IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Adaption of Best Methodology by Comparing Different Algorithm to Detect Parkinson Disease

[1]Vaibhav Kesarwani, [2]Neha Chauhan, [3]Rishabh Kumar Chauhan,[4]Sanjeev Ranjan,[5]Saurabh Kumar Singh
[1345]Student, [2]Assistant Professor
[12345]Computer Science and Engineering,
[12345]Babu Banarasi Das Institute Of Technology and Management, Lucknow, India

Abstract: Detecting Parkinson's disease (PD) at an early stage with precision holds great importance in enhancing its management and overall prognosis. The objective of this research was to determine the most effective methodology for predicting PD. To accomplish this, an extensive analysis was conducted, exploring diverse prediction techniques, including machine learning algorithms such as Support Vector Machines, Random Forests, and Neural Networks. The performance of these methods was assessed using key metrics like accuracy, sensitivity, and specificity. The outcomes revealed that combining Random Forest and Neural Networks proved to be the optimal approach for PD prediction, demonstrating remarkable accuracy and robust performance. This underscores the criticality of selecting an appropriate prediction method for PD and underscores the advantages of employing a combination of algorithms to achieve enhanced prediction outcomes.

IndexTerms **- Support Vector Machine (SVM), Random Forest, XGBoost, Decision Tree, Parkinson Disease**

I. Introduction

Parkinson's disease (PD) is a progressive neurological disorder that affects the central, peripheral, and enteric nervous systems in humans, as outlined in the seminal work by Braak (2000). The pathological process underlying PD develops gradually and continuously, impacting various neural systems. It specifically targets vulnerable types of nerve cells, leading to alterations in the neuronal cytoskeleton. Over time, these affected neurons give rise to abnormal protein aggregates known as Lewy bodies within their cell bodies and Lewy neurites within their neuronal processes, as detailed by Braak (2000). The treatment landscape for PD has significantly expanded, encompassing interventions for both early and advanced stages of the disease, alongside greater recognition of non-motor manifestations, as highlighted by Davie (2008). Sveinbjornsdottir (2016) revealed that up to 80% of dopaminergic cells in the Nigro-striatal pathway are typically destroyed before the cardinal motor symptoms of PD manifest. The diagnosis of PD is typically based on the presence of characteristic motor signs, as defined by the UK PD Brain Bank criteria. These signs include bradykinesia, which refers to a slowing down of voluntary movements in terms of speed and amplitude, along with at least one additional symptom, such as muscle rigidity, resting tremor, or postural instability. The cost of Parkinson's disease (PD) treatment poses a significant financial burden for many patients, rendering it unaffordable for a substantial portion of the affected population. Consequently, accurate prediction of PD has become a critical matter for clinical practitioners, enabling them to make informed decisions regarding the disease. In recent times, machine learning has emerged as a powerful tool for detecting PD, leveraging extensive platforms, and analyzing large volumes of medical data obtained from diverse clinical domains.

In summary, this research endeavors to contribute to the field of PD detection by employing advanced machine-learning techniques and signal-processing algorithms. By harnessing the potential of big data analysis and pattern recognition, we aim to develop a reliable and accurate method for early PD detection, enabling prompt intervention and improved patient outcomes.PD is a progressive neurodegenerative disorder with distinct pathological characteristics. The understanding of PD has expanded, leading to improved treatment options across various disease stages and increased recognition of non-motor symptoms. The diagnosis of PD primarily relies on the presence of specific motor signs, particularly bradykinesia, accompanied by other hallmark symptoms.

II. LITERATURE SURVEY

The application of machine learning algorithms has demonstrated significant potential in the detection of Parkinson's disease. These algorithms leverage diverse datasets comprising clinical assessments, brain imaging data, and physiological measurements to uncover patterns and make predictions regarding disease progression. The integration of information from multiple sources allows for a comprehensive analysis, enhancing the accuracy and effectiveness of the diagnostic process. Analysis of voice data is important in the present decade to understand diagnostic methods for human diseases. The present method provides the diagnosis of PD using a voice dataset through machine learning algorithms. The voice dataset for Parkinson's disease has been retrieved from the UCI Machine Learning repository from the Centre for Machine Learning and Intelligent Systems. The dataset range of

