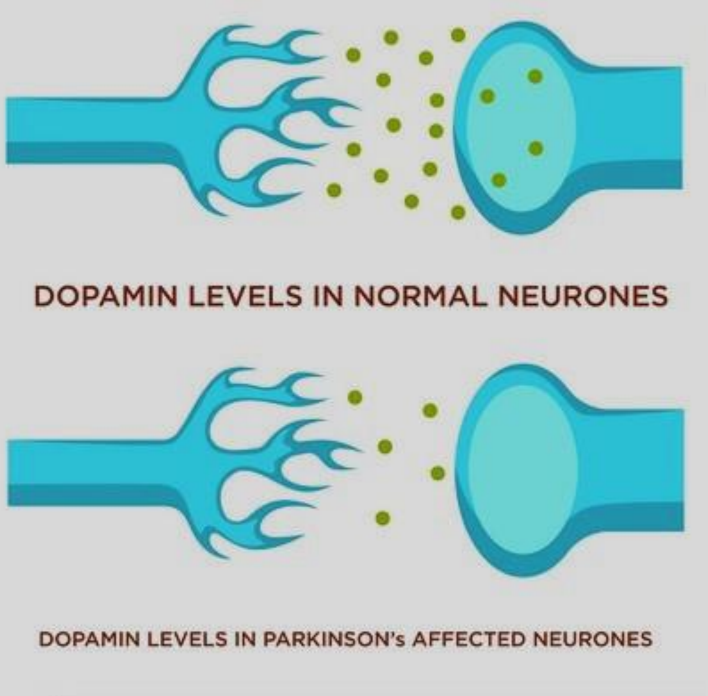
- ✓ **PD**: **P**arkinson’s **D**isease
- ✓ Parkinson's disease, is a long-term degenerative disorder of the central nervous system that mainly affects the motor system.

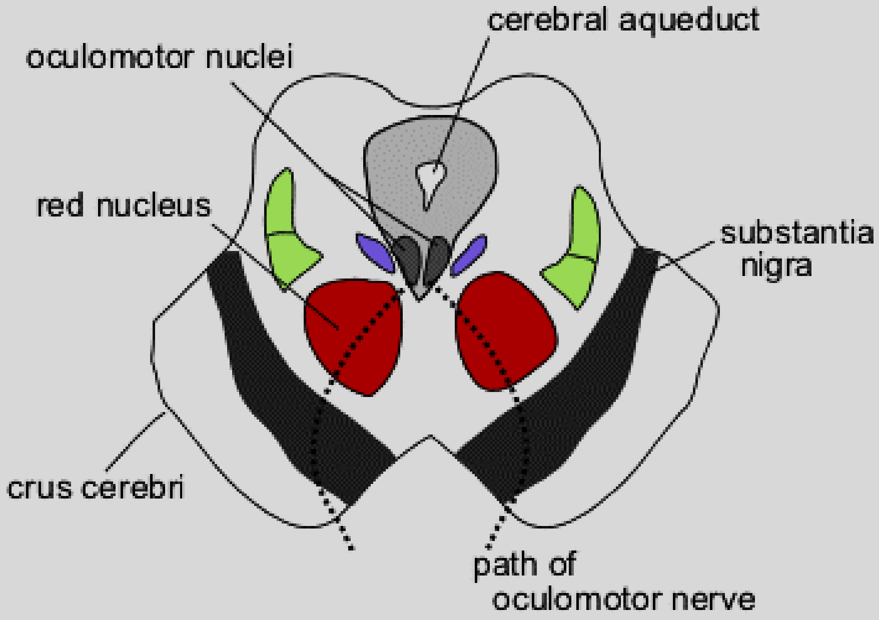
James Parkinson (1817)

What is PD?

✓ The motor symptoms of the disease result from the death of cells in the **substantia nigra**.

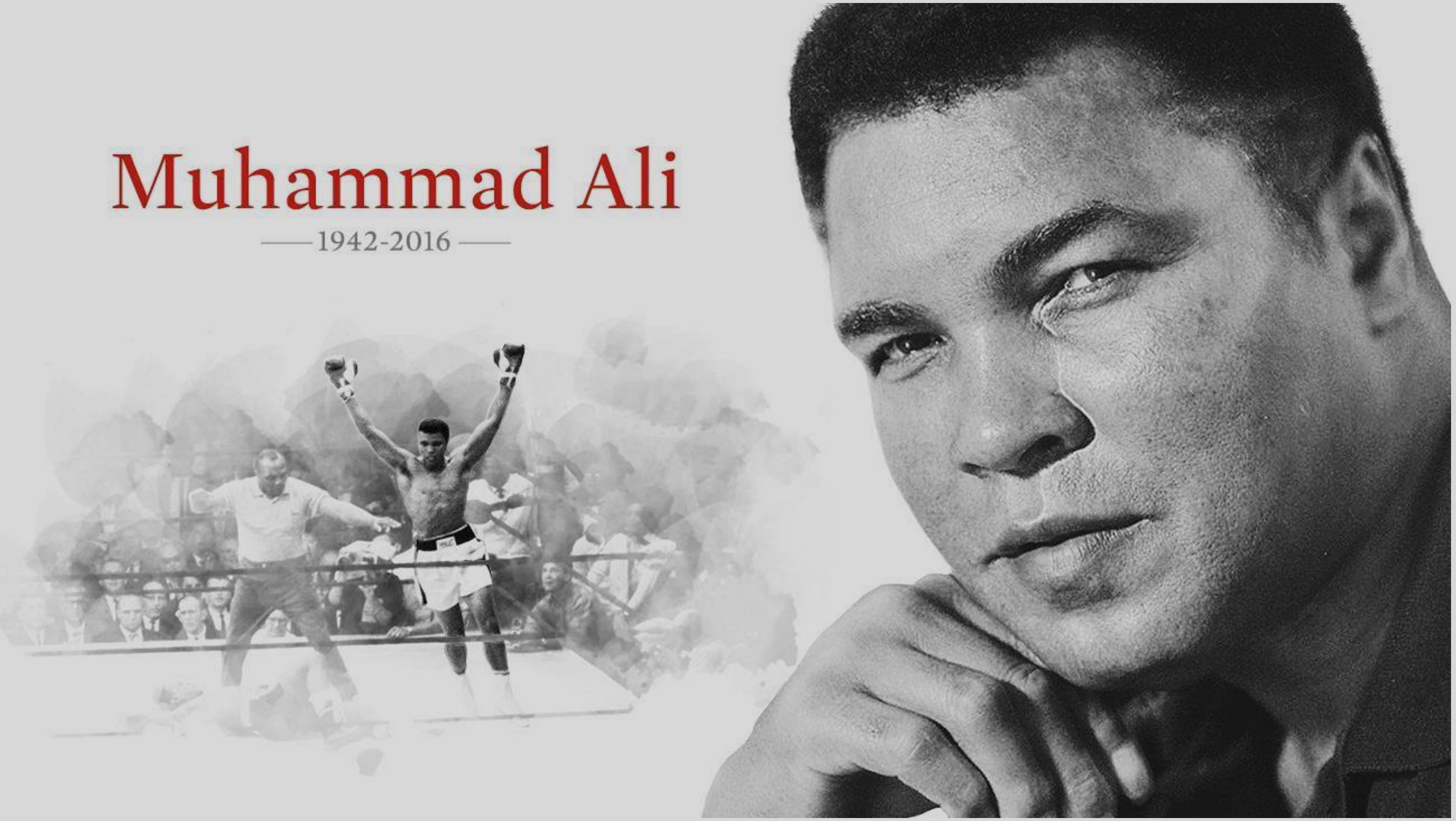
✓ A **lack of dopamine** causes Parkinson's disease.


DOPAMIN LEVELS IN NORMAL NEURONES
DOPAMIN LEVELS IN PARKINSON's AFFECTED NEURONES



What is PD?

✓ Environmental and genetic factors



What is PD?

PD symptoms:

- ✓ Tremor (trembling) in hands, arms, legs, jaw, or head.
- ✓ Stiffness of the limbs and trunk.
- ✓ Slowness of movement.
- ✓ Impaired balance and coordination, sometimes leading to falls.
- ✓ Changes in voice.
- ✓ Depression and other emotional changes.
- ✓ Sleep disruptions.

X%Q_?
SPEECH CHANGES
TREMOR
SLOWED MOVEMENT

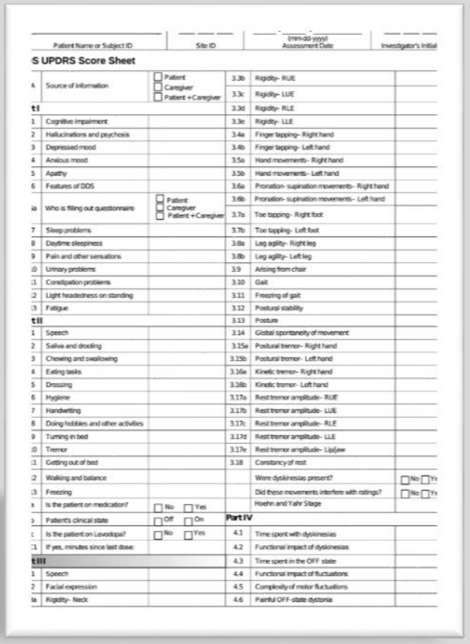
Clinical diagnosing PD

MDS-UPDRS: **M**ovement **D**isorder **S**ociety-**U**nified **P**arkinson **D**isease **R**ating **S**cale

- ✓ Used for the early diagnosis of PD.
- ✓ Certain drawbacks associated with these methods are as follows:

    1. Availability of skilled workforce

    2. Time and cooperation required from patients for a longer period.



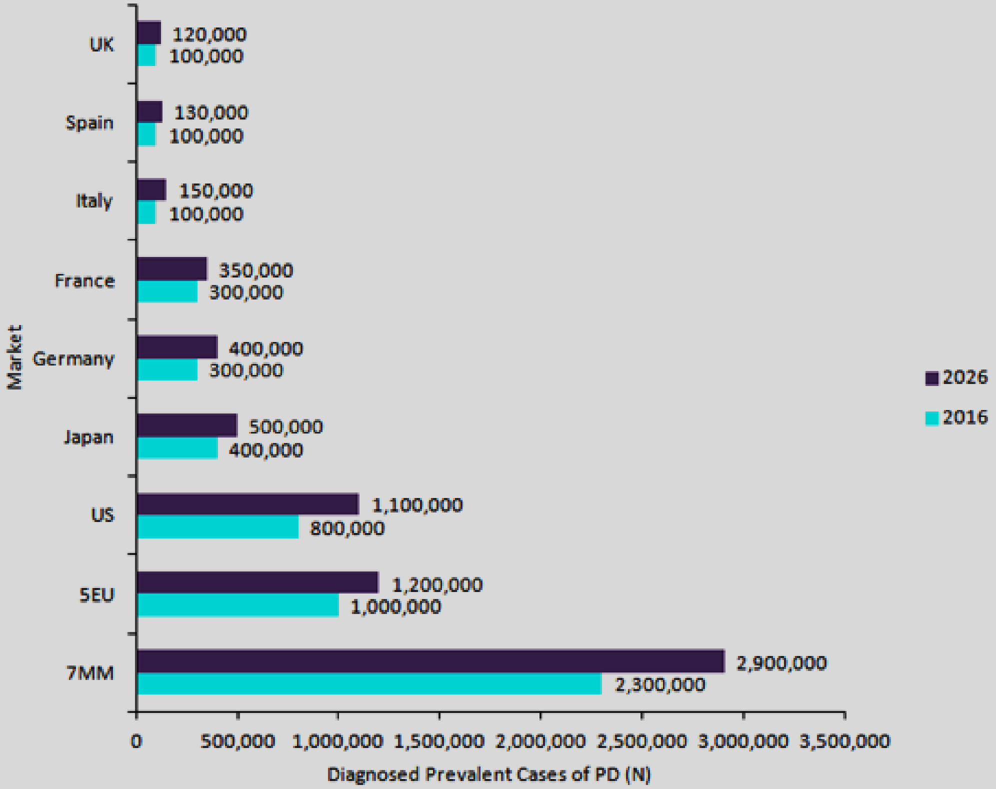
Patient Name or Subject ID	Site ID	(mm-dd-yyyy) Assessment Date	Investigator's Initial

S UPDRS Score Sheet

A	Source of information	□ Patient □ Caregiver □ Patient + Caregiver	3.3b	Rigidity- RUE	
			3.3c	Rigidity- LUE	
t I			3.3d	Rigidity- RLE	
1	Cognitive impairment		3.3e	Rigidity- LLE	
2	Hallucinations and psychosis		3.4a	Finger tapping- Right hand	
3	Depressed mood		3.4b	Finger tapping- Left hand	
4	Anxious mood		3.5a	Hand movements- Right hand	
5	Apathy		3.5b	Hand movements- Left hand	
6	Features of DDS		3.6a	Pronation- supination movements- Right hand	
a	Who is filling out questionnaire	□ Patient □ Caregiver □ Patient + Caregiver	3.6b	Pronation- supination movements- Left hand	
			3.7a	Toe tapping- Right foot	
7	Sleep problems		3.7b	Toe tapping- Left foot	
8	Daytime sleepiness		3.8a	Leg agility- Right leg	
9	Pain and other sensations		3.8b	Leg agility- Left leg	
10	Urinary problems		3.9	Arising from chair	
11	Constipation problems		3.10	Gait	
12	Light headedness on standing		3.11	Freezing of gait	
13	Fatigue		3.12	Postural stability	
t II			3.13	Posture	
1	Speech		3.14	Global spontaneity of movement	
2	Saliva and drooling		3.15a	Postural tremor- Right hand	
3	Chewing and swallowing		3.15b	Postural tremor- Left hand	
4	Eating tasks		3.16a	Kinetic tremor- Right hand	
5	Dressing		3.16b	Kinetic tremor- Left hand	
6	Hygiene		3.17a	Rest tremor amplitude- RUE	
7	Handwriting		3.17b	Rest tremor amplitude- LUE	
8	Doing hobbies and other activities		3.17c	Rest tremor amplitude- RLE	
9	Turning in bed		3.17d	Rest tremor amplitude- LLE	
10	Tremor		3.17e	Rest tremor amplitude- Lip/jaw	
11	Getting out of bed		3.18	Constancy of rest	
12	Walking and balance			Were dyskinesias present?	□ No □ Ye
13	Freezing			Did these movements interfere with ratings?	□ No □ Ye
a	Is the patient on medication?	□ No □ Yes		Hoehn and Yahr Stage	
b	Patient's clinical state	□ Off □ On	Part IV		
c	Is the patient on Levodopa?	□ No □ Yes	4.1	Time spent with dyskinesias	
1	If yes, minutes since last dose:		4.2	Functional impact of dyskinesias	
t III			4.3	Time spent in the OFF state	
1	Speech		4.4	Functional impact of fluctuations	
2	Facial expression		4.5	Complexity of motor fluctuations	
a	Rigidity- Neck		4.6	Painful OFF-state dystonia	

Importance of ML for diagnosing PD

- PD is one of the most common neurodegenerative diseases with a prevalence rate of **1%** in the population above 60 years old, affecting **1–2 people per 1,000**.
- The estimated global population affected by PD has more than doubled from 1990 to 2016 (from 2.5 million to 6.1 million), which is a result of increased number of elderly people and age-standardized prevalence rates.



