RESEARCH PAPER

Parkinson's Disease Detection Using Machine Learning

Ms. Syed Amirah Rizwan[1], Dr. Ritesh Shrivastava[2]

[1] Department of Computer Science and Engineering, Anjuman College of Engineering, Nagpur, Maharashtra, India.
Email: syedamirahrizwan@gmail.com

[2] Department of Computer Science and Engineering, Anjuman College of Engineering, Nagpur, Maharashtra, India.
Email: ritct@rediffmail.com

Received: 20th Feb, 2026 | **Revised:** 4th Mar, 2026 | **Accepted:** 25th Mar, 2026 | **Available Online:** 10th Apr, 2026

ABSTRACT

For this project, machine learning techniques are applied in order to build an accurate and early detection system for Parkinson's disease. The dataset consisting of clinical data along with patient information is pre-processed and separated into training and test sets. SVMs and XGBoost techniques are applied in this case to optimize the models based on several evaluation criteria. Finally, the developed models can greatly improve the early detection of Parkinson's disease leading to timely treatment. Finally, in order to evaluate and optimize the developed models, several performance metrics are used. In particular, accuracy, precision, recall, F1 score, ROC-AUC, and confusion matrix are some of the key metrics used in the evaluation process of the developed models. Alongside these performance metrics, feature importance analysis is carried out to identify the key factors behind Parkinson's disease detection. Feature importance visualizations are created for both models for a better interpretation of their decisions.

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

How to cite this article: Rizwan SA, Shrivastava R. Parkinson's Disease Detection Using Machine Learning. Int J Drug Deliv Technol. 2026;16(31s):665-672. DOI: 10.25258/ijddt.16.31s.73

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

I. INTRODUCTION

Parkinson's disease is a complicated disease, and it is where the gradual loss of functionality in the central nervous system in millions of people worldwide occurs. In 1817, Parkinson's Disease was found by James Parkinson. Symptoms of Parkinson's Disease are quite complex as they include motor symptoms and non-motor symptoms. Tremors while resting, bradykinesia, rigidity, and postural instability are some of the examples of motor symptoms. Some of the non-motor symptoms are cognitive problems, depression, anxiety, insomnia, and autonomic problems. Parkinson's Disease can be considered a neurodegenerative disorder that affects the patient's life adversely.

Parkinson's disease is caused by interactions among genetics and environmental factors; however, despite a lot of studies in the past, very few discoveries have been made in relation to etiology and pathogenesis of Parkinson’s disease. Although great success has been attained regarding finding out the pathophysiology of Parkinson’s disease, early diagnosis is difficult. Diagnosis of Parkinson’s disease can be achieved using a process called differential diagnosis by neurologists in view of motor symptoms and the patient's medical history. However, since the motor symptoms of Parkinson’s disease are always diagnosed at a later stage,As such, there is a dire need for more efficient ways of early detection of Parkinson’s disease. Here, Machine Learning (ML) comes in play. Machine learning is a subfield of artificial intelligence and has gained recognition in the healthcare sector in diagnosing and managing different diseases. The application of ML includes the capacity of analysing datasets and coming up with predictions based on data input. In the case of Parkinson’s disease, ML allows the algorithms to identify any change in the data, which might be too subtle for humans. Therefore, one of the key objectives of using ML in early detection of Parkinson’s disease is developing models of diagnosis using data inputs as features. This involves identification and selection of appropriate features before training the algorithms with the selected features. This process of feature selection requires very efficient techniques. In most cases, SVM and XGBoost algorithms are used.

