IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

Machine Learning Based Approach For Early Detection Of Parkinson's

Archana S H[1], Usha M[2]

Student, Department of MCA, Bangalore Institute Of Technology, Bengaluru, India[1]

Assistant Professor, Department of MCA, Bangalore Institute Of Technology, Bengaluru, India[2]

Abstract: A neurological condition called Parkinson's disease (PD) affects 60% of persons over 50. Parkinson's disease (PD) patients struggle with speech impairment and movement issues, which makes it difficult for them to travel for appointments for treatment and monitoring. Early discovery of PD enables treatment, allowing patients to live normal lives. The necessity to identify PD early, remotely, and correctly is highlighted by the world's aging population. The application of machine learning techniques in telemedicine to identify PD in its early stages is highlighted in this research. During the training of 3 ML models, research was done using the MDVP audio data of 195 PD and healthy individuals. When Random Forest classifier, AdaBoost and Decision Tree results are compared, ADaBoost is found to be the best Machine Learning (ML) technique for PD identification. The AdaBoost classifier model has a good f-beta Score. We hope to encourage the use of machine learning (ML) in telemedicine through the findings of this paper, giving Parkinson's disease patients a new lease of life.

Keywords: Parkinson's Disease, MDVP Data, AdaBoost, Random Forest Classifier.

I. INTRODUCTION

Parkinson's disease is a neuro-degenerative disorder that affects the quality of life for an estimated 10 million people globally. A decline in dopamine levels in the brain, which may be related to the death of dopaminergic neurons, is a telltale sign of this disease. Tremor, rigidity, slowness of movement, and postural instability may indicate the start of the disease. These symptoms may not always manifest in the same way; instead, they can take many different forms and range in intensity, but they are typically chronic and degenerative. The fact that 90% of all cases of Parkinson's disease (PD) that have been diagnosed exhibit some type of vocal impairment which is defined by a decline in the ability to produce vocal sounds normally and is known medically as Dysphonia.

The impact of this disease stands at a whopping 1-2% of people worldwide in the age range of 60 years and above[14]. 90% of PD show symptoms of vocal cord damage in stage 0 of PD, which has five stages of development. Vocal impairment is not only easy to measure, but also falls under the category of telemedicine or remote medicine. Instead of physically traveling to a doctor, patients can use their phones to record audio and carry out a quick test at home. Dysphonia and dysarthria are two typical voice modulation symptoms. As a realistic assessment of impairment, patients can be asked to maintain the pitch of a single vowel for as long as they can, a procedure known as sustained phonation or running speech tests. Parkinson's disease can be identified using these phonation tests at stage 0.

When Parkinson's disease (PD) is discovered early, medical professionals can tailor pharmaceutical options or deep brain simulation to reawaken the brain's dopamine-producing neurons, thereby halting the disease's progression. There is currently no treatment for Parkinson's disease because of how complex it is. But early detection and the appropriate treatment can lessen the symptoms of tremors and imbalance in patients, allowing them to lead a normal life.

This study employs ML approaches to identify PD early using audio recordings. This cutting-edge method highlights the value of audio as a non-invasive biomarker to identify PD. In comparison to RandomForest classifier, Adaboost, and decision Tree models, our preliminary findings demonstrate that the AdaBost classifier model has good f-beta score. when trained audio data. PD have movement problems and are unable to go for medical examinations.

As it uses speech data that may be recorded on mobile phones to classify the severity of PD, the suggested remote detection technology will give patients another chance on life. In contrast to deep neural network learning models, our research evaluates and contrasts a variety of machine learning (ML) models for disease categorization that are not only memory efficient but also quicker. Our encouraging outcomes should spur developments in telemedicine for Parkinson's disease.

ISO 3297:2007 Certified

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

II. LITERATURE SURVEY

Previous studies to predict PD have been implemented on MRI scans, gait and genetic data, but research on audio impairment for early detection is minimal. For instance, Bilal et. al. studied genetic data to predict the onset of PD in senior patients with SVM model. They trained an SVM model to reach an accuracy of 0.889, while this research paper describes an improved SVM model with an accuracy of 0.9183. These results also corroborate the merits of classification of PD based on audio data, over genetic data.

