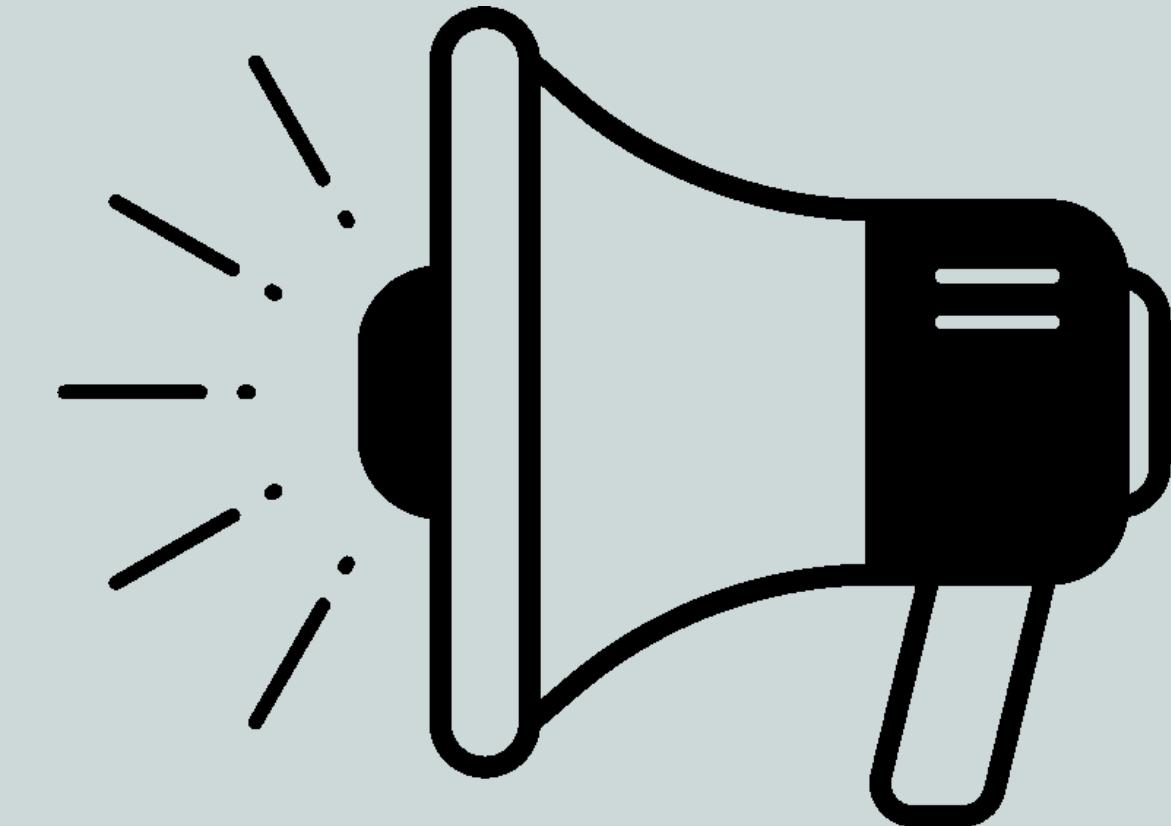


Music Genre Classification



By Flipping -A -Coin

Members:

- 1. Shambhavi Aggarwal**
- 2. Lakshay Chawla**
- 3. Devansh Shrestha**
- 4. Rommel Jalasutram**
- 5. Varnika Vatsyayan**

The Problem.

- Understand techniques used in the audio domain
- Build a machine learning model which classifies music into its respective genres.



