iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60268

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Parkinson’s Disease Detection Using Machine Learning

Chandraprabha Sahu[1], Maya Yadav[2], Parth Singh Dhruwe[3], Arpita Singh[4], Samiksha Shukla[5]
[1, 2, 3, 4]B.Tech Student, [5]Assistant Professor, Department of Information Technology, Government Engineering College, Bilaspur(C.G.).

Abstract: Parkinson’s disease is progressive nervous system disorder that affects movement leading to shaking, stiffness, and difficulty with walking, balance, and coordination. Parkinson’s symptoms begin gradually and get worse over time. It has 5 stages to it and estimated seven to 10 million people worldwide have Parkinson disease. This is chronic and still has no cure. It is a neurodegenerative disease that affects the neurons in the brain that contain dopamine. There is a model, in which using machine learning techniques we can detect Parkinson’s disease depending upon certain medical procedures. We used support vector machine for this and use the sklearn library to prepare the dataset. This gives the accuracy of 88%. In our model we used the dataset which contains biomedical voice measurements from 31 people. From the whole data 20% is used for testing and 80% is used for training. The data of any person is entered to check whether the person has Parkinson or not. The status column is set to 0 for healthy person and 1 for person having Parkinson disease.

I. INTRODUCTION

Parkinson’s disease is considered a neurodegenerative disease because it involves the degeneration and death of neurons. It is most frequently seen in adults over the age of 50. The most recognizable symptoms of Parkinson’s initially are movement-related and generally involve a tremor, slow movement and rigidity. The cause of Parkinson’s are not fully understood, but a combination of genetic and environmental factors are likely involved. In this model, we accurately detect the presence of Parkinson’s disease in an individual based on the huge data collected. By using machine learning algorithm, will predict the person is having Parkinson’s or not.in this model a huge amount of data is collected from the normal person and previously affected person by Parkinson’s disease.

II. RELATED WORKS

“Parkinson’s progression prediction using ml and serum cytokines” by Glenda-M Halliday and Niccholas, 25-july-2019. Serum samples from a clinic are tested to detect Parkinson’s disease, and the same samples are tested using ML algorithms for detecting Parkinson’s disease. Bhauwendraat, C., Bandres-Ciga, S., and Singleton, A.B. using voice to predict progression in Parkinson’s disease patients. Das R. “A comparison of multi classification methods for the diagnosis of Parkinson’s disease”. Methods used to test Parkinson’s disease include ML, DM neural, regression, and decision trees, with ML showing high performance.

III. PROPOSED METHODOLOGY

In this proposed methodology, we aim to develop a robust and accurate Support Vector Machine-based diagnostic tool for early detection and management of Parkinson's disease, thereby improving patient outcomes and facilitating personalized treatment strategies.

A. Data Flow Diagram

Data flow diagram show the flow of data between various elements of a system in graphical form. It also expresses the requirement of the system and shows the current system is implemented.

In the first step we will collect comprehensive dataset comprising clinical data, genetic information, and other relevant biomarkers from Parkinson’s disease patients.

After that, preprocess the data to ensure quality and consistency, including data cleaning, missing values imputation, normalization, and standardization.

Then classify the data into Train and Test datasets, apply Support Vector Machine Classifier to train the model. The dataset will be used to predict whether the patient has Parkinson’s disease or not.

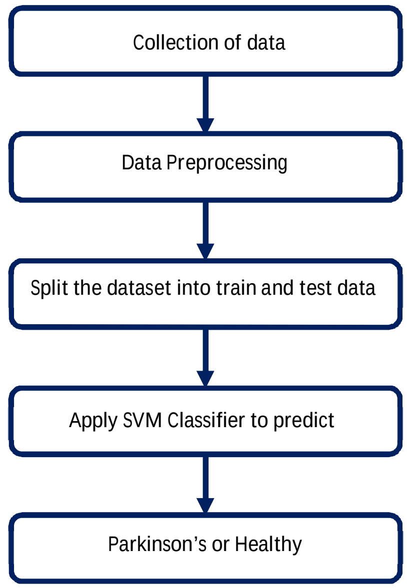


Fig. 1 Data Flow Diagram

B. *System Design*

The proposed architecture has four parts where it involves Data Acquisition, feature Extraction, Classification, and Output Production. By following this system design, the proposed Parkinson's disease detection system using Support Vector Machine algorithm can effectively leverage machine learning techniques to assist clinicians in early diagnosis and management of the disease, thereby improving patient outcomes and healthcare delivery.