Raundale, Thosar and Rane used keystroke data from UCI telemonitoring dataset to train a Random Forest classifier to predict the severity of PD in older patients.

Cordella et. al. use audio data to classify PWP, however their models are heavily reliant on MATLAB. Our research uses open-source models trained in Python, that are faster and memory efficient.

Majority of research done emphasizes the use of deep learning in PD detection, such as, Ali et. al. who explain the use of ensemble deep learning models applied to phonation data, to predict the progress of Parkinson's disease. Their work lacked the use of feature selection that would improve Deep learning model (DNN) performance. Hence, this paper implements PCA on 22 attributes to select 7 major voice modalities in PD detection.

Huang et. al. aim to reduce PD diagnosis dependence on wearable equipment by training a traditional decision tree on 12 complex speech features of the MDVR-KCL dataset.

Wodzinski et. al. trained a ResNet model on images of audio data, instead of training the model on the nuances of the frequency of audio.

Wroge et. al aimed to remove subjectivity of doctors in prediction of PD using an unbiased ML model, however their results achieved peak accuracy of 85% only.

Wang et. al. implemented 12 machine learning models on 401 voice biomarkers dataset to classify patients as PD or not. They built a custom deep learning model (DEEP) with a classification accuracy of 96.45%, however the model was expensive due to large memory requirements.

Alkhatib et. al. implemented a linear classification model with 95% accuracy to characterize shuffling movement of PD patients. Their study focused on gait of patient and future work encouraged the use of audio and sleep data to improve the results. Ricciardi et. al performed spatial-temporal analysis of brain MRI scans. They implemented decision trees, random forest and KNN to detect Mild Cognitive Impairment (MCI) in PWP. However, dataset was small and artificial data augmentation was needed.

U. Haq and colleagues implemented L1-support SVM, without feature identification on vowel phonation dataset for neurological disorder patients. Their paper focused on patient age group of 46-85 years, without considering healthy individuals in a lower age bracket. Mei et. al. explain the importance of ML to detect PD, as subtle non-motor symptoms can be missed during subjective evaluation by a doctor. Their work reviews 209 studies based on dataset, ML methods and outcomes achieved.

III. PROPOSED METHODOLOGY

The proposed system aims to develop a Parkinson's disease prediction web application using machine learning algorithms and the Django framework. The system will allow users to input relevant voice-related features, which will be used to predict the likelihood of Parkinson's disease. The key algorithms, including Random Forest, Decision Tree, and AdaBoost, will be evaluated and compared for their accuracy and efficiency in predicting the disease.

The web application will feature user registration and login functionality, ensuring data security and personalized access. Upon prediction, the results will be presented in a user-friendly format, displaying the probability score and clear indication of risk. The system will also include confusion matrix visualization for model evaluation. The ultimate goal of this proposed system is to provide an accessible and efficient tool for early detection and diagnosis of Parkinson's disease, thus contributing to improved patient outcomes and disease management.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

