Paralysis Agitans Disease Detection Using Machine Learning

Leena Supe[1], Manasvi Tayde[2], Sumit Patil[3], Divya Chaudhari[4], Dhanashree S. Tayade [5]
[1234]*Research Scholar, Department of Information Technology, SSBT COET, Jalgaon, Maharashtra, India*
[5]*Assistant Professor, Department of Information Technology, SSBT COET, Jalgaon, Maharashtra, India*

***Abstract* --- In this new technological cycle, current challenges demand a profound re-orientation of a global healthcare system. A more efficient system is required to cope with increased life expectancy which is associated with a prevalence of chronic neurological disorders such as Paralysis Agitans. The Paralysis Agitans disease is a neurodegenerative disorder affecting 60 percent of people. At this moment there is no such system available to detect the Paralysis Agitans disease with good accuracy. The detection of Paralysis Agitans disease is based on medical history, symptoms, and the neurological and physical Exam. The proposed system is used to detect Paralysis Agitans disease with the help of various machine learning techniques. The Paralysis Agitans detection system will achieve good results.**

Keywords - Algorithms, Handwriting analysis, Machine Learning, paralysis Agitans(PA) disease or Parkinson's Disease(PD).

I. INTRODUCTION

Paralysis agitans is also known as ' Parkinson's Disease'. It mostly affects the motor functions of humans. The main motor symptoms are called "parkinsonism" or "parkinsonian syndrome". It is a neurodegenerative disorder of the central nervous system that affects a lot only people significantly affecting their quality of life. The exact reason for Paralysis Agitans disease is unknown, but it is a concept that involves complex interactions between genetics, biology, and the environment, which not only leads to the heterogeneity of Paralysis Agitans signs and symptoms but also their charge of progression through the years.

Parkinson's disease (PD) is a complex neurodegenerative disorder commonly characterized by motor impairments such as tremors, bradykinesia, dyskinesia, and gait abnormalities.

Proper assessment of Paralysis Agitans motor impairments is vital for clinical management of the disease. Appropriate timing of dopaminergic medications to avoid sudden increases in symptom severity and selection for interventions such as deep brain stimulation both require precise understandings of symptom fluctuations in patients with PA. In addition, objective characterization of non-motor manifestations of PA such as sleep disorders, gastrointestinal symptoms, and psychiatric symptoms is needed to understand long-term disease progression. Addressing these flaws is vital to ensure proper diagnosis and management of patients. To that end, considerable efforts have been made to develop objective, at-home, and automated methods to monitor the main motor symptoms characteristic of PA. Leveraging motion sensors and, in some instances, video-based technologies can first enable physicians to take data-driven approaches to PA diagnoses.

Symptoms may include:-
1]Tremor = Rhythmic shaking, called tremor, usually begins in a limb, often your hand or fingers.
2]Slowed movement, known as bradykinesia = Over time, the disease may slow your movement, making simple tasks difficult and time-consuming.
3]Rigid muscles = Muscle stiffness may occur in any part of your body. The stiff muscles can be painful and limit your range of motion.
4]Impaired posture and balance = The posture may become stooped. Or you may fall or have balance problems as a result of Parkinson's disease.
5]Loss of automatic movements = A decreased ability to perform unconscious movements, including blinking, smiling swinging your arms when you walk.
6]Speech changes = You may speak softly or quickly, slur, or hesitate before talking. Your speech may be more of a monotone rather than have the usual speech patterns.

II. LITERATURE SURVEY

In a paper [1], F. Harrou (2020) introduced a model using the deep learning method for the early detection of Paralysis Agitans. To evaluate the accuracy, the system used two algorithms SVM and KNN for detection which randomly split the data, with the training data and the rest as testing data. The ratio of the patient with healthy people in training and testing data is kept the same as the original data. Hence, the performance is measured on the testing data using accuracy, specificity, and sensitivity.

In the paper [2], Satish J. Sharma (2021) introduced a system that is useful for the detection of Paralysis Agitans disease. The detection is done in two main stages, training and testing. The dataset introduced is the collection of samples between healthy people and people suffering from the disease. The SVM method is applied for classification between healthy people and people with Paralysis Agitans disease that provides accuracy in predicting whether the disease is present in the patient or not.

