THE PARKINSON'S DISEASE DETECTION BASED ON VOICE RECOGNITION USING MACHINE LEARNING

Aradhana Gupta[1], Avantika Paidalwar [2], Harshita Wasnik[3], Rushika Gajbhiye[4]
U.G. Student, Department of Information Technology, JD College of Engineering & Management, Nagpur, Maharashtra, India
Prof.Mirza Moiz Baig.[5]
Ass Prof, Department of Information Technology, JD College of Engineering & Management, Nagpur, Maharashtra, India

ABSTRACT:

In this project, we utilise machine learning techniques to develop accurate and early detection models for Parkinson's disease. The dataset, comprising clinical features and patient data, is preprocessed and split into training and testing sets. Support Vector Machines (SVM) and XGBoost algorithms are employed, optimised, and evaluated, with a focus on key performance metrics. Feature importance analysis aids in identifying influential factors, offering valuable insights for healthcare professionals. Ultimately, these models have the potential to significantly enhance early diagnosis and treatment of Parkinson's disease, thereby improving patient outcomes and quality of life. the models are trained and fine-tuned using techniques such as hyperparameter tuning and cross-validation to ensure robust performance. Performance metrics, including accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrices, are used to evaluate and compare the models. Additionally, feature importance analysis is conducted to identify the most influential factors in Parkinson's disease detection. Visualisations of feature importances are generated, aiding in the interpretation of model decisions.

KEYWORD: -Parkinson's disease, machine learning, early detection, Support Vector Machines (SVM), XGBoost, clinical features, data preprocessing.

I. INTRODUCTION

Parkinson's disease (PD) is a complex and progressive neurodegenerative disorder that affects millions of individuals globally, with a prevalence that increases with age. First described by James Parkinson in 1817, it is characterised by a wide spectrum of motor and non-motor symptoms. The cardinal motor symptoms include resting tremors, bradykinesia (slowness of movement), rigidity, and postural instability. Non-motor symptoms encompass cognitive impairment, depression, anxiety, sleep disturbances, and autonomic dysfunction, making PD a multi-faceted condition with a substantial impact on a patient's quality of life. The aetiology of Parkinson's disease is multifactorial, with a combination of genetic and environmental factors playing a role in its development. While substantial progress has been made in understanding the pathophysiological mechanisms involved in PD, early diagnosis remains a significant challenge. Currently, diagnosis primarily relies on clinical evaluation by neurologists, often based on the presence of motor symptoms and medical history. However, these symptoms typically manifest in the later stages of the disease when substantial damage to the brain's dopaminergic neurons has already occurred. Early diagnosis is critical as it enables the timely initiation of appropriate treatment and interventions, potentially slowing down the disease's progression and improving the patient's quality of life. Therefore, there is a pressing need for more accurate and efficient methods of early PD detection. This is where machine learning (ML) comes into playMachine learning, a subset of artificial intelligence (AI), is gaining prominence in the field of healthcare, including the diagnosis and management of neurological disorders like Parkinson's disease. ML algorithms have the ability to analyse complex datasets, identify patterns, and make predictions based on data inputs. In the context of PD, ML offers the potential to detect subtle disease-related changes in data, such as clinical assessments, patient histories, and even neuroimaging, that might not be apparent to the human eye. One of the primary goals of ML in PD detection is to develop accurate, data-driven diagnostic models that can recognize the disease in its early stages. This can be accomplished by feeding the algorithms with relevant features extracted from clinical data and leveraging powerful techniques for feature selection and model optimization. Two commonly used algorithms for this purpose are Support Vector Machines (SVM) and XGBoost.

II. LITERATURE SURVEY

Shrihari K Kulkarni, K R Sumana, the researchers use Decision Tree, Logistic Regression, and Naive Bayes, Deep Learning algorithms like Recurrent Neural Networks (RNN) by predicting the Performance Parameters to build the model. Machine learning approaches will be used to construct prediction models that can differentiate early PDfrom healthy normal using the Movement Disorder Society-Unified Parkinson's Disease Rating Scale (MDS UPDRS). For Subject and Record Validation, Logistic Regression, Random Forests, and Support Vector Machine were employed. Drawback of this paper were, Data Collection techniques are weakly regulated, resulting in unreliable results such as out-of-range or non-existent This Model purely relies on Evaluation of Motions which is not the only source of Data available on the Disease-bearers or Healthy Citizens. Yatharth Nakul, Ankit Gupta, Hritik Sachdeva, the researchers in used Supervised Learning Algorithms such as Random Forest, Support Vector and Naïve-Bayes are also compared. Confusion matrix was used for accuracy checking and different Classification methods were used.ML classification technique will improve the accuracy and reduce possible loopholes. Hyper parameter tuning is used to achieve the maximum accuracy. Achieved maximum accuracy of 98.30% using the K nearest neighbour classification The main drawbacks are Delay in Results derived and Output Progression is slow and Best Proposed Methodology used gives Higher error rate when Confusion Matrix is plotted.

