IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

ParkinSense: Deep Learning for Early Parkinson's Detection

Mr.Y.Manas Kumar[1], Chollangi Sathya Sri[2], Kancharla Prasanthi[3], Dunga Pranai[4], Pilli Ravi Santhosh[5], Mondepu Bhargavi[6]

manas.yogi@gmail.com[1] ,chsathyasri2505@gmail.com[2] ,kancharlaprashu@gmail.com[3] , dpranai820@gmail.com[4] , prvvsanthosh55555@gmail.com[5] , bhargaviim8@gmail.com[6]

Pragati Engineering College, 1-378, ADB Road, Surampalem Near Kakinada, Surampalem, Andhra Pradesh 533440

Abstract:

Parkinson's Disease (PD) is a progressive neurological disorder that impacts millions worldwide. Early detection is essential for managing its progression and improving patient outcomes. This paper presents a graphical user interface (GUI) application designed to detect PD using machine learning (ML) techniques. The application allows users to upload biomedical voice measurement datasets, preprocess the data by handling missing values, and apply the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. It also includes exploratory data analysis (EDA) to visualize data distribution and relationships.The system implements six machine learning algorithms—Decision Tree, Random Forest, Logistic Regression, Support Vector Machine (SVM), Naive Bayes, and k-Nearest Neighbors (KNN)—to classify PD. Models are evaluated based on accuracy, precision, recall, and F1 score, with results displayed through confusion matrices and comparison graphs. This interactive tool offers an accessible platform for researchers, clinicians, and data scientists to efficiently explore machine learning techniques for PD diagnosis.By providing detailed performance insights, the application helps identify the most effective algorithm for specific datasets. With its user-friendly interface and comprehensive functionality, this tool supports early detection efforts, potentially enhancing patient care through more accurate and timely diagnosis of Parkinson's Disease.

Keywords: Parkinson's Disease, Machine Learning, GUI, Biomedical Voice Measurements, EDA, SMOTE, Classification Algorithms, Model Evaluation, Early Detection, Healthcare Technology.

Introduction:

Parkinson's Disease (PD) is a chronic neurological disorder characterized by the degeneration of dopamine-producing neurons in the substantia nigra, a brain region responsible for movement control. Symptoms such as tremors, bradykinesia, rigidity, and postural instability progressively worsen, impairing motor functions and quality of life. Though the exact cause is unknown, both genetic and environmental factors are believed to contribute. PD affects approximately 1% of people over 60, with rising prevalence due to global aging.**Early and accurate diagnosis** is crucial for effective management. However, traditional methods relying on clinical assessments and patient history are often subjective and prone to error. Advancements in **machine learning (ML)** provide new opportunities by identifying subtle disease markers from biomedical data, such as voice recordings and movement measurements, enhancing diagnostic accuracy and reducing misdiagnosis.This project aims to develop an **ML-based application** for early PD detection. The system allows users to upload voice datasets, perform exploratory data analysis (EDA), and apply preprocessing techniques like handling missing values and class imbalance with **SMOTE**. Six ML models—Decision Tree, Random Forest, Logistic Regression, SVM, Naive Bayes, and KNN—are implemented and evaluated using metrics like accuracy, precision, recall, and F1 score. The **Random Forest model** provided the highest accuracy, making it the preferred choice for predictions.This interactive platform offers an accessible tool for clinicians and researchers, supporting early PD diagnosis and improving patient outcomes.

Literature review:

[1] Mohammad S Islam et al. conducted a comparative analysis to detect Parkinson's Disease using various classifiers. Support Vector Machine (SVM), Feedforward Back-Propagation Based Artificial Neural Network (FBANN), and Random Tree (RT) classifiers were used, and a comparison between them was made to differentiate between PD and healthy patients. The study utilized the UCI Machine Learning Repository. The dataset consisted of 195 voice samples from 31 individuals, comprising both males and females. Among them, 23 were determined to have PD, and 8 were healthy. To improve classification accuracy with minimal error rate, 10-fold cross-validation repeated 100 times was implemented for all three classifiers. The FBANN classifier achieved a 97.37% recognition accuracy, outperforming the other two classifiers.[2] R. Arefi Shirvan et al. proposed a system to detect PD using the KNN method. KNN is a simple method for grouping based on similarity and is used when data distribution facts are insufficient. The method involves: a) determining K close neighbors and b) classifying based on these neighbors. The study achieved 93.7% accuracy with 4 optimized features, 94.8% accuracy with 7 features, and 98.2% accuracy with 9 features, surpassing other studies. The UCI Repository dataset included 192 voice samples from 32 males and females, with 23