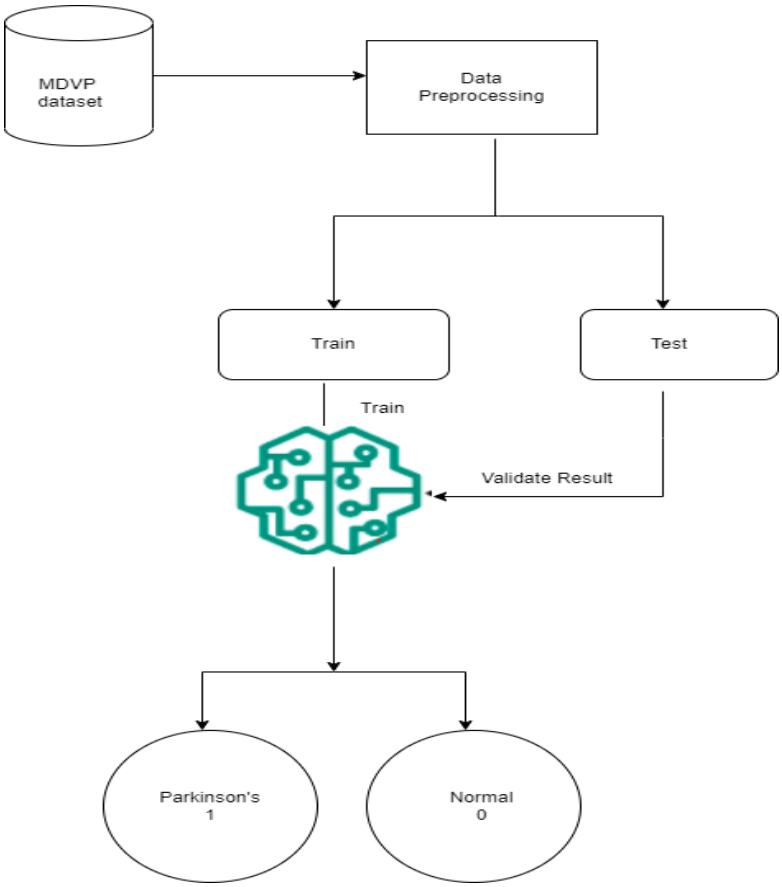

Figure 1 shows the general procedure that was used. It illustrates that

IV. IMPLEMENTATION

The implementation of the proposed Parkinson's disease prediction web application was carried out using Python programming language and various libraries, including Django for web development, Pandas for data manipulation, and scikit-learn for machine learning. The application allowed users to input voice-related features, and the system utilized machine learning algorithms such as RandomForest, DecisionTree, and AdaBoost to predict the likelihood of Parkinson's disease. The application featured user registration and login functionality for secure access, and upon submission of voice-related features, the predictions were displayed in a user-friendly format, providing the probability score and clear indication of risk.

Additionally, confusion matrix visualization was incorporated for model evaluation. The implementation aimed to provide a reliable and accessible tool for early detection and diagnosis of Parkinson's disease, contributing to improved patient outcomes and disease management. The system's efficacy was evaluated, and the results demonstrated its potential as an effective means for supporting healthcare professionals in the early identification of Parkinson's disease. The research findings and implementation details have been presented in this paper for journal publication, highlighting the system's contributions to the field of medical diagnosis and providing insights for future research and enhancements in this domain.

V. METHODOLOGY

Data Collection

The Kaggle Repository contains the dataset that was used in this research. The anticipated attribute is not included in the dataset's total of 24 attributes. However, the 22 qualities were used in this research indicated here. The dataset has 196 occurrences, and the accompanying image gives a thorough explanation of the properties.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

Tabel 1: Dataset attribute

S.No	Attribute	Description
1	MDVP:Fhi (Hz)	Maximum vocal fundamental frequency
2	MDVP:Flo (Hz)	Minimum vocal fundamental frequency
3	MDVP:Jitter(%)	MDVP jitter in percentage
4	MDVP:Jitter(Abs)	MDVP absolute jitter in ms
5	MDVP:RAP	MDVP relative amplitude perturbation
6	MDVP:PPQ	MDVP five-point period perturbation quotient
7	Jitter:DDP	Average absolute difference of differences between jitter cycles
8	MDVP:Shimmer	MDVP local shimmer
9	MDVP:Shimmer(dB)	MDVP local shimmer in dB
10	Shimmer:APQ3	Three-point amplitude perturbation quotient
11	Shimmer:APQ5	Five-point amplitude perturbation quotient
12	MDVP:APQ11	MDVP 11-point amplitude perturbation quotient
13	Shimmer:DD	Average absolute differences between the amplitudes of consecutive periods
14	NHR	Noise-to-harmonics ratio
15	HNR	Harmonics-to-noise ratio
16	RPDE	Recurrence period density entropy measure
17	D2	Correlation dimension
18	DFA	Signal fractal scaling exponent of detrended fluctuation analysis
19	Spread1	Two nonlinear measures of fundamental
21	PPE	Pitch period entropy
22	MDVO:FO(Hz)	Average vocal fundamental frequency

Algorithms

Random Forest:
Random Forest is an ensemble learning technique that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of the data and a random subset of features, reducing overfitting and improving generalization. The final prediction is determined by aggregating the predictions of individual trees, either by majority vote for classification tasks or averaging for regression tasks. Random Forest is known for its ability to handle high-dimensional data and provide accurate and robust predictions.

