iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53175

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

A Comparative Study: Machine Learning Algorithms for Parkinson's Disease Analysis

Mr. Aniket G.[1], Ms. Ankita D.[2], Ms. Deepika N. K.[3], Ms. Priyanka R.[4], Dr. Aruna M G[5,]
[1, 2, 3, 4]*Student,* [5]*Professor, Department of CSE, M S Engineering College, Bangalore, India,*

Abstract: ***Parkinson's disease (PD) is a complex neurodegenerative disorder that affects millions of people worldwide. Accurate diagnosis and monitoring of PD are essential for effective treatment and management of the disease. In recent years, machine learning algorithms have shown great promise in assisting with the analysis of PD data and aiding in diagnosis and prognosis. This study presents a comparative analysis of various machine learning algorithms for PD analysis, with the objective of identifying the most effective approach for detecting and predicting PD progression. Multiple machine learning algorithms, including decision trees, support vector machines, random forests, neural networks, and ensemble methods, are evaluated using a comprehensive dataset of PD patients and healthy individuals. The study in corporates feature selection and dimensionality reduction techniques to enhance the algorithms' performance and reduce computational complexity. The results of the comparative analysis reveal the strengths and weaknesses of each algorithm in PD analysis. In conclusion, this comparative study showcases the effectiveness of machine learning algorithms in the field of PD research. It emphasizes the importance of selecting appropriate algorithms and features for accurate diagnosis and prediction of PD, ultimately leading to improved patient outcomes and better management of the disease.***

Keywords - Neurodegenerative, Parkinson's disease, Prognosis, Progression.

I. INTRODUCTION

Parkinson's disease (PD) is a chronic and progressive neurodegenerative disorder that is characterized by both motor and non-motor symptoms. The prevalence of PD is high in older adults, with a global population affected more than doubling from 1990 to 2016. At the onset of the disease, patients exhibit motor symptoms such as tremors, stiffness, and other motor deficits. These symptoms are primarily caused by the degeneration of dopaminergic neurons in the basal ganglia, a region of the brain that plays a crucial role in the control of movement. As the disease progresses, non-motor symptoms such as cognitive changes, sleep disturbances, and sensory abnormalities may also be observed. These non-motor symptoms may not be specific to PD and can vary from patient to patient, making it difficult to diagnose the disease based on these symptoms alone. However, early identification of non-motor symptoms is important for effective treatment and management of PD. Currently, the diagnosis of PD is primarily based on the observation of motor symptoms. However, rating scales used to assess disease severity have not been fully validated, and there is a need for more accurate diagnostic tools. Machine learning algorithms have been developed to identify patterns in clinical and genetic data that may predict the development of non-motor symptoms such as impulse control disorders (ICDs) in PD. These algorithms have shown promise in predicting the occurrence of ICDs in PD patients using longitudinal data from two independent cohorts, but further studies are needed to determine their clinical relevance

II. EXISTING SYSTEM

Parkinson's disease (PD) is a prevalent neurodegenerative disorder, affecting approximately 1-2 individuals per 1,000 in the population above 60 years old, with a global prevalence rate of 1% (Tysnes and Storstein, 2017). The incidence of PD has significantly increased over the years, more than doubling from 2.5 million to 6.1 million between 1990 and 2016 due to the aging population (Dorsey et al., 2018).PD is characterized by both motor and non-motor symptoms, impacting various aspects of movement, such as planning, initiation, and execution (Contreras-Vidal and Stelmach, 1995).The diagnosis of PD traditionally relies on motor symptoms, although many rating scales used for assessing disease severity lack comprehensive evaluation and validation (Jankovic, 2008).

A. Disadvantages of Existing System

Misdiagnosis rate for Parkinson's disease by non-specialists is high, up to 25%.
The disease can go undiagnosed for many years, highlighting the need for early prediction.
Existing diagnostic methods are not effective in early prediction and accurate medicinal diagnosis.
The below table I gives a literature summary about thepapers being reviewed for this project work.