II. LITERATURE SURVEY

K Kulkarni, Shrihari, K R Sumana, the authors make use of the Decision tree, Logistic Regression, and Naive Bayes, Deep Learning algorithm i.e., Recurrent Neural Network (RNN) by predicting the performance parameters for model building purposes. Machine Learning algorithms would be used for the purpose of developing models that would differentiate early stage Parkinson’s disease from normal through the use of MDS UParkinson’s diseaseRS. In subject and record validation, Logistic regression, random forests, and support vector machine are made use of. Drawbacks in this paper are that there exists little control over data collection procedures, leading to errors such as out-of-bound values. This model depends heavily on the evaluation of motions, but this should not be the sole means of acquiring data from disease carriers or healthy people. Yatharth Nakul, Ankit Gupta, Hritik Sachdeva, the authors make use of supervised learning algorithms namely, Random Forest, Support Vector and Naïve-Bayes among others.For accuracy, confusion matrix method is used and there have been different classifications used here. ML classification will increase the accuracy and decrease loopholes that can occur. Hyperparameter tuning will help to achieve the maximum accuracy. I have been able to achieve the maximum accuracy of 98.30% using K nearest neighbour classification technique. The key limitations include Delay in Result Derived, slow Output Progression and highest Error rate of Best Proposed Methodology is found using the Confusion Matrix. According to authors Wu Wang, Junho Lee, Fouzi Harrou, and Ying Sun in SGD (Stochastic Gradient Descent) is applied for training data models. The FNN (Feed-Forward Neural Network) is employed. Linear discriminate analysis method employed sensitivity is the best, meaning it has the most probability of identifying the actual patient. The deep learning model accuracy is 96.45%. This because deep learning model capability in terms of learning both linear and nonlinear properties of Parkinson's disease data without the need of extracting manually. Limitations of the model include being black box algorithm, making it hard to analyze after the model is trained. Also, it becomes difficult theoretically to understand how the algorithm achieves 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 easy to understand. It can be modified to avoid overfitting. The sgd function can be utilized to fit linear models. The Algorithm and the Outputs of Binary Logistic Regression Have an Excellent Interpretation.Jayashree R. J, Ganesh S, Karanth S.C, Lalitha S has elaborated on how spectral features such as Spectral Contrast, STFT and temporal features such as Zero Crossing rate can be extracted and classification done using XGBoost and classifiers such as Random Forest and regression such as Logistic Regression. It has several advantages such as the Real Time Speech Analysis that perfectly captures the effect of noise and other factors on the detection of Parkinson's disease. It illustrates a higher Accuracy and a distinctive Characteristics approach that proves the effect of noise in the prediction of Parkinson's Disease.The major disadvantage of this methodology is that it works on a small sample size; if more sample sizes were available, then there would have been a more realistic approach. The Limited Classifier technique and Feature Analysis are done using only eight features.

III. PROBLEM DEFINITION

The issue being addressed through this project concerns the creation of an approach involving effective machine learning methodologies that will make it possible to accurately detect Parkinson's disease. This neurodegenerative condition is associated with the presence of motor and non-motor symptoms, which significantly complicates the process of diagnosing the illness. The goal of the research in question is to apply machine learning techniques in order to develop prediction algorithms for detecting Parkinson's disease at its initial stages, when motor symptoms have not become apparent yet. The project will entail activities including data gathering, preprocessing, selecting and training the algorithm, and analyzing the outcome in order to ensure accuracy of predictions produced.

IV. METHODOLOGY

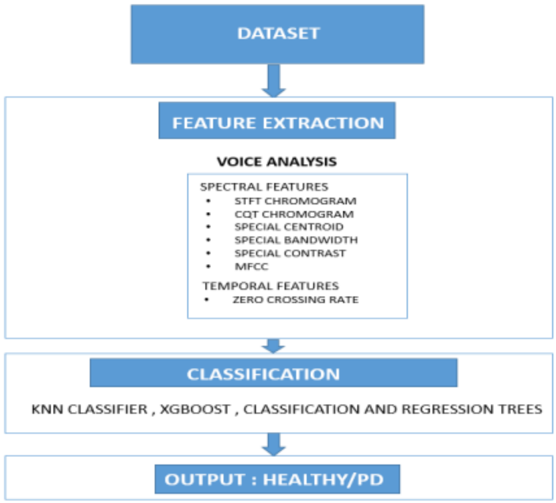


FIG 1: SYSTEM DESIGN

The design architecture for diagnosing early stages of Parkinson's disease uses a holistic approach, beginning with data acquisition and integration, then feature extraction and selection. Support vector machines and XGBoost machine learning algorithms are built and optimized, focusing on explainability. Ethics in this case include maintaining data privacy and fairness. The model will be thoroughly tested for accuracy before deployment to optimize integration with clinicians. Continuous improvement, remote monitoring, regulatory requirements, and the collaboration of data scientists with clinicians help improve the system and ensure accurate diagnosis of early stages of Parkinson's disease. Proposed Architectures is divided into 4 components where the first one includes the Dataset, i.e., Data Acquisition, Feature Extraction, Classification, and Output Production as illustrated in Figure 2 below.