Importance of ML for diagnosing PD

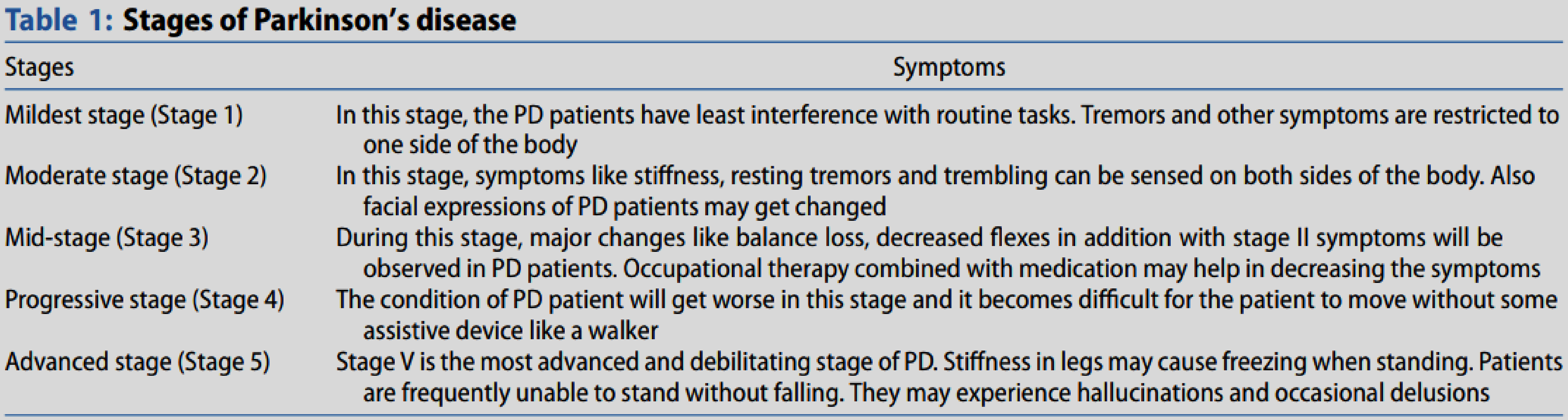


Table 1: Stages of Parkinson's disease

Stages	Symptoms
Mildest stage (Stage 1)	In this stage, the PD patients have least interference with routine tasks. Tremors and other symptoms are restricted to one side of the body
Moderate stage (Stage 2)	In this stage, symptoms like stiffness, resting tremors and trembling can be sensed on both sides of the body. Also facial expressions of PD patients may get changed
Mid-stage (Stage 3)	During this stage, major changes like balance loss, decreased flexes in addition with stage II symptoms will be observed in PD patients. Occupational therapy combined with medication may help in decreasing the symptoms
Progressive stage (Stage 4)	The condition of PD patient will get worse in this stage and it becomes difficult for the patient to move without some assistive device like a walker
Advanced stage (Stage 5)	Stage V is the most advanced and debilitating stage of PD. Stiffness in legs may cause freezing when standing. Patients are frequently unable to stand without falling. They may experience hallucinations and occasional delusions

Review of the past ML classifications

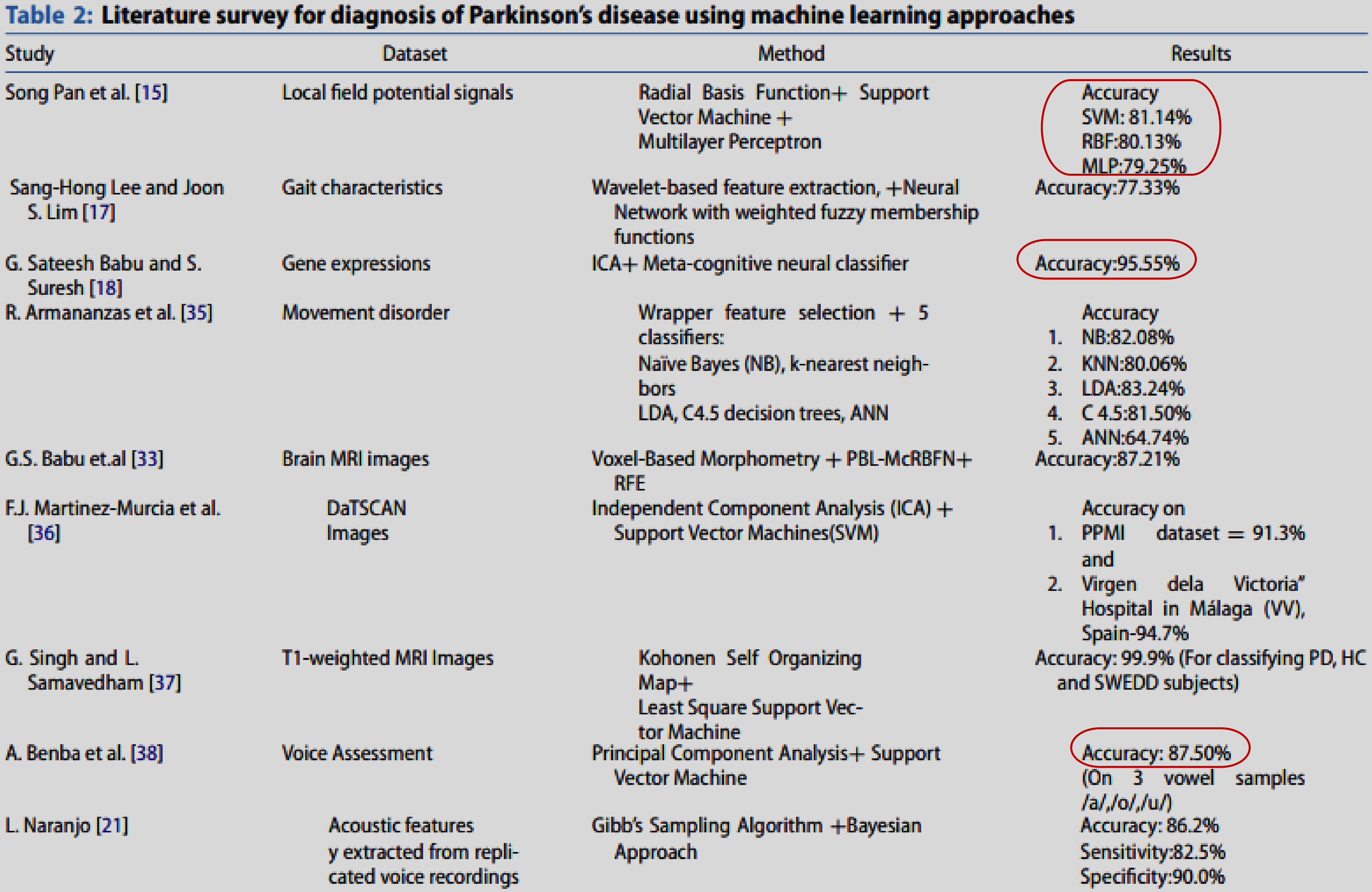


Table 2: Literature survey for diagnosis of Parkinson's disease using machine learning approaches

Study	Dataset	Method	Results
Song Pan et al. [15]	Local field potential signals	Radial Basis Function+ Support Vector Machine + Multilayer Perceptron	Accuracy SVM: 81.14% RBF:80.13% MLP:79.25%
Sang-Hong Lee and Joon S. Lim [17]	Gait characteristics	Wavelet-based feature extraction, +Neural Network with weighted fuzzy membership functions	Accuracy:77.33%
G. Sateesh Babu and S. Suresh [18]	Gene expressions	ICA+ Meta-cognitive neural classifier	Accuracy:95.55%
R. Armananzas et al. [35]	Movement disorder	Wrapper feature selection + 5 classifiers: Naïve Bayes (NB), k-nearest neighbors LDA, C4.5 decision trees, ANN	Accuracy 1. NB:82.08% 2. KNN:80.06% 3. LDA:83.24% 4. C 4.5:81.50% 5. ANN:64.74%
G.S. Babu et.al [33]	Brain MRI images	Voxel-Based Morphometry + PBL-McRBFN+ RFE	Accuracy:87.21%
F.J. Martinez-Murcia et al. [36]	DaTSCAN Images	Independent Component Analysis (ICA) + Support Vector Machines(SVM)	Accuracy on 1. PPMI dataset = 91.3% and 2. Virgen dela Victoria” Hospital in Málaga (VV), Spain-94.7%
G. Singh and L. Samavedham [37]	T1-weighted MRI Images	Kohonen Self Organizing Map+ Least Square Support Vector Machine	Accuracy: 99.9% (For classifying PD, HC and SWEDD subjects)
A. Benba et al. [38]	Voice Assessment	Principal Component Analysis+ Support Vector Machine	Accuracy: 87.50% (On 3 vowel samples /a/,/o/,/u/)
L. Naranjo [21]	Acoustic features y extracted from replicated voice recordings	Gibb's Sampling Algorithm +Bayesian Approach	Accuracy: 86.2% Sensitivity:82.5% Specificity:90.0%

Review of the past ML classifications

- **Feature Subset Selection (FSS) Techniques**

The diagnosis of neurodegenerative diseases through machine learning :

1. Data acquisition (Brain MRI images, gait movements, vocal data, local field potential etc.)
2. Feature extraction (extract the features suitable for training and testing a classifier).
3. Feature subset selection (to reduce the redundant features).
4. Training and validating the performance of the classifier.

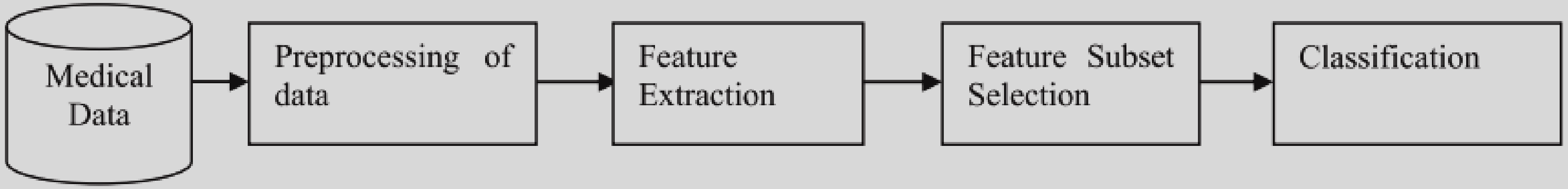


Figure 2: Steps involved in medical image processing (MIP) using machine learning techniques

Review of the past ML classifications

- **Classification**

Pattern recognition is defined as an act of taking raw data and classifying them into different categories based on machine learning algorithms such as **K-NN** rule, **SVM**, artificial neural networks (**ANN**) .