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

suffering from PD and the rest being healthy. Subjects were aged between 46 and 85 years. The main drawback of KNN is that it is a lazy learner, meaning it does not learn from the training data itself but classifies based on stored training instances.[3] K. Srilatha and V. Ulagamuthalvi et al. conducted a comparative study on tumor classification, highlighting the importance of classification in computer vision. Image classification refers to labeling images into predefined categories through processes such as image sensors, pre-processing, object detection, segmentation, feature extraction, and object classification. The study employed various classifiers combined with segmentation algorithms for tumor detection using image processing techniques.[4] S. R. Khonde and V. Ulagamuthalvi et al. proposed a feature selection method based on the average probability (AP) score of each feature. Features with low AP scores were removed from the training and testing sets. Performance parameters used were true positive, true negative, and accuracy. Various semi-supervised classifiers were employed for intrusion detection using the NSL-KDD dataset.[5] Resul Das et al. applied various classification methods to identify PD, including DMneural, neural networks, regression, and decision tree techniques. Multiple evaluation methods were used to assess classifier performance. The neural networks classifier yielded the best results. The input dataset was randomly partitioned into 65% training and 35% testing data. The adjustable parameters of each classifier were tuned. The BPNN algorithm was used in the feed-forward, single hidden layer neural network. The Levenberg–Marquardt (LM) algorithm was employed, achieving 92.9% accuracy.[6] Mercy Paul Selvam et al. used machine learning techniques to predict student dropout using data mining. The model employed a decision tree classifier, achieving 97.69% accuracy. The prediction was based on various parameters considered for each student.[7] Dr. R. Geetha Ramani et al. proposed a system to classify PD and non-PD patients using binary logistic regression, linear discriminant analysis (LDA), partial least squares regression (PLS), random tree (Rnd Tree), and SVM. The dataset was obtained from the UCI Repository and consisted of 197 samples with 22 features extracted from patients. The Fisher filtering feature selection algorithm was used for effective feature ranking. The Rnd Tree algorithm achieved 100% classification accuracy, while LDA, C4.5, CS-MC4, and KNN yielded accuracy results above 90%. The C-PLS algorithm achieved the lowest accuracy of 69.74%.[8] Anchana Khemphila et al. utilized a Multi-Layer Perceptron (MLP) with the back-propagation learning algorithm for PD classification. The study used the UCI Machine Learning Repository's PD dataset. Experiments were performed using Weka 3.6.6, with information gain used to filter features. Since continuous real numbers were used, the range of values was partitioned for discretization. The training dataset achieved 91.453% accuracy, while the validation dataset achieved 80.769% accuracy using the main model.

Proposed system:

To address the limitations of traditional diagnostic methods, the proposed system leverages machine learning algorithms for the early detection and diagnosis of Parkinson's Disease. This system integrates various data sources and analytical tools to provide a comprehensive and data-driven diagnostic solution. The key components of the proposed system include:

- **Data Integration and Preprocessing:** Collection of relevant datasets (e.g., voice recordings, movement measurements) and preprocessing steps such as handling missing values, normalization, and addressing class imbalances using techniques like SMOTE.
- **Exploratory Data Analysis (EDA):** Visualization tools for exploring data distributions, identifying correlations, and detecting patterns that inform model selection and preprocessing.
- **Machine Learning Algorithms:** Implementation of multiple machine learning models including Decision Trees, Random Forests, Logistic Regression, SVM, Naive Bayes, and KNN for PD detection. These models analyze the preprocessed data to identify patterns and markers indicative of Parkinson's Disease.
- **Model Training and Evaluation:** Training the models on the dataset and evaluating their performance using metrics such as accuracy, precision, recall, F1 score, and confusion matrices.
- **User Interface:** A graphical user interface (GUI) that allows users to load data, run analyses, visualize results, and compare the performance of different machine learning models.

