INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1818-1830
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

PARKINSON'S DISEASE PREDICTION SYSTEM

Mrs. Hemavathi R[1], Jheevashankar M[2], Pandiyan S[3]
[1,2,3]Sri Shakthi Institute Of Engineering And Technology, India.

ABSTRACT

arkinson's Disease (PD) is a chronic and progressive neurodegenerative disorder that primarily affects motor functions due to the loss of dopamine-producing neurons in the brain. Early detection of PD is vital for mitigating its effects and enhancing patient outcomes. This study explores the use of machine learning (ML) techniques for diagnosing PD through voice-based biomarkers. Voice features, such as pitch, jitter, shimmer, and harmonic-to-noise ratio, are extracted from patient speech recordings, which often reveal subtle changes indicative of PD.

Four ML models were evaluated for their effectiveness in classifying PD: Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN). Among these, the Random Forest classifier demonstrated superior performance with an accuracy of 91.83% and a sensitivity of 0.95, highlighting its potential for reliable early detection.

To facilitate practical implementation, a user-friendly web application was developed using the Streamlit framework. This application enables remote testing by allowing users to upload voice recordings and receive diagnostic predictions in real time. The system offers a cost-effective, scalable, and non-invasive diagnostic solution, which is especially advantageous for telemedicine applications in neurology. By integrating advanced ML techniques with an accessible digital platform, this project aims to revolutionize early PD detection, providing a significant step forward in personalized healthcare and remote diagnostics.

1. INTRODUCTION

Parkinson's Disease (PD) is a progressive neurological disorder that affects approximately 1% of individuals over the age of 60 worldwide, with prevalence increasing as the population ages. The disease is characterized by a range of motor symptoms, including tremors, muscle rigidity, bradykinesia (slowness of movement), and postural instability. Non-motor symptoms, such as speech impairments, also play a significant role in the progression of the disease. Despite advancements in medical research, there is currently no cure for PD. Treatment options primarily focus on managing symptoms and improving patients' quality of life.

The Importance of Early Diagnosis:

Early detection of PD is critical for slowing disease progression, optimizing treatment strategies, and providing better patient outcomes. However, traditional diagnostic methods, which often rely on clinical assessments and specialized tests, present several challenges:

Subjectivity: Diagnosis heavily depends on the expertise and observation of healthcare professionals.

Accessibility: Specialized testing facilities may not be readily available in rural or remote areas.

Cost and Time: Clinical evaluations can be expensive and time-intensive, limiting their scalability.

These challenges underscore the need for innovative, scalable, and cost-effective diagnostic solutions.

Machine Learning for Parkinson's Disease Diagnosis:

Emerging technologies, such as machine learning (ML), offer a transformative approach to PD diagnosis. Studies have demonstrated that subtle changes in voice patterns can serve as early biomarkers of the disease. Features such as pitch, jitter, shimmer, and harmonic-to-noise ratio reflect the neuromuscular impairments caused by PD. Machine learning algorithms can analyze these features to identify patterns indicative of the disease with high precision.

This project leverages machine learning to develop a diagnostic tool that utilizes voice-based biomarkers for PD detection. By comparing the performance of four ML models—Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN)—the study identifies the most effective algorithm for accurate classification. The Random Forest model emerged as the best performer, achieving an accuracy of 91.83% and a sensitivity of 0.95.

Integration with a Web-Based Application:

To enhance accessibility and usability, the diagnostic system is deployed as a web application using the Streamlit framework. This application enables users to upload voice recordings, which are processed by the trained ML model to provide a diagnostic prediction in real-time. The system is designed to be:

Non-invasive: Requires only a voice sample for analysis.

Scalable: Can be accessed globally with minimal infrastructure.

Cost-effective: Reduces reliance on expensive clinical tests.

By combining the power of machine learning with an intuitive digital platform, this project addresses the limitations of traditional diagnostic methods. It aims to revolutionize early PD detection, particularly in underserved regions, thereby contributing to advancements in telemedicine and personalized healthcare.

2. OVERVIEW

This project focuses on developing an innovative, reliable, and accessible diagnostic system for Parkinson's Disease (PD) by leveraging machine learning (ML) techniques and voice-based biomarkers. PD is a progressive neurodegenerative disorder characterized by motor and non-motor symptoms, with early detection being crucial for slowing its progression and improving patient outcomes. Traditional diagnostic methods are often subjective, time-consuming, costly, and inaccessible to individuals in underserved regions. To address these challenges, this system uses voice recordings to identify subtle speech impairments caused by PD. Key features such as pitch, jitter, shimmer, and harmonic-to-noise ratio are extracted from voice data and analyzed using ML models, including Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN). Among these, the Random Forest classifier achieved the highest accuracy (91.83%) and sensitivity (0.95), making it the most effective model for early detection. The system is integrated into a user-friendly web application built using the Streamlit framework, enabling users to upload voice recordings and receive diagnostic predictions in real time. This non-invasive, cost-effective, and scalable solution is designed to bridge the gap in PD diagnosis, particularly in remote and underserved areas, by providing an efficient telemedicine tool that could revolutionize neurology diagnostics and personalized healthcare.

