AUDIO EMOTION RECOGNITION

Introduction to Text Analytics

Instructor: Dr. Sajjad Haider

Syeda Maham Jafri, 22796
s.jafri.22796@khi.iba.edu.pk

Contents

Abstract

This project explores the domain of audio emotion recognition through a comprehensive analysis involving feature extraction, model training, and feature reduction techniques across multiple datasets. By leveraging a variety of machine learning models ranging from basic classifiers to advanced deep learning and ensemble models, the study aims to identify the most effective approaches for accurately detecting emotions in audio data. The implementation of feature reduction methods further refines the model performance, highlighting the balance between complexity and accuracy. The results demonstrate significant insights into the performance differentials among the models and underscore the potential for enhanced emotion recognition systems.

Introduction

Emotion recognition from audio signals has emerged as a critical component in various applications, including human-computer interaction, healthcare, and entertainment. The ability to accurately discern emotions from speech can significantly enhance the user experience in interactive systems and provide valuable insights in therapeutic settings. This project delves into the intricate process of recognizing emotions from audio data by employing a robust methodological framework.

The journey begins with the extraction of relevant features from the audio signals using diverse techniques to capture the nuances of emotional expressions. These features are then fed into a spectrum of machine learning models, spanning from traditional algorithms to sophisticated deep learning architectures and ensemble methods. Each model is rigorously trained and tested across four distinct datasets, ensuring a comprehensive evaluation of their performance.

To further optimize the models, feature reduction techniques are applied, aiming to reduce computational complexity while maintaining or improving accuracy. This multi-faceted approach not only enhances our understanding of the effectiveness of different models and techniques but also provides a roadmap for future developments in the field of audio emotion recognition.

Objectives

1. Feature Extraction Analysis:
   To systematically extract and analyze features from audio signals using various techniques, ensuring comprehensive capture of emotional characteristics.
2. Model Evaluation:
   To train and evaluate a wide range of machine learning models, from basic classifiers to advanced deep learning and ensemble models, on multiple datasets for emotion recognition.
3. Feature Reduction Implementation:
   To apply and assess feature reduction techniques in order to optimize model performance and reduce computational requirements.
4. Dataset Comparison:
   To compare and contrast the performance of models across four different datasets, identifying dataset-specific challenges and strengths.
5. Performance Benchmarking:
   To establish benchmarks for the effectiveness of different models and techniques in the context of audio emotion recognition, providing a foundation for future research and application.
6. Insight Generation:
   To generate insights into the interplay between feature complexity, model architecture, and performance, guiding the development of more efficient and accurate emotion recognition systems.

Features Extracted from Audio Signals

- Librosa

1.Time-Domain Features:

- Mean: Represents the average amplitude of the audio signal over time. It provides a measure of the overall loudness.
- Variance: Measures the spread of the amplitude values around the mean. It reflects the variability in the signal's intensity.
- Skewness: Indicates the asymmetry of the amplitude distribution. It can reveal whether the signal is skewed towards higher or lower amplitudes.
- Kurtosis: Describes the peakness or flatness of the amplitude distribution. It provides information about the tails of the distribution.
- Zero Crossing Rate: Counts the number of times the audio waveform crosses the zero-amplitude axis. It can be indicative of changes in the signal's frequency content.
- RMS Energy: Represents the root mean square of the signal's amplitude. It provides a measure of the signal's power.

2. Intensity and Pitch Features:

- Intensity (Average Energy): Represents the average energy of the audio signal, calculated as the mean absolute value of the waveform.
- Pitch (Fundamental Frequency): Reflects the perceived pitch of the audio, calculated from the pitch track.

3.Frequency-Domain Features:

- Spectral Centroid: Indicates the center of mass of the spectrum. It provides information about the "brightness" of the sound.
- Spectral Bandwidth: Measures the spread of the spectrum around the centroid. It reflects the range of frequencies present in the signal.
- Spectral Contrast: Describes the difference in amplitude between peaks and valleys in the spectrum. It captures the spectral texture of the signal.
- Spectral Rolloff: Determines the frequency below which a certain percentage of the total spectral energy lies. It can be indicative of the signal's spectral shape.

4.Mel-Frequency Cepstral Coefficients (MFCCs):

- MFCCs: Capture the spectral characteristics of the audio signal, mimicking the human auditory system's response. They are widely used as features for speech and audio processing tasks.

5.Chroma Features:

- Chroma: Represents the distribution of pitch classes (e.g., C, C#, D, etc.) in the audio signal. It provides information about the tonal content of the signal.

6.Spectral Flux, Density, and Flatness Features:

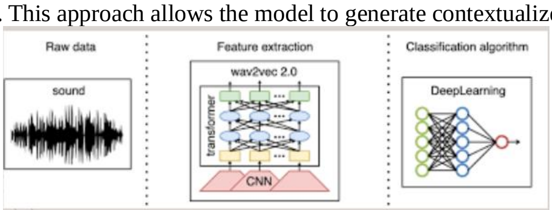
- Spectral Flux: Measures the rate of change in the spectrum over time. It can indicate sudden changes in the audio signal.
- Spectral Density: Represents the distribution of energy across different frequency bins. It provides information about the spectral content of the signal.
- Spectral Flatness: Quantifies the flatness of the spectrum. It can distinguish between noise-like and tonal sounds.

7.Wavelet Transform Coefficients:

- Wavelet Coefficients: Capture the time-frequency representation of the audio signal using wavelet analysis. They can reveal transient and frequency-specific features in the signal.

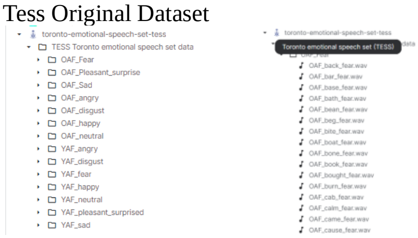
- Wav2vec

The wav2vec model extracts features from raw audio by using convolutional layers to capture short-term patterns and a Transformer-based context network to model long-range dependencies. In wav2vec 2.0, features are discretized using quantization to improve representation learning. This approach allows the model to generate contextualized embeddings suitable for speech recognition tasks.



Visualizing the Dataset

1. Tess Original Dataset



2. Emotion Classification

	Emotions	Path
0	fear	/kaggle/input/toronto-emotional-speech-set-tes...
1	fear	/kaggle/input/toronto-emotional-speech-set-tes...
2	fear	/kaggle/input/toronto-emotional-speech-set-tes...
3	fear	/kaggle/input/toronto-emotional-speech-set-tes...
4	fear	/kaggle/input/toronto-emotional-speech-set-tes...

3. After Extracting Features

	spectraldes800	spectraldes965	spectraldes1018	spectral_bandwidth	spectraldes311	spectraldes501	spectralflatness34_1	spectraldes169	spectralflatness63_1
0	(-1.92221-1.5739819j)	(3.209098+3.3247674j)	(0.00064413797+0.0005543969j)	2098.527546	(-0.006575377+0.025476627j)	(8.023978-0.9502358j)	0.004898	(5.4858284-7.9983535j)	NaN
1	(-0.3524934-1.4384637j)	(-1.4714466+0.51230013j)	(0.00024422153+0.0007641264j)	2541.522341	(-0.004673842+0.00084668584j)	(0.20690247-0.5132361j)	0.000484	(1.2699898+2.6835158j)	0.002525
2	(0.6470214-0.6753229j)	(4.464461+6.065009j)	(-0.0004122804-0.0001977338j)	2277.973682	(-0.0064187134+0.0035406854j)	(-0.56270653-1.3861161j)	0.000678	(-1.0598224-2.6220796j)	0.006471
3	(-1.2186965+1.0507218j)	(6.248013+4.5019j)	(-0.0005574594-0.0016202591j)	2298.334141	(0.0030216537+0.045358725j)	(-0.13301231-1.5885862j)	0.000523	(-22.305246-15.16855j)	0.128351
4	(0.617193+0.2507619j)	(-5.170534-1.7628525j)	(0.0001263773+0.0005818186j)	2371.323131	(0.0010483228-0.00020074192j)	(-0.6063541-2.7032049j)	0.000550	(-3.3541608-3.43196j)	0.027803

4. Converting the String columns to float

	spectraldes800	spectraldes965	spectraldes1018	spectral_bandwidth	spectraldes311	spectraldes501	spectralflatness34_1	spectraldes169	spectralflatness63_1	sp
0	-1.92221-1.5739819	3.209098+3.3247674	0.00064413797+0.0005543969	2098.527546	-0.006575377+0.025476627	8.023978-0.9502358	0.004898	5.4858284-7.9983535	NaN	2.21047
1	-0.3524934-1.4384637	-1.4714466+0.51230013	0.00024422153+0.0007641264	2541.522341	-0.004673842+0.00084668584	0.20690247-0.5132361	0.000484	1.2699898+2.6835158	0.002525	
2	0.6470214-0.6753229	4.464461+6.065009	-0.0004122804-0.0001977338	2277.973682	-0.0064187134+0.0035406854	-0.56270653-1.3861161	0.000678	-1.0598224-2.6220796	0.006471	0.769745
3	-1.2186965+1.0507218	6.248013+4.5019	-0.0005574594-0.0016202591	2298.334141	0.0030216537+0.045358725	-0.13301231-1.5885862	0.000523	-22.305246-15.16855	0.128351	1.83518
4	0.617193+0.2507619	-5.170534-1.7628525	0.0001263773+0.0005818186	2371.323131	0.0010483228-0.00020074192	-0.6063541-2.7032049	0.000550	-3.3541608-3.43196	0.027803	0.225922

	spectraldes800	spectraldes965	spectraldes1018	spectral_bandwidth	spectraldes311	spectraldes501	spectralflatness34_1	spectraldes169	spectralflatness63_1	spectraldes660	...	spectralflatness25_1
0	-3.496192	6.533865	0.001199	2098.527546	0.018901	7.073742	0.004898	-2.512525	NaN	1.604777	...	0.000174
1	-1.790957	-0.959146	0.001008	2541.522341	-0.003827	-0.306334	0.000484	3.953506	0.002525	-0.312741	...	0.003476
2	-0.028302	10.529470	-0.000610	2277.973682	-0.002878	-1.948823	0.000678	-3.681902	0.006471	0.351583	...	0.002312
3	-0.167975	10.749913	-0.002178	2298.334141	0.048380	-1.721599	0.000523	-37.473796	0.128351	3.731486	...	0.010373
4	0.867955	-6.933387	0.000708	2371.323131	0.000848	-3.309559	0.000550	-6.786121	0.027803	1.166919	...	0.003373

5. Handling NaN values in the dataset

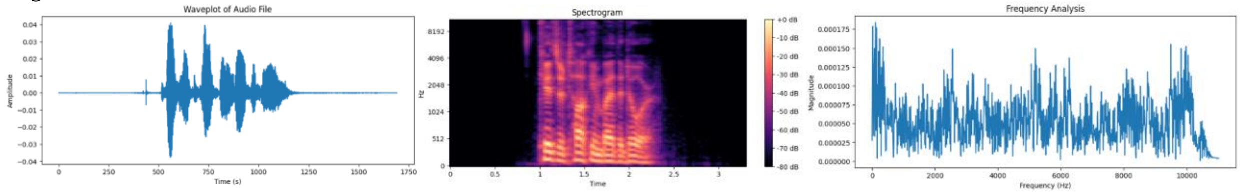
	spectraldes800	spectraldes965	spectraldes1018	spectral_bandwidth	spectraldes311	spectraldes501	spectralflatness34_1	spectraldes169	spectralflatness63_1	spectraldes660	...	spectralflatness25_1
0	-3.496192	6.533865	0.001199	2098.527546	0.018901	7.073742	0.004898	-2.512525	0.000000	1.604777	...	0.000174
1	-1.790957	-0.959146	0.001008	2541.522341	-0.003827	-0.306334	0.000484	3.953506	0.002525	-0.312741	...	0.003476
2	-0.028302	10.529470	-0.000610	2277.973682	-0.002878	-1.948823	0.000678	-3.681902	0.006471	0.351583	...	0.002312
3	-0.167975	10.749913	-0.002178	2298.334141	0.048380	-1.721599	0.000523	-37.473796	0.128351	3.731486	...	0.010373
4	0.867955	-6.933387	0.000708	2371.323131	0.000848	-3.309559	0.000550	-6.786121	0.027803	1.166919	...	0.003373

6. Final Representation of Feature Extracted Dataset

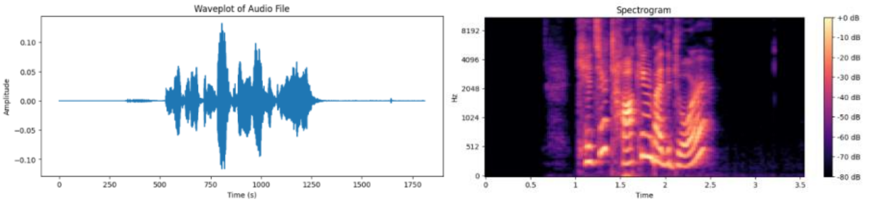
	intensity	pitch	mean	variance	skewness	kurtosis	rms_energy	zero_crossing_rate	mfcc1	spectralflux1	spectraldes1	spectralflatness34_1	spectral_bandwidth	spectral_centroid
0	0.082692	1583.7188	0.000012	0.013659	2098.527546	0.011349	0.105034	0.130083	-263.73022	0	-0.180903	0.004898	2098.527546	2313.922641
1	0.037956	1805.2482	-0.000042	0.003422	2541.522341	0.034405	0.047669	0.190693	-337.84518	0	-2.856732	0.000484	2541.522341	3348.274321
2	0.036141	2512.2668	-0.000051	0.002999	2277.973682	0.023356	0.046844	0.195319	-320.51890	0	-3.536198	0.000678	2277.973682	3173.867807
3	0.032188	2615.3618	0.000016	0.002413	2298.334141	0.042117	0.042772	0.296238	-300.41415	0	0.598392	0.000523	2298.334141	4072.295715
4	0.025229	2716.1520	-0.000041	0.001808	2371.323131	0.041530	0.033872	0.186067	-324.78800	0	-3.318592	0.000550	2371.323131	3170.919639

Visual Representation of Different Audio Signals

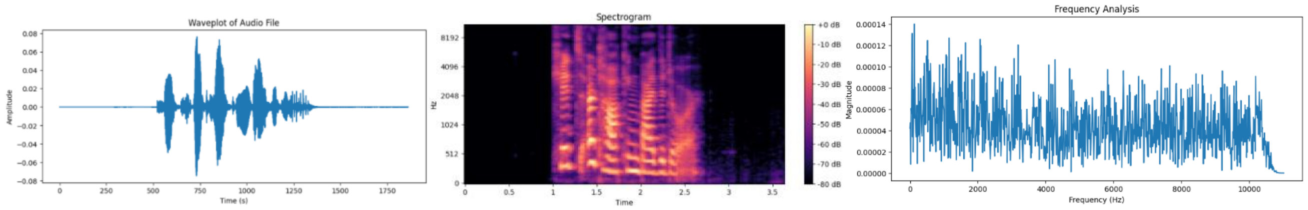
1. Anger



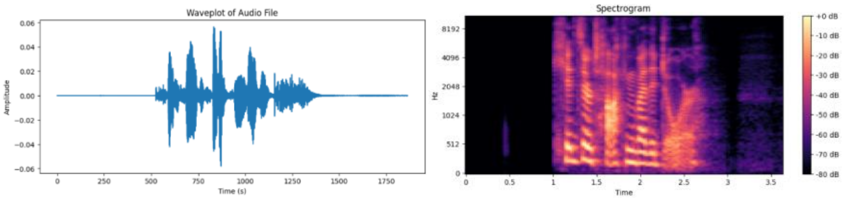
2. Surprise



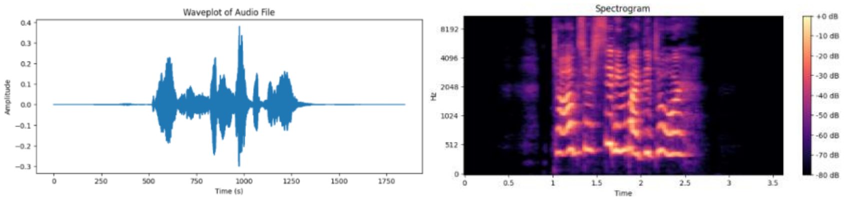
3. Neutral



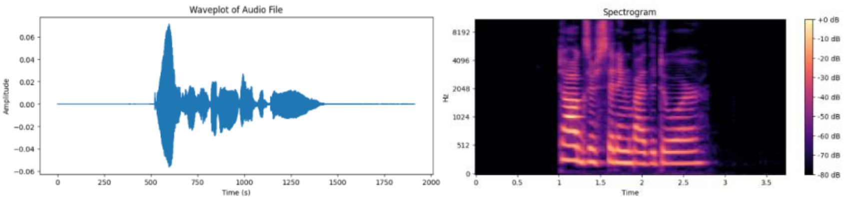
4. Disgust



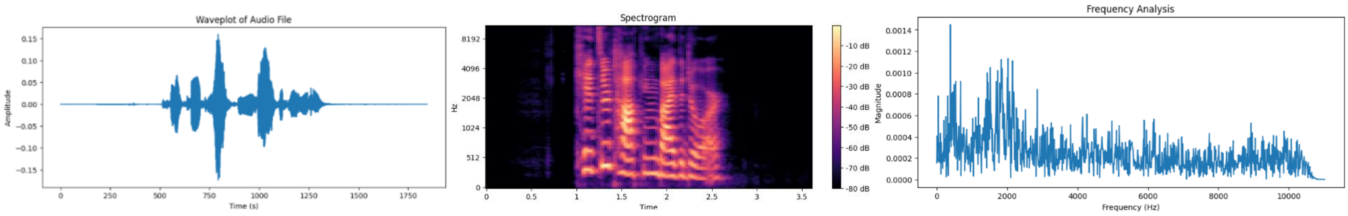
5. Fear



6. Sad



7.



Classification of Audio Signals

1. Tess

```
for directory in tess_directory_list:
    directory_path = os.path.join(Tess, directory)
    files = os.listdir(directory_path)
    for file in files:
        part = file.split('.')[0]
        part = part.split('_')[2]
        if part == 'ps':
            file_emotion.append('surprise')
        else:
            file_emotion.append(part)
        file_path.append(os.path.join(directory_path, file))
```

2. Savee

```python
for file in savee_directory_list:
    file_path.append(Savee + file)
    part = file.split('_')[1]
    ele = part[:-6]
    if ele=='a':
        file_emotion.append('angry')
    elif ele=='d':
        file_emotion.append('disgust')
    elif ele=='f':
        file_emotion.append('fear')
    elif ele=='h':
        file_emotion.append('happy')
    elif ele=='n':
        file_emotion.append('neutral')
    elif ele=='sa':
        file_emotion.append('sad')
    else:
        file_emotion.append('surprise')
```

3. Crema D

```python
for file in crema_directory_list:
    # storing file paths
    file_path.append(Crema + file)
    # storing file emotions
    part=file.split('_')
    if part[2] == 'SAD':
        file_emotion.append('sad')
    elif part[2] == 'ANG':
        file_emotion.append('angry')
    elif part[2] == 'DIS':
        file_emotion.append('disgust')
    elif part[2] == 'FEA':
        file_emotion.append('fear')
    elif part[2] == 'HAP':
        file_emotion.append('happy')
    elif part[2] == 'NEU':
        file_emotion.append('neutral')
```

4. Ravdess

```python
# Filter out directories that don't correspond to actor directories
actor_directories = [d for d in ravdess_directory_list if d.startswith('Actor_')]

for directory in actor_directories:
    actor_directory = os.path.join(ravdess, directory)  # Get the full path to the actor directory
    files = os.listdir(actor_directory)  # List all files for this actor
    for file in files:
        part = file.split('.')[0].split('-')
        # Ensure that the file name format is correct and contains at least three parts
        if len(part) >= 3:
            # Third part in each file represents the emotion associated with that file
            file_emotion.append(int(part[2]))
            file_path.append(os.path.join(actor_directory, file))  # Full path to the file
        else:
            print(f"Issue with filename: {file}")
```

Details of Dataset Used

1. RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song):

- Description: RAVDESS is a dataset designed for emotion recognition in speech and singing. It contains audio and video recordings of actors performing various emotional states.
- Number of Audio Files: RAVDESS consists of 7356 audio files in total.
- Gender Distribution: The dataset includes recordings from both male and female actors.
- Emotional States: Actors portray a range of emotional states, including neutral, calm, happy, sad, angry, fearful, surprise, and disgust.