biomedical voice measurements from 31 people, where 23 people are showing Parkinson's disease. The class column represents "status" which is set to 0 for healthy and 1 for PD. The data is in ASCII CSV format.[1]
This paper aimed to accurately predict the motor UPDRS and total UPDRS scores from the 16 voice features from the Parkinson's Telemonitoring Dataset. Various Machine Learning models such as SVM, Decision Trees, Linear Regression, and Resilient BP were trained on the dataset and their accuracy was measured. The ML algorithms were also compared and contrasted in light of the particular data. We were able to achieve desirable accuracy and predict the UPDRS scores in the expected way. The limitations of the current work would be that no matter how automated the process of Parkinson's prediction becomes, there still will be a need for human intervention, intelligence, and experience to make the diagnosis an accurate one. For future works, the dataset could be modelled on other more fitting Machine Learning models to improve the accuracy of prediction. [2]
The recently developed Meta-cognitive Fully Complex-valued Radial Basis Function (McFCRBF) network has been applied for predicting Parkinson's disease. Performance comparison of the meta-cognitive fully complex-valued RBF network (Mc-FCRBF) applied for Parkinson's disease prediction shows a better prediction of the disease when compared to a real-valued extreme learning machine and FC-RBF network. The improvement in performance is attributed to the self-regulatory learning mechanism of the meta-cognitive component. [3]
Support vector machines (SVMs), decision trees, and artificial neural networks (ANNs) are widely employed algorithms for the detection of Parkinson's disease. These algorithms have demonstrated their effectiveness in accurately identifying the presence of the disease. SVMs, in particular, have been recognized for their high accuracy rates, making them suitable for cases with substantial amounts of data. Moreover, SVMs exhibit remarkable performance in detecting early-stage Parkinson's disease, enabling early intervention and treatment.

III. METHODOLOGY

Our project aims to detect Parkinson's disease by assessing the motor function of patients, with a specific focus on achieving high accuracy in identifying the disease at its early stage. The dataset utilized for this prediction consists of patient data with 24 feature attributes, including columns such as Jitter and Pitch frequency. By preprocessing the dataset, which involves removing outlier values and standardizing the data, we ensure that it is suitable for model development. This research focuses on the application of six classification algorithms to a specific dataset. The algorithms employed include Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN), and XGBOOST (Extreme Gradient Boosting). The objective is to predict whether an individual is healthy or affected by Parkinson's disease based on voice input parameters.
A thorough comparison of the results obtained from these algorithms is conducted. The findings indicate that ensemble techniques exhibit more favorable outcomes compared to the base classification algorithms. Notably, ensemble techniques such as XGBOOST and Bagging Classifier yield more accurate predictions. Conversely, the base classification algorithms including Logistic Regression, SVM, Decision Tree, and KNN demonstrate satisfactory performance but are outperformed by the ensemble methods. These results underscore the effectiveness of ensemble techniques in enhancing prediction accuracy for Parkinson's disease detection. Further research and experimentation can be pursued to explore the potential of other ensemble models and refine the diagnostic process.. Additionally, we utilise sk-learn, a comprehensive library that provides essential tools for developing and implementing machine learning classifiers on the preprocessed dataset. By leveraging these advanced machine learning techniques and tools, we aim to achieve accurate and reliable detection of Parkinson's disease at its early stage, contributing to improved prognosis and timely intervention for patients.

A. Dataset Information

The dataset utilized in this study comprises a diverse range of biological voice measurements obtained from a group of 31 individuals, out of which 23 have been diagnosed with Parkinson's disease (PD). Each column in the dataset represents a distinct voice measure, while each row corresponds to one of the 195 voice recordings collected from these individuals, identified by the "name" column. The main objective of this dataset is to differentiate between individuals who are healthy and those who have Parkinson's disease, which is indicated by the "status" column. A value of 0 is assigned to healthy individuals, while a value of 1 is assigned to those diagnosed with PD.

B. Proposed System

The research project focused on developing a system to detect Parkinson's disease using supervised machine learning algorithms. The first step was to gather training data with labelled examples indicating whether individuals had Parkinson's disease or not. The dataset used for training the model was obtained from the Parkinson's dataset available at the University of California Irvine (UCI) machine learning repository. This dataset was created by Max Little of the University of Oxford in collaboration with the National Centre for Voice and Speech in Denver, Colorado. The dataset consists of speech signals recorded by the National Centre for Voice and Speech, and it was originally used to extract features for general voice disorders. The dataset is publicly accessible and provides valuable information for training the machine learning model. To develop the model, various supervised machine-learning algorithms were employed. These algorithms were trained on the labelled data from the Parkinson's dataset, enabling them to learn patterns and relationships between voice measurements and the presence of Parkinson's disease. The model was then able to predict whether an individual has Parkinson's disease or not by analyzing their voice measurements. By utilizing the dataset from the UCI machine learning repository and applying supervised machine learning techniques, the research aimed to create a reliable and accurate system for Parkinson's disease detection based on voice analysis