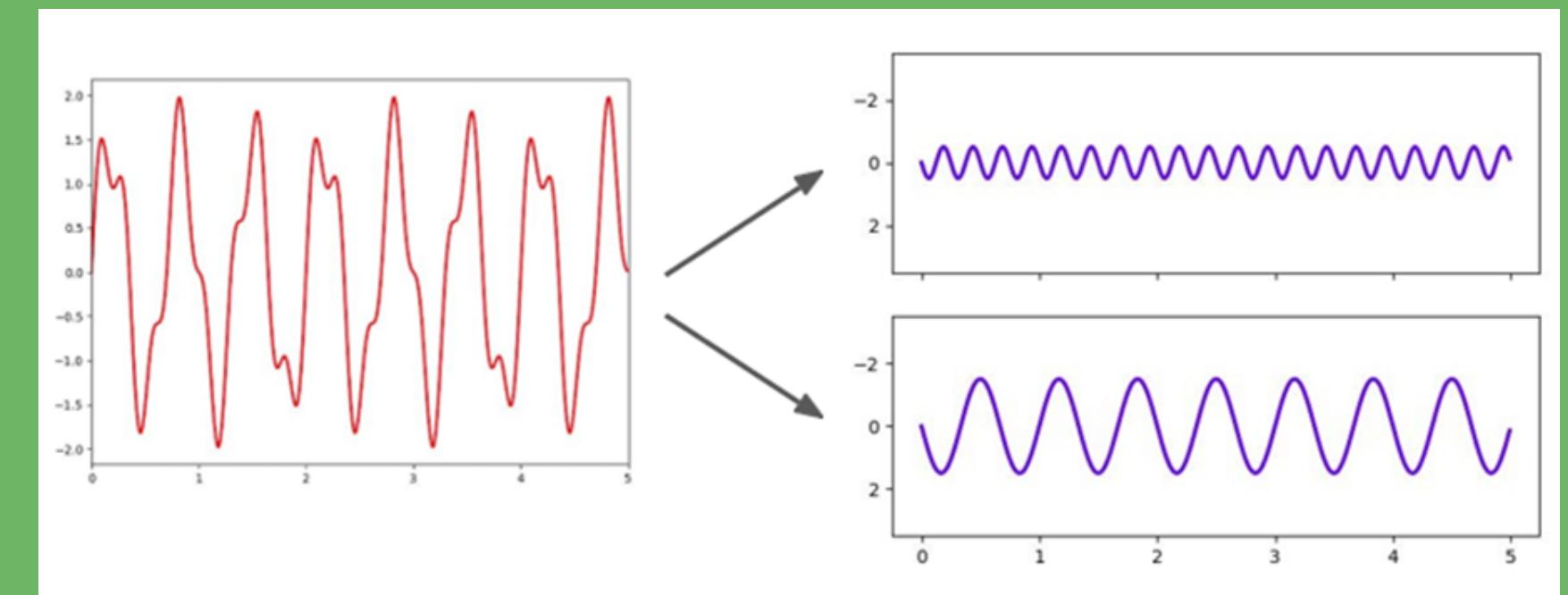
The Dataset.

