

Parkinson Disease Prediction

Miss. Amreen Khan¹, Abhishek Rajesh Shriwas², Shruti Vinod Shinde³, Dewam Katole⁴

Assistant Professor, CSE, G. H. Raisoni Institute of Engineering And Technology, Nagpur, India¹

Students, CSE, G. H. Raisoni Institute of Engineering and Technology, Nagpur, India^{2,3,4}

Abstract: Parkinson's disease is a progressive disorder of the central nervous system affecting movement and inducing tremors and stiffness. It has 5 stages to it and affects more than 1 million individuals every year in India. This is chronic and has no cure yet. It is a neurodegenerative disorder affecting dopamine-producing neurons in the brain. With the increase in the severity of the disease, the patient's voice gets more and more deteriorated. The non-invasive treatments for voice analysis are available that helps in ameliorating the life quality of a patient. Thus, for building the telemonitoring and tediagnosis models for prediction, the speech analysis has been tremendously increased. The proper interpretation of speech signals is one of the important classification problems for Parkinson's disease diagnosis. Deep learning and machine learning techniques have been used as a part of the discovery for the efficient classification of PD. The various classification models like support vector machines, naive Bayes, deep neural networks, decision tree and random forest are effectively employed for classification purposes. The analysis of results of different research works showed that both machine learning and deep learning algorithms have shown promising future and therefore paving a better way for the detection of Parkinson's disease at its earlier stages. The classification accuracy achieved by the machine learning classifier.

I. INTRODUCTION

Parkinson's disease (PD) is currently the second most frequent neurodegenerative disease, after Alzheimer disease. Generally, there are two kinds of symptoms of PD, motor symptoms and non-motor symptoms. The main motor symptoms of PD are tremor, slowness of movement (bradykinesia), stiffness (rigidity), and poor balance (postural instability). Non-motor symptoms mainly include mood disorders, cognitive dysfunction, pain, sensory dysfunction, and dysautonomia. Motor speech disorders are common among PD patients. Speech disturbances such as very quiet and hurried speech occur in more than half of the patients. Analysis of speech signals is considered as an important non-invasive method for PD identification.

Non-invasive identification and prediction technology of PD is attractive to clinicians and neuroscientist. In addition, detection of voice changes in PD patients would make it possible for early detection and intervention before the onset of disabling physical symptoms.

Parkinson's Disease (PD) is a debilitating neurodegenerative disease which cannot be diagnosed through standardized blood tests, so a faster, cheaper diagnostic tool is essential. Using machine learning algorithms to analyse the variations in voice patterns is a novel method of predicting the existence of PD in patients. This paper proposes a predictive model that effectively diagnoses PD with maximum accuracy using a dataset that consists of extrapolated data from voice recordings of Parkinson's patients and unaffected subjects.

Parkinson's Disease is a chronic, progressive disease which affects movement throughout the body. There are many symptoms of PD, including tremors, bradykinesia, muscle rigidity, impaired balance, micrographic, changes in facial expressions, orthostatic hypotension, and many more. PD is a devastatingly widespread movement disorder, with nearly one million Americans currently suffering from it scientists have been looking into creating machine learning models to predict the symptoms of PD without conducting invasive clinical tests. There has also been work regarding the diagnosis of PD by analysing existing symptoms using machine learning techniques. One symptom of PD that has not been extensively researched is the effects it has on vocal patterns.

Researchers hypothesize that the disease can be diagnosed with a short voice recording, done even on a common smartphone. With smartphones and cheap recording devices being as ubiquitous as they currently are, there exists a need for machine learning models that are compact, lightweight, and accurate enough to detect voice patterns from a limited amount of data.

II. PROBLEM DEFINATION

Prediction and Detection of Parkinson Disease using Machine Learning algorithms with the Dataset. So far existing systems are all about Parkinson Disease (PD) is detected at the secondary stage which leads to medical facilities and challenges. PD can generally be diagnosed with the following methods like MRI or CT scan but are not used to detect early signs of PD. Use Python and Jupyter Notebook for Data Visualization and Data Analytics.

III. ANALYSIS AND DESIGN

Hardware Requirements:

- Laptop with Windows 10 and above
- 8GB RAM and above
- Any Processor will be compatible (Intel or Ryzen)
- 250 MB HDD.