In the paper [3] P. Kumar and Vipul Narayan (2022), have introduced a system that is useful for the detection of Parkinson's disease. Parkinson's disease is a central nervous system condition that affects the body's motor processes. It's a long-term illness with symptoms that worsen over time. It usually affects the elderly, whose symptoms steadily worsen until they reach a peak. Hearing, walking, speech, and other basic bodily functions can all be affected by the condition. Generic machine learning methods that provide varied degrees of accuracy can be used to analyze this disease. As a result, the best one is picked, as it will provide the maximum level of accuracy in predicting whether or not the disease is present in the patient.

In paper [4], S. Palwe and A. Govindu (2023) introduced a model to predict Paralysis Agitans. The proposed model predicts no false positives in the result. The model also performs well for balanced datasets as classification into two categories without the presumption of data is favored. The trained model used SVM and a Logistic Regression model to reach the accuracy. The goal was to use feature selection techniques to sort the medical measures and select the most relevant characteristics.

III. METHODOLOGY

The proposed methodology collects audio data about Parkinson's patient's voice modulations. The dataset contains information about jitter, shimmer, and MDVP of vowel phonations. Data is preprocessed, analyzed, and visualized for a thorough understanding of the attributes. The initial approach is to the trained model – SVM on the data. Models are trained to classify given audio data into PD or healthy, based on variations in frequency.

The methodology followed can be divided into five steps. 1) Input data

2) Data Preprocessing 3

) Model Building and Training

4) Prediction and Optimization

5) Show Result

System Architecture

System architecture refers to the overall design and organization of a complex system, which includes hardware, software, and various components that interact with each other to achieve specific goals. It involves defining the system's structure, components, and interfaces, as well as the relationships between them. A good system architecture ensures the system is reliable, scalable, and maintainable while meeting the functional and non-functional requirements. Machine learning has given computer systems the ability to automatically learn without being explicitly programmed. The architecture diagram describes the high-level overview of major system components and important working relationships.

It represents the flow of execution and it involves the following five major steps:-

1) The architecture diagram is defined with the flow of the process which is used to refine the raw data and used for predicting the disease.
2) The next step is preprocessing the collected raw data into an understandable format.
3) The system trains the data by splitting the dataset into train data and test data.
4) The data is evaluated with the application of a machine learning algorithm and the classification accuracy of this model is found.
5) After training the data with these algorithms we have to test on the same algorithms.