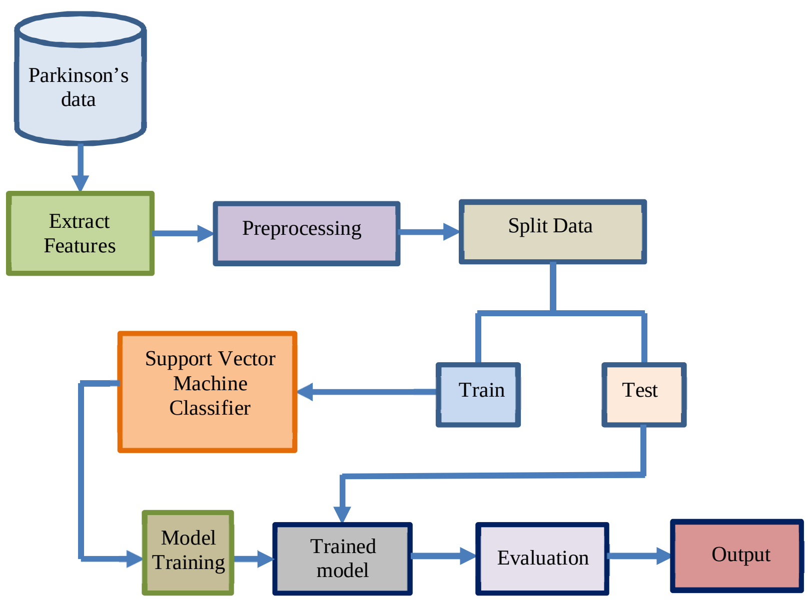


Fig. 2 System Design

C. *Modules*

1) *Data Extraction*

Import all the necessary dependencies, which is required for data cleaning, data analysis and model building. Read the Data into a DataFrame and get first five records. Get the features and labels from the DataFrame. Features are all the columns expect 'status', and the labels are those in the 'status' column. Fig. 3 shows importing the necessary dependencies.

```
In [1]: # importing the dependencies
        import numpy as np
        import pandas as pd

In [10]: from sklearn.model_selection import train_test_split
         from sklearn.preprocessing import StandardScaler
         from sklearn import svm
         from sklearn.metrics import accuracy_score
```

Data Collection & Analysis

```
In [2]: #loading the data from csv file to a pandas dataframe
        parkinsons_data=pd.read_csv("parkinsons.csv")

In [3]: #printing first five rows of dataframe
        parkinsons_data.head()
```

Out[3]:

	name	MDVP:Fo(Hz)	MDVP:Fhi(Hz)	MDVP:Flo(Hz)	MDVP:Jitter(%)	MDVP:Jitter(Abs)	MDVP:RAP	MDVP:PPQ	Jitter:DDP	MDVP:Shimmer	...	Shim
0	phon_R01_S01_1	119.992	157.302	74.997	0.00784	0.00007	0.00370	0.00554	0.01109	0.04374	...	
1	phon_R01_S01_2	122.400	148.650	113.819	0.00968	0.00008	0.00465	0.00696	0.01394	0.06134	...	
2	phon_R01_S01_3	116.682	131.111	111.555	0.01050	0.00009	0.00544	0.00781	0.01633	0.05233	...	
3	phon_R01_S01_4	116.676	137.871	111.366	0.00997	0.00009	0.00502	0.00698	0.01505	0.05492	...	
4	phon_R01_S01_5	116.014	141.781	110.655	0.01284	0.00011	0.00655	0.00908	0.01966	0.06425	...	

Fig. 3 Importing the dependencies

2) Data cleaning

The imported dataset is check for the missing values for each of the columns to build the error free model. The next step is to get the statistical measures about the data which gives us several statistical measures like mean, percentage value, and minimum value. Fig. 4 explains data analysis and Fig. 5 explains statistical measures about the data.

```
memory usage: 36.7+ KB

In [6]: # checking for missing value in each column
        parkinsons_data.isnull().sum()

Out[6]: name                0
        MDVP:Fo(Hz)         0
        MDVP:Fhi(Hz)        0
        MDVP:Flo(Hz)        0
        MDVP:Jitter(%)      0
        MDVP:Jitter(Abs)    0
        MDVP:RAP            0
        MDVP:PPQ            0
        Jitter:DDP          0
        MDVP:Shimmer        0
        MDVP:Shimmer(dB)    0
        Shimmer:APQ3        0
        Shimmer:APQ5        0
        MDVP:APQ            0
        Shimmer:DDA         0
        NHR                 0
        HNR                 0
        status              0
        RPDE                0
        DFA                 0
        spread1             0
        spread2             0
        D2                  0
        PPE                 0
        dtype: int64

In [7]: # getting some stastical measures about the data
        parkinsons_data.describe()
```

Out[7]:

	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)
count	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000

