Classification And Prediction Of Parkinson's Disease Using Machine Learning Algorithms

A.V.L. Prasuna[1], Sure Sree Charan Reddy[2], Toom Ritesh Reddy[3]

[1]*Assistant Professor, Mahatma Gandhi Institute of Technology*

[2,3]*UG Student, Mahatma Gandhi Institute of Technology*

Abstract- **Millions of people worldwide suffer from Parkinson's disease (PD), a degenerative neurological disorder. Better results and an efficient intervention depend on early and precise detection. The development of prediction models using machine learning methods like AdaBoost, Random Forest, and Decision Trees is the main goal of this study. Performance indicators such as accuracy, precision, recall, and F1-score were used to evaluate these models using datasets made up of spiral and wave drawings that were taken from motor skill assessments. Random Forest showed the best prediction performance and the highest dependability of all the models examined. These findings demonstrate how machine learning can be used to develop automated methods that aid in PD early diagnosis.**

Keywords: **AdaBoost, Random Forest, Decision Tree, Early Diagnosis, Machine Learning, Parkinson's Disease**

I. INTRODUCTION

Parkinson's disease is a progressive neurodegenerative disorder with very significant effects on the patients' quality of everyday life. It is defined by its motor symptoms: tremor, rigidity, and includes other complications such as cognitive impairment and sleep disorders. Early recognition and diagnosis are very important so that intervention can happen in a timely manner and to improve patient outcomes.

At present, PD diagnosis requires mostly subjective clinical observations and leads to inconsistent results as the reporting differs due to expertise differences. Recent developments in machine learning lend themselves to improving the diagnosis through developing methods that are automated and objective. One such promising avenue is by analyzing the patterns of drawing, such as spiral and wave designs that are often used for measuring motor functions. Those drawing patterns may then serve as inputs for machine learning models that learn to differentiate PD patients.

The study examines application of Decision Trees, Random Forest and AdaBoost algorithms in classifying patients as PD based on drawing data. These models will be evaluated for clinical use and possible augmentation to traditional diagnostic methodologies by accessing publicly available datasets. The aim is to provide an effective, scalable solution that can be adopted into pre-existing medical practices.

II. LITERATURE SURVEY

Latest innovations in machine learning (ML) and deep learning (DL) techniques depict great potential in the early identification as well as diagnosis of Parkinson's Disease (PD). Numerous approaches adopt different comparative views against various studies for model identification - finding the most promising effective methods. Models possess a wide range of capabilities regarding accuracy, reliability, and the inferences made for diagnosis. Random Forest and Support Vector Machines (SVM) have emerged as those models that generally improve performing models because they measure their higher accuracy and robustness against other techniques, unlike Naive Bayes and K-Nearest Neighbors, which gave low predictive reliability to worsen the picture of model selection necessary to best produce the diagnostic [1][5][8].

Handwriting analysis is another area of concern in PD. Hybrid fusion approaches based on offline images of handwriting can reach impressive results in diagnostic accuracy. For example, combining several visual features through pixel-level fusion and Laplacian transformation with the employment of pre-trained convolutional neural networks (CNNs) for feature extraction facilitates successful classifications using SVM [2]. Feature-level fusions that comprise handwriting and drawing activities improve diagnostic accuracy by sharing data across different architectures of a CNN-based feature extraction system [4].

Meta-learning frameworks have engaged cyber-physical systems to use contrastive learning and Siamese networks to scale better under the challenges posed by limited labeled data. This test on drawn datasets of spiral and wave pictures shows how well meta-learning performs regarding early detection and monitoring of PD [3].

III. PROPOSED SYSTEM

The proposed system for detecting Parkinson's disease relies on machine learning algorithms and image processing techniques to categorize medical images like spiral and wave drawings that fall into either the "Healthy" or "Parkinson" categories. First, it preprocesses images-resizes, normalizes, adds augmentation techniques to boost variability in the dataset. Feature extraction turns images into a numeric form compatible with machine learning through feature extraction. Several classifiers are then trained using the processed data. Such classifiers include Decision Tree, AdaBoost, and Random Forest. In assessing the performance of the models, accuracy, precision, recall, as well as F1-score are taken into consideration. The best performing model will be selected for onward use. Following the training and evaluation stages, this model would find use by the clinicians or researchers who would be able to input what new images they have to get predictions on Parkinson's disease. This will facilitate early detection and help in monitoring disease progression in order to achieve better patient outcomes. The model will also be frequently updated in order to keep it relevant at all times.