IV. IMPLEMENTATION

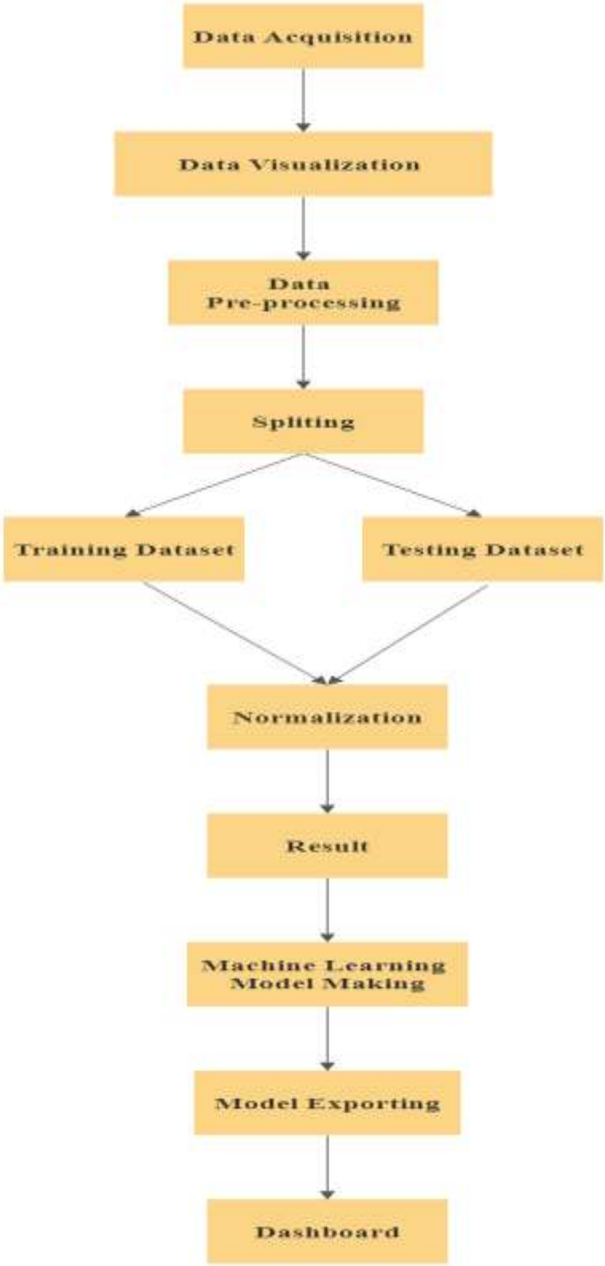

Fig.4.1 Workflow

1. Data Acquisition

We discovered that the data has 195 rows and 24 columns by using the "shape" command. Then we used '.info' to examine the data types of each column. Then we looked at the null values. Null values are nothing more than missing data () is mostly used to determine whether or not there is any missing data in our dataset. We discovered no missing values in the dataset. Then, using describe (), we examine certain statistical calculations of our dataset, such as the mean, median, count, and so on.

2. Data Visualization

We visualized how many rows (or people) had symptoms of Parkinson's in this area. We discovered that 147-row values (person) suggest Parkinson's disease, while 48 indicate normal. We have created a pie graph in order to better comprehend this unexpected finding. Following that, we created a heat map demonstrating the correlation between all dataset features and columns.

3. Data Pre-processing

Data Pre-processing is carried out to enhance the functionality or accuracy of the model that will determine if the subject has Parkinson's disease or not. The name column was originally removed since it wouldn't be helpful in building models because it didn't provide any meaningful information. The status column, which indicates whether a person is well or ill, was then divided. The data was then normalized using MinMaxScaler.

4. Normalization of Data

Normalization avoids problems with raw data and many datasets by creating new values and maintaining a wide distribution and ratio in the data. To improve the efficiency and accuracy of machine learning models, it also makes use of a variety of techniques and methods.

The data was then divided into training and testing data. We decided to divide the data into 70% training data and 30% testing data. Following the division, the training data now has 136 rows and 22 columns (since the status column was previously divided), while the test data has 59 rows and 22 columns.