Results:

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume 13, Issue 1, 2025

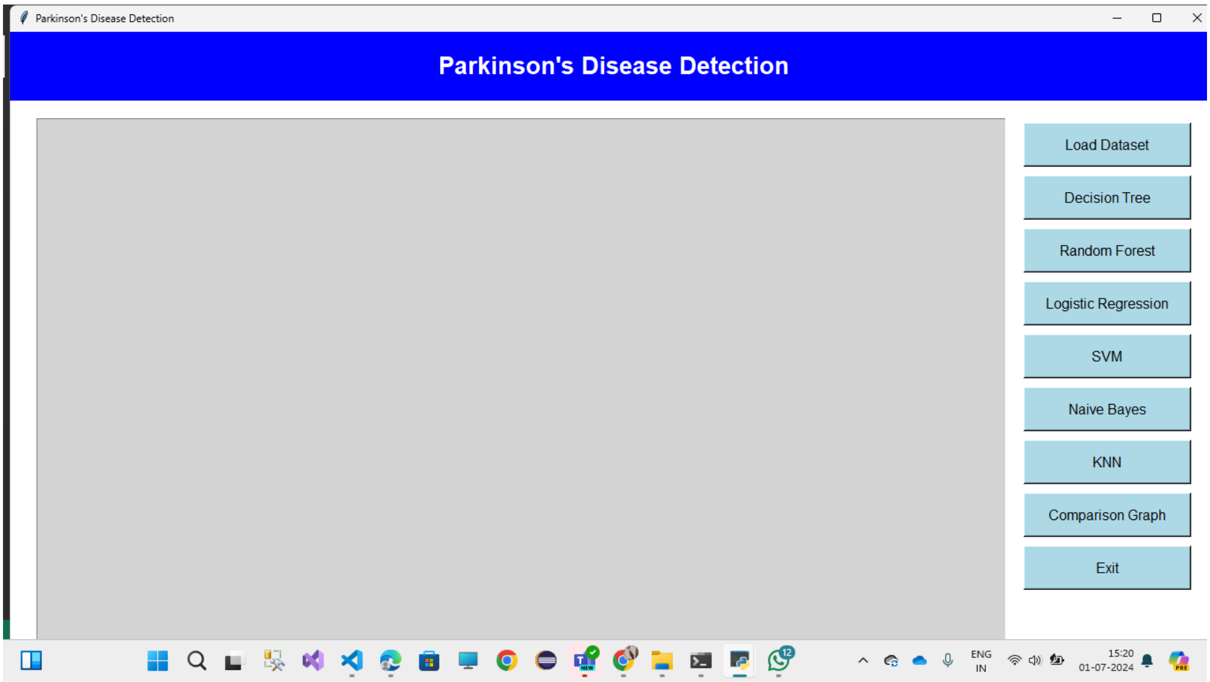

Fig.1 Output of the Application

With the fewest vocal features necessary to diagnose Parkinson's, 93.84% accuracy was attained. The results of the methods provided by Gil and Manuel (2009) based on artificial neural networks and support vector machines to aid specialists in the diagnosis of Parkinson's disease indicate a high accuracy of about 90%