- Feature Extraction through Librosa

Time Taken to Extract Feature	Number of Rows	Number of Columns
1435.2879178524017 seconds	1440	1527

- Feature Extraction through Wav2Vec Model:

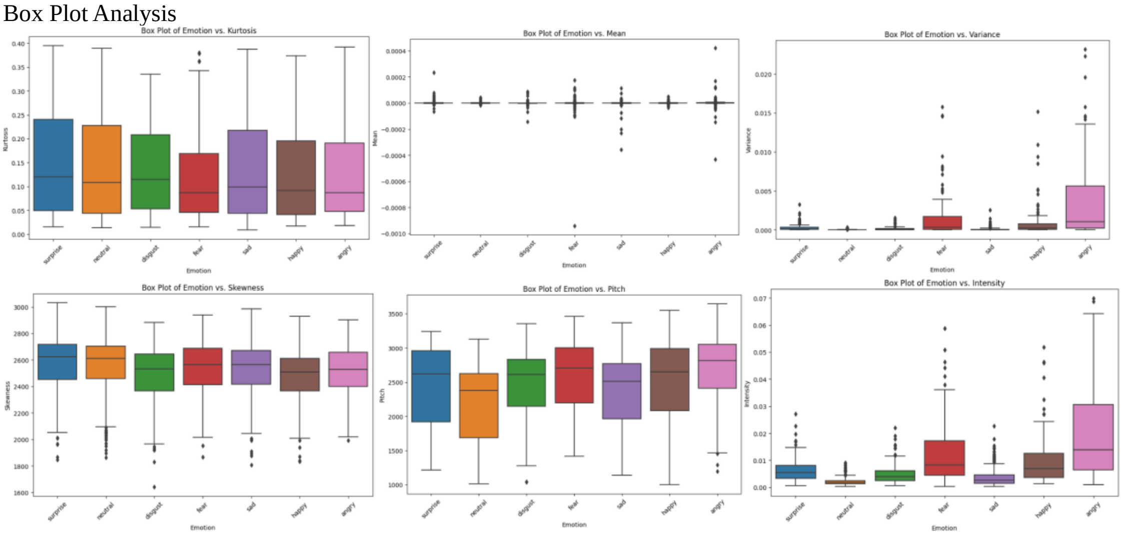
Time Taken to Extract Feature	Number of Rows	Number of Columns
6810.239865779877 seconds	1440	1025

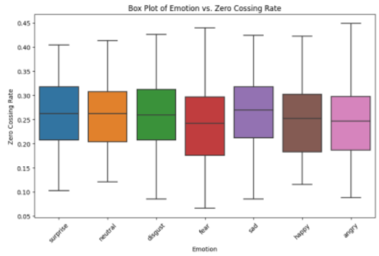
- Emotion Count:

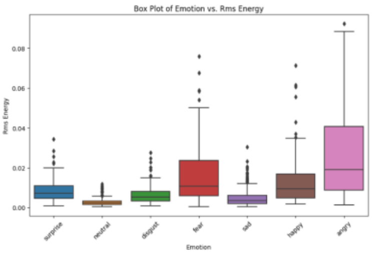
```
Emotions
neutral      288
surprise     192
disgust      192
fear         192
sad          192
happy        192
angry        192
Name: count, dtype: int64
```

- Relationship of Emotions with other features

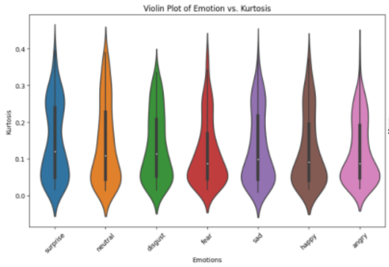
  - Box Plot Analysis

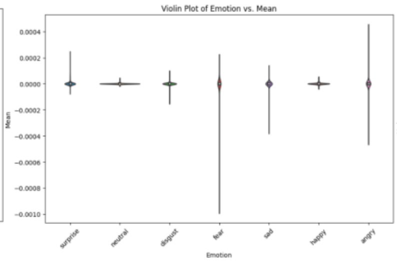


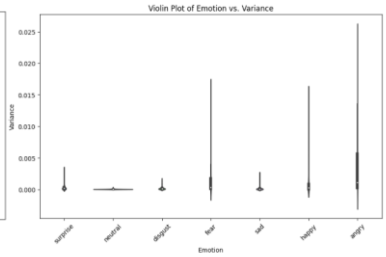


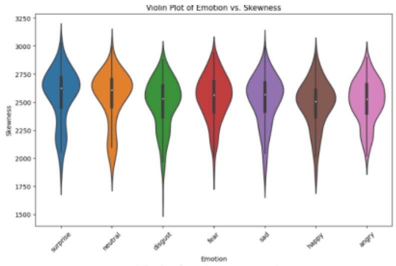


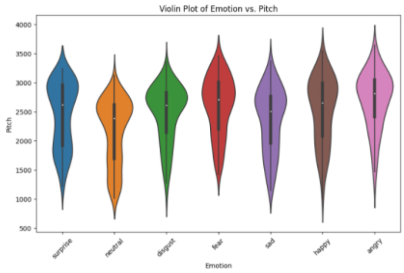
- Violin plot Analysis

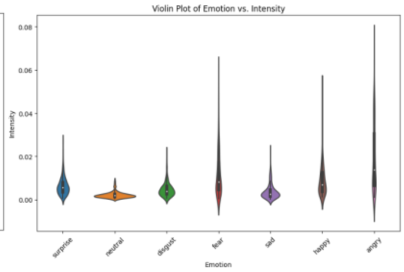


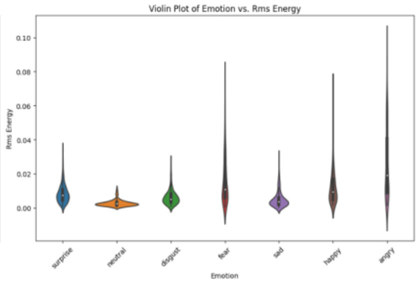












2. TESS (Toronto Emotional Speech Set):

- Description: TESS is a dataset created to study emotional speech perception. It contains audio recordings of a single female actor speaking short sentences with various emotional expressions.
- Number of Audio Files: TESS consists of 2800 audio files in total.
- Gender Distribution: The dataset includes recordings from a female actor.
- Emotional States: TESS covers a range of emotional expressions, including anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral.

- Feature Extraction through Librosa

Time Taken to Extract Feature	Number of Rows	Number of Columns
1859.3830995559692 seconds	2800	1329

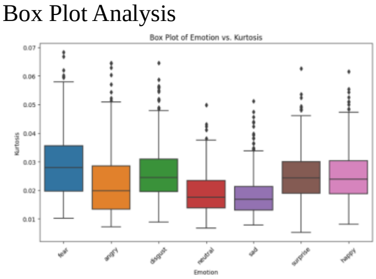
- Feature Extraction through Wav2Vec Model:

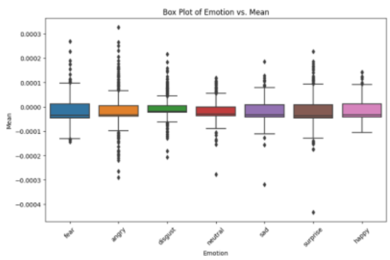
Time Taken to Extract Feature	Number of Rows	Number of Columns
2517.401907682419 seconds	2800	1025

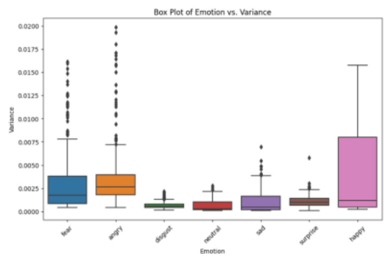
- Emotion Count:

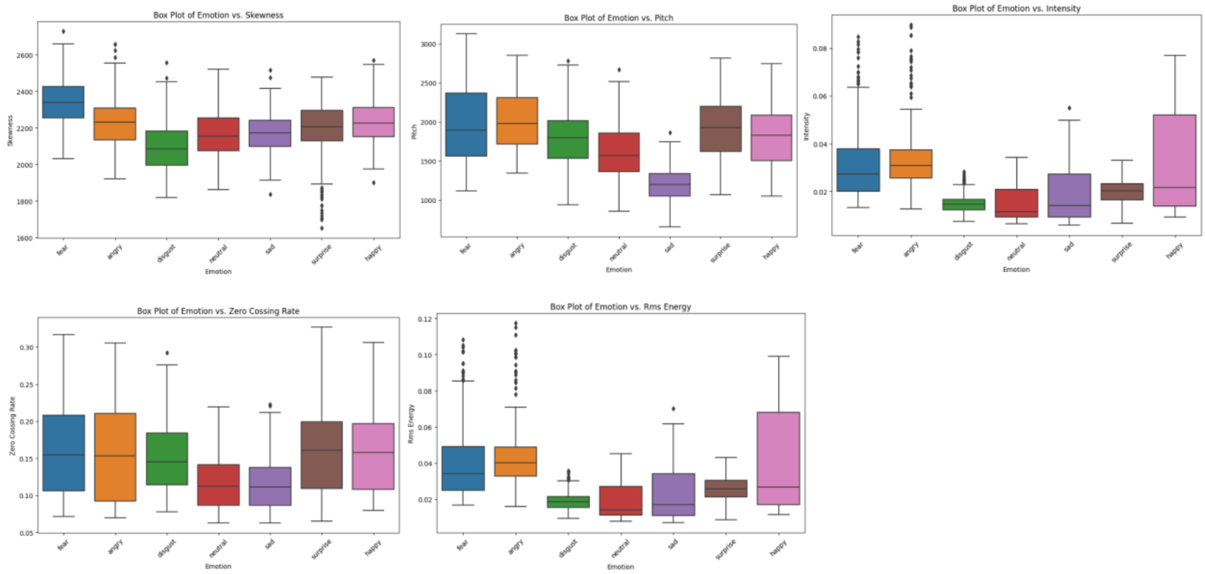
```
Emotions
fear        400
angry       400
disgust     400
neutral     400
sad         400
surprise    400
happy       400
Name: count, dtype: int64
```

- Relationship of Emotions with other features
  - Box Plot Analysis

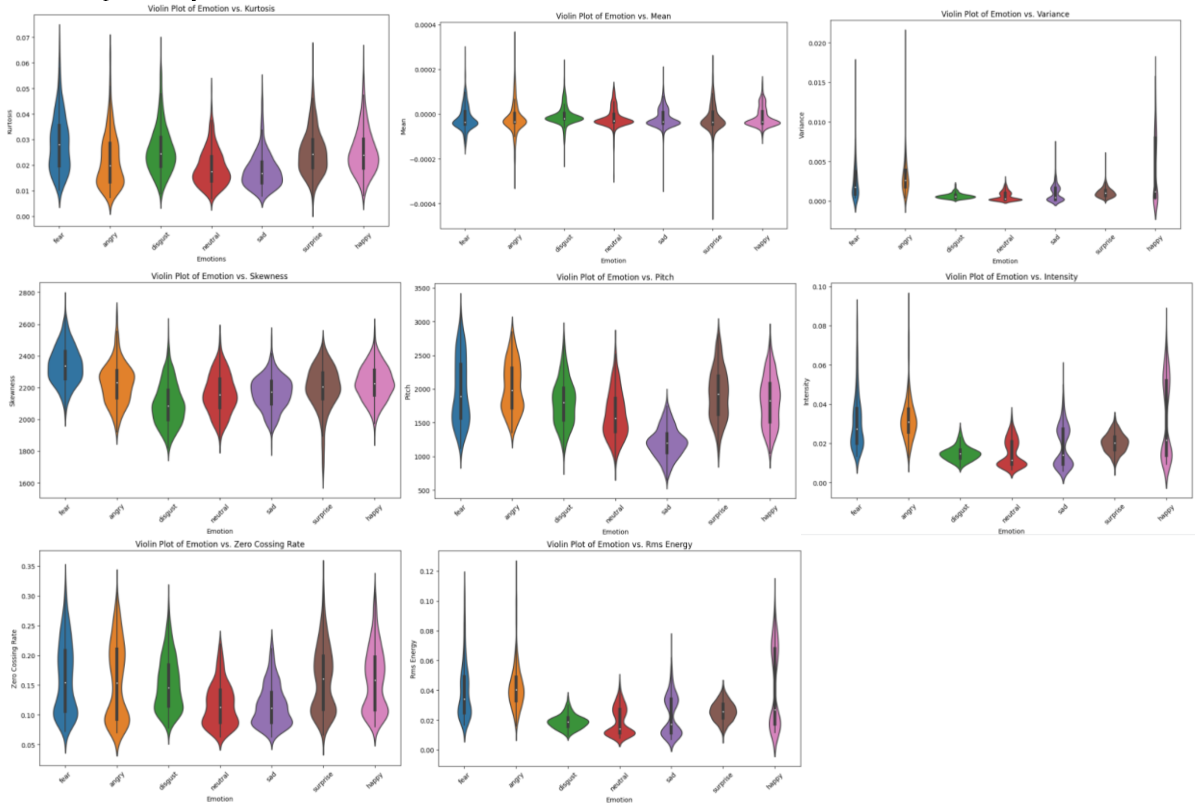








- Voilin plot Analysis



-

3. CREMA-D (CrowdEmotion Multimodal Affect Dataset):

- Description: CREMA-D is a dataset designed for emotion recognition research, containing audio and video recordings of actors portraying various emotional states.
- Number of Audio Files: CREMA-D consists of 7442 audio files in total.
- Gender Distribution: The dataset includes recordings from both male and female actors.
- Emotional States: CREMA-D covers a wide range of emotional expressions, including anger, disgust, fear, happiness, sadness, and neutral.

- Feature Extraction through Librosa

Time Taken to Extract Feature	Number of Rows	Number of Columns
5567.338131189346 second	7442	1503

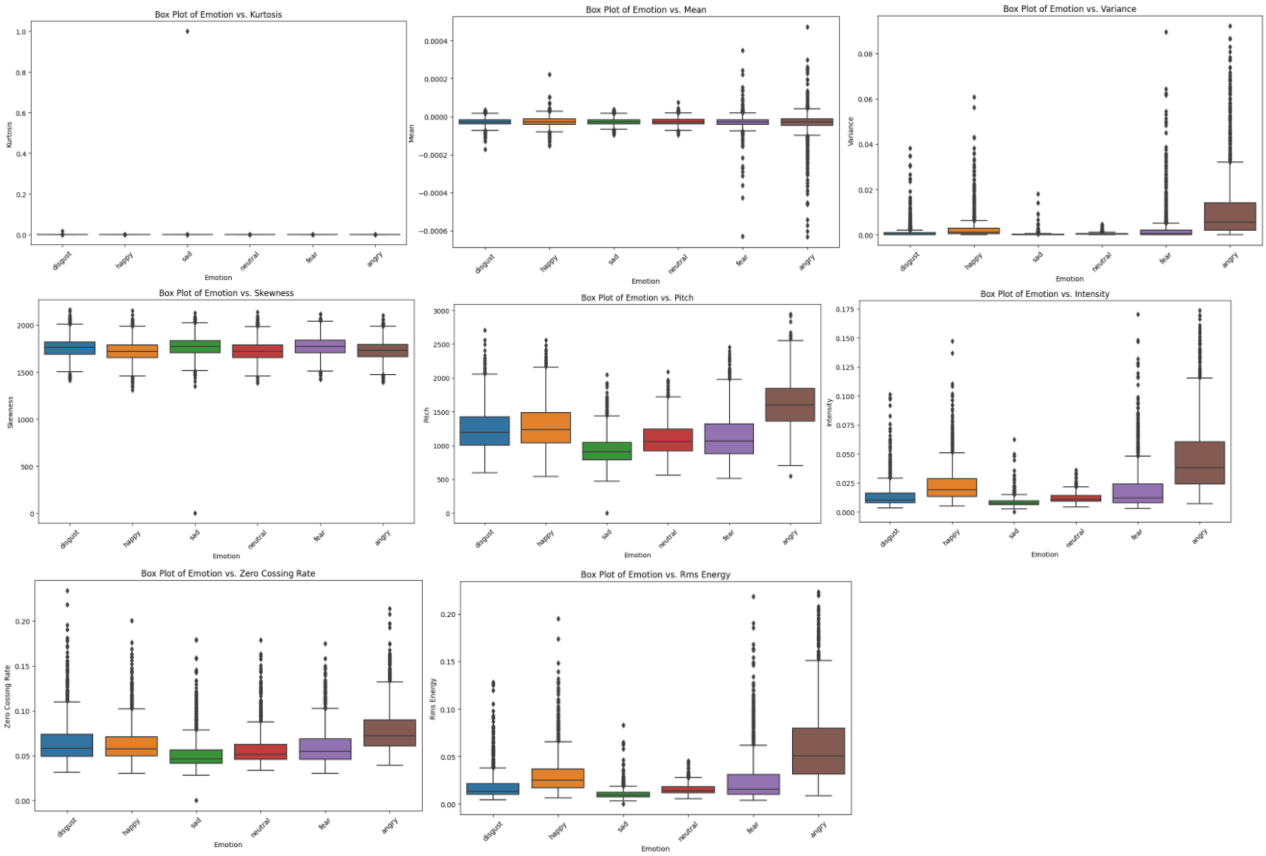
- Feature Extraction through Wav2Vec Model:

Time Taken to Extract Feature	Number of Rows	Number of Columns
6101.162845134735 seconds	7442	1025

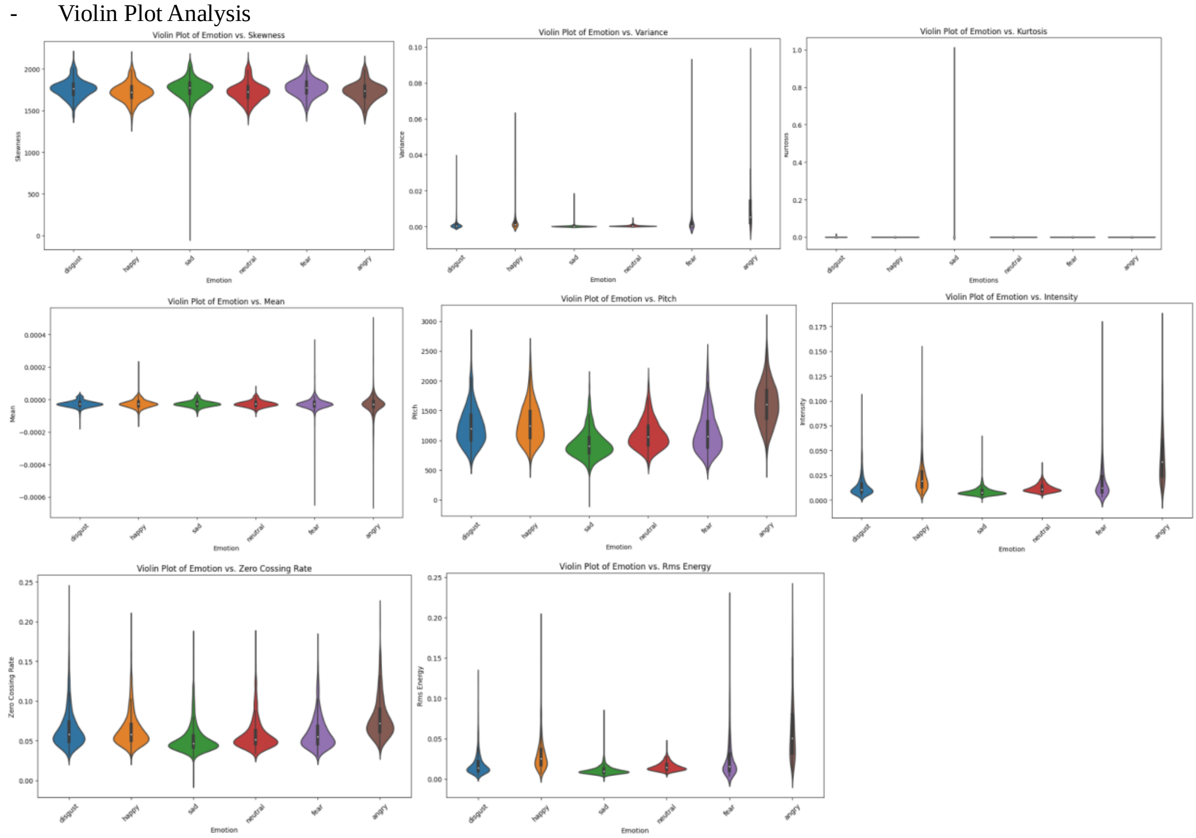
- Emotion Count:

```
Emotions
disgust    1271
happy      1271
sad        1271
fear       1271
angry      1271
neutral    1087
Name: count, dtype: int64
```