Software Requirements:

- PYTHON (IDLE VERSION 3.7 AND ABOVE)
- JUPYTER NOTEBOOK

Python is an interpreted high-level general-purpose programming language. Its design philosophy emphasizes code readability with its use of significant indentation. Its language constructs as well as its object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala. The best environment to capture this style of programming is Jupyter notebooks. You can work on bits of code, spell out your thoughts and when you are satisfied, put it all together.

Matplotlib You can generate static, dynamic, and interactive visuals with the Python programme Matplotlib. Matplotlib makes both challenging and basic tasks possible.

Seaborn A Python data visualisation toolkit called Seaborn uses the matplotlib library. It provides a sophisticated interface for designing eye-catching and illuminating statistical graphics. Seaborn helps you explore and understand your data.

Scikit or Sklearn Without a doubt, Python's most useful machine learning library is Scikit-learn. The sklearn toolkit for machine learning and statistical modelling includes a number of helpful features, including classification, regression, clustering, & dimensionality reduction. With sklearn, machine learning models were developed. Streamlit Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps. It is a Python-based library specifically designed for machine learning engineers. Data scientists or machine learning engineers are not web developers and they're not interested in spending weeks learning to use these frameworks to build web apps.

IV. SYSTEM FLOW

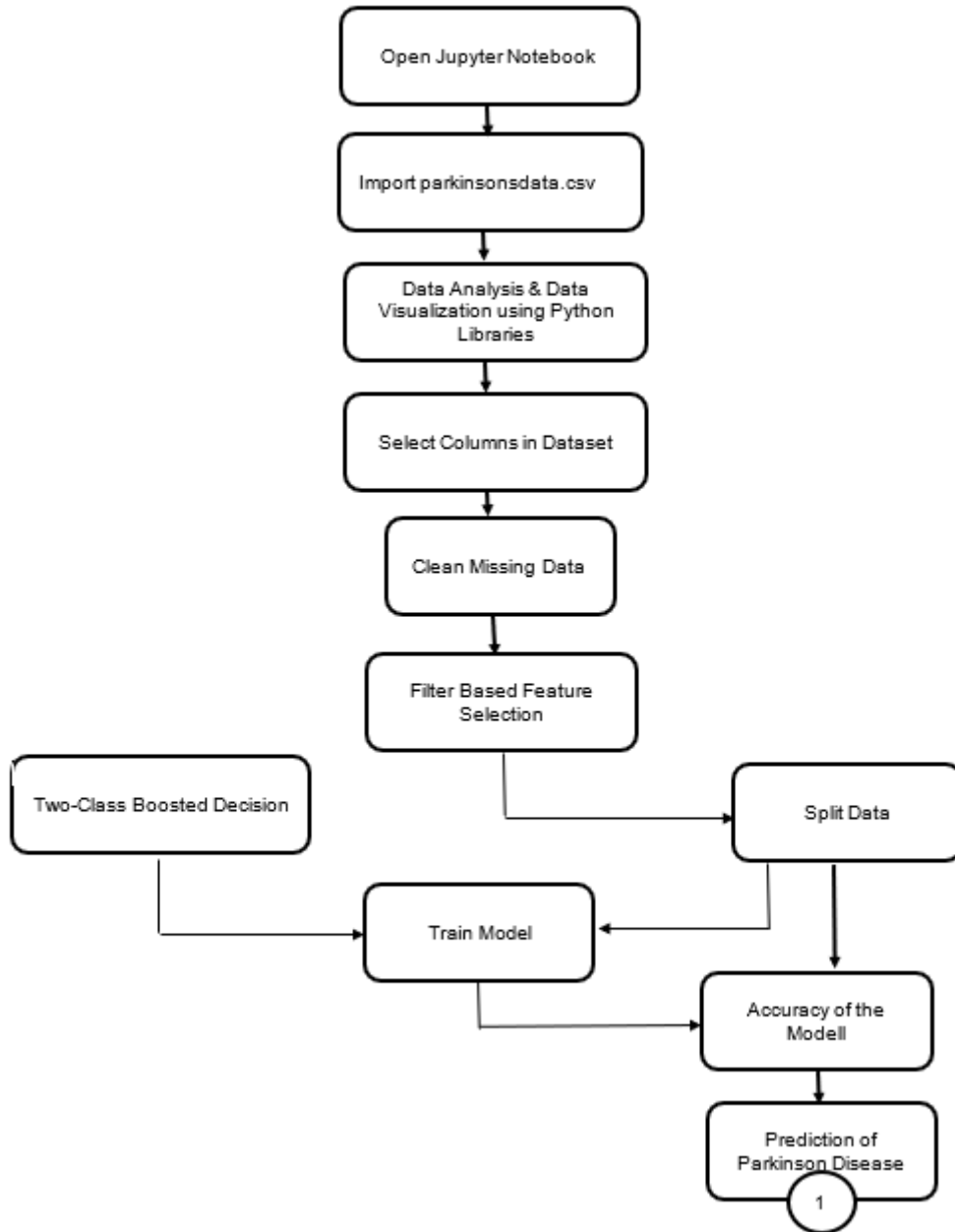


Figure 3.1: Flowchart of the Model

V. DATA FLOW DIAGRAM

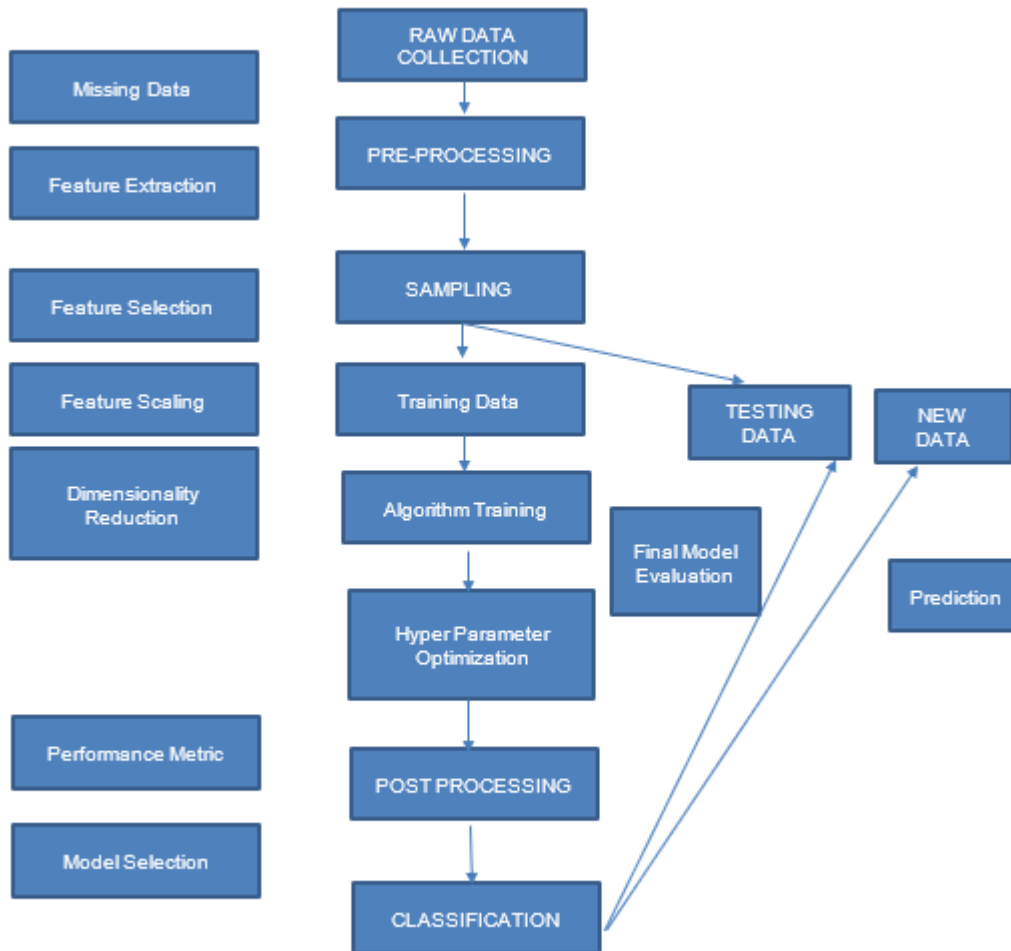


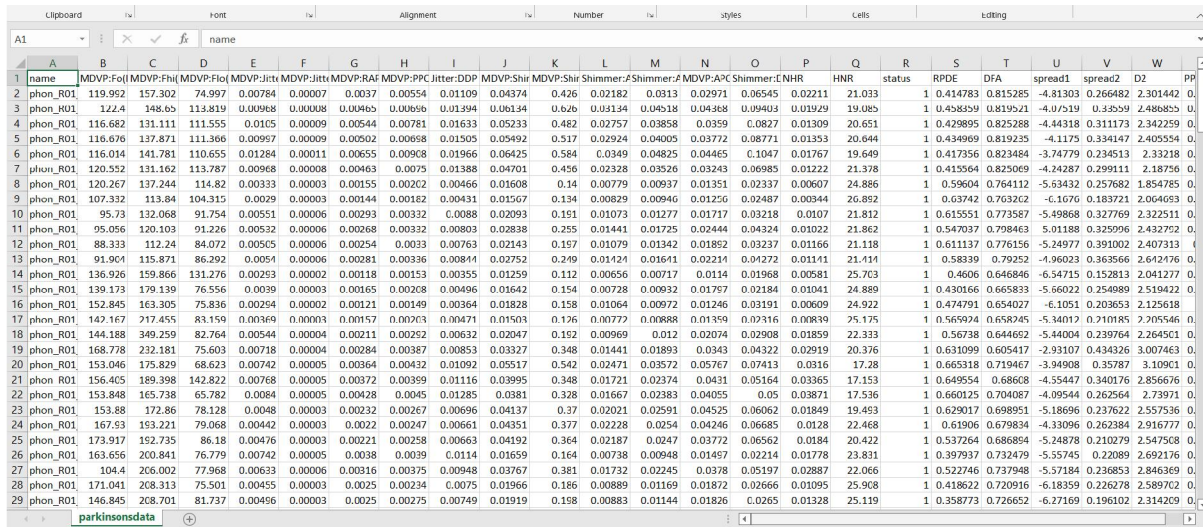
Figure 3.2: Dataflow Diagram of the Project

VI. IMPLEMENTATION & RESULTS

To explore the data, I utilised Pandas, NumPy, matplotlib, and seaborn libraries. We created a model using the Sklearn library as well. After imported everything, we chose a present resolution for a graph to display at while I was exploring the data. Then I used pandas to import our dataset.

Based on the characteristics offered, we can conclude that pandas are the finest data processing tool. It can operate with different file kinds, maintain missing data, clean up data, and manage missing data. This suggests that it has the ability to read or write the data in a wide range of formats, include CSV, Excel, SQL, and many others.

The biological voice measures in this collection come from 31 people, 23 of whom suffer from Parkinson's disease (PD). Each row in the table correlates to being one of the 195 voice recordings that these persons made, and each column in the table denotes a distinct voice measure ("name" column). The "status" column, which would be set to 0 for healthy individuals and 1 for those with Parkinson's disease, is used to distinguish between healthy individuals and those with the condition.