PROBLEM STATEMENT

Parkinson's Disease (PD) is a progressive neurodegenerative disorder affecting millions worldwide, with a marked increase in prevalence among people above 60 years of age. PD is characterized by both motor symptoms, including tremors, muscle rigidity, bradykinesia or slowness of movement, and postural instability, as well as non-motor symptoms such as speech impairments, cognitive decline, and mood disorders. Although there is no cure for PD, early detection is crucial in the management of symptoms, slowing the progression of the disease, and improving the quality of life of patients.

The current diagnostic methods for PD include clinical evaluations, imaging tests (e.g., MRI or PET scans), and symptom-based assessments. These methods are prone to subjectivity as results depend on the judgment or observation of clinicians. Early stages with slight symptoms cannot be detected or may undergo delay and misdiagnosis. Many advanced diagnostic tools cannot easily penetrate rural and underserved parts of a country due to their cost, special apparatus, and trained personnel needs. This creates major barriers to the timely diagnosis and treatment of many patients, especially in resource-poor settings.

Studies have demonstrated that voice impairments, including changes in pitch, jitter, shimmer, and harmonic-to-noise ratio, are early signs of PD. These vocal changes occur as a result of the neuromuscular effects of the disease even before the more overt motor symptoms manifest. However, current conventional diagnostic workflows do not harness these biomarkers systematically. There is a big requirement for a non-invasive, scalable, and cost-effective diagnostic approach that would easily aid in early detection, with increased accessibility to PD screening.

This project addresses all these challenges by using the power of machine learning techniques to analyze voice-based biomarkers in detecting PD. The models are integrated into web application, enabling the development of a reliable and remote diagnostic tool that could overcome the limitations of the traditional method, offering an innovative solution for early PD detection both in clinical and remote setting.

3. OBJECTIVE

This project aims to address the critical need for early and accessible diagnosis of Parkinson's Disease (PD) by leveraging the capabilities of machine learning (ML) and voice-based biomarkers. Below are the detailed objectives that guide the development and implementation of this innovative diagnostic system:

1. Develop an Accurate and Reliable Diagnostic Tool

Objective: Create an ML-based system that identifies Parkinson's Disease with high accuracy and reliability using voice data.

Details: Train machine learning models to detect subtle variations in voice patterns caused by the neuromuscular effects of PD, such as changes in pitch, jitter, shimmer, and harmonic-to-noise ratio. These voice changes often manifest early, making them a valuable diagnostic marker.

2**. Utilize Voice-Based Biomarkers for Non-Invasive Screening**

Objective: Analyze voice recordings to detect PD symptoms without invasive procedures.

Details: Focus on acoustic features derived from patients' voice samples, such as:

Jitter: Measures variations in pitch between vocal cycles.

Shimmer: Quantifies amplitude variations.

Harmonic-to-Noise Ratio (HNR): Evaluates voice signal quality.

Additional features, including fundamental frequency and signal stability metrics, ensure comprehensive analysis.

This ensures a simple, patient-friendly diagnostic process.

Optimize Machine Learning Model Performance

Objective: Identify the most effective ML algorithm for PD diagnosis.

Details: Compare and evaluate four key ML models:

Random Forest: Known for its robustness and ensemble decision-making.

Support Vector Machine (SVM): Effective in separating classes in a high-dimensional space.

Logistic Regression: A simple, interpretable baseline model.

K-Nearest Neighbors (KNN): Leverages distance-based pattern recognition.

Use metrics such as accuracy, sensitivity, specificity, and F1-score to select the optimal model.

Develop a User-Friendly Web Application for Accessibility Objective: Create a web-based platform for real-time PD diagnosis. Details: Design a Streamlit application with the following features:

Input Fields: Allow users to upload voice data and input acoustic feature values (as seen in the uploaded code).

Real-Time Analysis: Process user input through the trained ML model and return diagnostic results instantly.

Error Handling: Validate user inputs to ensure accurate predictions and avoid errors due to invalid data entries.

Visual Design: Implement an intuitive interface with clear and aesthetic appeal, accessible to non-experts.

Enhance Accessibility and Scalability

Objective: Ensure the diagnostic system is accessible to underserved and remote areas.

Details: The web-based deployment enables:

Remote Access: Users worldwide can test their voice samples without visiting healthcare facilities.

Scalability: The system can handle large-scale usage, making it suitable for mass screening programs.

Provide a Cost-Effective Diagnostic Alternative

Objective: Reduce reliance on expensive clinical tools and invasive procedures.

Details: By utilizing voice-based biomarkers and ML, the system offers a low-cost solution suitable for large-scale implementation. This makes it feasible for resource-constrained settings and individuals unable to afford conventional diagnostic methods.

Secure and Robust Implementation

Objective: Ensure the reliability, security, and ethical handling of user data.

Details: Implement:

Error Detection: Handle invalid or incomplete inputs with appropriate user feedback.

Data Security: Protect sensitive user data to comply with privacy standards maintaintrust.

Revolutionize Telemedicine in Neurology

Objective: Provide a transformative tool for telemedicine by enabling remote diagnosis.

