Major Project Phase - II
Review – 2

Detecting Logging of Forest Trees using Sound Event Detection

Team Members (Team - 17)

19K41A0594 - Boda Raju

19K41A0510 - Jinukala Vamshi

19K41A0517 - Mohammed Raamizuddin

19K41A05E9 - Bhonagiri Shreya

19K41A05F7 - Jupally Yochitha

Project Guide

Sallauddin Mohmmad

Assistant Professor, Dept of CSE

Contents

01. Introduction

02. Methodology

03. Results and Analysis

04. Conclusion

Introduction

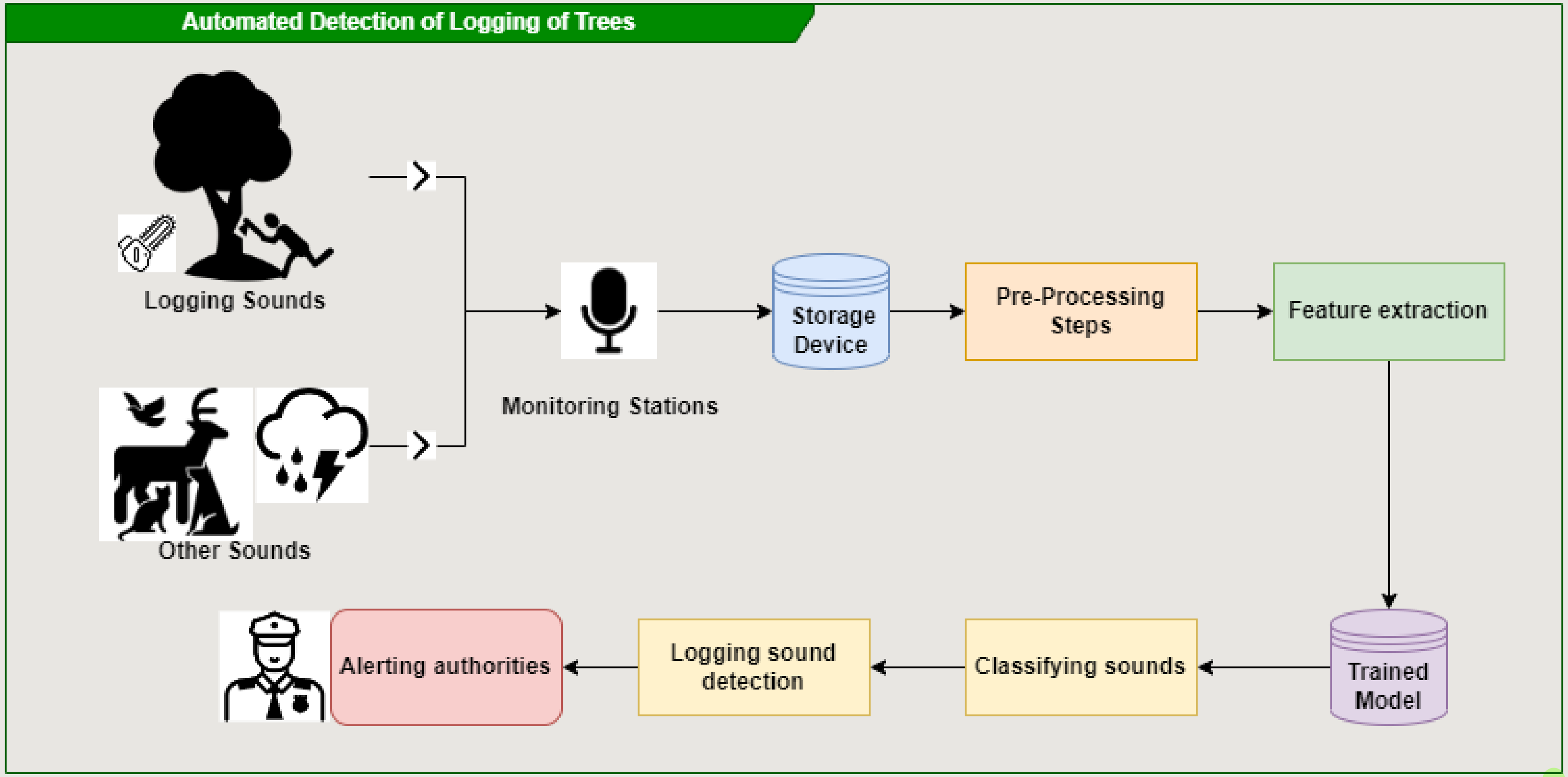
- Tree logging is an illegal activity.
- India is ranked third for illegally importing logged timber in the world.
- The issue must be dealt very seriously as it exhausts the forest assets and may increase deforestation.
- Monitoring the forest assets visually requires a lot of equipment.
- An acoustic signature can provide valuable information about the activities of any intruder inside the forest.