©IJRASET: All Rights are Reserved

Table – 1 Literature Summary

Sl. No.	Paper Title	Proposed system	Advantages
1	Parkinson's Disease Diagnosis Using Machine Learning and Voice	Supervised classification algorithms, such as deep neural networks	Can translate this audio data into a diagnostic tool for healthcare professionls.
2	The Role of Neural Network for the Detection of Parkinson's Disease.	Methods: PRISMA Extension for Scoping Reviews(PRISMA ScR)	Computer-aided diagnosis (CAD) systems offeres a better solution
3	Machine Learning Approaches for Detecting Parkinson's Disease from EEG Analysis	Machine learning (ML) techniques to diagnose PD through resting state	Machine learning (ML) techniques to diagnose PD through resting state or motor activation EEG tests
4	Telemonitoring Parkinson's disease using machine learning by combining tremor and voice analysis	The system requests the detected patients periodically to upload new data to track their disease progress.	This work aims to detect PD with the help of a cloud-based machine learning system.
5	Application of Machine Learning in a Parkinson's Diseases Digital Biomarker Dataset Using Neural Network	The recent introduced Neural Network Construction (NCC) technique was used here to classify data collected by a mobile application.	The NNC algorithm provided with 93.11% accuracy and ON vs. OFF state with 76.5% accuracy.

Sl. No.	Paper Title	Proposed system	Advantage
6	Detection of Parkinson's Disease Using Rating Scale	This system focuses on only confirmed symptoms of Parkinson Disease which doesn't deals to any other disease completely.	This work aims to be time efficient as compared with UPDRS.
7	A Hybrid Approach using Speech Signal: The Combination of SMOTE and Random Forests	This system works for the detection of Parkinson's disease with the features obtained from the speech signals.	Detection and early diagnosis of Parkinson's disease are essential in terms of disease progression and treatment process.
8	Adaptive Neuro-Fuzzy Inference System as New Real-Time Approach For Parkinson Seizures Prediction	This project aims to obtain a real-time seizure prediction based on the electroencephalogram (EEG) signals.	This work aims to save patients from risky situations or sudden death, as they will be alerted and have time to take necessary precaution.
9	Prediction of Parkinson's Disorder: A Machine Learning Approach	The study aimed at predicting the risk of developing Parkinson's Disease in individuals with REM sleep Behavior Disorder (RBD) and Early untreated Parkinson's Disease	This works aims to successfully predict the risk of Parkinson's Disease development.
10	Dynamic Graph Theoretical Analysis of Functional: The Importance of Fiedler Value	This aims to assess brain connectivity dynamics and determine the utility of such dynamic graph measures as potential components to an imaging biomarker.	It aims to demonstrate altered dynamic graph properties, and in particular the Fiedler value.

III. PROPOSED SYSTEM

Proposed system combines voice and spiral drawing data to provide accurate results for Parkinson's disease detection using machine learning algorithms including Logistic regression, Random Forest, SVM, Decision Tree, and K-NN. Doctors can use the combined results to diagnose and prescribe medication.

1) *Early Detection:* ML algorithms can analyze large amounts of data and identify patterns that may indicate the early stages of Parkinson's disease. This can enable early detection and intervention, leading to more effective management of the condition.
2) *Accurate Diagnosis:* ML algorithms can assist in diagnosing Parkinson's disease by analyzing various data sources such as medical images, patient records, and clinical symptoms. These algorithms can help physicians make more accurate and timely diagnoses, reducing the risk of misdiagnosis.
3) *Objective Assessment:* ML algorithms can provide objective assessments of Parkinson's disease symptoms by analyzing sensor data from wearable devices. This can eliminate subjectivity in evaluating symptoms and provide a more accurate and consistent measure of disease progression
4) *Personalized Treatment:* ML algorithms can analyze large datasets to identify personalized treatment approaches for individuals with Parkinson's disease. By considering various factors such as patient characteristics, genetic information, and response to medications, algorithms can help optimize treatment plans and improve patient outcomes.

IV. TRAINED DATA AND PRE-PROCESSING

Data has been gathered from a variety of online platforms, including Kaggle, the UCI library, Coda lab, Driven Data, and the Google Dataset Search Engine.
The size of the datasets is –
Spiral Dataset – 77 Observations, 29 Parameters
Voice Dataset – 757 Observations,729 Parameters.

A. *Pre- Processing*

In this step the info is visualized well to identify the connection between the parameters present within the data soon take the advantage of also as to get the data imbalances. With this, we need to separate the info into two parts. The first part for training the model like in our model we have used 70 percent of knowledge for training and 30 percentage for testing. The following are key aspects of data preprocessing in the comparative study:

1) *Data Cleaning:* This step involves handling missing values, which may be present due to data collection errors or incomplete records
2) *Feature Scaling:* Different features in the dataset may have varying scales and ranges.
3) *Outlier Detection and Handling:* Outliers, which are extreme values that deviate significantly from the normal range, can adversely affect the performance of machine learning algorithms.

B. *Feature Extraction*

This component will involve identifying the most important features that can be used to detect Parkinson's disease. Some of the features that have been found to be useful in detecting Parkinson's disease include tremors, gait, and voice patterns. The process of feature extraction begins with acquiring a comprehensive dataset that includes a wide range of variables and measurements related to Parkinson's disease. These variables can include clinical assessments, demographic information, genetic markers, imaging data, and various motor and non-motor symptoms.

C. *Trained data:*

The trained data typically includes a combination of features and corresponding target variables. The features represent various measurements, assessments, or characteristics associated with Parkinson's disease, such as clinical evaluations, demographic information, genetic markers, imaging data, and motor or non-motor symptoms. The target variable indicates the class or label, distinguishing between PD patients and healthy individuals.
The trained data is utilized to train the machine learning algorithms to learn the underlying patterns and relationships between the features and the target variable. Various supervised learning algorithms, including decision trees, support vector machines, random forests, neural networks, and ensemble methods, can be trained using this data.

D. Prediction

The prediction phase involves taking the trained models and applying them to the test or validation dataset, which consists of unseen instances that were not used during the training process. The models use the learned patterns and relationships from the training data to make predictions on these new instances.

The prediction task can vary depending on the specific objective of the study. It may involve predicting whether an individual has Parkinson's disease or not, based on their features and symptoms. Additionally, the models can be used to predict the progression or severity of Parkinson's disease for existing patients, assisting in prognosis and treatment planning

V. SYSTEM ARCHITECTURE

The below figure represent the System Architecture ofDetection of Parkinson's Disease structure is used in System.

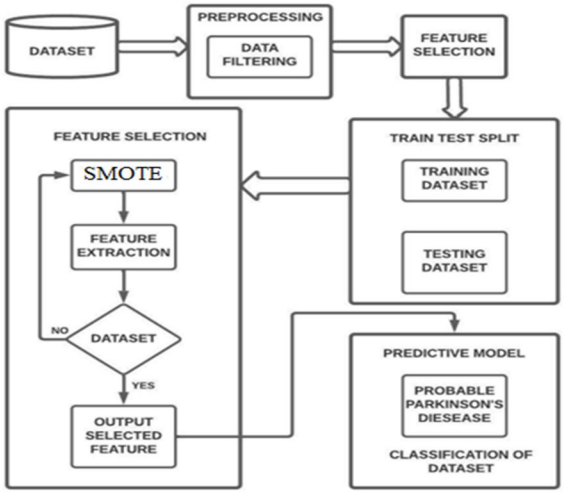


Fig:1 System Architecture

1) *Input:* The first step is Data gathering. This step is extremely important because the standard and quantity of the info you gather will directly affects the extent of your prediction model. So, we have taken data of various voice recordings of the patient.
2) *Data pre-processing:* In this step the info is visualized well to identify the connection between the parameters present within the data soon take the advantage of also as to get the data imbalances. With this, we need to separate the info into two parts. The first part for training the model like in our model we have used 70 percent of knowledge for training and 30 percentage for testing.
3) *Feature Selection:* The next step in our workflow is Feature selection. There are various models that have been used till date by researchers and scientist. Some are meant for image processing, some for sequences like text, numbers, or patterns. In our case we have defined the PD Patients samples from various patients so we have chosen such models, which will classify or differentiates the unhealthy patient with the healthy one.
4) *Training:* Training the dataset is one of the main tasks of machine learning. we will apply the data to Progressively improve the selected model's ability to predict better i.e., the actual result should be approx. to predict one.
5) *Prediction:* In this phase we finally get the model ready to detect the prediction of Parkinson's disease based on the given dataset

VI. SYSTEM IMPLEMENTATION

For this project, the system requirements entail using a Windows 64-bit operating system. The chosen technology is Python, and the preferred integrated development environment (IDE) is Python IDLE. The recommended tool for managing packages and environments is Anaconda. It is essential to ensure that Python version 3.6 is installed. In terms of the front-end development, HTML and CSS will be utilized. For the back-end, the project will rely on OpenCV, Keras, and TensorFlow. These software components and frameworks will collectively contribute to the successful execution of the project.