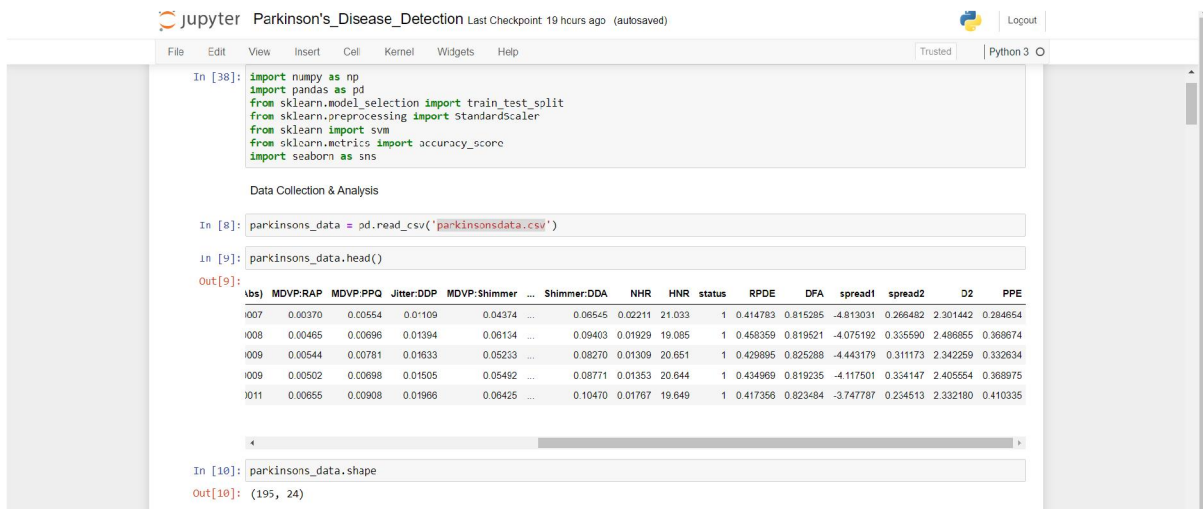


A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
name	MDVP:F0(F0)	MDVP:F1(F1)	MDVP:F2(F2)	MDVP:F3(F3)	MDVP:F4(F4)	MDVP:F5(F5)	MDVP:F6(F6)	MDVP:F7(F7)	MDVP:F8(F8)	MDVP:F9(F9)	MDVP:F10(F10)	MDVP:F11(F11)	MDVP:F12(F12)	MDVP:F13(F13)	MDVP:F14(F14)	MDVP:F15(F15)	MDVP:F16(F16)	MDVP:F17(F17)	MDVP:F18(F18)	MDVP:F19(F19)	MDVP:F20(F20)	MDVP:F21(F21)	MDVP:F22(F22)
phon_R01	119.952	157.302	74.997	0.00784	0.00007	0.0037	0.00554	0.01109	0.04374	0.426	0.02182	0.0313	0.02971	0.06545	0.02211	21.033	1	0.414783	0.815285	-4.81303	0.266482	2.301442	0.
phon_R01	122.4	148.65	113.819	0.00968	0.00008	0.00465	0.00656	0.01394	0.06134	0.626	0.03134	0.04518	0.04368	0.09403	0.01929	19.085	1	0.498359	0.819521	-4.07519	0.33559	2.486855	0.
phon_R01	116.662	131.111	111.555	0.0105	0.00009	0.00544	0.00761	0.01633	0.05233	0.482	0.02757	0.03858	0.0359	0.0827	0.01309	20.651	1	0.429895	0.825288	-4.44318	0.311173	2.342259	0.
phon_R01	116.676	127.871	111.366	0.00997	0.00009	0.00502	0.00658	0.01505	0.05492	0.517	0.02924	0.04005	0.03712	0.08771	0.01353	20.644	1	0.434969	0.819235	-4.1175	0.334147	2.405554	0.
phon_R01	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00968	0.01966	0.06425	0.584	0.0349	0.04825	0.04465	0.1047	0.01767	19.649	1	0.417356	0.823484	-3.74779	0.234513	2.332218	0.
phon_R01	120.552	131.162	113.787	0.00968	0.00008	0.00463	0.0075	0.01388	0.04701	0.456	0.02328	0.03526	0.03243	0.06985	0.01222	21.378	1	0.415564	0.825069	-4.24287	0.299111	2.18756	0.
phon_R01	120.267	137.244	114.82	0.00333	0.00003	0.00155	0.00202	0.00466	0.01608	0.14	0.00779	0.00937	0.01351	0.02337	0.00607	24.866	1	0.59604	0.764112	-5.53432	0.257682	1.854785	0.
phon_R01	107.332	113.84	104.315	0.0029	0.00003	0.00144	0.00162	0.00431	0.01567	0.134	0.00829	0.00946	0.01256	0.02487	0.00344	26.892	1	0.63742	0.763262	-6.1676	0.183721	2.004663	0.
phon_R01	95.73	132.068	91.754	0.00551	0.00006	0.00293	0.00332	0.0088	0.02093	0.151	0.01073	0.01277	0.01717	0.03218	0.0107	21.812	1	0.615551	0.773567	-5.49868	0.327769	2.252511	0.
phon_R01	95.056	120.103	91.226	0.00532	0.00006	0.00268	0.00332	0.00803	0.02838	0.255	0.01441	0.01725	0.02444	0.04324	0.01022	21.862	1	0.547037	0.798463	-5.01188	0.325996	2.432792	0.
phon_R01	88.333	112.24	84.072	0.00505	0.00006	0.00254	0.0033	0.00763	0.02143	0.157	0.01079	0.01342	0.01892	0.03237	0.01166	21.118	1	0.611137	0.776156	-5.24977	0.391002	2.407313	0.
phon_R01	91.904	115.871	86.292	0.0051	0.00006	0.00281	0.00336	0.00844	0.02752	0.249	0.01421	0.01611	0.02214	0.04272	0.01141	21.114	1	0.58339	0.79252	-4.96023	0.363566	2.642176	0.
phon_R01	136.926	159.866	131.276	0.00253	0.00002	0.00118	0.00153	0.00355	0.01259	0.112	0.00656	0.00717	0.0114	0.01968	0.00581	25.703	1	0.4606	0.646846	-6.54715	0.152813	2.041277	0.