SGD (Stochastic Gradient Descent) is utilized for training data models, according to Wu Wang, Junho Lee, Fouzi Harrou, and Ying Sun of . The FNN (Feed-Forward Neural Network) is put into action. The sensitivity of the linear discriminate analysis approach utilized is the best, which means it has the best likelihood of distinguishing a real patient. The proposed deep learning model had a 96.45% accuracy rate. This is owing to the deep learning model's favorable capabilities in learning linear and nonlinear features from PD data without the requirement for hand-crafted feature extraction. The biggest drawback is that Deep Learning is frequently employed as a Blackbox algorithm, the trained neural networks are difficult to evaluate. Theoretically, it's difficult to comprehend how deep learning generates good Results. Support Vector Machine (SVM), Feedforward BackPropagation Based Artificial Neural Network (FBANN) And Random Tree (RT) Classifiers, Binary Logistic Regression, Linear Discriminant Analysis (LDA), Convolutional Neural Network (CNN)Deep Belief Network

(DBN) Technique Deep Neural Network Classifiers were used by Muthumanickam S, Gayathri J, Eunice Daphne J, the researchers in . It has a higher level of accuracy than a deep neural network. Linear regression is simple to comprehend. It can be tweaked to prevent overfitting. The sgd command can be used to update linear models. The Algorithm and the Outputs of Binary Logistic Regression Have a Good Interpretation.Jayashree R. J, Ganesh S, Karanth S.C, Lalitha S has deeply explained how spectral features ex. Spectral Contrast, STFT and temporal features ex. Zero Crossing rate are extracted and classification done using XGBoost and Classifiers including Random Forest and Regression like Logistic Regression. It has Several Advantages including the Real Time Speech Analysis which perfectly shows how noise and other factors affect Parkinson's Disease Detection. It depicts better Accuracy and a different Characteristic approach on proving how noise affects in the Prediction of Parkinson's Disease. Its biggest drawback is it is based on a limited dataset, if more data was available, a more practical approach could be designed. Limited Classifiers are used and Analysis of Features are constrained to just eight features which is very limited comparatively.

III. PROBLEM DEFINITION

The problem addressed in this project is the development of accurate and efficient methods for early Parkinson's disease detection using machine learning. Parkinson's disease, a neurodegenerative disorder with a wide range of motor and non-motor symptoms, poses a significant diagnostic challenge, particularly in its early stages. The project's objective is to leverage machine learning techniques to create predictive models capable of identifying Parkinson's disease in its early phases, before prominent motor symptoms emerge. This encompasses data collection, preprocessing, model selection, training, evaluation, and ethical considerations, with the goal of producing accurate and interpretable models that can potentially enhance early diagnosis and patient care, thereby improving the quality of life for those affected by Parkinson's disease.

IV. METHODOLOGY

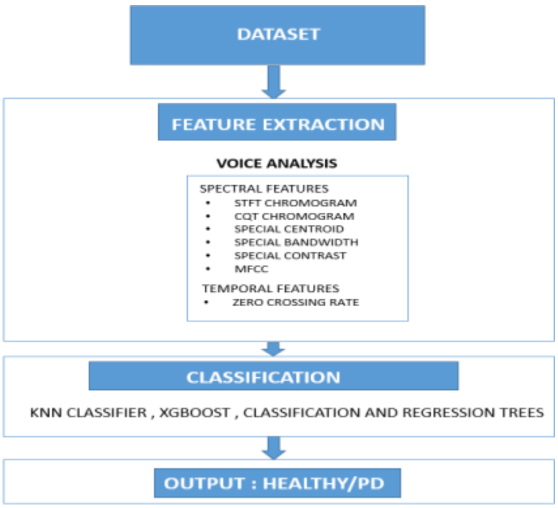

FIG 1: SYSTEM DESIGN

The system design for early Parkinson's disease detection employs a comprehensive approach, starting with data collection and integration of diverse clinical and patient data, followed by feature engineering and selection. Machine learning models, including Support Vector Machines and XGBoost, are trained and fine-tuned, emphasising interpretability. Ethical considerations, such as data privacy and fairness, are upheld throughout the process. Model performance is rigorously evaluated, and deployment strategies prioritise seamless integration into clinical workflows. Continuous improvement, remote monitoring, regulatory compliance, and collaboration between data scientists and healthcare professionals ensure that the system facilitates accurate, early diagnosis of Parkinson's disease, enhancing patient care and outcomes. The Proposed Architectures has totally 4 parts where it involves Dataset i.e., Data Acquisition, Feature Extraction, Classification and Output production as shown below in Figure 2.