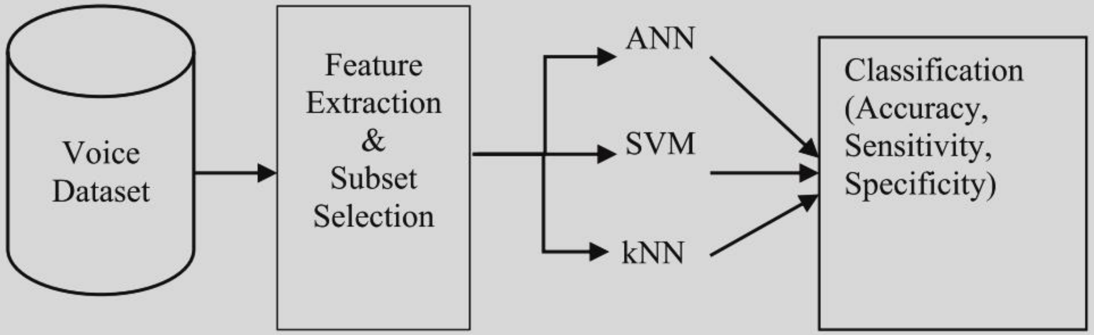

Voice Dataset
Feature Extraction & Subset Selection
ANN
SVM
kNN
Classification (Accuracy, Sensitivity, Specificity)

Figure 3: Methods applied for PD classification

Available Datasets

- **Gate dataset**

(Parkinson’s disease: 93 & Healthy: 73)

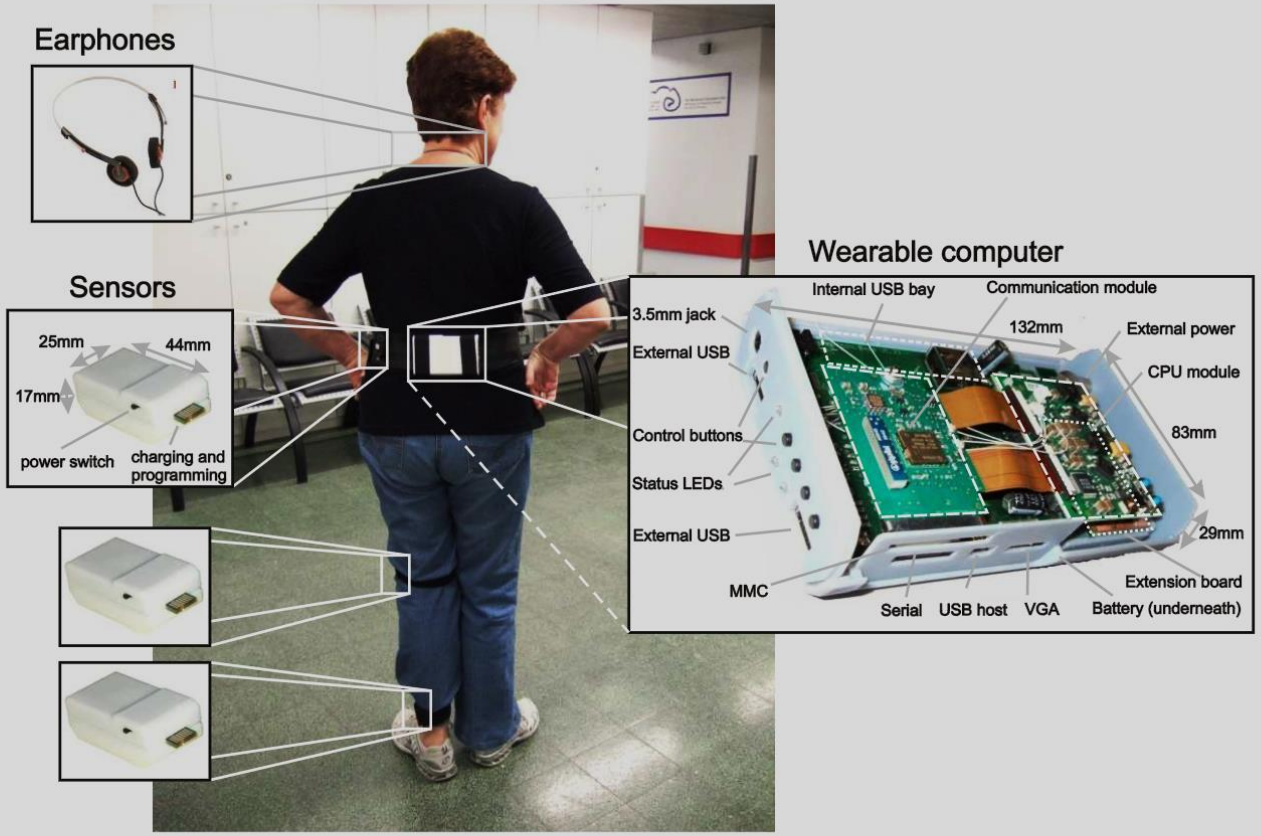
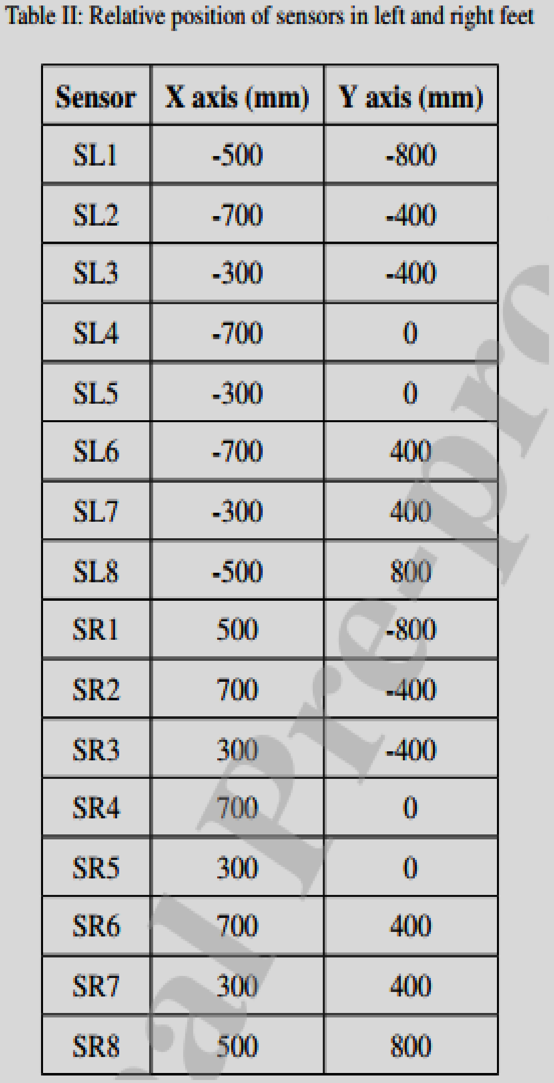


Table II: Relative position of sensors in left and right feet



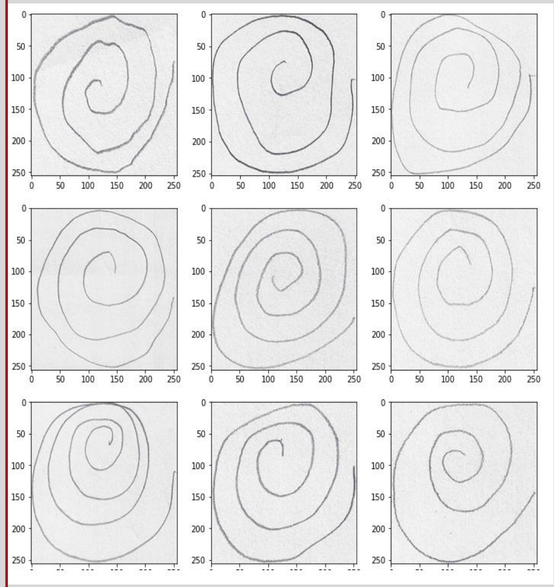
Sensor	X axis (mm)	Y axis (mm)
SL1	-500	-800
SL2	-700	-400
SL3	-300	-400
SL4	-700	0
SL5	-300	0
SL6	-700	400
SL7	-300	400
SL8	-500	800
SR1	500	-800
SR2	700	-400
SR3	300	-400
SR4	700	0
SR5	300	0
SR6	700	400
SR7	300	400
SR8	500	800

Available Datasets

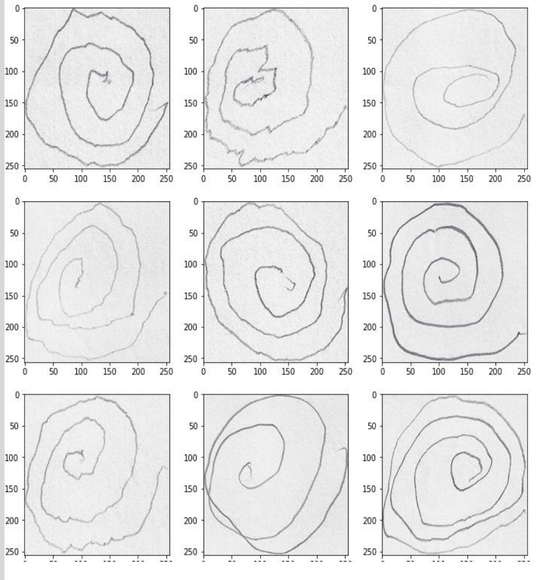
- **Spiral / wave drawings dataset**

(Parkinson’s disease: 62 & Healthy: 15)

Spiral drawings

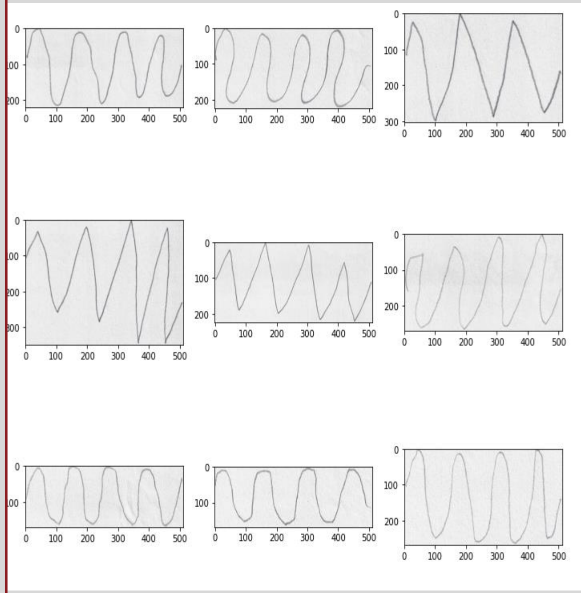


Healthy

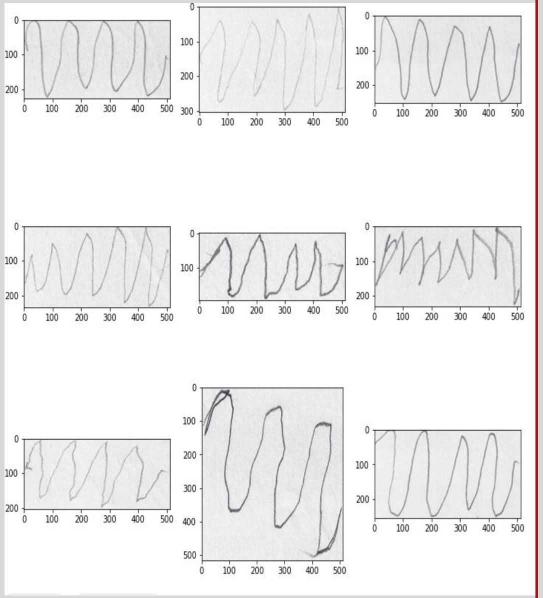


Parkinson’s disease

Wave drawings



Healthy



Parkinson’s disease

Available Datasets

- **Voice dataset**

There are six recordings per patient. The first column of the dataset specifies the name of the patient and the last column specifies the status which is set to **1 for PD** and **0 for healthy** subjects.

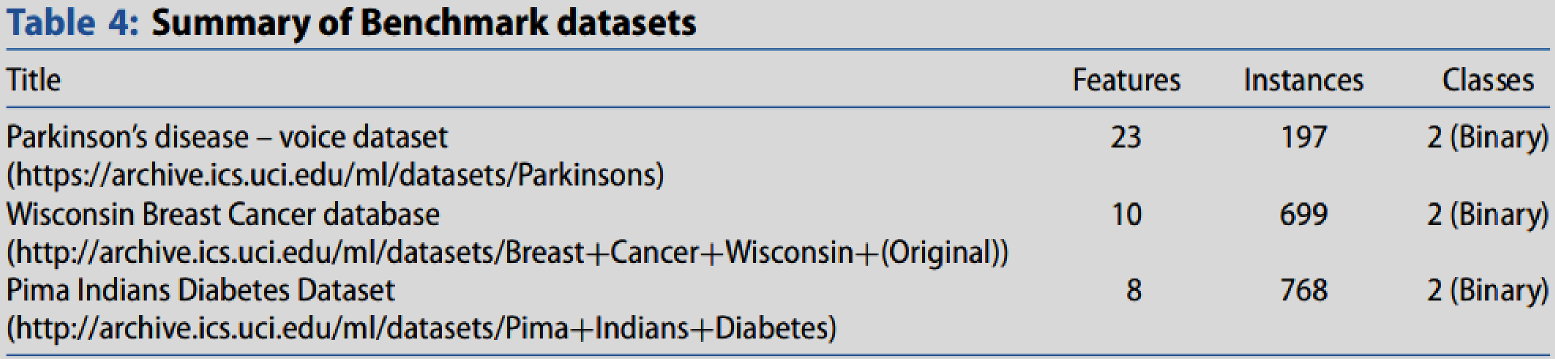


Table 4: Summary of Benchmark datasets

Title	Features	Instances	Classes
Parkinson's disease – voice dataset (https://archive.ics.uci.edu/ml/datasets/Parkinsons)	23	197	2 (Binary)
Wisconsin Breast Cancer database (http://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Original))	10	699	2 (Binary)
Pima Indians Diabetes Dataset (http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes)	8	768	2 (Binary)