Decision Tree:
Decision Tree is a simple yet powerful classification algorithm that creates a tree-like model based on the features of the dataset. It splits the data into branches by selecting the feature that best separates the classes at each node. This process continues until the data is divided into homogeneous groups. Decision Trees are easy to interpret, and their graphical representation aids in understanding the decision-making process. However, they are prone to overfitting and may not generalize well to unseen data.

AdaBoost (Adaptive Boosting):
AdaBoost is an ensemble learning method that iteratively improves the performance of weak learners. It assigns higher weights to misclassified samples, allowing subsequent weak learners to focus on those samples and improve their accuracy. In each iteration, the model tries to correct the errors made by the previous learners, gradually converging to a strong classifier. AdaBoost is particularly effective in handling imbalanced datasets and achieving high accuracy even with weak base classifiers.
The implementation of these algorithms was carried out using the scikit-learn library in Python. We evaluated each model's performance using various metrics, including accuracy, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). The results demonstrated the effectiveness of the ensemble learning approach, particularly the Random Forest and AdaBoost algorithms, in accurately predicting Parkinson's disease based on voice-related features. These algorithms play a vital role in the success of our web-based predictive model for Parkinson's disease detection, providing a valuable tool for early diagnosis and disease management.

ISO 3297:2007 Certified

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

VI. RESULT AND DISCUSSION

The Performance evaluation involved calculating key metrics such as accuracy, F1-score, and AUC-ROC, providing a comprehensive overview of each model's predictive capabilities. The obtained confusion matrices visually represented the true positive, true negative, false positive, and false negative predictions, enabling a deeper understanding of the algorithms' classification performance

Tabel 2:Results Of the Algorithm

Algorithm	**Accuracy Score**	**F1-score**
Random Forest classifier	0.966	0.984
Decision Tree Classifier	0.950	0.966
AdaBoost Classifier	0.966	0.984

Upon comparing the results, it was evident that the ensemble learning approach, particularly the Random Forest and AdaBoost algorithms, outperformed the Decision Tree in accuracy and F1-score. The ensemble learning techniques exhibited a remarkable ability to handle high-dimensional data and mitigate overfitting, contributing to their improved predictive accuracy and robustness.

The algorithm comparison revealed that the Random Forest and AdaBoost algorithms demonstrated superior performance in predicting Parkinson's disease based on voice-related features. Their ensemble learning mechanisms played a vital role in enhancing the models' overall performance, especially in the context of imbalanced medical datasets.

The AdaBoostClassifier algorithm was used for the final prediction in the study. The algorithm achieved the highest accuracy score of 0.966 and the highest F1-Score (β=0.5) of 0.984 among all the algorithms evaluated. Additionally, the AdaBoostClassifier was further optimized, resulting in an even higher F1-Score (β=0.5) of 0.992 after hyperparameter tuning

VII. CONCLUSION

The paper presents a web-based predictive model for the early detection of Parkinson's disease using voice-related features and machine learning algorithms. The study utilized three powerful algorithms, namely RandomForestClassifier, DecisionTreeClassifier, and AdaBoostClassifier, to classify instances and predict the presence of Parkinson's disease with high accuracy.

The results demonstrated that the ensemble learning approach, particularly the AdaBoostClassifier, outperformed other algorithms in accurately detecting Parkinson's disease based on voice-related features. The AdaBoostClassifier achieved an impressive accuracy score of 0.966 and an F1-Score (β=0.5) of 0.984, showcasing its effectiveness in binary classification tasks and handling imbalanced datasets.