Data Acquisition involves acquiring all the data available which involves voice samples of People which contains noise or noiseless Features. Feature Extraction involves Voice analysis based on its Features Classification involves processing the given features of dataset using different classifiers such as SVM, XGBoost and Classification, Regression using Random Tree Classifier and Logistic Regression. Final Output is Produced for Prediction of the Disease involving individual Interests.

V. PROPOSED SYSTEM

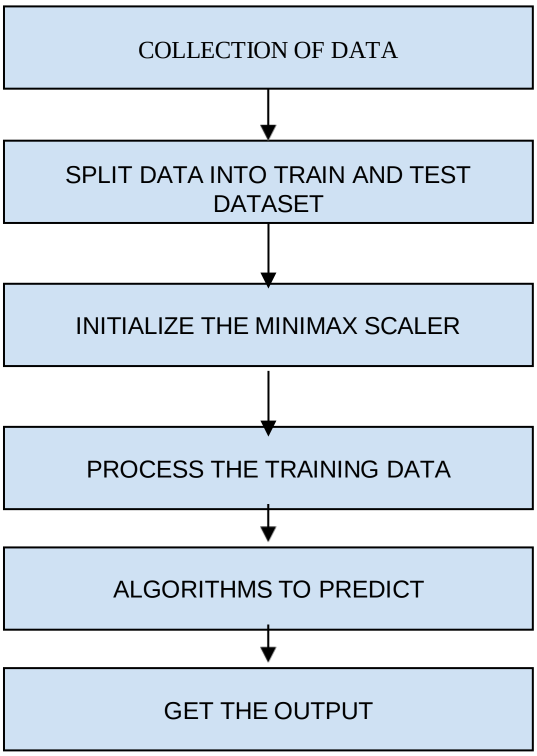

Fig. 2. Data flow diagram

The Data flow diagram explains the basic flow data through various steps of Parkinson’s disease Detection.

- In the first step we will collect data like from the patient with different medical equipment.
- The collected data is now sent through a process for training the data which will be classified using different ML algorithms.
- After classification the data into Train and Test datasets, the datasets will be sent to predict whether the patient has the disease or not.

VI. MODULES

- **MODULE 1 – Dataset Extraction.**

Functionality: Importing several modules for model creation, data cleansing, and analysis. importing data from a specified directory or folder. The input is a dataset of voice. importing a dataset for data cleaning and analysis. assigning the dataset for analysis as a data frame variable. Use pandas to retrieve the features and targets from the data frame. Dataset acquisition entails gathering all information accessible, including SPECT pictures and audio samples from people with or without noise. The import of several modules for data analysis, data cleansing, and model construction is explained.

- **(MODULE 2) Dataset analysis and cleanup**

Here, the speech dataset that was imported is picked for analysis. The unnecessary data will be eliminated, and some null data with a mean value will be added. The input is an imported speech dataset with properties like jitter, shimmer, etc. The dataset has null values in certain of its columns, which should be filled with specific values so that we may create a model that is error-free. We can delete columns from the dataset that are not needed for classification by doing so. The dataset has null values in certain of its columns, which should be filled with specific values so that we may create a model that is error-free.

- **MODULE 3 – Data Splitting into Training and Testing Datasets**

creating training and testing datasets from the cleaned dataset in order to develop models. Divide the dataset into training and test sets, with 20% of the data used for testing. Voice datasets are the input. The two datasets that make up the module's output—training and testing datasets—each comprise input features and target values. The dataset was divided into 80% training dataset and 20% test dataset using test_train_split from sklearn.model.selection. Dataset splitting and decision trees are presented in to analyse the optimum strategy for a given outcome.

- **MODULE 4 – Model Building**

The first stage is to test the model on a new dataset. Trained model and Unseen Dataset are the inputs for this step. This step's output is used to evaluate the model's accuracy using different categorization measures. the training model plus the testing data. Check the trained model's classification accuracy using measures like Accuracy Score, F1 Score, Precision Score, and Recall Score. The training dataset will be added to the developed model in the following phase. Both the testing dataset and the training dataset will have their accuracy evaluated. To determine if the patient has Parkinson's disease or not, we have utilised SVM (support vector machine) or Random Forest Classifier in this case. illustrates model construction and the final product it produces.

