

# Speech Emotion Recognition using CNN

By: Meshal Alamr  
&  
Norah Alkhalifah

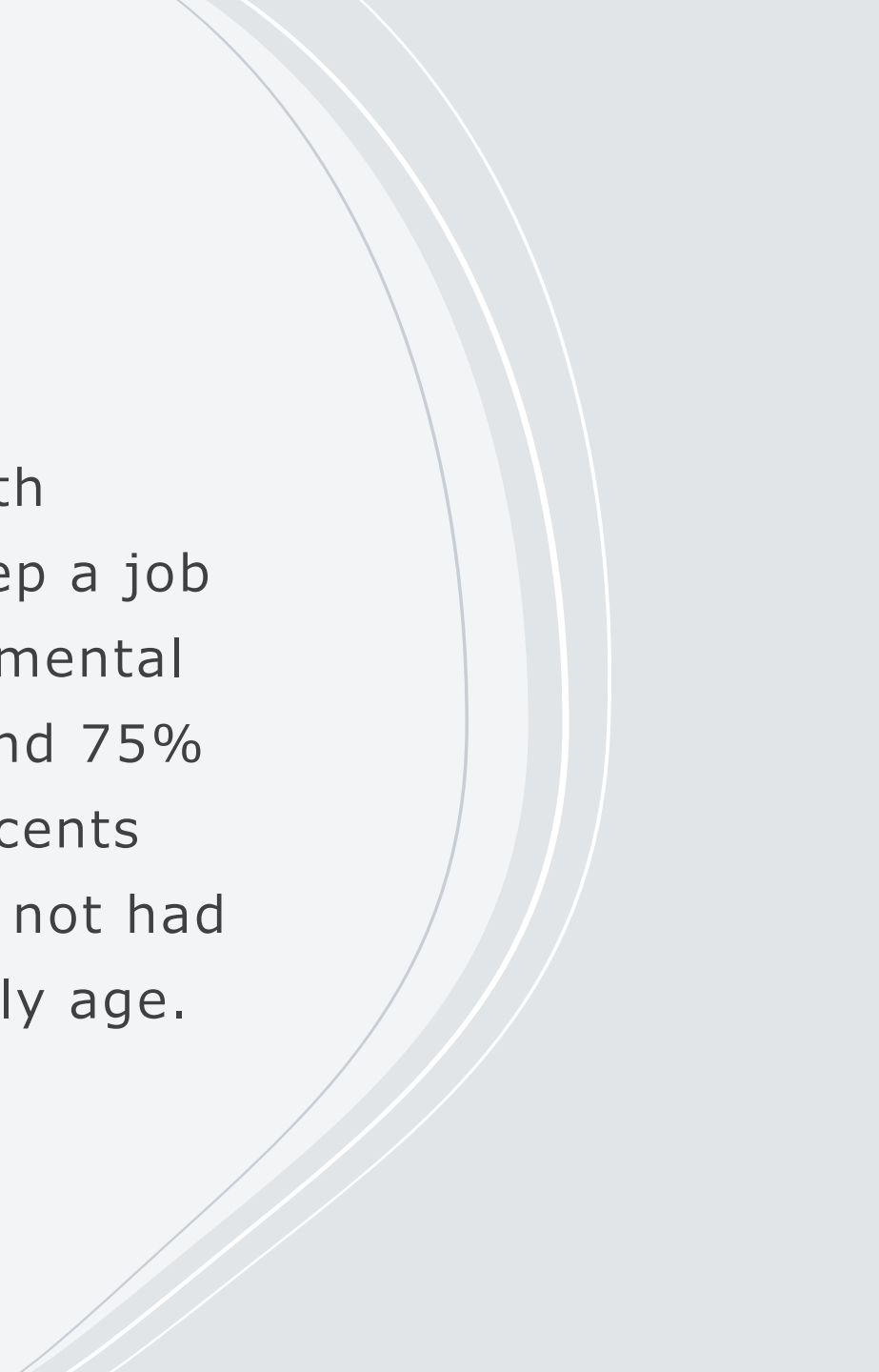


# Outline

- **Introduction**
- **EDA**
- **Data Augmentation**
- **Feature Extraction**
- **Model Building**
- **Conclusion**



# **Introduction**



Mental illness can make it difficult to cope with everyday life. It can inhibit your ability to keep a job and interact with others. Alarming, 50% of mental health problems are established by age 14, and 75% by age 24. Sadly, 70% of children and adolescents who experience mental health problems have not had appropriate interventions at a sufficiently early age.

**“The sooner the better” as told by many professionals.  
Having a lead to a problem may reduce the time and effort of  
solving that problem.**



**EDA**

# Datasets

---

## **RAVDESS**

- 1440 unique files with 8 emotions

---

## **Tess**

- 2800 unique files with 7 emotions

---

## **Crema**

- 7442 unique files with 6 emotions

---

## **Savee**

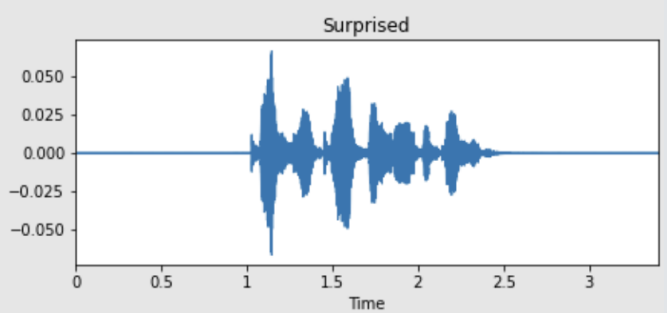
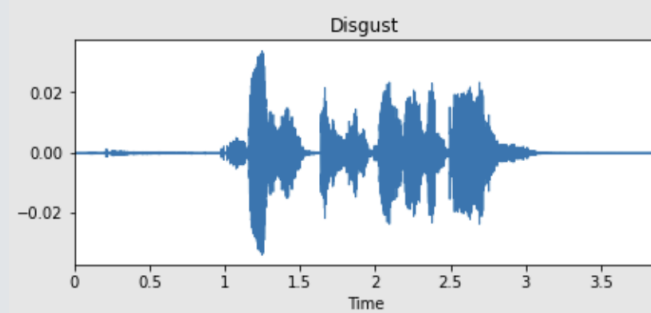
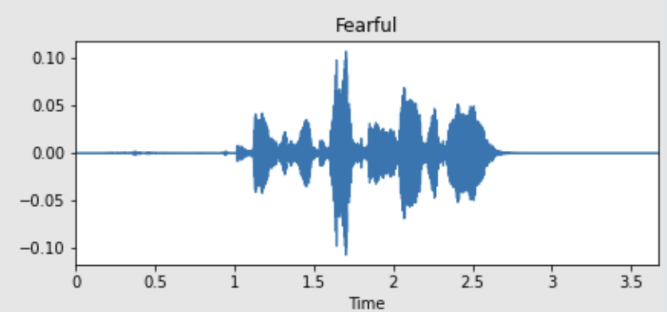
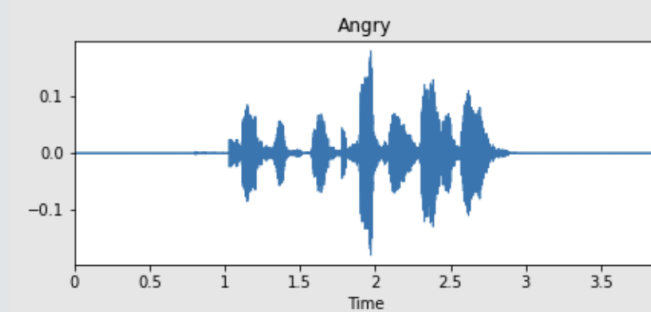
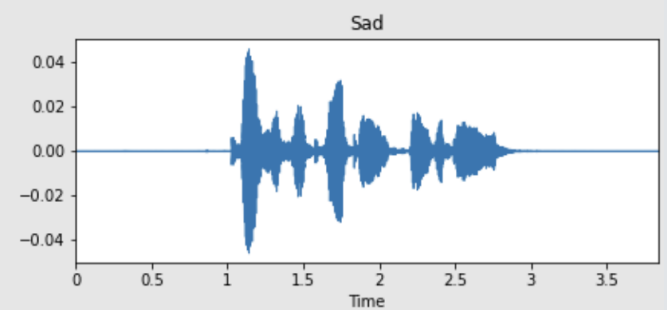
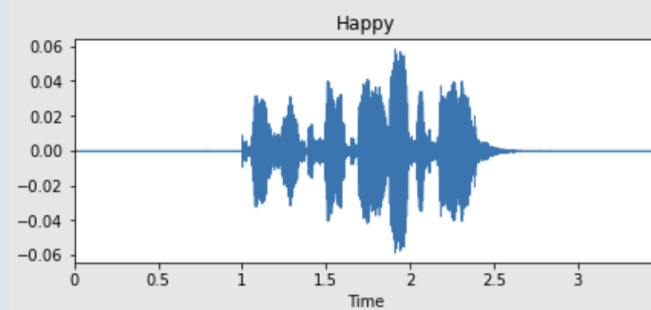
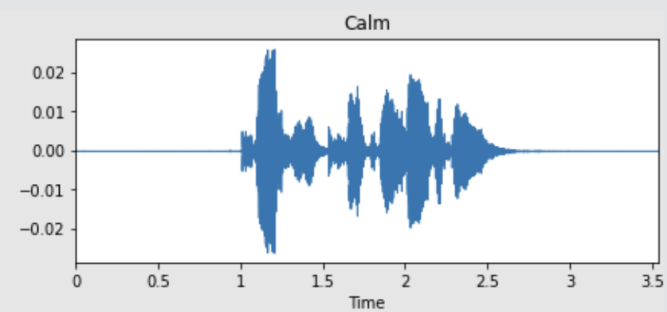
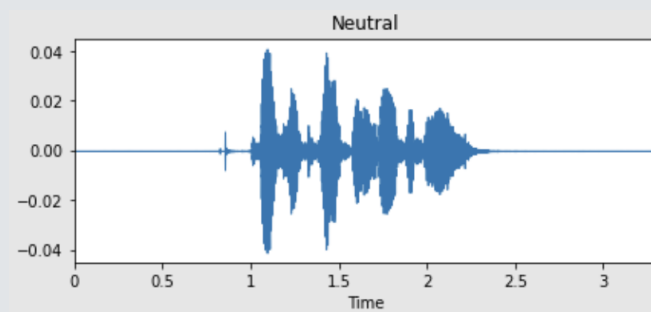
- 480 unique files with 7 emotions

**A total of 12,162 unique audio files with 8 emotion labels**

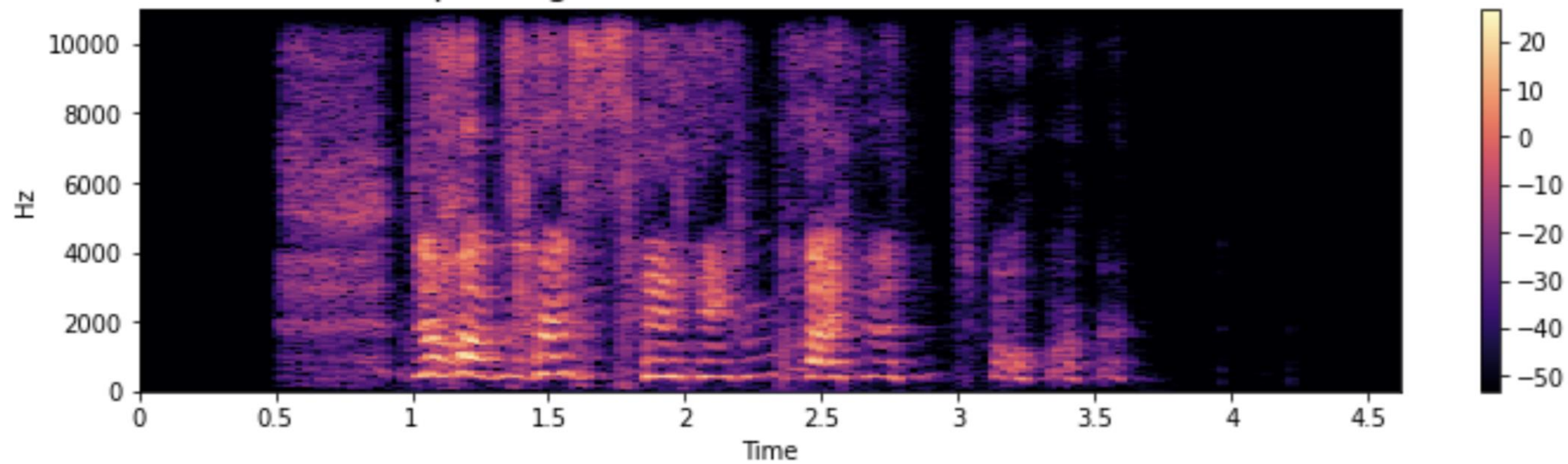
# The Combined Dataset

	Emotion	Path
0	fear	/content/Datasets/audio_speech_actors_01-24/Ac...
1	neutral	/content/Datasets/audio_speech_actors_01-24/Ac...
2	surprise	/content/Datasets/audio_speech_actors_01-24/Ac...
3	fear	/content/Datasets/audio_speech_actors_01-24/Ac...
4	neutral	/content/Datasets/audio_speech_actors_01-24/Ac...
...	...	...
12157	happy	/content/Datasets/Savee/KL_h08.wav
12158	angry	/content/Datasets/Savee/JE_a07.wav
12159	disgust	/content/Datasets/Savee/JE_d09.wav
12160	angry	/content/Datasets/Savee/KL_a04.wav
12161	disgust	/content/Datasets/Savee/JE_d07.wav
12162 rows x 2 columns		





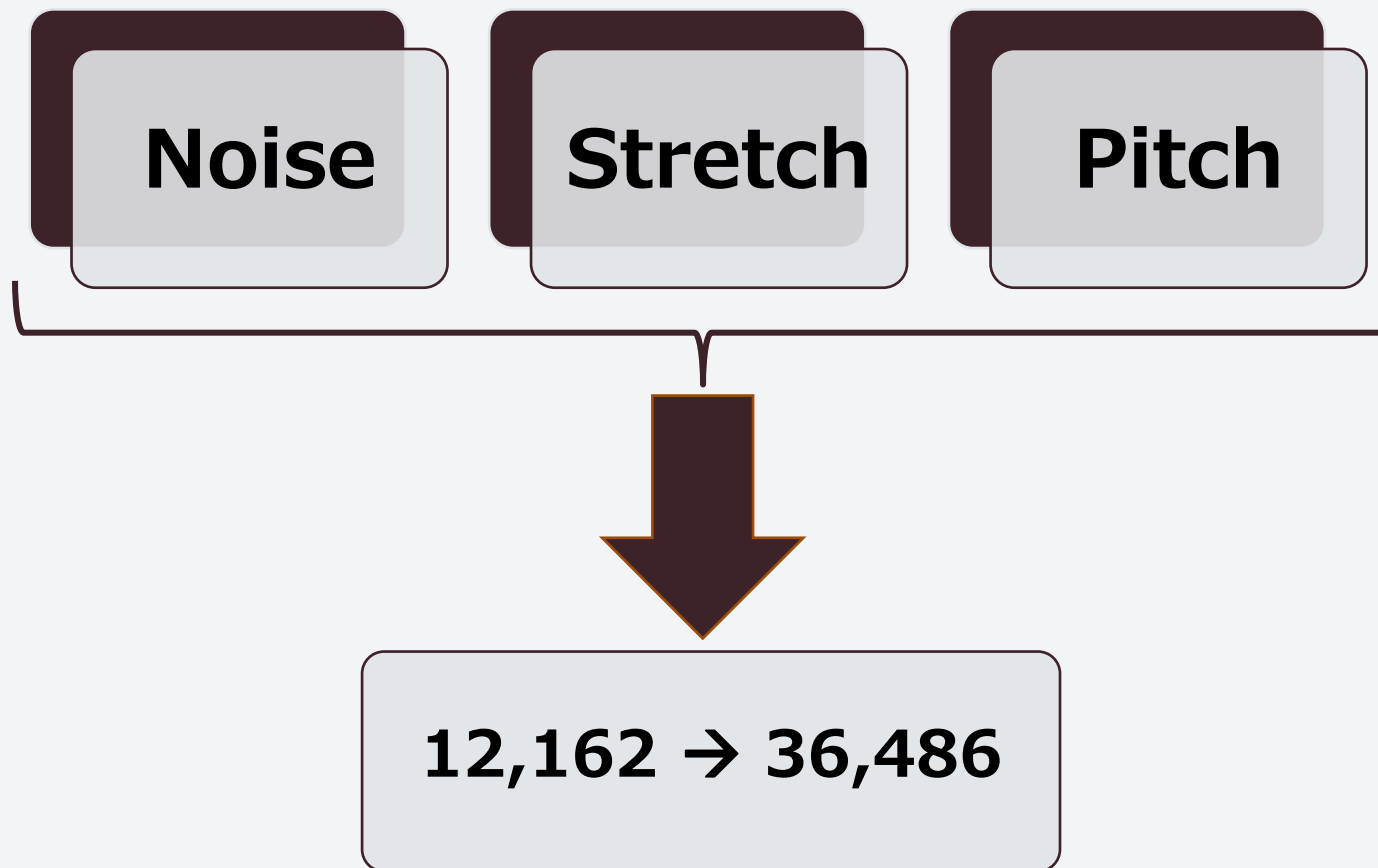
Spectrogram for audio with sad emotion





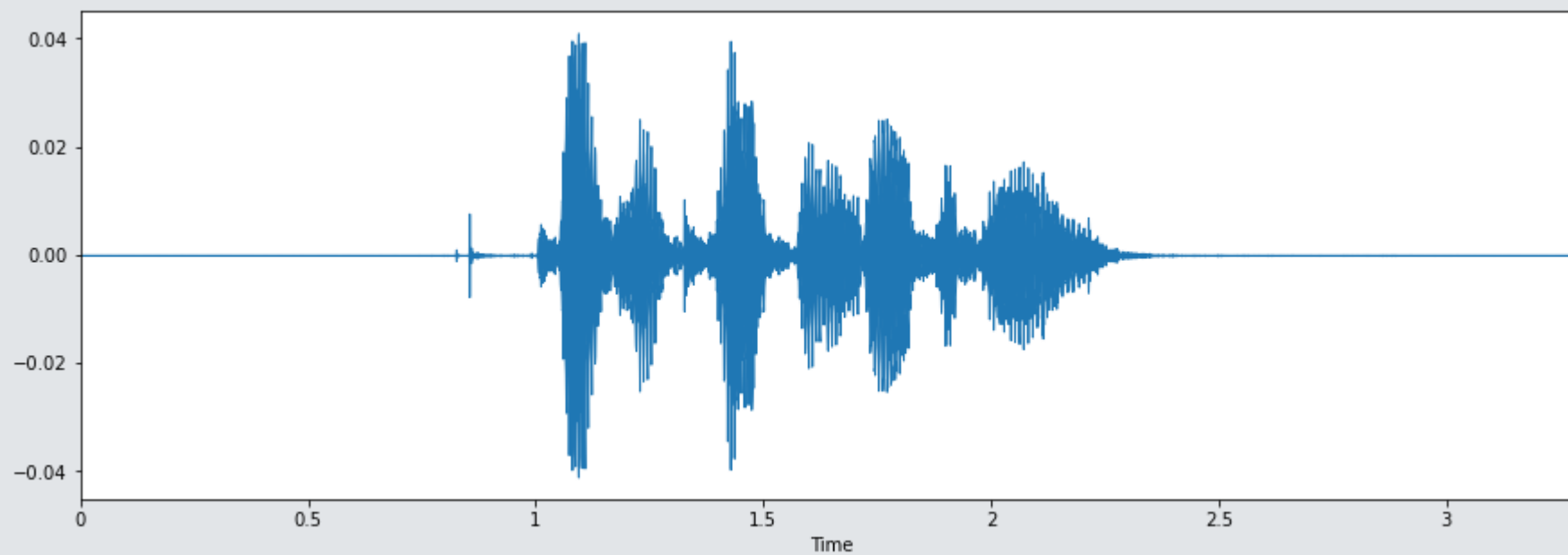