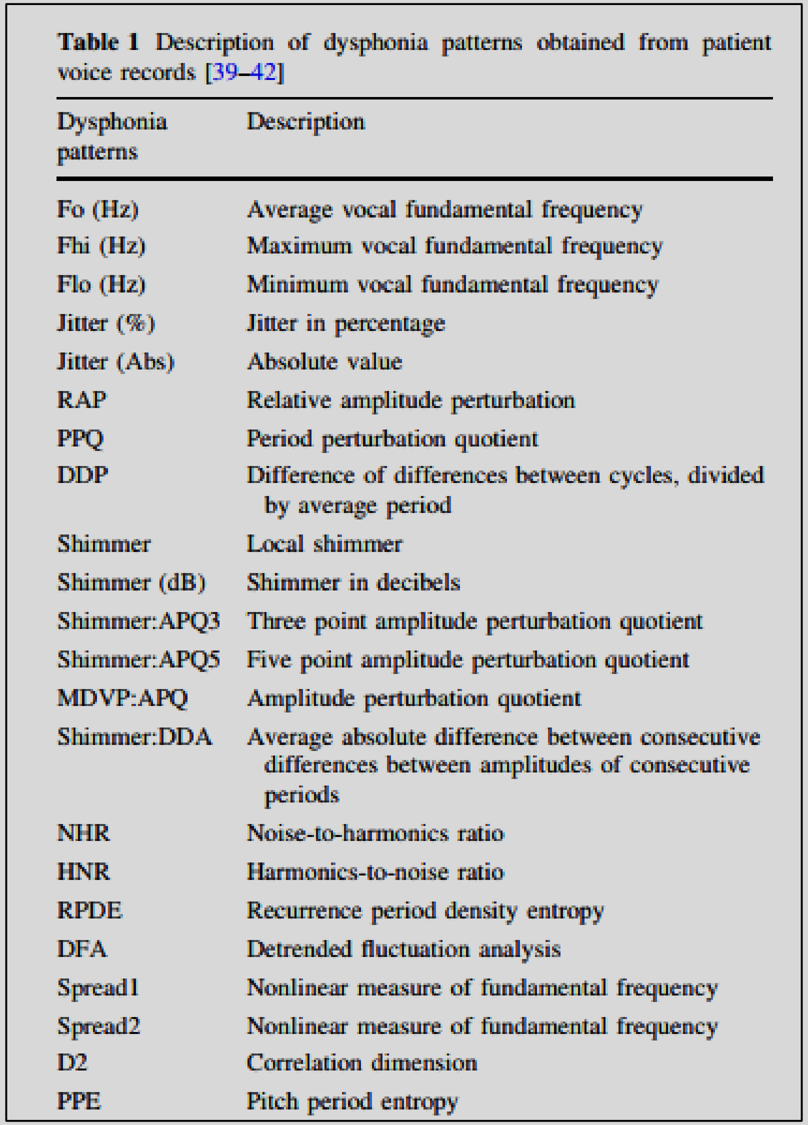


Table 1 Description of dysphonia patterns obtained from patient voice records [39–42]

Dysphonia patterns	Description
Fo (Hz)	Average vocal fundamental frequency
Fhi (Hz)	Maximum vocal fundamental frequency
Flo (Hz)	Minimum vocal fundamental frequency
Jitter (%)	Jitter in percentage
Jitter (Abs)	Absolute value
RAP	Relative amplitude perturbation
PPQ	Period perturbation quotient
DDP	Difference of differences between cycles, divided by average period
Shimmer	Local shimmer
Shimmer (dB)	Shimmer in decibels
Shimmer:APQ3	Three point amplitude perturbation quotient
Shimmer:APQ5	Five point amplitude perturbation quotient
MDVP:APQ	Amplitude perturbation quotient
Shimmer:DDA	Average absolute difference between consecutive differences between amplitudes of consecutive periods
NHR	Noise-to-harmonics ratio
HNR	Harmonics-to-noise ratio
RPDE	Recurrence period density entropy
DFA	Detrended fluctuation analysis
Spread1	Nonlinear measure of fundamental frequency
Spread2	Nonlinear measure of fundamental frequency
D2	Correlation dimension
PPE	Pitch period entropy

Methods

1. Artificial Neural Network (ANN)

Levenberg–Marquardt algorithm with 10 neurons in hidden layers.

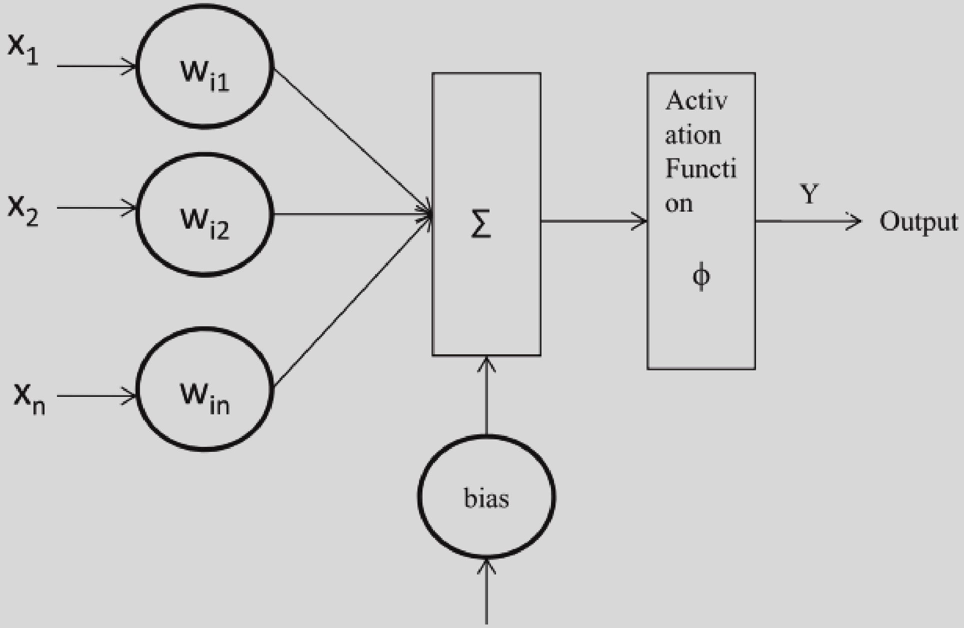


Figure 4: Artificial neural network architecture

Methods

2. Support Vector Machine (SVM)

SVM for binary classification. Binary classification is based on the concept of dividing the data into classes using a hyperplane.

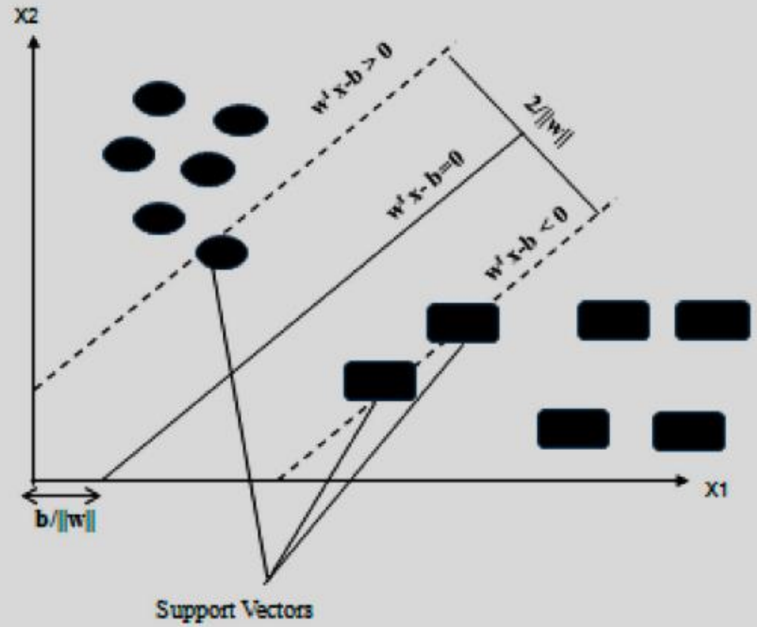
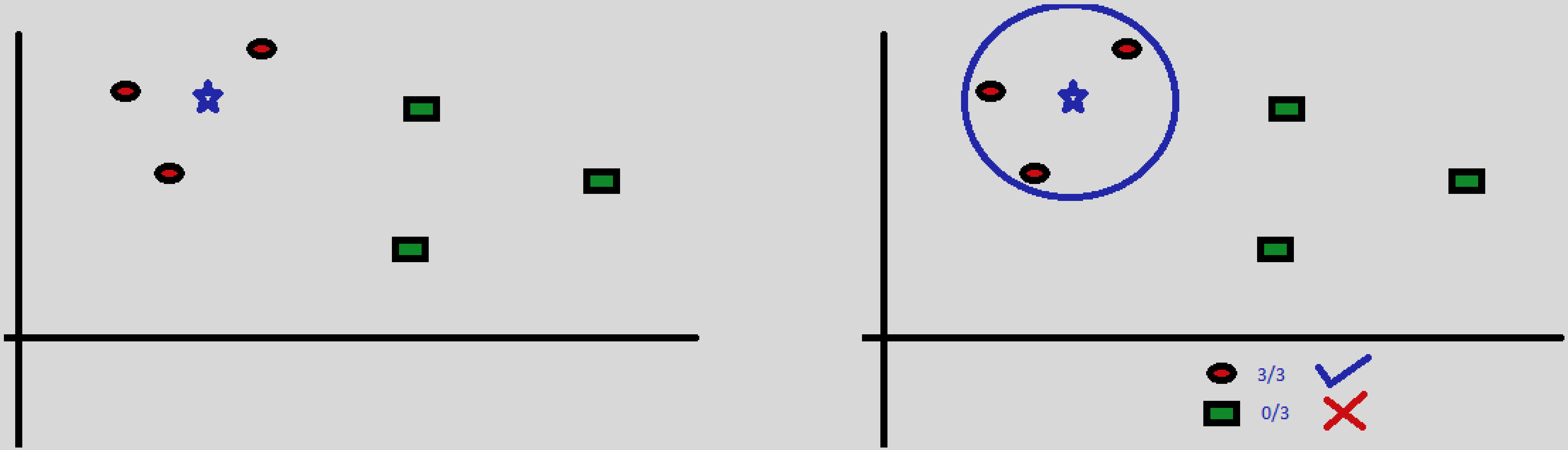


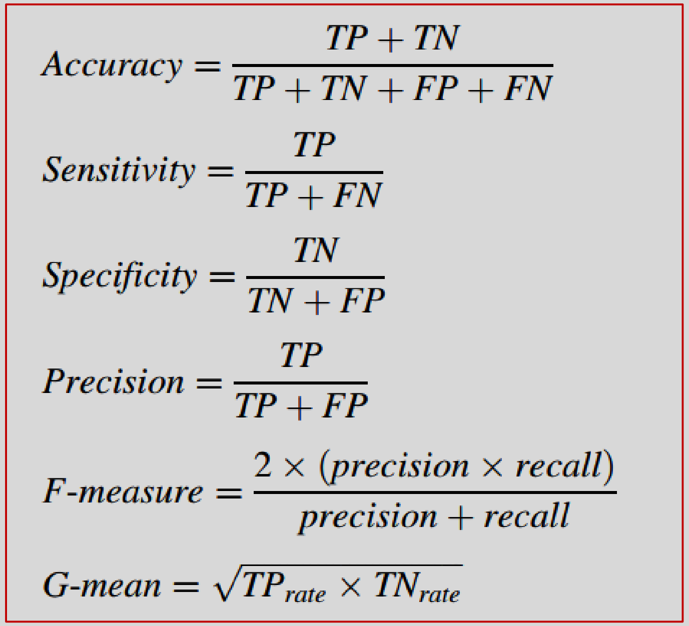
Figure 5: SVM trained with data/samples from 2 classes

Methods

2. k-Nearest Neighbor (kNN)



Results



$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

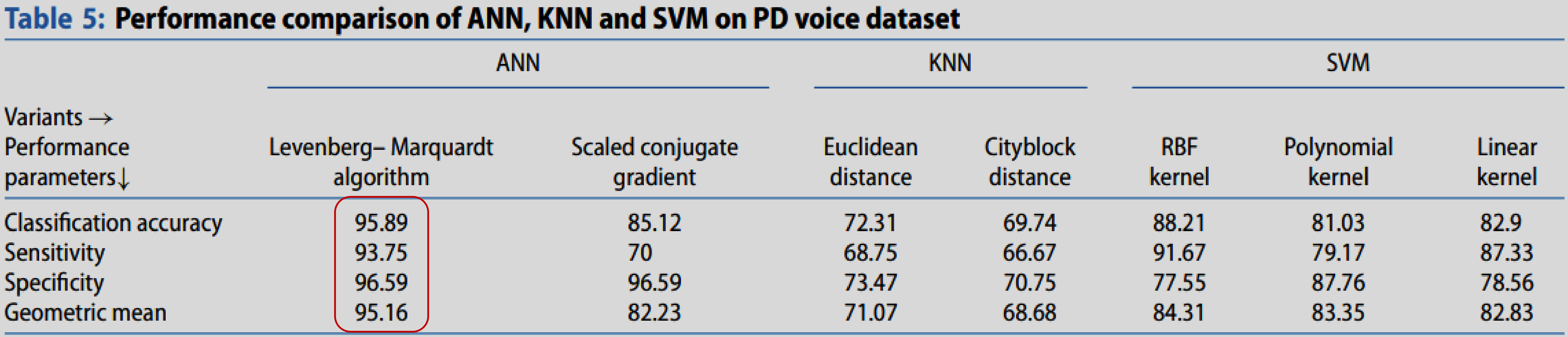
$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F\text{-}measure = \frac{2 \times (precision \times recall)}{precision + recall}$$

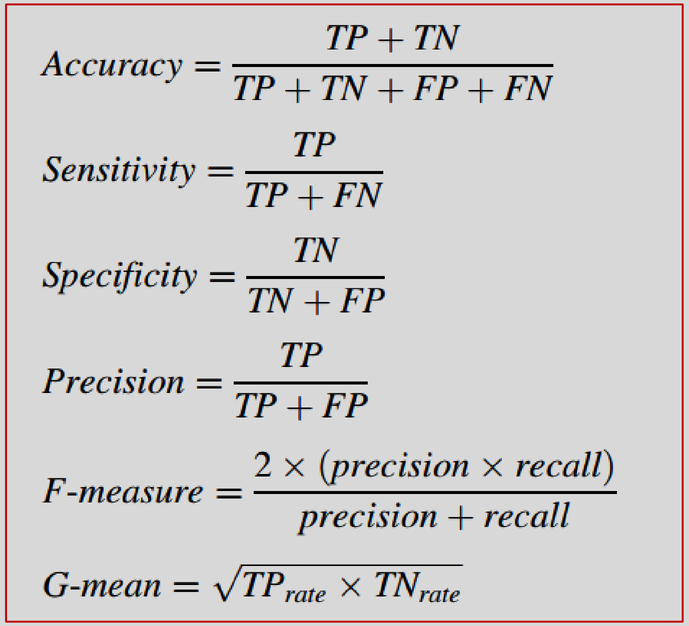
$$G\text{-}mean = \sqrt{TP_{rate} \times TN_{rate}}$$

