RAJALAKSHMI ENGINEERING COLLEGE
An AUTONOMOUS Institution
Affiliated to ANNA UNIVERSITY, Chennai

25 years OF ACADEMIC EXCELLENCE

Department of Computer Science and Engineering

Predictive Modelling for Parkinson's Disease Diagnosis Using Biomedical Voice Features

Mrs. M. Divya M.E

2116220701232
Rudrapriyan N

Problem Statement and Motivation

- Parkinson's Disease (PD) is a long-term neurodegenerative problem which has a profound impact on motor abilities and speech.
- Early diagnosis is problematic because the very early symptoms are often subtle and subjective clinical examinations are expensive. Voice-based biomarkers are an interesting non-invasive diagnostic approach, but robust models are required to interpret subtle variations in the voice. An accurate, interpretable and computationally efficient AI system for early PD revelation based on voice recordings.

Motivation

- Parkinson's disease is difficult to detect early with traditional methods.
- Voice analysis offers a non-invasive, low-cost diagnostic alternative.

Existing System

- **Clinical Diagnosis:**

1. Conventional method grounded in physical examination and subjective evaluations from neurologists.
2. Often results in delayed, or even wrong diagnosis, particularly in an early stage.

- **Machine Learning Approaches:**

1. Algorithms used: Support Vector Machines, Decision Trees, Logistic Regression, Random Forest, XGBoost, AdaBoost and Ensemble Methods.
2. Features: jitter; shimmer; RPDE; PPE; DFA; (Voice signal biomarkers).
3. Random Forest & XGBoost: ~94.8% accuracy

Objectives

- To develop an accurate and interpretable system for early detection of Parkinson’s disease using vocal biomarkers.
- To evaluate and compare the performance of traditional machine learning models and ensemble techniques for PD classification, with LightGBM.
- To implement and optimize TabNet, a deep learning model tailored for tabular data, to improve diagnostic accuracy.

Abstract

- Parkinson's Disease (PD) is a neurodegenerative disorder with challenges in early diagnosis due to subtle symptoms and reliance on subjective clinical evaluations. This project proposes a hybrid system that combines traditional machine learning models and LightGBM, along with TabNet, a deep learning architecture designed for tabular data. Using the UCI PD dataset based on vocal biomarkers, the models were preprocessed and trained with hyperparameter tuning for optimal performance. The system emphasizes model interpretability and clinical applicability, supporting early, accurate, and non-invasive detection of PD.