Details: This system aligns with the growing demand for digital health solutions, allowing clinicians and patients to monitor and manage PD symptoms efficiently from anywhere. It empowers early intervention and better management of the disease trajectory.

Address Future Development and Expansion

Objective: Create a foundation for expanding diagnostic capabilities and improving system performance.

Details: Plan future enhancements, such as:

Multimodal Analysis: Combine voice data with other non-invasive biomarkers like handwriting or gait

analysis. Deep Learning Integration: Use advanced neural networks for improved predictive accuracy and robustness. Healthcare Partnerships: Collaborate with clinics and hospitals to validate and deploy the system in real-world settings

Vol. 05, Issue 01, January 2025, pp : 1818-1830
e-ISSN : 2583-1062
Impact Factor : 7.001

4. TECHNOLOGY AND TOOLS

This project utilizes a comprehensive range of technologies, tools, and methodologies to build an efficient and accessible machine learning (ML)-based diagnostic system for Parkinson's Disease (PD). Below is a detailed breakdown of the components:

4.1. Machine Learning Algorithms

Multiple algorithms were explored for their ability to classify voice data as indicative of PD or healthy individuals. Each model was evaluated for accuracy, sensitivity, and specificity:

Support Vector Machine (SVM):

Constructs hyperplanes to separate data points, working efficiently in high-dimensional spaces. After applying Principal Component Analysis (PCA) for dimensionality reduction, SVM achieved near-comparable results to Random Forest.

Logistic Regression:

A probabilistic model that maps independent variables to categorical outcomes.

Provided a reliable baseline for performance comparison, though it was outperformed more sophisticated models.

K-Nearest Neighbors (KNN):

A distance-based classification technique.

Useful for understanding the dataset structure but less effective than Random Forest due to the high-dimensional feature space.

2.2 Dataset

Dataset Sources:

MDVP Dataset:

Biomedical voice measurements from the PPMI (Parkinson's Progression Markers Initiative) and UCI Machine Learning Repository.

Contains 195 records of voice data from 31 participants (23 with PD and 8 healthy individuals).

Each individual's voice is recorded multiple times, capturing a range of attributes related to speech quality.

Key Features in the Dataset:

The dataset includes 22 features derived from sustained vowel phonation tests. Notable features are:

Pitch-Related:

MDVP:Fo(Hz): Fundamental frequency (average pitch) **MDVP:Flo(Hz)**: Minimum pitch. **MDVP:Fhi(Hz):** Maximum pitch. J**itter** (Frequency Variability):

MDVP:Jitter(%), Jitter(Abs), RAP, PPQ: Reflect variations in pitch.

Shimmer (Amplitude Variability):

MDVP:Shimmer, Shimmer(dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:DDA: Measure amplitude stability.

Signal Quality:

HNR (Harmonic-to-Noise Ratio): Quantifies voice clarity.

NHR (Noise-to-Harmonic Ratio): Indicates noise levels in speech.

Other Nonlinear Features:

DFA, RPDE: Nonlinear dynamic variables indicating vocal cord stability.

PPE (Pitch Period Entropy): Reflects pitch regularity.

DATA PREPROCESSING:

Cleaning and Validation:

Removed missing or erroneous values to ensure data quality.

Feature Scaling:

Standardized features using StandardScaler to ensure consistency in model Training.

Principal Component Analysis (PCA):

Reduced dimensions to 5 principal components for SVM evaluation, optimizing computational efficiency without significant loss of information.

Class Balancing:

Addressed class imbalance (109 PD vs. 40 healthy cases) through oversampling techniques to equalize the dataset.

WEB DEVELOPMENT TOOLS

Programming Language:

Python:

Python was chosen for its extensive libraries, simplicity, and efficiency in machine learning and data science tasks.

Key Libraries and Frameworks:

scikit-learn:

Used for implementing ML models, feature scaling, and PCA.

Tools for model evaluation like confusion matrix, ROC-AUC, and accuracy score.

NumPy and Pandas:

NumPy for numerical computations.

Pandas for data manipulation and preprocessing.

Matplotlib and Seaborn:

For data visualization, including feature importance, distribution analysis, and performance metrics.

Streamlit:

Framework for creating a web-based diagnostic application.

Enabled seamless integration of ML models into an interactive platform for end-users.

Diagnostic Web Application

The system was deployed as a web application using Streamlit, providing a user-friendly interface for remote and real-time PD testing.

Key Features of the Application:

Input Fields:

Users manually enter acoustic feature values (e.g., Jitter, Shimmer) or upload pre-processed voice data.

Real-Time Processing:

The uploaded data is passed through the pre-trained Random Forest model, returning instant predictions.

Error Handling:

Includes validation to ensure input accuracy and avoid errors from invalid data formats.

User Interface:

Styled using CSS and Streamlit theming for an intuitive and visually appealing experience.

Backend and Integration:

Model Serialization

Pre-trained models were serialized using Pickle, allowing the application to directly use these models without retraining.

Deployment:

The application is lightweight, making it deployable on various platforms, including personal computers and cloud services.

Evaluates the number of true positives, false positives, true negatives, and false negatives.

