Digital Signal Processing and Machine Learning in Voice Analysis

By Siddharth Kumar (22/11/EE/028)

School Of Engineering, Jawaharlal Nehru University

Abstract: In the following project, a voice signal was analyzed based on some DSP and ML algorithms to judge the performance comparison of the different classification models. At the first step, the voice signal is recorded in .wav format, followed by adding artificial noise and phase distortion to generate signals simulating real-case scenarios. The distorted voice signal was decomposed by using Fourier decomposition to extract the features. Features thus extracted were used as inputs in the various machine learning classifiers, which included Logistic Regression, Support Vector Machines (SVM), Decision Trees, and Random Forests. The performance of the models was checked and compared in terms of the accuracy of the classification of the distorted voice features. This provided insight into how machine learning techniques can be used in voice signal processing.

Introduction: Voice signal processing is very important for several applications, including speech recognition, noise filtering, and voice-based security systems.

This project was aimed at researching the ability of various machine learning models in classification of voice signals distorted by noise and phase. In this process, feature extraction, which is a popular approach in DSP, decomposes a signal into its constituent parts in the frequency domain, was used. These extracted features were then used to train and test various machine learning models on the ability to classify.

Methodology:

Voice Signal Acquisition: A voice signal was recorded in.wav format. This is the basic input which will further undergo changes.

Signal Generation: The following types of signals were generated:

- Clean signals were duplicated from the original voice signal.
- Noisy signals were created by adding Gaussian noise to the clean signals.

- Phase-shifted signals were generated by introducing random phase shifts to the clean signals.

These signals were then labeled as 0 (clean), 1 (noisy), and 2 (phase-shifted), forming a labeled dataset for model training and evaluation.

Fourier Decomposition: Fourier decomposition was used to transform the disturbed voice signals into their frequency domain equivalents by converting time-domain signals into the frequency domain. Features were subsequently extracted in the frequency domain, which played a central role in the classification.

Machine Learning Models:

The extracted features were fed into a variety of machine learning models for classification:

- Logistic Regression
- Support Vector Machines (SVM)
- Decision Trees
- Random Forest

Random Forest Dimensionality reduction and standardization:

The features were standardized using StandardScaler to ensure that all features had zero mean and unit variance.

Using Principal Component Analysis (PCA), reduced the number of features to 20 components.