Fig. 4 Data Analysis

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

```
PPE                0
dtype: int64

In [7]: # getting some stastical measures about the data
        parkinsons_data.describe()

Out[7]:
```

	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)	...
count	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	195.000000	...
mean	154.228641	197.104918	116.324631	0.006220	0.000044	0.003306	0.003446	0.009920	0.029709	0.282251	...
std	41.390065	91.491548	43.521413	0.004848	0.000035	0.002968	0.002759	0.008903	0.018857	0.194877	...
min	88.333000	102.145000	65.476000	0.001680	0.000007	0.000680	0.000920	0.002040	0.009540	0.085000	...
25%	117.572000	134.862500	84.291000	0.003460	0.000020	0.001660	0.001860	0.004985	0.016505	0.148500	...
50%	148.790000	175.829000	104.315000	0.004940	0.000030	0.002500	0.002690	0.007490	0.022970	0.221000	...
75%	182.769000	224.205500	140.018500	0.007365	0.000060	0.003835	0.003955	0.011505	0.037885	0.350000	...
max	260.105000	592.030000	239.170000	0.033160	0.000260	0.021440	0.019580	0.064330	0.119080	1.302000	...

8 rows × 23 columns

```
In [8]: # distribution of target varible
        parkinsons_data['status'].value_counts()

Out[8]: 1    147
        0     48
        Name: status, dtype: int64
```

Fig. 5 Statistical measures about the data

1) *Splitting the Dataset into Training and Testing Data*

The data is split into training data and testing data were training data will be used to train the machine learning model and testing data is used to evaluate the model to check the performance of the model. We used train_test_split function so it will automatically split this data in a random way. Inside this we need to mention x and y, were x represents the training data features and y represents corresponding labels. The dataset is split into 80% of training data and 20% testing data. Fig. 6 shows the data splitting.

Splitting the data to training data & test data

```
In [14]: X_train, X_test, Y_train, Y_test=train_test_split(X, Y, test_size=0.2, random_state=2)

In [15]: print(X.shape, X_train.shape, X_test.shape)

         (195, 22) (156, 22) (39, 22)
```

Data Standardization

```
In [16]: s=StandardScaler()

In [17]: s.fit(X_train)

Out[17]: StandardScaler()
         In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
         On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [18]: X_train=s.transform(X_train)

         X_test=s.transform(X_test)

In [19]: print(X_train)

         [[ 0.63239631 -0.02731081 -0.87985049 ... -0.97586547 -0.55160318
            0.07769494]
          [-1.05512719 -0.83337041 -0.9284778  ...  0.3981808  -0.61014073
            0.39291782]
          [ 0.02996187 -0.29531068 -1.12211107 ... -0.43937044 -0.62849605
           -0.50948408]
```

Fig. 6 Data Splitting

2) *Model building*

SVM classifier is used to train the model and classify the data into classes here the two classes are person with Parkinson and without Parkinson. We trained support vector machine model with training data. The data is fit into the support vector machine model. The next step is to evaluating the model, in this we find the accuracy score of the model is 88%. Fig. 7 shows the model building.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Building a Predictive System

```
In [30]: input_data=(95.05600,120.10300,91.22600,0.00532,0.00006,0.00268,0.00332,0.00803,0.02838,0.25500,0.01441,0.01725,0.02444,0.04324,

# changing input data into numpy array
input_data_as_numpy_array=np.asarray(input_data)

#reshape the numpy array
input_data_reshaped=input_data_as_numpy_array.reshape(1,-1)

# standardized the data
std_data=s.transform(input_data_reshaped)

prediction=model.predict(std_data)
print(prediction)

if(prediction[0]==0):
    print("The Person does not have parkinson disease")

else:
    print("The Person has Parkinsons")

[1]
The Person has Parkinsons

C:\Users\owner\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:464: UserWarning: X does not have valid
feature names, but StandardScaler was fitted with feature names
  warnings.warn(
```

