International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Detection of Parkinson's disease by using Machine Learning

[1]Kavyasri.G,[2] Keerthana. D,[3]Keerthana.K ,[4]Keerthi Reddy. B ,[5]Keerthi.K, [6]KesavaAditya.J
Prof.SanjayKumar J.H
Department of AI-ML
Malla Reddy University
Maisammaguda,Hyd.

***Abstract*— This project proposes a machine learning-based approach to detect Parkinson's Disease using Python. The proposed system utilizes a dataset of biomedical signals acquired from patients diagnosed with Parkinson's Disease and healthy individuals. The dataset is preprocessed, and relevant features are extracted. Then, several machine-learning algorithms, including Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Networks (ANN), are trained on the extracted features. The performance of the trained models is evaluated using various metrics, including accuracy, sensitivity, specificity, and area under the curve (AUC). The results show that the proposed system achieves high accuracy in detecting Parkinson's Disease using the biomedical signals. The system can potentially be used in clinical settings to aid physicians in the early detection of Parkinson's Disease, leading to better patient outcomes. Parkinson's Disease is a neurodegenerative disorder that affects millions of people worldwide. Early detection of Parkinson's Disease is essential to prevent or delay its progression and to improve patient outcomes.**

Keywords—Parkinson's disease; machine learning (ML), XGBoost, Decision tree.

I. INTRODUCTION

Parkinson's disease is a disorder that affects small regions in the brain that control movement, posture and balance. It is a complex disease that has many different symptoms, so that not everyone with the condition suffers from the same problems. Parkinson's disease is named after the British doctor who wrote the first book about the disease, in 1817, that made it an easily recognized entity. Parkinson called it, "The Shaking palsy," or "paralysis agitans." In his day, the term "agitans" referred to tremors. "Palsy" meant weakness and "paralysis" meant paralyzed, so the condition was considered a disorder of weakness and tremors, which is not completely true, as we shallsee. Parkinson was famous in his day because of his political activities as an advocate for the poor, his scientific publications on geology and his invention of the truss, in the days before surgery for hernias was available.

The recent report of the World Health Organization shows a visible increase in the number and health burden of Parkinson's disease patients increases rapidly. In China, this disease is spreading so fast and estimated that it reaches half of the population in the next 10 years. Classification algorithms are mainly used in the medical field for classifying data into different categories according to the number of characteristics. Parkinson's disease is the second most dangerous neurological disorder that can lead to shaking, shivering, stiffness, and difficulty walking and balance. It caused mainly due by the breaking down of cells in the nervous system. Parkinson's can have both motor and non-motor symptoms. The motor symptoms include slowness of movement,

rigidity, balance problems, and tremors. If this disease continues, the patients may have difficulty walking and talking. The non-motor symptoms include anxiety, breathing problems, depression, loss of smell, and change in speech. If the above-mentioned symptoms are present in the person then the details are stored in the records. In this paper, the author considers the speech features of the patient, and this data is used for predicting whether the patient has Parkinson's disease or not. Neurodegenerative disorders are the results of progressive tearing and neuron loss in different areas of the nervous system. Neurons are functional units of the brain. They are contiguous rather than continuous. A good healthy looking neuron has extensions called dendrites or axons, a cell body, and a nucleus that contains our DNA. DNA is our genome and a hundred billion neurons contain our entire genome which is packaged into it. When a neuron gets sick, it loses its extension and hence its ability to communicate which is not good for it and its metabolism becomes low so it starts to accumulate junk and it tries to contain the junk in the little packages in little pockets. When things become worse and if the neuron is a cell culture it completely loses its extension, becomes round and full of vacuoles.

II. LITERATURE SURVEY

Glenda-M.halliday,Nichola,"Parkinson's progression prediction using ml and serum cytokines. The serum samples from a clinic are tested to find Parkinson's disease and the same samples are tested using ML algorithm to detect Parkinson's disease.Blauwendraat, C., Bandres-Ciga, S. & Singleton, A. B.Predicting the progression in patients with Parkinson's disease using their voice. Lancet Neurol.2017.•Voice change is also a symptom of Parkinson's disease by applying ML algorithm. Das R. "A comparison of multi-classification methods for diagnosis of Parkinson's disease". Expert Systems With Applications"; 37:1568-1572 2010.•For methods used for testing Parkinson's disease they are ML ,DM neural, regression, decision tree in those ML show high performance.

III. PROBLEM STATEMENT

The main aim is to predict the prediction efficiency that would be beneficial for the patients who are suffering from Parkinson and the percentage of the disease will be reduced. Generally in the first stage, Parkinson's can be cured by the proper treatment. So it's important to identify the PD at the early stage for the betterment of the patients.