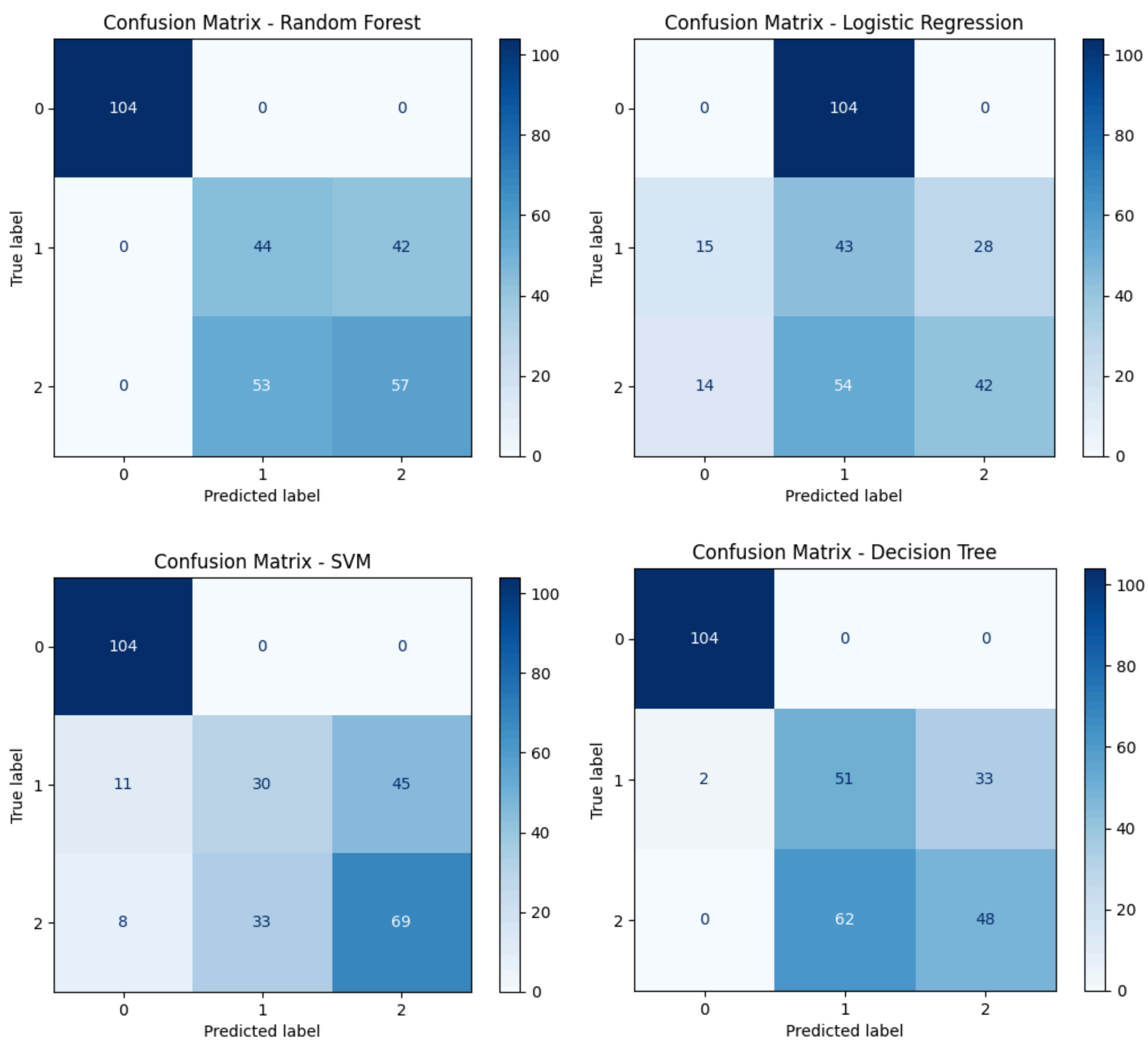
Model Training and Testing:

Random Forest: Hyperparameters tuning for Random Forest was applied with RandomizedSearchCV when searching over a range of parameters which included the number of estimators, the maximum depth, and minimum samples for splits and leaves.

