

DETECTION OF PARKINSON'S USING MACHINE LEARNING TECHNIQUES

Abstract

Technology plays a crucial role in our lifestyle, healthcare, and resource and asset maintenance in this global era. Technology has been developing at a rapid rate to combat the various diseases and symptoms that are currently affecting the world. Parkinson's disease is one such condition.

Parkinson's disease is a degenerative disorder that has far-reaching effects on the nervous system and all the parts of the body that it regulates. Large medical datasets are available in many data sources and are utilized to diagnose the illnesses. It is one of the most debilitating and chronic movement disorders. It is currently untreatable but is the second most disorders of the Nervous System that causes functional impairment and shortens life expectancy. Prediction is vital to patient recovery in the early phases. Machine learning can facilitate this procedure.

Keywords: Parkinson's disease, Machine learning, Decision tree, XGBoost

Authors

Anirudh Gupta

Department of Information Technology
Manipal University Jaipur
Jaipur
anirudh.189302167@mu.j.manipal.edu

Anurag Bhatnagar

Department of Information Technology
Manipal University Jaipur
Jaipur
anurag.bhatnagar@jaipur.manipal.edu

Nikhar Bhatnagar

Department of Information Technology
Manipal University Jaipur
Jaipur
nikharbhatnagar@gmail.com

I. INTRODUCTION

Parkinson's disease (PD) is a condition of the degenerating nervous system that impairs movement by producing stiffness, tremor, and issues with balancing, coordinating, and moving. In most cases, Parkinson's disease begins with very modest symptoms before gradually worsening over time. It affects mostly people who are more than 50 years of age, there are also cases when people below that age also get affected with it and that cases is called young-onset Parkinson's or YOPD.

Technology plays an essential role in our lifestyle and in the field of health care, it has been developing at a rapid rate to combat the various diseases and symptoms that are currently affecting the world. Parkinson's disease is one such condition. A neurological condition affecting the brain, Parkinson's Disease. It also makes the body stiff and causes tremors in the hands and body. There is no effective treatment or cure available currently. Treatment is only possible when the condition is discovered early or at its onset. In addition to lowering the cost of the disease, these may also save lives. As a result, we employ numerous machine learning strategies and Python libraries to create a model that can accurately detect the presence of sickness in an individual's body. To diagnose disease, the current models rely on either image or audio analysis. We encourage the creation of a new model that uses both.

1. **Objectives:** Understanding Parkinson's disease and identifying the disease's early signs are the key goals.
2. **Motivation of the Research:** Even though Parkinson's One of the most terminal, untreatable, and neurodegenerative conditions is Parkinson's disease. It can't be cured or prevented. People suffer from it for a long time before they are diagnosed.

According to the anticipated outcomes, In the world, 7–10 million people suffer with Parkinson's disease. Even though people over 50 have a higher risk of developing Parkinson's disease, only about 4% of people under 50 are diagnosed with the disease. Death results if these symptoms are not discovered early on.

As a result, researchers have only discovered the symptoms and the underlying causes of the disease up to this point. However, very few symptoms have been alleviated, and many others remain unsolved. a screening method for the disease is urgently needed. It's crucial that we develop a way to anticipate it in its earliest phases.

3. **Research Scope:** The goals of this research are to (1) validate existing PD diagnostic tool identification methods, and (2) classify people as either healthy or affected by PD utilizing machine learning-based algorithms and techniques.
4. **Proposed System:** Machine learning techniques can handle the issue with the fewest possible errors. The UCI device acquiring library's Parkinson's disease audio dataset is used as input. When the spiral design inputs of people without Parkinson's and those with the condition are combined, our suggested strategy yields reliable findings. Due to the combination of all the impacts, the doctor can determine if anything is normal or abnormal and then give medication based only on the afflicted stage.