Data Acquisition includes collecting all data, which includes both noisy and noiseless feature extraction of the voice samples of People. Feature Extraction includes analysing the voice samples according to their features. Classification includes classification and regression of the provided features from the dataset with the help of different classifiers, such as SVM, XGBoost, Classification, Regression, and Random Tree Classifier, Logistic Regression.

V. PROPOSED SYSTEM

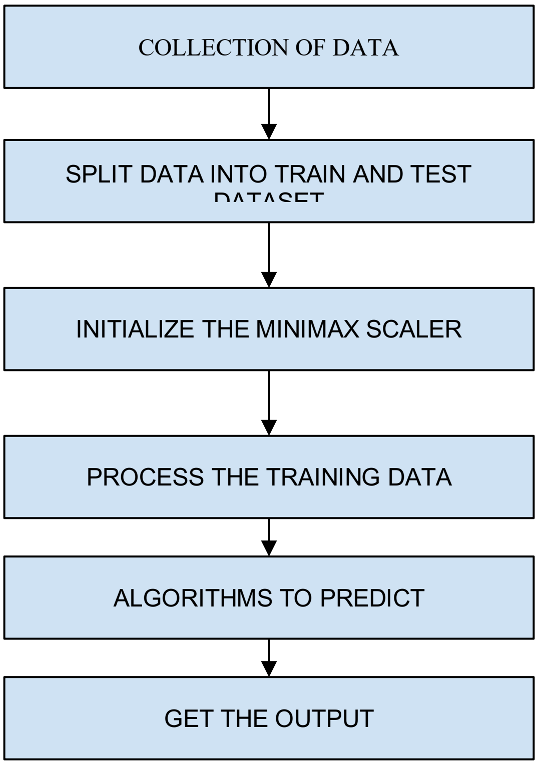


Fig. 2. Data flow diagram

The Data Flow Diagram illustrates the flow of data within the Parkinson's disease Detection system.

- First of all, we need to gather data from the patient by using various medical instruments.
- Then, the gathered data will be transferred through processing and training stages. At this point, machine learning algorithms will be used for classifying the data.
- Finally, after classifying the data into train and test datasets, they will be transferred for prediction purposes.

VI. MODULES

- **MODULE 1 – Dataset Extraction.**

Functionality: Importing several modules to create models, data cleansing, and data analysis.importing the data from a given directory or folder. The input data includes a dataset of voice.importing the dataset for cleansing and data analysis.assignment of the dataset used in the analysis as data frame variables.Under Dataset Acquisition, the data obtained include all information available such as SPECT images and sound samples of people who have noise or no noise.Importation of several modules for data analysis, data cleansing, and creation of models is explained above.

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

In this case, the speech dataset that was imported will be chosen to be analyzed. Unnecessary data will be removed, and null data with mean values will be included. The input will be a speech dataset imported that contains parameters such as jitter, shimmer, and so forth. There are null values in some columns of the dataset, which need to be replaced with particular values to ensure that we have a perfect model. Columns in the dataset that are not necessary in classification can be removed.

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

Creating training and testing data sets from the clean data set for the purpose of creating models. Split the data set into training data set and test data set, with 20% data allocated to the test data set. Voice data sets form the input. Training and test data sets, which are the outputs of the module, each include input attributes and targets. The dataset was split between 80% training dataset and 20% test dataset using test_train_split of sklearn.model.selection. Data splitting and decision trees are discussed in order to find the best method for a particular result.

- **MODULE 4 – Model Building**

Step one is running the test on a fresh dataset. For this step, the inputs include Trained Model and Unseen Dataset. The output from this step is used to test the performance of the algorithm by using various forms of categorization. This would include using Accuracy Score, F1 Score, Precision Score, and Recall Score. The training dataset is then included in the developed model in the next step. Both testing and training datasets will be tested in terms of accuracy. In order to check whether a patient is suffering from Parkinson's disease, we have considered SVM (Support Vector Machine) or Random Forest Classifier in this case.