IV. SYSTEM ARCHITECTURE

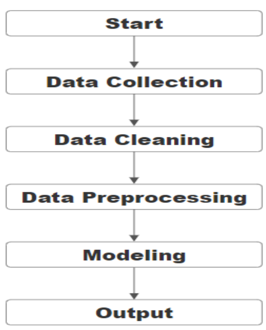


Figure1: System Architecture

Start: Initiates the project by defining the problem of detecting Parkinson's disease using medical images (e.g., spiral and wave drawings).

Data Collection: Gathers medical image data (e.g., spiral/wave drawings) related to Parkinson's disease, from trusted sources or public repositories, and organizes it for analysis.

Data Cleaning: Cleans the data by Removing duplicates and irrelevant data, handling missing values and inconsistencies, Filtering out any noise or outliers.

Data Preprocessing: Arrange the chosen data by cleaning, formatting, and extracting samples

Model Training: Decision Trees, Random Forests, and AdaBoost are trained using the data.

Model Evaluation: Performance is evaluated using metrics like accuracy, precision, recall, and F1-score.

V. MODULES

Data Collection Module: In this module, data for research on Parkinson's disease itself was collected, primarily medical images, such as spiral and waved drawings, from public datasets and medical repositories. All of this data represents the collection to be analyzed in future works, that will initialize the broad variety that is an essential part of training an effective model.

Data Cleaning Module: This module is designed to prepare a raw data set for further processing. It also assumes under this step the removal of duplicates and the irrelevant data and inconsistencies of the collected dataset itself. It also assumes the addressing of missing values and filtering out any noisy or outlier data points so that only clean, reliable data shall be used for training the model.

Data preprocessing modules: The term data preprocessing refers to the manipulation of raw data into the format needed by the machine learning package; for example: Normalization: This process rescales the data such that all features fall within a similar range, improving model performance. Feature Extraction: Some vital characteristics of images like textures, shapes, and patterns are extracted to help with identifying early signs of their development into Parkinson's Disease. Data Augmentation: The process artificially increases the dataset by application of transformations, such as rotation and flipping-the model generalizes better and reduces overfitting. Data Splitting: Divides the data set into training and testing to train the model on part and evaluate it on part.

Modelling module: Machine Learning is going to form a predictive model for the consideration of the Parkinson's disease. Training a Model: Algorithms based on the Decision Tree, Random Forest, and AdaBoost are now built on pre-processed data to learn from the feature extraction of medical images to see the different patterns identified with the disease.

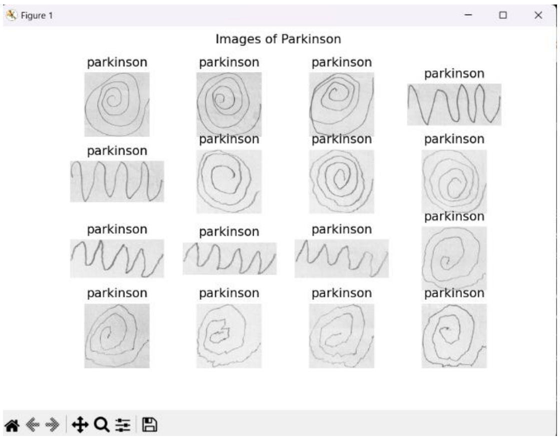
VI. ALGORITHMS

In order to augment the diagnostic precision towards the very early diagnosis of Parkinson's Disease (PD), several machine learning algorithms are deployed. The most commonly used algorithms for diagnostic purposes include these AdaBoost, Decision Trees and those Random Forest, each of which brings its unique strength in the entire detection process. AdaBoost is applied because it is good at merging weak classifiers into a strong classifier thus employing a better technique to increase accuracy in the model. It uses a technique of putting more attention on misclassified instances and assigns certain weights to the data such that it embraces robustness and prediction power of the system in extreme situations, especially in difficult or noisy datasets. Decision Trees are the other major algorithm among their simplicity and interpretability. They separate the dataset during the recursion of singularly splitting it along the feature values up until achieving a decision similar to those made by a tree. While the Decision Trees exhibit overfitting tendencies, the representation of complex relationships of the data makes them thus be useful in detecting PD. Random Forest is an ensemble method which builds several Decision Trees and averages the prediction by these trees. This makes Random Forest robust and generalized: as it will combine that magic when it comes to the individual trees and give a general opinion, thus averaging some errors of overfitting-the Random Forest is well fitted for larger and complicated databases, especially medical databases. Thus, the combination of such algorithms converges towards attaining high level classification accuracy, hence permitting more reliable and prompt detection of the Parkinson's Disease condition. This method enhances diagnostic precision through multi-algorithmic undertaking, contributing to better patient outcomes in path health because of being timely interventions.