- Relationship of Emotions with other features
  - Box Plot Analysis



- Violin Plot Analysis



-

4. SAVEE (Surrey Audio-Visual Expressed Emotion):

- Description: SAVEE is a dataset developed for emotion recognition research, containing audio recordings of male actors speaking short sentences with various emotional expressions.
- Number of Audio Files: SAVEE consists of 480 audio files in total.
- Gender Distribution: The dataset includes recordings from male actors only.
- Emotional States: SAVEE includes four emotional expressions: anger, happiness, sadness, and neutral.
- Feature Extraction through Librosa:

Time Taken to Extract Feature	Number of Rows	Number of Columns
540.6044182777405 seconds	480	1380

- Feature Extraction through Wav2Vec Model:

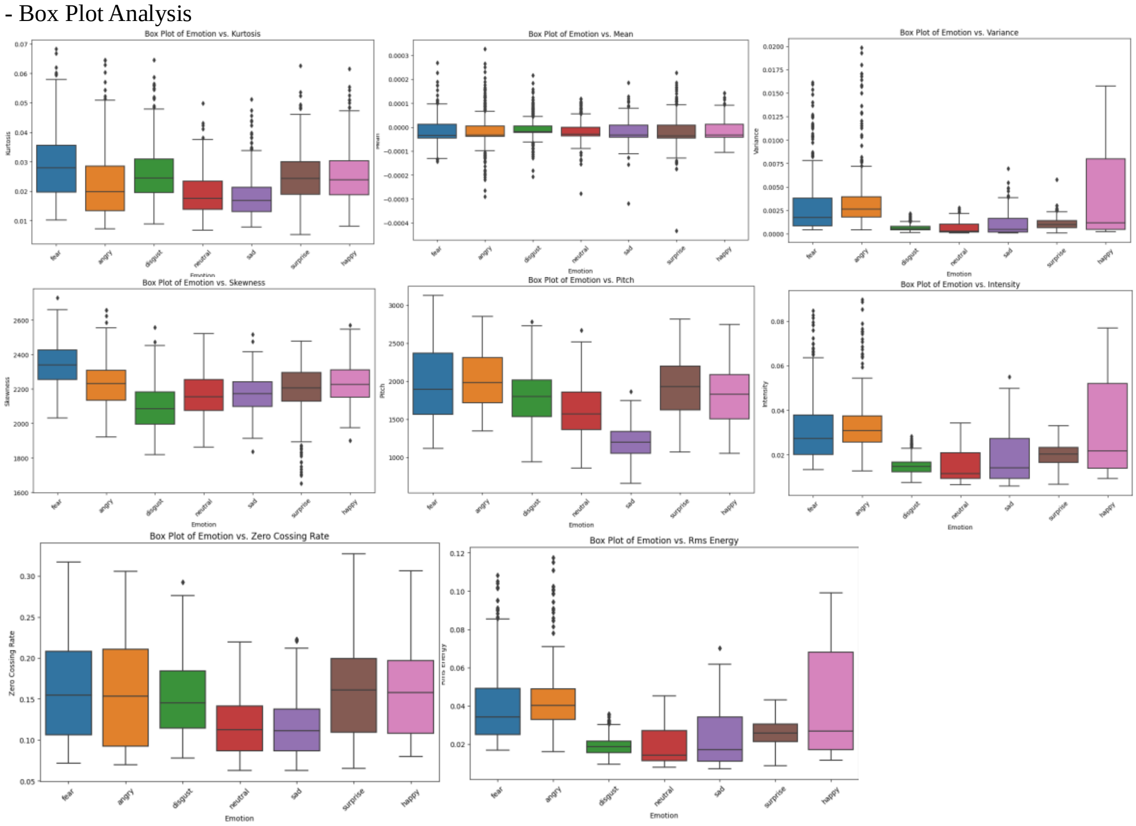
Time Taken to Extract Feature	Number of Rows	Number of Columns
540.6044182777405 seconds	480	1025

- Emotion Count:

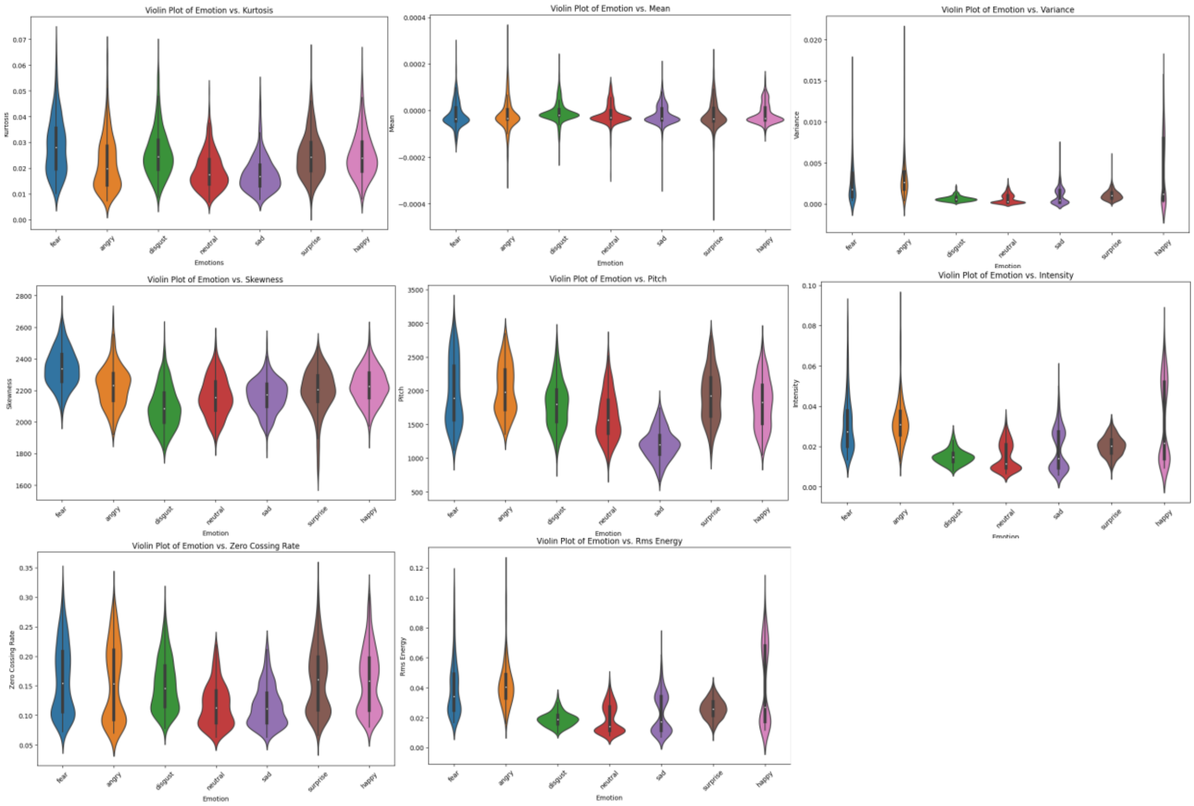
```
Emotions
fear        400
angry       400
disgust     400
neutral     400
sad         400
surprise    400
happy       400
Name: count, dtype: int64
```

- Relationship of Emotions with other features

  - Box Plot Analysis



  - Violin Plot Analysis



Methodology

1. Methods/ Models Used

- Logistic Regression:
  - Working:
    Logistic Regression is a linear model used for binary classification. It models the probability that an input belongs to a particular category.

- Steps:
  1. Initialize the model parameters (weights and bias).
  2. Compute the weighted sum of the input features.
  3. Apply the sigmoid function to the weighted sum to get the predicted probability.
  4. Threshold the predicted probability to make binary predictions.

- Naive Bayes:
  - Working:
    Naive Bayes is a probabilistic classifier based on Bayes' theorem with the "naive" assumption of feature independence.
  - Steps:
    1. Calculate the prior probability of each class.
    2. Calculate the likelihood of each feature given the class.
    3. Multiply the prior probability and likelihood to get the posterior probability.
    4. Choose the class with the highest posterior probability as the predicted class

- Decision Tree:
  - Working:
    Decision Tree is a tree-like model where each internal node represents a feature, each branch represents a decision, and each leaf node represents the outcome.
  - Steps:
    1. Select the best feature to split the data based on a criterion (e.g., Gini impurity, information gain).
    2. Split the data into subsets based on the selected feature.
    3. Recursively repeat steps 1 and 2 for each subset until a stopping criterion is met.
    4. Assign the majority class in each leaf node.

- Random Forest:
  - Working:
    Random Forest is an ensemble learning method that builds multiple decision trees and combines their predictions through voting.
  - Steps:
    Create a bootstrap sample from the training data.
    Build a decision tree using the bootstrap sample with a random subset of features at each node.
    Repeat steps 1 and 2 to build multiple decision trees.
    Aggregate the predictions of all trees through majority voting.

- K-Nearest Neighbours (KNN):
  - Working:
    KNN is a lazy learning algorithm that classifies an instance based on the majority class among its k nearest neighbours.
  - Steps:
  1. Calculate the distance between the test instance and all training instances.
  2. Select the k nearest neighbours based on the calculated distances.
  3. Assign the class label that is most frequent among the k nearest neighbours to the test instance.

- Gradient Boosting:
  - Working:
    Gradient Boosting is an ensemble learning method that builds multiple weak learners sequentially, with each new learner focusing on the errors of the previous one.
  - Steps:
  1. Initialize the model with a constant value.
  2. Fit a weak learner (e.g., decision tree) to the residual errors of the current model.
  3. Update the model by adding the predictions of the weak learner multiplied by a learning rate.
  4. Repeat steps 2 and 3 for a specified number of iterations.

- Voting Classifier:
  - Working:
    Voting Classifier combines the predictions of multiple individual models and predicts the class with the highest vote.
  - Steps:
  1. Train multiple base models on the same training data.
  2. For classification, each model predicts the class label.
  3. The Voting Classifier aggregates the predictions through majority voting (hard voting) or averaging the probabilities (soft voting).

- Feedforward Neural Network:
  - Working:
    Feedforward Neural Network, or Multilayer Perceptron, is a basic type of artificial neural network where information flows in one direction, from input to output.
  - Steps:
  1. Initialize the network architecture with input, hidden, and output layers.
  2. Connect the neurons between layers with weighted connections.
  3. Apply an activation function to the weighted sum of inputs at each neuron in the hidden layers.
  4. Calculate the output of the network using forward propagation.
  5. Update the weights using backpropagation and gradient descent.

- Deep Belief Network (DBN):
  - Working:
    DBNs are a type of deep neural network composed of multiple layers of stochastic, latent variables, with connections between the layers but not between units within each layer. They are trained in a greedy layer-wise manner followed by fine-tuning.
  - Steps:
  1. Pretrain each layer of the network as a Restricted Boltzmann Machine (RBM) using unsupervised learning.
  2. Greedily stack RBMs to form the deep belief network.
  3. Fine-tune the entire network using supervised learning with backpropagation.
  4. The network learns to represent complex patterns in the data and perform tasks such as classification or generation.

Convolutional Neural Network (CNN):
  - Working:

CNNs are deep learning models specifically designed for processing structured grid-like data such as images. They leverage convolutional layers to extract spatial hierarchies of features.

- Steps:
  1. Input images are passed through convolutional layers, where filters convolve over the input to extract features.
  2. The features are down sampled through pooling layers to reduce spatial dimensions and control overfitting.
  3. The resulting feature maps are flattened and connected to fully connected layers for classification.
  4. The network learns to classify images by adjusting the weights of the connections through backpropagation and gradient descent.

2. Data Preparation Techniques

- Data Collection:
  For audio emotion recognition, data collection involves gathering a diverse set of audio recordings that capture various emotional expressions. In this project, I used publicly available datasets such as RAVDESS, CREMA-D, SAVEE, and TESS from Kaggle, which include audio samples of different emotions spoken by multiple actors. These datasets provide a robust foundation for training and evaluating emotion recognition models.

- Data Preprocessing:
  Data preprocessing in audio emotion recognition involves preparing the raw audio data for analysis. This step includes loading the audio files, resampling them to a consistent sample rate, and trimming silence or unwanted noise. Preprocessing ensures that the audio data is in a suitable format for feature extraction and model training.

- Feature Extraction:
  Feature extraction is a crucial step where relevant characteristics are derived from the audio signals. Common features used in audio emotion recognition include Mel-frequency cepstral coefficients (MFCCs), chroma features, Mel-spectrogram, and zero-crossing rate. These features capture the essential qualities of the audio that are indicative of different emotions. I recorded the time taken to extract these features and the number of features extracted to evaluate the efficiency of this process.

- Data Normalization:
  Data normalization involves scaling the extracted features to ensure that they contribute equally during model training. Techniques such as min-max scaling or standardization are applied to normalize the features. This step helps in improving the convergence of the learning algorithms and achieving better model performance.

- Splitting the dataset:
  Splitting the dataset is necessary to evaluate the model's performance effectively. Typically, the dataset is divided into training, validation, and test sets. For this project, I used a standard split (e.g., 80-20 for training and testing) to ensure that the model is trained on a substantial portion of the data and tested on unseen data to gauge its generalization ability.

3. Evaluation Metrics

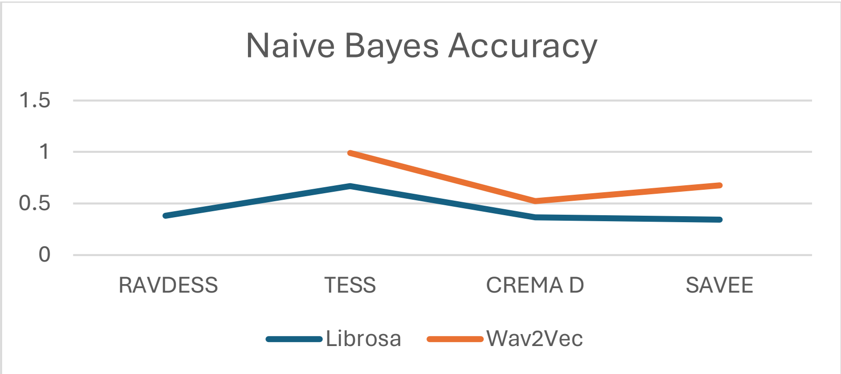
- Accuracy: Measures the overall correctness of the model.
- Precision: Assesses the model's ability to correctly identify specific emotions.
- Recall: Evaluates the model's completeness in capturing all relevant instances of specific emotions.
- F1 Score: The harmonic mean of precision and recall, useful for imbalanced datasets.
- Training Time: The time taken to train each model.
- Prediction Time: The time taken by each model to make predictions.
- Feature Extraction Time: The duration required to extract features from the audio data.
- Number of Features Extracted: The count of features extracted during preprocessing.

Results

1. Naïve Bayes

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.3819,	0.3591	3819	0.3474,	Training =0.1678 s Prediction =0.0608 s	Librosa
RAVDESS	Default	Min Max	0.3819,	0.3591	3819	0.3474,	Training = 0.1285s Prediction = 0.0546s	Librosa
RAVDESS	Default	Robust	0.3819,	0.3591	3819	0.3474,	Training = 0.1275s Prediction = 0.0569s	Librosa
RAVDESS	PCA =0.95	Robust	0.2257	0.2417	0.2257	0.1368	Training = 0.0038 s, Prediction = 0.0011 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3125	0.2706	0.3125	0.2422	Training = 0.0051 s, Prediction = 0.0022 s	Librosa
RAVDESS	LDA	Robust	0.2292	0.2826	0.2292	0.2213	Training = 0.0020 s, Prediction = 0.0005 s	Librosa
TESS	Default	Standard	0.6696	0.7482	0.6696	0.6704	Training =0.1620 s, Prediction =0.0582 s	Librosa
TESS	Default	Min Max	0.6696	0.7482	0.6696	0.6704	Training =0.1530 Prediction =0.0590	Librosa
TESS	Default	Robust	0.6696	0.7482	0.6696	0.6704	Training = 0.1427 s, Prediction = 0.0514 s	Librosa
CREMAD	Default	Standard	0.3660	0.4018	0.3660	0.3372,	Training = 0.3292 s, Prediction = 0.1009 s	Librosa
CREMAD	Default	Min Max	0.3660	0.4018	0.3660	0.3372,	Training = 0.3949 s, Prediction = 0.1534 s	Librosa
CREMAD	Default	Robust	0.3660	0.4018	0.3660	0.3372,	Training = 0.2804 s, Prediction = 0.1386 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3150	0.3229	0.3150	0.2831	Training = 0.0201 s,	Librosa

							Prediction = 0.0092 s	
CREMAD	PCA = 0.99	Robust	0.3224	0.3271	0.3224	0.2880	Training Time = 0.0374 s, Prediction = 0.0124 s	Librosa
SAVEE	Default	Standard	0.3438	0.3052	0.3229	0.3105	Training = 1.1236 s Prediction = 0.0482 s	Librosa
SAVEE	Default	Min Max	0.3438	0.3052	0.3229	0.3105	Training = 0.0728 s Prediction = 0.0465 s	Librosa
SAVEE	Default	Robust	0.3438	0.3052	0.3229	0.3105	Training = 0.0636 s Prediction = 0.0503 s	Librosa
SAVEE	PCA =0.95	Robust	0.3333	0.3367	0.3333	0.2427	Training = 0.0023 s Prediction = 0.0008 s	Librosa
SAVEE	PCA = 0.99	Robust	0.3333	0.2657,	0.3333	0.2236	Training = 0.0026 s Prediction = 0.0010 s	Librosa
SAVEE	LDA	Robust	0.2188	0.2921	0.2188	0.2403	Training = 0.0014 s Prediction = 0.0014	Librosa
TESS	Default	Robust	0.9911	0.9913	0.9911	0.9911	Training = 0.0976 s, Prediction = 0.0449 s	Wav2Vec
TESS	Default	Standard	0.9911	0.9911	0.9911	0.911	Training = 0.1388 s, Prediction = 0.0538 s	Wav2Vec
TESS	LDA	Robust	0.8000	0.8063	0.800	0.7943	Training = 0.0717 s, Prediction = 0.0003 s	Librosa
TESS	PCA = 0.95	Robust	0.4250	0.5252	0.4250	0.4351	Training = 0.0072 s, Prediction = 0.0032 s	Librosa
CREMAD	Default	Standard	0.4426	0.4369	0.4426	0.4164	Training = 0.1745 s, Prediction = 0.0785 s	Wav2Vec
CREMAD	Default	Min Max	0.4426	0.4369	0.4426	0.4164	Training = 0.2133 s, Prediction = 0.0715 s	Wav2Vec
CREMAD	Default	Robust	0.4426	0.4369	0.4426	0.4164	Training = 0.1823 s, Prediction = 0.0734 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.5238	0.5294	0.5238	0.5209	Training = 0.0566 s, Prediction = 0.0151 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.3150	0.3229	0.3150	0.2831	Training = 0.0201 s, Prediction = 0.0092 s	Wav2Vec
SAVEE	Default	Standard	0.5000	0.5090	0.5000	0.4848	Training = 0.0399 s, Prediction = 0.0352 s	Wav2Vec
SAVEE	Default	Min Max	0.5000	0.5090	0.5000	0.4848	Training = 0.0507 s, Prediction = 0.0461	Wav2Vec
SAVEE	Default	Robust	0.5000	0.5090	0.5000	0.4848	Training = 0.0386 s, Prediction = 0.0348 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.5938	0.5751	0.5938	0.5676	Training = 0.0028 s, Prediction = 0.0010 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5104	0.4137	0.5104	0.4456	Training = 0.0033 s, Prediction = 0.0014 s	Wav2Vec
SAVEE	LDA	Robust	0.6771	0.6912	0.6771	0.6655	Training = 0.0016 s Prediction = 0.0004 s	Wav2Vec