Support Vector Machine (SVM)

Support vector machine (SVM) [30] is a supervised learning technique which generates a hyperplane to divide N features, which are mapped into multiple dimensions. The SVM architecture is shown in figure. Since Parkinson's disease voice data cannot be linearly divided, SVM kernel has been used for transformation of data into higher dimension space. SVM works effectively for Parkinson's disease because of its memory effectiveness and support vectors generated from partial training points..

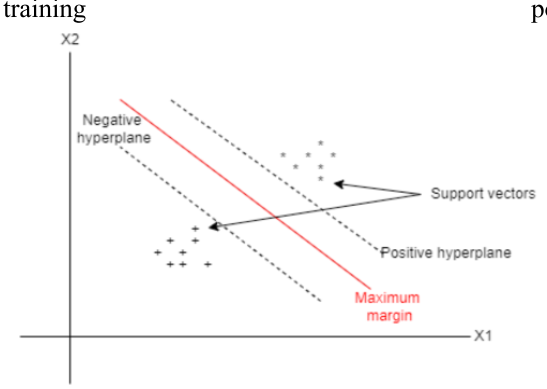


Fig:3 : Support Vector Machine (SVM)

K nearest neighbors(KNN)

The k-nearest neighbors [3] (KNN) method is a supervised non-parametric machine learning algorithm that forms clusters in accordance with similarities among data points. This technique works effectively on audio data sets having 109 entries, owing to its small size. Clusters formed are two in number, one for the PWP group and the other for the healthy group. It is a lazy learner technique, which means no assumptions about the data are made.

XGBOOST Algorithm

XGBoost is a machine learning algorithm that belongs to the ensemble learning category, specifically the gradient boosting framework. It utilizes decision trees as base learners and employs regularization techniques to enhance model generalization. Known for its computational efficiency, feature importance analysis, and handling of missing values, XGBoost is widely used for tasks such as regression, classification, and ranking.The algorithm works by sequentially adding weak learners to the ensemble, with each new learner focusing on correcting the errors made by the existing ones. It uses a gradient descent optimization technique to minimize a predefined loss function during training.

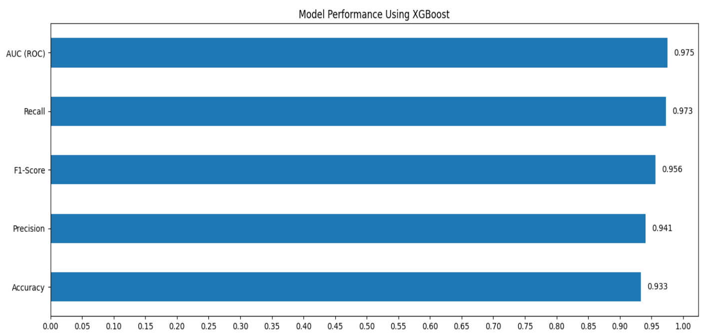


FIG:4 PERFORMANCE MODEL

- **MODULE 5 – Model Building**

To determine the model with the highest efficiency, we examine the outcome of three different approaches and nine models that have been trained. For this purpose, metrics considered are ROC-AUC graph, confusion matrix, accuracy, precision, recall, and F1-score. The formulae for the above-discussed metrics are provided in equations 1-3. Here, TP represents true positives, FP false positives, TN true negatives, and FN false negatives.

$$precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Accuracy = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

ROC or Receiver Operating Characteristics is a probabilistic plot. AUC indicates the area under this ROC curve. The metric provides information about separability or ability of the model to discriminate between classes.

Table 1 below depicts the results of models after approach 1 is applied, that is, models are trained on 22 attributes of MDVP dataset.

Metric	**Logistic Regression**	**Random Forest**	**SVM**	**KNN**
Accuracy	83.67%	91.83 %	85.71 %	85.71 %
Precision	1.0	0.95	1.0	0.95
Recall	0.83	0.86	0.84	0.86
ROC AUC curve	0.636	0.701	0.682	0.701

Table I:Results of Approach 1: 22 attributes training

Random Forest classifier is the best choice for the entire dataset since it is an ensemble method. This classifier analyzes the mean value of 100 decision trees before making any predictions. All features are given equal importance during the classification stage. The confusion matrix of this algorithm has been depicted in the figure below, where this algorithm predicts seven true negatives (noParkinson's disease), four false negatives, thirty-eight true positives (PWP), and zero false positives.

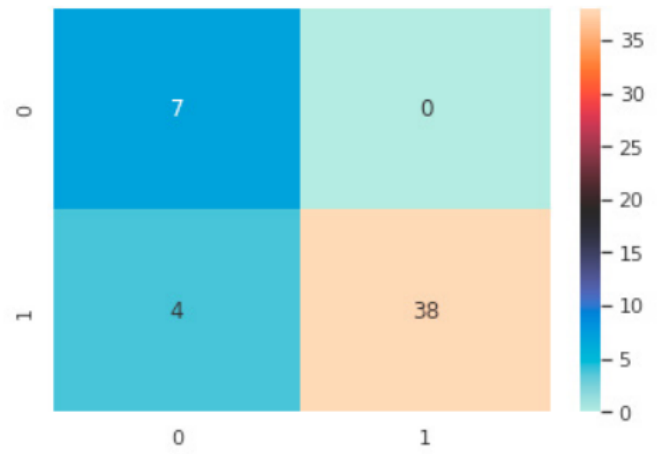


Fig 4:. Confusion matrix for Random Forest Model in approach 1

Approach 2 results after using PCA have been illustrated in Table 2. There are 5 prominent features found in PCA which are MDVP, Shimmer, Jitter, PPE, and RParkinson's diseaseE. After developing models on these features, we obtain

Metric	**Logistic Regression**	**Random Forest**	**SVM**	**KNN**
Accuracy	83.67%	83.67%	91.75 %	83.67 %
Precision	1.0	1.0	1.0	0.92
Recall	0.83	0.90	0.86	0.90
ROC AUC curve	0.636	0.818	0.727	0.779

Table II:Results of Approach 2: 5 attributes after PCA

Classifier using Support Vector with L1-support and Linear kernel is best suited for PCA dataset since it finds out perfect hyperplane quickly and with high precision. Confusion matrix of this model is illustrated in Figure 10, below, where SVM correctly classifies 5 TN (No Parkinson’s Disease), 6 FN, 38 TP (Parkinson’s with WPW) and 0 FP.

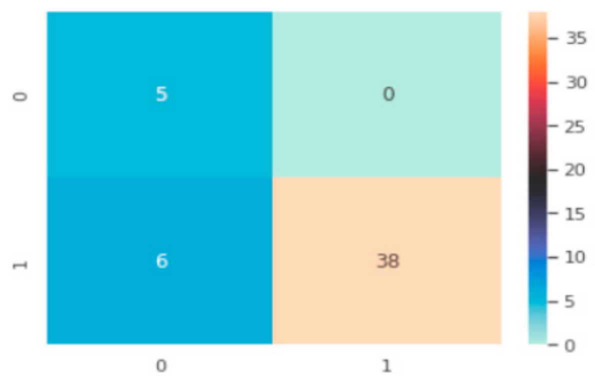


Fig 5: Confusion matrix for Support Vector Model in approach 2

Table 3 shows the output from approach 3 where a balanced dataset was used. Models were trained with an equal number of records of patients with normal cases and Parkinson’s cases. Balanced datasets ensure that each category gets the same weightage in the process. The following results have been obtained:

The KNN model came out top in all the parameters of accuracy, precision, recall, and ROC AUC compared to other models used in the approach 3 and balanced dataset. SVM model exhibited great results concerning recall and ROC AUC even though its accuracy was slightly less than others.

Metric	Logistic Regression	Random Forest	SVM	KNN
Accuracy	85.71%	85.71 %	81.63 %	91.83 %
Precision	0.89	0.89	0.82	0.95
Recall	0.92	0.92	0.94	0.95
ROC AUC curve	0.811	0.811	0.817	0.883

Table 3. Results of Approach 3: Balanced dataset

The k nearest neighbor algorithm has the highest accuracy and recall of 0.95. Because of the uniformity of data, similarities between PWP and non-Parkinson’s disease patients can be easily detected. Confusion matrix of classification is depicted in the figure below. The number of true positives (PWP) is 36; false positives (non-Parkinson’s disease) is 2; true negatives (no Parkinson’s disease) is 9, while false negatives (non-Parkinson’s disease) is 2.

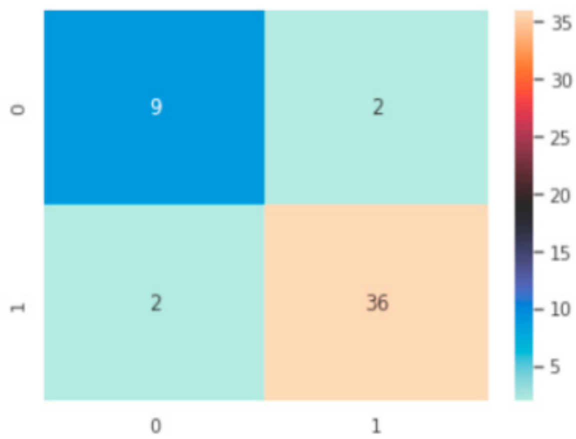


Figure 6. Confusion matrix for KNN model in approach

VII. RESULT

Based on the use of SVM for detecting Parkinson's disease, the efficiency of the model in differentiating Parkinson's patients from non-Parkinson's patients has been established through various evaluation criteria such as accuracy, precision, recall, and F1-score. Accuracy is a measure of the model's capability to classify data, whereby high values mean that the model effectively identifies the disease status. In addition, ROC and AUC measures the discrimination capability of the model, while the confusion matrix gives information regarding the error in classification. Interpretability is a crucial component that makes it possible to identify the key features and understand what causes the disease detection.

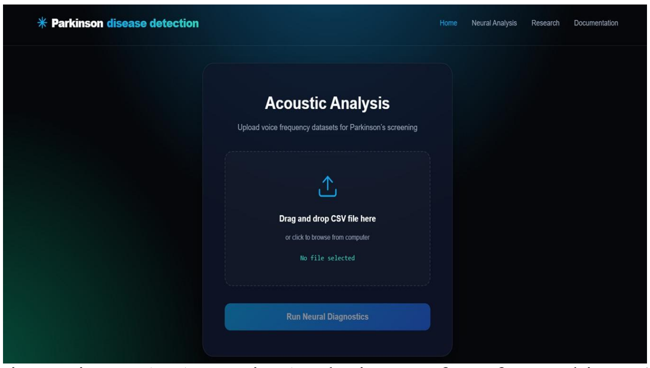


Fig 7:Figure 1: Acoustic Analysis Interface for Parkinson's Disease Detection

The figure illustrates the acoustic analysis interface of a machine learning-based system designed for the early detection of Parkinson's disease. The interface provides a user-friendly platform where users can upload a voice recording in .wav format. Once the audio file is uploaded, the system processes the speech signal to extract important acoustic features such as pitch variations, jitter, shimmer, and frequency irregularities, which are commonly affected in individuals with Parkinson's disease. After feature extraction, these parameters are analyzed using trained machine learning models like Support Vector Machines (SVM) and XGBoost to predict the presence of the disease. The interface also includes a "Run Voice Diagnosis" option that initiates the analysis and displays results. Overall, the diagram represents a non-invasive, efficient, and accessible approach for early detection of Parkinson's disease using voice-based data.

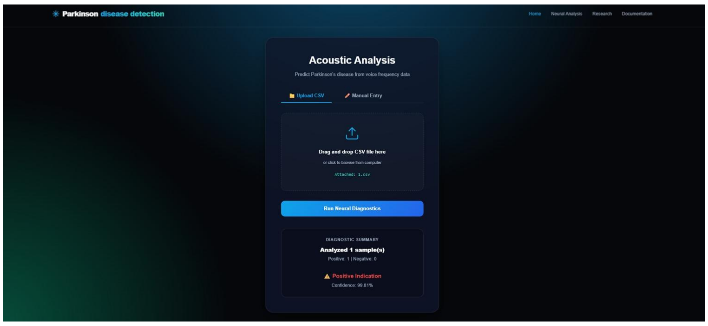
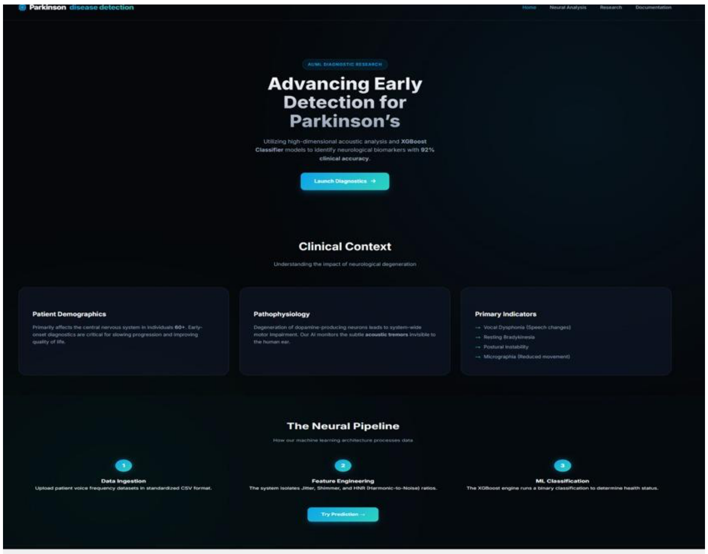


Fig 8 : Acoustic Analysis Result Interface for Parkinson's Disease Detection

The Fig 8 represents the enhanced acoustic analysis interface that not only accepts voice input but also displays the diagnosis results. The user uploads a voice sample in .wav format, after which the system extracts relevant speech features such as jitter, shimmer, and pitch variations. These features are then processed using trained machine learning models to determine whether the voice sample indicates signs of Parkinson's disease.The interface includes a "Run Neural Diagnostics" button to initiate the analysis. After processing, the system presents the diagnostic outcome, indicating whether the result is positive or negative, along with a confidence score. In this figure, the output shows a positive indication of Parkinson's disease with a certain confidence level, demonstrating the model's predictive capability.Overall, this interface highlights the complete workflow—from data input to result generation.

Figure 3: Homepage Interface of Parkinson's Disease Detection System

Figure 3 illustrates the homepage of the Parkinson's disease detection system, designed to provide an overview of the platform and guide users through its functionality. The interface prominently displays the objective of the system, which is advancing early detection of Parkinson's disease using machine learning techniques. It includes a central call-to-action button ("Launch Diagnostics") that directs users to the diagnostic module.Below the main section, the interface presents the clinical context, highlighting important aspects such as patient demographics, pathophysiology, and primary indicators of Parkinson's disease. These sections provide users



with background knowledge about the condition and its symptoms.At the bottom, the neural pipeline is illustrated, outlining the system workflow, including data ingestion, feature engineering, and machine learning classification. This gives users a clear understanding of how the system processes input data and generates predictions.

VIII. CONCLUSION

In this research project, we investigated the possibilities of implementing machine learning methods for early detection of Parkinson's disease, which remains one of the most critical problems in healthcare. Being a complex neurodegenerative disorder that substantially influences people's lives, the disease needs an early identification, as it is key for its proper treatment.

With machine learning algorithms such as Support Vector Machines (SVM) and XGBoost, we managed to develop several efficient models capable of detecting the disease at its early stages, which was achieved through training, fine-tuning and rigorous evaluation of the models.Ethical concerns associated with machine learning use in healthcare, data privacy and interpretable models were also considered throughout the course of our research, which is particularly important to provide healthcare providers with the reasons behind each prediction and ensure data integrity and confidentiality.Furthermore, even though our research brought us closer to developing a method for early identification of Parkinson's disease, we should not forget about numerous opportunities for further development. Multi-modal data and deep learning should be used in the future for better results. Working together with researches, hospitals, and other regulators will ensure the success of such endeavors. In the end, the possible influence that correct early detection can have is not limited to diagnosing patients. It can positively affect their further lives as well as increase their chances for recovery. With advances in technologies and the development of medicine, machine learning still holds much promise for the future.

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.

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

[16] 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.

[17] X. Yang, Q. Ye, G. Cai, Y. Wang and G. Cai, (2022), "Parkinson’s disease-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.

[18] 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.

[19] 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.

https://www.ppmi-info.org/access-data-specimens/download-data

[20] Amalia Luque, Alejandro Carrasco, Alejandro Martín, Ana de las Heras, (2019), “the impact of class imbalance in classification performance metrics based on the binary confusion matrix”, Pattern Recognition, Volume 91, Pages 216-231, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2019.02.023.

[21] J. R. Barr, M. Sobel and T. Thatcher (2022), "Upsampling, a comparative study with new ideas," 2022 IEEE 16th International Conference on Semantic Computing (ICSC), pp. 318-321, doi: 10.1109/ICSC52841.2022.00059.