Capstone Project:-
Speech Emotion Recognition

Deep Learning and MLE Project

Presented By:
Sagar Khekale

Contents

- Introduction
- Problem Statement
- Data Summary
- Dependencies
- Analysis of Data
- Data Augmentation
- Model
- Confusion Matrix
- Loss Accuracy Plot
- Challenges
- Conclusions

Introduction

- **What is Speech Emotion Recognition ?**

- **Why it is so important?**

- **What is the scopes of Speech Emotion Recognition?**

Problem Statement

•As human beings speech is amongst the most natural way to express ourselves.

•We depend so much on it that we recognize its importance when resorting to other communication forms like emails and text messages where we often use emojis to express the emotions associated with the messages.

•As emotions play a vital role in communication, the detection, and analysis of the same is of vital importance in today's digital world of remote communication.

•Emotion detection is a challenging task because emotions are subjective.

• We define an SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them

Data Summary

This portion of the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) contains 1440 files: 60 trials per actor x 24 actors = 1440. The RAVDESS contains 24 professional actors (12 female, 12 male), vocalizing two lexically-matched statements in a neutral North American accent. Speech emotions includes calm, happy, sad, angry, fearful, surprise, and disgust expressions. Each expression is produced at two levels of emotional intensity (normal, strong), with an additional neutral expression.

Filename example: 03-01-06-01-02-01-12.wav

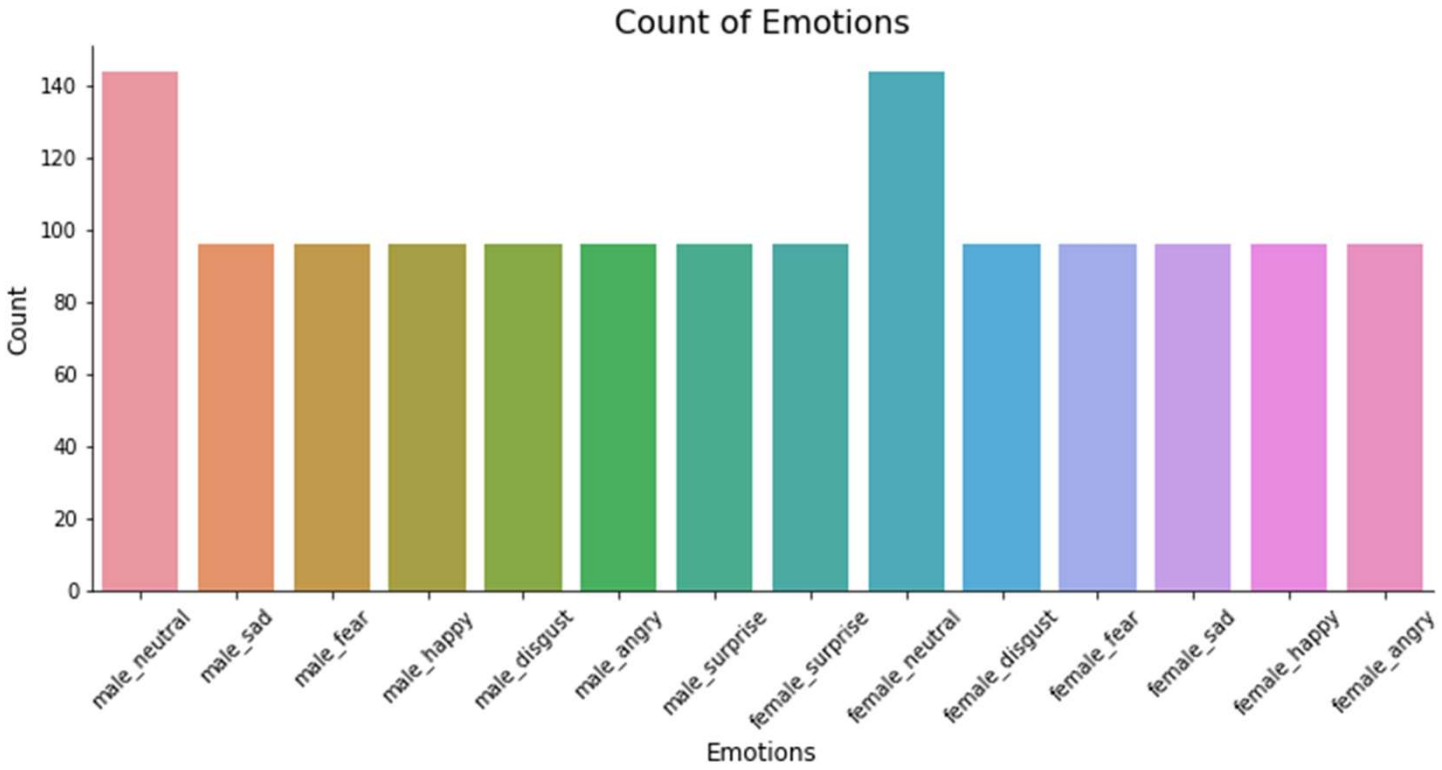
Dataset Link: https://www.kaggle.com/uwrfkaggler/ravdess-emotional-speech-audio