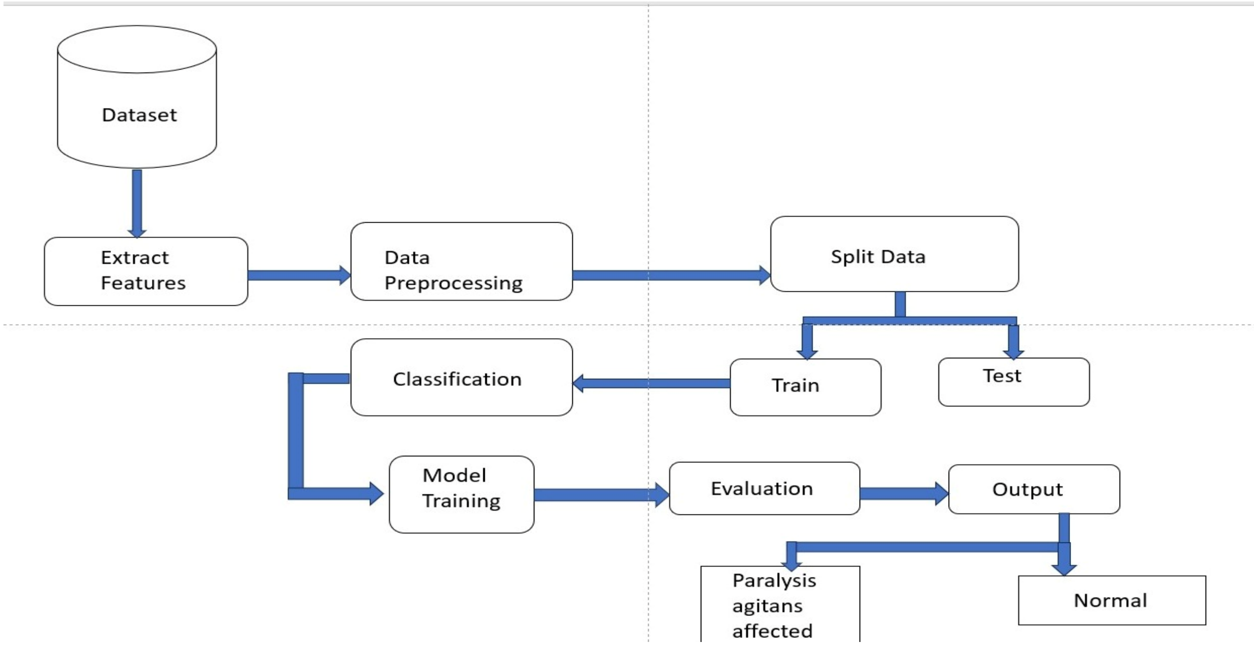

Fig 1. System architecture

Algorithm - Support Vector Machine

A Support Vector Machine (SVM) is defined as a machine learning algorithm that uses supervised learning models to solve complex classification, regression, and outlier detection problems by performing optimal data transformations that determine boundaries between data points based on predefined classes, labels, or outputs. Technically, the primary objective of the SVM algorithm

is to identify a hyperplane that distinguishably segregates the data points of different classes. The hyperplane is localized in such a manner that the largest margin separates the classes under consideration.

In the mathematical context, an SVM refers to a set of ML algorithms that use kernel methods to transform data features by employing kernel kernel functions. Kernel functions rely on the process of mapping complex datasets to higher dimensions in a manner that makes data point separation easier. The function simplifies the data boundaries for non-linear problems by adding higher dimensions to map complex data points. A hyperplane is defined as a line that tends to widen the margins between the two closest tags or labels. The distance of the hyperplane to the most immediate label is the largest, making the data classification easier. The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

The SVM algorithm is implemented with kernel that transforms an input data space into the required form.

1] Linear Kernel :-

• It can be used as a dot product between any two observations. The formula of linear kernel is as below : K(x , xi)=sum(x∗ xi)

• From the above formula, we can see that the product between two vectors say x & xi is the sum of the multiplication of each pair of input values.

2] Polynomial Kernel :-

• It is more generalized form of linear kernel and distinguish curved or nonlinear input space. Following is the formula for polynomial kernel-

K(x , xi)= 1+sum(x , xi)^d

• Here d is the degree of polynomial, which we need to specify manually in the learning algorithm.

3] Radial Basis Function (RBF):-

Kernel RBF kernel, mostly used in SVM classification, maps input space in indefinite dimensional space. Following formula explains it mathematically –

K(x , xi)= exp(-ɣ|| x - xi ||2)

Here, gamma ranges from 0 to 1. We need to manually specify it in the learning algorithm. A good default value of gamma is 0.1.

There are some basic steps of support vector machine:-

1) The first step is Data Collection and Preprocessing.

2) After data collection, Feature Selection is the next step in which useful features are extracted from the data.
3) Split the dataset into training and testing sets.
4) Choose the SVM model and kernel function.
5) Evaluate the model's performance using the testing data and appropriate metrics.
6) Use the trained SVM model to make predictions on new data.

VI. IMPLEMENTATION

This section introduced the techniques, used dataset and the confusion matrix and working of the algorithm.

a) Techniques

1) NumPy: NumPy serves as a Python module for numerical computing that supports arrays, matrix structures, and mathematical functions. Key features of the NumPy library are NumPy arrays, mathematical functions, algebraic operations, and broadcasting.
2)Pandas: Pandas serve as a freely available data analysis and manipulation package written in Python. It makes working with organized data like tabular, time series, & observational information simple and efficient.
3)Sklearn: A well-liked Python package for machine learning is sci-kit-learn, sometimes referred to Sklearn. It works nicely with other Python ecosystem libraries and is developed on top of the SciPy, NumPy, and Matplotlib packages.
4)Streamlit: The Open-source Python module Streamlit can be utilized to easily build web-based applications for machine learning projects and data science.
5)Pickle: An option for serializing and deserializing objects in Python into a continuous sequence of bytes is provided by the pickle library module in Python.