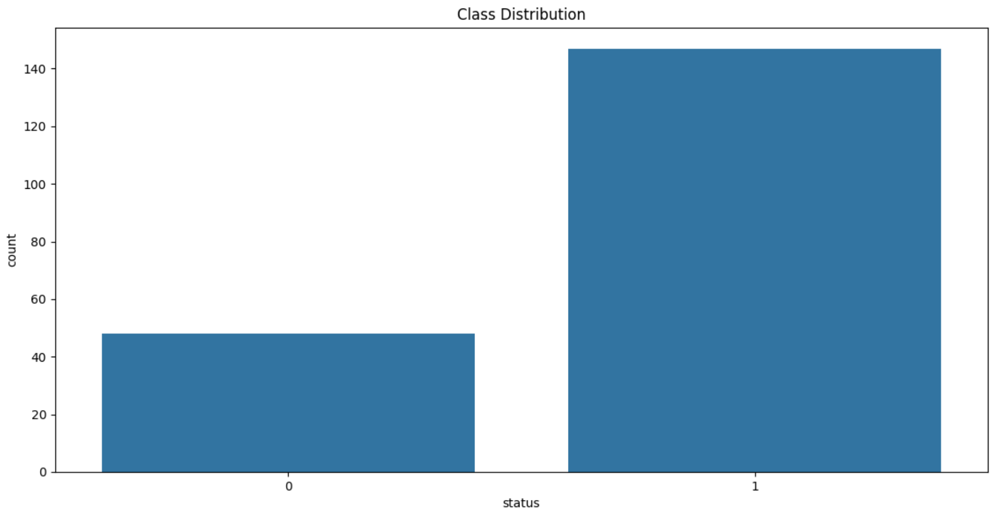

Fig.2 Class Distribution For the Load Data set

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 1, 2025

The image shows a bar chart representing the distribution of two classes. Class 1 is significantly more represented than class 0. This indicates that the dataset is imbalanced, with one class dominating the other.

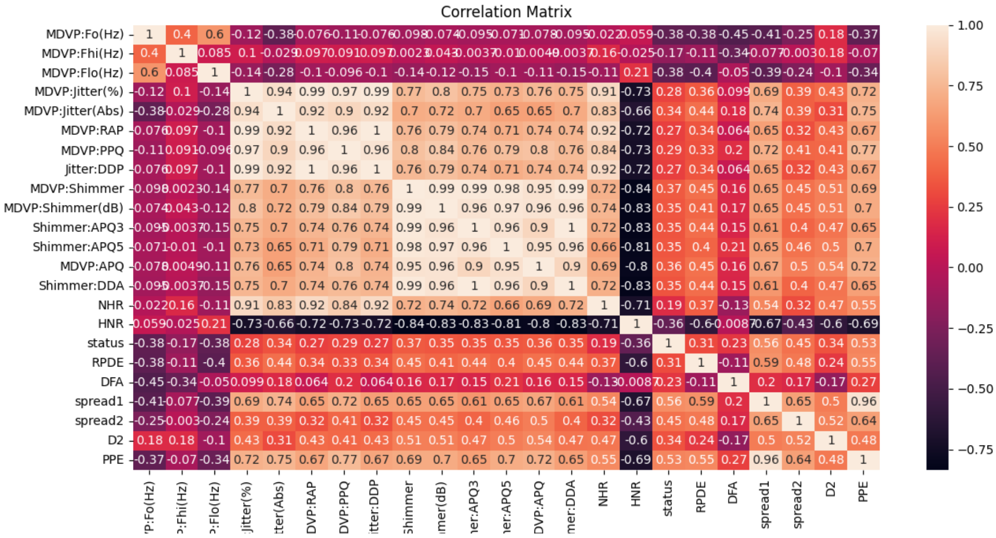

Fig.3 Correlation matrix

The image shows a correlation matrix visualized as a heatmap, representing the relationships between different features in a Parkinson's disease dataset. The color gradient ranges from dark purple (negative correlation) to bright orange (positive correlation), with values closer to 1.0 indicating a strong positive correlation and values near -1.0 showing a strong negative correlation. Features such as MDVP:Jitter(%), MDVP:Shimmer, and Shimmer:APQ3 display high inter-correlation, indicating potential redundancy. The matrix reveals distinct clusters of highly correlated features, which might indicate multicollinearity. The color bar on the right provides a reference for interpreting the correlation strength.