Accuracy:

Measures the proportion of correctly classified samples.

Sensitivity (Recall):

The model’s ability to identify actual PD cases correctly.

ROC-AUC Curve:

Demonstrates the trade-off between sensitivity and specificity.

F1-Score:

Provides a balance between precision and recall for overall performance assessment.

Significance of Tools and Technologies

The selected tools and technologies provide a robust foundation for achieving project goals. Key benefits include:

Accuracy and Reliability:

The Random Forest model's high sensitivity ensures reliable early detection of PD.

Scalability:

Streamlit enables global accessibility, making it suitable for mass screening **programs.**

Cost-Effectiveness:

By relying solely on voice data, the system eliminates the need for expensive imaging and physical tests.

User-Friendly Deployment:

The intuitive web interface ensures accessibility for both healthcare providers and non-technical users

5. LITERATURE REVIEW

Parkinson's disease (PD) is a neurodegenerative disorder marked by motor and non-motor symptoms, with research spanning its pathoanatomy, diagnostics, treatment, and predictive modeling. Braak and Braak (2000) described the progressive degeneration of dopaminergic neurons in the substantia nigra, while Del Tredici and Braak (2008) proposed a neuroanatomical staging system detailing the sequential involvement of brain regions, from the medulla oblongata to cortical areas. Recent advancements in noninvasive diagnostics, as highlighted by Alatas et al. (2022), focus on imaging, biomarkers, and clinical data for early and accurate detection. Machine learning (ML) and deep learning (DL) techniques have also shown promise, with Raundale et al. (2021) demonstrating their potential for early PD diagnosis using datasets from sources like the UCI Machine Learning Repository and the PPMI Database.

Speech impairments, a common symptom of PD, have been explored extensively in diagnostic and therapeutic contexts. Tsanas et al. (2012) developed innovative speech signal processing algorithms for high-accuracy classification, while Little et al. (2007) utilized nonlinear techniques to detect voice disorders. Ramig et al. (2008) emphasized the importance of speech therapy in improving vocal communication. Beyond motor symptoms, PD significantly impacts patients' mental health and quality of life. Schrag et al. (2000) highlighted the strong association between depression and PD, stressing the need for psychological support in patient care. Meanwhile, Prabhavathi and Patil (2022) focused on hallmark motor symptoms such as tremors and bradykinesia, underscoring the need for targeted therapies to enhance motor function and overall well-being.

Together, these studies underscore the multifaceted nature of PD research, highlighting advancements in understanding its pathology, improving diagnostics, and developing holistic approaches to treatment. Despite progress, ongoing efforts remain critical to addressing early detection challenges and enhancing comprehensive care for patients.

6. METHODOLOGY

This section explains the key steps and processes used to develop the Parkinson's Disease diagnostic system, including dataset preparation, feature selection, machine learning model training, and application development**.**

6.1 Dataset

Dataset Overview:

Source:

The dataset was obtained from the MDVP (Multidimensional Voice Program) dataset, available through the PPMI (Parkinson's Progression Markers Initiative) and UCI Machine Learning Repository. The dataset contains 195 voice recordings from 31 participants.

23 participants were diagnosed with Parkinson's Disease (PD), and 8 were healthy individuals.

The dataset includes 22 acoustic features derived from sustained vowel phonation tests.

6.2 Data Preprocessing:

Data Cleaning:

Removed any missing or invalid values.

Feature Scaling:

Used StandardScaler to standardize the range of features, ensuring consistent model performance.

Class Balancing:

The dataset was imbalanced, with more data points for PD-positive cases than healthy cases.

Applied oversampling techniques to ensure an equal number of samples for both classes.

6.3 Model Training and Evaluation

Selected Machine Learning Models:

Three machine learning models were trained and tested:

Logistic Regression:

A basic classification algorithm used as a baseline for comparison.

Support Vector Machine (SVM):

Effective in handling high-dimensional data and finding the optimal boundary between classes.

K-Nearest Neighbors (KNN):

A simple, distance-based classifier that uses the closest neighbors to predict outcomes.

Training Process:

Split the dataset into 75% training and 25% testing subsets.

Applied grid search for hyperparameter tuning (e.g., the number of trees in Random Forest).

Used cross-validation to prevent overfitting.

Evaluation Metrics:

Accuracy: Measures the overall correctness of the model.

Sensitivity (Recall): Evaluates how well the model identifies PD-positive cases.

Confusion Matrix: Displays the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Results:

SVM:

Achieved comparable accuracy (91.83%) after PCA but slightly lower sensitivity (0.94).

