Early Detection Of Parkinson's Disease Using Machine Learning

Mridul Madnani |Devansh Bajpai| Srijan Singh Somvanshi| Sayan Sikder| SCOPE

Introduction

Research aims to apply ensemble learning for Parkinson's disease diagnosis, addressing the gap of integrating feature selection, ensemble learning, and diverse ML algorithms. The literature gap lies in underutilization of ensemble learning and lack of comprehensive systems for PD diagnosis.

Motivation

The motivation for this project stems from the potential to improve patients' quality of life through early intervention and treatment of Parkinson's disease. Additionally, it offers valuable learning experience in the field of ML.

SCOPE of the Project

The project scope involves developing a system for early Parkinson's disease detection using ensemble stacking techniques. It includes data collection, preprocessing, and feature selection tailored to PD datasets. Ensemble learning algorithms will be employed to combine diverse base classifiers for improved accuracy. Model performance will be evaluated rigorously through cross-validation and testing on independent dataset. Collaboration with domain experts will ensure clinical relevance and validation of the system. The project will explore the potential of biomarker identification for enhanced diagnostic accuracy.

Methodology

The methodology adapted for this project is as follows:

Data Preprocessing:

- Importing necessary libraries and dataset.
- Exploring dataset content, dropping irrelevant columns.
- Scaling features and splitting data into training and testing sets.

Model Development and Training:

- Selection of machine learning algorithms like logistic regression, SVM, XGBoost.
- Fine-tuning hyperparameters through grid search or random search.
- Training models on preprocessed data, using cross-validation techniques.
- Iteratively optimizing model architectures and hyperparameters based on validation results.

Comparison and Stacking Ensemble Techniques:

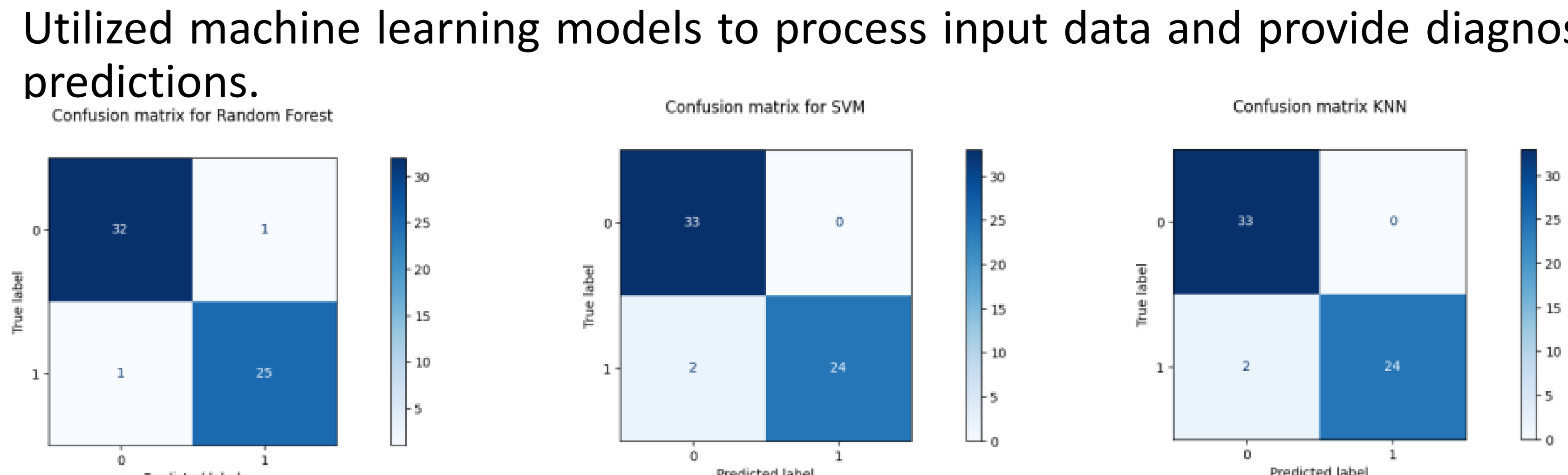
- Evaluating baseline model performance using cross-validation.
- Implementing stacking ensemble methods with diverse base models.
- Assessing ensemble performance and comparing it with individual models.
- Interpreting ensemble model decisions and visualizing decision-making processes.