- GTZAN Dataset
- 10 genres
- 100 songs per genre
- 30 seconds per song



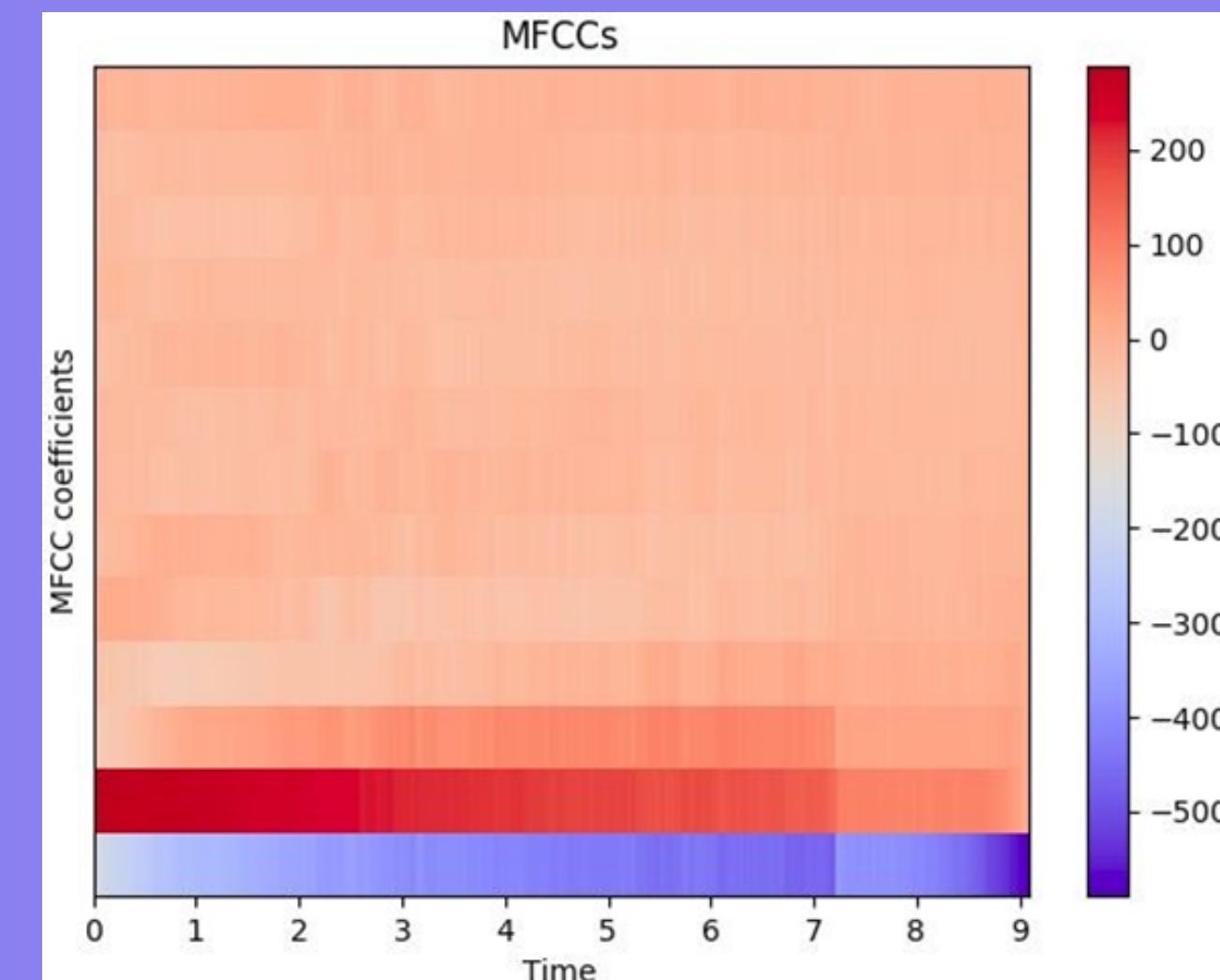
Audio PreProcessing

- Converted audio files to digital format.
- Perform FFT (Fast Fourier Transform)
- Perform STFT (Short Time Fourier Transform)
- Extract MFCCs
- Used Librosa Library



Mel Frequency Cepstral Coefficients (MFCCs)

- Capture timbral/textural aspects of sound
- Frequency domain feature
- Approximate human auditory system
- 13 to 40 coefficients
- Calculated at each frame