Dependencies

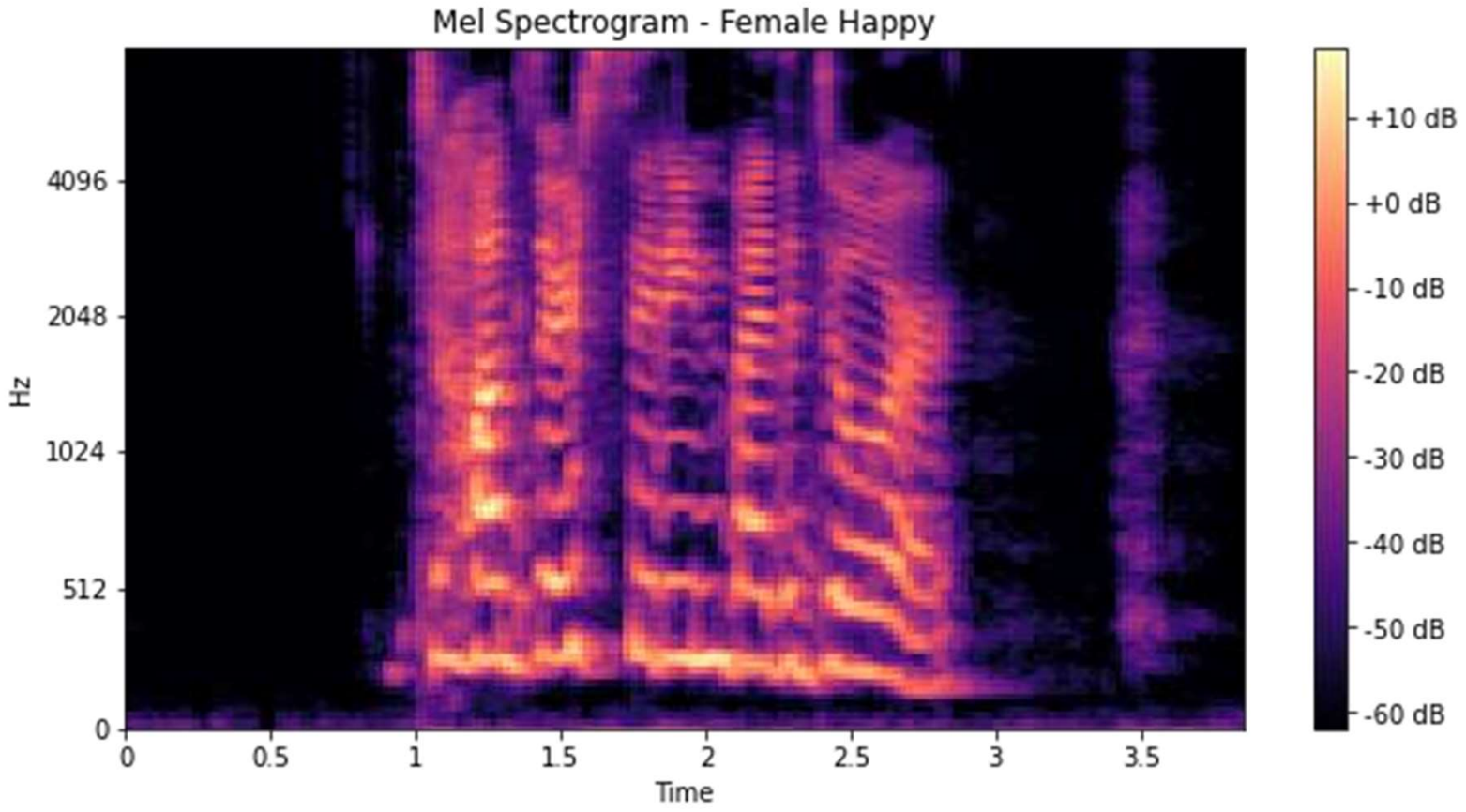
- **Numpy**
- **Pandas**
- **Matplotlib**
- **Pillow**
- **librosa**
- **Streamlit**
- **tensorflow-cpu**
- **Keras**