phon_R01	139.173	179.139	76.556	0.0039	0.00003	0.00165	0.00208	0.00496	0.01642	0.154	0.00728	0.00932	0.01767	0.02184	0.01041	24.889	1	0.430166	0.665833	-5.66022	0.254989	2.519422	0.
phon_R01	152.845	163.305	75.836	0.00254	0.00002	0.00121	0.00149	0.00364	0.01828	0.158	0.01064	0.00972	0.01246	0.03191	0.00609	24.922	1	0.474791	0.654027	-6.1051	0.203653	2.125618	0.
phon_R01	142.167	217.455	83.159	0.00369	0.00003	0.00157	0.00203	0.00471	0.01503	0.126	0.00777	0.00888	0.01359	0.02316	0.00839	25.175	1	0.565924	0.682495	-5.34017	0.210185	2.705546	0.
phon_R01	144.188	349.259	82.764	0.00544	0.00004	0.00211	0.00252	0.00632	0.02047	0.152	0.00969	0.012	0.02074	0.02908	0.01859	22.333	1	0.56738	0.644662	-5.44004	0.239764	2.264561	0.
phon_R01	168.778	232.181	75.603	0.00718	0.00004	0.00284	0.00387	0.00853	0.03327	0.348	0.01441	0.01893	0.0343	0.04322	0.02919	20.376	1	0.631099	0.605417	-2.93107	0.434326	3.007463	0.
phon_R01	153.046	175.829	68.623	0.00742	0.00005	0.00364	0.00432	0.01092	0.05517	0.542	0.02471	0.03572	0.05767	0.07413	0.0316	17.28	1	0.665318	0.719467	-3.94908	0.35787	3.10901	0.
phon_R01	156.405	169.398	142.822	0.00768	0.00005	0.00372	0.00359	0.01116	0.03995	0.348	0.01721	0.02374	0.0431	0.05164	0.03365	17.153	1	0.649554	0.68668	-4.55447	0.340176	2.856676	0.
phon_R01	153.848	165.738	65.782	0.0068	0.00005	0.00428	0.0045	0.01285	0.0381	0.328	0.01667	0.02383	0.04055	0.05	0.03871	17.536	1	0.660125	0.704087	-4.09544	0.262564	2.73971	0.
phon_R01	153.88	172.86	78.138	0.00648	0.00003	0.00232	0.00267	0.00696	0.04137	0.37	0.02021	0.02591	0.04525	0.06062	0.01849	19.453	1	0.629017	0.698951	-5.18696	0.237622	2.557536	0.
phon_R01	167.63	153.221	79.058	0.00442	0.00003	0.0022	0.00247	0.00661	0.04351	0.377	0.02228	0.0254	0.04246	0.06685	0.0128	22.468	1	0.61906	0.679834	-4.33096	0.262384	2.916777	0.
phon_R01	173.917	152.735	86.18	0.00476	0.00003	0.00221	0.00258	0.00663	0.04192	0.364	0.02187	0.0247	0.03772	0.06562	0.0184	20.422	1	0.537264	0.686884	-5.24878	0.210279	2.547508	0.
phon_R01	163.566	200.841	76.779	0.00742	0.00005	0.0038	0.0039	0.0114	0.01659	0.164	0.00738	0.00948	0.01457	0.02214	0.01778	23.831	1	0.397937	0.732479	-5.55745	0.22089	2.692176	0.
phon_R01	104.4	206.002	77.968	0.00633	0.00006	0.00316	0.00375	0.00948	0.03767	0.381	0.01372	0.02245	0.0378	0.05197	0.02887	22.066	1	0.522746	0.737948	-5.7184	0.236853	2.846369	0.
phon_R01	171.041	208.313	75.501	0.00455	0.00003	0.0025	0.00234	0.0075	0.01966	0.186	0.00889	0.01169	0.01872	0.02666	0.01095	25.908	1	0.418622	0.720916	-6.18359	0.226278	2.589702	0.
phon_R01	146.845	208.701	81.737	0.00456	0.00003	0.0025	0.00275	0.00749	0.01919	0.158	0.00883	0.01144	0.01826	0.02665	0.01328	25.119	1	0.358773	0.726652	-6.27169	0.196102	2.314209	0.