a) Used dataset

In the present work, the dataset was taken from kaggle.com. This dataset is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson's disease (PD).
This paper uses a set of voice measurement of data includes 195 sustained vowel phonations. It consists of 147 people with Parkinson's and 48 is data of healthy people. Patients are in the age that ranges between 46 to 85 years, while normal readings are from people of 23 years of age. An average of 6 phonation's were recorded 195 times for every person, ranging from 1 to 36 seconds in duration. The attributes of 195 records are elaborated in table 1 below:

Attribute	Purpose
Name	Data is stored in ASCII CSV format where patient name and recording number is stored
MDVP: Fo (Hz)	Fundamental frequency of pitch period
MDVP: Fhi (Hz)	Upper limit of fundamental frequency or maximum threshold of voice modulation
MDVP: Flo (Hz)	Lower limit or minimal vocal fundamental frequency
MDVP: Jitter, Abs, RAP, PPQ, DDP	These are various Kay Pentax's multi-dimensional voice program (MDVP) measures. MDVP is a traditional measure of frequency of vibrations in vocal folds at pitch period to vibrations at start of next cycle called pitch mark [25]
Jitter and Shimmer	Measures of absolute difference between frequencies of each cycle, after normalizing the average
NHR and HNR	Signal to noise and tonal ratio measures, that indicate robustness of environment to noise
Status	0 indicates healthy person while 1 indicates PWP.
D2	Correlation dimension is used to identify dysphonia in speech using fractal objects. It is a nonlinear, dynamic attribute.
RPDE	Recurrence Period Density Entropy quantifies the extent to which signal is periodic
DFA	Detrended Fluctuation Analysis or DFA measures the extent of stochastic self-similarity of noise in speech signals.
PPE	Pitch Period entropy is used to assess abnormal variations in speech on a logarithmic scale
Spread1, spread2	Analysis of extent or range of variations in speech with respect to MDVP: Fo(Hz)

Table 1: Data set attributes

b) Confusion Matrix

A confusion matrix is a tool used in machine learning and data analysis to evaluate the performance of a classification model. By analyzing the values in the confusion matrix, researchers can calculate performance metrics such as accuracy, precision, recall, and F1 score, which provide insights into the model's ability to accurately classify Parkinson's affected patients.

c) Working

1)Data pre-processing includes separating the features and target.
2)To separate training and testing data four arrays are created using train test split function.
3)Data is next standardised, that is all data is transformed to same range without changing actual meaning. The transformed data is displayed which lies in the range of -1 to +1.
4)To train the model we have chosen support vector machine. SVM plots several data points in the graph. It distributes between two set of data type. SVM tries to separate the data with a line called as hyperplane. The data points close to SVM hyperplane line are called support vectors. The location of support vector changes

when new data point is given. It decides on which side of hyperplane to be placed after training the model. The model hence, classifies into two Parkinson positive and Parkinson negative. The model is trained with training data.

5)For evaluating the model accuracy score of training data is needed. If the accuracy score of training data and testing data has huge variation it may lead to overfitting or underfitting.

6)Finally, to build a predictive system, we should change the input data into NumPy array. This will be presented in form of tuple which should be reshaped and standardised.

7)Conclusively print statement must be added according to conditions to detect whether the person is affected or healthy.

VII. RESULT

In the proposed system, early detection and intervention of Parkinson's' illness is facilitated with an accuracy rate of 84.3 percent on a dataset of 195 voice samples with 147 samples from healthy people and 48 samples from people having Paralysis Agitans disease or Parkinson's disease. Support Vector Machine (SVM) classifier had been trained on 80 percent of the data set and evaluated on 20 percent. Overall, these findings indicate that SVM can detect disease illness using voice data parameters.

Result Analysis

Accuracy refers how close a measurement is to its true value. In the above graph the F1 score, accuracy and MCC(Mathews Correlation Coefficient) is measured. The process ensures that the prediction can be done accurately by the proposed system .