5. **Overview of the Research:** Movement is disrupted by Parkinson's disease, an emerging neurodegenerative disease, Parkinson's disease. A slight tremor in one hand may be the first indicator of a progressive set of symptoms. While tremors are characteristic of this disorder, stiffness or slowness of movement are also common. In the beginning of Parkinson's, your facial expressions may be flat. Maybe you shouldn't wander about with your hands outstretched. It may also cause your speech to become slurred or weak. Parkinson's disease symptoms grow as the disease progresses. Medications may dramatically increase your symptoms, and Parkinson's disease is currently untreatable. In severe cases, surgery to treat symptoms and regain function in affected areas of the brain may be advised by your doctor.
6. **Problem Definition:** In many cases, a neurological record of the patient and motor skill observations are necessary to find out if someone has Parkinson's Disease. When the diagnosis is first made and the motor consequences are not yet severe, it becomes more challenging for the clinician. In order to monitor the disease's progression over time, a patient needs to return to the clinic frequently. The efficient screening procedure can be more beneficial and does not necessitate a medical visit. Voice recordings are regarded as a useful diagnostic tool because Parkinson's disease sufferers have distinctive voice characteristics. Before seeing a doctor, using AI computations on the discourse dataset to precisely determine the illness might be a helpful screening step.

II. LITERATURE REVIEW

Parkinson's disease (PD) is typically identified through medical intervention and clinical signs, such as the description of motor symptoms. Traditional diagnostic techniques may be subject to subjectivity and misclassification since they depend on the evaluation of motions that are frequently imperceptible to human eye and hence challenging to discern.

As a result, these symptoms are frequently disregarded, making it challenging to make the importance of early Parkinson's disease. These problems must be resolved to enhance the PD diagnosis and assessment processes, machine learning algorithms have been applied to categories PD and healthy controls or patients with similar clinical presentations. The primary objective of the suggested system is to get beyond the shortcomings of the current one and develop a system that is accurate enough to identify Parkinson disease in its earliest stages.[1]

In order to accurately and objectively diagnose the initial stages of Parkinson's disease have been studied utilizing deep dense ANNs (Artificial Neural Networks) applied to voice recordings and large- scale picture categorization of gait data turned into spectrograms.[2] Using Logistic Regression, Random Forests, Boosted Trees, and Support Vector Machine, four machine learning methods that are gaining popularity in biomedicine, we develop a predictive model that can differentiate between early PD and healthy normal.

To assess the robotic learning methods, we performed validation on both a subject- and record-level. We discover that these techniques accurately and effectively (both >95%) discriminate between early PD and healthy normal. A statistically significant fit of the data by the logistic model indicated that predictions could be made using it. By using machine learning to combine the items in a questionnaire, these prediction models. may be able to help physicians with the diagnostic process.[3]

We described a unique approach to Parkinson's disease identification based on the expression of facial features. First, we discussed the characteristics that could distinguish Parkinson's from the healthy control group.

The characteristics that described variations in how people expressed fear at the time were the most significant in

Terms of statistics. In comparison to other classifiers, the XGBoost classifier excelled, achieving 0.69 total accuracy.

III. METHODOLOGY/ MATERIALS

Exploratory Data Analysis is the first attempt, A loaded and pre-processed dataset is ready for analysis. identify the crucial characteristics necessary for accurate models.

For this purpose, Feature Importance analysis is used. This is a group of methods that assign scores to each feature that is used in a prediction from the input data. The dataset can be better understood through feature importance analysis.

The features that are more relevant to the target are highlighted by the relative scores. A dataset-fitted predictive model determines the scores that are most significant. After setting up the information and acquiring significant understanding about significant highlights from the information is executing different demonstrating strategies. Using both the training data and the test data, the modelling techniques are put to the test to see how accurate the models are.

The following is a list of the various modelling methods suggested for use:

- 1. Support Vector Machine:** A new type of learning program called the Support Vector Machine is based on the most recent developments in mathematical learning vision. It is your algorithm, and it works with both linear and indirect data. Converts the original data to a higher magnitude, from which it uses important training tuples known as support vectors to locate the topflight of data segmentation. Separating the hyper plane officially defines the discriminatory category known as the Support Vector Machine. This distance is called a functional margin because it usually reduces the separator generalization error. The useful data that is created is then used with the preloaded SVM algorithm.
- 2. XGBOOST:** XGBOOST is a software algorithm. Recently, this has dominated applied technology education. An implementation of gradient-boosted choice timber is the XGBoost set of rules. That evolved into the design for overall performance and speed.
- 3. Code:** Importing the NumPy, Pandas, Sklearn, and XGBoost libraries

IV. IMPLEMENTATION

The basic flow of data through the various Parkinson's disease detection steps is depicted in the Data flow diagram. In the initial step we will gather information like from the patient with various clinical hardware.

The process for training the data, which will be classified using various ML algorithms, is now sending the collected data. The Train and Test datasets will be sent to determine whether a patient has the disease after they have been classified.

In this article, we use machine learning to forecast Parkinson's disease by explaining several different approaches.