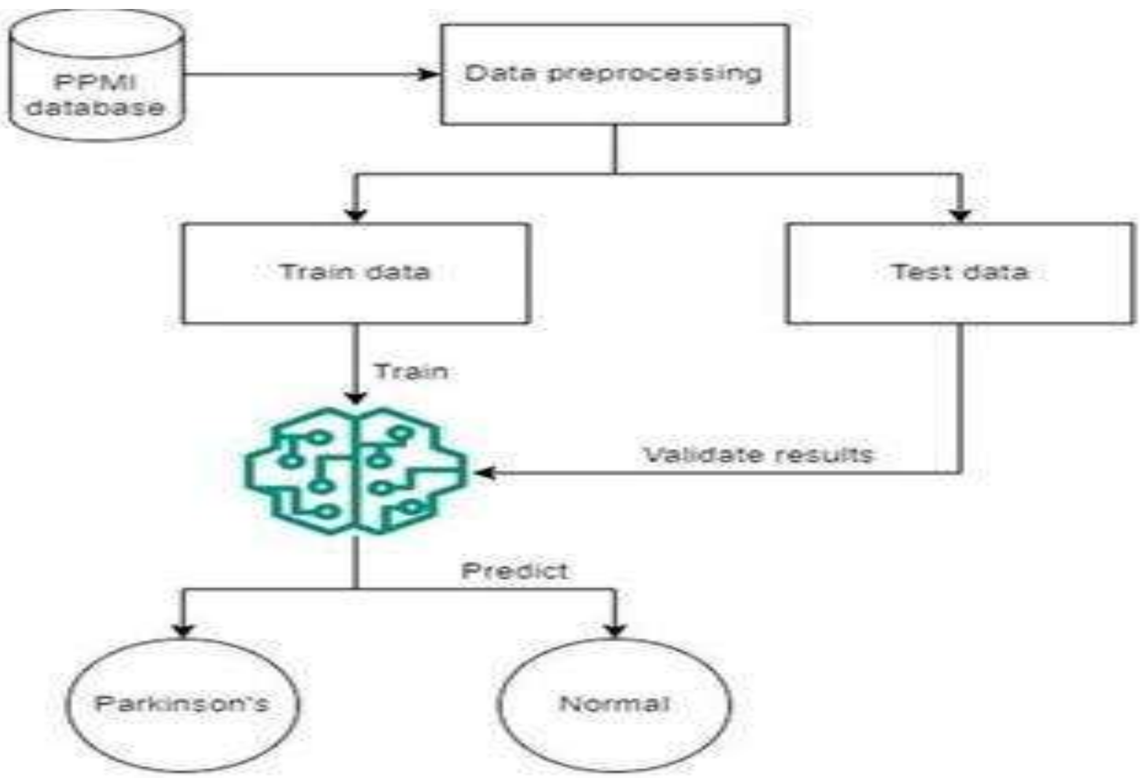
Logistic Regression and KNN:

Performed well but were less accurate and sensitive than Random Forest.

6.4 Application Development

A web application was developed using the Streamlit framework to make the diagnostic system user-friendly and accessible.

7. DIAGRAM



Vol. 05, Issue 01, January 2025, pp : 1818-1830
e-ISSN : 2583-1062
Impact Factor : 7.001

PROPOSED ARCHITETURE

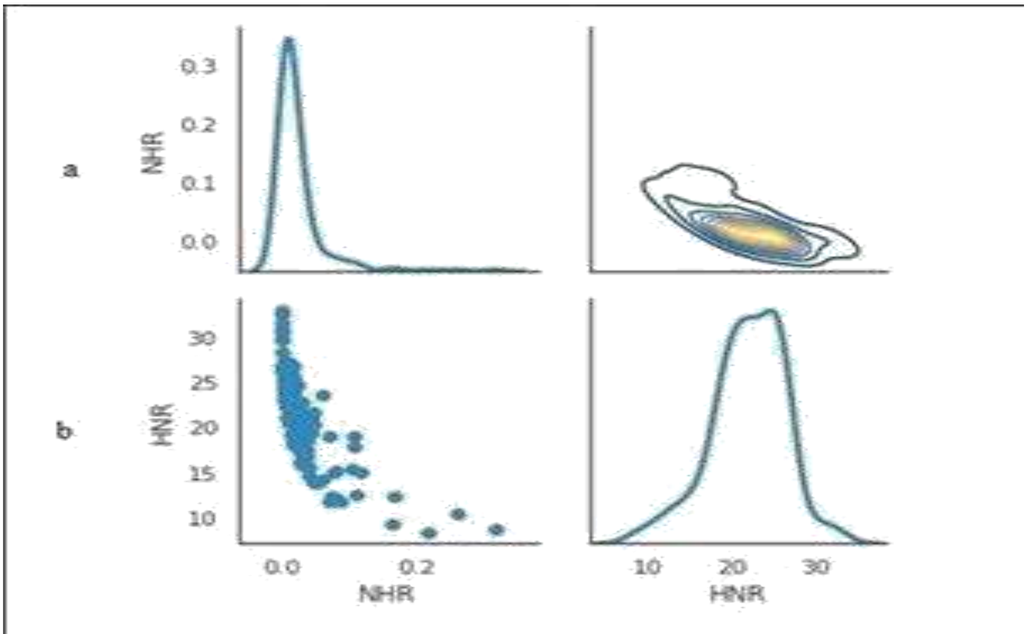
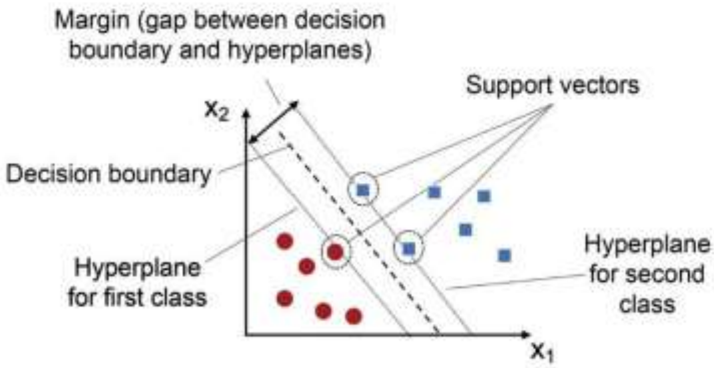


Figure 3. (a) NHR plot; (b) HNR plot

NHR AND HNR PLOT



SVM

LOGISTIC REGRESSION CLASSIFIER

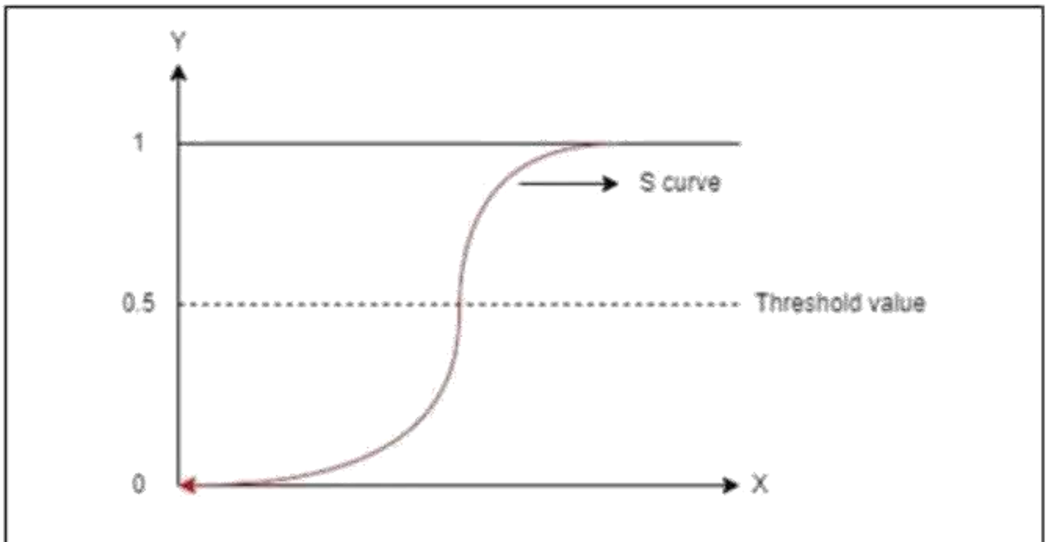
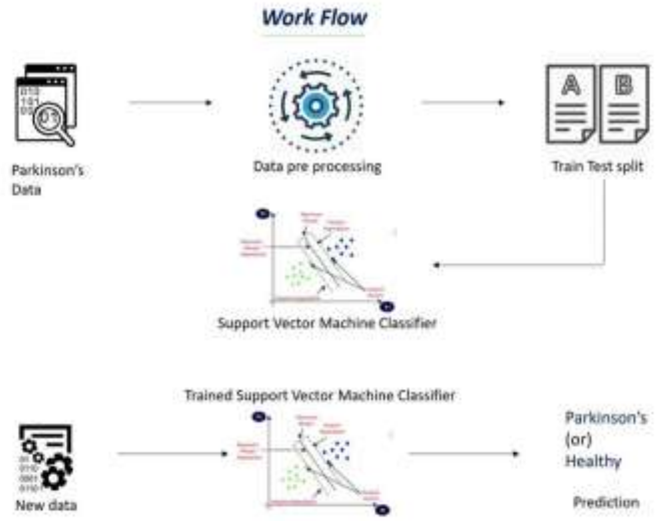


Figure 6: Logistic regression classifier

8. WORKFLOW



The Parkinson's Disease Prediction Project outlines the structured workflow to develop an accurate and efficient diagnostic tool. It starts with the collection of biomedical voice data based on relevant features from trusted sources such as the UCI Machine Learning Repository and the PPMI database. The dataset contains key features such as jitter, shimmer, and the noise-to-harmonic ratio (NHR), which are essential to the detection of voice impairments associated with Parkinson's disease. The preprocessing of the data, during the raw data preparation for training, takes care of missing values, normalizes the data, and also takes care of any imbalance that would help improve model performance. Then comes exploratory data analysis, which is essential in understanding feature distributions and relationships; visualization tools like Matplotlib and Seaborn can be used to determine correlations between features and even identify outliers. For feature selection, PCA is used to reduce dimensionality with critical features. This makes the dataset more efficient for modeling. In model training, various machine learning models such as SVM, Logistic Regression, and KNN are trained on the data. The dataset is divided into training and testing subsets, and performance is measured using metrics like accuracy, precision, and sensitivity. The evaluation stage of the model chooses the best model according to these measures. Then, the application development phase develops a user-friendly interface using Streamlit. It allows users to upload their voice data and retrieve real-time predictions from the trained model. The model deployment is done either by opening the system locally with Streamlit or on cloud-based systems such as AWS, GCP, or Heroku for scalability. Finally, the project focuses on feedback and continuous improvement, where user feedback is collected to refine both the interface and prediction accuracy. The model is periodically retrained with new data to enhance its performance, and future plans include incorporating additional biomarkers and exploring advanced deep learning models. This comprehensive workflow ensures a robust Parkinson's disease diagnostic system that evolves to meet real-world needs.

9. CONCLUSION

This paper highlights the potential of machine learning in revolutionizing the diagnosis of Parkinson’s disease. By utilizing voice biomarkers and integrating ML models into a telemedicine platform, the proposed system provides a cost-effective, accessible, and efficient solution. Future developments aim to incorporate additional biomarkers, such as gait analysis, and deploy the application on cloud platforms for global accessibility. Expanding the dataset to include diverse demographics and more voice attributes could further enhance model accuracy and reliability. Partnering with healthcare institutions for validation with real-world patient data is essential to ensure practical applicability. The use of wearable devices for continuous monitoring can complement this system by tracking disease progression in real-time. Incorporating deep learning techniques could uncover more complex patterns in the data, pushing diagnostic accuracy to new heights. The integration of this tool with electronic health record (EHR) systems can streamline data collection and improve patient management. Supporting multi-language functionality would increase global accessibility, especially in underserved regions. Extending the system to diagnose other neurodegenerative disorders can make it a versatile diagnostic platform. Moreover, hosting the application on scalable cloud platforms can handle a higher user base and ensure data security. Through continuous improvement and collaboration with medical experts, this system can serve as a cornerstone in personalized healthcare. Its ability to provide non-invasive, remote, and reliable diagnostics is a testament to the transformative power of machine learning in modern medicine.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 01, January 2025, pp : 1818-1830
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

10. FUTURE WORK

Future work for this Parkinson's Disease Prediction system can focus on several key improvements, particularly enhancing the integration between the database and the machine learning model. One area for improvement is to optimize database queries for better performance, especially when dealing with large datasets, by introducing indexing or more efficient querying mechanisms. The system could also expand its data handling capabilities to accommodate more complex patient records, such as adding additional biomarkers or medical history that could improve prediction accuracy.

The database architecture can be improved to handle data more dynamically, for example, by introducing real-time data insertion or updates, which would allow the model to be trained periodically with new patient data. This can be paired with a model retraining pipeline that automatically updates the machine learning model when new data is added, ensuring the system stays accurate over time.

Additionally, the web interface could be extended to allow healthcare professionals to view not only the predictions but also detailed reports on the patient's historical data and the features used for prediction. Implementing authentication and user access controls would improve data security, ensuring that only authorized personnel can access patient data and model predictions.

To further enhance usability, integrating the system with Electronic Health Record (EHR) systems would streamline the prediction process by allowing seamless patient data imports directly from healthcare institutions. Furthermore, improving the error handling and data validation processes would ensure more reliable interactions with the database and the model. Overall, these enhancements would make the system more robust, secure, and effective in diagnosing Parkinson's disease at scale.