Conclusion:

Leveraging machine learning techniques like Naive Bayes and k-Nearest Neighbors (k-NN) significantly enhances the accuracy and early diagnosis of Parkinson's disease by analyzing datasets containing patient symptoms, voice recordings, and other biomarkers. Naive Bayes, known for its computational efficiency, handles high-dimensional data and small sample sizes effectively, making it suitable for identifying patterns even with missing or undefined features. Despite its simplicity and the assumption of feature independence, it has shown promising results in early-stage diagnosis by analyzing individually correlated features. On the other hand, k-NN captures local patterns by measuring the proximity of new instances to labeled examples, making it effective for non-linear decision boundaries, though it can be computationally expensive due to distance calculations. Both algorithms benefit from proper preprocessing, such as normalization and feature selection, to enhance predictive performance. Hybrid models combining these techniques with advanced methods like neural networks or support vector machines can further improve accuracy and robustness. While many existing models use classifiers like SVM, ANN, and KNN, ensemble techniques have shown superior performance. This predictive approach aids in identifying Parkinson's symptoms early, potentially reducing death rates by enabling timely intervention. Although the model was not deployed, future plans include creating a web app that takes voice recordings as input and detects speech inconsistencies indicative of Parkinson's, making diagnosis more accessible for elderly individuals. This would reduce the need for frequent hospital visits, allowing for remote, early symptom detection. Additionally, incorporating deep learning algorithms in future iterations could enhance the model's accuracy and reliability, contributing to more effective and accessible Parkinson’s disease diagnosis.

Future scope:

Future research and development in Parkinson’s disease prediction should focus on several key areas to enhance the accuracy, reliability, and clinical applicability of predictive models. Integrating multi-modal data, such as genetic information, imaging data (MRI, PET scans), and detailed clinical records with traditional biomarkers like voice recordings and motor assessments, could offer a more comprehensive understanding of the disease’s progression, leading to improved prediction and personalized treatment strategies. Advanced feature engineering techniques should be developed to capture subtle motor and non-motor changes, enhancing model sensitivity and specificity, particularly for early-stage diagnosis. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), can be explored for analyzing complex data patterns and temporal sequences, enabling more accurate predictions by identifying intricate disease markers. Ensuring explainability and interpretability through explainable AI (XAI) techniques will be crucial for building clinician trust and facilitating seamless integration into clinical decision-making. Real-time monitoring using wearable sensors and mobile health technologies could enable continuous symptom tracking, providing dynamic risk assessments and timely interventions. Model personalization, by tailoring predictive models to individual patients based on their unique clinical profiles, could improve treatment efficacy and patient outcomes. Large-scale validation studies across diverse populations will be necessary to confirm model generalizability, identify biases, and ensure robustness. Integration with electronic health records (EHRs) can further streamline clinical workflows, facilitating early detection during routine check-ups and enhancing data-driven decision-making. Addressing ethical and regulatory considerations, such as data privacy, informed consent, and compliance with healthcare standards, will be essential for the responsible deployment of predictive models. Finally, longitudinal studies that track the disease over time can provide insights into Parkinson's progression, enabling models to predict not only diagnosis but also disease trajectory, leading to more effective early interventions and personalized management plans.

References:

[1] Islam, M. S., et al. (n.d.). *Comparative analysis to detect Parkinson's disease using various classifiers*. UCI Machine Learning Repository.

[2] Shirvan, R. A., et al. (n.d.). *Parkinson's disease detection using K-Nearest Neighbors (KNN) method*. UCI Machine Learning Repository.

[3] Srilatha, K., & Ulagamuthalvi, V. (n.d.). *Comparative study on tumor classification using image processing techniques*.

[4] Khonde, S. R., & Ulagamuthalvi, V. (n.d.). *Feature selection based on average probability score for intrusion detection*. NSL-KDD dataset.

[5] Das, R., et al. (n.d.). *Classification methods for identifying Parkinson's disease*. UCI Machine Learning Repository.

[6]] Selvam, M. P., et al. (n.d.). *Predicting student dropout using machine learning techniques*.

[7] Ramani, R. G., et al. (n.d.). *Classification of Parkinson's disease using machine learning techniques*. UCI Machine Learning Repository.

[8] Khemphila, A., et al. (n.d.). *Multi-layer perceptron with back-propagation learning algorithm for Parkinson's disease classification*. UCI Machine Learning Repository.