VII. RESULTS

Figure 2: Virtualization of Healthy Images

Figure 3: Virtualization of Parkinson Images



The Random Forest, Decision Tree, and AdaBoost classifiers' performance in distinguishing between healthy and Parkinson images produced some insights worth discussing. The dataset is sufficiently visualized to understand the quality and distribution of the images-it shows grayscale representations of both. This visualization would serve to identify the challenge the models have to undergo toward Tackling Changes in image features.

Figure 4: Algorithms vs Precision

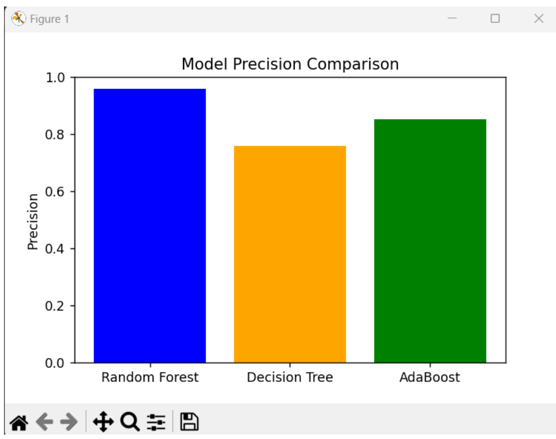
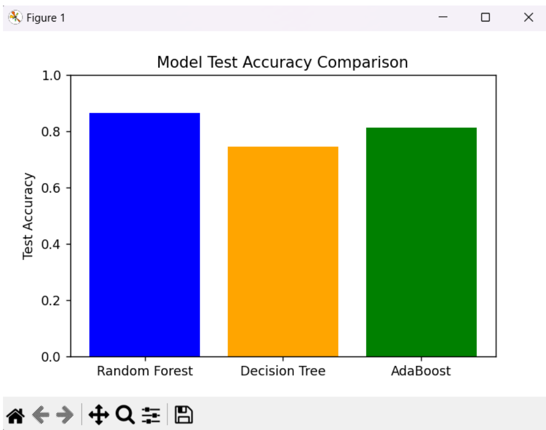


Figure 5: Algorithms vs Recall

Figure 6: Algorithms vs F1 Score

Figure 7: Algorithms vs Accuracy



Their evaluation was based on four important measures; accuracy, precision, recall, and F1 score. Random Forest, among the models, was found to perform well overall against all four metrics, thus proving the model to be capable of learning complex patterns even from high-dimensional data. Decision Tree, a simpler model, performed slightly less at accuracy and F1 scores, which could either mean over-fitting or that it may be sensitive to the relative size of the dataset on which it has been trained. Though AdaBoost displayed competitive precision, it had low recall which indicates it could be quite sensitive to noise in the dataset.

Figure 8: Model Comparison

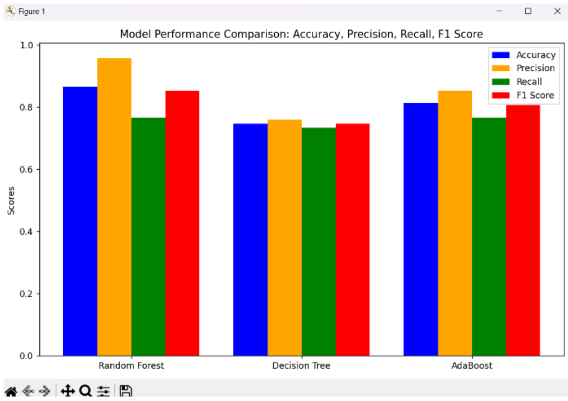
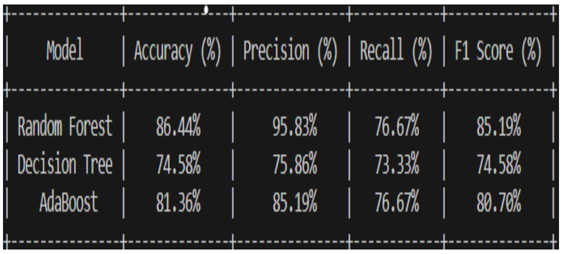


Figure 9: Value Analysis



Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Random Forest	86.44%	95.83%	76.67%	85.19%
Decision Tree	74.58%	75.86%	73.33%	74.58%
AdaBoost	81.36%	85.19%	76.67%	80.70%

In fact, all the performance differences in the classifiers were brought out in bar charts and comparison plots. Random Forest had the highest accuracy, while AdaBoost was particularly remarkable for its precision. The Decision Tree also scores steadily, but surprisely irrelevant when it comes to recall and F1 score. There is quite a clear trade-off among the metrics, with one model excelling in terms of some but not being able to perform well on others.