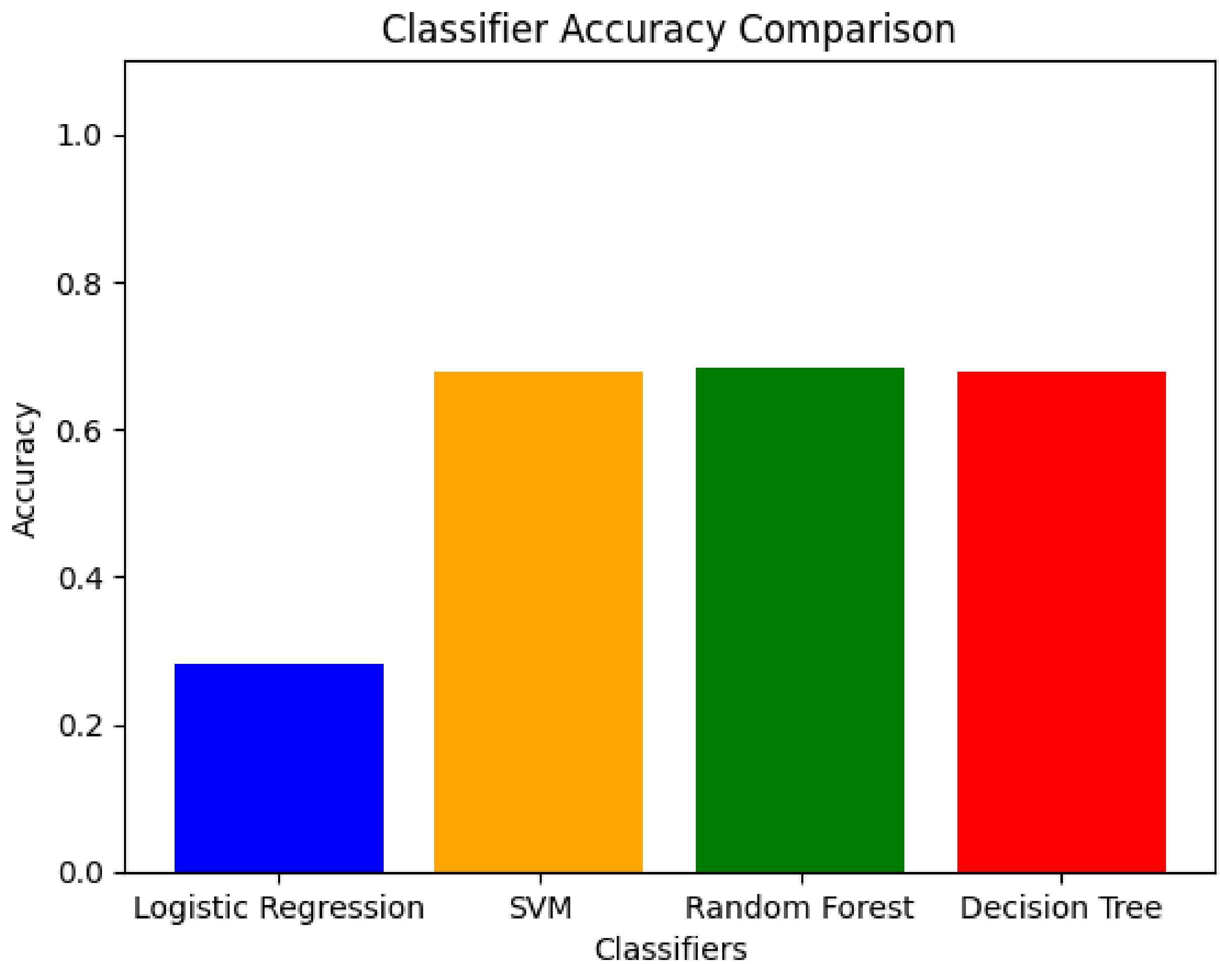
Each model was trained using the processed features and assessed for accuracy to see how that model performed. For visualization, classification performance plots for the confusion matrices of all the classifiers were generated. Results: The machine learning models were trained on the Fourier features and performance was evaluated using accuracy. The confusion matrix of all classifiers was plotted, and all the models tested were compared to display the accuracy results for each.

Results:

```
Logistic Regression Accuracy: 0.2833
SVM Accuracy: 0.6767
Best Hyperparameters for Random Forest: {'max_depth': 32, 'max_features': None, 'min_samples_leaf': 9, 'min_samples_split': 9, 'n_estimators': 61}
Random Forest Accuracy: 0.6833
Decision Tree Accuracy: 0.6767
```



Label 0 represents clean signal, Label 1 represents noisy signal & Label 2 represents phase shifted signal



Conclusion:

This project showed the feasibility of using Fourier decomposition for feature extraction from voice signals affected by noise and phase distortions. The machine learning models, including Logistic Regression, SVM, Decision Trees, and Random Forest, were tested for their performance in classifying distorted voice features. The results provided valuable insights into the suitability of these models for various signal processing tasks and their robustness in handling real-world distortions in voice signals

Source Code:

https://github.com/siddharthk7704/DSP-FDM

Source Code of Fourier Decomposition Method used:

https://github.com/udawat/Fourier-Decomposition-Method