Feed Forward Neural Network

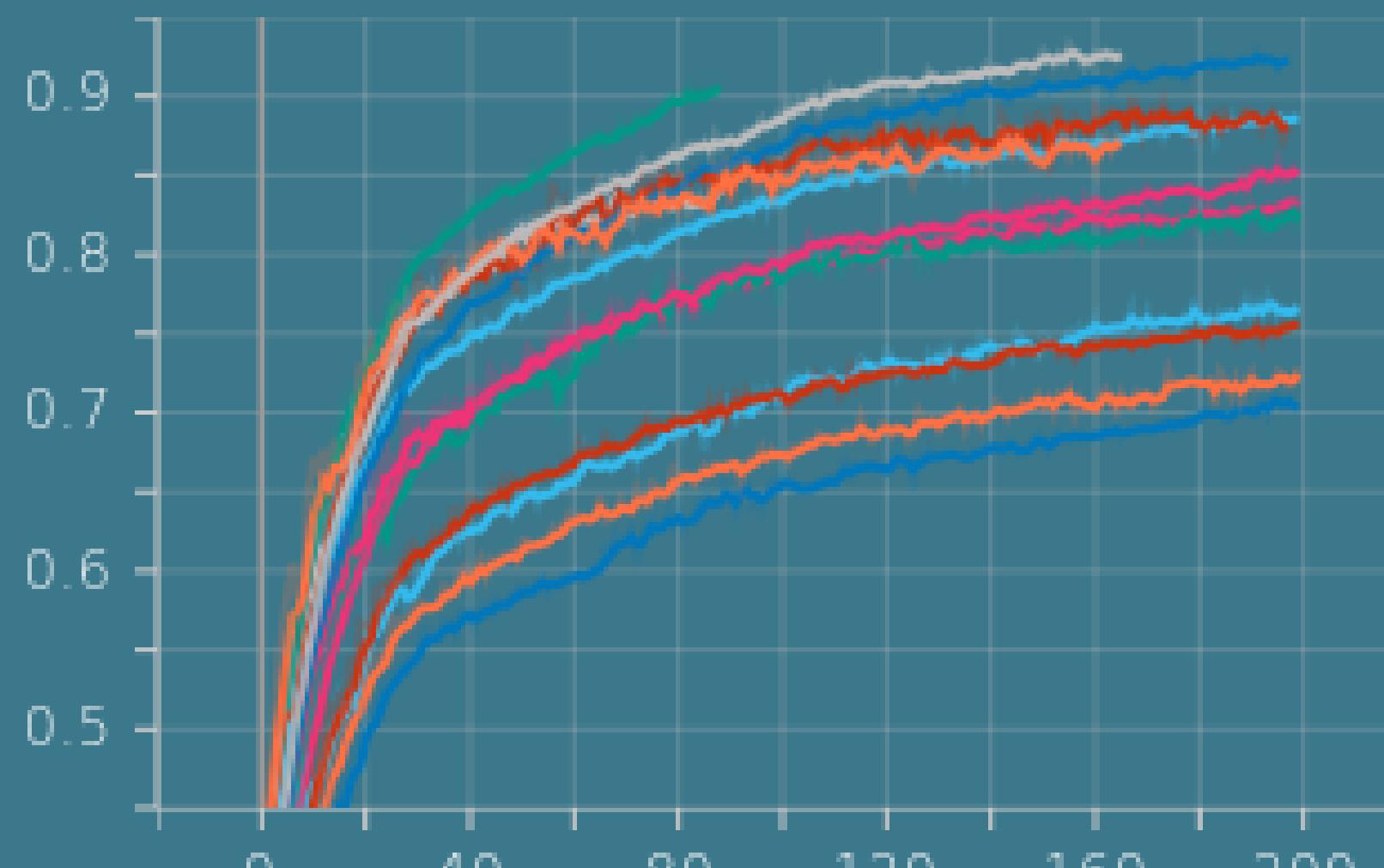
- Layers Used - Dense, Dropout and Batch normalization.
- Experimented with L1 and L2 regularizers.
- Best Validation Accuracy - **65.7%**