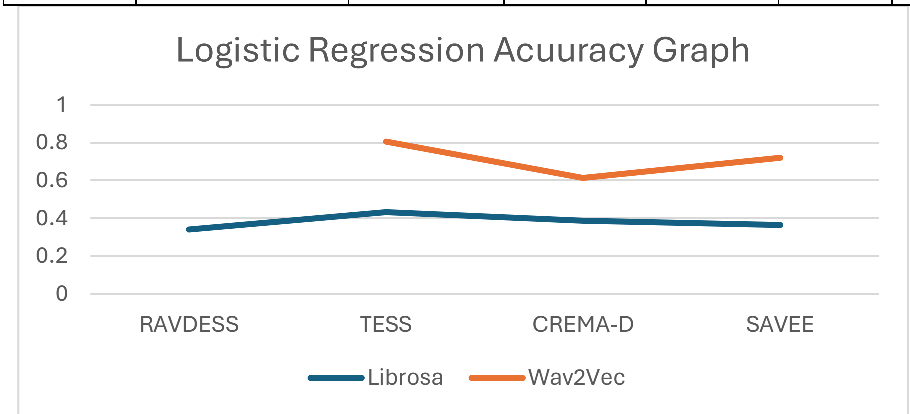


2. Logistic Regression

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.3264	0.2875	0.3264	0.2800	Training = 2.1172 s Prediction =0.1272 s	Librosa
RAVDESS	Default	Min Max	0.3403	0.3132	0.3403	0.3101	Training = 19.4166 s Prediction = 0.0867 s	Librosa
RAVDESS	Default	Robust	0.3403	0.3132	0.3403	0.3101	Training = 2.1172 s Prediction =0.1272 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2292	0.2024	0.2292	0.1746	Training = 0.0712 s, Prediction = 0.0008 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.2674,	0.2537	0.2674	0.2518	Training = 0.1522 s, Prediction = 0.0005 s	Librosa
RAVDESS	LDA	Robust	0.2778	0.2944	0.2778	0.2779	Training = 0.0271 s, Prediction = 0.0003 s	Librosa
TESS	Default	Standard	0.4321	0.4208	0.4321	0.4202	Training = 1.7261 s, Prediction = 0.0491 s	Librosa
TESS	Default	Min Max	0.4321	0.4208	0.4321	0.4202	Training = 1.7138 s, Prediction = 0.0493 s	Librosa
TESS	Default	Robust	0.4321	0.4208	0.4321	0.4202	Training = 1.6655 s,	Librosa

							Prediction = 0.0478 s	
CREMAD	Default	Standard	0.3694	0.3632	0.3694	0.3472	Training = 3.2753 s, Prediction = 0.0589 s	Librosa
CREMAD	Default	Min Max	0.3694	0.3632	0.3694	0.3472	Training = 3.1698 s, Prediction = 0.0597 s	Librosa
CREMAD	Default	Robust	0.3694	0.3632	0.3694	0.3472	Training = 3.0676 s, Prediction = 0.0412 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3036	0.2853	0.3036	0.2820	Training = 0.6202 s, Prediction = 0.0012 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3862	0.3858	0.3862	0.3816	Training = 0.9376 s, Prediction = 0.0013 s	Librosa
SAVEE	Default	Standard	0.3229	0.3153	0.3229	0.3144	Training = 1.1236 s, Prediction = 0.0482 s	Librosa
SAVEE	Default	Min Max	0.3229	0.3153	0.3229	0.3144	Training = 0.5769 s, Prediction = 0.0473 s	Librosa
SAVEE	Default	Robust	0. 3229	0.3153	0.3229	0.3144	Training = 0.5416 s, Prediction = 0.0458 s	Librosa
SAVEE	max_iter=1000, penalty="l2"	Robust	0.3125	0.3218	0.3125	0.3157	Training = 3.4741s, Prediction = 0.07272	Librosa
SAVEE	max_iter=1000, penalty="none	Robust	0.3021	0.3364	0.3021	0.3153	Training = 3.6147s, Prediction = 0.0710s	Librosa
SAVEE	max_iter=1000, solver="liblinear", penalty="l1"	Robust	0.3646	0.3659	0.3646	0.3626	Training = 6.9868s, Prediction = 0.0317s	Librosa
SAVEE	max_iter=1000, solver="liblinear", penalty="l2	Robust	0.3021	0.3009	0.3021	0.3012	Training = 31.0323s, Prediction = 0.0318s	Librosa
SAVEE	(max_iter=1000, solver="newton-cg", penalty="l2"	Robust	0.3333	0.3422	0.3333	0.3354	Training = 198.5196s, Prediction = 0.0723s	Librosa
SAVEE	max_iter=1000, solver="newton-cg", penalty="none"	Robust	0.3125	0.3270	0.3125	0.3170	Training = 8.6335s, Prediction = 0.0685s	Librosa
SAVEE	max_iter=1000,solve r="newton-cholesky" , penalty="l2"	Robust	0.2917	0.2953	0.2917	0.2925	Training = 22.0542s, Prediction = 0.0737s	Librosa
SAVEE	max_iter=1000, solver="newton-cholesky", penalty="none	Robust	0.3229	0.2942	0.3229	0.3003	Training = 6.4618s, Prediction = 0.0677s	Librosa
SAVEE	max_iter=1000, solver="sag", penalty="l2"	Robust	0.3438	0.3413	0.3438	0.3375	Training = 20.6752s, Prediction = 0.0324s	Librosa
SAVEE	max_iter=1000,solve r="sag", penalty="none"	Robust	0.3438	0.3413	0.3438	0.3375	Training = 20.6752s, Prediction = 0.0324s	Librosa
SAVEE	max_iter=1000, solver="saga", penalty="l2"	Robust	0.3125	0.3158	0.3125	0.3066	Training = 24.6310 s, Prediction = 0.0313 s	Librosa
SAVEE	max_iter=1000, solver="saga", penalty="elasticnet", l1_ratio=0.5	Robust	0.3125	0.3158	0.3125	0.3066	Training = 59.6154 s, Prediction = 0.0338 s	Librosa
SAVEE	max_iter=1000, solver="saga", penalty="l1"	Robust	0.3125	0.3158	0.3125	0.3066	Training = 57.6706 s, Prediction = 0.0324 s	Libroса
SAVEE	max_iter=1000, solver="saga", penalty="none"	Robust	0.3125	0.3158	0.3125	0.3066	Training = 24.4497 s, Prediction = 0.0316 s	Librosa
SAVEE	max_iter=1000, class_weight="balan ced	Robust	0.3229	0.3519	0.3329	0.3340	Training = 3.4886 s, Prediction = 0.0707 s	Librosa
SAVEE	max_iter=1000, C=10	Robust	0.2917	0.3302	0.2917	0.3065	Training = 3.5867 s, Prediction = 0.0680 s	Librosa
SAVEE	max_iter=1000, C=0.001	Robust	0.3542	0.3484	0.3542	0.3499	Training = 3.4920 s, Prediction = 0.0702 s	Librosa
SAVEE	max_iter=1000, C=0.1	Robust	0.3229	0.3283	0.3229	0.3242	Training = 3.4800 s, Prediction = 0.0697 s	Librosa
SAVEE	Default, PCA = 0.95	Robust	0.2812	0.2478	0.2812	0.2562	Training = 0.0565 s, Prediction = 0.0007 s	Librosa
SAVEE	Default, PCA = 0.99	Robust	0.2604	0.2544	0.2604	0.2519	Training = 0.0612 s, Prediction = 0.0006 s	Librosa
SAVEE	Default, LDA	Robust	0.2292	0.2595	0.2292	0.2404	Training = 0.0196 s, Prediction = 0.0003 s	Librosa
TESS	Default	Robust	0.7768	0.7737	0.7768	0.7356	Training = 0.1730 s, Prediction = 0.0327 s	Wav2Vec
TESS	Default	Standard	0.7768	0.7377	0.7768	0.7356	Training = 0.2579 s, Prediction = 0.0545 s	Wav2Vec
TESS	LDA	Robust	0.8054	0.8210	0.8054	0.8053	Training = 0.0717 s, Prediction = 0.0003 s	Wav2Vec
TESS	PCA = 0.95	Robust	0.4518,	0.4874	0.4518	0.4498	Training = 0.2070 s, Prediction = 0.0007 s	Librosa
TESS	LDA	Robust	0.4250	0.5252	0.4250	0.4351	Training = 0.0072 s, Prediction = 0.0032 s	Librosa

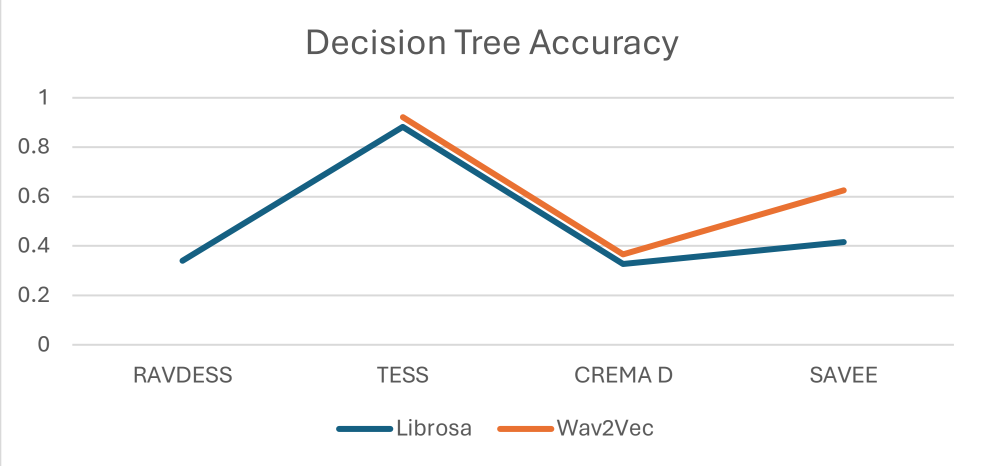
CREMAD	Default	Standard	0.5850	0.5809	0.5850	0.5811	Training = 1.8937 s, Prediction = 0.0736 s	Wav2Vec
CREMAD	Default	Min Max	0.5850	0.5809	0.5850	0.5811	Training = 1.8937 s, Prediction = 0.0736 s	Wav2Vec
CREMAD	Default	Robust	0.5850	0.5809	0.5850	0.5811	Training = 1.8531 s, Prediction = 0.0348 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.6132	0.6102	0.6132	0.6107,	Training = 0.9978 s, Prediction = 0.0027 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.5944	0.5933	0.5944	0.5934	Training = 1.4321 s, Prediction = 0.0027 s	Wav2Vec
SAVEE	Default	Standard, Min Max, Robust	0.5833	0.5912	0.5833	0.5663	Training = 0.2736, Prediction = 0.0308	Wav2Vec
SAVEE	max_iter=1000, penalty="none	Robust	0.60420	0.6330	0.6042	0.6100	Training = 0.2231 s, Prediction = 0.0293 s	Wav2Vec
SAVEE	max_iter=1000, solver="liblinear", penalty="l1"	Robust	0.5521	0.5546		0.5209	Training = 0.2900 s, Prediction = 0.0190 s	Wav2Vec
SAVEE	max_iter=1000, solver="liblinear", penalty="l2	Robust	0.5938	0.6077	0/5938	0.5697	Training = 0.6210 s, Prediction = 0.0176 s	Wav2Vec
SAVEE	(max_iter=1000, solver="newton-cg", penalty="l2"	Robust	0.5833	0.5912	0.5833	0.5663	Training = 0.1932 s, Prediction = 0.0329	Wav2Vec
SAVEE	max_iter=1000, solver="newton-cg", penalty="none"	Robust	0.6458	0.6669	0.6458	0.6506	Training = 0.3375 s, Prediction = 0.0291 s	Wav2Vec
SAVEE	max_iter=1000,solver="newton-cholesky", penalty="l2"	Robust	0.6042	0.6244		0.5829	Training = 5.1243 s, Prediction = 0.0356 s	Wav2Vec
SAVEE	max_iter=1000, solver="newton-cholesky", penalty="none	Robust	0.6146	0.6667	0.6146	0.6144	Training = 1.1338 s, Prediction = 0.0311 s	Wav2Vec
SAVEE	max_iter=1000, solver="sag", penalty="l2"	Robust	0.5833	0.5912	0.5833	0.5663	Training = 1.1137 s, Prediction = 0.0172 s	Wav2Vec
SAVEE	max_iter=1000,solver="sag", penalty="none"	Robust	0.6354	0.6559	0.6354	0.6354	Training = 16.4420 s, Prediction = 0.0222 s	Wav2Vec
SAVEE	max_iter=1000, solver="saga", penalty="l2"	Robust	0.5833	0.5912	0.5833	0.5663	Training = 2.9444 s, Prediction = 0.0170	Wav2Vec
SAVEE	max_iter=1000, solver="saga", penalty="elasticnet", l1_ratio=0.5	Robust	0.5938	0.6174		0.5771	Training = 3.7982 s, Prediction = 0.0174 s	Wav2Vec
SAVEE	max_iter=1000, solver="saga", penalty="l1"	Robust	0.5833	0.6037	0.5833	0.5625	Training = 14.8000 s, Prediction = 0.0211 s	Wav2Vec
SAVEE	max_iter=1000, solver="saga", penalty="none"	Robust	0.6354	0.6559	0.6354	0.6354	Training = 19.4504 s, Prediction = 0.0171 s	Wav2Vec
SAVEE	max_iter=1000, class_weight="balanced	Robust	0.5521	0.5685	0.5521	0.5505	Training = 0.3484 s, Prediction = 0.0291 s	Wav2Vec
SAVEE	max_iter=1000, C=10	Robust	0.6354	0.6727	0.6354	0.6411	Training = 1.1049 s, Prediction = 0.0366 s	Wav2Vec
SAVEE	max_iter=1000, C=0.001	Robust	0.3125	0.0977	0.3125	0.1488	Training = 0.1248 s, Prediction = 0.0299 s	Wav2Vec
SAVEE	max_iter=1000, C=0.1	Robust	0.4375	0.2976	0.4375	0.3271	Training = 0.3052 s, Prediction = 0.0310 s	Wav2Vec
SAVEE	Default, PCA = 0.95	Robust	0.6771	0.6844	0.6771	0.6763	Training =0.0617s, Prediction = 0.0006s	Wav2Vec
SAVEE	Default, PCA = 0.99	Robust	0.6875	0.7136	0.6875	0.6914	Training = 0.0774s, Prediction = 0.0006s	Wav2Vec
SAVEE	Default, LDA	Robust	0.7188	0.7402	0.7188	0.7151	Training = 0.0175 s, Prediction = 0.0002 s	Wav2Vec



3. Decision Tre

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.3021	0.3143	0.3021	0.3055	Training = 1.6986 s Prediction =0.0376 s	Librosa
RAVDESS	Default	Min Max	0.3403	0.3572	0.3403	0.3461	Training = 1.7352 s Prediction = 0.0374 s	Librosa
RAVDESS	Default	Robust	0.3229	0.3336	0.3229	0.3245	Training = 1.7710 s Prediction = 0.0350 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2847	0.2889	0.2847	0.2731	Training = 0.0142 s, Prediction = 0.0004 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3056	0.2864	0.3056	0.2724	Training = 0.1219 s, Prediction = 0.0004 s	Librosa
RAVDESS	LDA	Robust	0.1979	0.2212	0.1979,	0.1861	Training = 0.0061 s, Prediction = 0.0008 s	Librosa
TESS	Default	Standard	0.8339	0.8831	0.8839	0.8829	Training =2.8709 Prediction =0.0236	Librosa
TESS	Default	Min Max	0.8661	0.8673	0.8661	0.8655	Training =2.8739 s, Prediction = 0.0233 s	Librosa
TESS	Default	Robust	0.8821	0.8836	0.8821	0.8823	Training = 2.9229 s, Prediction = 0.0235 s	Librosa
TESS	PCA =0.95	Robust	0.4946	0.5297	0.4946	0.4885	Training = 0.1285 s, Prediction = 0.0004 s	Librosa
CREMAD	Default	Standard	0.3217	0.3205	0.3217	0.3209	Training = 12.0937 s, Prediction = 0.0307 s	Librosa
CREMAD	Default	Min Max	0.3271	0.3270	0.3271	0.3269	Training = 12.0868 s, Prediction = 0.0329 s	Librosa
CREMAD	Default	Robust	0.3156	0.3132	0.3156	0.3139	Training = 12.0809 s, Prediction =0.0230s	Librosa
CREMAD	PCA = 0.95	Robust	0.3183	0.3022	0.3183	0.2521	Training = 0.5402 s, Prediction = 0.0008 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3069	0.2196	0.3069	0.2447,	Training = 0.9872 s, Prediction = 0.0010 s	Librosa
SAVEE	Default	Standard	0.3646	0.4196	0.3646	0.3782	Training = 0.4087 s Prediction = 0.0217s	Librosa
SAVEE	Default	Min Max	0.3646	0.3951	0.3646	0.3656	Training = 0.4138s, Prediction = 0.0221s	Librosa
SAVEE	Default	Robust	0.3229	0.3413	0.3229	0.3187	Training = 0.4017s, Prediction = 0.0212s	Librosa
SAVEE	criterion='gini', splitter='best'	Robust	0.3333	0.3669	0.3333	0.3412	Training = 0.3851s, Prediction = 0.0308s	Librosa
SAVEE	criterion='gini', splitter='random	Robust	0.3229	0.3346	0.3229	0.3186	Training = 0.0795s, Prediction = 0.0310s	Librosa
SAVEE	(criterion='entropy', splitter='best'	Robust	0.4167	0.4373	0.4167	0.4234	Training = 0.7014s, Prediction = 0.0332s	Librosa
SAVEE	(criterion='entropy', splitter='random'	Robust	0.4062	0.3872	0.4062	0.3890	Training = 0.0842s, Prediction = 0.0313s	Librosa
SAVEE	max_depth=10	Robust	0.3229	0.3669	0.3229	0.3349	Training = 0.3620s, Prediction = 0.0318s	Librosa
SAVEE	max_depth=20	Robust	0.3750	0.4276	0.3750	0.3900	Training = 0.3650s, Prediction = 0.0322s	Librosa
SAVEE	max_depth=5	Robust	0.3021	0.4495	0.3021	0.3104	Training = 0.2625s, Prediction = 0.0321s	Librosa
SAVEE	min_samples_split=1 0	Robust	0.2917	0.3454	0.2917	0.3049	Training = 0.3537s, Prediction = 0.0314s	Librosa
SAVEE	min_samples_split=5	Robust	0.3333	0.3775	0.3333	0.3461	Training = 0.3550s, Prediction = 0.0296s	Librosa
SAVEE	min_samples_split=2	Robust	0.3542	0.3811	0.3542	0.3549	Training = 0.3172s, Prediction = 0.0325s	Librosa
SAVEE	min_samples_leaf=2	Robust	0.3021	0.3292	0.3021	0.3108	Training = 0.3599s, Prediction = 0.0317s	Librosa
SAVEE	min_samples_leaf=5	Robust	0.3438	0.3858	0.3430	0.3521	Training = 0.3284s, Prediction = 0.0345s	Librosa
SAVEE	min_samples_leaf=1 0	Robust	0.3646	0.4612	0.3646	0.3626	Training = 0.2823s, Prediction = 0.0333s	Librosa
SAVEE	min_weight_fraction _leaf=0.1	Robust	0.4062	0.3469	0.4062	0.3551	Training = 0.1784s, Prediction = 0.0321s	Librosa
SAVEE	min_weight_fraction _leaf=0.2	Robust	0.3333	0.1849	0.3333	0.3551	Training = 0.1417s, Prediction = 0.0320s,	Librosa
SAVEE	max_features='sqrt’	Robust	0.2604	0.2919	0.2604	0.2713s	Training = 0.0443s, Prediction = 0.0338s	Librosa
SAVEE	max_features='log2'	Robust	0.3438	0.3549	0.3438	0.3428	Training = 0.0359s, Prediction = 0.0326s	Librosa
SAVEE	max_features=0.5,	Robust	0.3958	0.4178	0.3958	0.3974	Training = 0.1871s, Prediction = 0.0338s	Librosa
SAVEE	random_state=42	Robust	0.3229	0.3779	0.3229	0.3376	Training = 0.3664s, Prediction = 0.0321s	Librosa
SAVEE	random_state=123	Robust	0.3125	0.3581	0.3125	0.3240	Training = 0.3587s, Prediction = 0.0320s	Librosa
SAVEE	PCA = 0.95	Robust	0.2604	0.3072	0.3604	0.2327	Training = 0.0133s, Prediction = 0.0002s	Librosa
SAVEE	PCA = 0.99	Robust	0.2604	0.2116	0.2604	0.2142	Training = 0.0303s, Prediction = 0.0003s	Librosa
SAVEE	LDA	Robust	0.1875	0.2884	0.1875	0.1934	Training = 0.0022s,	Librosa