Table 5: Performance comparison of ANN, KNN and SVM on PD voice dataset



	ANN		KNN		SVM		
Variants → Performance parameters↓	Levenberg– Marquardt algorithm	Scaled conjugate gradient	Euclidean distance	Cityblock distance	RBF kernel	Polynomial kernel	Linear kernel
Classification accuracy	95.89	85.12	72.31	69.74	88.21	81.03	82.9
Sensitivity	93.75	70	68.75	66.67	91.67	79.17	87.33
Specificity	96.59	96.59	73.47	70.75	77.55	87.76	78.56
Geometric mean	95.16	82.23	71.07	68.68	84.31	83.35	82.83

Results



$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F\text{-}measure = \frac{2 \times (precision \times recall)}{precision + recall}$$

$$G\text{-}mean = \sqrt{TP_{rate} \times TN_{rate}}$$

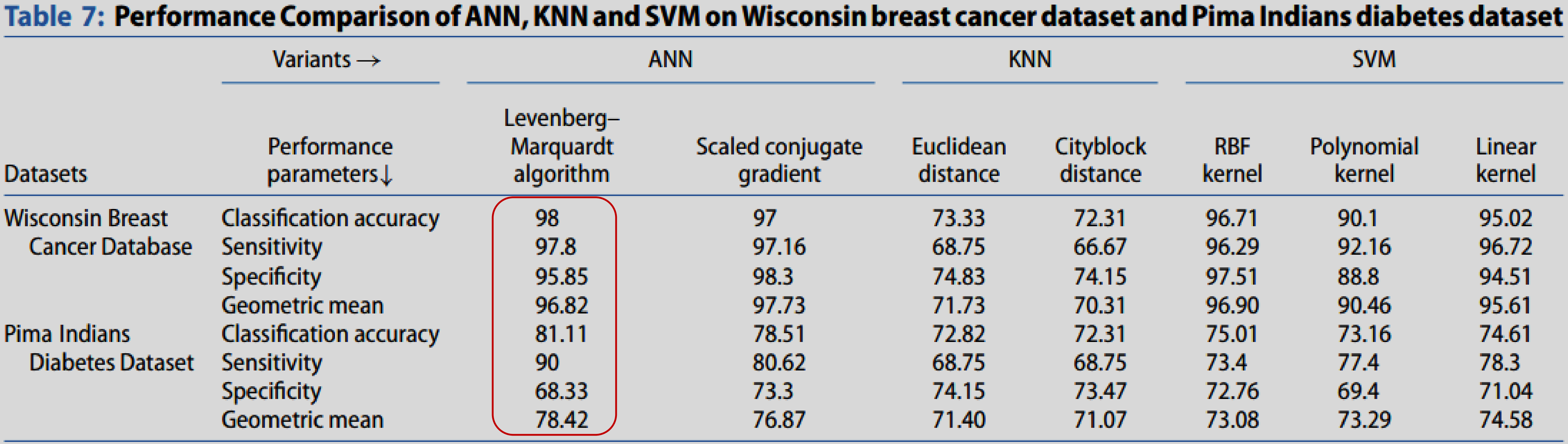


Table 7: Performance Comparison of ANN, KNN and SVM on Wisconsin breast cancer dataset and Pima Indians diabetes dataset

	Variants →	ANN		KNN		SVM		
Datasets	Performance parameters↓	Levenberg–Marquardt algorithm	Scaled conjugate gradient	Euclidean distance	Cityblock distance	RBF kernel	Polynomial kernel	Linear kernel
Wisconsin Breast Cancer Database	Classification accuracy	98	97	73.33	72.31	96.71	90.1	95.02
	Sensitivity	97.8	97.16	68.75	66.67	96.29	92.16	96.72
	Specificity	95.85	98.3	74.83	74.15	97.51	88.8	94.51
	Geometric mean	96.82	97.73	71.73	70.31	96.90	90.46	95.61
Pima Indians Diabetes Dataset	Classification accuracy	81.11	78.51	72.82	72.31	75.01	73.16	74.61
	Sensitivity	90	80.62	68.75	68.75	73.4	77.4	78.3
	Specificity	68.33	73.3	74.15	73.47	72.76	69.4	71.04
	Geometric mean	78.42	76.87	71.40	71.07	73.08	73.29	74.58

Conclusion

✓ It is observed that Artificial Neural Networks with Levenberg–Marquardt algorithm gives the highest classification **accuracy of 95.89%** for voice dataset.

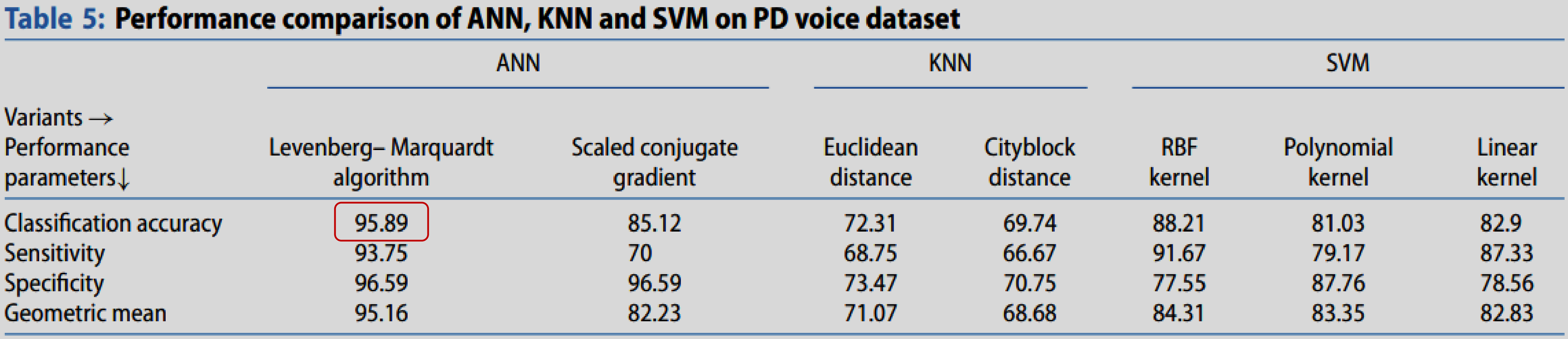


Table 5: Performance comparison of ANN, KNN and SVM on PD voice dataset

	ANN		KNN		SVM		
Variants → Performance parameters↓	Levenberg– Marquardt algorithm	Scaled conjugate gradient	Euclidean distance	Cityblock distance	RBF kernel	Polynomial kernel	Linear kernel
Classification accuracy	95.89	85.12	72.31	69.74	88.21	81.03	82.9
Sensitivity	93.75	70	68.75	66.67	91.67	79.17	87.33
Specificity	96.59	96.59	73.47	70.75	77.55	87.76	78.56
Geometric mean	95.16	82.23	71.07	68.68	84.31	83.35	82.83

Part II

The Project:

Early Stage Prediction of Parkinson's Disease

using Machine Learning Algorithms

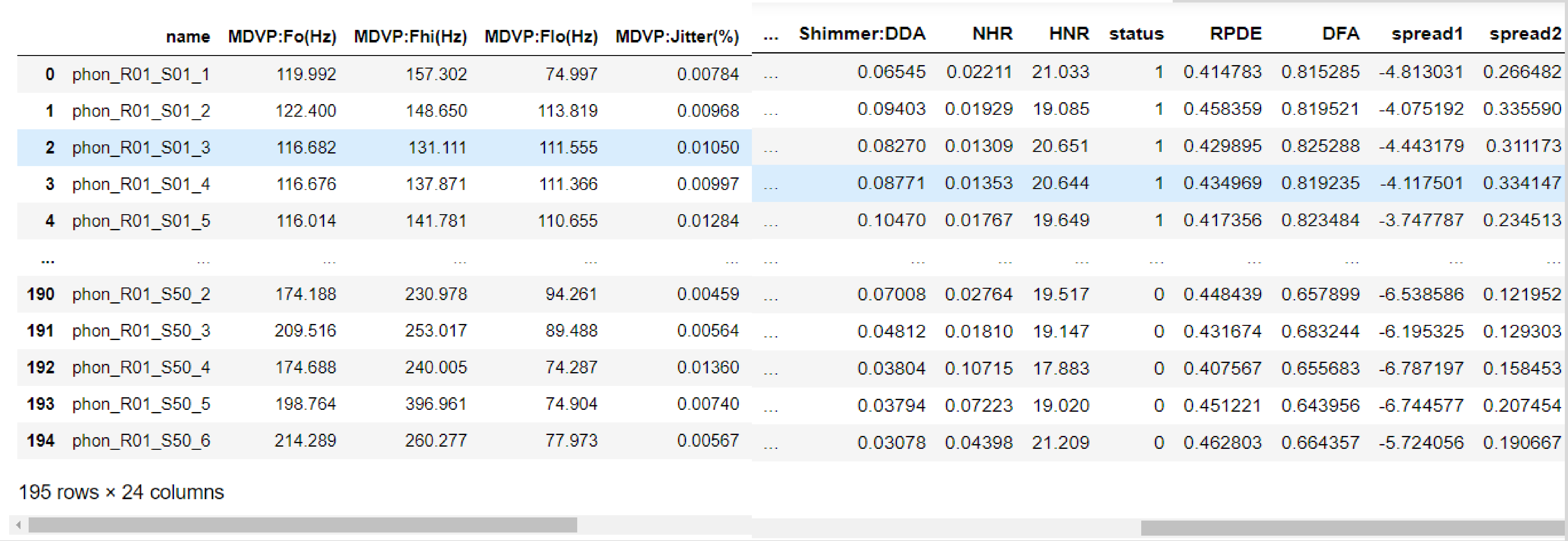
Our Purpose

The goal of this project:

To provide simple, low-cost, high-accuracy methods for the early diagnosis of Parkinson's disease.

Dataset

- **Voice dataset**

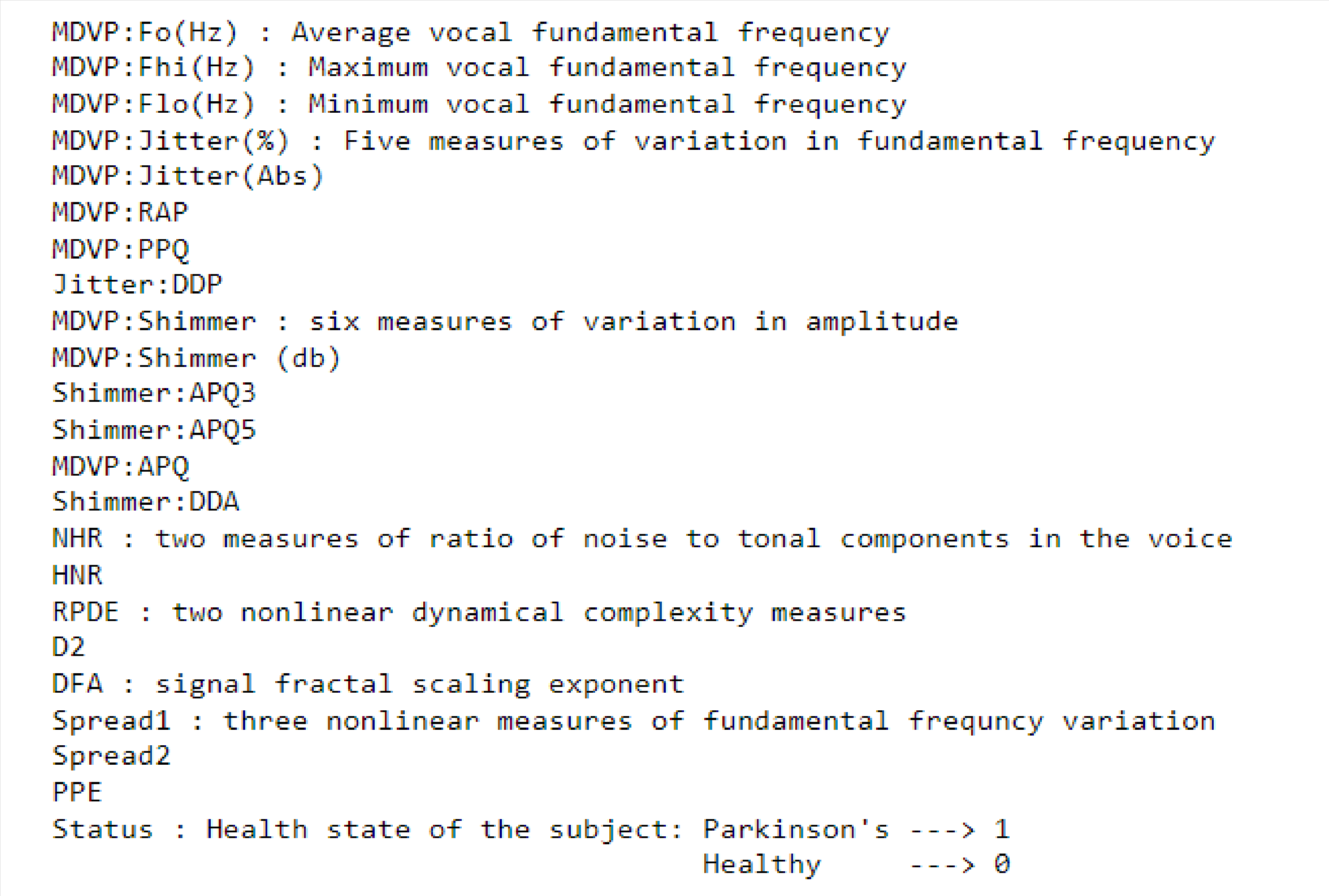


	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	...	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	...	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.813031	0.266482
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	...	0.09403	0.01929	19.085	1	0.458359	0.819521	-4.075192	0.335590
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	...	0.08270	0.01309	20.651	1	0.429895	0.825288	-4.443179	0.311173
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	...	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.117501	0.334147
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	...	0.10470	0.01767	19.649	1	0.417356	0.823484	-3.747787	0.234513
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
190	phon_R01_S50_2	174.188	230.978	94.261	0.00459	...	0.07008	0.02764	19.517	0	0.448439	0.657899	-6.538586	0.121952
191	phon_R01_S50_3	209.516	253.017	89.488	0.00564	...	0.04812	0.01810	19.147	0	0.431674	0.683244	-6.195325	0.129303
192	phon_R01_S50_4	174.688	240.005	74.287	0.01360	...	0.03804	0.10715	17.883	0	0.407567	0.655683	-6.787197	0.158453
193	phon_R01_S50_5	198.764	396.961	74.904	0.00740	...	0.03794	0.07223	19.020	0	0.451221	0.643956	-6.744577	0.207454
194	phon_R01_S50_6	214.289	260.277	77.973	0.00567	...	0.03078	0.04398	21.209	0	0.462803	0.664357	-5.724056	0.190667