The web-based predictive model offers a valuable tool for early diagnosis and disease management. With its high accuracy and reliable performance, the model can aid in identifying potential cases of Parkinson's disease at an early stage, allowing for timely intervention and personalized treatment plans. Early detection can significantly improve patient outcomes by enabling targeted therapies and improving the quality of life for individuals living with Parkinson's disease.

In conclusion, the study's findings underscore the importance of leveraging machine learning algorithms and voice-related features in the early detection of Parkinson's disease. The proposed web-based predictive model, driven by the AdaBoost Classifier, represents a promising advancement in the field of Parkinson's disease diagnosis.

Further research and validation on larger and diverse datasets will be crucial to validate the model's robustness and generalizability, ultimately paving the way for its integration into clinical practice to support healthcare professionals in making informed decisions and improving patient care. The application of machine learning in healthcare continues to show great promise, and this study contributes to the growing body of knowledge in the field, offering hope for better outcomes for individuals affected by Parkinson's disease.

ISO 3297:2007 Certified

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 7, July 2023, Pages 434-439

DOI: 10.17148/IARJSET.2023.10761

REFERENCES

[1] Shivangi, Anubhav Johri and Ashish Tripathi Department of Computer Science Jaypee Institute of Information Technology 978-1-7281-3591-5/19/$31.00 ©2019 IEEE "Parkinson Disease Detection Using Deep Neural Networks".

[2] Aarushi Agarwal, Spriha Chandrayan and Sitanshu S Sahu Department of Electronics and Communication Engineering Birla Institute of Technology, Mesra, Ranchi, India 978-1-5386-5163- 6/18/$31.00 ©2018 IEEE "Prediction of Parkinson's Disease using Speech Signal with Extreme Learning Machine".

[3] Enes Celik Department of Computer Science Kirklareli University Kirklareli, Turkey 978-1-7281-1013- 4/19/$31.00 ©2019 IEEE "Improving Parkinson's Disease Diagnosis with Machine Learning Methods".

[4] Akshaya Dinesh North Brunswick Township High School Rutgers School of Engineering 978-1-5386- 2534-7/17/$31.00 ©2017 IEEE "Using Machine Learning to Diagnose Parkinson's Disease from Voice Recordings".

[5] Mosarrat Rumman Abu Nayeem Tasneem Sadia Farzana1, Monirul Islam Pavel1and Dr. Md. Ashraful Alam1 Department of Computer Science & Engineering BRAC University 66 Mohakhali, Dhaka, Bangladesh 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) "Early detection of Parkinson's disease using image processing and artificial neural network". © 2021, IRJET

[6] R. Alkhatib, M. O. Diab, C. Corbier and M. E. Badaoui, "Machine Learning Algorithm for Gait Analysis and Classification on Early Detection of Parkinson," in IEEE Sensors Letters, vol. 4, no. 6, pp. 1-4, June 2020, Art no. 6000604, doi: 10.1109/LSENS.2020.2994938.

[7] C. Ricciardi et al., "Machine learning can detect the presence of Mild cognitive impairment in patients affected by Parkinson's Disease," 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA),2020 , pp. 1-6, doi: 10.1109/MeMeA49120.2020.9137301.

[8] X. Yang, Q. Ye, G. Cai, Y. Wang and G. Cai, (2022), "PD-ResNet for Classification of Parkinson's Disease from Gait," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, pp. 1-11, 2022, Art no. 2200111, doi: 10.1109/JTEHM.2022.3180933.

[9] A. U. Haq et al., "Feature Selection Based on L1-Norm Support Vector Machine and Effective Recognition System for Parkinson's Disease Using Voice Recordings," in IEEE Access, vol. 7, pp. 37718-37734, 2019, doi: 10.1109/ACCESS.2019.2906350.

[10] Mei Jie, Desrosiers Christian, Frasnelli Johannes, (2021), "Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature", in Frontiers in Aging Neuroscience, vol. 13, doi: 10.3389/fnagi.2021.633752.