VIII. CONCLUSION

The analysis of the three algorithms in the initial diagnosis of Parkinson's disease from spiral and wave images-Namely Random Forest, AdaBoost, and Decision Tree-proved to generate some favorable insights on their efficacy. Random Forest gave the highest test accuracy of 86.44%, probably showing that the algorithm is very powerful in working on a complex dataset and does not cause overfitting since it uses ensemble learning. And so, this turns out to be the best efficient in this model, the most reliable. AdaBoost, on the other hand, achieved test accuracy under 81.36%, which is pretty high in reducing misclassifications after several iterations, but is still less efficient than Random Forest under this circumstance. Decision tree-the one with the least accuracy of 74.58 percent-would overfit when used singly, thus stressing on the need for boosting or bagging methods to better its overall performance. Random Forest emerged as the most reliable model having high accuracy with robustness. Excellent in giving fingerprints feature and patterns with Parkinson disease prospective predictive. Future works can take further steps towards the involvement of deep learning for an even more accurate and rapid diagnosis.

IX. FUTURE SCOPE

In future, it is intended to extend this functionality towards deep learning models like Convolutional Neural Networks for improving image-based feature extraction and accuracy in the detection of Parkinson's disease. Techniques such as image flipping or rotation will also be employed for data augmentation so that the overall diversity of the dataset increases and improves generalization and performance on unseen data. Such hybrid models additionally combine machine learning skills with current and forthcoming deep learning for a better feature selection through Random Forest along with previous classification in CNN for accurate prediction. The aforementioned proposed future improvements will make the entire fast application, more accurate, and robust with great capability in real-world clinical potential and also towards early diagnosis of advancement in the case of Parkinson's disease.

REFERENCE

[1] S. Sharma, K. Guleria and R. Sobti, "An Early Detection of Parkinson’s Disease using Machine Learning and Deep Learning Models," 2024 International Conference on Computational Intelligence and Computing Applications (ICCICA), Samalkha, India, 2024, pp. 154-159, doi: 10.1109

[2] S. Dong, J. Liu and J. Wang, "Diagnosis of Parkinson's Disease Based on Hybrid Fusion Approach of Offline Handwriting Images," in IEEE Signal Processing Letters, vol. 31, pp. 3179-3183, 2024, doi: 10.1109

[3] M. S. Hossain and M. Shorfuzzaman, "MetaParkinson: A Cyber-Physical Deep Meta-Learning Framework for n-Shot Diagnosis and Monitoring of Parkinson's Patients," in IEEE Systems Journal, vol. 17, no. 4, pp. 5251-5260, Dec. 2023, doi: 10.1109

[4] S. Naz, I. Kamran, S. Gul, F. Hadi and F. Khalifa, "Multi-Model Fusion of CNNs for Identification of Parkinson’s Disease Using Handwritten Samples," in IEEE Access, vol. 11, pp. 135600-135608, 2023, doi: 10.1109

[5] S. M. Abdullah et al., "Deep Transfer Learning Based Parkinson’s Disease Detection Using Optimized Feature Selection," in IEEE Access, vol. 11, pp. 3511-3524, 2023, doi: 10.1109

[6] J. A. Nolazco-Flores, M. Faundez-Zanuy, V. M. De La Cueva and J. Mekyska, "Exploiting Spectral and Cepstral Handwriting Features on Diagnosing Parkinson’s Disease," in IEEE Access, vol. 9, pp. 141599-141610, 2021, doi: 10.1109

[7] J. P. Devarajan, V. R. Sreedharan and G. Narayanamurthy, "Decision Making in Health Care Diagnosis: Evidence From Parkinson's Disease Via Hybrid Machine Learning," in IEEE Transactions on Engineering Management, vol. 70, no. 8, pp. 2719-2731, Aug. 2023, doi: 10.1109