Proposed System

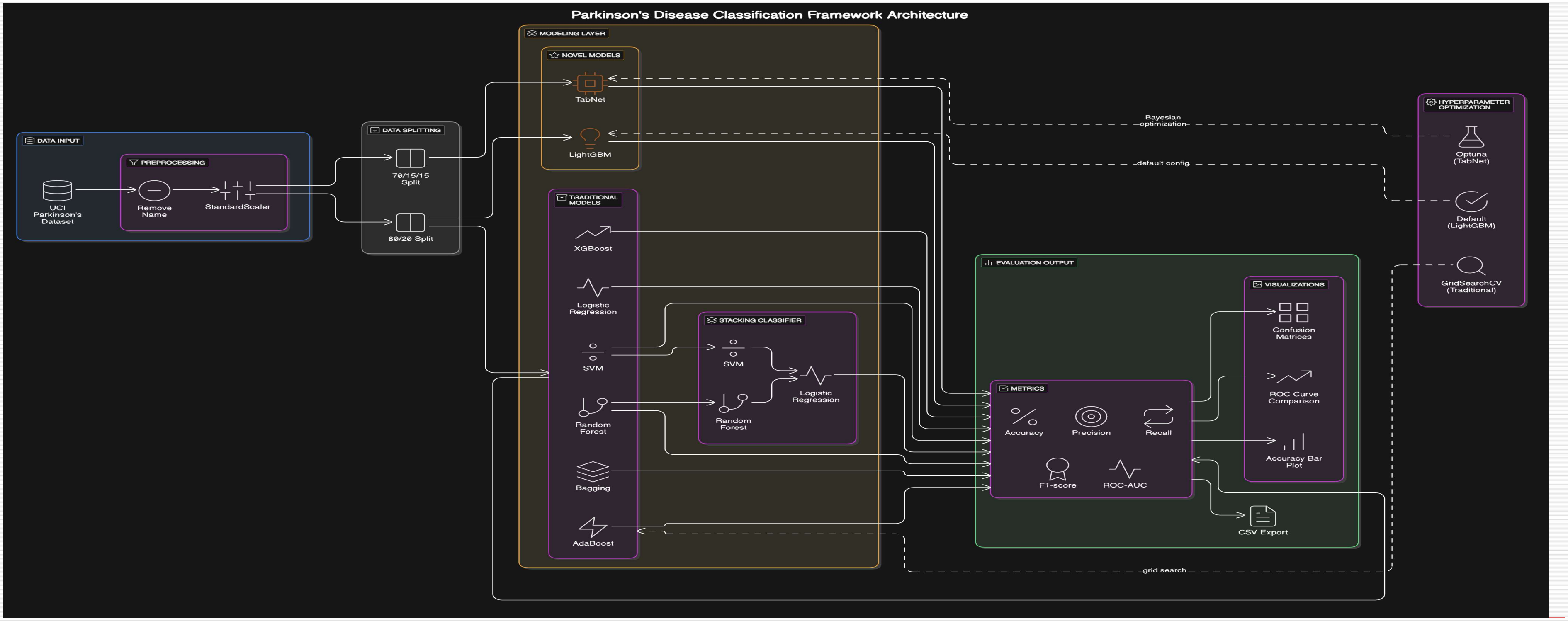
1. **LightBGM**: Efficient implementation of decision tree-based learning.

Key Features : Fast training, low memory usage, supports categorical features.

2. **TabNet**: Uses attention-based neural networks, designed specifically for tabular data**.** Key Features: Interpretable, flexible, and efficient.

Optimization: Hyperparameter tuning via **Optuna** to achieve best model performance.

