**Parkinson’s Disease Diagnosis: ML Algorithm Comparison (Additional Material) Feature Selection using L1 (Lasso) and Univariate Method:**

| **Method** | **Features Selected** | **Total Count** |
| --- | --- | --- |
| **LASSO** | DFA, b4, GNE\_NSR\_TKEO, IMF\_SNR\_SEO, mean\_MFCC\_1st\_coef, mean\_MFCC\_6th\_coef, mean\_MFCC\_7th\_coef, mean\_MFCC\_8th\_coef, mean\_MFCC\_10th\_coef, mean\_delta\_log\_energy, mean\_2nd\_delta, mean\_10th\_delta\_delta, mean\_11th\_delta\_delta, std\_MFCC\_2nd\_coef, std\_8th\_delta, std\_9th\_delta, std\_delta\_delta\_log\_energy, std\_6th\_delta\_delta, std\_7th\_delta\_delta, std\_9th\_delta\_delta, det\_LT\_entropy\_shannon\_4\_coef, tqwt\_energy\_dec\_11, tqwt\_energy\_dec\_15, tqwt\_energy\_dec\_19, tqwt\_energy\_dec\_25, tqwt\_entropy\_shannon\_dec\_23, tqwt\_entropy\_shannon\_dec\_25, tqwt\_entropy\_shannon\_dec\_34, tqwt\_entropy\_shannon\_dec\_35, tqwt\_entropy\_log\_dec\_26, tqwt\_entropy\_log\_dec\_28, tqwt\_entropy\_log\_dec\_33, tqwt\_TKEO\_mean\_dec\_17, tqwt\_TKEO\_mean\_dec\_21, tqwt\_TKEO\_mean\_dec\_27, tqwt\_TKEO\_std\_dec\_11, tqwt\_TKEO\_std\_dec\_12, tqwt\_medianValue\_dec\_5, tqwt\_medianValue\_dec\_10, tqwt\_medianValue\_dec\_25, tqwt\_medianValue\_dec\_29, tqwt\_medianValue\_dec\_33, tqwt\_medianValue\_dec\_36, tqwt\_meanValue\_dec\_5, tqwt\_meanValue\_dec\_7, tqwt\_meanValue\_dec\_11, tqwt\_meanValue\_dec\_15, tqwt\_meanValue\_dec\_16, tqwt\_meanValue\_dec\_22, tqwt\_meanValue\_dec\_36, tqwt\_minValue\_dec\_12, tqwt\_minValue\_dec\_17, tqwt\_maxValue\_dec\_6, tqwt\_maxValue\_dec\_25, tqwt\_skewnessValue\_dec\_9, tqwt\_skewnessValue\_dec\_17, tqwt\_skewnessValue\_dec\_27, tqwt\_skewnessValue\_dec\_28, tqwt\_kurtosisValue\_dec\_18, tqwt\_kurtosisValue\_dec\_20, tqwt\_kurtosisValue\_dec\_27, tqwt\_kurtosisValue\_dec\_28, tqwt\_kurtosisValue\_dec\_33, tqwt\_kurtosisValue\_dec\_36 | **64** |
| **KBest** | mean\_MFCC\_2nd\_coef, std\_8th\_delta, std\_6th\_delta\_delta, std\_7th\_delta\_delta, std\_8th\_delta\_delta, std\_9th\_delta\_delta, tqwt\_entropy\_shannon\_dec\_11, tqwt\_entropy\_log\_dec\_11, tqwt\_entropy\_log\_dec\_12, tqwt\_TKEO\_std\_dec\_12, tqwt\_stdValue\_dec\_11, tqwt\_stdValue\_dec\_12, tqwt\_stdValue\_dec\_13, tqwt\_minValue\_dec\_11, tqwt\_minValue\_dec\_12, tqwt\_minValue\_dec\_13, tqwt\_maxValue\_dec\_11, tqwt\_maxValue\_dec\_12, tqwt\_maxValue\_dec\_13, tqwt\_kurtosisValue\_dec\_27 | **20** |
| **Common Features** | std\_6th\_delta\_delta, std\_9th\_delta\_delta, tqwt\_kurtosisValue\_dec\_27, tqwt\_minValue\_dec\_12, tqwt\_TKEO\_std\_dec\_12, std\_7th\_delta\_delta, std\_8th\_delta | **7** |