5. Training and Testing data

Testing or validation data are used to assess the correctness of your model, whereas training data is the first dataset you use to teach a machine learning programme to recognise patterns or perform to your criteria.

6. Machine Learning Model's development

We developed several models, including KNN, XGBoost, Random Forest, Support Vector Machine (SVM), and Logistic Regression.
The Random Forest model outperformed the others in terms of accuracy and AUC/ROC score.

```
display(chart)
```

	Metric	DT	RF	LR	SVM	NB	KNN	XGB
0	Accuracy	0.932203	0.966102	0.830508	0.966102	0.762712	0.966102	0.915254
1	F1-Score	0.920000	0.961538	0.782609	0.960000	0.650000	0.960000	0.909091
2	Recall	0.884615	0.961538	0.692308	0.923077	0.500000	0.923077	0.961538
3	Precision	0.958333	0.961538	0.900000	1.000000	0.928571	1.000000	0.862069
4	R2-Score	0.724942	0.862471	0.312354	0.862471	0.037296	0.862471	0.656177

Fig. 4.2 Accuracy Details of Trained & Tested Machine Learning Algorithms

7. Exporting the Model

In order to make data-driven business choices, deployment is the act of integrating a machine-learning model into an already-existing production environment. It takes the most time and is one of the final phases in the machine learning process. One of the most challenging steps in getting the most out of machine learning is model deployment. Simply stated, model deployment is utilized and crucial for offering a better intuitive user interface and visuals for the model.

How should our random forest model be deployed?

The simplest method to implement a machine learning model is to build a prediction web service.

Pickle the model by saving it as a pickle file. This makes it simple for us to access the model and make predictions. But since we cannot directly access the pickle file, doing this alone is insufficient. So I used Pickle to first store my model code file.

8. Designing Dashboard

Next, we created a Python file in which a link between the Python file and the stored model pickle file was created using Streamlit. Additionally, we connected the prediction.py file to the ipynb file. A "python" file is called prediction.py. We created this file, which is connected to the Streamlit frame module file, using this python file. Our web service page has a design thanks to the Streamlit module file, which also improves the appearance of websites.