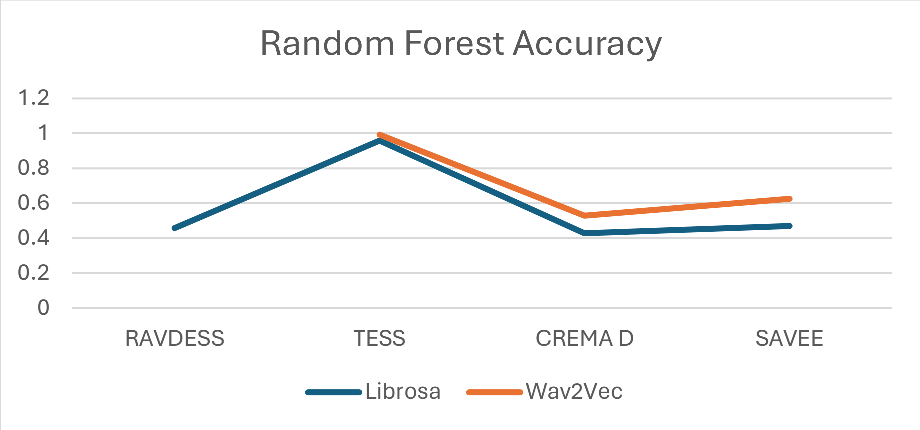
							Prediction = 0.0002s	
TESS	Default	Robust	0.9214	0.9231	0.9214	0.9220	Training = 1.7056 s, Prediction = 0.0187 s	Wav2Vec
TESS	Default	Standard	0.9196	0.9241	0.9196	0.9206	Training = 1.7385 s, Prediction = 0.0254 s	Wav2Vec
TESS	LDA	Robust	0.7536	0.7669	0.7536	0.7488	Training = 0.0115 s, Prediction = 0.0005 s	Librosa
CREMAD	Default	Standard	0.3499	0.3473	0.3499	0.3482	Training = 8.7460 s, Prediction = 0.0200 s	Wav2Vec
CREMAD	Default	Min Max	0.3472	0.3465	0.3472	0.3465	Training = 8.8007 s, Prediction = 0.0200 s	Wav2Vec
CREMAD	Default	Robust	0.3586	0.3571	0.3586	0.3575	Training = 8.6584 s, Prediction = 0.0201 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.3754	0.3618	0.3754	0.3449,	Training = 1.3319 s, Prediction = 0.0015 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.3660	0.3519	0.3660	0.3388,	Training = 2.7107 s, Prediction = 0.0023 s	Wav2Vec
SAVEE	Default	Standard	0.3646	0.3888	0.3646	0.3707	Training = 0.2790 s, Prediction = 0.0173 s	Wav2Vec
SAVEE	Default	Min Max	0.3646	0.4081	0.3646	0.3719	Training = 0.2797 s, Prediction = 0.0168 s	Wav2Vec
SAVEE	Default	Robust	0.3958	0.4447	0.3958	0.4109	Training = 0.2842 s, Prediction = 0.0173 s	Wav2Vec
SAVEE	criterion='gini', splitter='best'	Robust	0.3125	0.3377	0.3125	0.3176	Training = 0.2670 s, Prediction = 0.0168 s	Wav2Vec
SAVEE	criterion='gini', splitter='random	Robust	0.3646	0.3840	0.3646	0.3646	Training = 0.0621 s, Prediction = 0.0168 s	Wav2Vec
SAVEE	(criterion='entropy', splitter='best'	Robust	0.3229	0.3151	0.3229	0.3159	Training = 0.5484 s, Prediction = 0.0168 s	Wav2Vec
SAVEE	(criterion='entropy', splitter='random'	Robust	0.4062	0.3989	0.4062	0.3859	Training = 0.0660 s, Prediction = 0.0170 s	Wav2Vec
SAVEE	max_depth=10	Robust	0.3438	0.3737	0.3438	0.3500	Training = 0.2692 s, Prediction = 0.0169 s	Wav2Vec
SAVEE	max_depth=20	Robust	0.3333	0.3737	0.3333	0.3411	Training = 0.2681 s, Prediction = 0.0169 s	Wav2Vec
SAVEE	max_depth=5	Robust	0.3854	0.4439	0.3854	0.3992	Training = 0.2150 s, Prediction = 0.0174 s	Wav2Vec
SAVEE	min_samples_split=10	Robust	0.4062	0.4580	0.4062	0.4159	Training = 0.2576 s, Prediction = 0.0178 s	Wav2Vec
SAVEE	min_samples_split=5	Robust	0.3958	0.4372	0.3958	0.4036	Training = 0.2627 s, Prediction = 0.0169 s	Wav2Vec
SAVEE	min_samples_split=2	Robust	0.4479	0.5170	0.4479	0.4588	Training = 0.2433 s, Prediction = 0.0168 s	Wav2Vec
SAVEE	min_samples_leaf=2	Robust	0.3125	0.3321	0.3125	0.3186	Training = 0.2587 s, Prediction = 0.0173 s	Wav2Vec
SAVEE	min_samples_leaf=5	Robust	0.4375	0.4859	0.4375	0.4452	Training = 0.2440 s, Prediction = 0.0169 s	Wav2Vec
SAVEE	min_samples_leaf=10	Robust	0.3646	0.4162	0.3646	0.3747	Training = 0.2017 s, Prediction = 0.0180 s	Wav2Vec
SAVEE	min_weight_fraction_leaf=0.1	Robust	0.4167	0.3443	0.4167	0.3704	Training = 0.1219 s, Prediction = 0.0170 s	Wav2Vec
SAVEE	min_weight_fraction_leaf=0.2	Robust	0.3438	0.2853	0.3438	0.2994	Training = 0.0992 s, Prediction = 0.0167 s	Wav2Vec
SAVEE	max_features='sqrt’	Robust	0.3438	0.4415	0.3438	0.3699	Training = 0.0288 s, Prediction = 0.0177 s	Wav2Vec
SAVEE	max_features='log2'	Robust	0.3125	0.3407	0.3125	0.3218	Training = 0.0211 s, Prediction = 0.0182 s	Wav2Vec
SAVEE	max_features=0.5,	Robust	0.3750	0.3991	0.3750	0.3769	Training = 0.0963 s, Prediction = 0.0169 s	Wav2Vec
SAVEE	random_state=42	Robust	0.3958	0.4316	0.3958	0.4039	Training = 0.2672 s, Prediction = 0.0167 s	Wav2Vec
SAVEE	random_state=123	Robust	0.4479	0.5135	0.4479	0.4604	Training = 0.2669 s, Prediction = 0.0166 s	Wav2Vec
SAVEE	PCA = 0.95	Robust	0.4583	0.4564	0.4583	0.4390	Training = 0.0316s, Prediction =0.0002s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.4583	0.4564	0.4583	0.4390	Training = 0.0545s, Prediction = 0.0003s	Wav2Vec
SAVEE	LDA	Robust	0.6250	0.6458	0.6250	0.6148	Training = 0.0018s, Prediction = 0.0002s	Wav2Vec



4. Random Forest

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0. 4375	0.4146	0.4375	0.4035	Training = 3.5694 s Prediction = 0.0533 s	Librosa
RAVDESS	Default	Min Max	0.4583	0.4338	0.4583	0.4242	Training = 3.6570 s, Prediction = 0.0530 s	Librosa
RAVDESS	Default	Robust	0.4306	0. 4089	0.4306	0.4001	Training = 3.5993 s Prediction = 0.0506 s	Librosa
RAVDESS	PCA =0.95	Robust	0.3056	0.3732	0.3056	0.2534	Training = 0.3953 s, Prediction = 0.0087 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3333	0.3058	0.3333	0.2718	Training = 0.7050 s, Prediction = 0.0087 s	Librosa
RAVDESS	LDA	Robust	0.2188	0.2368	0.2188	0.2196	Training = 0.2736 s, Prediction = 0.0100 s	Librosa
TESS	Default	Standard	0.9536	0.9550	0.9536	0.9534	Training = 6.2261 s, Prediction = 0.0464 s	Librosa
TESS	Default	Min Max	0.9589	0.9601	0.9589	0.9589	Training = 6.2194 s, Prediction = 0.0444 s	Librosa
TESS	Default	Robust	0.9482	0.9503	0.9482	0.9484	Training = 6.1732 s, Prediction = 0.0469 s	Librosa
TESS	PCA = 0.95	Robust	0.5643	0.5645	0.5643	0.5396	Training = 1.0363 s, Prediction = 0.0141 s	Librosa
CREMAD	Default	Standard	0.4238	0.4012	0.4238	0.4012	Training = 23.3002 s, Prediction = 0.1110 s	Librosa
CREMAD	Default	Min Max	0.4184	0.3829	0.4184	0.3898	Training = 23.3146 s, Prediction = 0.1000 s	Librosa
CREMAD	Default	Robust	0.4285	0.4015	0.4285	0.4050	Training = 22.9618 s, Prediction = 0.1047 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3620	0.3314	0.3620	0.2718,	Training = 3.2362 s, Prediction = 0.0203 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3559	0.3196	0.3559	0.2606,	Training = 4.3176 s, Prediction = 0.0211 s	Librosa
SAVEE	Default	Standard	0.4688	0.4698	0.4688	0.4030	Training = 1.0606 s Prediction = 0.0312s	Librosa
SAVEE	Default	Min Max	0.4583	0.3851	0.4583	0.3845	Training = 1.1071 s, Prediction = 0.0345 s	Librosa
SAVEE	Default	Robust	0.4688	0.4650	0.4688	0.4108	Training = 1.0661 s, Prediction = 0.0321 s	Librosa
SAVEE	PCA =0.95	Robust	0.3333	0.2304	0.3333	0.2276	Training = 0.3636 s, Prediction = 0.0083 s	Librosa
SAVEE	PCA = 0.99	Robust	0.3438	0.1660	0.3438	0.2051	Training = 0.3892 s, Prediction = 0.0081 s	Librosa
SAVEE	LDA	Robust	0.2083	0.2504	0.2083	0.2200	Training = 0.2299 s, Prediction = 0.0069 s	Librosa
TESS	Default	Robust	0.9929	0.9929	0.9929	0.9929	Training = 0.0303 s, Prediction = 0.1770 s	Wav2Vec
TESS	LDA	Robust	0.7875	0.7869	0.7875	0.7821	Training = 0.4224 s, Prediction = 0.0131	Wav2Vec
CREMAD	Default	Standard	0.5285	0.5227	0.5285	0.5162	Training = 18.9548 s, Prediction = 0.0842 s	Wav2Vec
CREMAD	Default	Min Max	0.5225	0.5158	0.5225,	0.5084	Training = 19.1707 s, Prediction ime = 0.0905 s	Wav2Vec
CREMAD	Default	Robust	0.5292	0.5214	0.5292	0.5147	Training = 18.6303 s, Prediction = 0.0846 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.4668	0.4892	0.4668	0.4448,	Training = 5.0934 s, Prediction = 0.0258 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.4419	0.4572	0.4419	0.4142	Training = 7.1109 s, Prediction = 0.0286 s	Wav2Vec
SAVEE	Default	Standard	0.6146,	0.6068,	0.6146	0.5743	Training = 0.8207 s, Prediction = 0.0246 s	Wav2Vec
SAVEE	Default	Min Max	0.5417	0.5650	0.5417	0.4866	Training = 0.8288 s, Prediction = 0.0238 s	Wav2Vec
SAVEE	Default	Robust	0.6042	0.6521	0.6042	0.5741	Training = 0.8288 s, Prediction = 0.0238 s	Wav2Vec

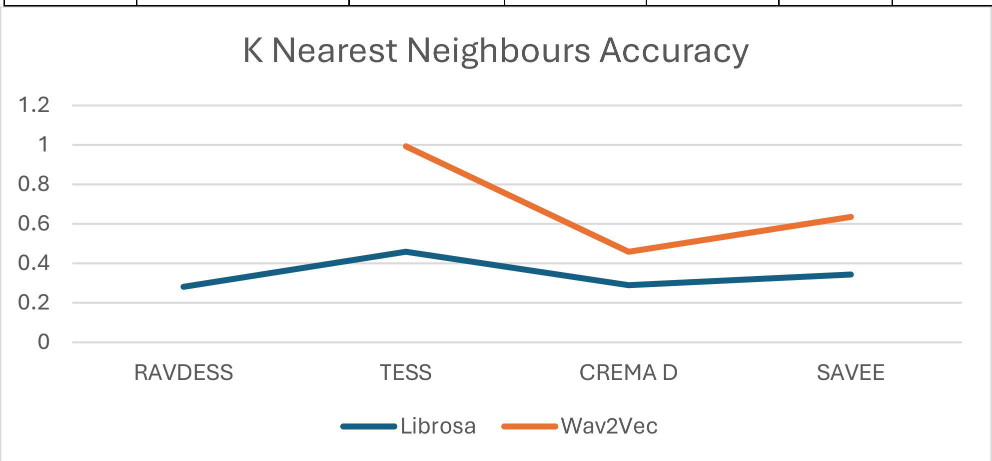
SAVEE	PCA =0.95	Robust	0.5000	0.4698	0.5000	0.4029	Training = 0.3953 s, Prediction = 0.0082 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.4271	0.4499	0.4271	0.3186	Training = 0.4354 s, Prediction = 0.0074 s	Wav2Vec
SAVEE	LDA	Robust	0.6250	0.6069	0.6250	0.6010	Training = 0.2289 s, Prediction = 0.0072 s	Wav2Vec
TESS	Default	Robust	0.9929	0.9929	0.9929	0.9929	Training = 0.0303 s, Prediction = 0.1770 s	Wav2Vec



5. KNN

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.2743	0.2463	0.2743	0.2470	Training =0.0592s Prediction =0.2481s	Librosa
RAVDESS	Default	Min Max	0.2743	0.2463	0.2743	0.2470	Training = 0.0539 s Prediction =0.0880 s	Librosa
RAVDESS	Default	Robust	0.2743	0.2463	0.2743	0.2470	Training = 0.0504s Prediction = 0.0966s	Librosa
RAVDESS	PCA =0.95	Robust	0.2431	0.2429	0.2431	0.2363	Training = 0.0012 s, Prediction = 0.0201 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.2500	0.2161	0.2500	0.2057	Training = 0.0013 s, Prediction = 0.0204 s	Librosa
RAVDESS	LDA	Robust	0.2812	0.2886	0.2812	0.2821	Training = 0.0026 s, Prediction = 0.0224 s	Librosa
TESS	Default	Standard	0.4589	0.5088	0.4589	0.4477	Training = 0.0448 Prediction = 0.1959	Librosa
TESS	Default	Min Max	0.4589	0.5088	0.4589	0.4477	Training =0.0435 s, Prediction = 0.1359	Librosa
TESS	Default	Robust	0.4589	0.5088	0.4589	0.4477	Training =0.0459 Prediction =0.2339	Librosa
TESS	PCA = 0.95	Robust	0.4518	0.4588	0.4518	0.4486	Training = 0.0014 s, Prediction = 0.1166 s	Librosa
CREMAD	Default	Standard	0.2901	0.3062	0.2901	0.2873	Training = 0.1184 s, Prediction = 0.6747 s	Librosa
CREMAD	Default	Min Max	0.2901	0.3062	0.2901	0.2873	Training = 0.1383 s, Prediction = 0.5989 s	Librosa
CREMAD	Default	Robust	0.2901	0.3062	0.2901	0.2873	Training = 0.0913 s, Prediction = 0.5956 s	Librosa
CREMAD	PCA = 0.95	Robust	0.2512	0.2842	0.2512	0.2128	Training = 0.0042 s, Prediction = 0.1717 s	Librosa
CREMAD	PCA = 0.99	Robust	0.2277	0.2715	0.2277	0.1882	Training = 0.0071 s, Prediction = 0.1517 s	Librosa
SAVEE	Default	Standard	0.3021	0.2336	0.3021	0.2418	Training = 0.0277 s, Prediction = 0.0343s	Librosa
SAVEE	Default	Min Max	0.3021	0.2336	0.3021	0.2418	Training = 0.0255 s, Prediction = 0.0397 s	Librosa
SAVEE	Default	Robust	0.3021	0.2336	0.3021	0.2418	Training = 0.0269 s, Prediction = 0.0384 s	Librosa
SAVEE	n_neighbors=3, weights='uniform'	Robust	0.3333	0.3699	0.3333	0.3015	Training = 0.0388 s, Prediction = 0.1121 s	Librosa
SAVEE	n_neighbors=5, weights='uniform'	Robust	0.3021	0.2336	0.3021	0.2418	Training = 0.0346 s, Prediction = 0.0414 s	Librosa
SAVEE	n_neighbors=10, weights='uniform'	Robust	0.3125	0.1915	0.3125	0.2155	Training = 0.0337 s, Prediction = 0.0409 s	Librosa
SAVEE	n_neighbors=15, weights='uniform'	Robust	0.3229	0.2121	0.3229	0.2235	Training = 0.0336 s, Prediction = 0.0428 s	Librosa
SAVEE	n_neighbors=5, weights='uniform', algorithm='kd_tree'	Robust	0.3021	0.2336	0.3021	0.2418	Training = 0.0603 s, Prediction = 0.0929 s	Librosa
SAVEE	n_neighbors=5, weights='uniform', algorithm='ball_tree'	Robust	0.3021	0.2336	0.3021	0.2418	Training = 0.0506 s, Prediction = 0.0881 s	Librosa
SAVEE	_neighbors=5, weights='uniform', algorithm='brute'	Robust	0.3021	0.2336	0.3021	0.2418	Training = 0.0348 s, Prediction = 0.0416 s	Librosa
SAVEE	n_neighbors=3, weights='distance'	Robust	0.3438	0.2673	0.3438	0.2830	Training = 0.0345 s, Prediction = 0.0342 s	Librosa
SAVEE	n_neighbors=5,	Robust	0.3021	0.2163	0.3021	0.2214	Training = 0.0375 s,	Librosa