CONVOLUTIONAL NEURAL NETWORKS

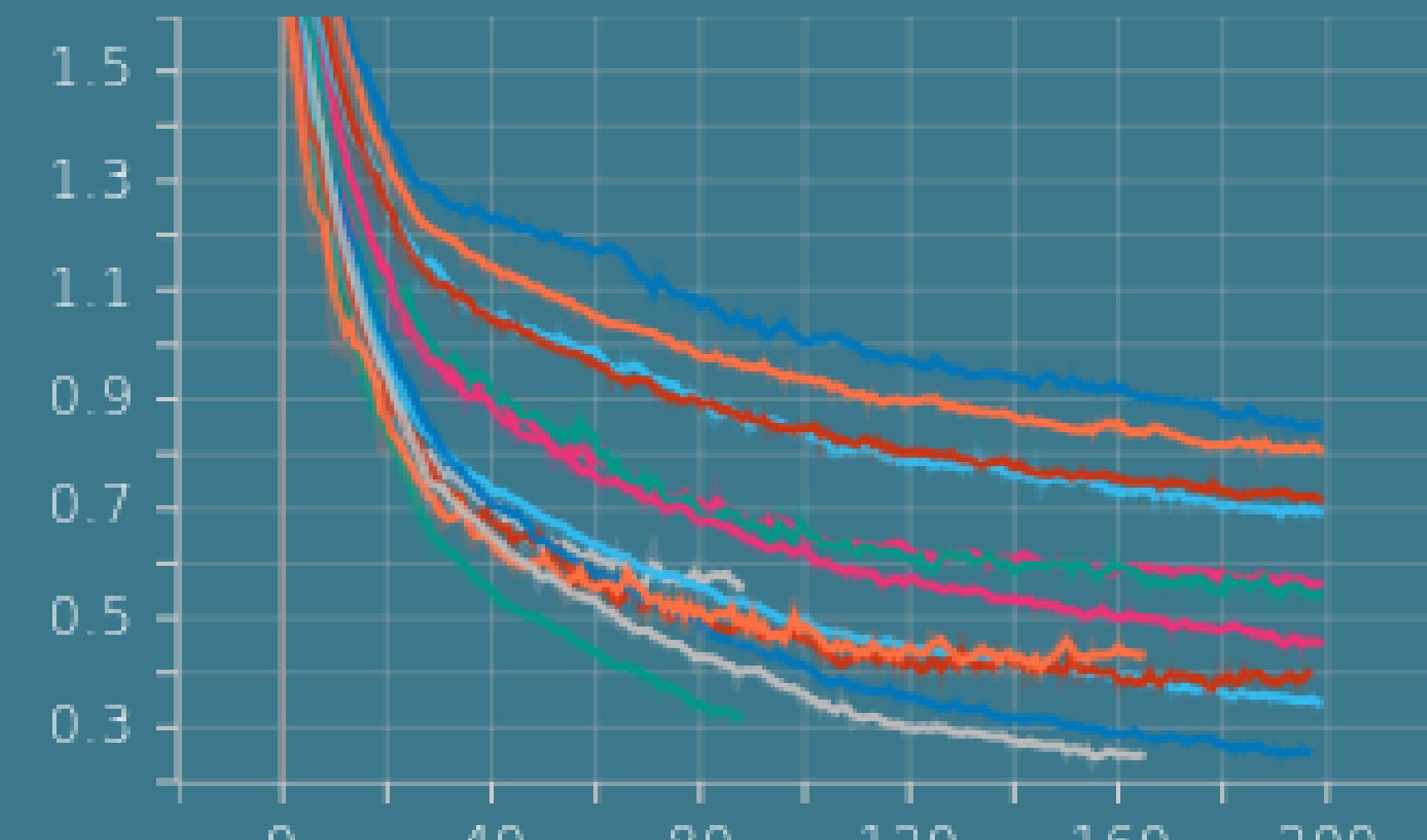
- Layers Used - Convolution, Max-pooling, Dense, Dropout and Batch normalization.
- Experimented with L2 regularizer, early stopping, lr scheduler.

Filters per Conv2D layer	Accuracy	Loss
[128, 64, 64, 32]	82.78	0.516
[256, 128, 64, 64]	85.53	0.518
[256, 128, 64, 32]	86.63	0.477
[32, 64, 128, 256]	79.98	0.640
[64, 128, 256, 256]	85.08	0.574