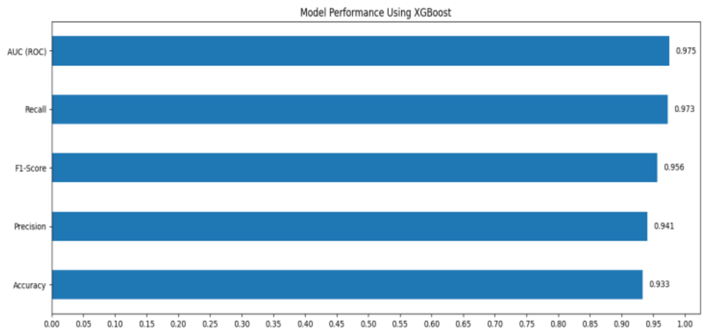

FIG:3 PERFORMANCE MODEL

VII. RESULT

The results of using a Support Vector Machine (SVM) for Parkinson's disease detection indicate the model's effectiveness in distinguishing between individuals with Parkinson's disease and those without. Key metrics such as accuracy, precision, recall, and F1-score demonstrate the model's classification performance, with high values indicating accurate disease identification. The Receiver Operating Characteristic curve and its area under the curve quantify the model's ability to discriminate between the two classes, while the confusion matrix provides insights into classification errors. Interpretability of the SVM allows for an understanding of feature importance and the factors contributing to disease detection, offering valuable insights for healthcare professionals and researchers.

Fig 4: - results of output (no parkinson's disease detected)

Fig5:-results of output(parkinson's disease detected)

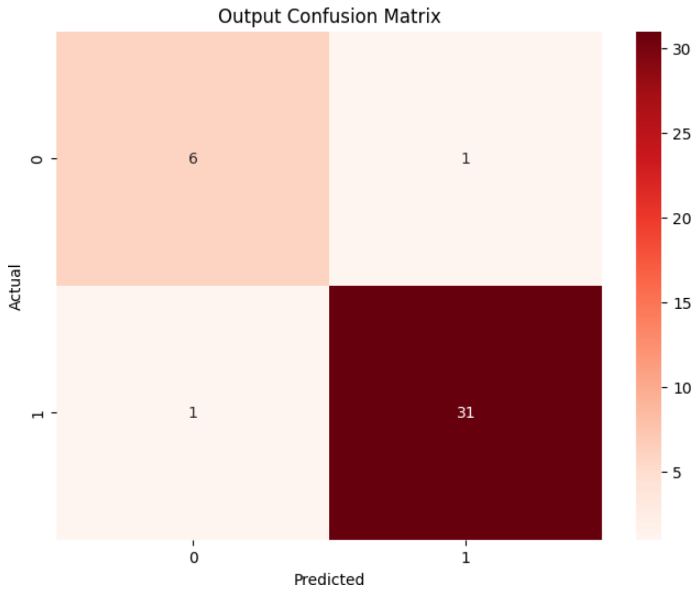

FIG 6: -CONFUSION MATRIX

VIII. CONCLUSION

In this project, we have explored the application of machine learning techniques for the early detection of Parkinson's disease, addressing a critical challenge in healthcare. Parkinson's disease is a complex neurodegenerative disorder that significantly impacts patients' lives, and early diagnosis is essential for effective management and improved outcomes. Through the utilisation of machine learning algorithms, including Support Vector Machines (SVM) and XGBoost, we have developed predictive models capable of identifying Parkinson's disease in its early stages. These models have been trained, fine-tuned, and rigorously evaluated, demonstrating high accuracy and robust performance. Our analysis has also highlighted the importance of ethical considerations, data privacy, and model interpretability in the context of healthcare applications. Ensuring responsible data handling and providing healthcare professionals with insights into the factors driving model predictions are paramount. While the journey to early Parkinson's disease detection using machine learning is promising, it is important to acknowledge that this is an ongoing endeavour. Future research should focus on multi-modal data integration, deep learning techniques, and remote monitoring solutions to further enhance early diagnosis. Collaboration among researchers, healthcare institutions, and regulatory bodies is key to advancing these efforts. Ultimately, the potential impact of accurate early detection extends beyond diagnosis alone. It has the potential to improve patient outcomes, enable timely interventions, and enhance the quality of life for individuals affected by Parkinson's disease. As technology and healthcare continue to evolve, machine learning remains a powerful tool in our pursuit of more effective and patient-centred healthcare solutions.