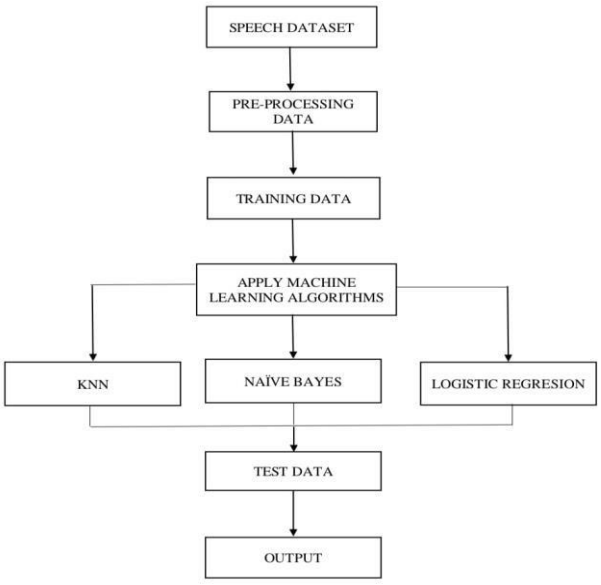
The main purpose of this research work is to find the best prediction model i.e. the best machine learning technique which will distinguish the Parkinson's patient from the healthy person. The techniques used in this problem are KNN, Naïve Bayes, and Logistic Regression. The experimental study is performed on the voice dataset of Parkinson's patients which is downloaded from the Kaggle. The prediction is evaluated using evaluation metrics like confusion matrix, precision, recall accuracy, and f1-score. The author used feature selection where the important features are taken into consideration to detect Parkinson's.

IV. SYSTEM DESIGN

Machine learning has given computer systems the ability to automatically learn without being explicitly programmed. In this, the author has used three machine learning algorithms (Logistic Regression, KNN, and Naïve Bayes). 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:

- The architecture diagram is defined with the flow of the process which is used to refine the raw data and used for predicting the Parkinson's data.
- The next step is preprocessing the collected raw data into an understandable format.
- Then we have to train the data by splitting the dataset into train data and test data.
- The Parkinson's data is evaluated with the application of a machine learning algorithm that is Logistic Regression, KNN, and Naïve Bayes algorithm, and the classification accuracy of this model is found.
- After training the data with these algorithms we have to test on the same algorithms.
- Finally, the result of these three algorithms is compared on the basis of classification accuracy.

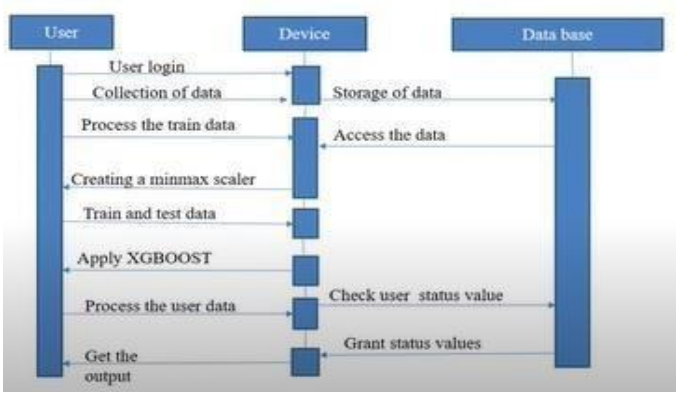
ARCHITECTURE DIAGRAM:



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. This is often one among the crucial steps of knowledge preprocessing as by doing this, we will enhance the performance of our machine learning model.Suppose, if we've given training to our machine learning model by dataset and that we test it by a totally different dataset. Then, it'll create difficulties for our model to know the correlations between the models.If we train our model alright and its training accuracy is additionally very high, but we offer a replacement dataset there to, then it'll decrease the performance. So we always attempt to make a machine learning model which performs well with the training set and also with the test dataset.

SEQUENCE DIAGRAM:



DATA PRE-PROCESSING:

The main aim of this step is to study and understand the nature of data that was acquired in the previous step and also to know the quality of data. A real-world data generally contains noises, missing values, and maybe in an unusable format that cannot be directly used for machine learning models. Data pre-processing is a required task for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. Identifying duplicates in the dataset and removing them is also done in this step.Actually, in this dataset, we have 755 features out of which some may not be useful in building our model. So, we have to leave out all those unnecessary features which are not responsible to produce the output. If we take more features in this model the accuracy we got is less. When we check the correlation of the features, some of them are the same. correlation of the columns where two of the columns have similar values. So, one of them is removed.

DATA FLOW DIAGRAM:

V. EXPERIMENT RESULT

To demonstrate the results of our project, we take the remaining test data and it is tested using three algorithms. After that our trained model to ready to predict the disease is present or not. The test accuracy is done in the Google colab which is our python notebook.First, KNN algorithm is trained with the tested with the remaining test data Third, the Logistic Regression algorithm is trained with the training dataset and later it was tested with the remaining test data.
From this three techniques, we got Naïve Bayes with more accuracy and this model is used in front end. The model is loaded into the pickle file and that file is opened in the frontend and compares the user input values with this corresponding model. Finally it results with a text message displaying that either the patient having Parkinson's disease or not.