195 rows × 24 columns

Dataset

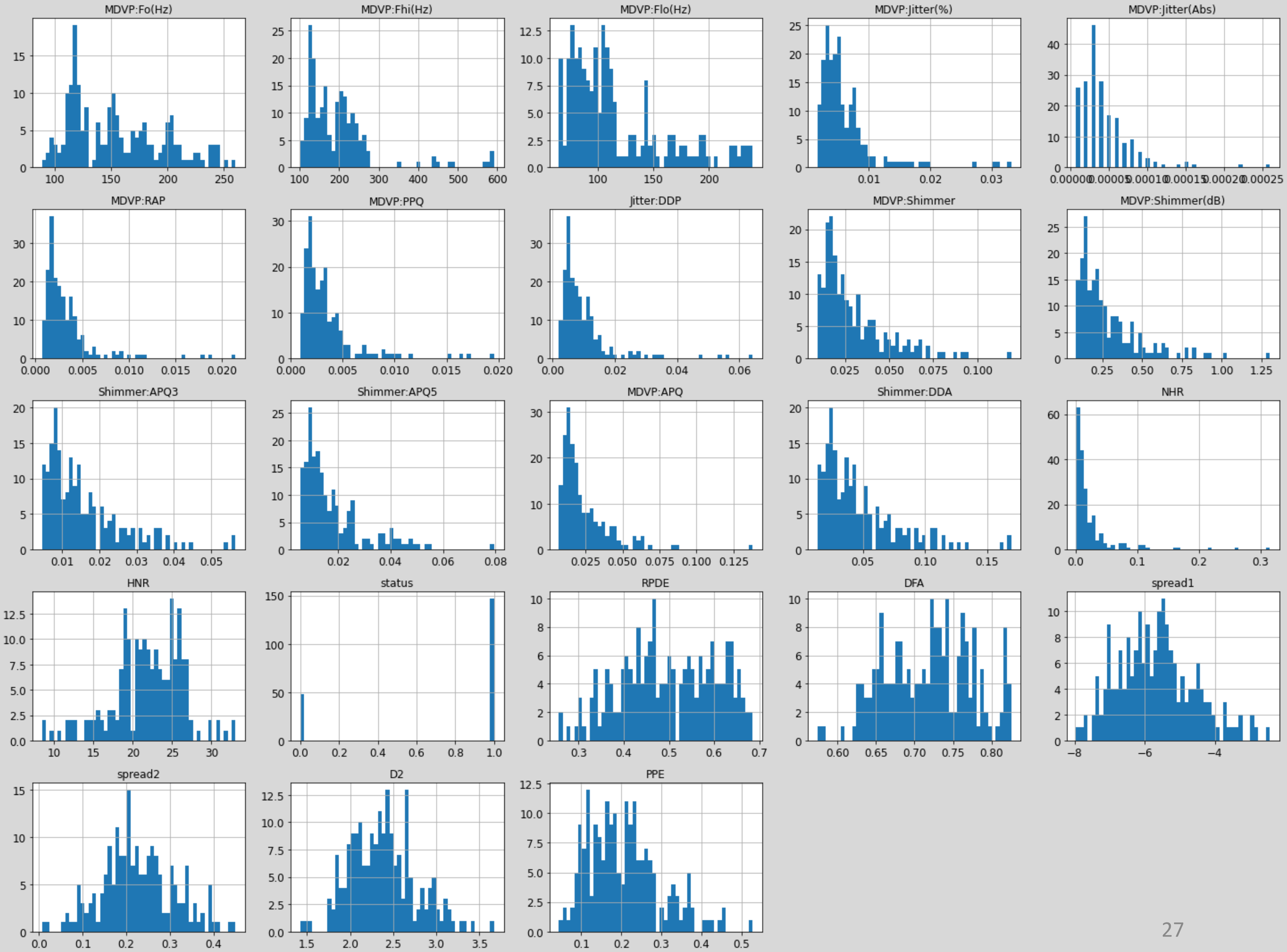
Features:



```
MDVP:Fo(Hz) : Average vocal fundamental frequency
MDVP:Fhi(Hz) : Maximum vocal fundamental frequency
MDVP:Flo(Hz) : Minimum vocal fundamental frequency
MDVP:Jitter(%) : Five measures of variation in fundamental frequency
MDVP:Jitter(Abs)
MDVP:RAP
MDVP:PPQ
Jitter:DDP
MDVP:Shimmer : six measures of variation in amplitude
MDVP:Shimmer (db)
Shimmer:APQ3
Shimmer:APQ5
MDVP:APQ
Shimmer:DDA
NHR : two measures of ratio of noise to tonal components in the voice
HNR
RPDE : two nonlinear dynamical complexity measures
D2
DFA : signal fractal scaling exponent
Spread1 : three nonlinear measures of fundamental frequncy variation
Spread2
PPE
Status : Health state of the subject: Parkinson's ---> 1
                                      Healthy     ---> 0
```

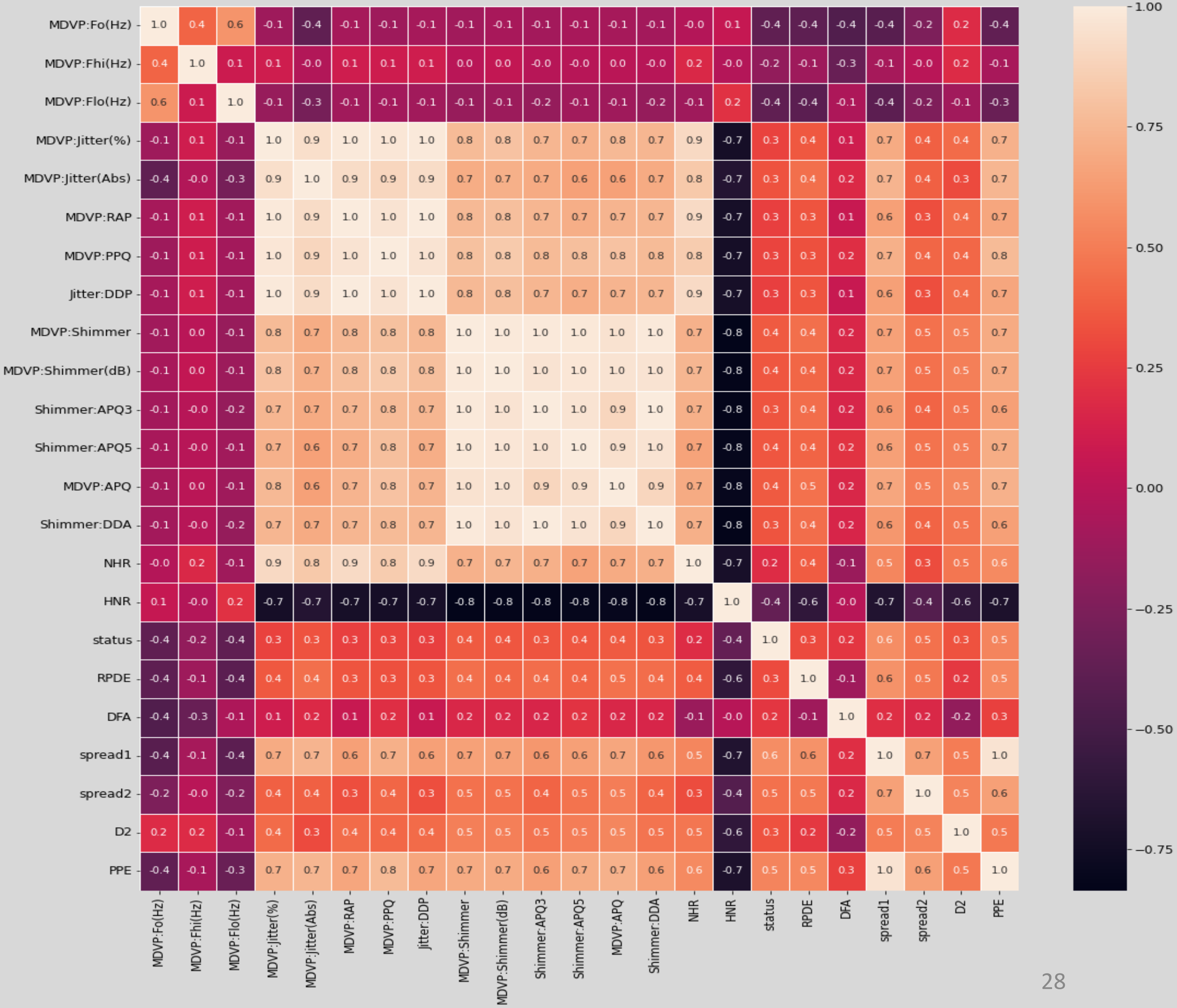
Dataset

Features Histograms:



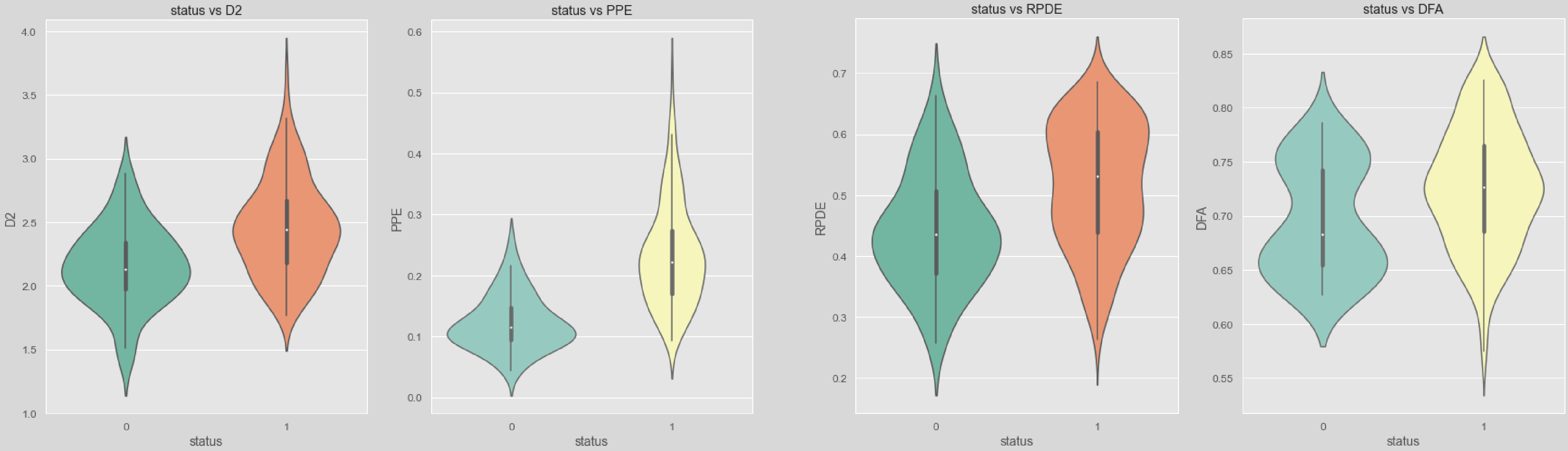
Dataset

Data Visualization of Correlation matrix:

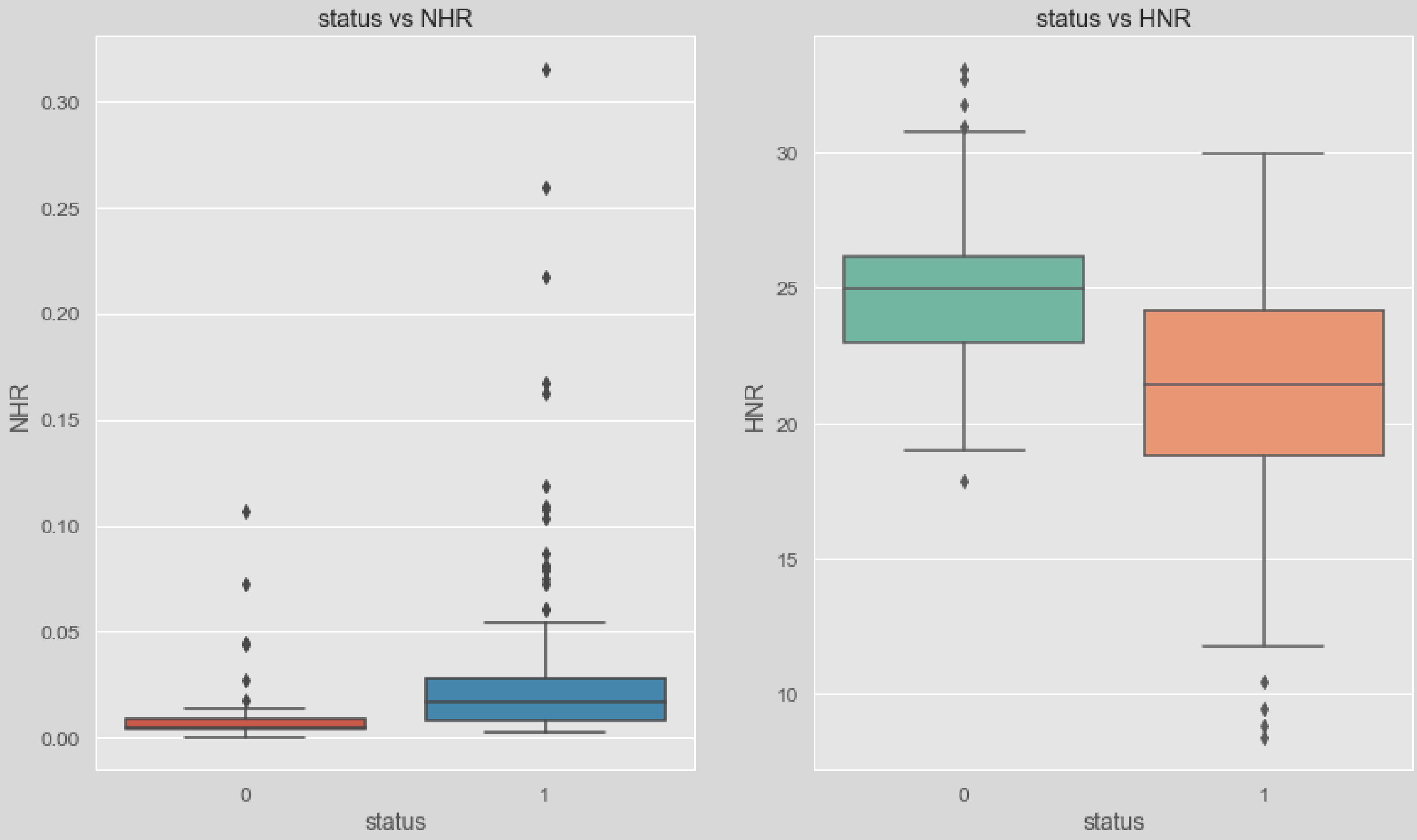


	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	MDVP:Shimmer(dB)	Shimmer:APQ3	Shimmer:APQ5	MDVP:APQ	Shimmer:DDA	NHR	HNR	status	RPDE	DFA	spread1	spread2	D2	PPE
MDVP:Fo(Hz)	1.0	0.4	0.6	-0.1	-0.4	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.0	0.1	-0.4	-0.4	-0.4	-0.4	-0.2	0.2	-0.4
MDVP:Fhi(Hz)	0.4	1.0	0.1	0.1	-0.0	0.1	0.1	0.1	0.0	0.0	-0.0	-0.0	0.0	-0.0	0.2	-0.0	-0.2	-0.1	-0.3	-0.1	-0.0	0.2	-0.1
MDVP:Flo(Hz)	0.6	0.1	1.0	-0.1	-0.3	-0.1	-0.1	-0.1	-0.1	-0.1	-0.2	-0.1	-0.1	-0.2	-0.1	0.2	-0.4	-0.4	-0.1	-0.4	-0.2	-0.1	-0.3
MDVP:Jitter(%)	-0.1	0.1	-0.1	1.0	0.9	1.0	1.0	1.0	0.8	0.8	0.7	0.7	0.8	0.7	0.9	-0.7	0.3	0.4	0.1	0.7	0.4	0.4	0.7
MDVP:Jitter(Abs)	-0.4	-0.0	-0.3	0.9	1.0	0.9	0.9	0.9	0.7	0.7	0.7	0.6	0.6	0.7	0.8	-0.7	0.3	0.4	0.2	0.7	0.4	0.3	0.7
MDVP:RAP	-0.1	0.1	-0.1	1.0	0.9	1.0	1.0	1.0	0.8	0.8	0.7	0.7	0.7	0.7	0.9	-0.7	0.3	0.3	0.1	0.6	0.3	0.4	0.7
MDVP:PPQ	-0.1	0.1	-0.1	1.0	0.9	1.0	1.0	1.0	0.8	0.8	0.8	0.8	0.8	0.8	0.8	-0.7	0.3	0.3	0.2	0.7	0.4	0.4	0.8
Jitter:DDP	-0.1	0.1	-0.1	1.0	0.9	1.0	1.0	1.0	0.8	0.8	0.7	0.7	0.7	0.7	0.9	-0.7	0.3	0.3	0.1	0.6	0.3	0.4	0.7
MDVP:Shimmer	-0.1	0.0	-0.1	0.8	0.7	0.8	0.8	0.8	1.0	1.0	1.0	1.0	1.0	1.0	0.7	-0.8	0.4	0.4	0.2	0.7	0.5	0.5	0.7
MDVP:Shimmer(dB)	-0.1	0.0	-0.1	0.8	0.7	0.8	0.8	0.8	1.0	1.0	1.0	1.0	1.0	1.0	0.7	-0.8	0.4	0.4	0.2	0.7	0.5	0.5	0.7
Shimmer:APQ3	-0.1	-0.0	-0.2	0.7	0.7	0.7	0.8	0.7	1.0	1.0	1.0	1.0	0.9	1.0	0.7	-0.8	0.3	0.4	0.2	0.6	0.4	0.5	0.6
Shimmer:APQ5	-0.1	-0.0	-0.1	0.7	0.6	0.7	0.8	0.7	1.0	1.0	1.0	1.0	0.9	1.0	0.7	-0.8	0.4	0.4	0.2	0.6	0.5	0.5	0.7
MDVP:APQ	-0.1	0.0	-0.1	0.8	0.6	0.7	0.8	0.7	1.0	1.0	0.9	0.9	1.0	0.9	0.7	-0.8	0.4	0.5	0.2	0.7	0.5	0.5	0.7
Shimmer:DDA	-0.1	-0.0	-0.2	0.7	0.7	0.7	0.8	0.7	1.0	1.0	1.0	1.0	0.9	1.0	0.7	-0.8	0.3	0.4	0.2	0.6	0.4	0.5	0.6
NHR	-0.0	0.2	-0.1	0.9	0.8	0.9	0.8	0.9	0.7	0.7	0.7	0.7	0.7	0.7	1.0	-0.7	0.2	0.4	-0.1	0.5	0.3	0.5	0.6
HNR	0.1	-0.0	0.2	-0.7	-0.7	-0.7	-0.7	-0.7	-0.8	-0.8	-0.8	-0.8	-0.8	-0.8	-0.7	1.0	-0.4	-0.6	-0.0	-0.7	-0.4	-0.6	-0.7
status	-0.4	-0.2	-0.4	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.4	0.4	0.3	0.2	-0.4	1.0	0.3	0.2	0.6	0.5	0.3	0.5
RPDE	-0.4	-0.1	-0.4	0.4	0.4	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.5	0.4	0.4	-0.6	0.3	1.0	-0.1	0.6	0.5	0.2	0.5
DFA	-0.4	-0.3	-0.1	0.1	0.2	0.1	0.2	0.1	0.2	0.2	0.2	0.2	0.2	0.2	-0.1	-0.0	0.2	-0.1	1.0	0.2	0.2	-0.2	0.3
spread1	-0.4	-0.1	-0.4	0.7	0.7	0.6	0.7	0.6	0.7	0.7	0.6	0.6	0.7	0.6	0.5	-0.7	0.6	0.6	0.2	1.0	0.7	0.5	1.0
spread2	-0.2	-0.0	-0.2	0.4	0.4	0.3	0.4	0.3	0.5	0.5	0.4	0.5	0.5	0.4	0.3	-0.4	0.5	0.5	0.2	0.7	1.0	0.5	0.6
D2	0.2	0.2	-0.1	0.4	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.5	-0.6	0.3	0.2	-0.2	0.5	0.5	1.0	0.5
PPE	-0.4	-0.1	-0.3	0.7	0.7	0.7	0.8	0.7	0.7	0.7	0.6	0.7	0.7	0.6	0.6	-0.7	0.5	0.5	0.3	1.0	0.6	0.5	1.0

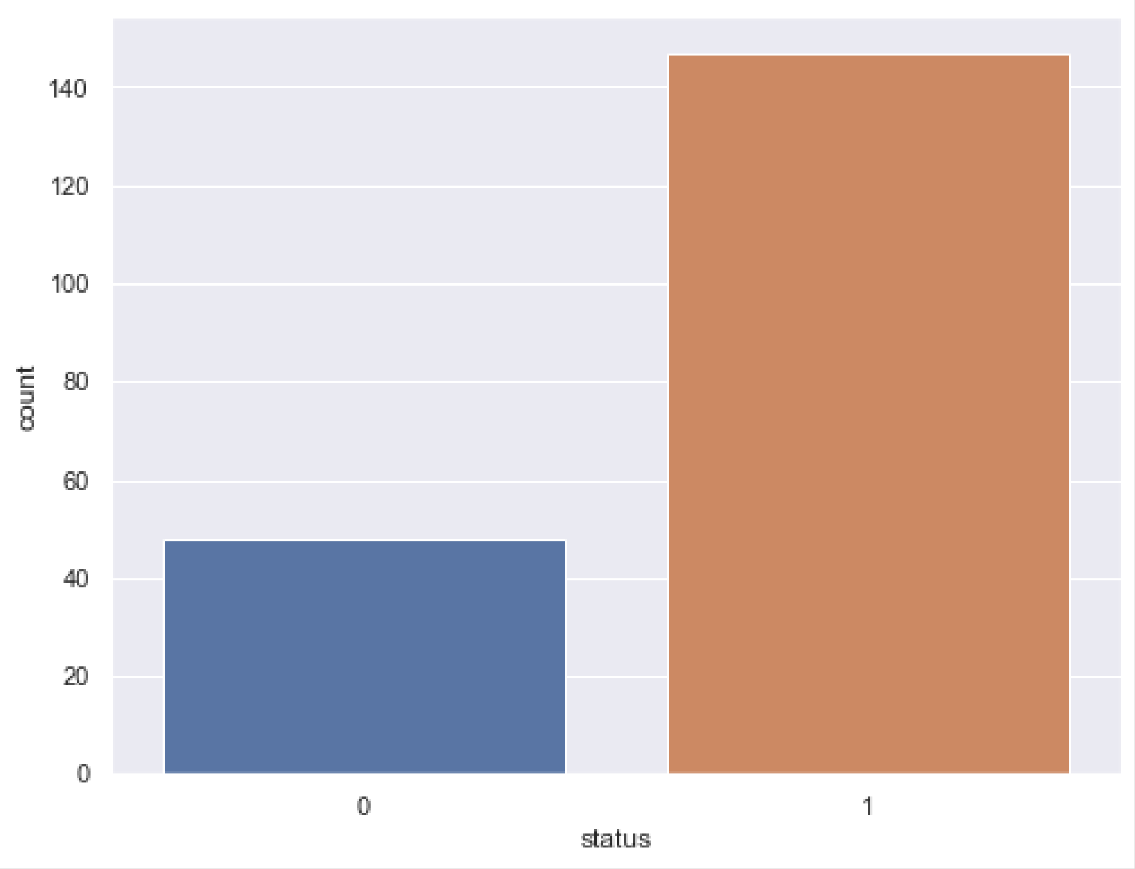
Dataset

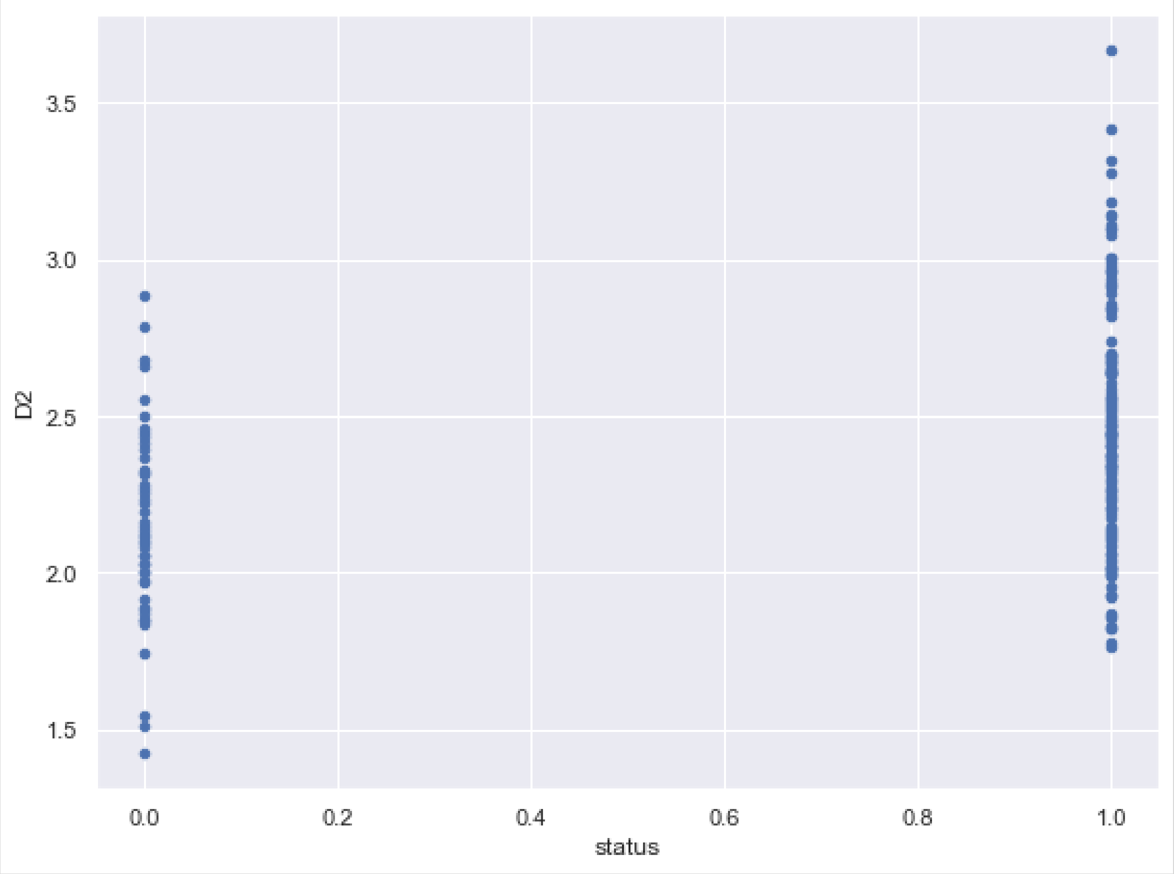

status vs D2
D2
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
1
status
status vs PPE
PPE
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
status
status vs RPDE
RPDE
0.7
0.6
0.5
0.4
0.3
0.2
0
1
status
status vs DFA
DFA
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0
1
status