Figure: Dataset of Parkinson Disease

VII. OUTPUT FORMS WITH DATA

We discovered that the data consists of 195 rows and 24 columns by using "shape" Then I used the.info extension to validate the types of data of each column. We then looked at the null values. Null values are merely the absence of data. () is essentially utilized to determine whether or not our dataset contains the 11-missing data. There were no missing values in the dataset, we discovered. We then witness several statistical calculations of our dataset, such as mean, median, count, etc., using. Describe ().



```

In [38]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
import seaborn as sns

Data Collection & Analysis

In [8]: parkinsons_data = pd.read_csv('parkinsonsdata.csv')

In [9]: parkinsons_data.head()

Out[9]:
  (bs) MDVP:RAP MDVP:PPQ Jitter:DDP MDVP:Shimmer ... Shimmer:DDA NHR NHR status RPDE DFA spread1 spread2 D2 PPE
007 0.00370 0.00554 0.01109 0.04374 ... 0.06545 0.02211 21.033 1 0.414783 0.815285 -4.813031 0.266482 2.301442 0.284654
008 0.00465 0.00696 0.01394 0.06134 ... 0.09403 0.01929 19.085 1 0.458359 0.819521 -4.075192 0.335550 2.486855 0.368674
009 0.00544 0.00781 0.01633 0.05233 ... 0.08270 0.01309 20.651 1 0.429895 0.825288 -4.443179 0.311173 2.342259 0.332634
009 0.00502 0.00698 0.01505 0.05492 ... 0.08771 0.01353 20.644 1 0.434969 0.819235 -4.117501 0.334147 2.405554 0.368975
011 0.00655 0.00908 0.01966 0.06425 ... 0.10470 0.01767 19.649 1 0.417356 0.823484 -3.747787 0.234513 2.332218 0.410335

In [10]: parkinsons_data.shape
Out[10]: (195, 24)

```

Figure: Importing Dataset into Jupyter Notebook

The Support Vector Machine, or SVM, is a popular Supervised Learning method that may be used to solve both classification and regression issues. However, it is mostly utilised in Machine Learning for Classification difficulties. The goal of the SVM algorithm is to determine the best decision boundary or line for classifying n-dimensional space into groups so that future additions of data points can be quickly assigned to the appropriate group. The border of the ideal choice is called a hyperplane.

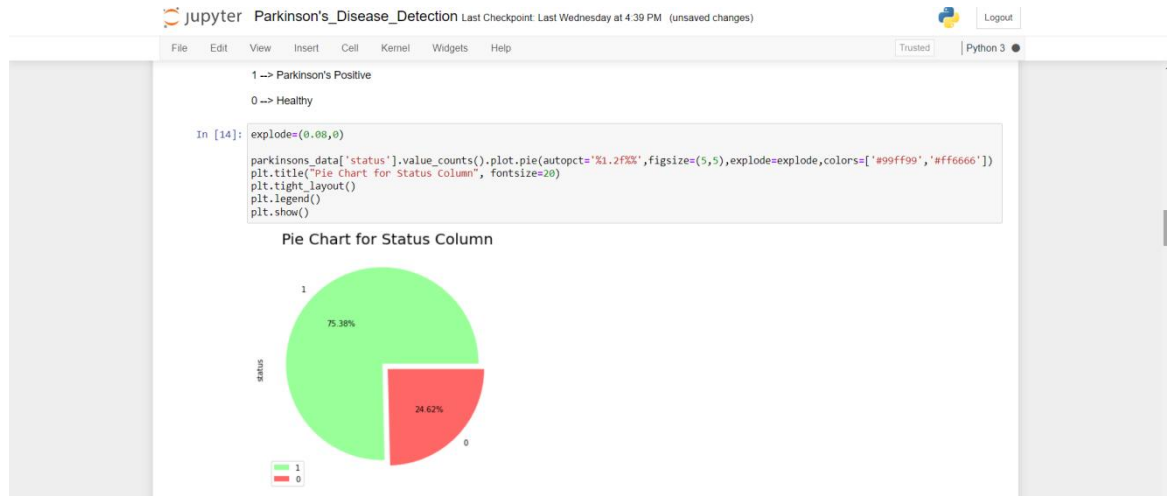


Figure 4.3: Data Visualization

Support vector machines (SVMs) were robust yet adaptable supervised machine learning techniques used for outlier detection, regression, and classification. SVMs are frequently employed in classification issues and are extremely effective in large dimensional spaces. Because they only use a portion of the training points inside the decision function, SVMs are well-liked and memory-efficient algorithms.

To determine the maximum marginal hyperplane (MMH), SVMs divide datasets into a number of classes, which can be accomplished in the following two steps: Support Vector Machines would first produce hyperplanes that best divide the classes through iterative generation. The hyperplane that properly separates the classes will then be selected.

Data Splitting is done to avoid a small concept called overfitting, a process where machine learning dataset is split into further sets to avoid wrong prediction.