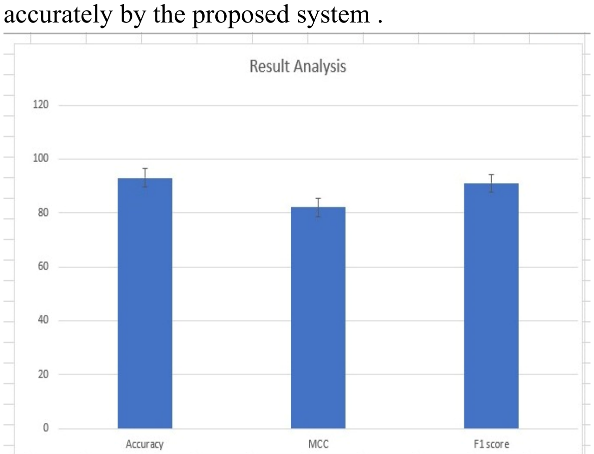

Fig 2. Accuracy of model

VIII. CONCLUSION

The proposed system aims to facilitate remote monitoring of motor and non motor symptoms of Paralysis Agitans disease patients by predicting the features . As a result, additional study is required to evaluate SVM's usefulness in detecting Parkinson's illness utilizing speech data on larger sets of data as well as real-world scenarios. SVM may improve the accuracy and efficacy of the diagnosis of Parkinson's illness.

REFERENCE

[1] Nihar M. Ranjan and Gitanjali Mate, "Detection of Parkinson's Disease using Machine Learning algorithms and Handwriting Analysis", Volume-8, Issue-1, e-ISSN: 2456-9437

[2] Wu Wang, Junjo Lee, Fouzi Harrou ," Early Detection of Parkinson Disease using Machine Learning." In 2020 IEEE , this work was supported by KAUST University, OSR(office of sponsored research, under award OSR-2019-CRG7-3800.

[3] Satish J. Sharma, Shraddha V. Bante, Shweta M. Barhate, "Parkinson's Disease Detection using Machine Learning.", 2022 JETIR ,Vol. 9 issue-6, (ISSN2349-5162).

[4] Rudramurthy V C , Bhagyashri R Hanji, Manoj C , Suhas M S, "Parkinson's Disease Detection Using Machine Learning",Volume:09 Issue:06, e-ISSN : 2395-0056, p-ISSN : 2395-0072.

[5] Ms Shanavi Sharma, Ms Pranjali Singh, Dr Seema Gard, Dr Pratima Singh," Machine Learning Based Health Care In Neurology", Vol.7, ISSN:0974-5823.

[6] Rajesh Kumar Yadav, Arun kumar Rai, Vipul Narayan, " Early Warning Signs Of Parkinson's Disease Prediction Using Machine Learning". Volume 13.

[7] Yunfeng Wu, Jian Chen, Meihong Wu, "Dsyphonic Voice Pattern Analysis of Patients in Parkinson's Disease Using Minimum Interclass Probability Risk Feature Selection and Bagging Ensemble Learning Methods",PMCID:PMC5434464,PMID:28553366.

[8] Soumanas Kar and Shreya Chakraborty, "Detection of Parkinson's Disease using Image Classification and Machine Learning",Volume:03/Issue:05,Impact Factor-5.354,e-ISSN: 2582-5208.

[9] Dr. C. K. Gomathy and Mr. B. Dheeraj Reddy, "The Parkinson's Disease Detection using Machine Learning Techniques",Volume:08,Issue:10, e-ISSN:2395-0056,p-ISSN:2395-0072.
[10] Uwishema, O. et al.: "The understanding of Parkinson's disease through genetics and new therapies. Brain Behav. 12, 5, e2577 (2022).
[11] Dr Pratima Singh, Dr Seema Garg, Ms Shanavi Sharma, "Machine Learning Based Healthcare In neurology", Vol. 7, ISSN : 0974-5823(2022).
[12] Rehman, R.Z. et al.: "Comparison of Walking Protocols and Gait Assessment Systems for Machine Learning-Based Classification of Parkinson's Disease", (2020).
[13] Anitha R1, Nandhini T2, Sathish Raj S3 and Nikitha V4, "Early detection of parkinson's disease using machine learning",6(2),pp. 505-511, 2020.