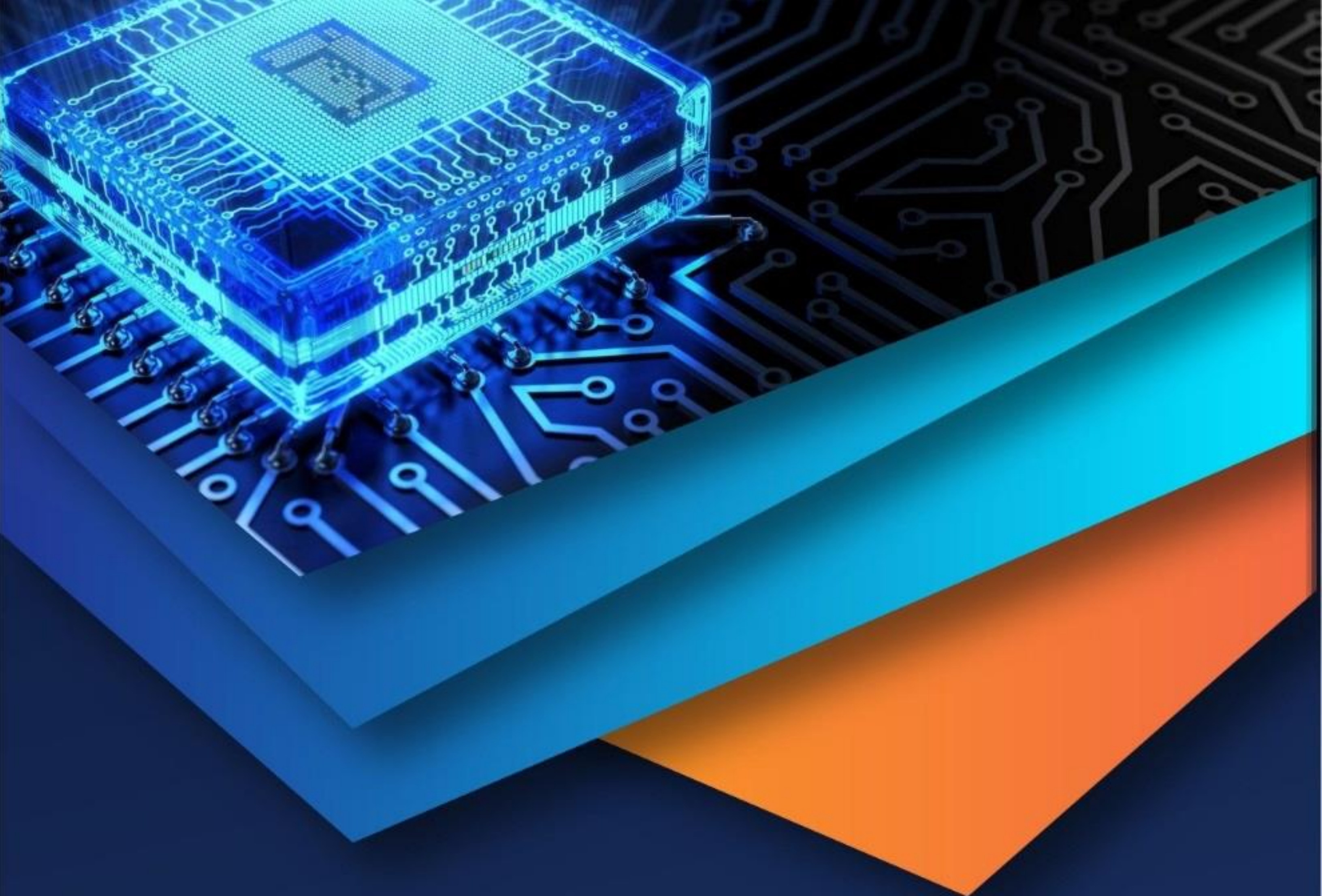
Fig. 7 Model Building

IV. CONCLUSION

Parkinson's disease is a progressive disorder that affects the nervous system and the parts of the body controlled by the nerves. Parkinson's disease can't be cured. Therefore it is important to diagnose it early. We used SVM classifier for early detection of Parkinson's disease. To diagnose Parkinson disease in patients we use machine learning model. This model provides diagnostics on time, in this way it reduce treatment costs. The algorithm that is used in this ml model is better than other algorithms. This model is also beneficial for medical students like they can use this as a teaching tool. This ML model can predict Parkinson's disease with high accuracy by the help of different data like spiral images, medical records etc.

REFRENCES

[1] 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.
[2] Z. A. Dastgheib, B. Lithgow, and Z. Moussavi, "Diagnosis of Parkinson's disease using electrovestibulography," Medical & Biological Engineering & Computing, vol. 50, no. 3, pp.365-373,2017.
[3] Yatharth Nakul1 , Ankit Gupta2, Hritik Sachdeva3,,, "Parkinson Disease Detection Using Machine Learning Algorithims" International Journal of Science and Research (ISJR) ISSN: 2319-7064 SJIF (2020): 7.803 Volume 10 Issue 6, June 2021.
[4] M. Abdar and M. Zomorodi-Moghadam, "impact of Patient's Gender on Parkinson's disease using Classification Algorithim" Journal of AI and Data Mining, vol. 6, 2018.
[5] Dr. Anupam Bhatia and RaunakSulekh, "Predictive model for Parkinson's Disease through Naïve Bayes Classification" International Journal of Computer Science & Communication vol. 9, March 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY