International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning in Neurodegenerative Disease Prediction: Focus on Parkinson’s Disease- A Review.

Tanishka Ahirrao[1], Maheshwari Bhadane[2], Mitali Gaikwad[3], Asst. Prof. Mahesh Dhande[4]

[123]Student, AI&DS, Matoshri College of Engineering and Research Centre, Nashik.
[4]Asst. Professor, AI&DS, Matoshri College of Engineering and Research Centre, Nashik.
[1]ahirraotanishka14@gmail.com, [2]maheshwaribhadane714@gmail.com, [3]mitaligaikwad06@gmail.com, [4]mahesh.dhande@matoshri.edu.in

ABSTRACT –

Parkinson’s disease (PD) is one of the major public health diseases in the world which is progressively increasing day by day and had its effect on many countries. Thus, it is very important to predict it in early age which has been challenging task among researchers because the symptoms of disease come into existence in either middle or late middle age. So this Project focuses on the Spiral Test difficulty symptoms of PD affected people and formulates the model using various machine learning techniques such as adaptive boosting, RNN, Convolutional Deep Neural Networks, support vector machine, decision tree, Convolutional Neural Networks and linear regression. Performance of these classifiers is evaluated using various metrics, that is the accuracy, receiver operating characteristic curve (ROC), Sensitivity, precision, specificity. At last, Feature selection technique is used to find the most important features among all the feature to predict disease.

Keywords: Parkinson's Disease Prediction, Spiral Test Symptoms, Machine Learning Techniques, Convolutional Neural Networks (CNN), Feature Selection, Performance Evaluation Metrics

I. INTRODUCTION

Neurodegenerative disorders occur when neurons in different parts of the nervous system become damaged and die over time. Neurons are the brain's working units. They are separate but connected. A healthy neuron has parts called dendrites, axons, a cell body, and a nucleus that holds our DNA. DNA is our genetic blueprint, and each neuron contains it. When a neuron is unhealthy, it loses its connections and can't communicate well, which is harmful. Its metabolism slows, causing waste to build up, which it tries to store in small pockets. If things get worse, especially in lab settings, the neuron loses shape, becomes round, and fills with vacuoles.

This work looks at predicting Parkinson's disease, a rapidly growing, incurable disease. Parkinson's is widespread and named after James Parkinson, who first called it "shaking palsy." Later, it became known as PD, after his last name. It mostly affects neurons that control the body's movements. Key chemicals involved are dopamine and acetylcholine, which are crucial for brain function.

Several environmental factors contribute to PD development:

A. Environmental Factors: The environment is where we live and affects not just the human brain but all living things around. Research shows the environment plays a significant role in disorders like Alzheimer's and Parkinson’s. Important factors include:

a. Exposure to heavy metals (like lead and aluminum) and pesticides.

b. Air Quality: Pollution can cause breathing problems.

c. Water Quality: Contaminants in water lead to pollution.

d. Unhealthy Lifestyle: Can cause obesity and inactivity.

e. Psychological Stress: Increases stress hormones that harm neurons.

B. Brain Injuries or Biochemical Factors: The brain is the body's control center. Trauma can cause brain injuries, resulting in biochemical changes. These include enzymes that help stabilize neurons and maintain some chromosomes and genes.

C. Aging Factor: Aging contributes to Parkinson's development. In India, 11,747,102 people out of 1,065,070,607 are affected by it.

D. Genetic Factors: Genetics are a major cause of neurodegenerative disorders. The size, impact, and function of different genes determine the severity or level of these diseases, often increasing over time. Genetic factors related to these disorders are divided into pharmacodynamics and pharmacokinetics.

E. Spiral Test Factors: Due to rigidity and slowness (bradykinesia) in Parkinson's, changes in speech, drawing, and swallowing can occur:

i. Writing becomes breathy and softer.

ii. Speech may become slurred.

iii. Difficulty finding words makes speech slower.

II, LITERATURE REVIEW

Parkinson's Disease (PD) is a progressive neurodegenerative disorder that significantly affects motor functions. With advancements in machine learning (ML) and artificial intelligence (AI), numerous studies have explored automated approaches for early diagnosis and progression prediction of PD. This literature review summarizes notable research works that have applied diverse ML techniques to improve the accuracy, efficiency, and reliability of PD detection.

[1] **Arora, S., Venkataraman, S., Park, M. (2019) – "Using SVM for Early Detection of Parkinson's Disease from Speech Data."**

This study focused on using Support Vector Machines (SVMs) for early detection of Parkinson's Disease by analyzing speech data, proposed the use of Support Vector Machines (SVM) for early detection of Parkinson's disease using speech data. The researchers extracted vocal features such as jitter, shimmer, and pitch variation, which are known to be affected in Parkinson's patients. Their method showed effectiveness in distinguishing Parkinsonian speech patterns, emphasizing the utility of SVM in biomedical applications. The SVM model was trained to classify between healthy individuals and PD patients, achieving promising accuracy. This method showed the potential of speech as a non-invasive biomarker and highlighted SVM's effectiveness in binary classification tasks with biomedical datasets.

[2] **Li, H., Zhang, X., and Sun, Y. (2020) –**

"Multi-Biomarker Approach for Accurate Parkinson's Disease Diagnosis Using Random Forest and Neural Networks."

Introduced a multi-biomarker approach combining Random Forest and Neural Networks for accurate PD diagnosis. The study utilized EEG signals and demonstrated improved performance by using multiple data features. The authors proposed a multi-biomarker approach that combines Random Forest and Neural Networks for accurate PD diagnosis. EEG signals were used to extract relevant features, capturing both spatial and temporal brain activity patterns. Random Forests helped with feature selection and interpretation, while Neural Networks enhanced classification performance. This hybrid approach proved effective in capturing complex data relationships and provided a robust framework for disease detection.

[3] **Saravanan, S. et al. (2023) – "Explainable AI Model for Early Prediction of PD Based on Spiral and Wave Drawings."**

This research introduced an explainable artificial intelligence (XAI) model that predicts Parkinson's Disease using spiral and wave sketches drawn by patients. Designed an explainable AI model leveraging spiral and wave- based input patterns for early detection of Parkinson's disease, an explainable AI model for early PD prediction using spiral and wave drawings. It Combined biometric and clinical data with AI and deep learning models, demonstrating state-of-the-art performance. The model utilized deep learning techniques, particularly CNNs, to detect abnormalities in the patterns drawn—an effective method of identifying motor dysfunction. The system aimed to maintain high accuracy while also being interpretable, which is crucial for clinical adoption. The paper emphasized explain ability in AI, making it easier for healthcare professionals to trust and use such models in practice.

[4] **Srividhya, S. et al. (2023) – "PD Detection using Wearable Data."**

The study highlighted the effectiveness of capturing real-time data for symptom tracking and using it for automated detection with reliable accuracy. Focused on PD detection using wearable sensors, real-time detection of Parkinson's symptoms using wearable sensor data.

The system collected movement data such as gait and tremor information and applied ML models for classification. The study stressed the importance of continuous monitoring and highlighted the potential for integrating such systems into smart devices for remote healthcare. The use of wearable technology provided a scalable and user-friendly approach for PD detection.

[5] **Prashanth, R. et al. (2016) – "Parkinson's Disease Prediction using Machine Learning Approaches."**

Compared several ML algorithms for Parkinson's prediction, including SVMs and decision trees. Found ML-based methods to be superior in identifying PD-related patterns, emphasizing the importance of algorithm selection. Conducted a comparative analysis of multiple machine learning algorithms including Decision Trees, SVMs, and k-NN for Parkinson's Disease prediction. Their study revealed that machine learning algorithms could outperform traditional diagnostic methods in terms of accuracy and early detection. The researchers used biomedical voice measurements to train their models and found that algorithm selection plays a crucial role in optimizing prediction results.

[6] **K. Ang, Z. Zhang, and C. Guan (2020) – "Parkinson's Disease Detection and Classification Using EEG Based on Deep Learning."**

Applied deep learning models like CNN and LSTM on EEG data to classify Parkinson's disease. The work demonstrated high accuracy and emphasized the use of temporal features for better classification results.

Developed a method using deep learning algorithms such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to analyze EEG signals for Parkinson's classification. CNNs helped in learning spatial hierarchies from the EEG data, while LSTMs captured temporal dependencies. Their work demonstrated that deep learning can achieve high diagnostic accuracy by extracting subtle and complex patterns from brain signals, offering an advanced direction for neurological disorder diagnosis.

The reviewed studies clearly demonstrate that machine learning, particularly when integrated with diverse data types and wearable technology, can significantly enhance the accuracy and effectiveness of Parkinson's Disease diagnosis and monitoring. Future directions across these works include expanding datasets, integrating IoT for real-time monitoring, and employing more advanced and explainable AI techniques to facilitate clinical adoption and personalized care.

III. PROBLEM STATEMENT

Most of the studies reported in the literature survey focused on the usage of machine learning techniques like Logistic regression, Decision Tree, Support vector machine, Convolutional Neural Networks. Very few studies performed Adaptive boosting, RNN and Deep Neural Networks. The study evaluated and compared various machine learning techniques for the early prediction of disease. Our study is proposed with the aim to perform feature selection and to provide the comparative study of machine learning technique algorithms i.e. adaptive boosting, RNN, Deep Neural Networks, Support vector machine, Convolutional Neural Networks, Decision Tree. So our study will focus on finding the best model to provide an automated method to extract the necessary biomarkers which will help in the prediction disease.

IV. DEPENDENCY

The use of machine learning for predicting Parkinson's disease is greatly influenced by many aspects including the methods to ascertain the precision and dependability of these approaches. One of the problems is the availability and reliability of the data. Machine learning models require vast amounts of data inclusive of voice recordings, motor symptoms, gait analysis, case histories as well as clinical, demographic and biometric data in order to discern patterns suggestive of the ability to generalize to new instances. Additionally, feature selection also has a significant impact on the efficiency of the model. Machine learning models rely not only on identifying features but also on the extraction of relevant information from the data that contains various indicators such as tremor severity, motor function, speech deviations, and others, which assist in making the proper diagnosis. Incorrect features selection entails inadequate selection and cleansing approaches that causes the model to fail.

V. STRATEGY PLAN TO SOLVE PROBLEM

In order to address the issue of Parkinson's disease prediction with machine learning, the approach would be a methodical one. To begin with, a wide and inclusive dataset would have to be gathered, including clinical, biometric, and demographic information, along with non- invasive data like voice recordings, handwriting analysis, or gait patterns.

Once data quality has been ensured by preprocessing steps(missing data handling, normalization, etc.), feature selection methods would be used to determine the most appropriate predictors for Parkinson's disease.

Various machine learning algorithms, including support vector machines (SVM), random forests, or deep models, would be trained and calibrated to maximize prediction accuracy. Cross- validation and hyperparameter tuning would assist in improving the robustness of the model with no overfitting predictions.

In addition, the model would be assessed on various metrics (precision, recall, F1-score) to reduce false positives and negatives. Lastly, a pipeline for actual implementation, including ongoing monitoring and revision with additional data, would guarantee that the model continues to be effective and responsive to varying patient populations.

VI. DATASETS

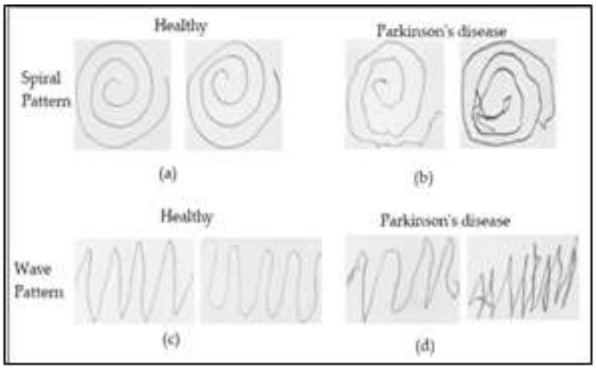

Fig 01. Spiral images samples.

The handwriting database for Parkinson's Disease (PD) includes data from 62 individuals with Parkinson's and 15 healthy controls. These participants visited the Neurology Department at Carapace Faculty of Medicine, Istanbul University, for this study. Researchers collected three types of handwriting samples from each participant: the Static Spiral Test (SST), the Dynamic Spiral Test (DST), and the Stability Test on Certain Point (STCP). Additionally, the dataset contains images of spiral drawings from those with Parkinson's, allowing the dataset to be used for different analyses like regression.

The public Parkinson Disease Spiral Drawings Using Digitized Graphics Tablet dataset was utilized, featuring drawings from 77 participants. Of these, 62 have Parkinson's, and 15 are healthy controls. Data was gathered using the Wacom Cintiq 12WX graphics tablet, a device that displays a computer screen and interacts with a digital pen. Each drawing session generated five sets of data, capturing details like X, Y, Z coordinates, pressure applied by the pen, and the angle at which the pen was held.