Analysis Of Data

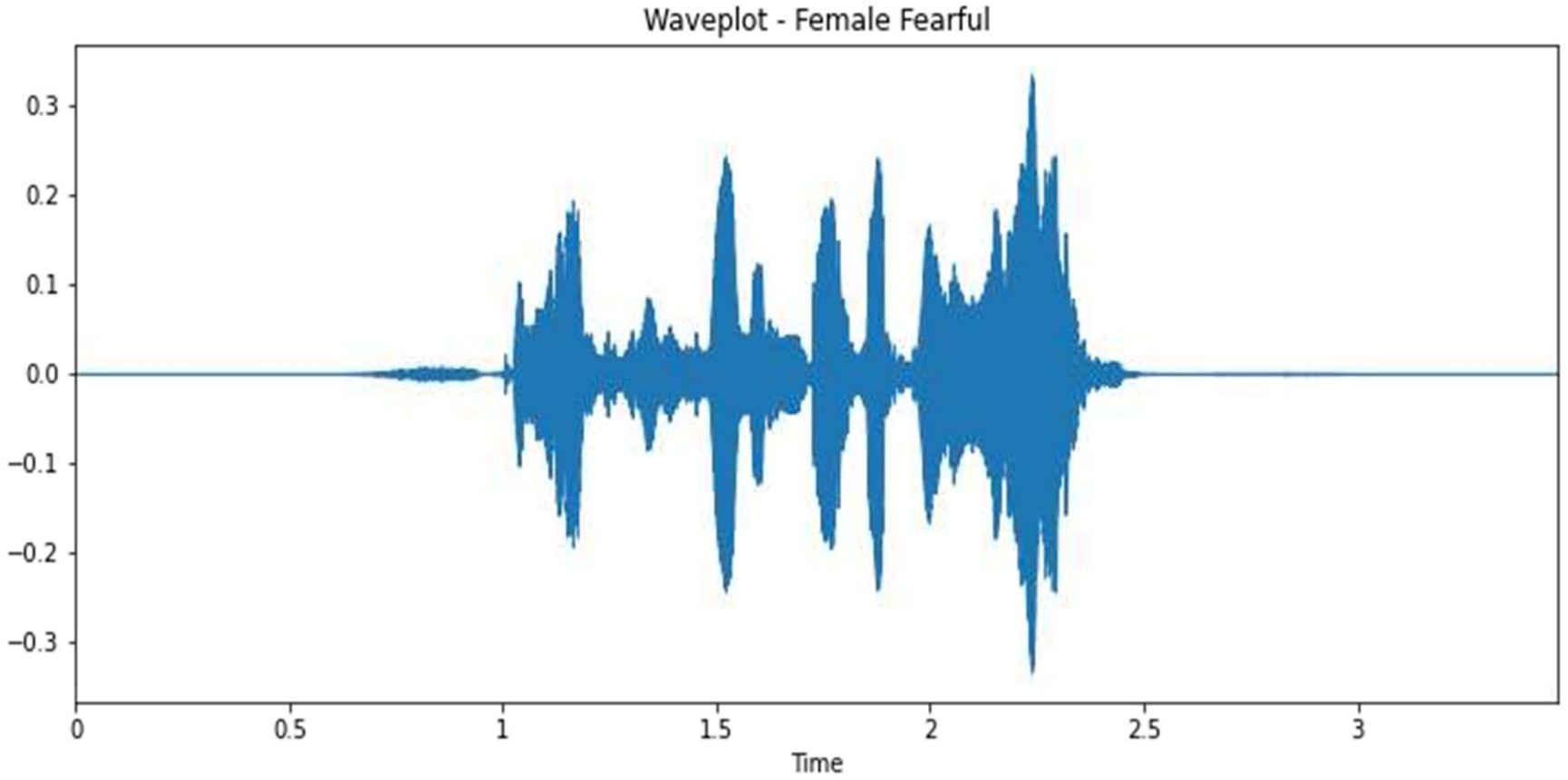
1)EDA



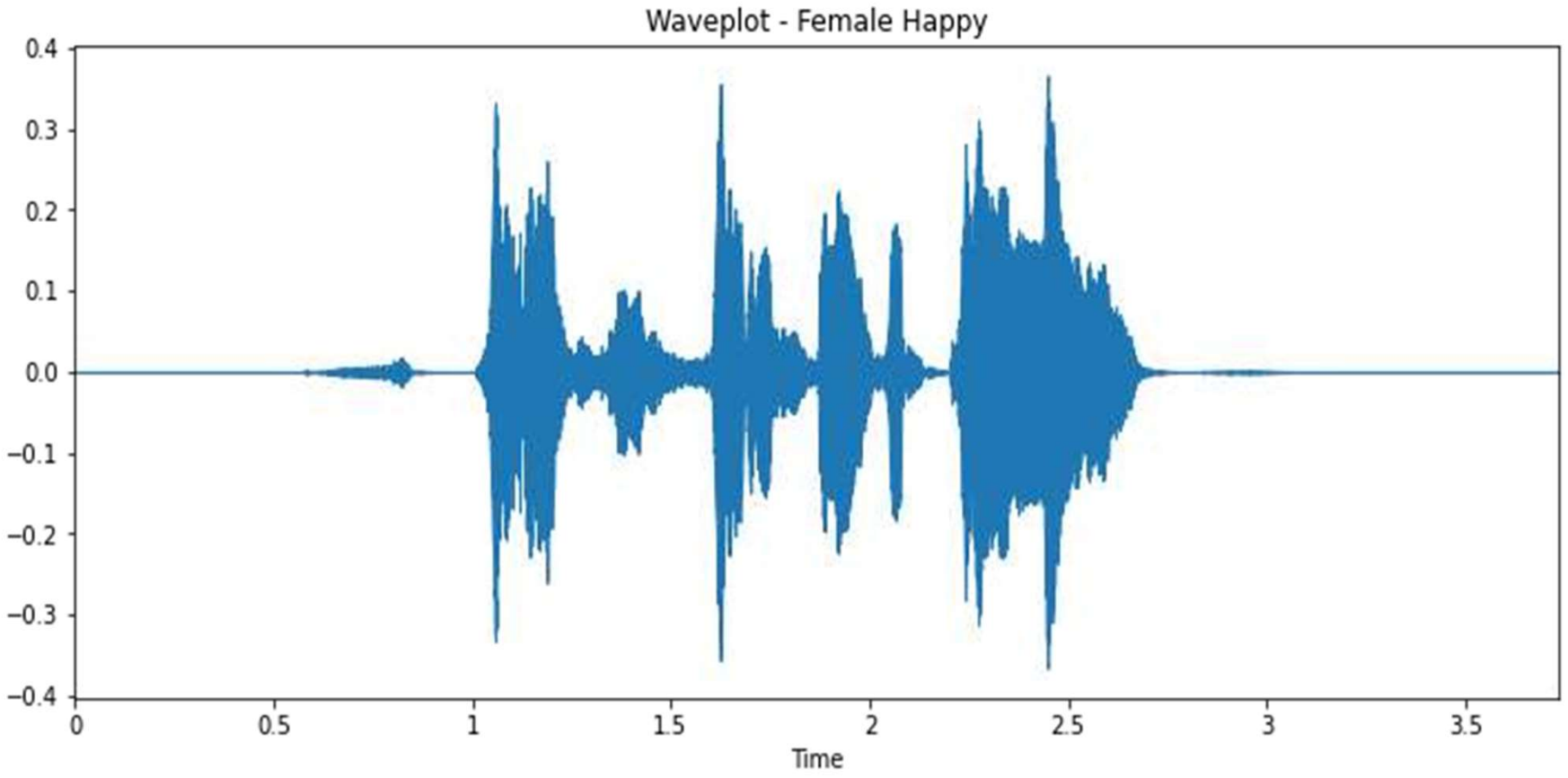
2) Mel Spectrogram – Female Happy



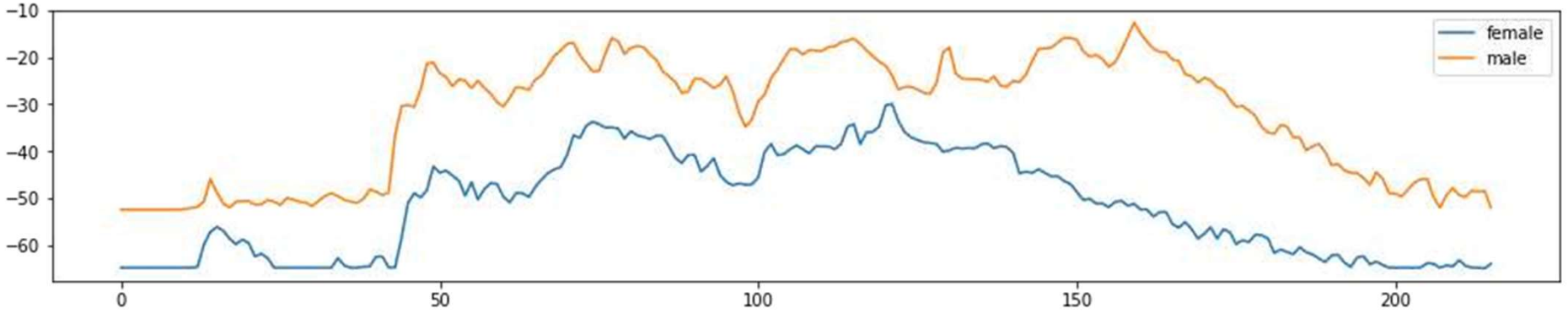
3) Fear Track



4) Happy Track

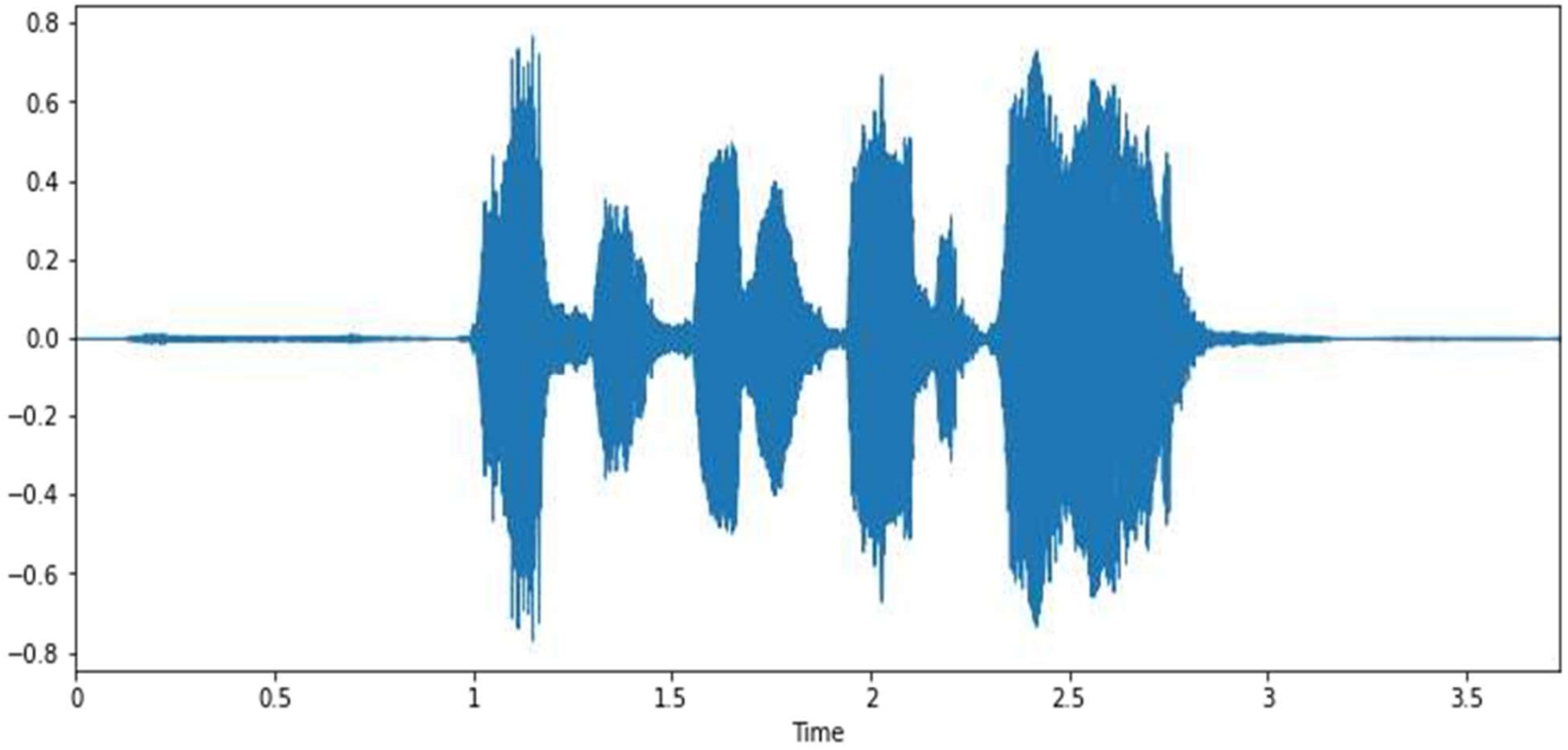


Compare Male and Female Voice

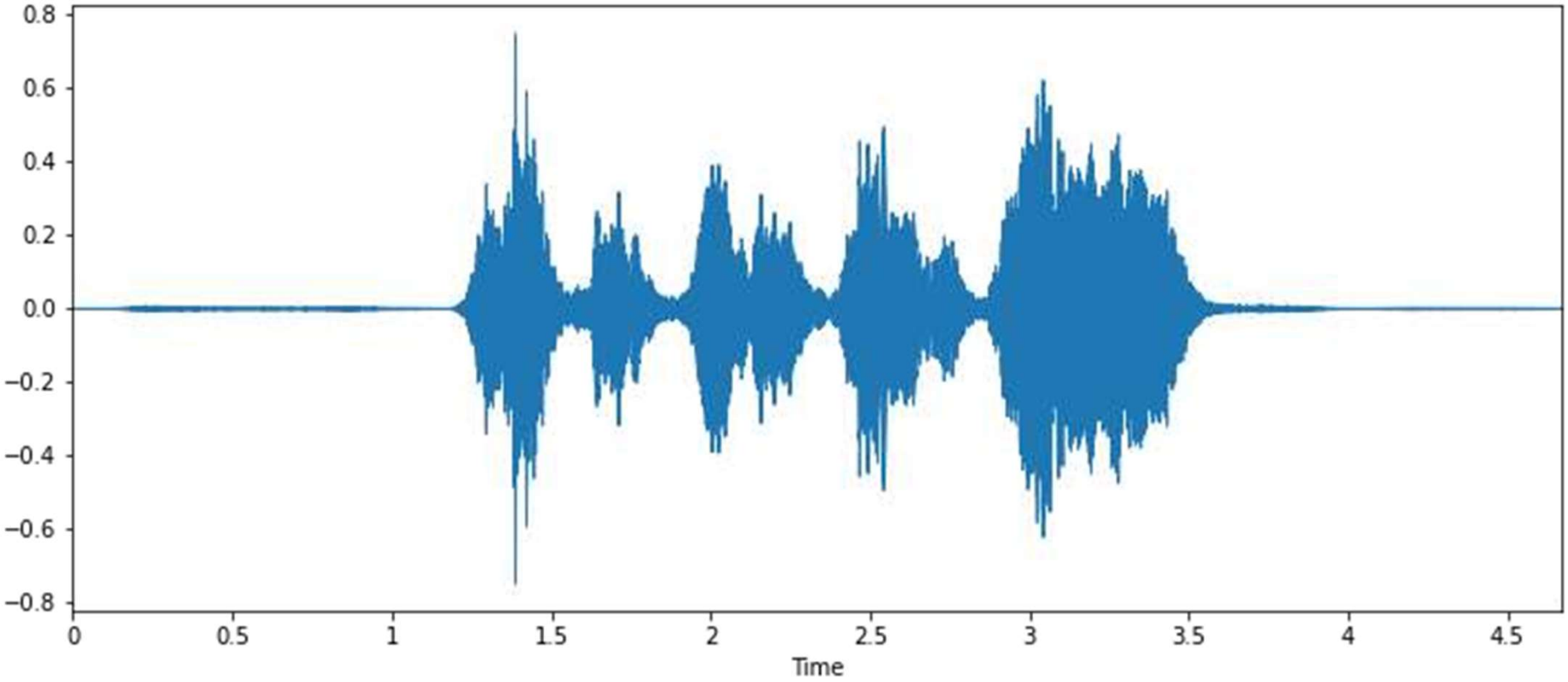


Data Augmentation

1) Normal Audio

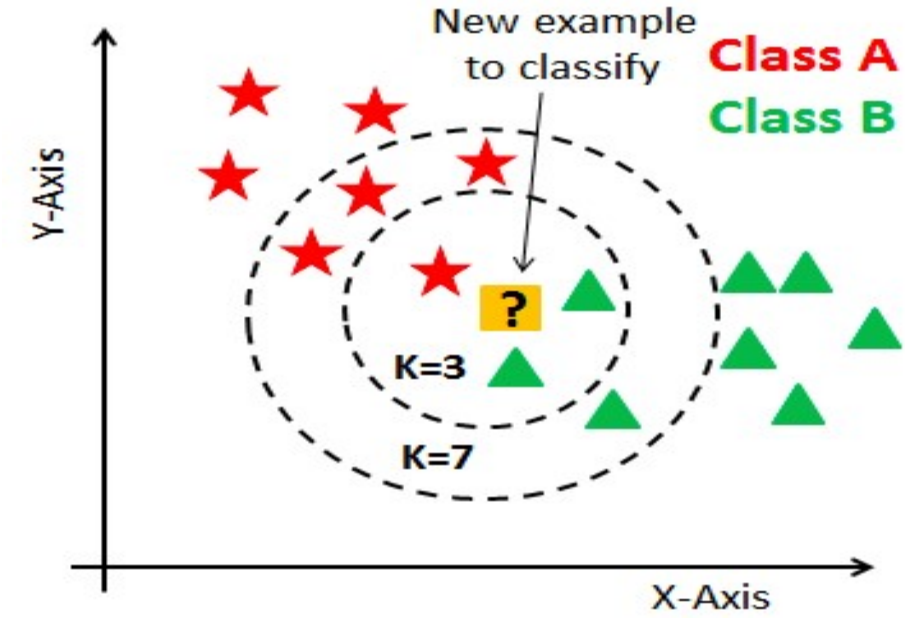


1) Stretch Audio

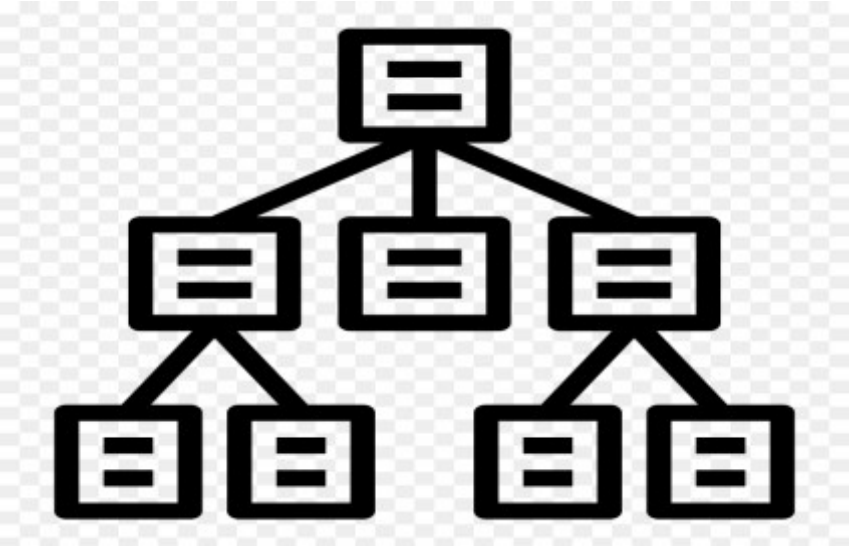


Model

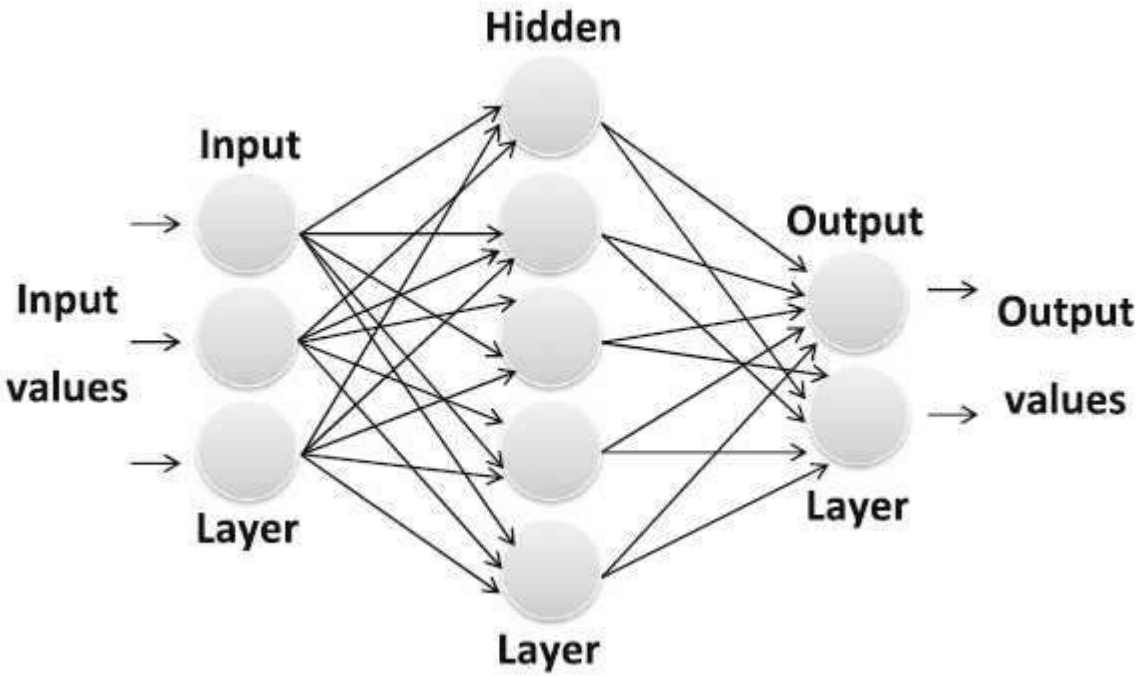
1. KNN(K-nearest neighbor)



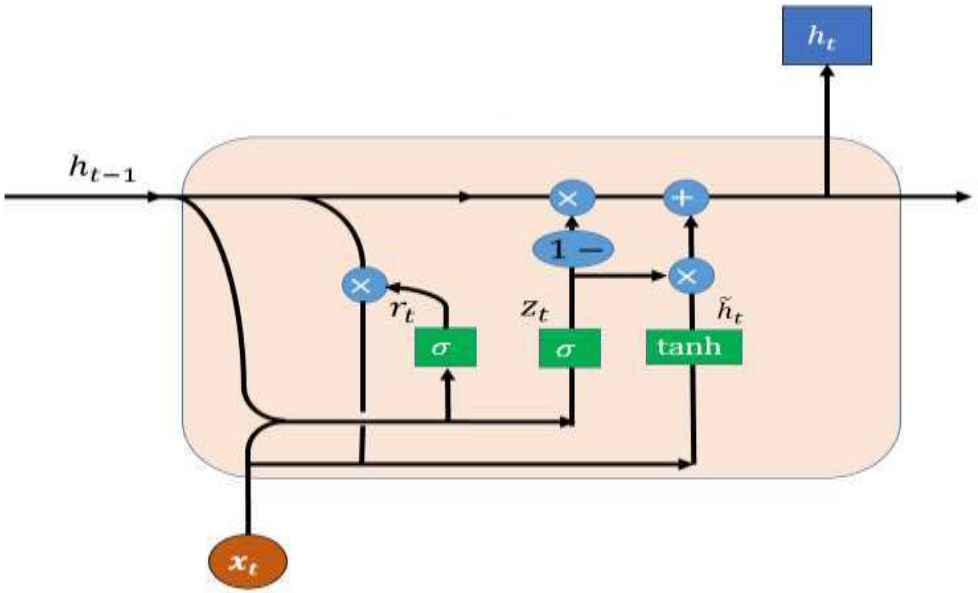
2. Decision Tree



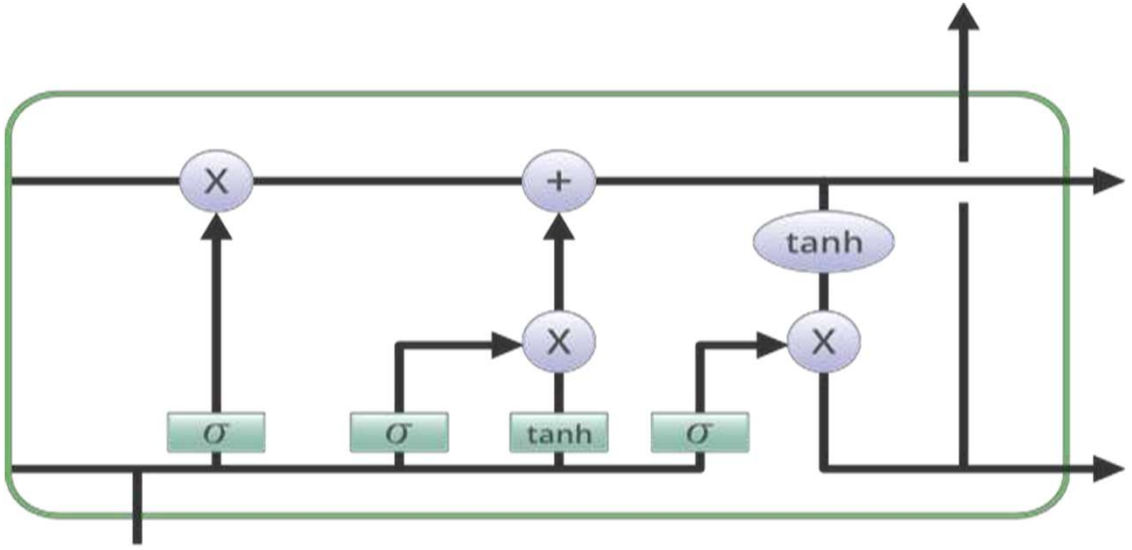
3. MLP(Multilayer Perceptron) Classifier



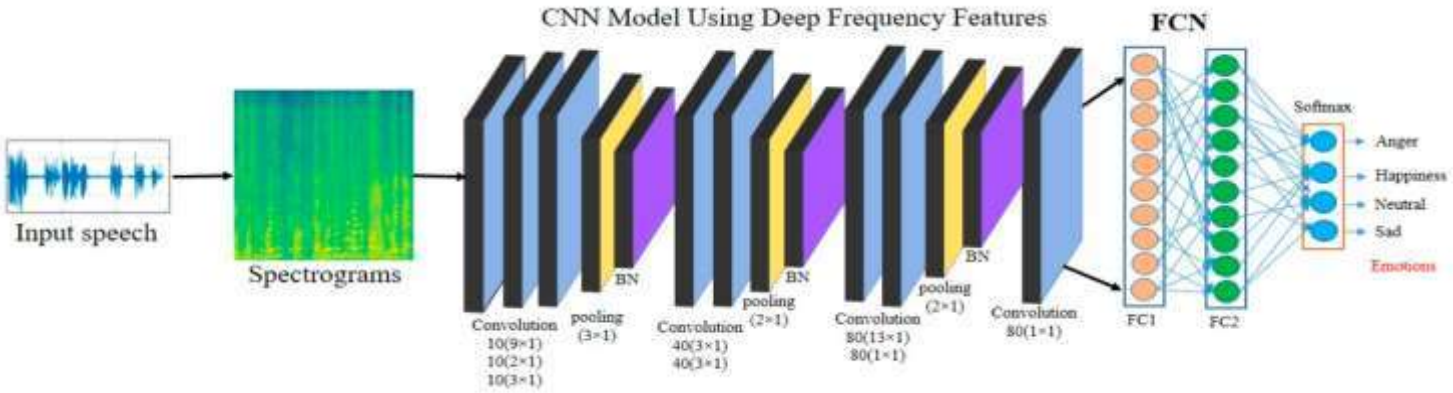
4. GRU(Gated Recurrent Unit)