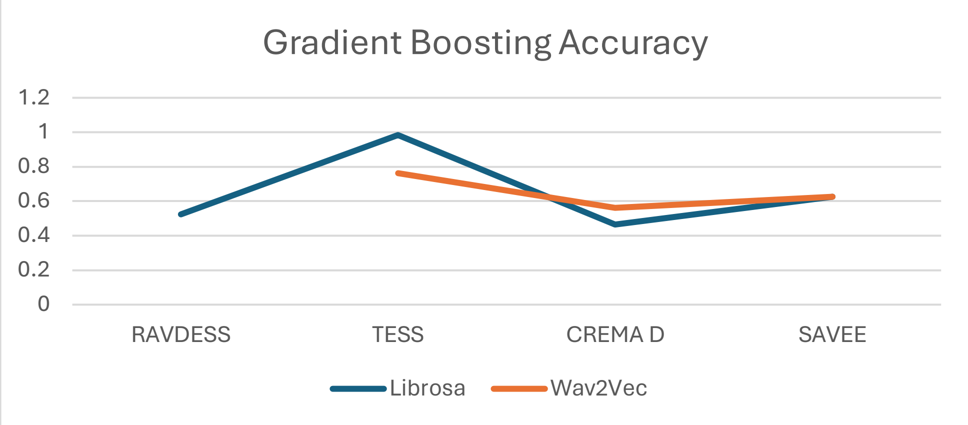
	weights='distance'						Prediction = 0.0343 s	
SAVEE	n_neighbors=10, weights='distance'	Robust	0.3021	0.2103	0.3021	0.2128	Training = 0.0373 s, Prediction = 0.0390 s	Librosa
SAVEE	n_neighbors=15, weights='distance'	Robust	0.3229	0.2140	0.3229	0.2215	Training = 0.0384 s, Prediction = 0.0375 s	Librosa
SAVEE	PCA =0.95	Robust	0.2604	0.2465	0.2604	0.2398	Training = 0.0011 s, Prediction = 0.1371 s	Librosa
SAVEE	PCA = 0.99	Robust	0.2396	0.2137	0.2396	0.2168	Training = 0.0009 s, Prediction = 0.0095 s	Librosa
SAVEE	LDA	Robust	0.1979	0.2624	0.1979	0.2167	Training = 0.0013 s, Prediction = 0.0072 s	Librosa
TESS	Default	Standard	0.9929	0.9929	0.9929	0.9929	Training = 0.0303 s, Prediction = 0.1770 s	Wav2Vec
TESS	Default	Min Max	0.9929	0.9929	0.9929	0.9929	Training = 0.0373 s, Prediction = 0.1790 s	Wav2Vec
TESS	Default	Robust	0.9929	0.9929	0.9929	0.9929	Training = 0.0303 s, Prediction = 0.1770 s	Wav2Vec
TESS	LDA	Robust	0.8161	0.8156	0.8161	0.8113	Training = 0.0028 s, Prediction = 0.0401 s	Wav2Vec
CREMAD	Default	Standard	0.4587	0.4489	0.4587	0.4443	Training = 0.0609 s, Prediction = 0.4499 s	Wav2Vec
CREMAD	Default	Min Max	0.4587	0.4489	0.4587	0.4443	Training = 0.0607 s, Prediction = 0.4155 s	Wav2Vec
CREMAD	Default	Robust	0.4587	0.4489	0.4587	0.4443	Training = 0.0574 s, Prediction = 0.4397 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.4291	0.4266	0.4291	0.4134	Training = 0.0096 s, Prediction = 0.1807 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.4285	0.4248	0.4285	0.4127	Training = 0.0233 s, Prediction = 0.2758 s	Wav2Vec
SAVEE	Default	Standard	0.6250	0.6502	0.6250	0.6226	Training = 0.0200 s, Prediction = 0.0268 s	Wav2Vec
SAVEE	Default	Min Max	0.6250	0.6502	0.6250	0.6226	Training = 0.0196 s, Prediction = 0.0268 s	Wav2Vec
SAVEE	Default	Robust	0.6250	0.6502	0.6250	0.6226	Training = 0.0198 s, Prediction = 0.0261 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.5312	0.5289	0.5312	0.4874	Training = 0.0010 s, Prediction = 0.0101 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5417	0.5419	0.5417	0.5036	Training = 0.0010 s, Prediction = 0.0091 s	Wav2Vec
SAVEE	LDA	Robust	0.6354	0.6454	0.6354,	0.6202,	Training = 0.0011 s, Prediction = 0.0074 s	Wav2Vec



6. Gradient Boosting

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.5104	0.5044	0.5104	0.5089	Training = 372.2324s Prediction =0.0546s	Librosa
RAVDESS	Default	Min Max	0.5174	0.5100	0.5174	0.5078	Training =366.3957 s Prediction =0.0479 s	Librosa
RAVDESS	Default	Robust	0.5243	0.5280	0.5243	0.5174,	Training =365.4907 s Prediction =0.0467 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2743	0.2910	0.2743	0.2709	Training = 6.3346 s, Prediction = 0.0054 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3819	0.3614	0.3819	0.3619	Training = 34.7406 s, Prediction = 0.0062 s	Librosa
RAVDESS	LDA	Robust	0.2535	0.2631	0.2535	0.2497	Training = 2.5556 s, Prediction = 0.0054 s	Librosa
TESS	Default	Standard	0.9839	0.9842	0.9839	0.9840	Training = 704.1230 s, Prediction = 0.0402 s	Librosa
TESS	Default	Min Max	0.9839	0.9842	0.9839	0.9839	Training = 687.6743 s, Prediction = 0.0345 s	Librosa
TESS	Default	Robust	0.9839	0.9842	0.9839	0.9840	Training = 685.1359 s, Prediction = 0.0372 s	Librosa
TESS	PCA = 0.95	Robust	0.6464,	0.6432	0.6464	0.6402	Training = 36.2285 s, Prediction = 0.0098 s	Librosa
CREMAD	Default	Standard	0.4647	0.4579	0.4647	0.4573	Training = 1722.1660 s, Prediction = 0.0633 s	Librosa

CREMAD	Default	Min Max	0.4634	0.4564	0.4634	0.4561	Training = 1709.5834 s, Prediction = 0.0797 s	Librosa
CREMAD	Default	Robust	0.4647	0.4580	0.4647	0.4575	Training = 1720.3580 s, Prediction = 0.0813 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3788	0.3564	0.3788	0.3620	Training = 174.1038 s, Prediction = 0.0215 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3942	0.3774	0.3942	0.3810	Training = 320.5539 s, Prediction = 0.0244 s	Librosa
SAVEE	Default	Standard	0.5938	0.5996	0.5938	0.5768	Training =105.044s Prediction = 0.0240s	Librosa
SAVEE	Default	Min Max	0.5833	0.5922	0.5833	0.5739	Training = 106.0897 s, Prediction = 0.0261 s	Librosa
SAVEE	Default	Robust	0.6250	0.6177	0.6250	0.6048	Training = 103.6968 s, Prediction = 0.0249 s	Librosa
SAVEE	n_estimators= 25	Robust	0.5417	0.5368	0.5417	0.5236,	Training = 23.2525 s, Prediction = 0.0332 s	Librosa
SAVEE	n_estimators= 50	Robust	0.6250	0.6341	0.6250	0.6067	Training = 46.3236 s, Prediction = 0.0314 s	Librosa
SAVEE	n_estimators= 100	Robust	0.6042	0.5992	0.6042	0.5950	Training = 93.3287 s, Prediction = 0.0345 s	Librosa
SAVEE	n_estimators= 300	Robust	0.5833	0.5958	0.5833	0.5782	Training = 465.0516 s, Prediction = 0.0419 s	Librosa
SAVEE	n_estimators= 1200	Robust	0.6250	0.6440	0.6250	0.6169	Training = 280.2834 s, Prediction = 0.0369 s	Librosa
SAVEE	learning_rate= 0.001	Robust	0.3229	0.1518	0.3229	0.1686,	Training = 92.2564 s, Prediction = 0.0362 s	Librosa
SAVEE	learning_rate= 0.01	Robust	0.5417	0.5387	0.5417	0.5090	Training = 91.0064 s, Prediction = 0.0324 s	Librosa
SAVEE	learning_rate= 0.05	Robust	0.6042	0.6083	0.6042,	0.5843	Training = 92.5703 s, Prediction = 0.0337 s	Librosa
SAVEE	learning_rate= 0.2	Robust	0.6146	0.6301	0.6146	0.6000	Training Time = 92.4034 s, Prediction Time = 0.0327 s	Librosa
SAVEE	max_depth=5	Robust	0.5833	0.5923	0.5833	0.5707	Training = 149.3616 s, Prediction = 0.0342 s	Librosa
SAVEE	max_depth=7	Robust	0.4896	0.4685	0.4896	0.4670	Training = 198.6530 s, Prediction = 0.0361 s	Librosa
SAVEE	max_depth=10	Robust	0.3958	0.3884	0.3958	0.3842	Training = 258.6429 s, Prediction = 0.0368 s	Librosa
SAVEE	max_depth=15	Robust	0.4167	0.4059	0.4167	0.3911	Training = 340.4915 s, Prediction = 0.0373 s	Librosa
SAVEE	PCA =0.95	Robust	0.3333	0.3232	0.3333	0.3089	Training = 5.6070 s, Prediction = 0.0026 s	Librosa
SAVEE	PCA = 0.99	Robust	0.2500	0.2034	0.2500	0.2208	Training = 12.8178 s, Prediction = 0.0028 s	Librosa
SAVEE	LDA	Robust	0.1562	0.2092	0.1562	0.1685	Training = 1.1494 s, Prediction = 0.0025 s	Librosa
TESS	LDA	Robust	0.7625	0.7641	0.7625	0.7579	Training = 4.6005 s, Prediction = 0.0096 s	Wav2Vec
CREMAD	Default	Standard	0.5615	0.5589	0.5615	0.5585	Training = 1335.8269 s, Prediction = 0.0539 s	Wav2Vec
CREMAD	Default	Min Max	0.5615	0.5592	0.5615	0.5584	Training = 1350.0676 s, Prediction = 0.0512 s	Wav2Vec
CREMAD	Default	Robust	0.5608	0.5583	0.5608	0.5576	Training = 1330.5700 s, Prediction = 0.0501 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.5292	0.5238	0.5292	0.5223	Training = 459.6395 s, Prediction = 0.0248 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.5124	0.5054	0.5124	0.5051	Training = 941.3102 s, Prediction = 0.0290 s	Wav2Vec
SAVEE	Default	Standard	0.5938	0.5920	0.5938	0.5803	Training = 81.7035 s, Prediction = 0.0203 s	Wav2Vec
SAVEE	Default	Min Max	0.5729	0.5496	0.5729	0.5538	Training = 81.2110 s, Prediction = 0.0203 s	Wav2Vec
SAVEE	Default	Robust	0.5938	0.5633	0.5938	0.5695	Training = 81.7567 s, Prediction = 0.0200 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.5625	0.5947	0.5625	0.5554	Training = 13.6509 s, Prediction = 0.0027 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5208	0.5245	0.5208	0.5073	Training = 22.4514 s, Prediction = 0.0026 s	Wav2Vec
SAVEE	LDA	Robust	0.6250	0.6285	0.6250	0.6213	Training = 1.2057 s, Prediction = 0.0024 s	Wav2Vec



7. XGBoost

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Min Max	0.5312	0.5152	0.5312	0.5124	Training = 59.7684 Prediction = 0.2426 s	Librosa
RAVDESS	Default	Robust	0.5312	0.5152	0.5312	0.5124	Training = 86.2239 s, Prediction = 0.2480 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2847	0.2909	0.2847	0.2823	Training = 1.2084 s, Prediction = 0.0071 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3611	0.3376	0.3611	0.3407	Training = 7.9718 s, Prediction = 0.0077 s	Librosa
TESS	Default	Robust	0.9839	0.9840	0.9839	0.9840	Training = 111.6926bs, Prediction = 0.2248 s	Librosa
CREMAD	Default	Standard	0.4621	0.4526	0.4526	0.4517	Training = 282.0845 s Prediction = 0.3131 s	Librosa
CREMAD	Default	Min Max	0.4621	0.4526	0.4526	0.4517	Training = 283.2889 s, Prediction = 0.2496 s	Librosa
CREMAD	Default	Robust	0.4621	0.4526	0.4526	0.4517	Training = 282.9373 s, Prediction = 0.2423 s	Librosa
SAVEE	LDA	Robust	0.6146	0.6352	0.6146	0.6067	Training = 0.2854 s, Prediction = 0.0356 s	Wav2Vec

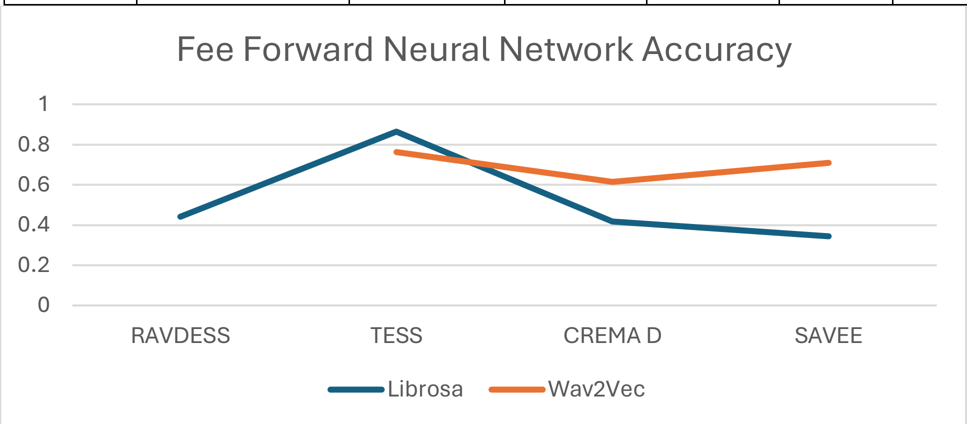
8. LGBoost

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Min Max	0.5451	0.5226	0.5451	0.5209	Training = 72.2895 s Prediction = 0.0368 s	Librosa
RAVDESS	Default	Robust	0.5451	0.5226	0.5451	0.5209	Training = 128.2181 s Prediction = 0.0390 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2847	0.2891	0.2847	0.2791	Training = 1.4213 s, Prediction = 0.0186 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3750	0.3665	0.3750	0.3558	Training = 5.6555 s, Prediction = 0.0211s	Librosa
TESS	Default	Robust	0.9857	0.9858	0.9857	0.9857	Training = 73.0391 s, Prediction = 0.0537 s	Librosa
CREMAD	Default	Standard	0.4587	0.4535	0.4587	0.4534	Training = 205.0562 s, Prediction = 0.1435 s	Librosa
CREMAD	Default	Min Max	0.4587	0.4535	0.4587	0.4534	Training = 189.6366 s, Prediction = 0.1399 s	Librosa
CREMAD	Default	Robust	0.4587	0.4535	0.4587	0.4534	Training = 195.8373 s, Prediction = 0.1484 s	Librosa
SAVEE	LDA	Robust	0.6354,	0.6679	0.6354	0.6227	Training = 0.2095 s, Prediction = 0.0032 s	Wav2Vec

9. Feedforward Neural Network

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.4410	0.4359	0.4410	0.4295	Training = 4.2429s, Prediction = 0.1755s	Librosa
RAVDESS	PCA =0.95	Robust	0.2292	0.1695	0.2292	0.1732	Training = 2.7569 s, Prediction = 0.2062 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3438	0.3432	0.3438	0.3267	Training = 2.4138 s, Prediction = 0.2000 s	Librosa
RAVDESS	LDA	Robust	0.2674	0.2851	0.2674	0.2701	Training = 2.5782 s, Prediction = 0.1347 s	Librosa
TESS	Default	Standard	0.8643	0.8681	0.8643	0.8644	Training = 4.6779 s, Prediction = 0.2368 s	Librosa
TESS	Default	Min Max	0.8143	0.8364	0.8143	0.8051	Training = 3.7338 s, Prediction = 0.2736 s	Librosa
TESS	Default	Robust	0.7250	0.7215	0.7250	0.7216	Training = 4.7384 s, Prediction = 0.2314 s	Librosa
TESS	PCA = 0.95	Robust	0.5000	0.5039	0.5000	0.4945	Training = 3.6634 s, Prediction = 0.1917 s	Librosa

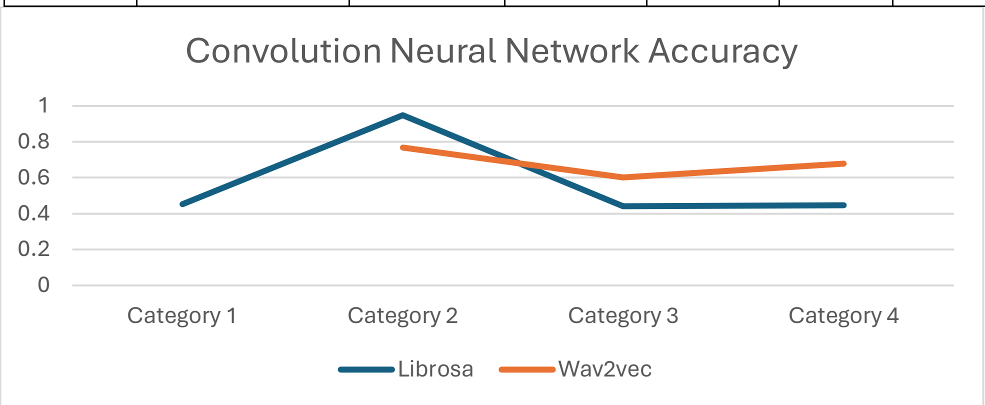
CREMAD	Default	Standard	0.4030	0.4052	0.4030	0.4024	Training = 6.3520 s, Prediction = 0.3831 s	Librosa
CREMAD	Default	Min Max	0.4184	0.3458	0.4184	0.3686	Training = 6.0320 s, Prediction = 0.3817 s	Librosa
CREMAD	Default	Robust	0.3794	0.3785	0.3794	0.3780	Training = 6.3059 s, Prediction = 0.3810 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3566	0.3363	0.3566	0.3396	Training = 5.0692 s, Prediction = 0.2376 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3546	0.3519	0.3546	0.3499,	Training = 6.6987 s, Prediction = 0.5339 s	Librosa
SAVEE	Default	Standard	0.2604	0.2435	0.2604	0.2491	Training = 2.2774 s Prediction = 0.1555s	Librosa
SAVEE	Default	Min Max	0.3438	0.1733	0.3438	0.2188	Training = 1.9432 s, Prediction = 0.1364 s	Librosa
SAVEE	Default	Robust	0.3229	0.2990	0.3229	0.2892	Training = 1.8582 s, Prediction = 0.1260 s	Librosa
SAVEE	PCA =0.95	Robust	0.2812	0.3000	0.2812	0.2464	Training = 3.7547 s, Prediction = 0.2722 s	Librosa
SAVEE	PCA = 0.99	Robust	0.2812	0.2551	0.2812	0.2580	Training = 4.2979 s, Prediction = 0.2659 s	Librosa
SAVEE	LDA	Robust	0.1979	0.2080	0.1979	0.2020	Training = 3.3975 s, Prediction = 0.2067 s	Librosa
TESS	Default	Robust	0.7625	0.7641	0.7625	0.7579	Training = 4.6005 s, Prediction = 0.0096 s	Wav2Vec
CREMAD	Default	Standard	0.6017	0.6049	0.6017	0.6028	Training = 6.5222 s, Prediction = 0.2658 s	Wav2Vec
CREMAD	Default	Min Max	0.5769	0.5879	0.5769	0.5725	Training = 6.9187 s, Prediction = 0.2240 s	Wav2Vec
CREMAD	Default	Robust	0.6145	0.6113	0.6145	0.6120,	Training = 4.7517 s, Prediction = 0.2265 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.6044	0.6025	0.6044	0.6031	Training = 7.4187 s, Prediction = 0.2576 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.6125	0.6100	0.6125	0.6102	Training = 8.0123 s, Prediction = 0.2494 s	Wav2Vec
SAVEE	Default	Standard	0.6771	0.7211	0.6771	0.6734	Training = 1.5035 s, Prediction = 0.1223 s	Wav2Vec
SAVEE	Default	Min Max	0.5938	0.6162	0.5938	0.5576	Training = 1.9479 s, Prediction = 0.1095 s	Wav2Vec
SAVEE	Default	Robust	0.5938	0.6099	0.5938	0.5989	Training = 1.5438 s, Prediction = 0.1023 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.6667	0.6667	0.6667	0.6625	Training = 4.2153 s, Prediction = 0.2650 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5833,	0.5761	0.5833	0.5721	Training = 4.2277 s, Prediction = 0.7586 s	Wav2Vec
SAVEE	LDA	Robust	0.7083	0.7236	0.7083	0.7037	Training = 3.3476 s, Prediction = 0.2098 s	Wav2Vec