Introduction

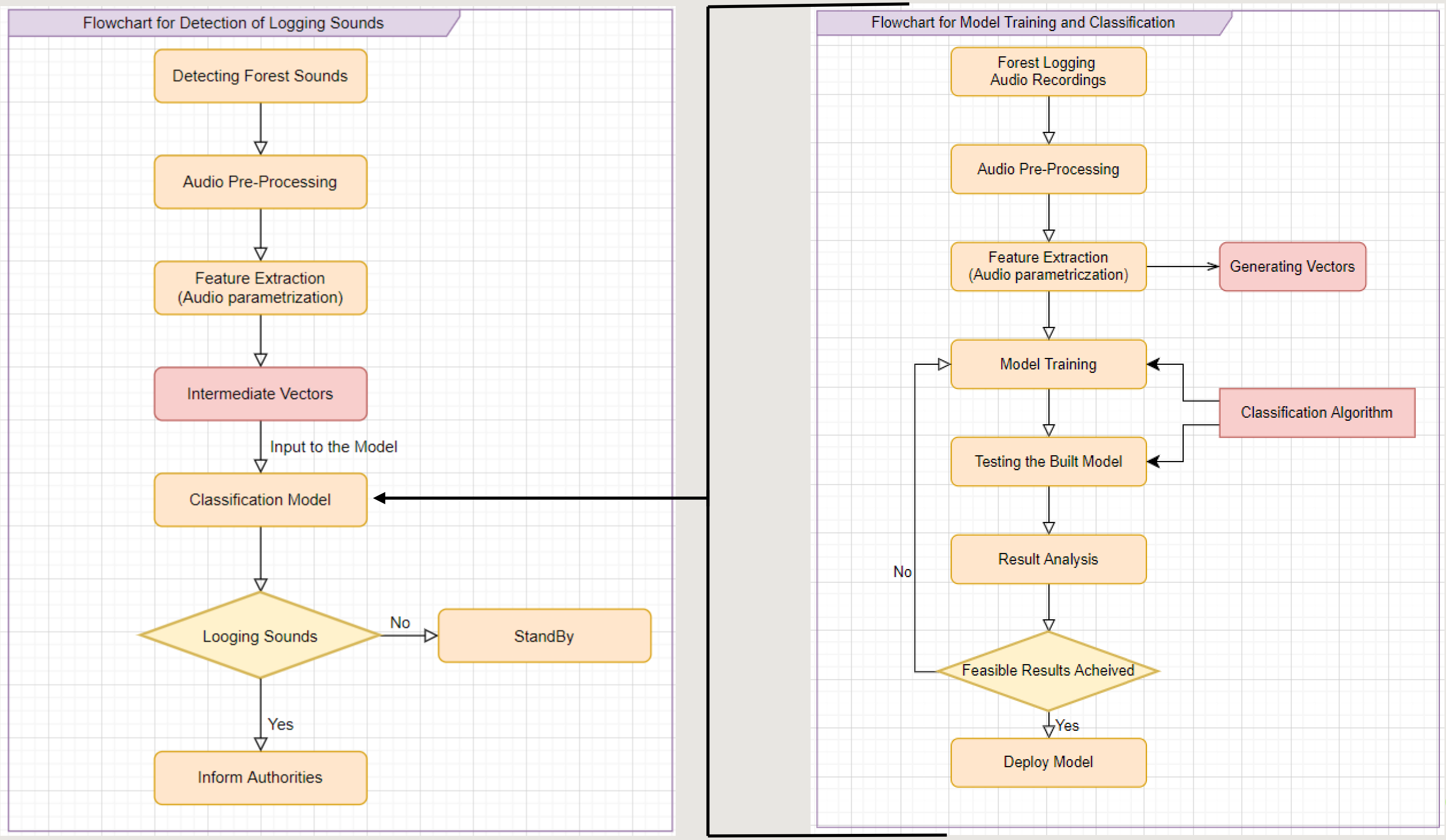
- Domain:
  - Sound Event Detection.
  - Prevention of illegal logging of forest trees.
- Implementation Domain:
  - Sound Feature Extraction Techniques.
  - Deep learning techniques for Classification.

Methodology

- The solution to the mentioned problem can be achieved by using the below workflow :