The dataset was created with the Wacom Cintiq 12WX tablet, which combines a graphics capability and an LCD screen. It allows users to interact directly with the screen using a digital pen. Specially-developed software was used to record the handwriting tests and check motor skills in people with Parkinson's. This software was coded in the C# language and runs on Windows systems. Those interested in the software can reach out to the researchers to request access.

The study conducted three specific tests using this tablet. The first is the Static Spiral Test (SST), widely used in research to evaluate motor function, detect tremors, and diagnose Parkinson's. During this test, participants trace three spirals displayed on the tablet using the digital pen, and their performance data is recorded.

The second test is called the Dynamic Spiral Test (DST). In this test, the spiral appears and disappears at intervals, causing it to "blink." This requires participants to remember the pattern and complete it from memory. The test assesses changes in drawing performance and measures pause times. Results show that while most participants continued trying to draw, nearly all had difficulty maintaining the pattern.

The third test is the Stability Test on Certain Point (STCP). This test involves a red dot at the center of the screen, where participants are asked to hover the digital pen over the dot without touching it for a specific duration. The goal is to measure hand stability and identify levels of hand tremor.

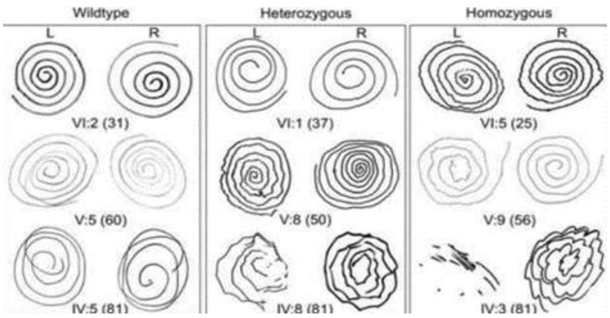

Fig 02: Sample dataset of biomedical drawing measurements.

VII. PROPOSED METHODOLOGY

The chapter focuses on the description of the dataset used and the methods used to enable early prediction of disease in a PD patient; the selected methods were meant to differentiate a patient with diseases from one who is healthy. The plan is to evaluate some performance metrics accuracy, ROC, AAE, and ARE by running various models on the chosen dataset and so finding the best machine learning technique among them by means of a comparative study of several machine learning technique.

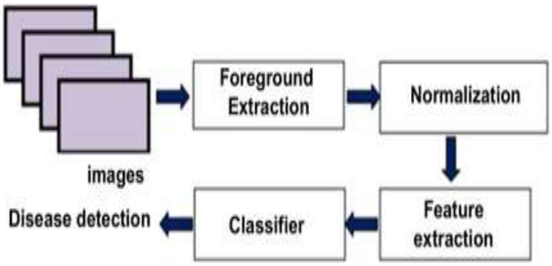

Fig 03: System Architecture

The process for predicting Parkinson's disease with spiral images using machine learning works by examining drawing trends. Patients are required to draw a spiral; then these photographs are analysis sets got in input data. Employing image processing methods, the system retrieves traits like tremors intensity, line irregularities, and fluctuations in forces.

Usually a convolutional neural network (CNN), a machine learning model is trained on known spiral pictures from healthy people and Parkinson's patients give data sets. The model starts to recognize subtle characteristics to differentiate between typical and Parkinsonian drawing patterns variations linked to motor handicaps.

Once trained, the system can predict whether a new spiral drawing shows early Parkinson's stage, helping with initial diagnosis and disease stage tracking.

ALGORITHMS

A. Deep Neural Network:

Deep Neural Networks (DNNs) are inspired by how our brain's neurons work and are widely used to make predictions. They start with input neurons that receive information. For example, in Parkinson’s disease detection, they use drawings of spirals. These inputs are given specific weights and processed through functions like the sigmoid to produce results. When many neurons are organized in layers, they form what is called a multilayer perceptron (MLP). This setup helps in recognizing complex patterns. For predicting Parkinson’s, DNNs examine spiral images to distinguish between healthy individuals and those with symptoms by identifying small inconsistencies and distortions caused by tremors in the drawings. This arrangement allows DNNs to understand complex relationships, but this also makes them harder to compute. This complexity can lead to issues like overfitting, which is when a model learns too much from the training data and doesn't work well on new data. This can be addressed by using methods like dropout, where certain parts of the network are ignored during training to prevent overfitting. DNNs are very powerful because they can learn complex details directly from raw data, making them promising tools for creating easy and non-invasive diagnostic methods.