Figure: Splitting Dataset into Training and Testing

Accuracy of our Test Data and Train data is shown below where we reached accuracy of 87.00%

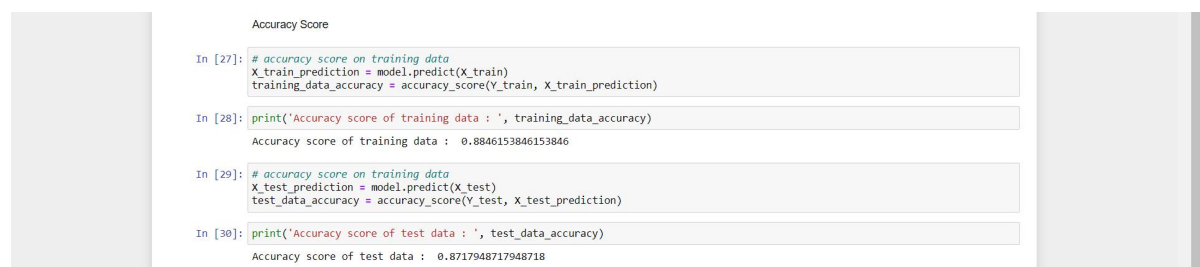


Figure: Accuracy of the Model

VIII. TESTING

By feeding the values manually our machine learning model using Support vector machine will tell us whether the person is suffering from Parkinson Disease or not and the accuracy of our model is also accurate and good.

REFERENCES

- [1] C. W. Olanow, M. B. Stern, and K. Sethi, "The scientific and clinical basis for the treatment of Parkinson disease," *Neurology*, vol. 72(21 Supplement 4), pp. S1-S136, 2009.
- [2] M. C. de Rijk, L. J. Launer, K. Berger, M. M. Breteler, J. F. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. M. Lage, C. Trenkwalder, and A. Hofman, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group," *Neurology*, vol. 54(11 Suppl 5), pp. S21-3, 2000.
- [3] A. E. Lang, and A. M. Lozano, "Parkinson's disease—First of two parts," *New. Engl. J. Med.*, vol. 339, pp. 1044-1053, 1998.
- [4] J. W. Langston, "Parkinson's disease: current and future challenges," *Neurotoxicology*, vol. 23(4-5), pp. 443-450, 2002.
- [5] S. K. Van Den Eeden, C. M. Tanner, A. L. Bernstein, R. D. Fross, A. Leimpeter, D. A. Bloch, and L. M. Nelson, "Incidence of Parkinson's disease: variation by age, gender, and race/ ethnicity," *American journal of epidemiology*, vol. 157(11), pp. 1015-1022, 2003.
- [6] D. M. Huse, K. Schulman, L. Orsini, J. Castelli-Haley, S. Kennedy, and Lenhart, "Burden of illness in Parkinson's disease. Movement disorders: official journal of the Movement Disorder Society," vol. 20(11), pp. 1449-1454, 2005.
- [7] J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients," *Journal of Speech and hearing Disorders*, vol. 43(1), pp. 47-57, 1978.
- [8] D. J. Gelb, E. Oliver, and S. Gilman, "Diagnostic criteria for Parkinson disease," *Arch Neurol.*, vol. 56, pp. 33-39, 1999.
- [9] A. Tsanas, M. A. Little, P. E. McSharry, and L. O Ramig, "Accurate telemonitoring of Parkinson's disease progression by non-invasive speech tests," *IEEE transactions on Biomedical Engineering*, vol. 57(4), pp. 884-893, 2009.
- [10] K. S. Perez, L. O. Ramig, M. E. Smith, and C. Dromey, "The Parkinson larynx: tremor and video stroboscopic findings," *Journal of Voice*, vol. 10(4), pp. 354- 361, 1996.
- [11] P. Bugalho, and M. Viana-Baptista, "REM sleep behavior disorder and motor dysfunction in Parkinson's disease - A longitudinal study," *Park Relat Disord*, vol. 19(12), pp.1084-1087, 2013.
- [12] N. Singh, V. Pillay, and Y. E. Choonara, "Advances in the treatment of Parkinson's disease," *Progress in neurobiology*, vol. 81(1), pp. 29-44, 2007.
- [13] B. E.Sakar, M. E. Isenkul, C. O. Sakar, A. Sertbas, F. Gurgun, S. Delil, and O. Kursun, "Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 17(4), pp. 828-834, 2013.
- [14] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56(4), pp. 1010 -1022, 2009.