- Data Collection
- Pre-processing
- Model Selection
- Choosing a model through training
- Evaluation
- Forecasting

We will construct a model employing an XGB Classifier in this Python machine learning project by making use of the Python libraries scikit-learn, numpy, pandas, and XG boost. It is an illness that affects the nerve system. This disease affects approximately 1 million people worldwide today. Neurodegenerative dopamine-producing neurons are produced in the brain because of this disorder. The body's symptoms can be detected by the following system.

The project will be made by another AI calculation called the XG Boost. The XG Boost algorithm was created with a focus on optimising speed and performance within the realm of machine learning. XG Boost represents Outrageous Slope Supporting and depends on decision trees. We will import the XGB Classifier from the XG boost library into this project; to develop a model that can precisely determine whether a person has Parkinson's disease.

1. Architecture Diagram

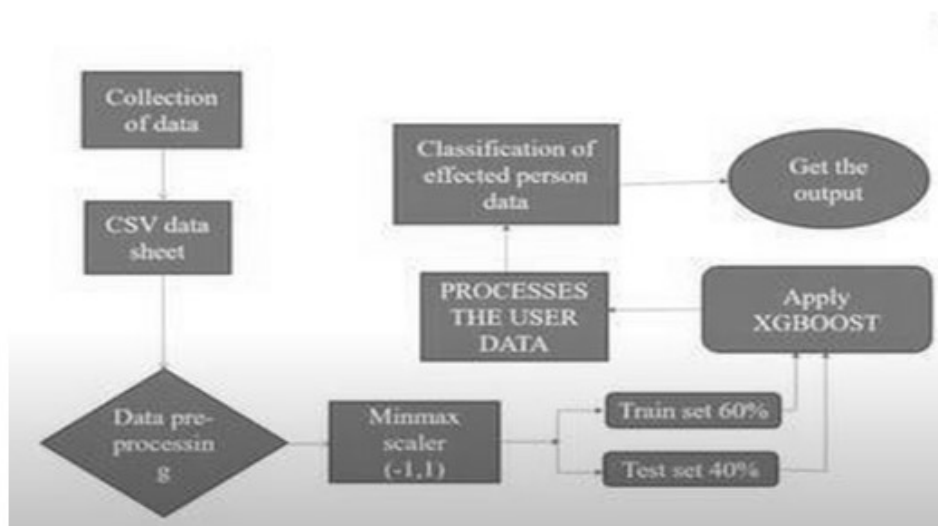


Figure 1: Architecture Diagram

2. Data Flow Diagram

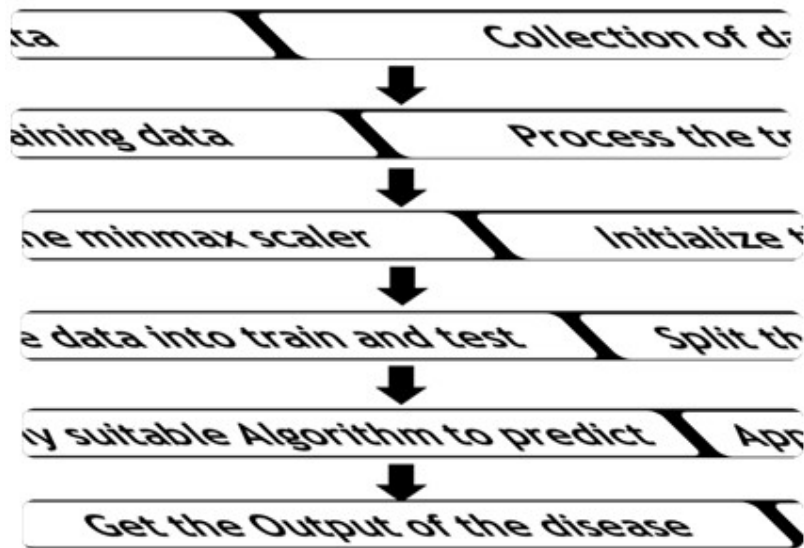


Figure 2: Data Flow Diagram

3. ER Diagram

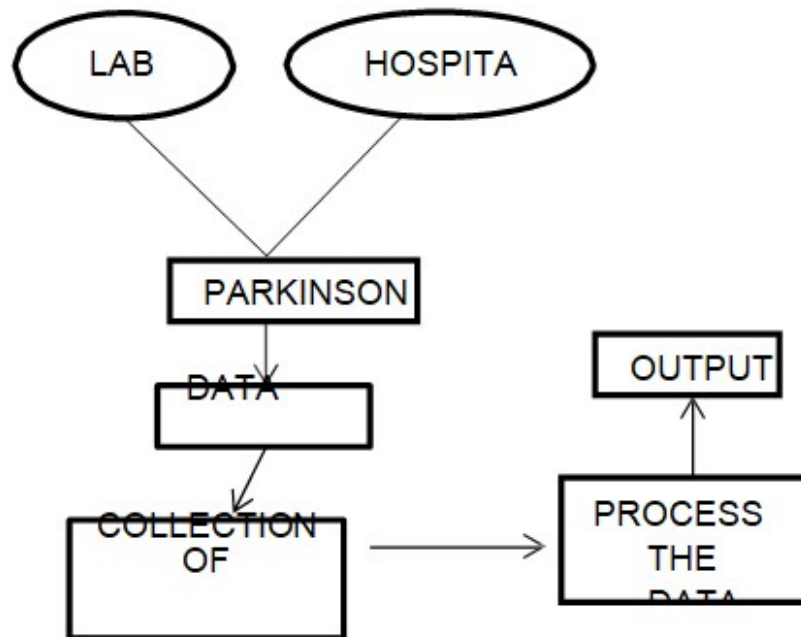


Figure 3: ER Diagram

4. Sequence Diagram

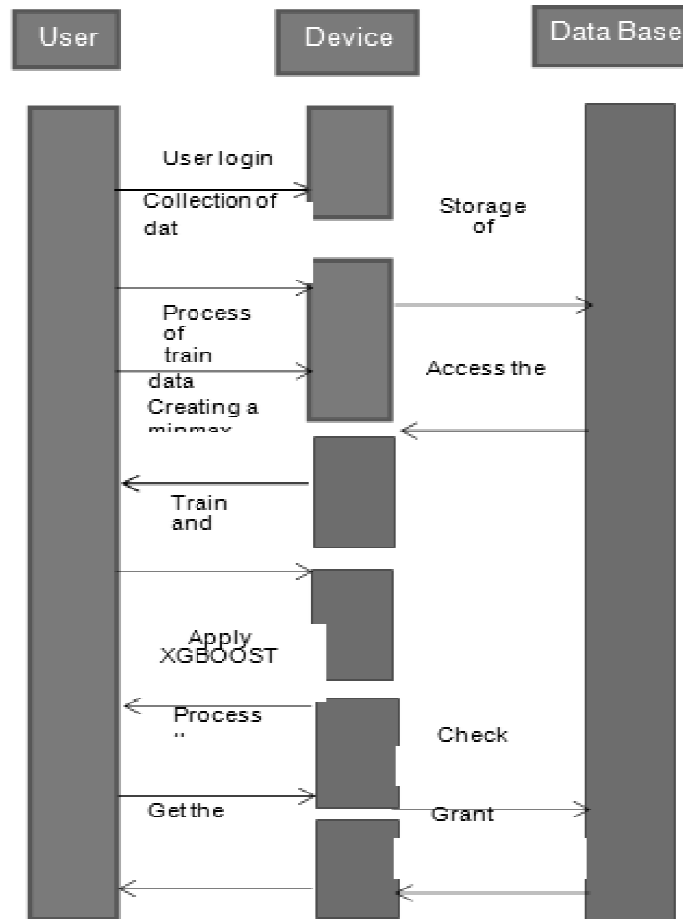


Figure 4: Sequence Diagram

V. TESTING

1. **Testing of Units:** During this phase of software testing, a thorough examination of individual modules and components is conducted. The proposed initiative involves the collection and evaluation of an individual's data. Upon examination of data from a singular entity, the precision is recorded at 100%.
2. **Testing for Integraion:** This could refer to a stage of software testing where discrete components are integrated and evaluated as whole. In the proposed framework, the information is collected and analysed. The precision is 94.87%. This testing will examine the entire endeavour at once. It reduces the duration and difficulty of integration examination.
3. **Functional Evaluations:** Software testing of this type confirms the program's compliance with its functional specifications and requirements. This test uses a machine learning algorithm to identify Parkinson's disease. The ML algorithm will boost productivity.

The following steps are often included in functional testing:

- Determining the expected functionality of the software.
- Generate input data in accordance with the function's requirements.
- The parameters of the function are used to determine the outcome.
- Run the test instance.
- Evaluate actual and predicted outcomes.