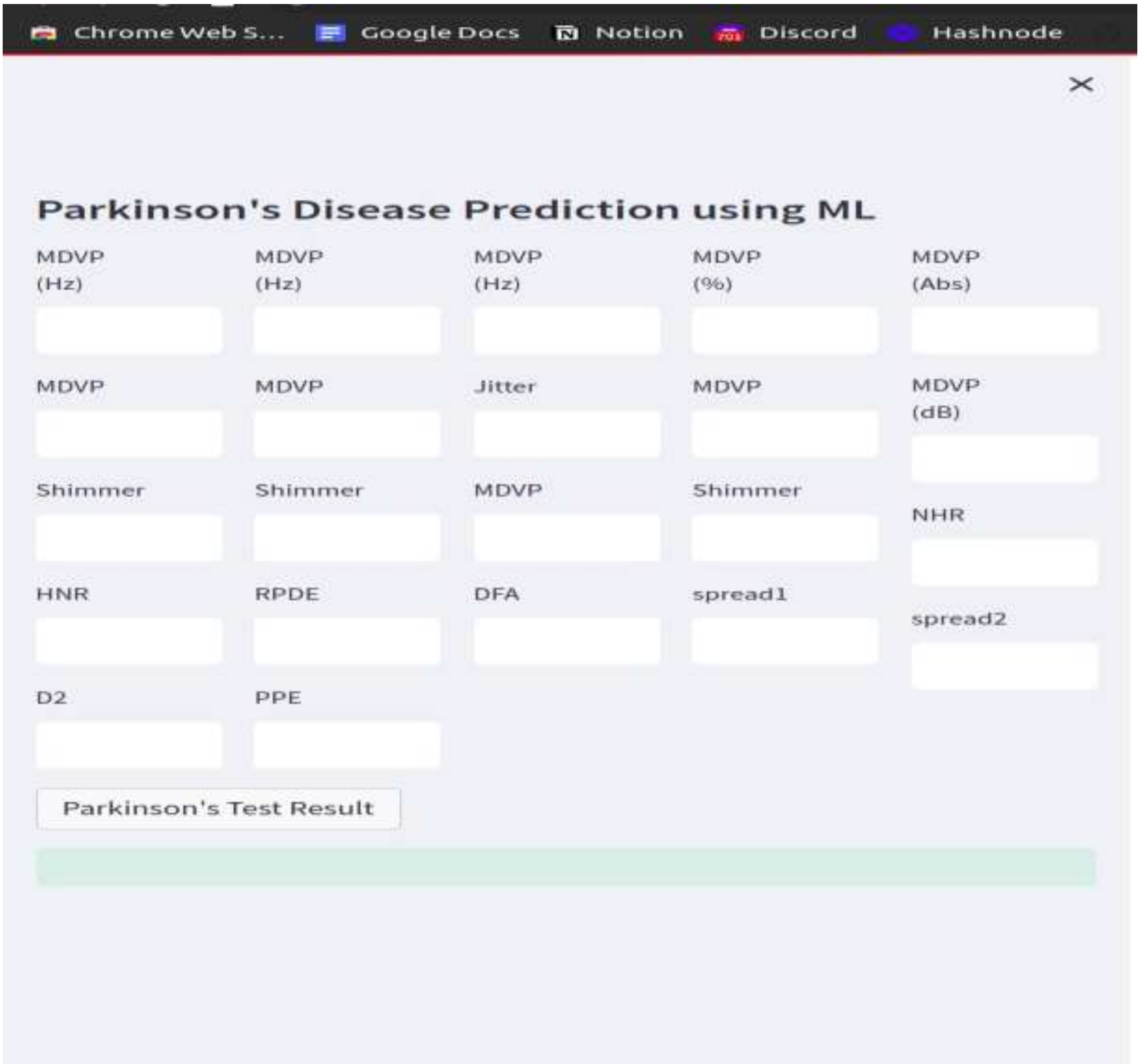

Fig. 4.3 Dashboard

V. CONCLUSION

We have created a powerful machine-learning model that can accurately diagnose Parkinson's disease in people based on given details. The use of machine learning has significantly helped both the healthcare industry and the general public detect disorders such as Parkinson's more simply and affordably as digitalization continues to revolutionize several industries, including healthcare. Furthermore, the use of machine learning goes well beyond the diagnosis of diseases and has significant applications in the pharmaceutical and healthcare industries. One aspect of the many applications where machine learning may be used to detect diseases is only one of them.

VI. References

1. B K Varghese, G. B. Amali D*, Uma Devi K S "Parkinson's Disease Prediction Using Machine Learning Approaches" 2013 Fifth International Conference on Advanced Computing (ICoAC)
2. Gokul.S, Sivachitra.M, "Parkinson's Disease Prediction Using Machine Learning Approaches" 2013 Fifth International Conference on Advanced Computing (ICoAC)
3. Early Prediction of Parkinson Disease Using Machine Learning and Deep Learning Approaches Harshvardhan Tiwari, Shiji K Shridhar, Preeti V Patil, K R Sinchana and G Aishwarya.
4. A Comparative Analysis Of Parkinson Disease Prediction Using Machine Learning Approaches F.M. Javed Mehedi Shamrat, Md. Asaduzzaman, A.K.M. Sazzadur Rahman, Raja Tariqul Hasan Tusher, Zarrin Tasnim.

5. Comparative Analysis Of The Early Detection Of Parkinson's Disease J. Olivia Capitola Assistant professor, Computer Application Department & Affiliated to Bharathidasan University, Bishop Heber College, Tiruchirappalli
6. Support Vector Regression and its Mathematical Implementation, Rahul Rastogi, Medium, 2020.
7. Parkinson’s Disease Detection using Machine Learning - Python, 2021, Siddhardhan, YouTube.
8. Timothy J. Wroge, Yasin Özkanca, C. Demiroğlu, Dong Si, David C. Atkins and R. Ghomi (2018), “Parkinson’s Disease Diagnosis Using Machine Learning and Voice”, Computer Science, IEEE Signal Processing in Medicine and Biology Symposium (SPMB).
9. Saloni, R. K. Sharma, and A. K. Gupta, “Detection of Parkinson Disease Using Clinical Voice Data Mining,” vol. 9, pp. 320–326, 2015.
10. Tarigoppula V.S Sriram , M. Venkateswara Rao , G V Satya Narayana3 , DSVGK Kaladhar4 , T Pandu Ranga Vital, “Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms” International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 3, September 2013
11. Aniruddha, 2020. *AUC-ROC Curve in Machine Learning Clearly Explained.* [Online] Available at: https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine- learning/.
12. Braak, H. a. B. E., 2000. Pathoanatomy of Parkinson’s disease. *Journal of neurology*, 247(2), pp. II3--II10.