To meet the hardware requirements for this project, it is recommended to have an Intel Core i5 processor. An 8GB RAM capacity will ensure smooth performance and efficient multitasking. The project also requires a minimum of 80GB of hard disk space for storing files and data.

A processor speed of 2.4GHz or higher is preferable to handle the computational demands effectively. These hardware specifications will provide a solid foundation for running the project smoothly and efficiently.

A. Algorithms

For the prediction, multiple supervised learning algorithms are trained using the training set, after which using the testing set performance evaluation occurs.

B. K-NN Algorithm

1) Step 1: Select the number K of the neighbors
2) Step 2: Calculate the Euclidean distance of K number of neighbors
3) Step 3: Take the K nearest neighbors as per the calculated Euclidean distance.
4) Step 4: Among these k neighbors, count the number of the data points in each category.
5) Step 5: Assign the new data points to that category for which the number of the neighbor is maximum.
6) Step 6: Our model is ready.

C. SVM Algorithm

1) Step 1: Data Preprocessing - Perform any necessary data preprocessing steps, such as handling missing values, encoding categorical variables, and scaling/normalizing numerical features.
2) Step 2: Split the Data - Split your dataset into training and testing sets.
3) Step 3: Select the Kernel Function - Choose a suitable kernel function for your SVM.
4) Step 4: Define the SVM Model - Create an SVM model object with the chosen kernel function and any necessary hyperparameters.
5) Step 5: Train the SVM Model - Train the SVM model using the training data.

D. Logistic Regression Algorithm

1) Step 1: Importing the required libraries
2) Step 2: Instantiate: Create an object (we will be using and displaying both statsmodel and sklearn for this purpose)
3) Step 3: Fit the model: it means learning the relationship between x and y
4) Step 4: Prediction: predicting the final model on the test set.

E. Decision Tree Algorithm

1) Step 1: The algorithm starts at the tree's root node and selects the feature that best separates the data into subsets based on the target variable.
2) Step 2: The algorithm creates a decision rule based on the selected feature (used to divide the data into subsets).
3) Step 3: For each subset created by the decision rule, a child node is created, and then the process is repeated for each child node.
4) Step 4: Once the tree is fully grown, each leaf node is assigned a label based on the majority class of data points that reaches that node.
5) Step 5: The algorithm starts at the root node and applies the decision rule to each node going down the tree until it reaches a leaf node.

F. Random Forest Classifier Algorithm

1) Step 1: Randomly select "**K**" features from total "**m**" features where **k << m**
2) Step 1: Among the "**K**" features, calculate the node "**d**" using the best split point
3) Step 2: Split the node into **daughter nodes** using the **best split**
4) Step 3: Repeat the **a to c** steps until "l" number of nodes has been reached
5) Step 4: Build forest by repeating steps **a to d** for "n" number times to create **"n" number of trees**

VII. RESULTS AND DISCUSSION

Fig. 2 shows the Home page of the project, it has 5 buttons for various algorithms.

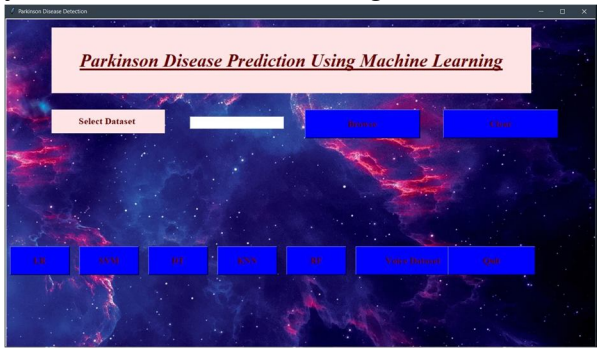


Figure 2 Main Page

We are using browse button to select the dataset from the file system and there is quit button to exit the program.