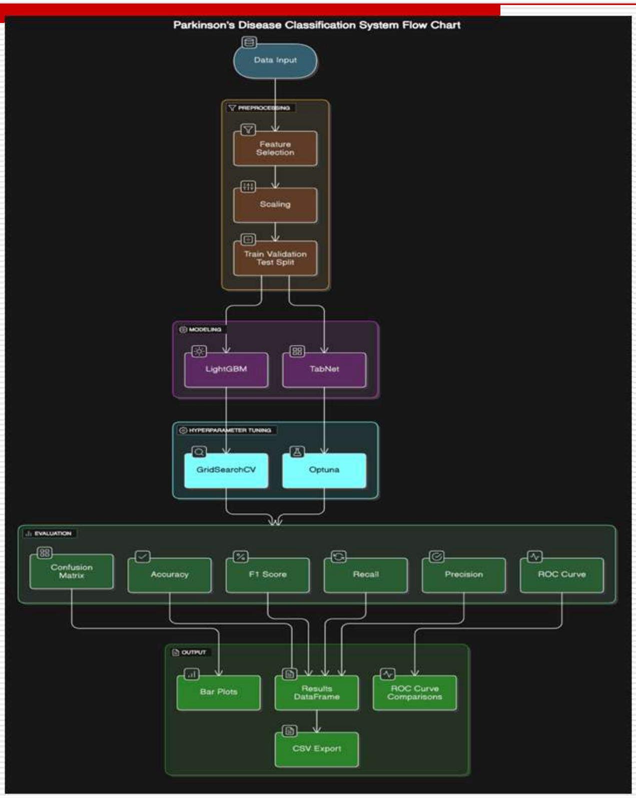
System Architecture

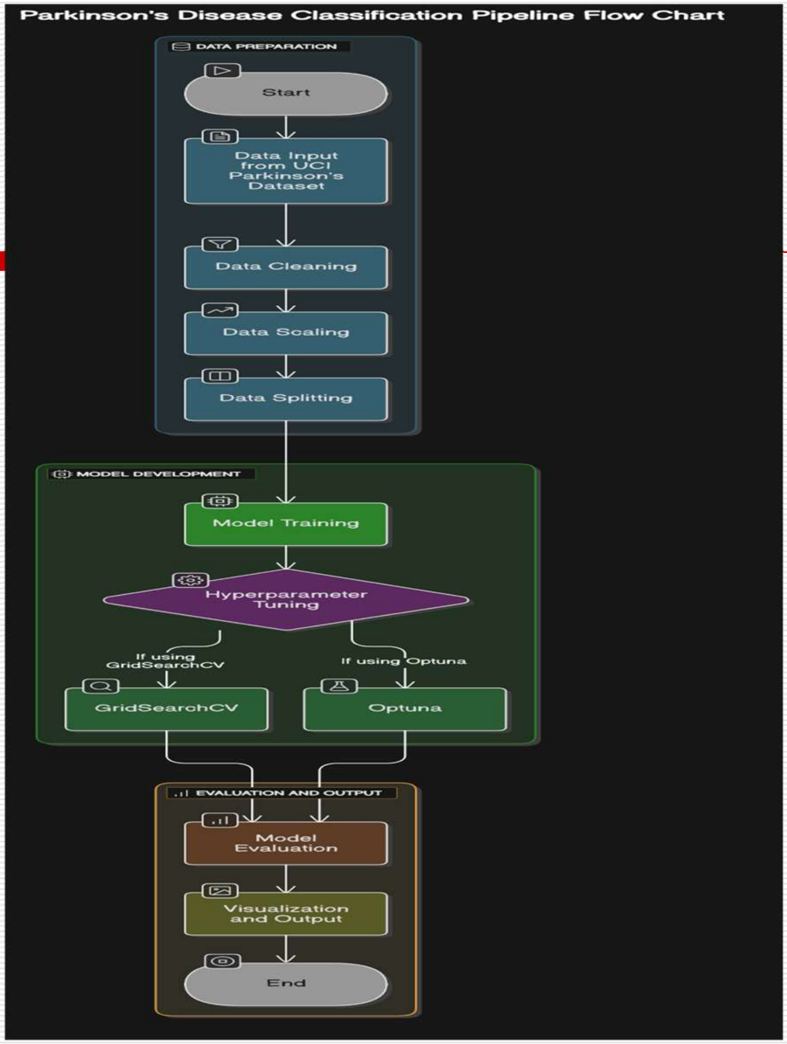


List of Modules

- **Data Input & Preprocessing:** Loads and preprocesses the dataset, splits data into training, validation, and test sets, and applies feature scaling.
- **Modeling:** Implements various classifiers including Logistic Regression, Random Forest, SVM, XGBoost, LightGBM, AdaBoost, Bagging, and Stacking, TabNet.
- **Hyperparameter Tuning:** Uses GridSearchCV for traditional models and Optuna for TabNet to find the best hyperparameters.
- **Model Evaluation:** Evaluates models using accuracy, precision, recall, F1-score, confusion matrix, and ROC curve, Output & Visualization.