B. Convolutional Neural Network:

Convolutional Neural Networks, known as CNNs, excel at sorting images, especially for tasks like analyzing spiral drawings to predict Parkinson's disease. These networks work by examining pictures through special layers called convolutional filters, which spot key details such as tremor patterns or unusual areas in sketches. After finding these features, the information goes through activation functions, such as ReLU, and pooling layers. These layers make the data smaller while keeping the important parts. Next, the data passes through flattening and fully connected layers, allowing the network to decide if a drawing came from a person with Parkinson's or someone healthy. CNNs learn using datasets that are labeled, and they continually adjust their settings to become more accurate. This type of network requires a lot of computing power because it has many layers and settings. However, this depth allows CNNs to automatically understand relationships and important features in images. Because of their strength in dealing with complex image data, CNNs are a powerful and effective tool for early and non-invasive detection of Parkinson's disease.

C. Linear Regression:

Linear regression is a straightforward method for predicting Parkinson's disease. It uses continuous features like age, symptom severity scores, or biomarkers. This method assumes there’s a straight-line relationship between these inputs and the result. It helps show how the disease might progress or estimate the risk score. By finding the best line, it minimizes the difference between predicted and actual values. Despite its simplicity, linear regression is useful for identifying basic trends and relationships in medical data. However, its effectiveness is limited because Parkinson’s symptoms are complex and not always linear. While linear regression is easy to use and understand, it may not always predict accurately by itself. It is better as a starting point or when used alongside more advanced models.

D. AdaBoost:

AdaBoost, which stands for Adaptive Boosting, is a method in machine learning designed to improve prediction accuracy by combining several simple models into one strong model. In the study of Parkinson’s disease, AdaBoost can be very useful as it helps to find small patterns in clinical and genetic data. The method works by focusing more on where the simple models make mistakes and then adjusting them to improve accuracy. Each simple model, also known as a weak learner, plays a role in the final decision based on how well it performs individually. The end result is a combination of these models, which helps the system make better distinctions between people who have the disease and those who are healthy. Though AdaBoost requires more computing power and careful adjustments, especially if the data is not balanced, it is excellent for scenarios that need high precision. Its biggest strength is learning from its mistakes and improving its predictions, making it a valuable tool for early detection of diseases like Parkinson’s.

E. Decision Tree:

Decision trees are a useful tool in medical diagnosis because they are easy to understand and can handle complex relationships. For predicting Parkinson’s disease, we look at factors such as age, UPDRS scores, and genetic markers to create the tree. The tree has decision points where choices are made based on these factors, and the final points tell us whether the person might have Parkinson’s.The tree operates by dividing data at each decision point. It picks the factor that gives the most useful information to make these divisions, leading to a decision about the disease. Although decision trees are straightforward, they can become complicated with too many layers and factors, which might cause problems, especially with messy clinical data. Despite potential complexity, decision trees remain helpful because they help identify important rules for making decisions in medical information.

F. Support Vector Machine:

Support Vector Machines (SVM) are excellent tools for deciding between two groups, especially in medical data analysis. SVMs work by identifying the best boundary to separate data points of different groups with the widest margin possible. In Parkinson’s disease prediction, we use features like voice measurements, tremor analysis, and movement scores. These features are transformed into a higher-level space, allowing SVMs to effectively identify who is healthy and who has Parkinson’s. This method is particularly useful for small to medium-sized datasets and can also manage complex patterns with the help of special functions called kernels. Although choosing the right kernels and adjusting settings are important for SVMs to work their best, they are very effective at handling complex data and providing accurate outcomes. This makes them crucial for early detection and monitoring of Parkinson’s disease.

G. Recurrent Neural Network:

Recurrent Neural Networks, known as RNNs, are computer models made to handle data that changes over time. This includes things like how someone walks or changes in their speech, which is useful in studying Parkinson’s disease. RNNs can remember past information thanks to loops in their design. This memory helps them notice patterns and changes that happen with time. Such an ability is important for tracking issues like muscle stiffness or voice tremors that gradually change. Although RNNs can be challenging to train and might face issues like vanishing gradients, improved versions like Long Short- Term Memory networks, or LSTMs, can address these challenges. LSTMs are strong at learning from changing sequences, which helps in predicting when and how a disease like Parkinson’s might develop. This knowledge allows doctors to improve and tailor treatments for patients.