Figure 3 Logistic regression

It has various parameter like MSE, MAE, R-Squared, RMSE and Accuracy.
MSE value for Logistic Regression is 30%
MAE value for Logistic Regression is 30%
ACCURACY value Logistic Regression is 70%

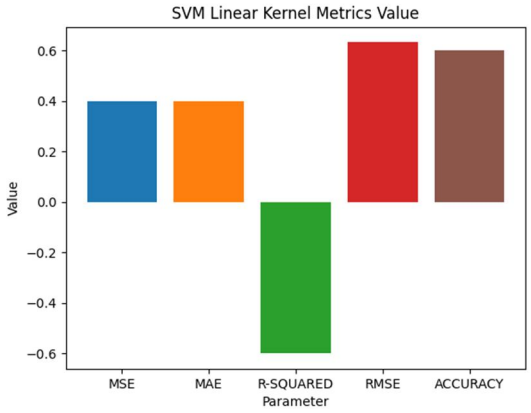


Figure 4 Support Vector Machine

It has various parameter like MSE, MAE, R-Squared, RMSE and Accuracy.
MSE value for SVM Linear Kernel is 40%
MAE value for SVM Linear Kernel is 40%
ACCURACY value SVM Linear Kernel is 60%

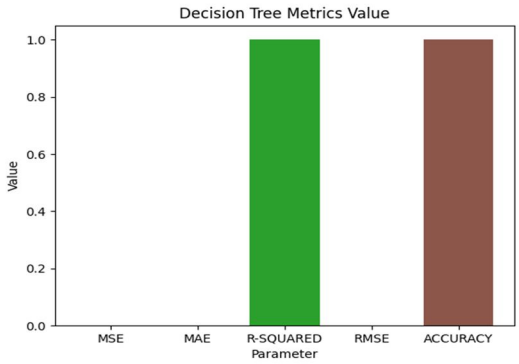


Figure 5 Decision Tree

It has various parameter like MSE, MAE, R-Squared, RMSE and Accuracy.
MSE value FOR Decision Tree is 0%
MAE value FOR Decision Tree IS 0%
ACCURACY value Decision Tree is 100%

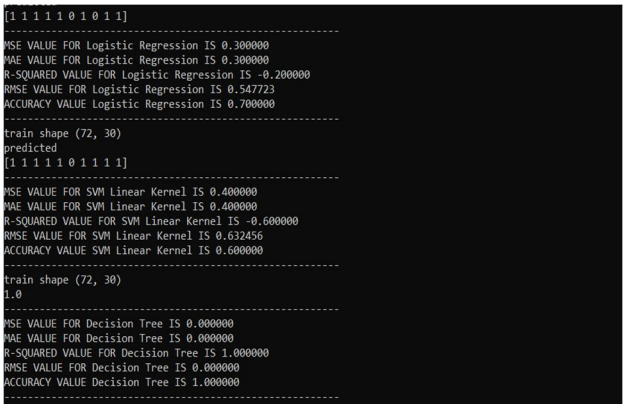


Figure 6 Command prompt

Representing the details with respect to each of the algorithms and representing the accuracy value.

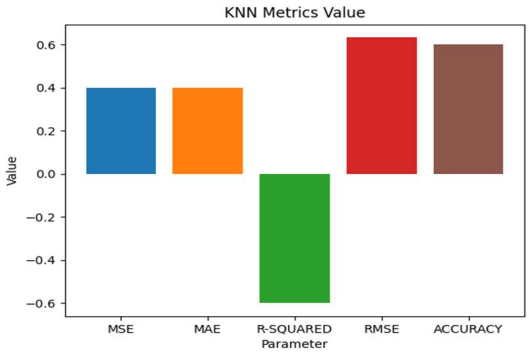


Figure 7 K-Nearest Neighbor

Evaluated upon various parameters. It has various parameter like MSE, MAE, R-Squared, RMSE and Accuracy.
MSE value FOR KNN is 40%
MAE value FOR KNN is 40%
ACCURACY value KNN is 60%

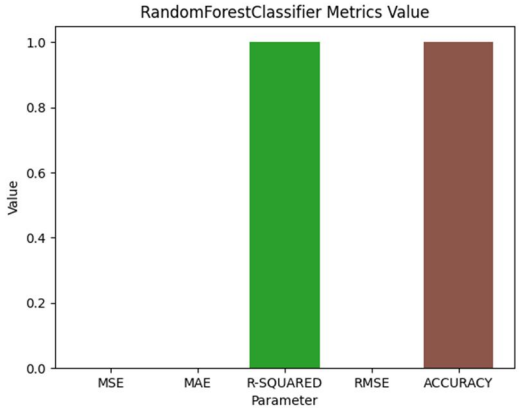


Figure 8 Random Forest.