VI. THE INPUT - OUTPUT

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	119.92	117.32	119.81	0.0078	0.0077	0.0073	0.0074	0.0119	0.0278	0.0282	0.0212	0.0213	0.0251	0.0245	0.0211	21.23	1	0.41472	0.01285	4.07321	0.06422	
2	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
3	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
4	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
5	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
6	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
7	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
8	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
9	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
10	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
11	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
12	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
13	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
14	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
15	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
16	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
17	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
18	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
19	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
20	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
21	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
22	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
23	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
24	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
25	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
26	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
27	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
28	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
29	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
30	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
31	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	
32	119.92	117.32	119.81	0.0062	0.0058	0.0065	0.0066	0.0134	0.0278	0.0278	0.0212	0.0213	0.0251	0.0245	0.0211	19.85	1	0.42839	0.01221	4.07192	0.03060	

Figure 5: Data set

```
df.apply(lambda x: sum(x.isnull()),axis=0)

Out[12]: name          0
NDVP:Fo(Hz)    0
NDVP:Phi(Hz)   0
NDVP:Flo(Hz)   0
NDVP:Jitter(S) 0
NDVP:Jitter(Abs) 0
NDVP:RAP       0
NDVP:PPQ       0
Jitter:DOP     0
NDVP:Shimmer   0
NDVP:Shimmer(dB) 0
Shimmer:APQ3   0
Shimmer:APQ5   0
NDVP:APQ       0
Shimmer:DDA   0
HNR           0
HNR           0
status        0
RPDE          0
DFA           0
spread1       0
spread2       0
D2            0
PPE           0
dtype: int64

There is no missing values in dataset
```

Figure 6: Inspecting missing values checks

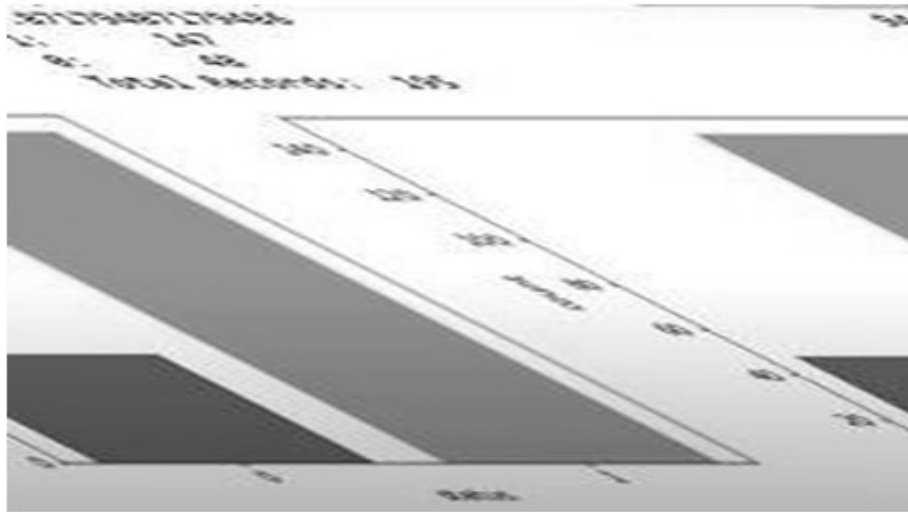


Figure 7: Graph

It represents the characteristics and names where labels act as the data's output and features act as its input. After applying features and labels, we receive 147 1s, which indicates that the person has Parkinson's disease, and 48 0s, which indicates that the person is not impacted by the condition.

```
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=7)

In [16]: # y_train
Out[16]: array([[1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 1,
1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0,
1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1,
1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1,
1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0,
1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1,
1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,
1, 0], dtype=int64)

In [17]: # y_test
Out[17]: array([[0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1,
1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1], dtype=int64)
```