Tensorboard Visualisations



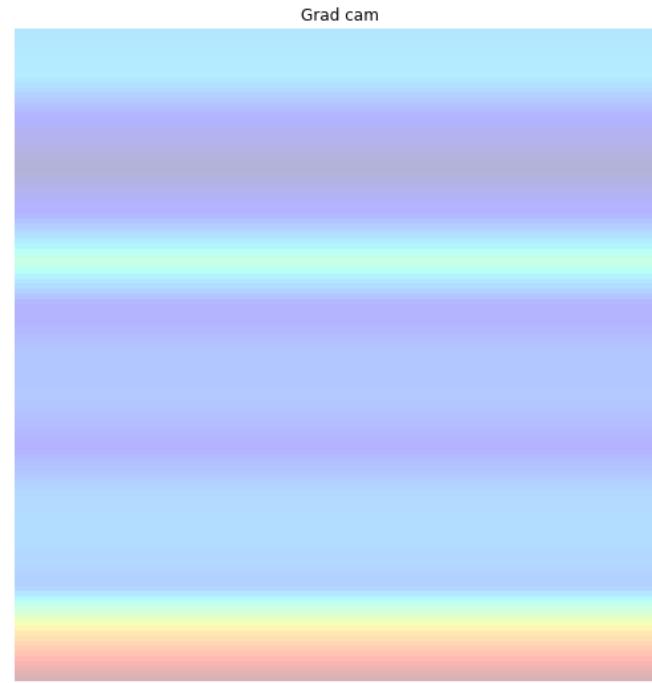
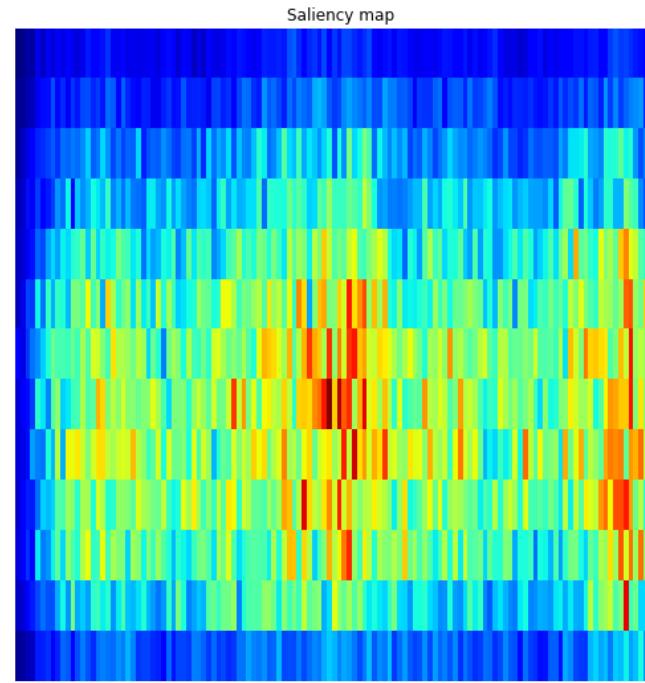
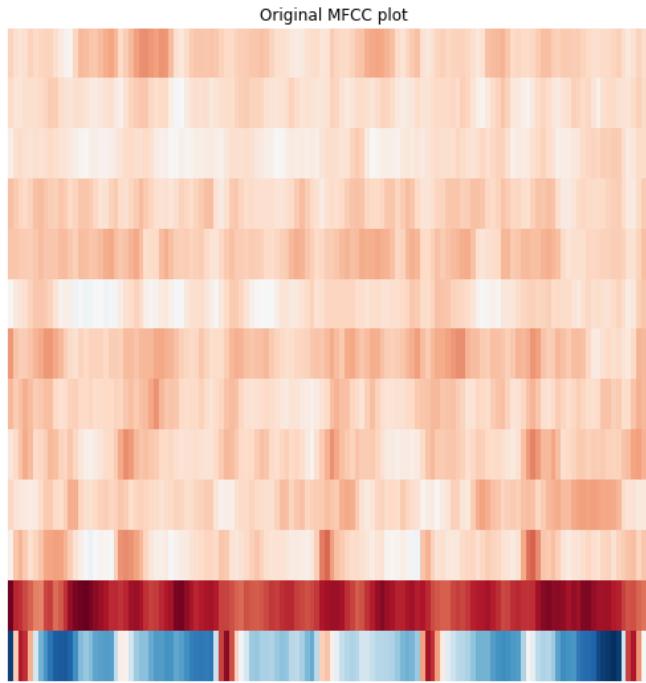
Epochs vs Accuracy