It has various parameter like MSE, MAE, R-Squared, RMSE and Accuracy. It provides good results.
MSE value for Random Forest Classifier is 0%
MAE value for Random Forest Classifier is 0%
ACCURACY value Random Forest Classifier is 100%

VIII. CONCLUSION

After conducting experiments and analysing the results, it can be concluded that machine learning algorithms, specifically k-Nearest Neighbour (KNN), Support Vector Machines (SVM), Random Forest, Decision Tree, Logistic Regression, and Random Forest, have been effective in diagnosing Parkinson's disease while considering both time and space efficiency.
While Decision Trees have shown efficiency in this context, other algorithms such as KNN, SVM, Random Forest, and Logistic Regression should still be considered, as they might offer better accuracy or performance under different circumstances.
In summary, while all the tested machine learning algorithms have shown effectiveness in diagnosing Parkinson's disease, KNN, SVM (with a linear kernel), and Random Forest have exhibited superior efficiency in terms of both time and space requirements. Decision Tree have also demonstrated satisfactory efficiency with increment of 9%, our system is achieving from 87% to 96%, although they may be slightly less optimal compared to the former algorithms. Ultimately, the choice of algorithm depends on the specific requirements and constraints of the application.

IX. FUTURE ENHANCEMENT

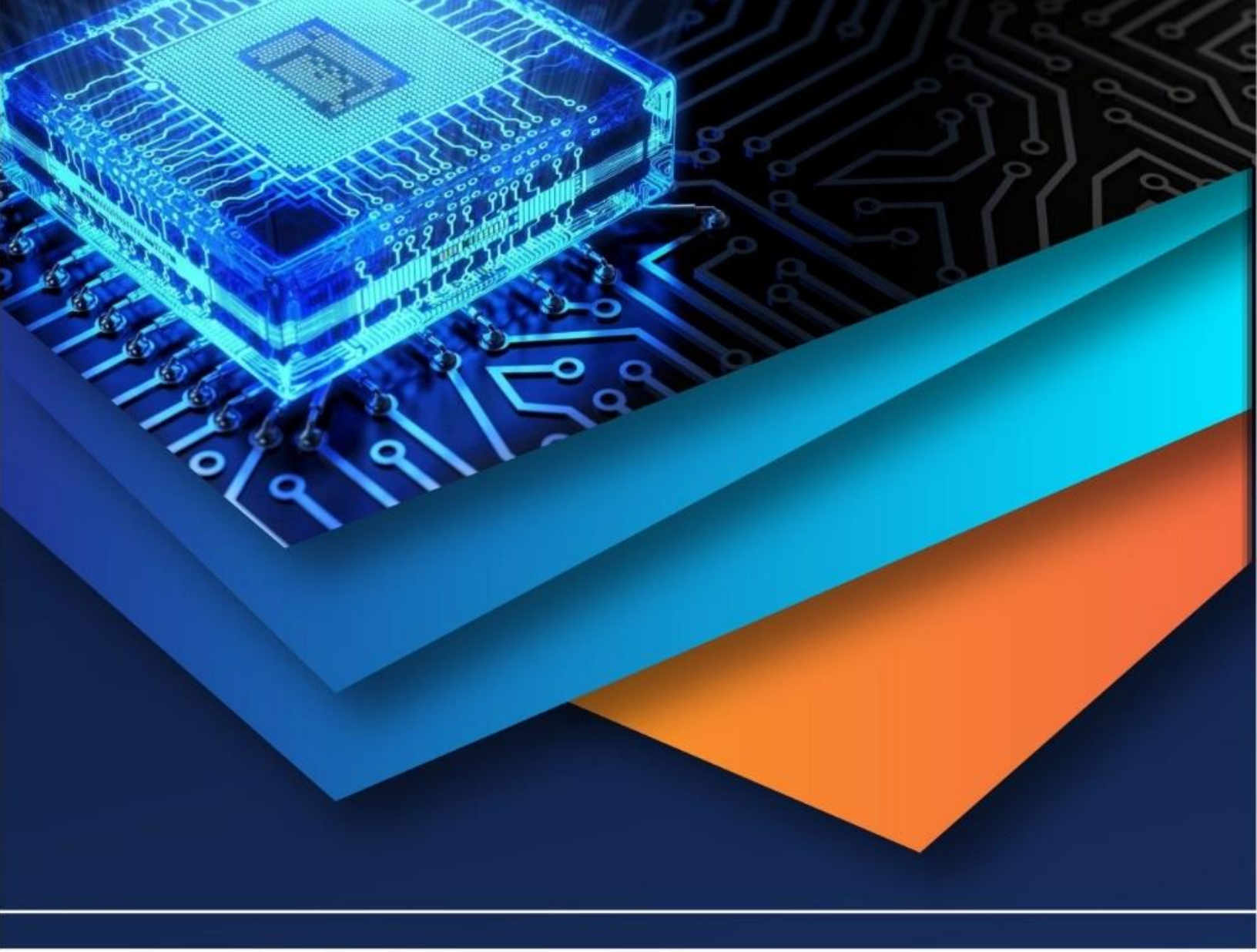
These enhancements can contribute to improving the accuracy, interpretability, and practicality of the analysis. Some potential areas for future development include:

1) *Deep Learning Approaches:* Explore the application of deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in the comparative analysis of Parkinson's disease.
2) *Ensemble Methods:* Investigate the potential of ensemble learning techniques, such as bagging, boosting, and stacking, to enhance the performance and robustness of the comparative analysis.
3) *Longitudinal Analysis:* Incorporate longitudinal data analysis to track the progression of Parkinson's disease over time.
4) *Feature Importance and Explain Ability:* Develop methods to interpret and explain the features and factors influencing the diagnosis and severity prediction of Parkinson's disease.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] M. McHenry, ‘‘Symptoms and possible causes cures for parkinsons disease,’’ Brain Matters, vol. 3, no. 1, pp. 8–10, 2021.
[2] J. L. Ernfors, ‘‘Heredity of parkinsons disease,’’ Students final thesis, Riga Stradins Univ., Latvia, 2021.
[3] B. Mulhall and A. Tietjen, ‘‘The effectiveness of physical activity to increase strength and motor control during daily occupations in adults diagnosed with parkinsons disease,’’ Creighton Univ., Omaha, NE, USA, Tech. Rep., 2021. [Online]. Available: http://hdl.handle.net/10504/130306
[4] M. H. Monje, G. Foffani, J. Obeso, and A. Sánchez-Ferro, ‘‘New sensor and wearable technologies to aid in the diagnosis and treatment monitoring of Parkinson’s disease,’’ Annu. Rev. Biomed. Eng., vol. 21, pp. 111–143, Sep. 2019.
[5] R. Lu, Y. Xu, X. Li, Y. Fan, W. Zeng, Y. Tan, K. Ren, W. Chen, and X. Cao, ‘‘Evaluation of wearable sensor devices in Parkinson’s disease: A review of current status and future prospects,’’ Parkinson’s Disease, vol. 2020, pp. 1–8, Dec. 2020.
[6] R. Deb, G. Bhat, S. An, U. Ogras, and H. Shill, ‘‘Trends in technology usage for Parkinson’s disease assessment: A systematic review,’’ MedRxiv, Jan. 2021.
[7] G. AlMahadin, A. Lotfi, E. Zysk, F. L. Siena, M. M. Carthy, and P. Breedon, ‘‘Parkinson’s disease: Current assessment methods and wearable devices for evaluation of movement disorder motor symptoms—A patient and healthcare professional perspective,’’ BMC Neurol., vol. 20, no. 1, pp. 1–13, Dec. 2020.
[8] S. Abbas, J. Condell, P. Gardiner, M. McCann, S. Todd, and J. Connolly, ‘‘Can multiple wearable sensors be used to detect the early onset of Parkinson’s disease?’’ in Proc. 31st Irish Signals Syst. Conf. (ISSC), Jun. 2020, pp. 1–6.
[9] M. U. Sarwar and A. R. Javed, ‘‘Collaborative health care plan through crowdsource data using ambient application,’’ in Proc. 22nd Int. Multitopic Conf. (INMIC), Nov. 2019, pp. 1–6.
[10] J. G. V. Habets, M. Heijmans, A. F. G. Leentjens, C. J. P. Simons, Y. Temel, M. L. Kuijf, P. L. Kubben, and C. Herff, ‘‘A long-term, reallife parkinson monitoring database combining unscripted objective and subjective recordings,’’ Data, vol. 6, no. 2, p. 22, Feb. 2021



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)