Figure 8: Data set for training and testing

1. **XGBoost:** The XGBoost algorithm is a machine learning technique. Recently, applied gadget learning has been dominating the field. The XGBoost algorithm is a computational implementation of gradient boosting decision trees. The aforementioned alteration was made with the intention of enhancing both the swiftness and efficiency of the design.

```
y_pred= xg.predict(x_test)
print(y_pred)

[1 1 1 1 1 1 0 1 1 1 1 0 1 1 0 1 1 1 0 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1
 1 1]

In [28]: M y_pred1= xg.predict(x_train)
print(y_pred1)

[1 1 1 0 1 0 1 1 0 1 1 1 1 1 1 1 1 1 0 1 0 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 0 1
 1 1 1 0 1 0 0 1 1 1 0 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 0
 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 0 1 1 0 1 0 1 0 1 0 0 1
 0 1 0 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 0 1 0 1 0 1 1 1 1
 1 0 1 1 1 1 1 0]
```

```
In [29]: M y_test

Out[29]: array([[0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1,
 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1], dtype=int64)
```

Figure 9: Using XGB classifier to predict data values

```
from sklearn.metrics import confusion_matrix
xg_cm_test=confusion_matrix(y_test,y_pred)
print(xg_cm_test)
xg_cm_train=confusion_matrix(y_train,y_pred1)
print(xg_cm_train)

[[ 6  1]
 [ 1 31]]
[[ 41  0]
 [  0 115]]
```

Figure 10: Confusion grid

It depicts a confusion grid, which is a tabular representation of the accuracy of a classification model. The model's accuracy is indicated by the number of correct and incorrect predictions it produces, thereby providing a comprehensive evaluation of its performance. Six out of thirty-one predicted values were accurate, while one out of one predicted values were inaccurate.

```
M sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,y_pred)
print(accuracy)

0.8717948717948718
```

Figure 11: XGBoost Accuracy

- 2. **Decision Tree Classifier:** It is an algorithm for machine learning in which the data is always split up based on certain factors. Nodes and leaves, which we have also brought in, are the main parts:

```
from sklearn.tree import DecisionTreeClassifier
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)

# Predicting on test set
y_pred = clf.predict(X_test)

# Predicting on training set
y_train_pred = clf.predict(X_train)

# Print the predicted values
print(y_pred)
print(y_train_pred)
```

Figure 12: Predicting data set of Decision Tree Classifier

It shows how the Decision Tree Classifier predicts the data set.

```
from sklearn.metrics import confusion_matrix
cm_test = confusion_matrix(y_test, y_pred)
cm_train = confusion_matrix(y_train, y_train_pred)

print(cm_test)
print(cm_train)
```

Figure 13: Model Confusion matrix

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test, y_pred)
print(accuracy)
```

Figure 14: Decision tree model accuracy

- 3. Navie Bayes:** It's an AI technique for addressing classification issues and making predictions; it's grounded in Bayes' theorem. Here, we employed a Gaussian distribution to make predictions across three distinct models. Therefore, this is also a sklearn library import.

```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test,y_pred)
cm_test)
cm=confusion_matrix(y_train,y_pred)
cm_train)

[[ 5  2]
 [ 0 24]
 [19  2]
 [44 71]]

In [40]: M from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,y_pred)
print(accuracy)

Out[40]: 0.7435897435897436
```

Figure 15: Navie Bayes data prediction

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,y_pred)
print(accuracy)

Out[40]: 0.7435897435897436
```

Figure 16: Confusion grid

```
from sklearn.metrics import accuracy_score
accuracy = accuracy_score(y_test,y_pred)
print(accuracy)

Out[40]: 0.7435897435897436
```

Figure 17: Navie Bayes Accuracy

VII. HARDWARE AND SOFTWARE REQUIREMENTS

1. Hardware specification

- **Processor:** Any Processor above 500 MHz Ram: 4 GB
- **Hard Disk:** 4 GB
- **Input device:** Standard Keyboard and Mouse. Output device: VGA and High Resolution Monitor.

2. Software Specification

- **Operating System:** Windows 7 or higher Programming language: Python and related libraries. Software: Anaconda Version 3.6

VIII. RESULTS / DISCUSSION

The Random Forest classification model demonstrates a 91.835% accuracy rate and a sensitivity of 0.95 in the classification of Parkinson's disease utilizing vowel phonation data. The Random Forest algorithm yields optimal outcomes as it assigns equal importance to all 22 attributes present in the MDVP dataset. The present report provides an analysis of the accuracy and sensitivity of the SVM model, which yielded results of 91.836% and 0.94, respectively, subsequent to the application of PCA to the dataset.

According to the models, the probability of obtaining false positives in the outcomes is zero. The K-nearest neighbor (KNN) model is effective for datasets with balanced class distribution, as it allows for classification into two categories without making any underlying data assumptions. This treatment modality is non-invasive, straightforward, and precise, and has demonstrated efficacy in affording enduring alleviation to individuals with Parkinson's disease across the globe.

It is suggested that in forthcoming studies, the utilization of both audio and REM sleep data be considered for enhanced outcomes, as the use of solely audio data is deemed inadequate as a biomarker for the classification of Parkinson's disease. It is anticipated that the outcomes of this study will encourage the utilization of mobile recorded audio for the purpose of Parkinson's disease classification through telemedicine.

IX. CONCLUSION

In this study, we attempt to create a Parkinson's disease diagnostic model. In this research, we evaluated different feature measurements and classifiers to identify healthy individuals from Parkinson's sufferers. There is currently no recognized treatment for Parkinson's disease, a brain disorder that affects the central nervous system (CNS). A lack of treatment and death result from late detection. Consequently, early detection is essential. For early disease detection, we utilized machine learning algorithms like SVM (Support Vector Machine), Decision Tree, Random Tree Classifier, and neural networks due to their efficiency and quick retrieval. Most importantly, speech processing has a lot of potential for detecting, categorizing, and diagnosing Parkinson's disease. We anticipate that more medical

tools and technologies based on machine learning will soon be available to save people from this disease.

REFERENCES

- [1] "Parkinson Disease Detection Using Deep Neural Networks", Shivangi, Anubhav Johri and Ashish Tripathi, Department of Computer Science, Jaypee Institute of Information Technology Noida, India.
- [2] Prashanth R, Dutta Roy S, Early Detection of Parkinson's Disease through Patient Questionnaire and Predictive Modelling, International Journal of Medical Informatics (2018), <https://doi.org/10.1016/j.ijmedinf.2018.09.008>
- [3] Justyna Skibińska, Radim Burget, "Parkinson's Disease Detection based on Changes of Emotions during Speech", Brno University of Technology Brno, Czech Republic, 2020 12th International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops (ICUMT)
- [4] Zahari Abu Bakar, Nooritawati Md Tahir and Ihsan MYassin, "Classification of Parkinson's Disease Based on Multilayer Perceptron Neural Network", 2010 6th International Colloquium on Signal Processing & Its Applications (CSPA).
- [5] Parkinson's disease clinical features and diagnosis by Annett Blochberge and Shelley Jones, Vol 3, December 2011. [6] <https://www.kaggle.com/margot234/parkinsons-disease-dataset>
- [6] Kalia, L.V.; Lang, A.E. Parkinson's Disease. *Lancet* 2015, 386, 896–912.
- [7] D. Heisters, "Parkinson's: symptoms treatments and research", vol. 20, no. 9, pp. 548-554, 2011.
- [8] Srishti Grover, Saloni Bhartia, Akshama, Abhilasha Yadav, Seeja K. R. "Predicting Severity of Parkinson's Disease Using Deep Learning". International Conference on Computational Intelligence and Data Science (ICCIDIS 2018). *Procedia Computer Science* 132 (2018) 1788–1794.
- [9] <https://www.kaggle.com/margot234/parkinsons-disease-dataset>
- [10] <https://archive.ics.uci.edu/ml/datasets/parkinsons>
- [11] Shrihari K Kulkarni1, K R Sumana2, "Detection of Parkinson's Disease Using Machine Learning and Deep Learning Algorithms" *International Journal of Engineering Science Invention (IJESI)* ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 VOLUME 8 ISSUE:8, (AUG 2021)-Page No :1189-1192.
- [12] Yatharth Nakul1, Ankit Gupta2, Hritik Sachdeva3,, " Parkinson Disease Detection Using Machine Learning Algorithms" *International Journal of Science and Research (IJSR)* ISSN: 2319-7064 SJIF (2020): 7.803 Volume 10 Issue 6, June 2021-Page no 314-318.
- [13] Wu Wang1, Junho Lee2, Fouzi harrou3 and Ying sun4,, " Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning" *IEEE ACCESS Digital Object Identifier* 10.1109/ACCESS.2020.3016062 Volume 8,2020-Page no 147635- 147646.
- [14] Muthumanickam S1, Gayathri J2, Eunice Daphne J3,, "Parkinson's Disease Detection And Classification Using Machine Learning And Deep Learning Algorithms– A Survey", *International Journal of Engineering Science Invention (IJESI)* ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org ,Volume 7 Issue 5 Ver. 1, May 2018 || PP 56-63.
- [15] Timothy J. Wroge1, Yasin Ozkanca 2, Cenk Demiroglu3, Dong Si4, David C. Atkins5 and Reza Hosseini Ghomi6,, " Parkinson's Disease Diagnosis Using Machine Learning and Voice", Conference: 2018 IEEE Signal Processing in Medicine and Biology Symposium||DOI:10.1109/SPMB.2018.8615607.
- [16] Jayashree R.J., Ganesh S., Karanth S.C., Lalitha S. (2021) Automatic Detection of Parkinson Speech Under Noisy Environment. In: Thampi S.M., Gelenbe E., Atiquzzaman M., Chaudhary V., Li KC. (eds) *Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering*, vol 736. Springer, Singapore. https://doi.org/10.1007/978-981-33-6987-0_16.
- [17] Mahlknecht, P.; Krismer, F.; Poewe, W.; Seppi, K. Meta- Analysis of Dorsolateral Nigral Hyperintensity on Magnetic Resonance Imaging as a Marker for Parkinson's Disease. *Mov. Disord.* 2017, 32, 619–623.
- [18] Dickson, D.W. Neuropathology of Parkinson disease. *Parkinsonism Relat. Disord.* 2018, 46 (Suppl. 1), S30–S33. [20] Kalia, L.V.; Lang, A.E. Parkinson's Disease. *Lancet* 2015, 386, 896–912.
- [19] D. Heisters, "Parkinson's: symptoms treatments and research", vol. 20, no. 9, pp. 548-554, 2011.
- [20] Arvind Kumar Tiwari, "Machine Learning based Approaches for Prediction of Parkinson's Disease," *Machine Learning and Applications- An International Journal (MLAU)* vol. 3, June 2016.
- [21] Gil, D. and Johnson, M. (2009) Diagnosing Parkinson by Using Artificial Neural Networks and Support Vector Machines. *Global Journal of Compute Science and Technology*, 9, 63-71
- [22] M. Abdar and M. Zomorodi-Moghadam, "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" *Journal of AI and Data Mining*, vol. 6, 2018.