10. Convolutional Neural Network

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.4514	0.4616	0.4514	0.4519	Training = 25.5527s Prediction = 0.2850s	Librosa
RAVDESS	PCA =0.95	Robust	0.1632	0.2029	0.1632	0.1467	Training = 3.1467 s, Prediction = 0.2012 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3333,	0.3252	0.3333	0.3246	Training = 5.5793 s, Prediction = 0.2004 s	Librosa
RAVDESS	LDA	Robust	0.2361	0.2526	0.2361	0.2371	Training = 2.2986 s, Prediction = 0.2008 s	Librosa
TESS	Default	Standard	0.9482	0.9490	0.9482	0.9482	Training = 35.1142 s, Prediction = 0.4171 s	Librosa
TESS	Default	Min Max	0.4304	0.3656	0.4304	0.3760	Training = 26.2663 s, Prediction = 0.4153 s	Librosa
TESS	Default	Robust	0.8179	0.8307	0.8179	0.8152	Training = 25.1400 s, Prediction = 0.7162 s	Librosa
TESS	PCA = 0.95	Robust	0.6036,	0.6196	0.6036	0.5987	Training = 6.8818 s, Prediction = 0.2206 s	Librosa
CREMAD	Default	Standard	0.4412	0.4370	0.4412	0.4356	Training = 74.2722 s, Prediction = 0.6992 s	Librosa

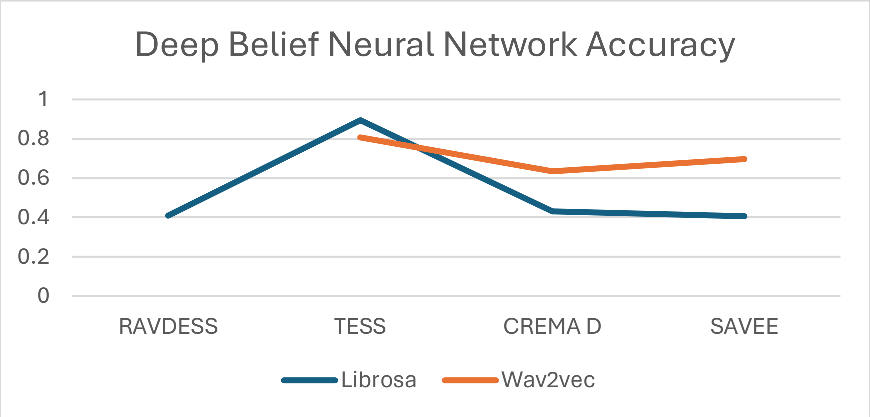
CREMAD	Default	Min Max	0.3929	0.4133	0.3929	0.3618	Training = 67.2560 s, Prediction = 0.6988 s	Librosa
CREMAD	Default	Robust	0.4224	0.4176	0.4224	0.4156	Training = 63.7391 s, Prediction = 0.7017 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3687	0.3608	0.3687	0.3548	Training = 20.5862 s, Prediction = 0.2978 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3909	0.3989	0.3909	0.3735	Training = 42.9928 s, Prediction = 0.3744	Librosa
SAVEE	Default	Standard	0.4479	0.4648	0.4479	0.4386	Training = 6.7852 s, Prediction = 0.2060s	Librosa
SAVEE	Default	Min Max	0.1771	0.2855	0.1771	0.1619	Training = 6.7334 s, Prediction = 0.1733 s	Librosa
SAVEE	Default	Robust	0.3438	0.3377	0.3438	0.3192	Training = 6.7334 s, Prediction = 0.1733 s	Librosa
SAVEE	PCA =0.95	Robust	0.3229	0.2917	0.3229	0.2864	Training = 2.2903 s, Prediction = 0.5193 s	Librosa
SAVEE	PCA = 0.99	Robust	0.3125	0.2649	0.3125	0.2686	Training = 3.1389 s, Prediction = 0.4124 s	Librosa
SAVEE	LDA	Robust	0.1771	0.1950	0.1771	0.1818	Training = 1.9951 s, Prediction = 0.2153 s	Librosa
TESS	LDA	Robust	0.7679	0.7806	0.7679	0.7668	Training = 3.1075 s, Prediction = 0.2060	Wav2Vec
CREMAD	Default	Standard	0.6017	0.6041	0.6017	0.5988	Training = 47.4573 s, Prediction = 0.5266 s	Wav2Vec
CREMAD	Default	Min Max	0.5950	0.6346	0.5950	0.5965	Training = 49.4549 s, Prediction = 0.4858 s	Wav2Vec
CREMAD	Default	Robust	0.5870	0.5866	0.5870	0.5774	Training = 46.1418 s, Prediction = 0.4456 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.5413	0.5760	0.5413	0.5391	Training = 46.5058 s, Prediction = 0.6968 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.5702	0.5715	0.5702	0.5706,	Training = 92.2166 s, Prediction = 0.5587 s	Wav2Vec
SAVEE	Default	Standard	0.6667	0.6618	0.6667	0.6558	Training = 3.9731 s, Prediction = 0.2038 s	Wav2Vec
SAVEE	Default	Min Max	0.6562	0.6529	0.6562	0.6383	Training = 4.3484 s, Prediction = 0.1417 s	Wav2Vec
SAVEE	Default	Robust	0.6771	0.7168	0.6771	0.6746	Training = 4.0426 s, Prediction = 0.1362 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.6354	0.6513	0.6354	0.6320	Training = 3.2108 s, Prediction = 0.4240 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.6667	0.6890,	0.6667	0.6568	Training = 3.2575 s, Prediction = 0.4561 s	Wav2Vec
SAVEE	LDA	Robust	0.6562	0.6758	0.6562	0.6518	Training = 1.9872 s, Prediction = 0.2258 s	Wav2Vec



11. Deep Belief Neural Network

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.4097	0.3929	0.4097	0.3878	Training = 4.7402s, Prediction = 0.1780s	Librosa
RAVDESS	PCA =0.95	Robust	0.2257	0.2447	0.2257	0.1868	Training = 2.4784 s, Prediction = 0.2036 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.2812	0.2626	0.2812	0.2283	Training = 3.1426 s, Prediction = 0.2024 s	Librosa
RAVDESS	LDA	Robust	0.2743,	0.2837	0.2743	0.2764	Training = 2.5665 s, Prediction = 0.1327 s	Librosa
TESS	Default	Standard	0.8946	0.8945	0.8946	0.8938	Training = 4.6750 s, Prediction = 0.2243 s	Librosa
TESS	Default	Min Max	0.6429	0.5995	0.6429	0.5937	Training = 4.0520 s, Prediction = 0.2114 s	Librosa
TESS	Default	Robust	0.7446	0.7450	0.7446	0.7415	Training = 3.9428 s, Prediction = 0.4177 s	Librosa
TESS	PCA = 0.95	Robust	0.3089	0.3608	0.3089	0.2893	Training = 3.6697 s, Prediction = 0.2106 s	Librosa
TESS	Default	Robust	0.8071	0.8095	0.8071	0.8032	Training = 3.8969 s, Prediction = 0.1935 s	Wav2Vec
CREMAD	Default	Standard	0.4305	0.4238	0.4305	0.4257	Training = 6.7865 s, Prediction = 0.3783 s	Librosa
CREMAD	Default	Min Max	0.3324	0.1963	0.3324	0.2278	Training = 6.5043 s, Prediction = 0.3765 s	Librosa

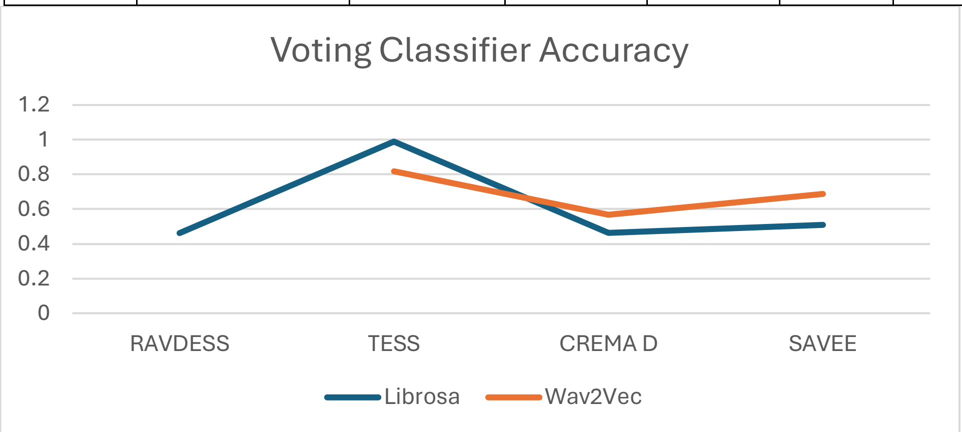
CREMAD	Default	Robust	0.4023	0.3888	0.4023	0.3915	Training = 6.6692 s, Prediction = 0.3798 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3465	0.3480	0.3465	0.3132	Training = 5.3659 s, Prediction = 0.2407 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3506	0.3615	0.3506	0.3297	Training = 6.9446 s, Prediction = 0.2308 s	Librosa
SAVEE	Default	Standard	0.4062	0.3420	0.4062	0.3547	Training = 2.2866 s, Prediction = 0.1274s	Librosa
SAVEE	Default	Min Max	0.3125	0.0977	0.3125	0.1488	Training = 2.3791 s, Prediction = 0.1246 s	Librosa
SAVEE	Default	Robust	0.3750	0.3670	0.3750	0.3310	Training = 2.0614 s, Prediction = 0.1322 s	Librosa
SAVEE	PCA =0.95	Robust	0.2604	0.1793	0.2604	0.1904	Training = 5.4803 s, Prediction = 0.2363 s	Librosa
SAVEE	PCA = 0.99	Robust	0.2917	0.2013	0.2917	0.2151	Training = 5.5858 s, Prediction = 0.2349 s	Librosa
SAVEE	LDA	Robust	0.2188	0.2547	0.2188	0.2297	Training = 5.0899 s, Prediction = 0.2183 s	Librosa
CREMAD	Default	Standard	0.6219	0.6243	0.6219	0.6185	Training = 6.6726 s, Prediction = 0.2287 s	Wav2Vec
CREMAD	Default	Min Max	0.4553	0.4640	0.4553	0.4076	Training = 6.6133 s, Prediction = 0.2217 s	Wav2Vec
CREMAD	Default	Robust	0.6246	0.6249	0.6246	0.6215	Training = 5.0722 s, Prediction = 0.2210 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.6192	0.6181	0.6192	0.6170	Training = 8.2463 s, Prediction = 0.2846 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.6347	0.6380	0.6347,	0.6353	Training = 12.3881 s, Prediction = 0.3718 s	Wav2Vec
SAVEE	Default	Standard	0.6562	0.7225	0.6562	0.6552	Training = 1.6253 s, Prediction = 0.1046 s	Wav2Vec
SAVEE	Default	Min Max	0.4062	0.3019	0.4062	0.3038	Training = 1.7014 s, Prediction = 0.1095 s	Wav2Vec
SAVEE	Default	Robust	0.6562	0.6741	0.6562	0.6468	Training = 1.5748 s, Prediction = 0.1060 s	Wav2Vec
SAVEE	PCA =0.95	Robust	0.6146	0.5951	0.6146,	0.5901	Training = 5.6185 s, Prediction = 0.2307 s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5729	0.6409	0.5729	0.5434	Training = 5.6892 s, Prediction = 0.2413 s	Wav2Vec
SAVEE	LDA	Robust	0.6979,	0.7209	0.6979	0.6975	Training = 4.5138 s, Prediction = 0.2299 s	Wav2Vec
TESS	Default	Min Max	0.6429	0.5995	0.6429	0.5937	Training = 4.0520 s, Prediction = 0.2114 s	Librosa



12. Voting Classifier

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	Default	Standard	0.4583	0.4311	0.4583	0.4170	Training = 385.2786s Prediction = 0.3984s	Librosa
RAVDESS	Default	Min Max	0.4618	0.4399,	0.4618	0.4213	Training = 392.5543 s, Prediction = 0.3909 s	Librosa
RAVDESS	Default	Robust	0.4375	0.4031	0.4375	0.3945	Training = 388.9106 s, Prediction = 0.3887 s	Librosa
RAVDESS	PCA =0.95	Robust	0.2917	0.2894	0.2917	0.2396	Training = 6.6993 s, Prediction = 0.0341 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3646	0.3569	0.3569	=0.3099	Training = 35.7626 s, Prediction = 0.0389 s	Librosa
RAVDESS	LDA	Robust	0.2812	0.2946,	0.2812	0.2777	Training = 2.8794 s, Prediction = 0.0359 s	Librosa
TESS	Default	Standard	0.9661	0.9684	0.9661	0.9662	Training = 706.3745 s, Prediction = 0.4020 s	Librosa
TESS	Default	Min Max	0.9750	0.9760	0.9750	0.9751	Training = 707.8623 s, Prediction = 0.3828 s	Librosa
TESS	Default	Robust	0.9893	0.9893	0.9893	0.9893	Training = 896.8598 s, Prediction = 0.7133 s	Librosa
CREMAD	Default	Standard	0.4627	0.4528	0.4627	0.4491	Training = 2241.1449 s, Prediction = 1.9094 s	Librosa

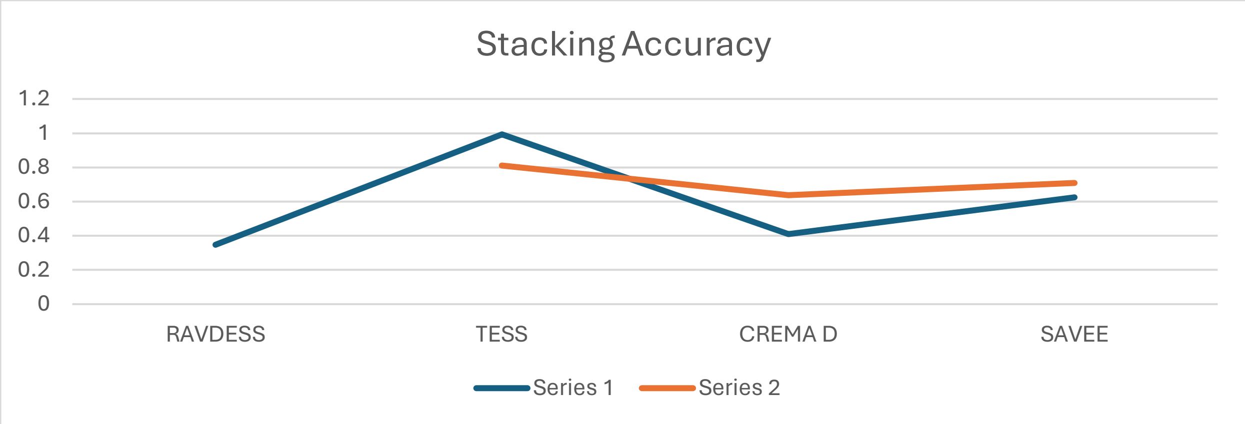
CREMAD	Default	Min Max	0.4567	0.4455	0.4567	0.4426	Training = 2233.5620 s, Prediction = 1.8066 s	Librosa
CREMAD	Default	Robust	0.4627	0.4534	0.4627	0.4487	Training = 2234.1295 s, Prediction = 1.7767 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3694	0.3576	0.3694	0.3169	Training = 327.3507 s, Prediction = 0.2691 s	Librosa
SAVEE	Default	Standard	0.5104	0.5123	0.5104	0.4738	Training =107.217s Prediction = 0.2211s	Librosa
SAVEE	Default	Min Max	0.4896	0.4863	0.4896	0.4441	Training = 108.3979s Prediction = 0.2355s	Librosa
SAVEE	Default	Robust	0.5000	0.5305	0.5000	0.4608	Training = 1106.6418s Prediction = 0.2318s	Librosa
SAVEE	PCA = 0.95	Robust	0.3646	0.3850	0.3646	0.2989	Training = 5.8201s Prediction = 0.0227s	Librosa
SAVEE	PCA = 0.99	Robust	0.3438	0.2320	0.3438	0.2440	Training =13.1225s Prediction = 0.0229s	Librosa
SAVEE	LDA	Robust	0.1979	0.2441	0.1979	0.2099	Training = 1.4340s Prediction = 0.0185s	Librosa
TESS	Default	Robust	0.8179	0.8189	0.8179	0.8142	Training = 5.1407 s, Prediction = 0.0701 s	Wav2Vec
CREMAD	Default	Standard	0.5520	0.5458	0.5520	0.5423	Training = 1359.6522 s, Prediction = 0.6914 s	Wav2Vec
CREMAD	Default	Min Max	0.5500	0.5453	0.5500	0.5409	Training = 1382.6576 s, Prediction = 0.7207 s	Wav2Vec
CREMAD	Default	Robust	0.5561	0.5507	0.5561	0.5472	Training = 1354.6760 s, Prediction = 0.6897 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.5675	0.5678	0.5675	0.5565	Training = 465.7612 s, Prediction = 0.3314 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.5413	0.5441	0.5413	0.5301	Training = 951.1881 s, Prediction = 0.4368 s	Wav2Vec
SAVEE	Default	Standard	0.6250	0.6276	0.6250	0.6000	Training = 83.1712s Prediction = 0.1828s	Wav2Vec
SAVEE	Default	Min Max	0.6146	0.6337	0.6146	0.5893	Training = 82.5836s Prediction = 0.1850s	Wav2Vec
SAVEE	Default	Robust	0.6771	0.6802	0.6771	0.6571	Training =82.6354s Prediction = 0.1913s	Wav2Vec
SAVEE	PCA = 0.995	Robust	0.5938	0.6299	0.5938	0.5683	Training =14.2042s Prediction = 0.0237s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.6458	0.7085	0.6458	0.6289	Training =23.2082s Prediction =0.0242s	Wav2Vec
SAVEE	LDA	Robust	0.6875	0.7049	0.6875	0.6794	Training = 1.4377s Prediction = 0.0191s	Wav2Vec



