ASEC 2024 Conference
The 5th International Electronic Conference on Applied Sciences
04–06 December 2024 | Online
MDPI

Using machine learning (ML) algorithms based on voice disorders to identify Parkinson's disease

Krishna Dharavathu, M Ranga Rao , Sankara chinmai sri padma, Dinesh Reddy Rikka, Lakshmi Priya Kuruhuri, Meghana Addagarla
krishnadharavath4u@gmail.com, prof.mrrao1@gmail.com, chinmaipostbox@gmail.com, dineshreddy1621@gmail.com, lakshmipriyakuruhuri@gmail.com, meghanaaddagarla2004@gmail.com.
Potti Sriramulu Chalavadi Mallikarjuna Rao College of Engineering and Technology, Vijayawada, Andhra Pradesh, India

INTRODUCTION & AIM

Introduction

Diseases like breast cancer and diabetes receive significant research focus and visibility. Parkinson's disease (PD), despite being the second most common neurodegenerative disorder, lacks similar awareness [1]. PD affects 1% of the population globally, with a higher prevalence in men than women [2].•Characterized by slow progression, PD initially manifests subtle symptoms, such as impaired sense of smell, constipation, and depression, which double the risk of developing the disease [3].• Early symptoms often include speech-related issues like hoarse voice, breathiness, monotone, and imprecise articulation, affecting communication [3-4].• Early diagnosis can significantly improve patient outcomes by enabling timely intervention [2]• The main objective of this work is to focus on early detection of Parkinson's disease with voice analysis by using machine learning algorithms

METHOD

Dataset

• Source: UC Irvine Machine Learning Repository.
• Composition: 195 voice samples with 23 attributes related to voice disorders in Parkinson's disease patients.2.

Data Preprocessing

• Class Imbalance: Addressed using SMOTE (Synthetic Minority Oversampling Technique).
• Feature Scaling: Applied standard scaling techniques.
• Dimensionality Reduction: Principal Component Analysis (PCA) is used to retain significant variance and simplify the dataset.

Model Selection

• Machine Learning Algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF), AdaBoost.

Validation and Testing

• Train-test split: 80-20 ratio.
• Cross-validation: 5-fold stratified K-Folds.

Performance Metrics

• Evaluated using metrics such as Accuracy, Precision, Recall, and F1-score.Confusion matrix, ROC-AUC curves.

RESULTS & DISCUSSION

From the experiment results, recognized K-Nearest Neighbors (KNN) emerged as the best performer, achieving 98.31% accuracy, ideal precision for the normal class (1.00), and perfect recall for Parkinson's cases (1.00), ensuring no missed diagnoses. Its F1-score of 0.98 highlights a strong balance between precision and recall. While AdaBoost matched KNN in accuracy, its slightly lower recall for Parkinson's cases (0.97) makes K-Nearest Neighbors (KNN) the preferred choice. Consequently, K-Nearest Neighbors (KNN) is proposed as the most reliable model for robust and accurate Parkinson's disease classification, showing outstanding performance compared to other models. The overall performance of the models is presented in Tables 1 and 2. Figures 1 to 5 illustrate the ROC curves for KNN, AdaBoost, LR, SVM, and Random Forest algorithms, while Figures 6 to 10 display the confusion matrices for these techniques. Figure 11 visualizes the correlation among variables using heat maps.

Table 1: Model Evaluation through various metrics, 0 – Normal, 1 - Parkinson's

Model	Accuracy (%)	Precision		Recall		F1-Score		Cross-Validated ROC-AUC
		0	1	0	1	0	1	
KNN	0.98	1.00	0.97	0.97	1.00	0.98	0.98	0.9813
AdaBoost	0.98	0.97	1.00	1.00	0.97	0.98	0.98	0.9698
LR	0.85	0.84	0.86	0.87	0.83	0.85	0.84	0.9193
RF	0.95	1.00	0.94	0.71	1.00	0.83	0.97	0.7653
SVM	0.95	0.91	1.00	1.00	0.90	0.95	0.95	0.9609

Table 2: Model Evaluation through Confusion metrics.

Model	Confusion Metrics			
	TP	TN	FP	FN
KNN	29	29	1	0
AdaBoost	28	30	0	1
LR	24	26	4	5
RF	32	5	2	0
SVM	26	30	0	3

TP – True Positives
TN – True Negatives
FP – False Positives
FN – False Negatives

Receiver Operating Characteristic curves (ROC):

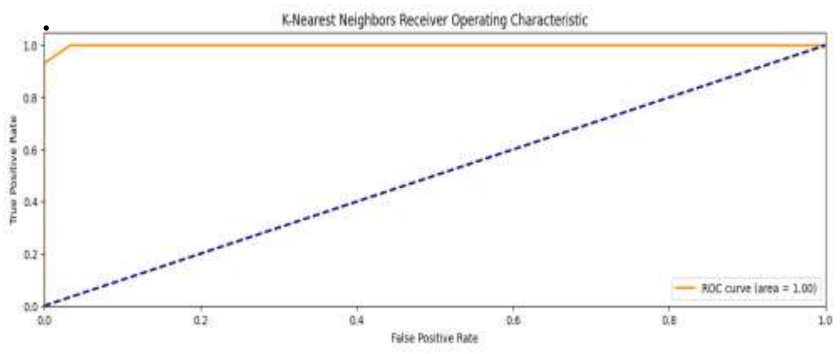

Fig.1: ROC Curve for KNN

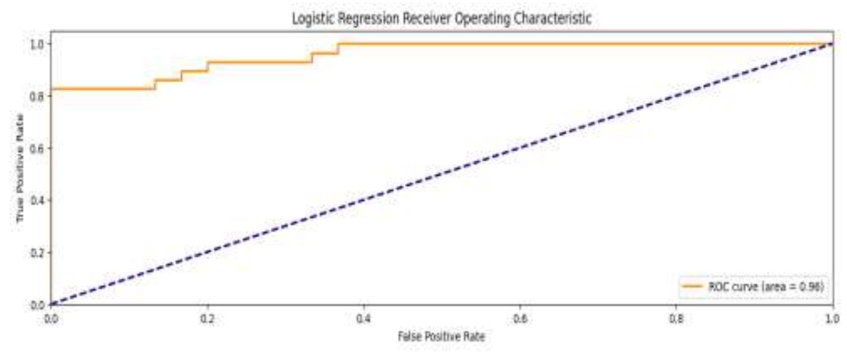

Fig.3: ROC Curve for Logistic Regression

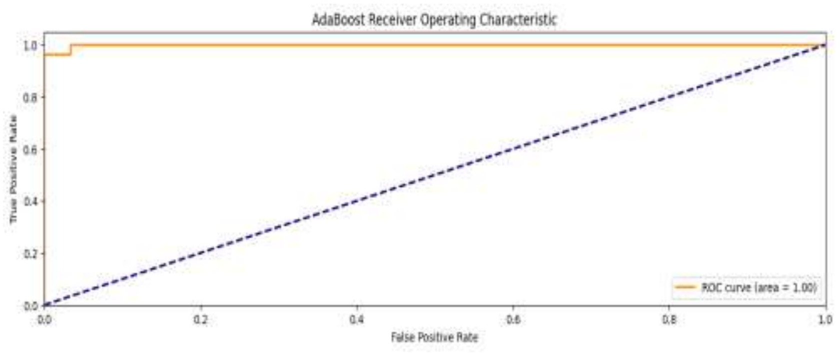

Fig.2: ROC Curve for AdaBoost

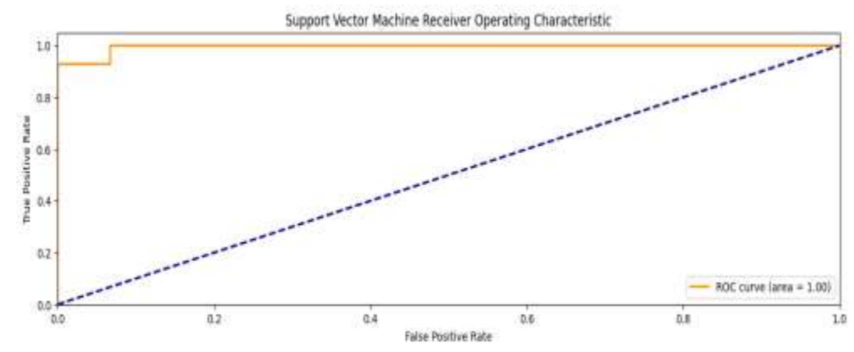

Fig.4: ROC Curve for SVM

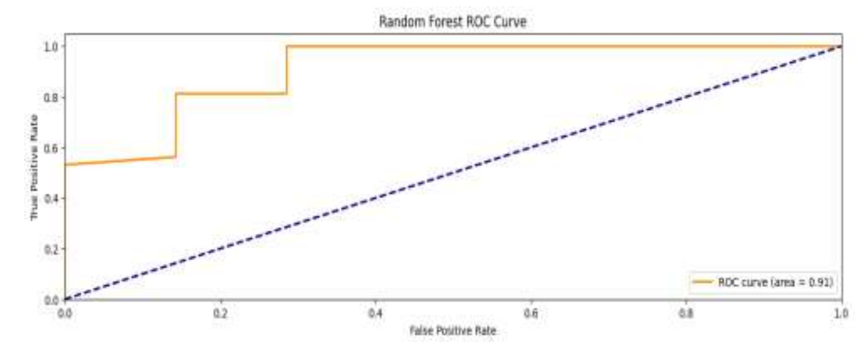

Fig.5: ROC Curve for Random Forest

Confusion Matrices:

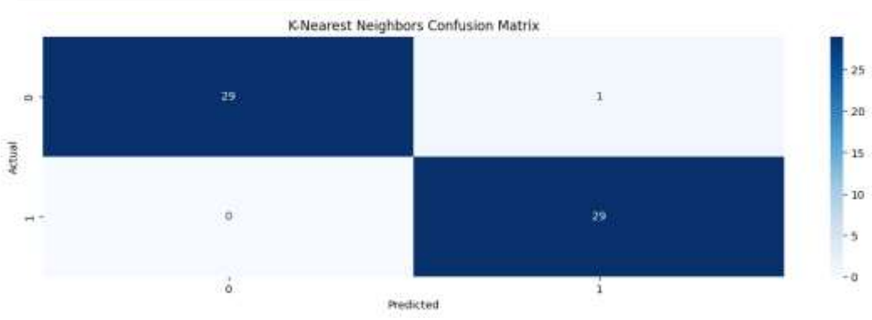

Fig.6: Confusion Matrix for KNN model

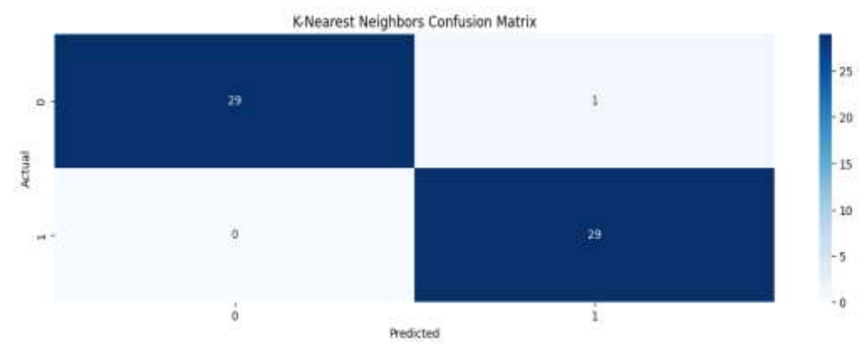

Fig.9: Confusion Matrix for SVM model

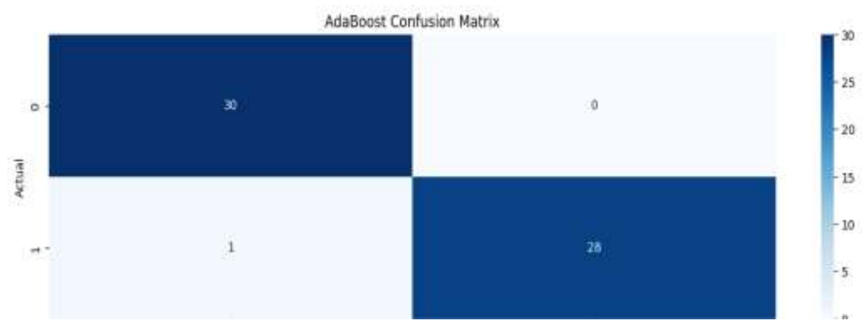

Fig.7: Confusion Matrix for AdaBoost model

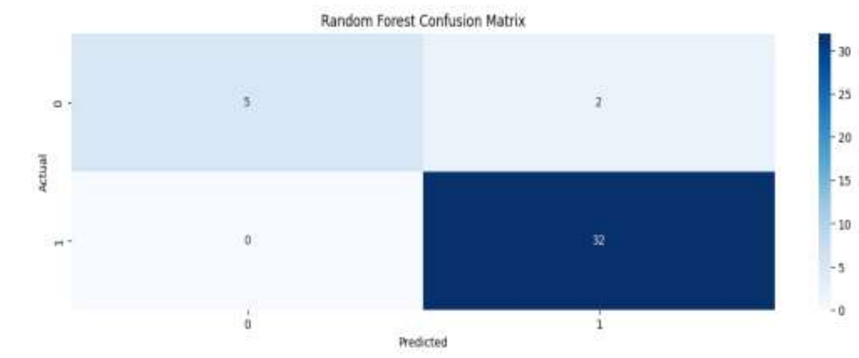

Fig.10: Confusion Matrix for Random Forest Model

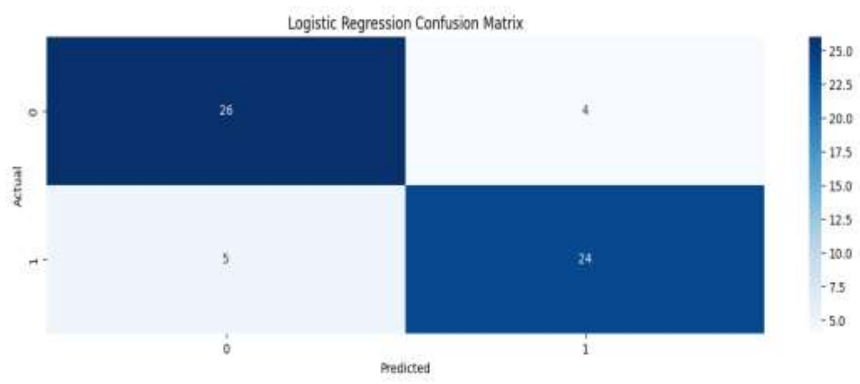

Fig.8: Confusion Matrix for LR Model

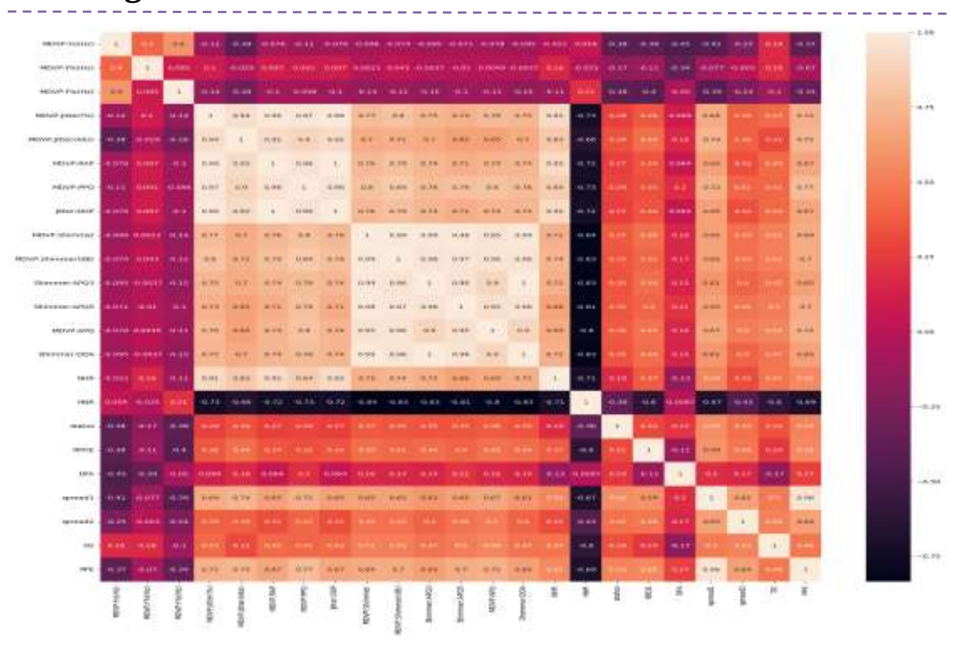

Fig.11: Visualizing Correlation Using Heat maps

CONCLUSION & FUTURE WORKS

This work utilises several Machine Learning techniques, which are k-NN, SVM, RF, LR, and AdaBoost boosting models, to distinguish between individuals with PD and those under normal conditions. ML classifiers have a strong performance when applied to speech data that involve the extraction of many phonetic characteristics. From the experiment results, recognized K-Nearest Neighbors (KNN) emerged as the best performer, achieving 98.31% accuracy. The early detection of PD has the potential to facilitate accurate diagnosis and ease the progression of symptoms. This study has the potential to be applied to various Machine Learning techniques and telemonitoring datasets to enhance the accuracy of classifiers. In future investigations, this work can contribute to the development of an improved prototype using Field Programmable Gate Arrays (FPGA) to analyse voice pathology.

REFERENCES

1. Elizabeth Leritz , Chris Loftis , Greg Crucian , William Friedman & Dawn Bowers (2004) "*Self-Awareness of Deficits in Parkinson Disease, The Clinical Neuropsychologist*", 18:3, 352-361, DOI: 10.1080/1385404049052412
2. Asmat Ullah Khan, Muhammad Akram, Muhammad Daniyal & Rida Zainab (2018): Awareness and current knowledge of Parkinson disease: "*A Neurodegenerative disorder, International Journal of Neuroscience*", DOI: 10.1080/00207454.2018.1486837
3. Asmat Ullah Khan, Muhammad Akram, Muhammad Daniyal & Rida Zainab (2018): "*Awareness and current knowledge of Parkinson disease: A Neurodegenerative disorder*", DOI:10.1080/00207454.2018.1486837, ISSN: 0020-7454 (Print) 1543-5245 (Online).
4. Nicola Amoroso, Marianna La Rocca, Alfonso Monaco, Roberto Bellotti, Sabina Tangaro, "*Complex networks reveal early MRI markers of Parkinson's disease, Medical Image Analysis" (2018)*, DOI: 10.1016/j.media.2018.05.004 ,ISSN 1361-8415