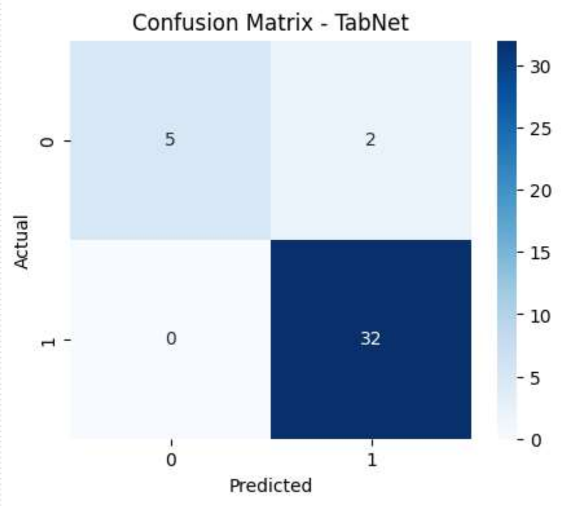
Functional Description for each modules with DFD and Activity Diagram



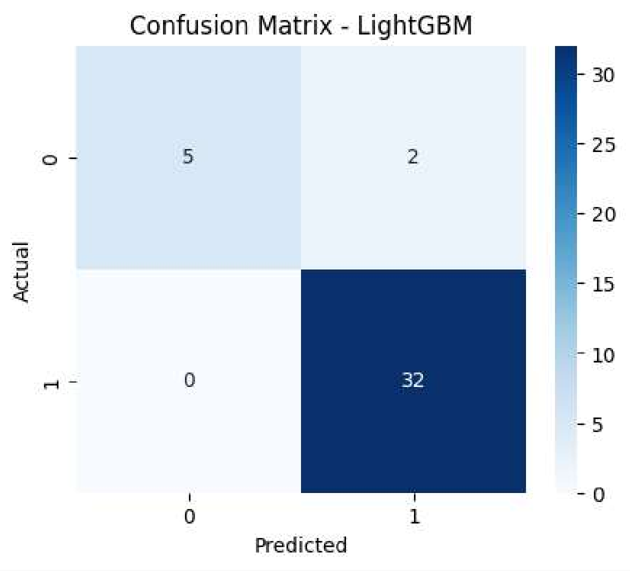

Parkinson's Disease Classification Pipeline Flow Chart
DATA PREPARATION
Start
Data Input from UCI Parkinson's Dataset
Data Cleaning
Data Scaling
Data Splitting
MODEL DEVELOPMENT
Model Training
Hyperparameter Tuning
If using GridSearchCV
If using Optuna
GridSearchCV
Optuna
EVALUATION AND OUTPUT
Model Evaluation
Visualization and Output
End