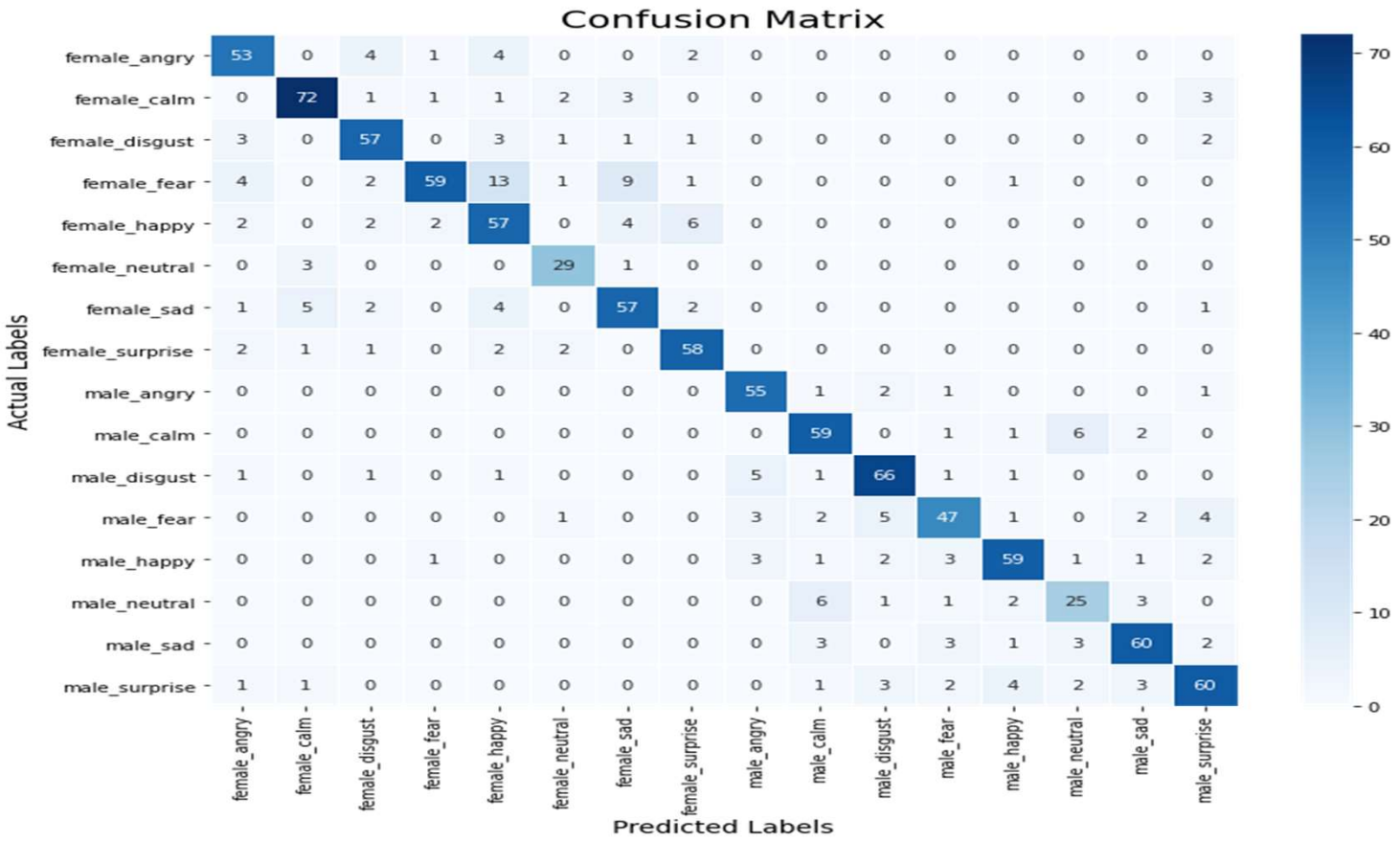
5. LSTM



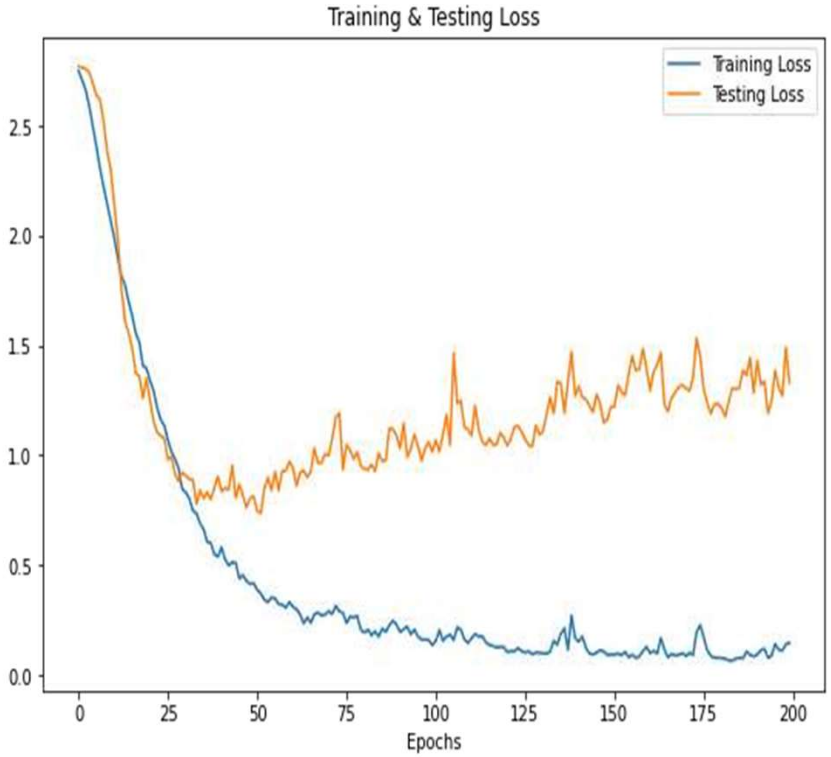
6.CNN

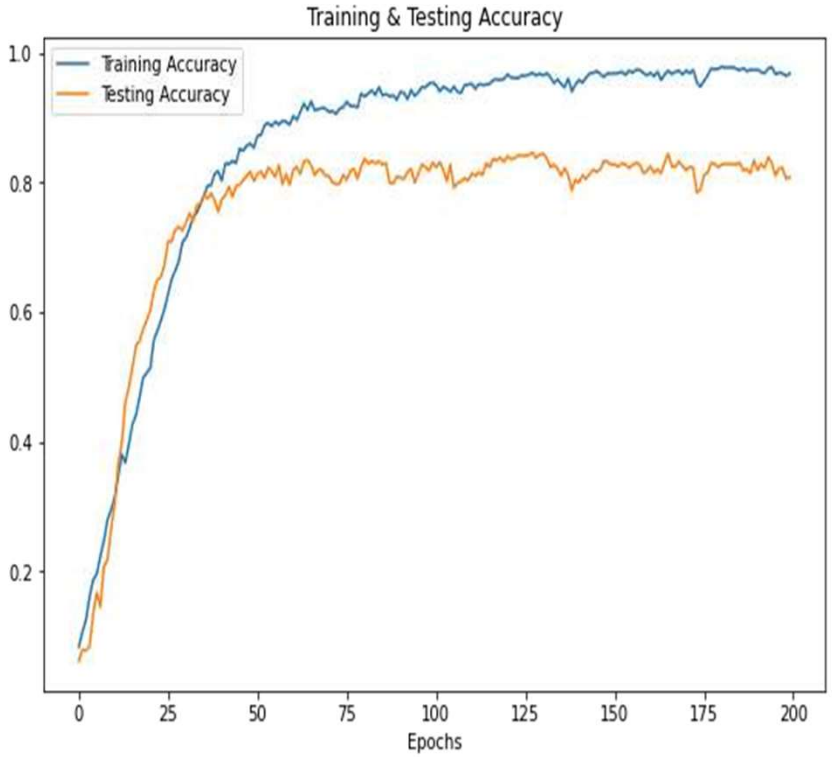

CNN Model Using Deep Frequency Features
FCN
Input speech
Spectrograms
Convolution
10(9×1)
10(2×1)
10(3×1)
pooling
(3×1)
BN
Convolution
40(3×1)
40(3×1)
pooling
(2×1)
BN
Convolution
80(13×1)
80(1×1)
pooling
(2×1)
BN
Convolution
80(1×1)
FC1
FC2
Softmax
Anger
Happiness
Neutral
Sad
Emotions

Confusion Matrix



Loss Accuracy Plot





Challenges

- Large speech Dataset to handle
- Annotating an audio recording is challenging. Should we label a single word , sentence or a whole conversation? How many emotions should we define to recognize
- It is not an easy task to convert speech to emotion
- Emotions are subjective, people would interpret it differently. It is hard to define the notion of emotions

Conclusion

- Through this project, we showed how we can leverage Machine learning to obtain the underlying emotion from speech audio data and some insights on the human expression of emotion through voice.

- This system can be employed in a variety of setups like Call Centre for complaints or marketing, in voice-based virtual assistants or chatbots, in linguistic research, etc

- Codes which are deployed are in Github Repository.

- It was such an amazing and interesting project. We learnt a lot from this.

THANK YOU

Q & A