Dataset

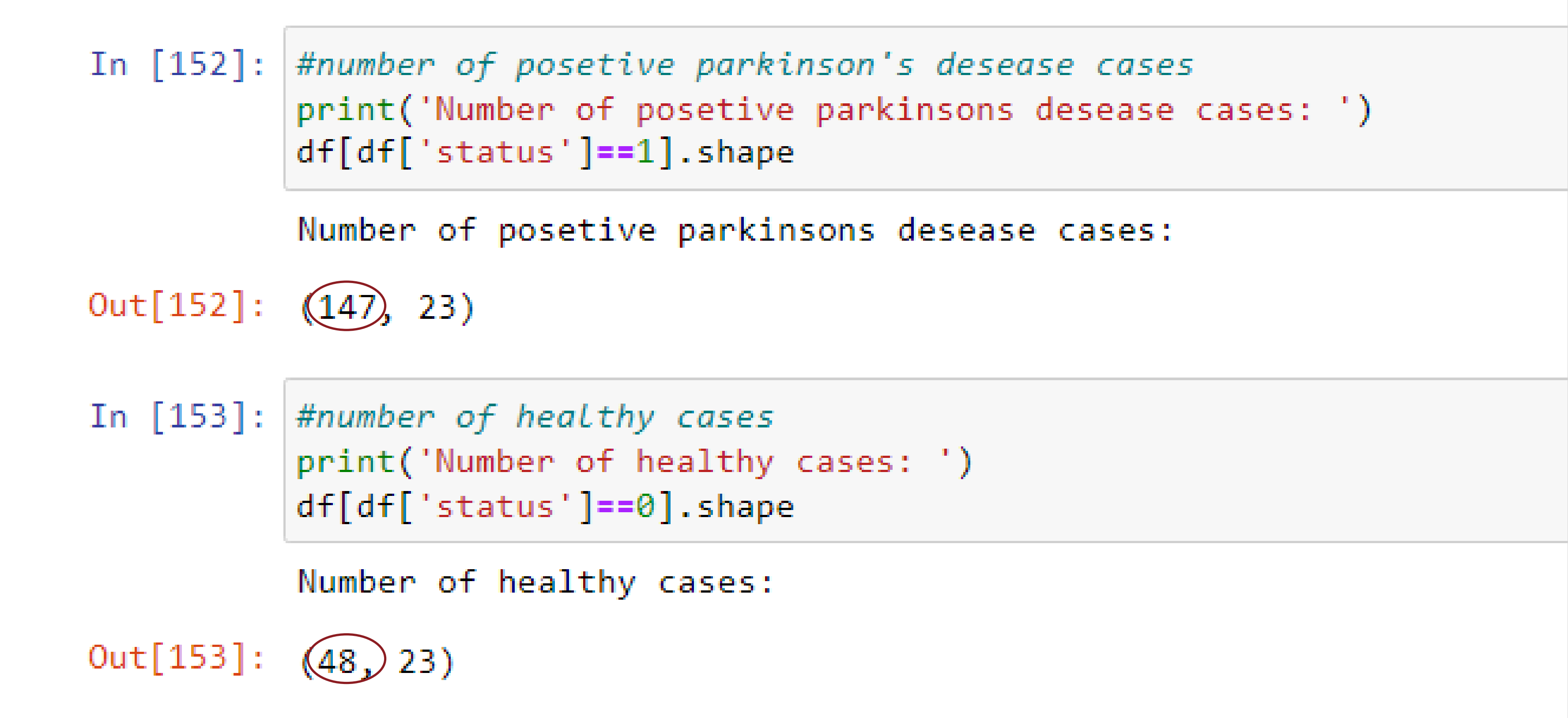

status vs NHR
NHR
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1
status
status vs HNR
HNR
30
25
20
15
10
0
1
status

Dataset


count
140
120
100
80
60
40
20
0
0
1
status

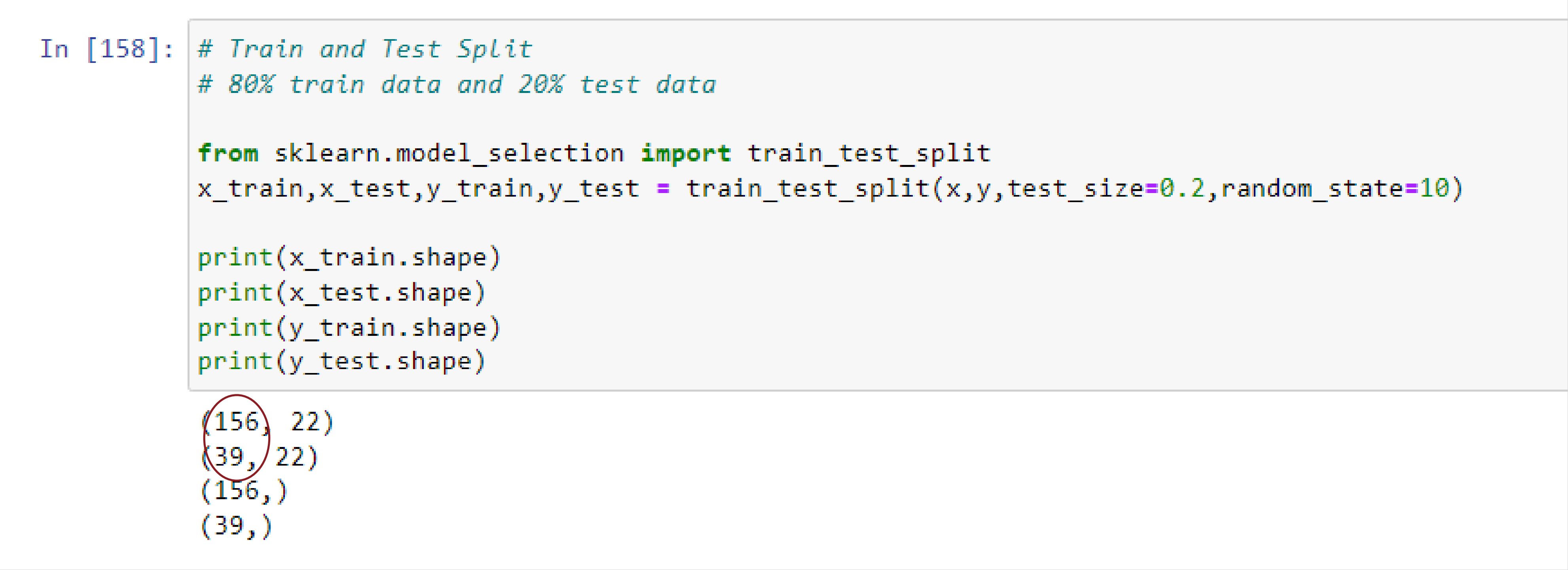

D2
3.5
3.0
2.5
2.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
status

Dataset



```
In [152]: #number of posetive parkinson's desease cases
          print('Number of posetive parkinsons desease cases: ')
          df[df['status']==1].shape

          Number of posetive parkinsons desease cases:

Out[152]: (147, 23)

In [153]: #number of healthy cases
          print('Number of healthy cases: ')
          df[df['status']==0].shape

          Number of healthy cases:

Out[153]: (48, 23)
```

Dataset



```
In [158]: # Train and Test Split
          # 80% train data and 20% test data

          from sklearn.model_selection import train_test_split
          x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=10)

          print(x_train.shape)
          print(x_test.shape)
          print(y_train.shape)
          print(y_test.shape)
```

```
(156, 22)
(39, 22)
(156,)
(39,)
```

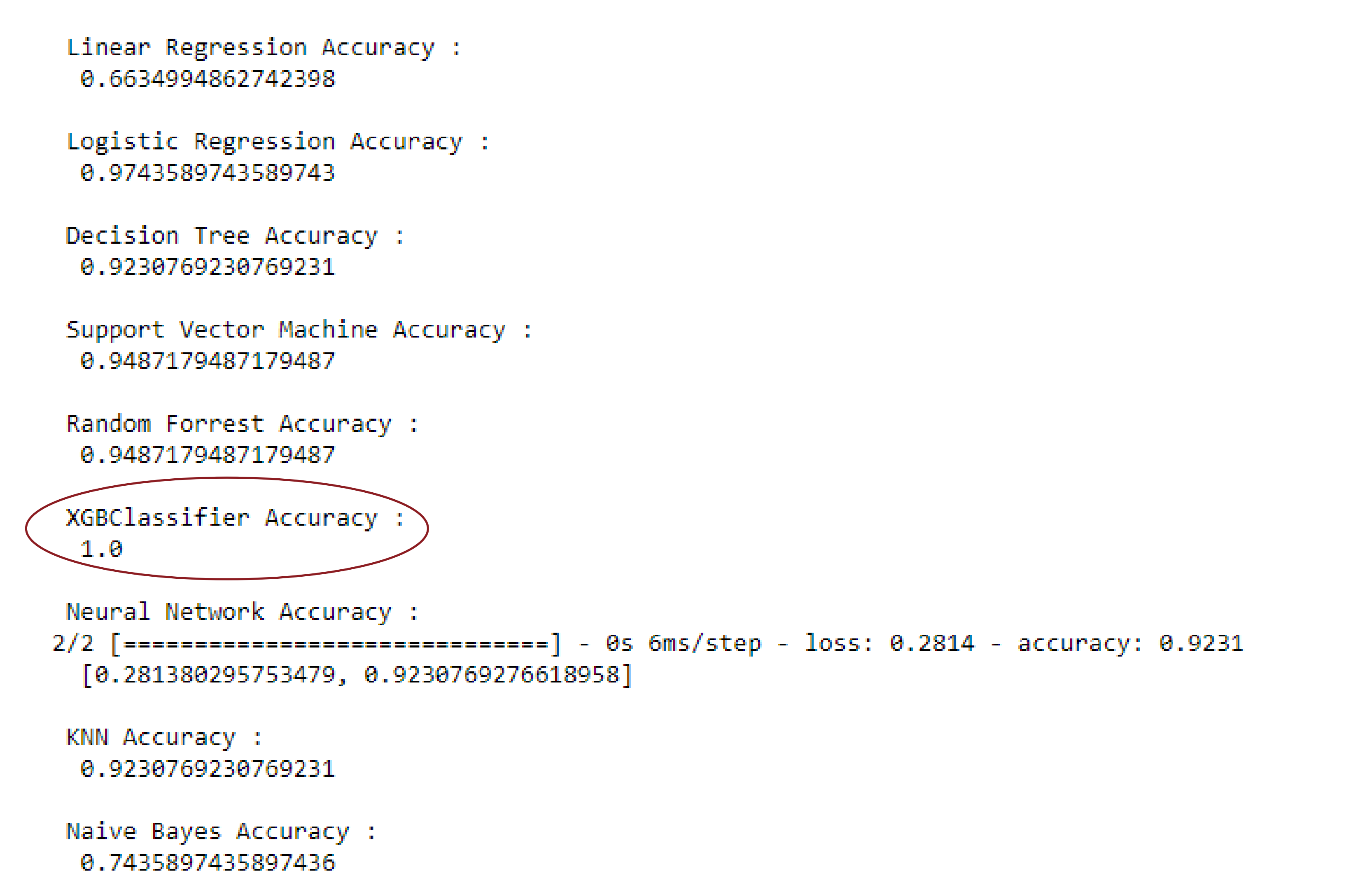
Models

1. **Linear Regression**
2. **Logistic Regression**
3. **Support Vector Machine**
4. **Decision Tree**
5. **Random Forrest**
6. **Extreme Gradient Boosting**
7. **K-Nearest Neighbors**
8. **Naïve Bayes**
9. **Neural Network**

Let’s see the code in python…

Conclusion

Algorithms Comparision



```
 Linear Regression Accuracy :
  0.6634994862742398

 Logistic Regression Accuracy :
  0.9743589743589743

 Decision Tree Accuracy :
  0.9230769230769231

 Support Vector Machine Accuracy :
  0.9487179487179487

 Random Forrest Accuracy :
  0.9487179487179487

 XGBClassifier Accuracy :
  1.0

 Neural Network Accuracy :
2/2 [==============================] - 0s 6ms/step - loss: 0.2814 - accuracy: 0.9231
  [0.281380295753479, 0.9230769276618958]

 KNN Accuracy :
  0.9230769230769231

 Naive Bayes Accuracy :
  0.7435897435897436
```

References

1. *M. Hariharan, K. Polat, and R. Sindhu, "A new hybrid intelligent systems for accurate detection of Parkinson's disease," Comp. Methods Prog. Biomed., Vol. 113, pp. 904–13, 2014.*

2. *G. Singh, and L. Samavedham, "Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of Parkinson disease," J. Neurosci.Methods, Vol. 256, pp. 30–40, 2015.*

3. *A. H. Hadjahmadi and T. J. Askari, "A decision support system for Parkinson's disease diagnosis using classification and regression tree," J. Math. Comp. Sci., Vol. 4, pp. 257–63, 2012.*

4. *Sadek, R. M., Mohammed, S. A., Abunbehan, A. R. K., Ghattas, A. K. H. A., Badawi, M. R., Mortaja, M. N., ... & Abu-Naser, S. S. (2019). Parkinson's Disease Prediction Using Artificial Neural Network.*