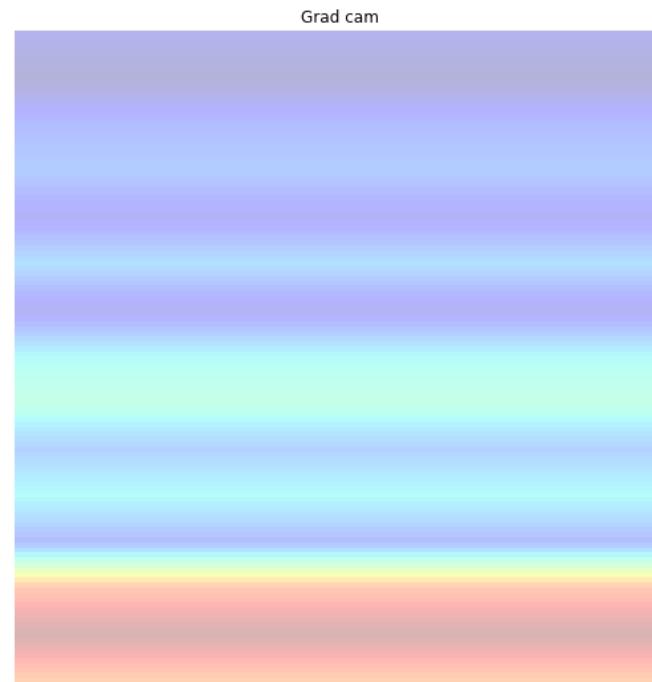
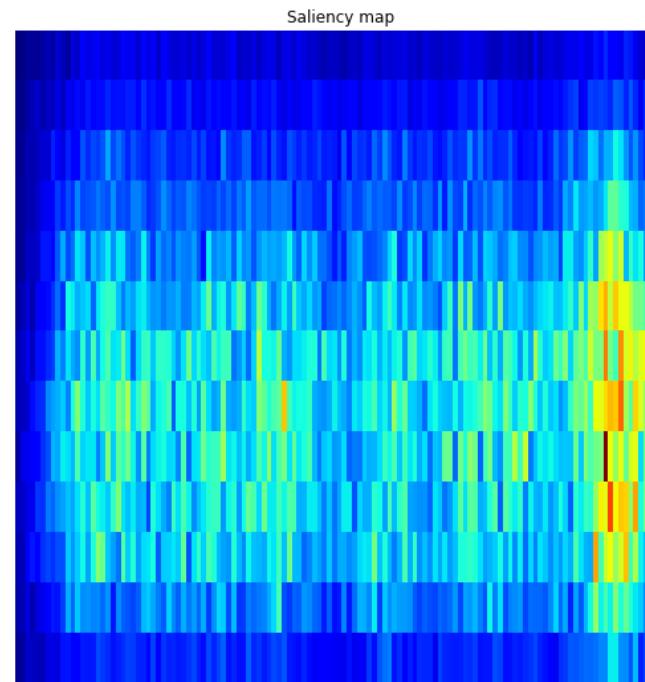
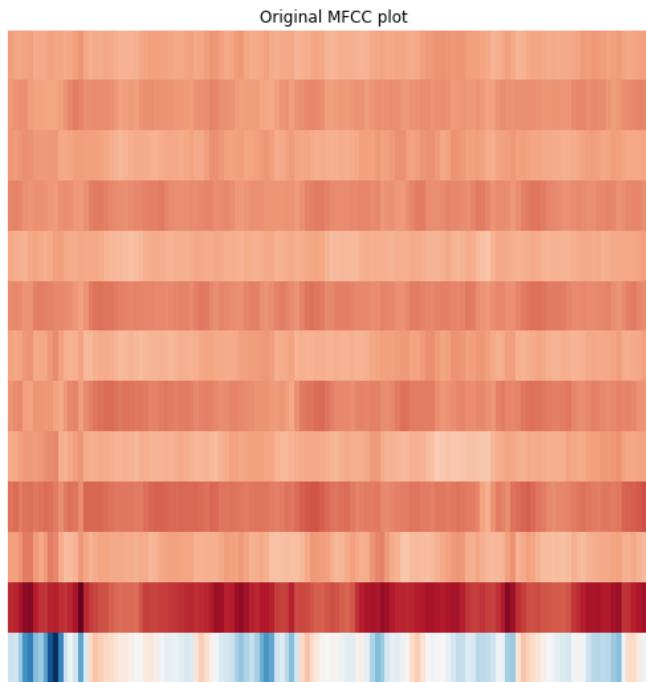
Epochs vs Loss

Saliency Maps

Correctly classified



Incorrectly classified



Experiments.

- Different variations of layers
- Used STFT and signal data
- Reconstruct audio from saliency maps



Future Scope.

- Use Auto-Encoders
- Use RNNs
- Use skip connections
- Use Transfer Learning



THANK YOU