Flowchart for Model Training

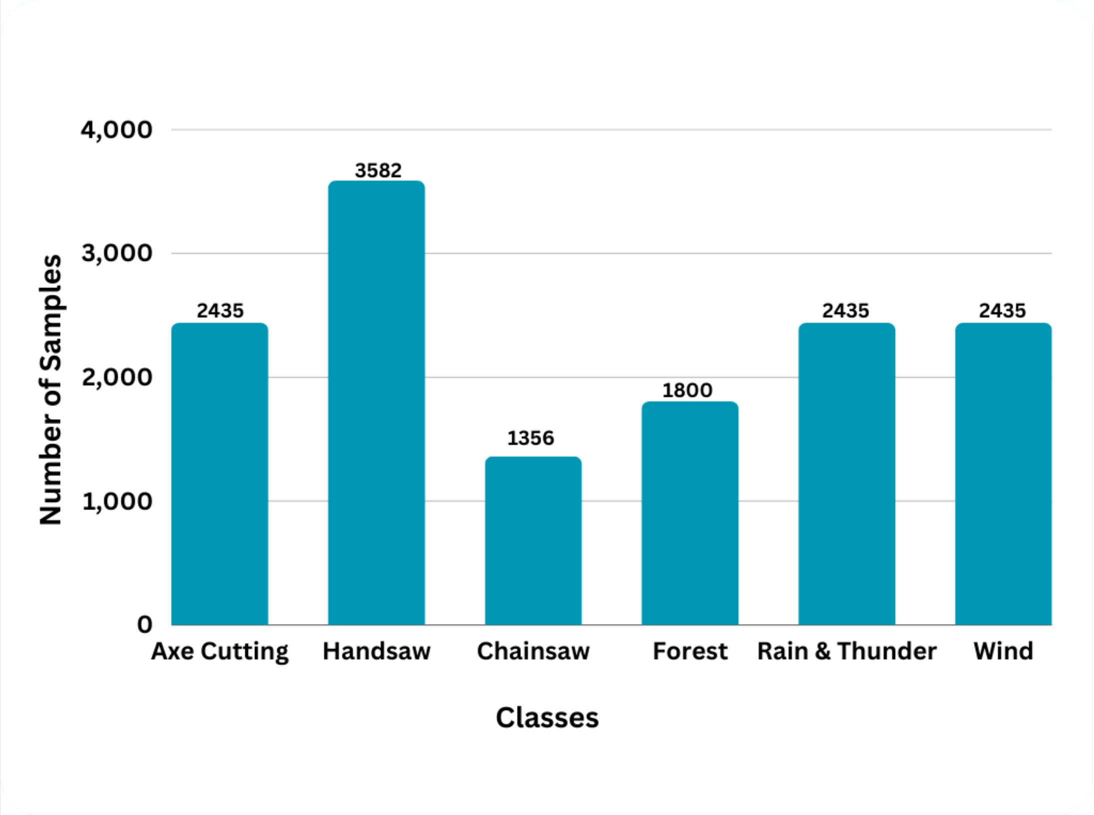


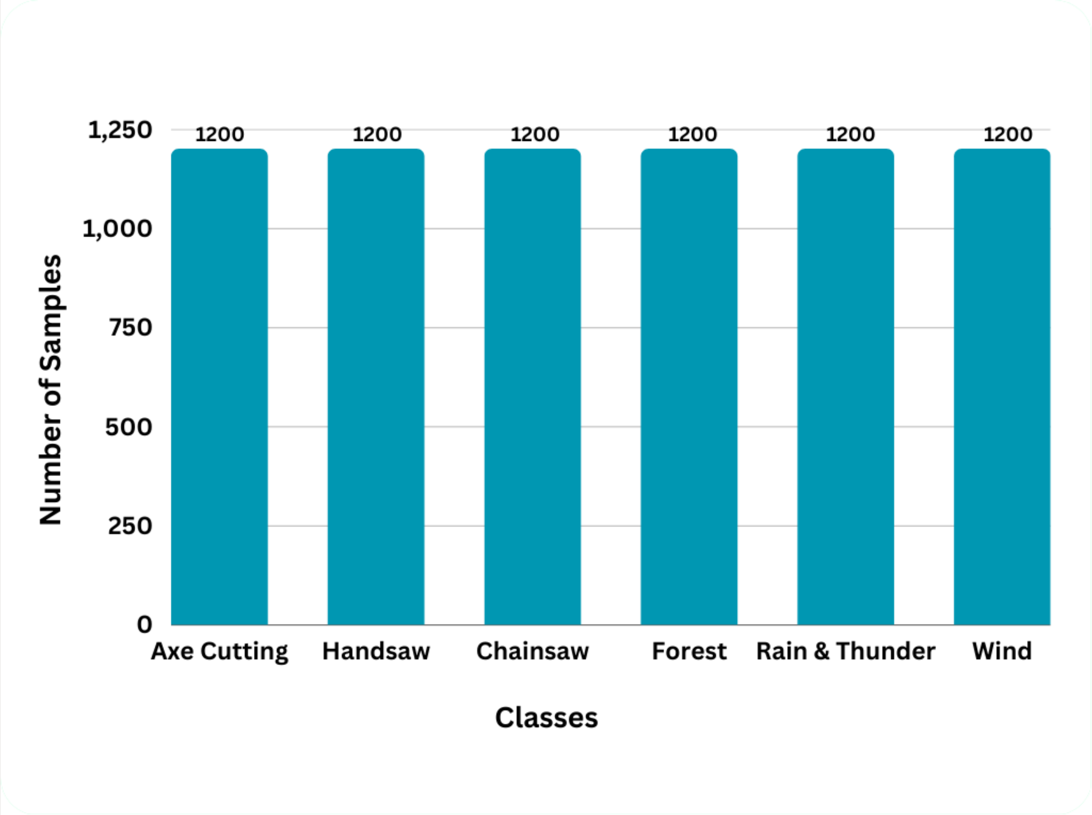
Dataset

- The dataset contains equal positive and negative classes, and the orthogonality of the data is high.
- There are six annotated classes in the dataset (Chainsaw, Handsaw, Axe cutting, Wind, Forest, Rain and Thunder Sounds).
- Each class has 1200 samples. Therefore, 7200 audio files.
- The audio files are 10 seconds each.

Dataset (Contd.)

- The number of samples gathered and the number of samples utilized for training, testing and validation are shown below



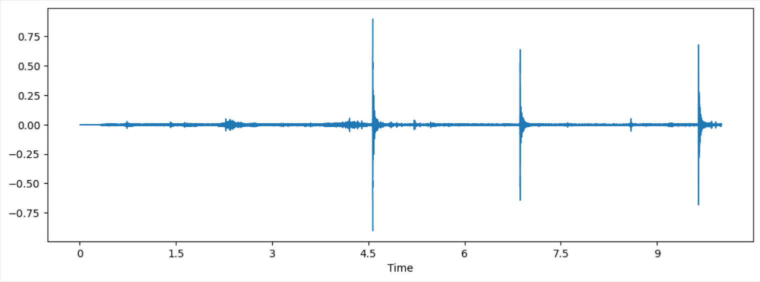


Audio Pre-Processing

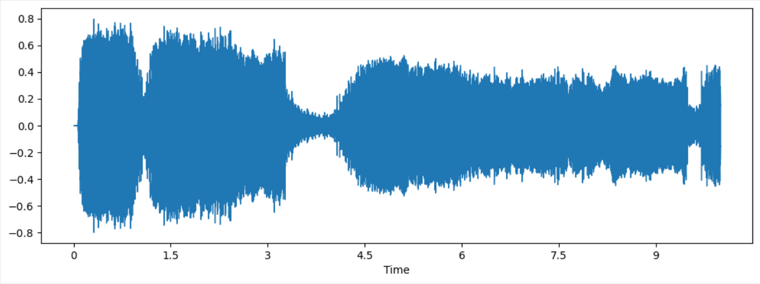
- The audio samples collected were of various durations.
- All the samples were divided into several samples of ten seconds each in order to keep the samples of same duration.
- The class with least samples was chainsaw sounds with 1356 samples and the class with highest samples was handsaw sounds with 3582 samples.
- 1200 samples were selected at random in order to avoid imbalance of data.
- Audio sampling was performed, and data was loaded into the python.