- [23] Prediction of Parkinson's Disease using Machine Learning Techniques on Speech dataset, Basil K Varghese, Geraldine Bessie Amali D, Uma Devi K S, Volume - 12, Issue - 2, Year – 2019.
- [24] W. Wang, J. Lee, F. Harrou and Y. Sun, "Early Detection of Parkinson's Disease Using Deep Learning and Machine Learning," in *IEEE Access*, vol. 8, pp. 147635-147646, 2020, doi: 10.1109/ACCESS.2020.3016062.
- [25] R. Alkhatib, M. O. Diab, C. Corbier and M. E. Badaoui, "Machine Learning Algorithm for Gait Analysis and Classification on Early
- [26] Detection of Parkinson," in *IEEE Sensors Letters*, vol. 4, no. 6, pp. 1-4, June 2020, Art no. 6000604, doi: 10.1109/LSENS.2020.2994938.
- [27] P. Raundale, C. Thosar and S. Rane (2021), "Prediction of Parkinson's disease and severity of the disease using Machine Learning and Deep Learning algorithm," 2021 2nd International Conference for Emerging Technology (INCET), pp. 1-5, doi: 10.1109/INCET51464.2021.9456292. [29] Y. Guan (2021), "Application of logistic regression algorithm in the diagnosis of expression disorder in Parkinson's disease," 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), 2021, pp. 1117-1120, doi: 10.1109/ICIBA52610.2021.9688135.
- [28] D. Yadav and I. Jain (2022), "Comparative Analysis of Machine Learning Algorithms for Parkinson's Disease Prediction," 2022 6th International Conference on 10.1109/ICICCS53718.2022.9788354
- [29] Hamid KarimiRouzbahani and Mohammad Reza Daliri, "Diagnosis of Parkinson's Disease in Human Using Voice Signals", Biomedical Engineering Department Iran University of Science and Technology (IUST), Tehran, Iran, 2(3), pp 12- 20,2011.
- [30] Anitha R1, Nandhini T2, Sathish Raj S3 and Nikitha V4, "Early detection of parkinson's disease using machine learning",6(2), pp.
- [31] C K GOMATHY, B. DHEERAJ KUMAR REDDY, Ms. B. VARSHA and B. VARSHINI, "The parkinson's disease detection using machine learning techniques", *International Research Journal of Engineering and Technology (IRJET)*, 8(10), pp 440- 444,2021.