Validation and Evaluation:

- Comparing developed models with baseline algorithms.
- Visualizing model performance using ROC curves, confusion matrices.
- Assessing interpretability and explainability of models.
- Selecting promising models for further refinement based on evaluation results.

Building a User Interface:

- Developed a user-friendly website for Parkinson's disease detection.
- Implemented a feature for users to input voice data.
- Utilized machine learning models to process input data and provide diagnosis predictions.

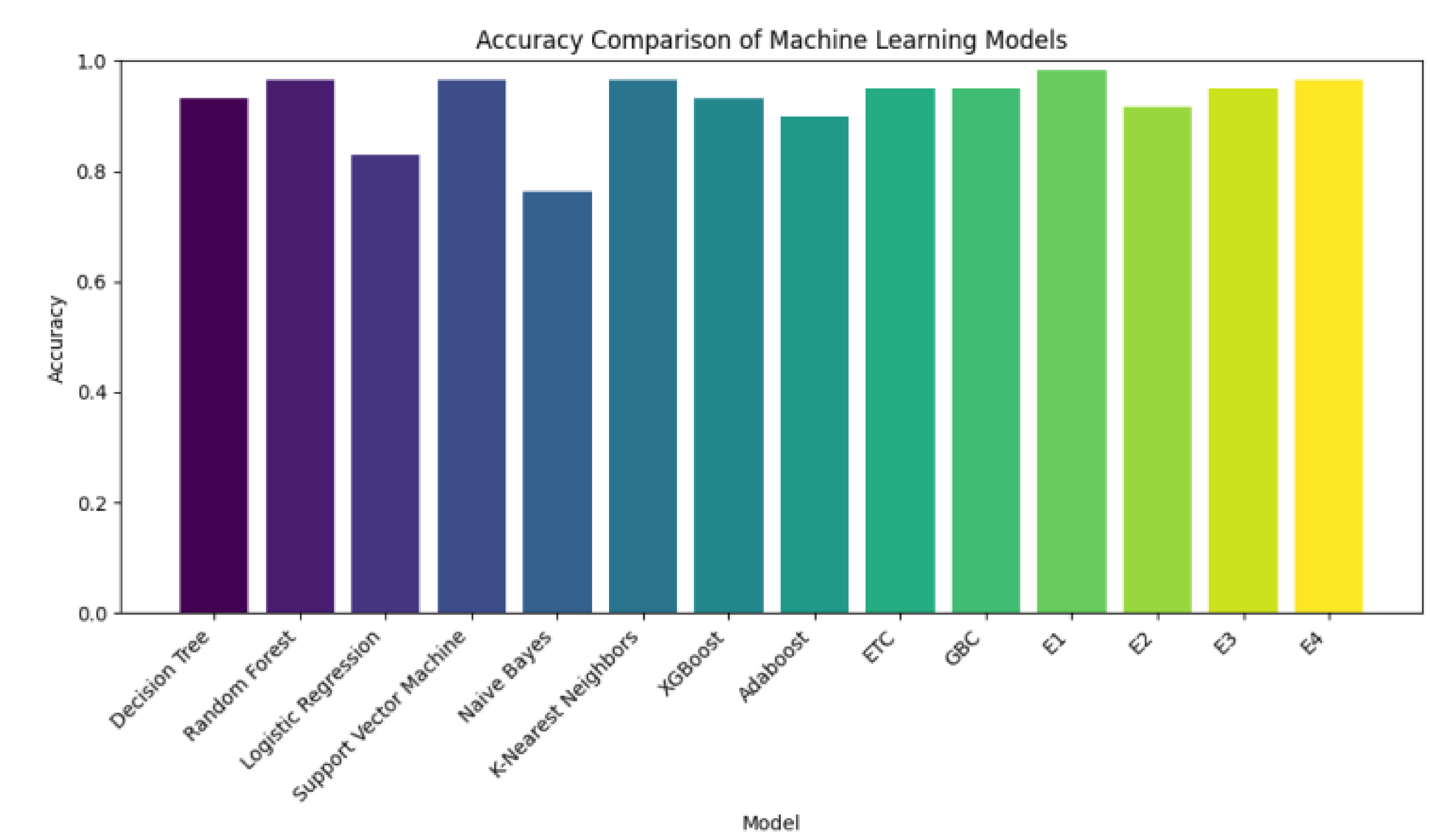


Results

	Metric	DT	RF	LR	SVM	NB	KNN	XGB	ADA	ETC	GBC	E2	E3	E4	E1
0	Accuracy	0.932203	0.966102	0.830508	0.966102	0.762712	0.966102	0.932203	0.898305	0.949153	0.949153	0.915254	0.949153	0.966102	0.983051
1	F1-Score	0.920000	0.961538	0.782609	0.960000	0.650000	0.960000	0.925926	0.888889	0.941176	0.943396	0.897959	0.941176	0.961538	0.981132
2	Recall	0.884615	0.961538	0.692308	0.923077	0.500000	0.923077	0.961538	0.923077	0.923077	0.961538	0.846154	0.923077	0.961538	1.000000
3	Precision	0.958333	0.961538	0.900000	1.000000	0.928571	1.000000	0.892857	0.857143	0.960000	0.925926	0.956522	0.960000	0.961538	0.962963
4	R2-Score	0.724942	0.862471	0.312354	0.862471	0.037296	0.862471	0.724942	0.587413	0.793706	0.793706	0.656177	0.793706	0.862471	0.931235

Evaluation results summarizing model performance using key metrics.

We've developed a Parkinson's disease detection model using machine learning techniques. Through meticulous comparison, we identified Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) as the best-performing models, each boasting accuracies around 96%. Leveraging the ensemble stacking method, we combined these top-performing models into a unified ensemble model, which remarkably achieved an accuracy of approximately 98% in early Parkinson's disease detection. This outcome highlights the exceptional effectiveness and robustness of the ensemble learning approach. By integrating Random Forest, SVM, and KNN classifiers, our ensemble model capitalizes on the diverse strengths of individual models, surpassing the performance of standalone baseline classifiers like logistic regression or decision trees. This unified approach offers superior accuracy and robustness, significantly enhancing the identification of Parkinson's disease cases.



Visualization showing the accuracy of different ML models & their Ensembles.

Conclusion

In conclusion, our ensemble stacking model, combining Random Forest, SVM, and KNN classifiers, achieved an impressive early detection accuracy of around 98% for Parkinson's disease. This underscores the robustness of ensemble methods in leveraging diverse classifier strengths. Future endeavors may center on validating the model with larger datasets, analyzing feature importance for enhanced interpretability, and exploring alternative ensemble techniques or domain-specific knowledge incorporation. Extending applicability to clinical settings and mitigating potential biases would further bolster its practical utility in healthcare decision-making.

References

T. Velmurugan and J. Dhinakaran, "A Novel Ensemble Stacking Learning Algorithm for Parkinson's Disease Prediction," Mathematical Problems in Engineering, vol. 2022, pp. 1–10, Jul. 2022, doi: https://doi.org/10.1155/2022/9209656.

H. V. Nguyen and H. Byeon, "Prediction of Parkinson's Disease Depression Using LIME-Based Stacking Ensemble Model," Mathematics, vol. 11, no. 3, p. 708, Jan. 2023, doi: https://doi.org/10.3390/math11030708

S. A. Doumari, K. Berahmand, and M. J. Ebadi, "Early and High-Accuracy Diagnosis of Parkinson's Disease: Outcomes of a New Model," Computational and Mathematical Methods in Medicine, vol. 2023, p. 1493676, 2023, doi: https://doi.org/10.1155/2023/1493676.

Johann Faouzi, Olivier Colliot, and Jean-Christophe Corvol, "Machine Learning for Parkinson's Disease and Related Disorders," Neuromethods, pp. 847–877, Jan. 2023, doi: https://doi.org/10.1007/978-1-0716-3195-9_26.