VI. CONCLUSION

Parkinson's disease has been plaguing humans for thousands of years and was described in detail in ancient medical writings. Early sufferers from it effects were treated with varying results by a variety of plant-based treatments, some of which are still in use today. With the discovery of dopamine in the twentieth century and the subsequent development of dopamine replacement therapy, plus surgical techniques such as deep brain stimulation (DBS), many of the debilitating symptoms are now successfully treated—at least for a time.Despite the increased attention on Parkinson's, there is still no diagnostic test that is definitive. Diagnosis is made based on presenting symptoms and tested by medicating with levodopa. Only on postmortem can the diagnosis be confirmed.There is an ever-increasing understanding that PD is more than a motor disorder. Research into the nonmotor symptoms of PD is the focus of intense research, and there is hope of developing treatments that not only arrest the progress of the disease but stop it in its tracks.While research into the genetic basis of PD continues, pharmacologic treatment remains the mainstay. and may even affect the course of the disease by stimulating the production of protective neurotransmitters.Despite these advances, the medical management of PD is complex, requiring knowledge of multiple medications that interact in sometimes unforeseen ways. Deep brain stimulation has helped some patients control some symptoms but does not provide across-the-board improvement. A number of gene therapy trials are under way and are showing promise, most focusing on the dopamine pathway. Stem cell therapy appears promising but results are currently inconclusive.As PD progresses to the advanced stage, care becomes increasingly complicated. The side effects of years of PD medications begin to take their toll, requiring additional medications to address worsening sleep disorders, gastric dysfunction, and a host of other difficulties. In light of these challenges, research into neuroprotective therapies is occurring at a feverish pace. The hope is to find the cause of PD, along with treatments that stop the disease from progressing. Of particular interest, PD research is uncovering what may turn out to be a common pathophysiologic mechanism underlying dementia and PD. For now, healthcare providers must continue to educate themselves about currently available treatments and hope for better alternatives in the near future.

VII. FUTURE WORK

In future, these models can be trained with different datasets that have best features and can be predicted more accurately. If the accuracy rate increases, it can be used by the laboratories and hospitals so that it is easy to predict in early stages. This models can be also used with different medical and disease datasets. In future the work can be extended by building a hybrid model that can find more than one disease with an accurate dataset and that dataset has common features of two diseases. In future the work can extended to build a model that may extract more important features among all features in the dataset so that it produce more accuracy

VIII. REFERENCES

[1] A. Ozcift, "SVM feature selection based rotation forest ensemble classifiers to improve
computer-aided diagnosis of Parkinson disease" Journal of medical systems, vol-36, no. 4,
pp. 2141-2147, 2012.
[2] Anila M Department of CS1, Dr G Pradeepini Department of CSE, "DIAGNOSIS OF
PARKINSON'S DISEASE USING ARTIFICIAL NEURAL NETWORK", JCR, 7(19):
7260-7269, 2020.
[3] 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.
[4] Carlo Ricciardi, et al, "Using gait analysis' parameters to classify Parkinsonism: A data
mining approach" Computer Methods and Programs in Biomedicine vol. 180, Oct. 2019.
[5] Dr. Anupam Bhatia and Raunak Sulekh, "Predictive Model for Parkinson's Disease
through Naive Bayes Classification" International Journal of Computer Science &
Communication vol-9, Dec. 2017, pp. 194- 202, Sept 2017 - March 2018.
[6] Dr. R.GeethaRamani, G.Sivagami, ShomonaGraciajacob "Feature Relevance Analysis
and Classification of Parkinson's Disease TeleMonitoring data Through Data Mining"
International Journal of Advanced Research in Computer Science and Software
Engineering, vol-2, Issue 3, March 2012.
[7] Dragana Miljkovic et al, "Machine Learning and Data Mining Methods for Managing
Parkinson's Disease" LNAI 9605, pp. 209-220, 2016.
[8] FarhadSoleimanianGharehehopogh, PeymenMohammadi, "A Case Study of
Parkinson's Disease Diagnosis Using Artificial Neural Networks" International Journal of
Computer Applications, Vol-73, No.19, July 2013.
[9] Heisters. D, "Parkinson's: symptoms, treatments and research". British Journal of
Nursing, 20(9), 548–554. doi:10.12968/bjon.2011.20.9.548, 2011.
[10] 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.
[11] M. A. E. Van Stiphout, J. Marinus, J. J. Van Hilten, F. Lobbezoo, and C. De Baat,
"Oral health of Parkinson's disease patients: a case-control study" Parkinson's disease, vol-2018, Article ID 9315285, 8 pages, 2018