Audio Waveforms

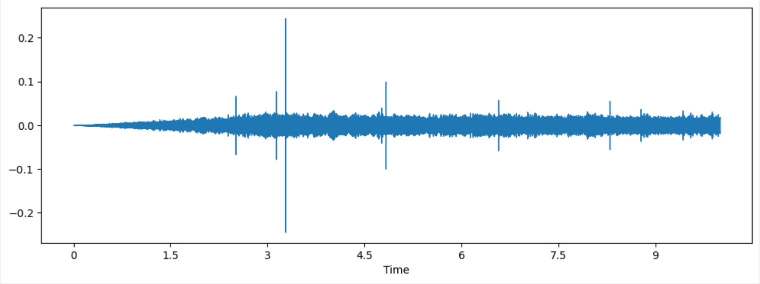
- Below are the audio waveform representations of some audio class



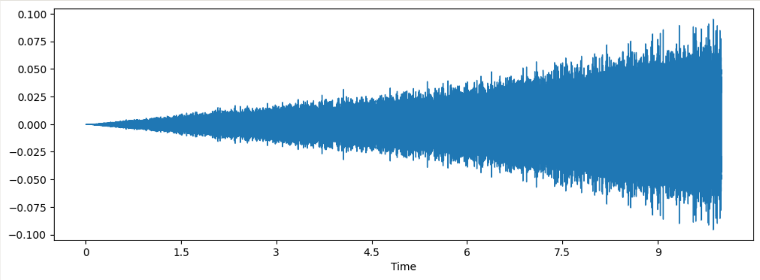
Axe cutting Sounds



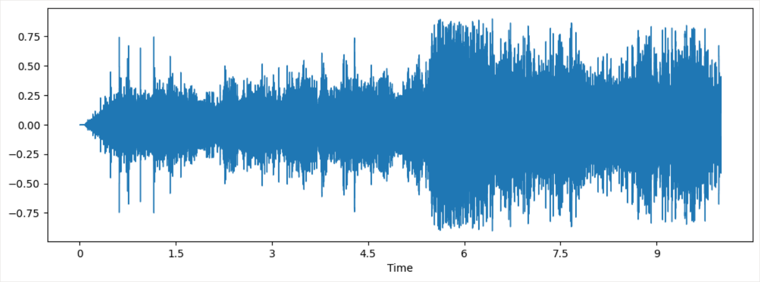
Chainsaw Sounds



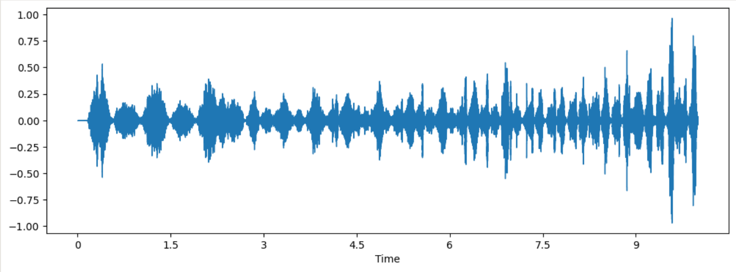
Forest Sounds



Wind Sounds



Rain and Thunder



Handsaw

Feature Extraction Techniques Used

- **MFCC**: The Mel-frequency Cepstrum coefficient is a representation of the short-term power spectrum of a sound.
- **SPECTRAL CONTRAST**: It is the decibel difference between peaks and valleys, that helps for echancing of sounds.
- **MEL-SPECTROGRAM**: It is used for rendering the frequencies above a certain threshold frequency.
- **CHROMA**: The value of an audio which represents the intensity of twelve distinctive pitch classes that are used to analyse music.
- **TONNETZ**: It is used for computing the tonal centroid features of sound.

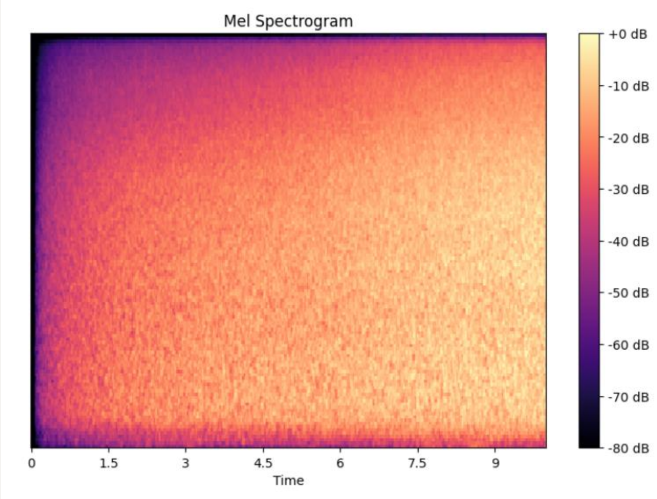
Feature Extraction (Contd.)

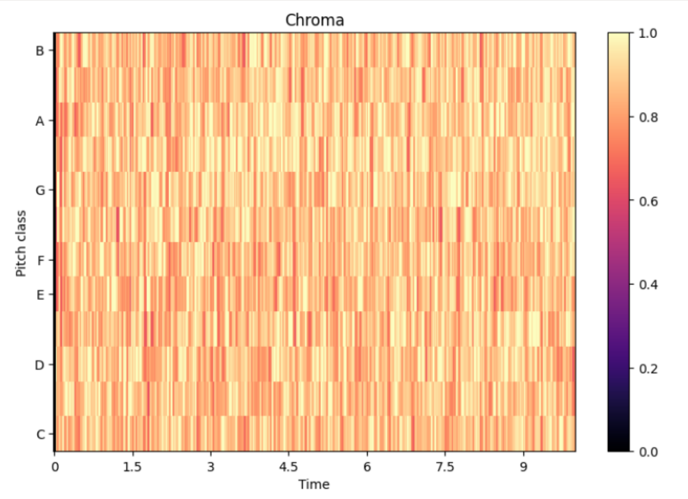
Data/Feature	Dimensions
Sampling Rate	22050
Audio File	(220500,)
MFCC Matrix	(40, 431)
STFT	(1025, 431)
Chroma gram Matrix	(12, 431)
Mel Spectrogram Matrix	(128, 431)
Spectral Contrast Matrix	(7, 431)
Tonal Centroid Features Matrix	(6, 431)