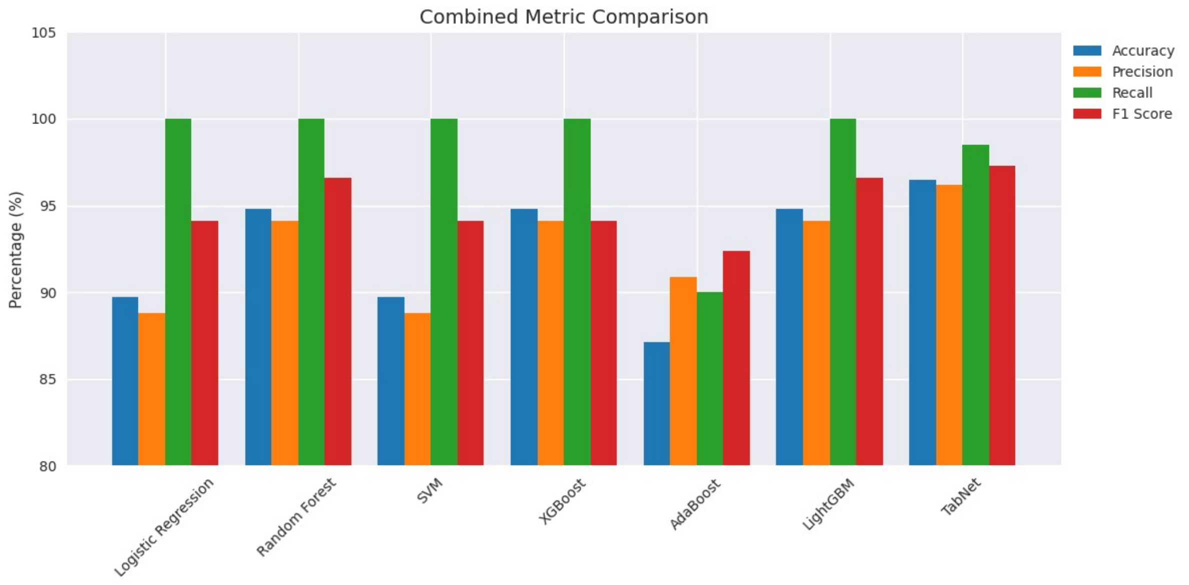
Implementation & Results of Module

Model	Accuracy	Precision	Recall (%)	F1 Score
Logistic Regression	89.7	88.8	100	94.1
Random Forest	94.8	94.1	100	96.6
SVM	89.7	88.8	100	94.1
XGBoost	94.8	94.1	100	94.1
AdaBoost	87.1	90.9	90	92.4
LightGBM	94.8	94.11	100	96.6
TabNet	96.5	96.2	98.5	97.3



Implementation & Results of Module





Conclusion & Future Work

- Machine learning models, particularly TabNet, show great promise in enhancing Parkinson's disease diagnosis by achieving high accuracy and providing interpretable insights. Future advancements in feature engineering, raw signal processing, and ensemble architectures can further improve model precision and clinical utility.

1. Advanced Feature Engineering with TabNet
2. Class Imbalance Mitigation
3. Hierarchical Ensemble Architectures
4. Longitudinal Data Integration

References

[1] L. A. P. A. J. E. R. Deeks, "Parkinson's Disease: Diagnosis and Management," *British Medical Journal*, vol. 355, p. i5073, 2016.

[2] K. A. Prakash and G. C. T. M. Y. T. Su, "Parkinson's Disease Diagnosis Using Machine Learning Algorithms: A Comparative Study," *Journal of Medical Systems*, vol. 43, no. 3, pp. 93-106, 2019.

[3] M. P. C. S. L. K. S. M. S. Adil, "Machine Learning Algorithms for Parkinson's Disease Classification Using UCI Parkinson's Dataset," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 150-158, 2019.

[4] T. R. S. A. M. V. T. M. P. J. K. T. L. A. Bozic, "Prediction of Parkinson's Disease Using XGBoost and Feature Engineering," *Computers in Biology and Medicine*, vol. 104, pp. 27-34, 2018.

[5] J. R. J. M. C. A. O. K. L. F. M. T. D. G. D. Schapire, "Boosting Algorithms for Machine Learning and Classification," *International Journal of Computational Intelligence*, vol. 34, pp. 1835-1847, 2017.

[6] Y. M. L. G. C. A. P. X. Wang, "Support Vector Machines in Parkinson's Disease Diagnosis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1481-1490, 2020.

[7] M. M. Y. Z. W. S. Z. F. X. F. L. G. C. Li, "Bagging and Random Forests for Parkinson's Disease Classification," *Pattern Recognition Letters*, vol. 106, pp. 14-20, 2018.

[8] J. S. L. T. T. T. R. Scherzer, "Evaluating the Performance of Deep Learning Models for Parkinson's Disease Diagnosis," *Neurocomputing*, vol. 345, pp. 123-132, 2019.

[9] Z. Y. Y. G. L. Y. R. M. T. G. X. Chen, "Parkinson's Disease Classification Using Ensemble Techniques: A Systematic Review," *Journal of Artificial Intelligence in Medicine*, vol. 108, pp. 36-45, 2020.

[10] R. R. S. W. T. S. P. P. K. M. H. B. K. A. T. R. K. Lee, "Parkinson's Disease Diagnosis: Comparison of Logistic Regression and Random Forest," *Journal of Health Informatics Research*, vol. 16, no. 3, pp. 56-64, 2018.

Thank You