13. Stacking

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	PCA =0.95	Robust	0.2986	0.2774	0.2986	0.2702	Training = 33.9547 s, Prediction = 0.0308 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.3472	0.3053	0.3472	0.3112	Training = 179.6770 s, Prediction = 0.0392 s	Librosa
RAVDESS	LDA	Robust	0.2882,	0.3050,	0.2882	0.2897	Training = 15.0409 s Prediction = 0.0325 s	Librosa
TESS	Default	Robust	0.9929	0.9929	0.9929	0.9929	Training = 4550.4837 s, Prediction = 0.7389 s	Librosa
CREMAD	Default	Standard	0.4030	0.3967	0.4030	0.3972	Training = 1693.7533 s, Prediction = 0.1898 s	Librosa
CREMAD	Default	Min Max	0.4030	0.3967	0.4030	0.3972	Training = 1693.7533 s, Prediction = 0.1898 s	Librosa
CREMAD	Default	Robust	0.4010	0.3907	0.4010	0.3936	Training = 1532.7533 s, Prediction = 0.1693 s	Librosa
CREMAD	PCA = 0.99	Robust	0.4110	0.3967	0.4110	0.3936	Training = 1673.7533 s, Prediction = 0.1998 s	Librosa
SAVEE	Default	Standard	0.6250	0.6343	0.6250	0.6160	Training = 525.9915s, Prediction = 0.2183 s	Librosa
SAVEE	Default	Min Max	0.6250	0.6433	0.6250	0.6206	Training = 527.8180s Prediction = 0.2234s	Librosa
SAVEE	Default	Robust	0.6042	0.6269	0.6042	0.5973	Training =524.0122s Prediction = 0.2227s	Librosa

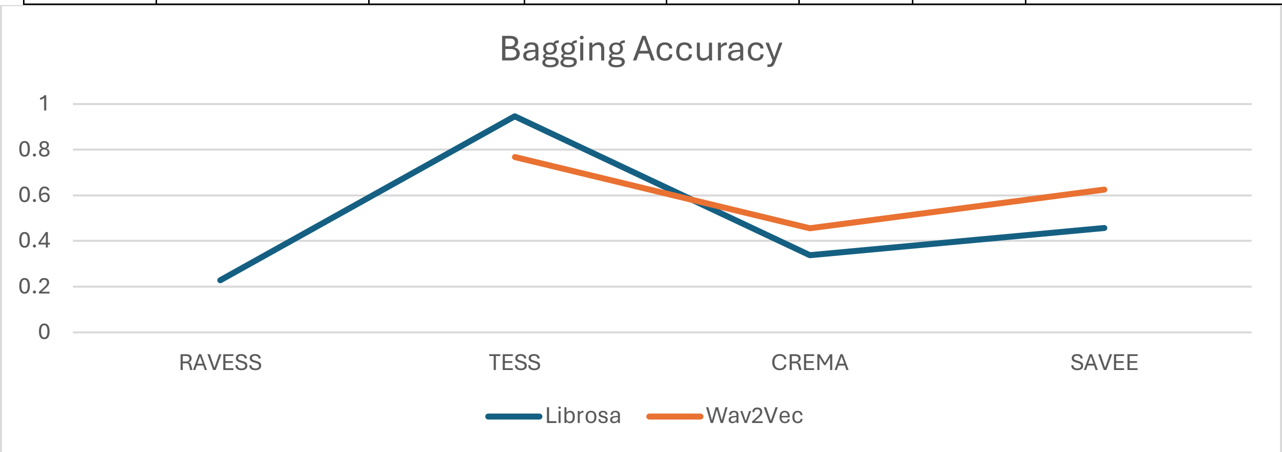
SAVEE	PCA = 0.95	Robust	0.3438	0.3263	0.3438	0.2920	29.8840s, 0.0142s	Librosa
SAVEE	PCA = 0.99	Robust	0.3542	0.2885	0.3542	0.2974	66.0065S, 0.0149s	Librosa
SAVEE	LDA	Robust	0.2396	0.2985	0.2396	0.2559	Training = 8.0092s Prediction = 0.0119s	Librosa
TESS	LDA	Robust	0.8107	0.8118	0.8107	0.8071	Training = 26.5531 s, Prediction =0.0612 s	Wav2Vec
CREMAD	Default	Standard	0.5937	0.5907	0.5937	0.5899	Training = 6819.1376 s, Prediction = 0.7795 s	Wav2Vec
CREMAD	Default	Min Max	0.5944	0.5924	0.5944	0.5914,	Training = 6792.8304 s, Prediction = 0.7919 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.6373	0.6354,	0.6373	0.6354	Training = 2404.2927 s, Prediction = 0.2590 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.6105,	0.6087	0.6105,	0.6089	Training = 4978.1185 s, Prediction = 0.3736 s	Wav2Vec
SAVEE	Default	Standard	0.6250	0.6013	00.6250	0.5959	Training = 407.1540s Prediction = 0.1728s	Wav2Vec
SAVEE	Default	Min Max	0.6354	0.5947	0.6354	0.6016	Training = 408.4183s Prediction = 0.1750s	Wav2Vec
SAVEE	Default	Robust	0.6562	0.6428	0.6562	0.6398	Training = 406.3441s Prediction = 0.1751s	Wav2Vec
SAVEE	PCA = 0.95	Robust	0.6875	0.7051	0.6875	0.6714	Training = 71.3641 Prediction = 0.0140s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.7083	0.7257	0.7083	0.7088	Training = 116.4605s Prediction = 0.0154s	Wav2Vec
SAVEE	LDA	Robust	0.6562	0.6716	0.6562	0.6442	Training = 8.1432s Prediction = 0.0120s	Wav2Vec



14. Bagging

Dataset	Parameter	Scaler	Accuracy	Precision	Recall	F1 score	Time	Extraction
RAVDESS	PCA =0.95	Robust	0.2500	0.2375	0.2500	0.2396	Training = 0.2168 s, Prediction = 0.0025 s	Librosa
RAVDESS	PCA = 0.99	Robust	0.2882	0.2717	0.2882	0.2756	Training = 1.2476 s, Prediction = 0.0039 s	Librosa
RAVDESS	LDA	Robust	0.2188,	0.2211,	0.2188	0.2076	Training = 0.0850 s, Prediction = 0.0022 s	Librosa
TESS	Default	Standard	0.9411	0.9421	0.9411	0.9413	Training Time = 17.6097 s, Prediction Time = 0.0571 s	Librosa
TESS	Default	Min Max	0.9464	0.9467	0.9464	0.9462	Training = 18.7146 s, Prediction = 0.0517 s	Librosa
TESS	Default	Robust	0.9393	0.9405	0.9393	0.9396	Training = 17.9860 s, Prediction = 0.0529 s	Librosa
CREMAD	Default	Standard	0.3173	0.3286,	0.3173	0.3315	Training = 16.1003 s, Prediction = 0.0178 s	Librosa
CREMAD	Default	Min Max	0.3182	0.3286,	0.3182	0.3315	Training = 13.267 s, Prediction = 0.0228 s	Librosa
CREMAD	Default	Robust	0.3185	0.3286,	0.3185	0.3315	Training = 16.9144 s, Prediction = 0.0234 s	Librosa
CREMAD	PCA = 0.95	Robust	0.3286	0.3182,	0.3286	0.3310	Training = 16.9144 s, Prediction = 0.0238 s	Librosa
CREMAD	PCA = 0.99	Robust	0.3378	0.3286,	0.3378	0.3315	Training = 16.9144 s, Prediction = 0.0238 s	Librosa
SAVEE	Default	Standard	0.4583	0.4383	0.4583	0.4378	Training = 2.3041 s, Prediction = 0.0271s	Librosa
SAVEE	Default	Min Max	0.3646	0.4130	0.3646	0.3710	Training =2.3994s Prediction = 0.0257s	Librosa
SAVEE	Default	Robust	0.4271	0.4425	0.271	0.4138	Training =2.3653s Prediction = 0.0276s	Librosa
SAVEE	PCA= 0.95	Robust	0.2604	0.2806	0.2604	02664	Training = 0.1568s, Prediction = 0.0021a	Librosa
SAVEE	PCA =0.99	Robust	0.3021	0.3309	0.3021	0.3140	Training = 0.3380S, Prediction = 0.0026S	Librosa
SAVEE	LDA	Robust	0.2396	0.2558	0.2396	0.2409	Training = 0.0315s, Prediction = 0.0015	Librosa
TESS	LDA	Robust	0.7679	0.7749	0.7679	0.7369	Training = 0.0957 s, Prediction = 0.0026 s	Wav2Vec
CREMAD	Default	Standard	0.4527	0.4438,	0.4527	0.4463	Training = 62.4587 s,	Wav2Vec

								Prediction = 0.0766 s	
CREMAD	Default	Min Max	0.4560	0.4496	0.4560	0.4504	Training = 59.4287 s, Prediction = 0.0869 s	Wav2Vec
CREMAD	PCA = 0.95	Robust	0.3747	0.3671,	0.3747	0.3689	Training = 19.1054 s, Prediction = 0.0334 s	Wav2Vec
CREMAD	PCA = 0.99	Robust	0.3936	0.3888	0.3936	0.3894	Training = 38.3173 s, Prediction = 0.0517 s	Wav2Vec
SAVEE	Default	Standard	0.4583	0.4189	0.4583	0.4233	Training = 1.6756s Prediction = 0.0216s	Wav2Vec
SAVEE	Default	Min Max	0.5833	0.6005	0.5833	0.5782	Training = 1.737s Prediction = 0.0211a	Wav2Vec
SAVEE	Default	Robust	0.4688	0.4772	0.4688	0.4581	Training = 1.6561s Prediction = 0.0222s	Wav2Vec
SAVEE	PCA = 0.95	Robust	0.5000	0.4835	0.5000	0.4690	Training = 0.3086s Prediction = 0.0022s	Wav2Vec
SAVEE	PCA = 0.99	Robust	0.5104	0.4861	0.5104	0.4763	Training = 0.4634s Prediction = 0.0032s	Wav2Vec
SAVEE	LDA	Robust	0.6250	0.6308	0.6250	0.6189	Training = 0.0330s Prediction = 0.0016s	Wav2Vec



Findings

My analysis revealed that the performance of the same model with identical parameters varied significantly across different datasets. This variation underscores the impact of dataset characteristics on model efficacy. Furthermore, the two feature extraction techniques I employed—wav2vec and Librosa—produced different accuracies for each dataset and model combination. Notably, feature extraction using wav2vec consistently outperformed Librosa across all datasets. The highest accuracy results were achieved on the TESS dataset, indicating its suitability for emotion recognition tasks. Overall, these findings highlight the importance of choosing the right feature extraction method and dataset for optimizing model performance in audio emotion recognition.

Future Work

- Explore advanced deep learning architectures such as RNNs for improved model accuracy.
- Investigate multimodal approaches by integrating audio with visual cues for a more comprehensive understanding of emotions.
- Apply transfer learning techniques to fine-tune pre-trained models on large-scale datasets for emotion recognition tasks.
- Expand the diversity of training data by incorporating more varied and real-world audio samples to enhance model robustness.
- Develop real-time emotion recognition systems for practical applications such as virtual assistants and healthcare.

Conclusion

In conclusion, this project has demonstrated the feasibility and effectiveness of using machine learning techniques for audio emotion recognition. Through the analysis of various datasets and feature extraction methods, we have observed the importance of dataset characteristics and feature representation in model performance. While wav2vec emerged as a promising feature extraction technique, achieving superior accuracy compared to Librosa, the variability in model performance across datasets underscores the need for further exploration and optimization. Moving forward, future research should focus on refining model architectures, incorporating multimodal data, and deploying real-time applications to advance the field of audio emotion recognition and its practical implementations.

Code Snippets

Feature Extraction through Librosa

```python
import librosa
import numpy as np
import csv
import time
import pywt  # Adding import statement for pywt

def extract_features(audio_file):
    start_time = time.time()
    # Load audio file
    y, sr = librosa.load(audio_file)

    # Time-domain features
    mean = np.mean(y)
    variance = np.var(y)
    skewness = np.mean(librosa.feature.spectral_bandwidth(y=y, sr=sr))
    kurtosis = np.mean(librosa.feature.spectral_flatness(y=y))
    zero_crossing_rate = np.mean(librosa.feature.zero_crossing_rate(y))
    rms_energy = np.mean(librosa.feature.rms(y=y))

    # Intensity (average energy)
    intensity = np.mean(np.abs(y))

    # Pitch (fundamental frequency)
    pitch = librosa.piptrack(y=y, sr=sr)[0]
    pitch_freqs = [max(pitch[:, i]) for i in range(pitch.shape[1]) if max(pitch[:, i]) > 0]
    if pitch_freqs:
        pitch = np.mean(pitch_freqs)
    else:
        pitch = 0

    # Frequency-domain features
    spectral_centroid = librosa.feature.spectral_centroid(y=y, sr=sr)[0].mean()
    spectral_bandwidth = librosa.feature.spectral_bandwidth(y=y, sr=sr)[0].mean()
    spectral_contrast = librosa.feature.spectral_contrast(y=y, sr=sr).mean(axis=1)
    spectral_rolloff = librosa.feature.spectral_rolloff(y=y, sr=sr)[0].mean()

    # Mel-frequency cepstral coefficients (MFCCs)
    mfccs = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=13)

    # Chroma features
    chroma = librosa.feature.chroma_stft(y=y, sr=sr)
    chroma_features = {f"chroma{i}": np.mean(chroma[i-1]) for i in range(1, chroma.shape[0] + 1)}

    # Spectral features
    spectral_flux = librosa.onset.onset_strength(y=y, sr=sr)
    spectral_flatness = librosa.feature.spectral_flatness(y=y)
    spectral_density = np.sum(librosa.stft(y), axis=1)  # Summing up magnitudes across frequency bins

    # Extract each spectral flux coefficient separately
    spectralflux_features = {}
    for i, flux_value in enumerate(spectral_flux):
        spectralflux_features[f"spectralflux{i+1}"] = np.mean(flux_value)

    # Extract each spectral density coefficient separately
    spectraldes_features = {}
    for i, density_value in enumerate(spectral_density):
        spectraldes_features[f"spectraldes{i+1}"] = np.mean(density_value)

    # Extract each spectral flatness coefficient separately
    spectralflatness_features = {}
    for i, flatness_value in enumerate(spectral_flatness):
        for j, value in enumerate(flatness_value):
            spectralflatness_features[f"spectralflatness{j+1}_{i+1}"] = value

    # Wavelet transform (example code, may need customization)
    coeffs, _ = pywt.dwt(y, 'db1')  # Example of using wavelet transform (you might need to adjust parameters)

    # Extract each wavelet coefficient separately
    waveletcoeffs_features = {}
    for i, coeff_value in enumerate(coeffs):
        waveletcoeffs_features[f"waveletcoeff{i+1}"] = np.mean(coeff_value)

    end_time = time.time()
    time_taken = end_time - start_time
    # Extract each MFCC coefficient separately
    mfcc_features = {}
    for i in range(1, 14):
        mfcc_features[f"mfcc{i}"] = np.mean(mfccs[i-1])

    return {
        "mean": mean,
        "variance": variance,
        "skewness": skewness,
        "kurtosis": kurtosis,
        "zero_crossing_rate": zero_crossing_rate,
        "rms_energy": rms_energy,
        "intensity": intensity,
        "pitch": pitch,
        "spectral_centroid": spectral_centroid,
        "spectral_bandwidth": spectral_bandwidth,
        **{f"spectralcontrast{i+1}": spectral_contrast[i] for i in range(len(spectral_contrast))},  # Unpack Spectral contrast features
        "spectral_rolloff": spectral_rolloff,
        **mfcc_features,  # Unpack MFCC coefficients
        **chroma_features,  # Unpack Chroma features
        **spectralflux_features,  # Unpack Spectral flux features
        **spectraldes_features,  # Unpack Spectral density features
        **spectralflatness_features,  # Unpack Spectral flatness features
        "time_taken": time_taken
    }
# Read audio file paths and their corresponding emotion labels from CSV
def read_audio_paths(csv_file):
    audio_paths = []
    with open(csv_file, mode='r') as file:
        reader = csv.DictReader(file)
        for row in reader:
            audio_paths.append((row['Path'], row['Emotions']))  # Assuming 'Path' and 'Emotion' are the column names
    return audio_paths

# Example usage:
csv_file = "/kaggle/working/classified_crema.csv"
audio_paths = read_audio_paths(csv_file)

# List to store features of all audio files
all_features = []

# Start time for the entire dataset
total_start_time = time.time()

# Extract features for each audio file
for audio_path, emotion_label in audio_paths:
    features = extract_features(audio_path)
    features["Path"] = audio_path
    features["Emotions"] = emotion_label
    all_features.append(features)

# End time for the entire dataset
total_end_time = time.time()

# Total time taken for the entire dataset
total_time_taken = total_end_time - total_start_time

# Collect all feature names
all_feature_names = set()
for features in all_features:
    all_feature_names.update(features.keys())

# Create fieldnames list
fieldnames = list(all_feature_names)

# Write features to CSV file
output_csv_file = "cremafeatures.csv"
with open(output_csv_file, mode='w', newline='') as file:
    writer = csv.DictWriter(file, fieldnames=fieldnames)
    writer.writeheader()
    writer.writerows(all_features)

print("Features extracted and stored in:", output_csv_file)
print("Total time taken for the entire dataset:", total_time_taken, "seconds")
```

Feature Extraction through Wav2Vec

```python
import os
import torch
import torch.nn as nn
import pandas as pd
from transformers import Wav2Vec2Processor, Wav2Vec2Model
import torchaudio
from torchaudio.transforms import Resample
import numpy as np
import time

class EmotionModel(nn.Module):
    def __init__(self, config):
        super().__init__()
        self.config = config
        self.wav2vec2 = Wav2Vec2Model(config)

    def forward(self, input_values):
        outputs = self.wav2vec2(input_values)
        hidden_states = outputs.last_hidden_state.squeeze().mean(axis=0)
        return hidden_states

def process_func(audio_dir):
    device = 'cpu'
    model_name = 'audeering/wav2vec2-large-robust-12-ft-emotion-msp-dim'
    processor = Wav2Vec2Processor.from_pretrained(model_name)
    model = Wav2Vec2Model.from_pretrained(model_name)
    model.to(device)

    features_list = []
    paths_list = []

    resampler = Resample(orig_freq=24414, new_freq=16000)

    start_time = time.time()

    for root, dirs, files in os.walk(audio_dir):
        for file in files:
            audio_path = os.path.join(root, file)
            array, fs = torchaudio.load(audio_path)
            array_resampled = resampler(array)
            input_values = processor(array_resampled.squeeze(), sampling_rate=16000, return_tensors="pt")
            input_values = input_values.input_values.to(device)
            with torch.no_grad():
                hidden_states = model(input_values)
            features_list.append(hidden_states.last_hidden_state.squeeze().mean(axis=0).cpu().numpy().tolist())
            paths_list.append(audio_path)

    end_time = time.time()
    elapsed_time = end_time - start_time
    print("Time taken to extract features:", elapsed_time, "seconds")

    return features_list, paths_list

# Example usage:
audio_dir = '/kaggle/input/cremad/AudioWAV/'
features_list, paths_list = process_func(audio_dir)

# Convert lists to DataFrame
data = {'Path': paths_list}
for i in range(len(features_list[0])):
    data[f'Feature_{i}'] = [feature[i] for feature in features_list]

df = pd.DataFrame(data)

# Save DataFrame to CSV
csv_file = '/kaggle/working/featuresforCremadWav2vec.csv'
df.to_csv(csv_file, index=False)
print("Features saved to:", csv_file)
```

Platform

For all the activities described above, the evaluation of datasets was conducted using Kaggle Notebook. Kaggle Notebook provided a convenient and efficient environment for data analysis, model development, and evaluation. The integration of code, visualizations, and annotations within the notebook facilitated a clear and structured approach to my analysis, enhancing my understanding and decision-making processes.