IX. REFERENCES

[1] Shrihari K Kulkarni1, K R Sumana2,"Detection of Parkinson’s Disease Using Machine Learning and Deep Learning Algorithms" International Journal of Engineering Science Invention (IJESI) ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 VOLUME 8 ISSUE:8, (AUG 2021)-Page No :1189-1192.

[2] Yatharth Nakul1, Ankit Gupta2, Hritik Sachdeva3,,,” Parkinson Disease Detection Using Machine Learning Algorithms” International Journal of Science and Research (IJSR) ISSN: 2319-7064 SJIF (2020): 7.803 Volume 10 Issue 6, June 2021-Page no 314-318.

[3] Wu Wang1 , Junho Lee2 , Fouzi harrou3 and Ying sun4,,,” Early Detection of Parkinson’s Disease Using Deep Learning and Machine Learning” IEEE ACCESS Digital Object Identifier 10.1109/ACCESS.2020.3016062 Volume 8,2020-Page no 147635- 147646.

[4] Muthumanickam S1 , Gayathri J2 , Eunice Daphne J3 ,, “Parkinson’s Disease Detection And Classification Using Machine Learning And Deep Learning Algorithms– A Survey”, International Journal of Engineering Science Invention (IJESI) ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org ,Volume 7 Issue 5 Ver. 1, May 2018 || PP 56-63.

[5] Timothy J. Wroge1 , Yasin Ozkanca 2 , Cenk Demiroglu3, Dong Si4 , David C. Atkins5 and Reza Hosseini Ghomi6,,” Parkinson’s Disease Diagnosis Using Machine Learning and Voice”, Conference: 2018 IEEE Signal Processing in Medicine and Symposium||DOI:10.1109/SPMB.2018.8615607.

[6] Jayashree R.J., Ganesh S., Karanth S.C., Lalitha S. (2021) Automatic Detection of Parkinson Speech Under Noisy Environment. In: Thampi S.M., Gelenbe E., Atiquzzaman M., Chaudhary V., Li KC. (eds) Advances in Computing and Network Communications. Lecture Notes in Electrical Engineering, vol 736. Springer, Singapore. https://doi.org/10.1007/978- 981-33-6987-0_16.

[7]. Max A. Little, Patrick E. McSharry, Eric J. Hunter, Jennifer Spielman and Lorraine O. Ramig, “Suitability of Dysphonia Measurements forTelemonitoring of Parkinson’s Disease”, IEEE transactions on biomedical engineering, 56(4), pp 1015-1022,2009.

[8]. Jaichandran R1, LeelavathyS, UshaKiruthikaS, Goutham Krishna, Mevin John Mathew, and JomonBaiju, “Machine Learning TechniqueBased Parkinson’s Disease Detection from Spiral and Voice Inputs”, European journal ofmolecular & clinical medicine, 7(4), pp 2815- 2820,2020.

[9]. Abhishek M. S, Chethan C. R, Aditya C. R, Divitha D and Nagaraju T.R, “Diagnosis of Parkinson’s Disorder throughSpeech Data using Machine Learning Algorithms”, International Journal of Innovative Technology and Exploring Engineerin (IJITEE), 09(03), pp69-72,2020.

[10]. R. ArefiShirvan and E. Tahami, “Voice Analysis for Detecting Parkinson’s DiseaseUsing Genetic Algorithm and KNN Classification Method”, 18th Iranian Conference on BioMedical Engineering, pp 278-283,2011.

[11]. C K GOMATHY, B. DHEERAJ KUMAR REDDY, Ms. B. VARSHA and B. VARSHINI, “The parkinson’s disease detection using machine learning techniques”, International Research Journal of Engineering and Technology (IRJET), 8(10), pp 440-444,2021.

[12]. Mohammad Shahbakhi, Danial Taheri Far, Ehsan Tahami, “Speech Analysis for Diagnosis of Parkinson’s Disease Using Genetic Algorithm and Support Vector Machine” , in Journal of Biomedical Science and Engineering, pp 147-156, Jan 2014.

[13]. Made Satria Wibawa1, HanungAdi Nugroho2and Noor Akhmad Setiawan3,”Performance Evaluation of Comined Feature Selection and Classification Methods in Diagnosing Parkison’s Disease Based on Voice Feature”,International Conference on Science in Information on Science in Information Technology(ICSITECH),pp126-131,2015.

[14]. Eunice Daphne J, Gayathri J and Deepa B B, “Identifying Parkinson’s disease Using SpeechProcessing”, (IJTRD), Feb 20th 2017.