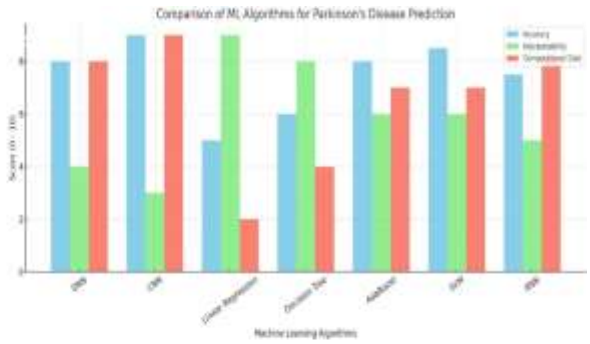

Fig 04. Graph showing comparison between algorithms

This graph compares how well different machine learning algorithms predict Parkinson's disease. It examines three key aspects: accuracy (how often they get the right result), interpretability (how easy it is to understand how they work), and computational cost (how much computer power is needed to run them).

IX. CONCLUSION

This proposed paper envisions a meaningful blend of technical exploration and real-world impact. The goal is to develop a system that can assist in detecting Parkinson's disease using a simple spiral drawing. The approach involves leveraging deep learning techniques— particularly convolutional neural networks (CNNs)—to train a model capable of distinguishing between healthy individuals and those with Parkinson's, based on patterns in their spiral sketches. This image-based method aims to provide a non-invasive and accessible tool for early screening, especially in regions with limited medical resources. The plan includes phases such as data collection and preprocessing, model design and training, and evaluation of accuracy and reliability. Each stage is expected to contribute to building a system that meaningfully integrates healthcare and artificial intelligence. The promise of this approach lies not only in its predictive accuracy but also in the simplicity of the input—a hand-drawn image. Ultimately, this project aspires to lay the groundwork for supportive diagnostic tools in the medical field. Future directions may include improving dataset diversity and integrating the model into a user- friendly application, thereby opening up new possibilities for tech-assisted healthcare.

Disease Detection and Classification Using EEG Based on Deep Learning.” This paper presents a methodology for classifying Parkinson's disease using EEG signals and machine learning algorithms like convolutional neural networks (CNN) and long short-term memory (LSTM) networks for time-series data analysis.

X. REFERENCES

[1] Arora, S. Venkataraman, and M. Park, "Using SVM for Early Detection of Parkinson's Disease from Speech Data, IEEE Transactions on Biomedical Engineering, vol. 66, no. 5, pp. 1390- 1395, 2019. DOI: 10.1109/TBME.2019.2892817.

[2] H. Li, X. Zhang, and Y. Sun, "Multi- Biomarker Approach for Accurate Parkinson's Disease Diagnosis Using Random Forest and Neural Networks," Journal of Neuroscience Methods, vol. 330, pp. 108-116, 2020. DOI: 10.1016/j.jneumeth.2020.108438.

[3] S. Saravanan, Kannan Ramkumar, K Narasimhan,Subramaniyaswamy Vairavasundaram, Ketan Kotech "Explainable Artificial Intellige Based on Spiral and Wave L Abraham (Senior Member IEEE), Detection of Parkinson's Disease xplainable artificial intelligence, 108% Parkinson's disease and deep learnmgvor 11, 2023, p10.1109/ACCESS.2023.3291406.

[4] Srividhya S., Sabarikrishnan S., Sandhiya S., Shajitha Begam S., "PD Detection using wearable data" PDdetection 10.7511/j.jneumth.2023.104858. vol 304,pp 104- 111,2023.DOI:

[5] S. Prashanth, H. D. Roy, S. Mandal, and R. Ghosh, "Parkinson's disease prediction using machine learning approaches," IEEE Journal of Biomedical and Health Informatics, vol. 20, no. 4, pp. 1088-1095, July 2016, doi: 10.1109/JBHI.2016.2521232.

[6] K. K. Ang, Z. Zhang, and C. Guan. 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). “Parkinson’s