**Train-Test Split for ‘Lasso’ Selected Features**

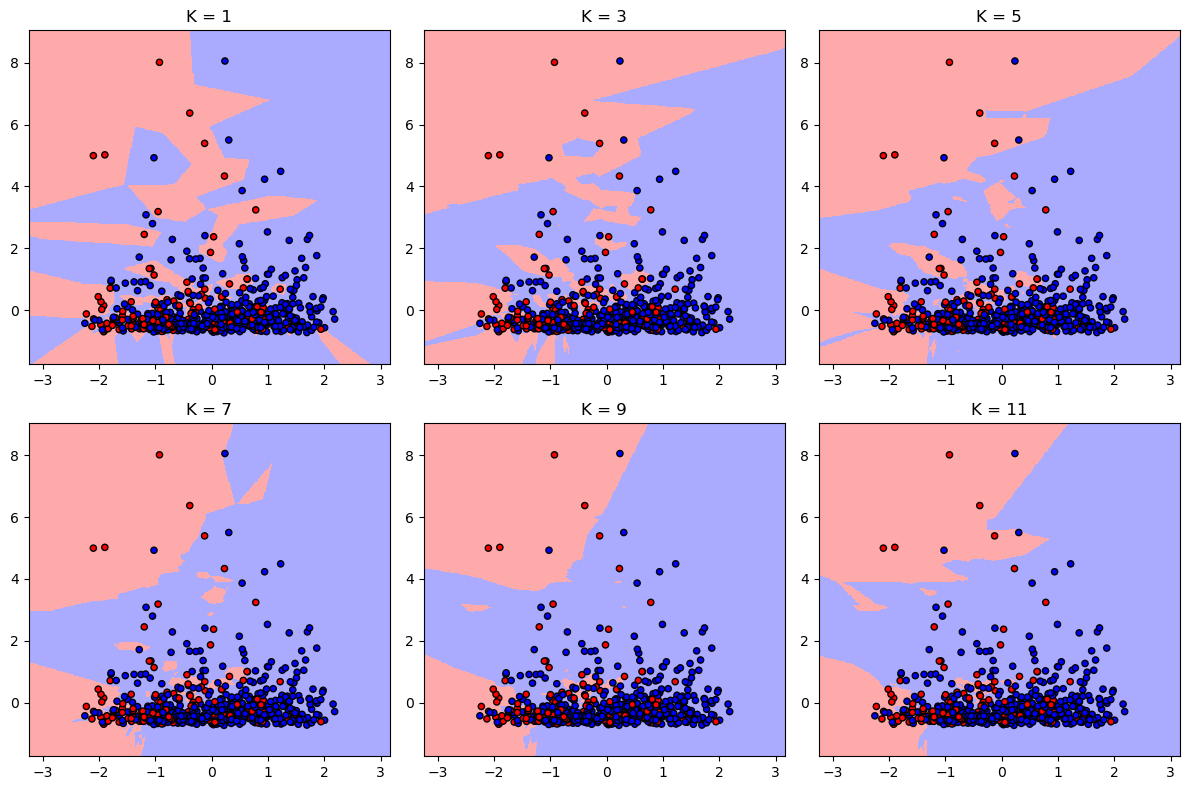
| **Data splitting** | **Shape** |
| --- | --- |
| X\_train | (604, 64) |
| X\_test | (152, 64) |
| y\_train | (604,) |
| y\_test | (152,) |

**Visualizing Decision Boundaries: SVM**

A screen shot of a diagram

Description automatically generated

**Decision Boundaries Using KNN classifier with variable values of K along with Mean cross validation Scores**



| **K** | **Mean Cross Validation** |
| --- | --- |
| 1 | 0.8792 |
| 3 | 0.8642 |
| 5 | 0.8675 |
| 7 | 0.8592 |
| 9 | 0.8559 |
| 11 | 0.8658 |

**Decision Tree Visualization for Random Forest Classifier with limited depth**A diagram of a diagram

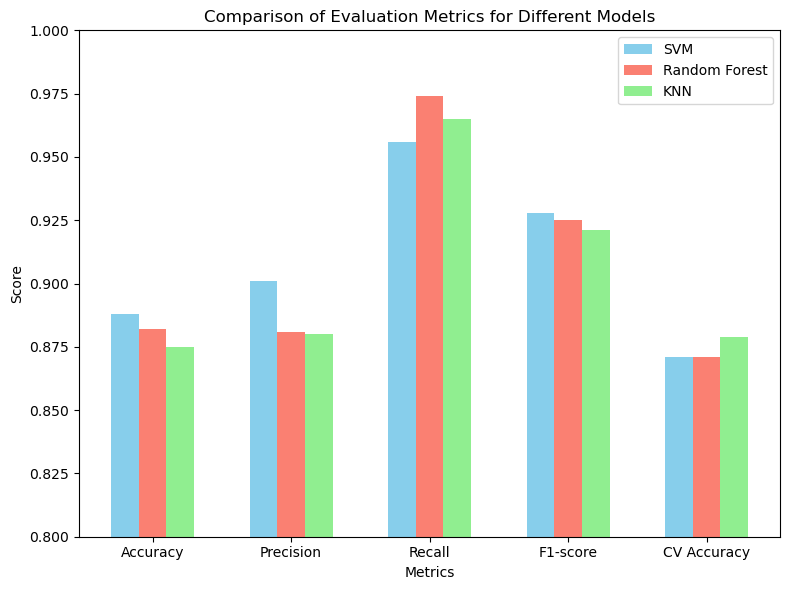
Description automatically generated

**Comparison of Confusion Matrices Across Different Models**

| **Model** | **True Positives (TP)** | **True Negatives (TN)** | **False Positives (FP)** | **False Negatives (FN)** |
| --- | --- | --- | --- | --- |
| **SVM** | 109 | 26 | 1 | 5 |
| **RF** | 111 | 23 | 15 | 3 |
| **KNN** | 112 | 22 | 16 | 2 |

**Comparison of Evaluation Metrices Across Different Models**

| **Metric** | **SVM** | **RF** | **KNN** |
| --- | --- | --- | --- |
| **Accuracy** | 0.8882 | 0.8816 | 0.8816 |
| **Precision** | 0.9008 | 0.8810 | 0.8750 |
| **Recall** | 0.9561 | 0.9737 | 0.9824 |
| **F1 score** | 0.9277 | 0.9250 | 0.9256 |
| **Cross-Validation** | 0.8709 | 0.8708 | 0.8792 |
| **AUC-ROC** | 0.9367 | 0.9421 | 0.9158 |



**Comparing ROC-AUC Curves Across Different Models**

A graph of a curve

Description automatically generated A graph of a curve

Description automatically generated

A graph of a curve

Description automatically generated

Introduction:

Parkinson's disease (PD) presents a significant challenge for early diagnosis, given its complex array of motor and non-motor symptoms. Leveraging machine learning (ML) algorithms, particularly in the analysis of speech features, holds promise for improving diagnostic accuracy. This repository serves as a comprehensive exploration into the efficacy of three prominent ML algorithms—Support Vector Machine (SVM), Random Forest (RF), and k-Nearest Neighbors (KNN)—in diagnosing PD based on speech parameters.

The dataset utilized in this study encompasses information from 756 patients, including diverse numerical features alongside binary labels indicating the presence or absence of PD. Through meticulous preprocessing, including numerical feature scaling and handling missing values, the dataset was primed for analysis. Feature selection techniques such as LASSO and univariate methods were employed to identify the most informative features for model training.

The repository offers detailed insights into the methodological approach, including data preprocessing, feature selection, model training, and evaluation. Various evaluation metrics, visualizations, and decision boundaries were utilized to assess the performance of each ML algorithm comprehensively.

By delving into the nuances of each algorithm's strengths and considerations, this repository aims to provide valuable guidance for researchers and practitioners in the field of Parkinson's disease diagnosis. Furthermore, it opens avenues for future research directions, including the exploration of ensemble methods to enhance diagnostic accuracy and clinical utility.

Overall, this repository serves as a foundational resource for understanding the role of ML algorithms in early PD detection based on speech features, offering valuable insights and opportunities for advancement in the field.