International Journal of Research Publication and Reviews

Improved Model of Parkinson's Disease Prediction using Machine Learning Approach

Lamkhade Anil[1], Malgunde Mahesh[2], Nachan Rushikesh[3], Pandule Tushar[4], Prof. Unde Suvarna[5]

[1,2,3,4,5] Rajiv Gandhi College of Engineering, Ahmednagar, Maharashtra

ABSTRACT

Parkinson's disease (PD) is a neurodegenerative disease in which symptoms appear gradually, starting with a slight tremor in one hand and a feeling of stiffness in the body, leading to physical deterioration. It affects more than 6 million people worldwide. Currently, lay doctors do not have good results in this disease, especially in the early stages of the disease, when early detection of symptoms is very difficult. The project uses machine learning algorithms to analyze changes in speech patterns to predict whether a patient has Parkinson's disease. Find the best features using the Pearson correlation method and combined model (XGBoost) to diagnose Parkinson's disease with the highest accuracy using data containing words recorded from patients with and without Parkinson's disease. These differences in data can be added to the sample and results to find people affected by PD and display the results.

Keywords*: -* Back propagation, Severity Prediction, Parkinson's disease, voice patterns, XGBoost.

I. INTRODUCTION

Parkinson's disease is a disease that affects panic attacks. These symptoms begin gradually due to low dopamine levels in the brain. Dopamine is a chemical that is a neurotransmitter that carries signals from the body to the brain. The decrease in neurons responsible for dopamine production leads to a decrease in dopamine, resulting in decreased coordination between the mind and body. The disease is caused by a deficiency of dopamine producing neurons in the

Substantia nigra, a part of the brain today, and is characterized by motor symptoms such as tremors, rigidity, and bradykinesia. Parkinson's disease also affects non motor abilities such as sleep, cognition and thinking. Currently, most predictions are based on scientific research and historical research.

II. Literature Review

Parkinson's disease (PD) is a debilitating neurodegenerative disease that cannot be diagnosed with blood tests, so faster, cheaper diagnostic tools are needed. Using machine learning algorithms to analyze changes in speech patterns is a new way to predict whether a patient has Parkinson's disease. This article presents a predictive model that successfully diagnoses Parkinson's disease with the highest possible accuracy using data that incorporates evidence from Parkinson's patients' records and is irrelevant to the context. Experimental results show that Boosted Decision Tree (a hybrid model of gradient boosting regression trees) is the best model using the data with an accuracy of 91-95%. The filter based search also found that the heaviest weight was spread, spread 2, an PPE, which were three independent measurements of the significant difference in the record. These findings can be applied to PD biometrics [1].

Tarigopra et al. (Sriram, Rao, Narayana, Kaladhar and Vital, 2013) 6 © Daffodil International University Presented a comparative study on Naive Bayes, Random Forest, Logistic Regression, Support Vector Machine to diagnose Parkinson's disease. In detecting Parkinson's disease, SVM (mean 88.9%) performed better than NB (mean 69.23%), while RF (mean 90.26%) performed better than SVM. Also LR (i.e. 83.66%) performs well. 86%)[2]

M. Abdar and M. Zomorodi-Moghadam proposed a paper "Impact of Patients' Gender on Parkinson's disease using Classification Algorithms" [10]. In this paper, the author chooses the UCI PD dataset for finding the accuracy of Parkinson's using SVM and Bayesian Network algorithms[3]

III. METHODOLOGY

1) Data Collection:- Parkinson disease data from provided by the UCI Machine Learning Repository. In addition, this dataset is consisting of 62 people with Parkinson disease and 15 peoples were healthy.

2) Data Pre-processing:- In this section, firstly we extracted features from the Parkinson disease datasets. We picked the 30 columns and 77 entries of data. Then, we conducted several experiments to checking missing values, redundant values.

3) Training data: Splitting the dataset into Training set and testing set: In machine learning data preprocessing, we have to break our dataset into both training set and test set.

4) Apply Machine Learning Algorithms: Now, we've both the train and test data. The subsequent step is to spot the possible training methods and train our models. As this is often a classification problem, we've used three different classification methods KNN, Naïve Bayes, and Logistic Regression.

5) Testing Data: - Once Parkinson's disease Prediction model has been trained on the pre-processed dataset, then the model is tested using different data points. In this testing step, the model is checked for correctness and accuracy by providing a test dataset to it.
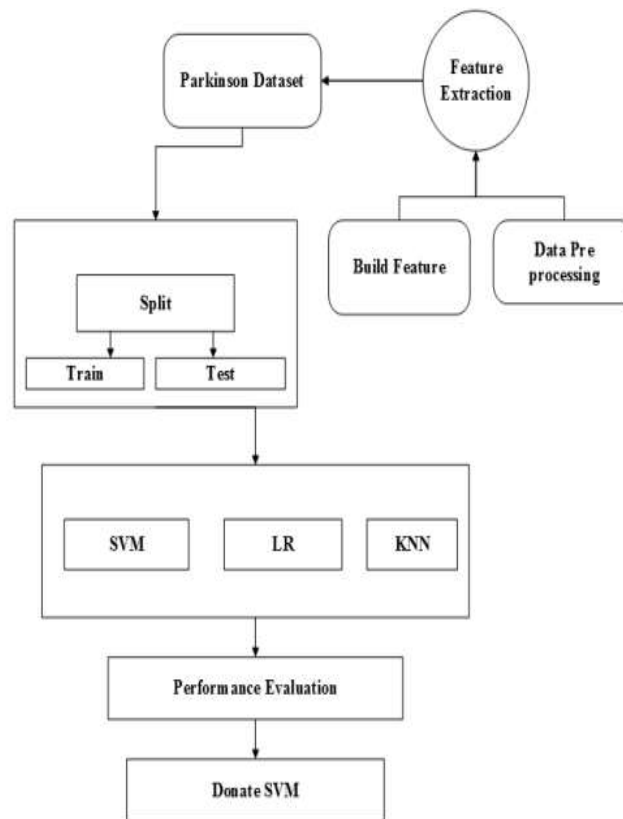
IV. WORKING PROCEDURE:-