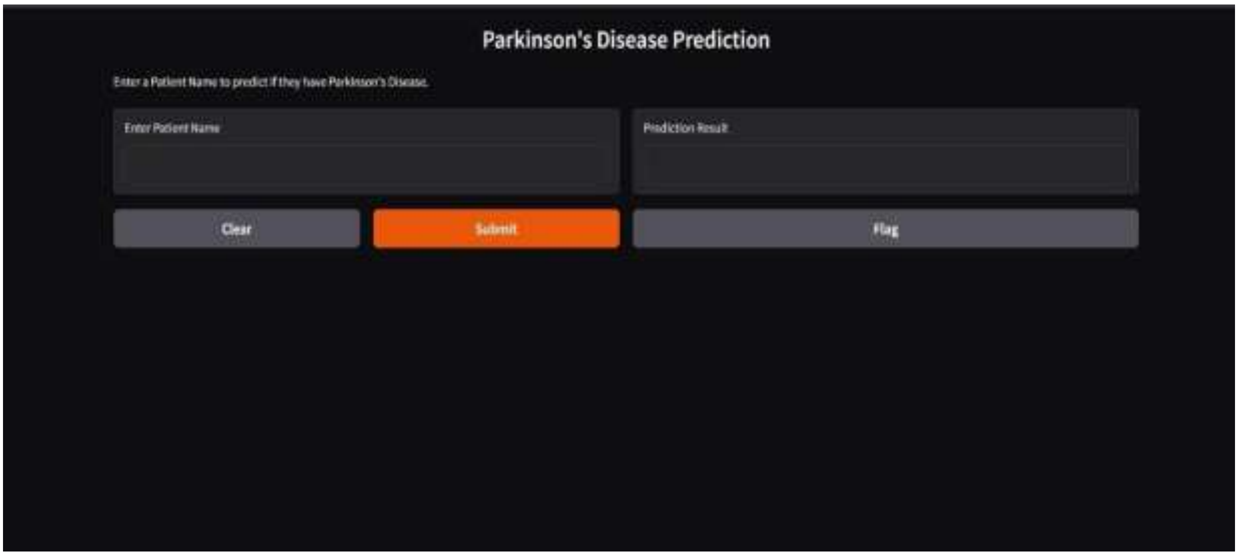
11. APPENDIX

PROGRAM:

```
import os
import pickle
import gradio as gr
# Load the saved model
parkinsons_model = pickle.load(open(r'C:\Users\jheev\Music\New folder\parkinsons_model.sav', 'rb'))
# Define the prediction function
def predict_parkinsons(fo, fhi, flo, Jitter_percent, Jitter_Abs, RAP, PPQ, DDP, Shimmer, Shimmer_dB, APQ3, APQ5, APQ, DDA, NHR, HNR, RPDE, DFA, spread1, spread2, D2, PPE):
try:
# Convert inputs to float
user_input = [float(fo), float(fhi), float(flo), float(Jitter_percent), float(Jitter_Abs), float(RAP), float(PPQ), float(DDP), float(Shimmer), float(Shimmer_dB), float(APQ3), float(APQ5), float(APQ), float(DDA), float(NHR), float(HNR), float(RPDE), float(DFA), float(spread1), float(spread2), float(D2), float(PPE)]
# Make prediction
prediction = parkinsons_model.predict([user_input])[0]
if prediction == 1:
return "The person has Parkinson's disease"
else:
return "The person does not have Parkinson's disease"
except ValueError:
return "Please enter valid numerical values for all fields."
# Create Gradio interface
with gr.Blocks(theme="default") as interface:
gr.Markdown("<h1 style='color:#bc5090;'>Parkinson's Disease Prediction using ML</h1>")
#	Organize inputs into a grid layout with gr.Row():
fo = gr.Textbox(label="MDVP:Fo(Hz)") fhi = gr.Textbox(label="MDVP:Fhi(Hz)") flo = gr.Textbox(label="MDVP:Flo(Hz)")
```

```
Jitter_percent = gr.Textbox(label="MDVP:Jitter(%)") Jitter_Abs = gr.Textbox(label="MDVP:Jitter(Abs)")
with gr.Row():
RAP = gr.Textbox(label="MDVP:RAP")
PPQ = gr.Textbox(label="MDVP:PPQ")
DDP = gr.Textbox(label="Jitter:DDP")
Shimmer = gr.Textbox(label="MDVP:Shimmer")
Shimmer_dB = gr.Textbox(label="MDVP:Shimmer(dB)")
with gr.Row():
APQ3 = gr.Textbox(label="Shimmer:APQ3")
APQ5 = gr.Textbox(label="Shimmer:APQ5")
APQ = gr.Textbox(label="MDVP:APQ")
DDA = gr.Textbox(label="Shimmer:DDA")
NHR = gr.Textbox(label="NHR")
with gr.Row():
HNR = gr.Textbox(label="HNR")
RPDE = gr.Textbox(label="RPDE")
DFA = gr.Textbox(label="DFA")
spread1 = gr.Textbox(label="spread1")
spread2 = gr.Textbox(label="spread2")
with gr.Row():
D2 = gr.Textbox(label="D2")
PPE = gr.Textbox(label="PPE")
# Prediction button and output display
predict_button = gr.Button("Parkinson's Test Result")
output = gr.Textbox(label="Diagnosis Result", interactive=False)
#     Link inputs, button, and output to the prediction function predict_button.click(
predict_parkinsons,
inputs=[fo, fhi, flo, Jitter_percent, Jitter_Abs, RAP, PPQ, DDP, Shimmer, Shimmer_dB, APQ3, APQ5, APQ, DDA, NHR, HNR, RPDE, DFA, spread1, spread2, D2, PPE],
outputs=output
)
Launch the Gradio app
interface.launch('share=True')
```

OUTPUT



12. REFERENCE

[1] Alatas, B., et al. “Noninvasive Diagnostics Biomarkers in Parkinson’s Disease.” BioMed Research International, 2022.

[2] Braak, H., Braak, E. “Pathoanatomy of Parkinson’s Disease.” Journal of Neurology, 2000.

[3] Del Tredici, K., Braak, H. “Review: Neuroanatomical Staging of Parkinson’s Disease.” Acta Neuropathologica, 2008.

[4] Little, M.A., et al. “Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection.” BioMed Engineering Online, 2007.

[5] Prabhavathi, K., Patil, S. “Tremors and Bradykinesia.” Springer, 2022.

[6] Ramig, L.O., et al. “Speech Treatment for Parkinson’s Disease.” Expert Reviews in Neurotherapeutics, 2008.

[7] Raundale, P., et al. “Prediction of Parkinson’s Disease Using ML and DL Algorithms.” INCET, 2021.

[8] Schrag, A., et al. “Quality of Life and Depression in Parkinson’s Disease.” Movement Disorders, 2000.

[9] Tsanas, A., et al. “Novel Speech Signal Processing Algorithms for High-Accuracy Classification of

[10] Parkinson’s Disease.” IEEE Transactions on Biomedical Engineering, 2012.

[11] Dataset References: UCI Machine Learning Repository, PPMI Database.