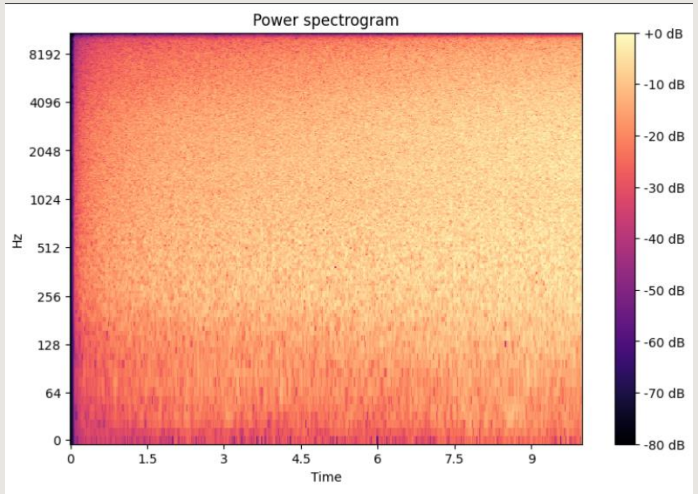
Feature Dimensions

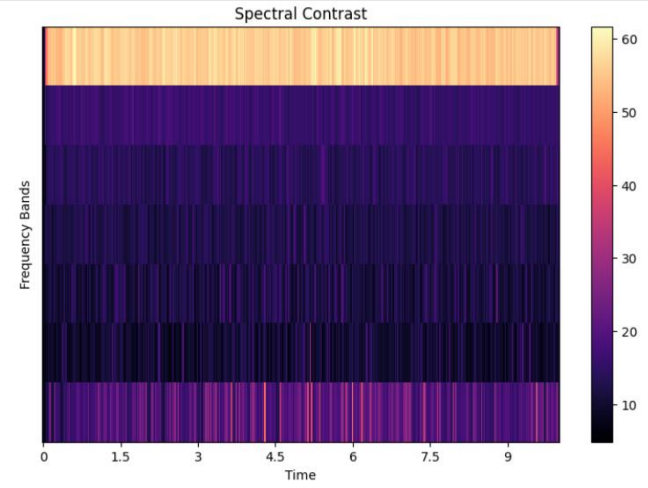
Feature Extraction Image Representation

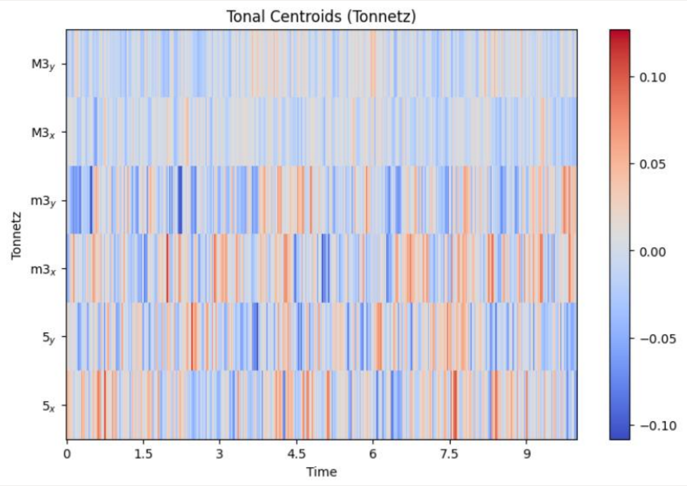
- Below are the representations of various feature extraction techniques:











Models Used

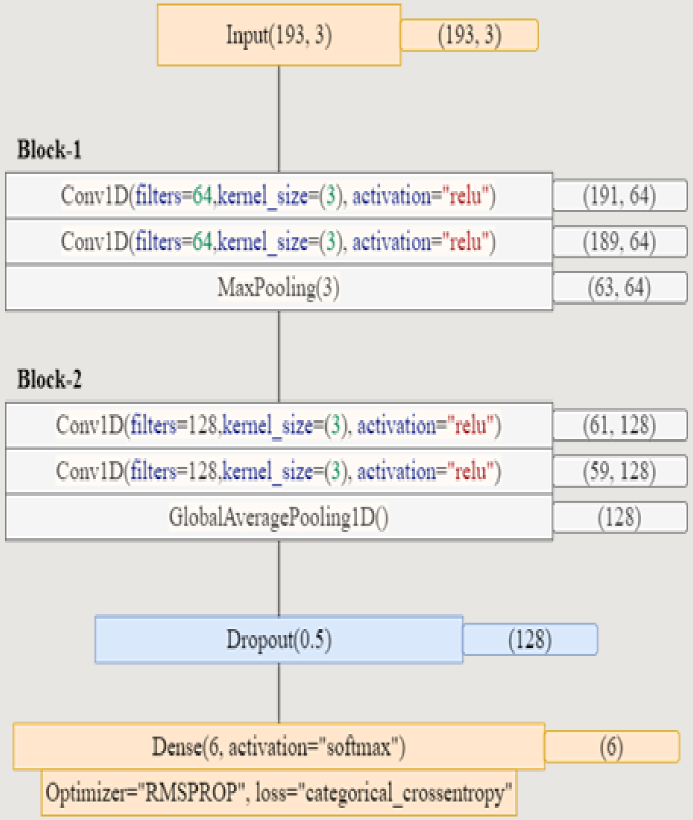
- Various CNN models were trained using the extracted features.
- The models considered for analysis at this stage are:

    1. Convolutional Neural Network (CNN)
    2. Convolutional Recurrent Neural Network (BICRNN)
    3. Bi-directional Convolutional Recurrent Neural Network (BICRNN)

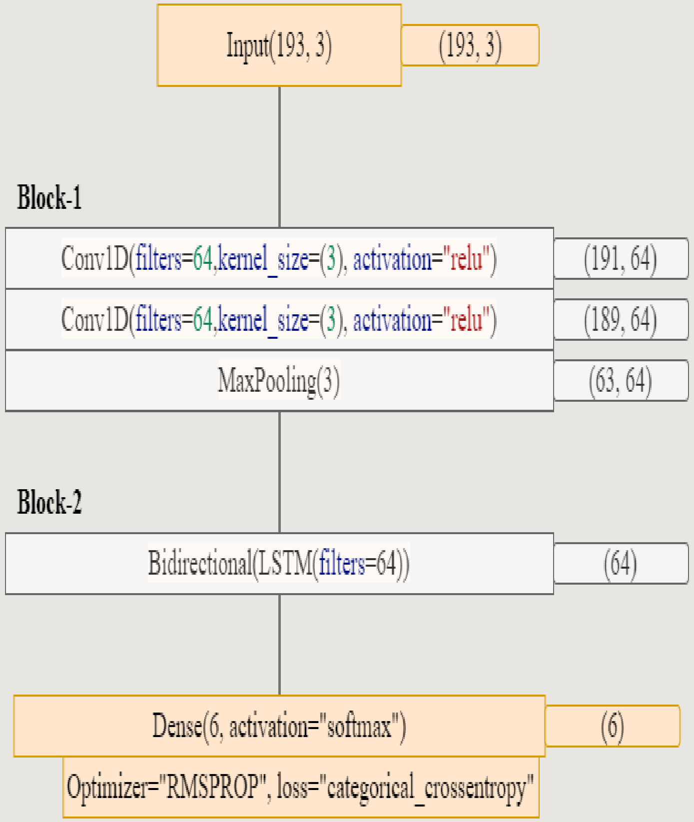
- These models are considered for analyzing the results and improving the efficiency of classification further by altering the architecture or tuning the hyperparameters.

Model Architecture

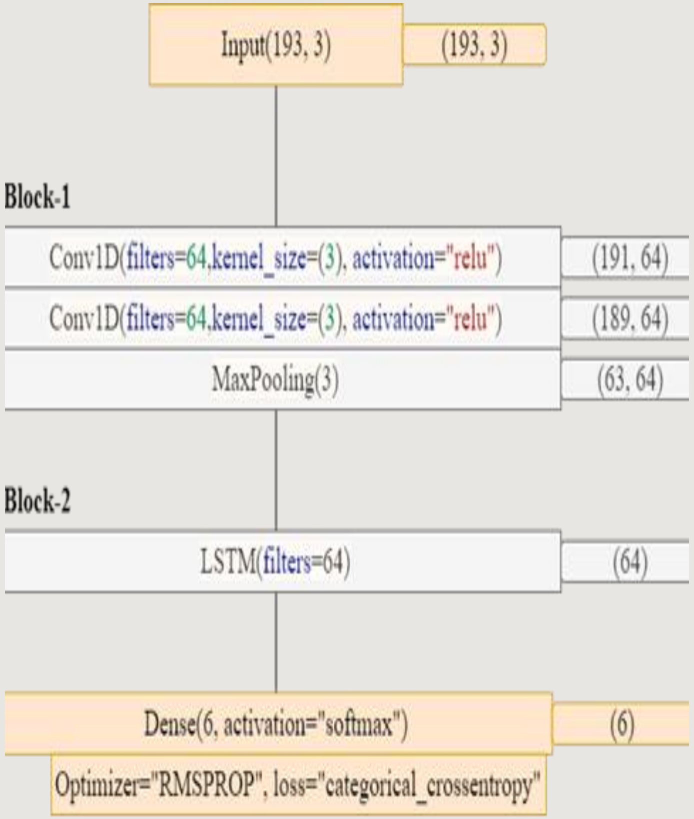
- The architectures of models are represented below:



CNN



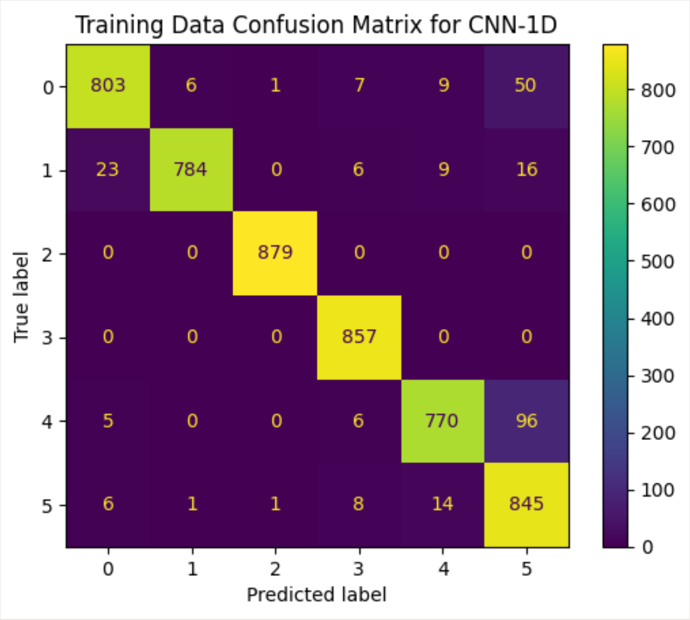
BICRNN

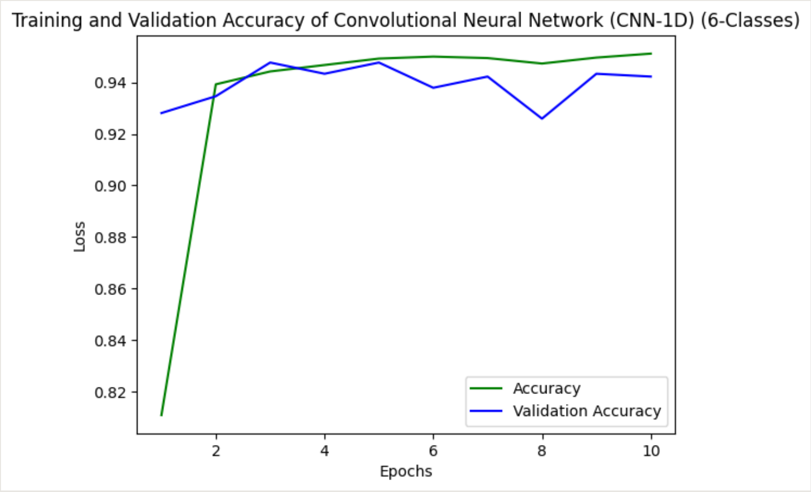


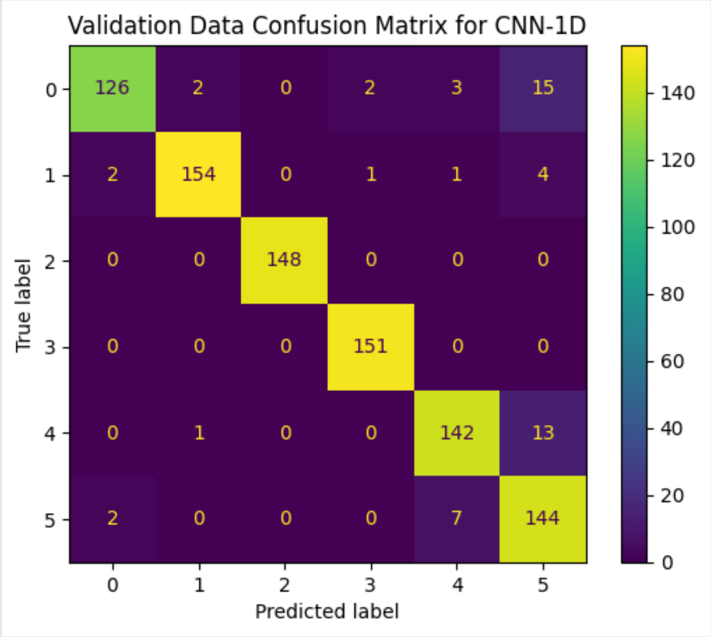
CRNN

Result Analysis

- The Confusion matrices of Convolutional Neural Network (CNN) is given below:

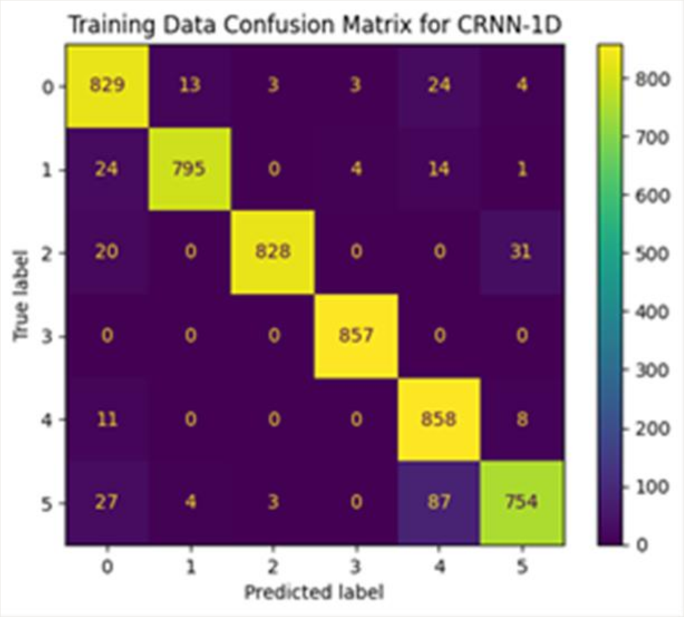


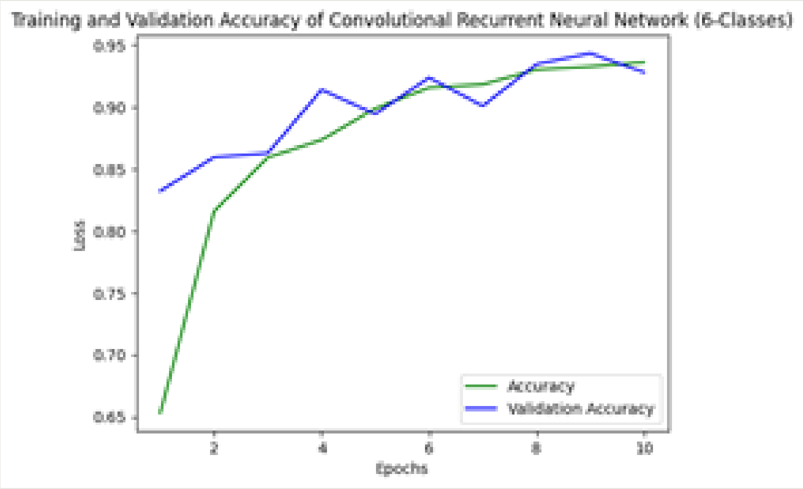


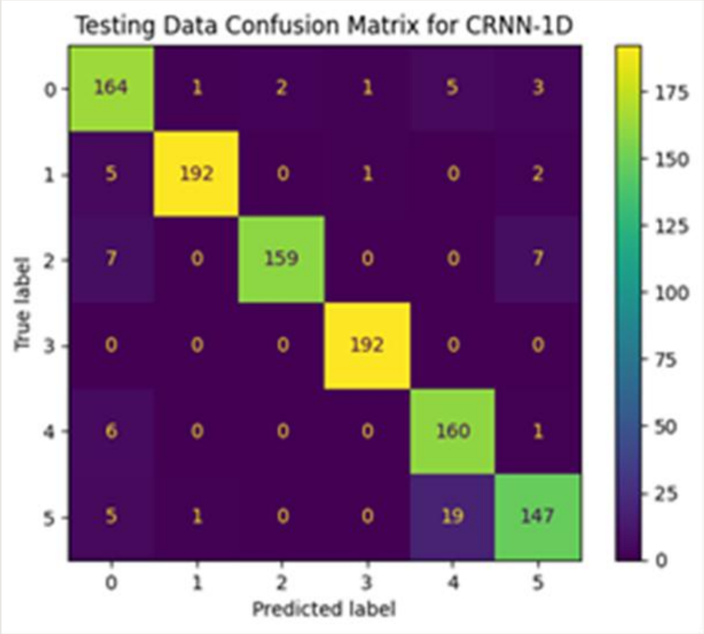


Result Analysis

- The Confusion matrices of Convolutional Recurrent Neural Network (CRNN) is given below:

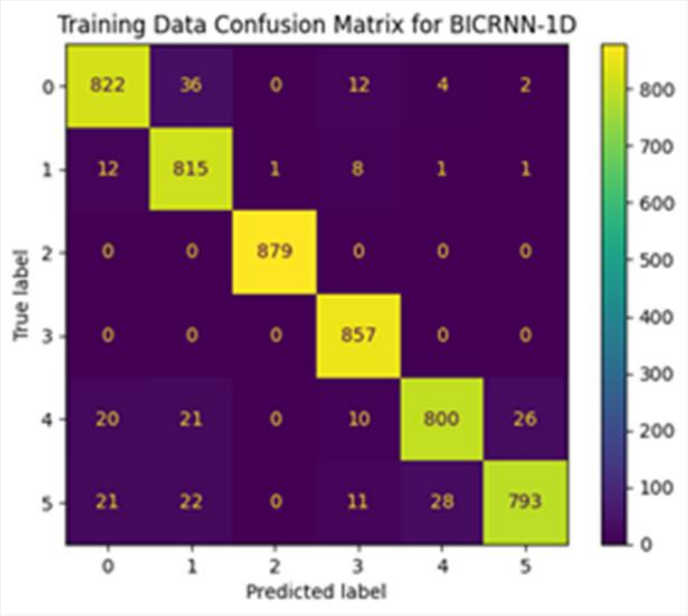


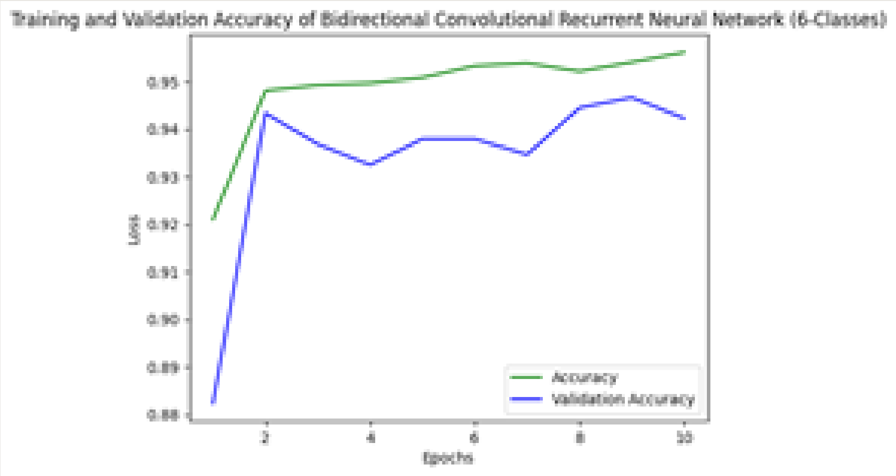


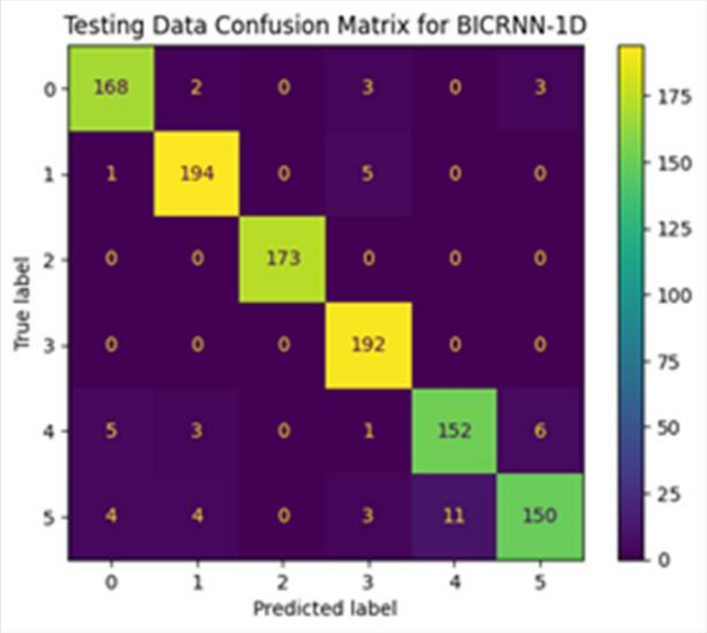


Result Analysis

- The Confusion matrices of Convolutional Neural Network (CNN) is given below:







Result Analysis (Contd.)

- Accuracy and loss for the models are shown below

Type	CNN		CRNN		Bi-CRNN	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
Training	0.951	0.204	0.938	0.150	0.956	0.132
Testing	0.950	1.680	0.938	0.150	0.952	0.159
Validation	0.942	0.204	0.928	0.182	0.942	0.200

Conclusion

- Audio pre-processing techniques were studied.
- The various feature extraction techniques were discussed.
- Extracted features are classified using various models and the results were compared.
- Multi-labelled audio classification is being researched such that the system can be made more accurate.

Thank You