K-Nearest Neighbor: The k-nearest neighbors (KNN) algorithm may b e a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand. It belongs to the supervised learning domain.
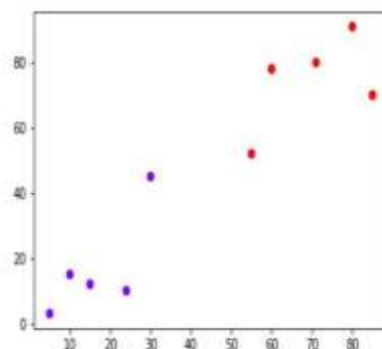
Naïve Bayes: Naive Bayes may be a statistical classification technique supported Bayes Theorem. It is one of the simplest supervised learning algorithms.

$$P\left(\frac{h}{D}\right) = \frac{P\left(\frac{D}{h}\right) \cdot P(h)}{P(D)} \longrightarrow (1)$$

Naive Bayes classifier may be a fast, accurate, and reliable algorithm. Naive Bayes classifiers have high accuracy and speed on large datasets. Naive Bayes classifier assumes that the effect of a specific feature during a class is independent of other features.

Logistic Regression: Logistic regression is additionally one among the foremost popular Machine Learning algorithms, which comes under the Supervised Learning technique. it's used for predicting the specific variable employing a given set of independent variables. It becomes a classification technique only a choice threshold is brought into the image.

XG Boost –

XXGBoost (Extreme Gradient Boosting) is a popular machine learning algorithm widely used for supervised learning such as classification and regression. It is a learning algorithm that combines multiple weak models (decision trees) to create a strong model that can accurately predict new data.

SVM -

SVM (Support Vector Machine) is a popular machine learning algorithm used for classification and regression analysis. It works by finding the best hy perplane that aligns the edges of two data points.
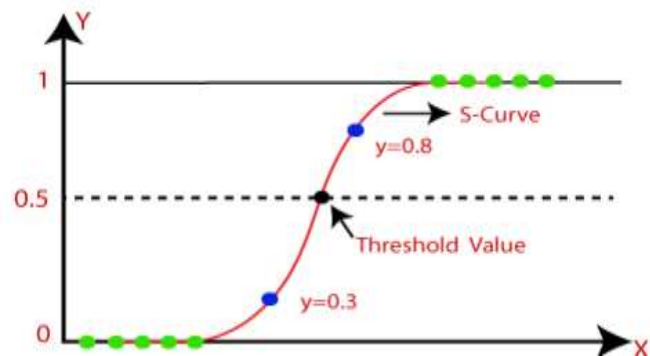
VI. CONCLUSION

The program aims to cover broad areas of brain diagnostics and machine learning so that researchers in the field can identify the latest tools in the field. We also emphasize the importance of early diagnosis and prognosis of Parkinson's disease in order to provide treatment and support to patients as soon as possible.

VII. REFERENCES

[1] Iqra Nissar, Danish Raza Rizvi, Sarfaraz Masood, Aqib Nazir Mir Year: 2019 VoiceBased Detection of Parkinson's Disease through Ensemble Machine Learning Approach: A Performance Study PHAT EAI. DOI: 10.4108/eai.13-7-2018.162806.

[2] Sharma, Kirti and Mishra, Ashutosh, Prediction of Parkinson's Disease Using Machine Learning Techniques (December 4, 2020). e-journal - First Pan IIT International Management Conference– 2018.

[3] Amit Kumar Patra et al 2019,"Prediction of Parkinson's disease using Ensemble Machine Learning classification from acoustic analysis" ,J. Phys.: Conf. Ser. 1372 012041.

[4] Adams, Warwick R. "High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing." PloS one vol. 12,11 e0188226. 30 Nov.2017, doi:10.1371/journal.pone.0188226.

[6] Muhlenbach, F. and Rakotomalala, R. (2015) Discretization of Continuous Attributes. In: Wang, J., Ed., Encyclopedia of Data Warehousing and Mining, Idea Group Reference, 397-402.

[7] Olanrewaju, R.F., Sahari, N.S., Musa, A.A. and Hakiem, N. (2014) Application of Neural Networks in Early Detection and Diagnosis of Parkinson's Disease. International Conference on Cyber and IT Service Management.

[8] Saman Sarraf, Danielle D. DeSouza, John Anderson, Ghassem Tofighi, DeepAD: Alzheimer′s Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI, Cold Spring Harbor Laboratory Press.