# **Data Augmentation**

# Data Augmentation Techniques



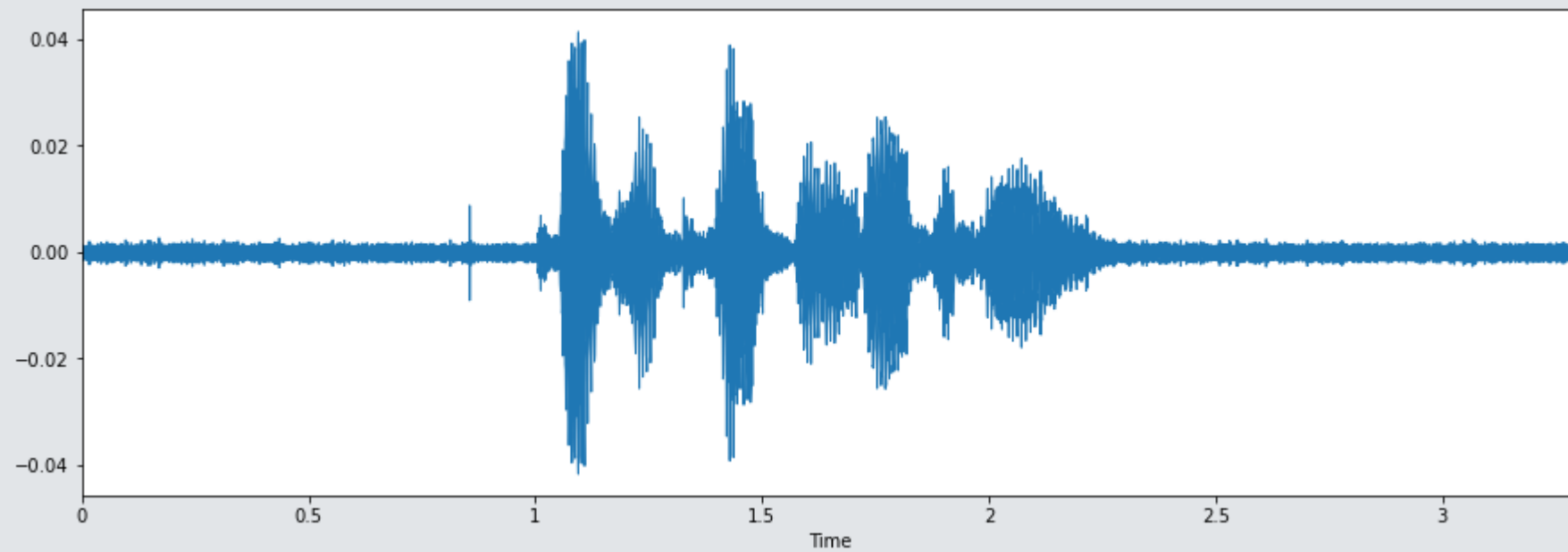
# Original

---

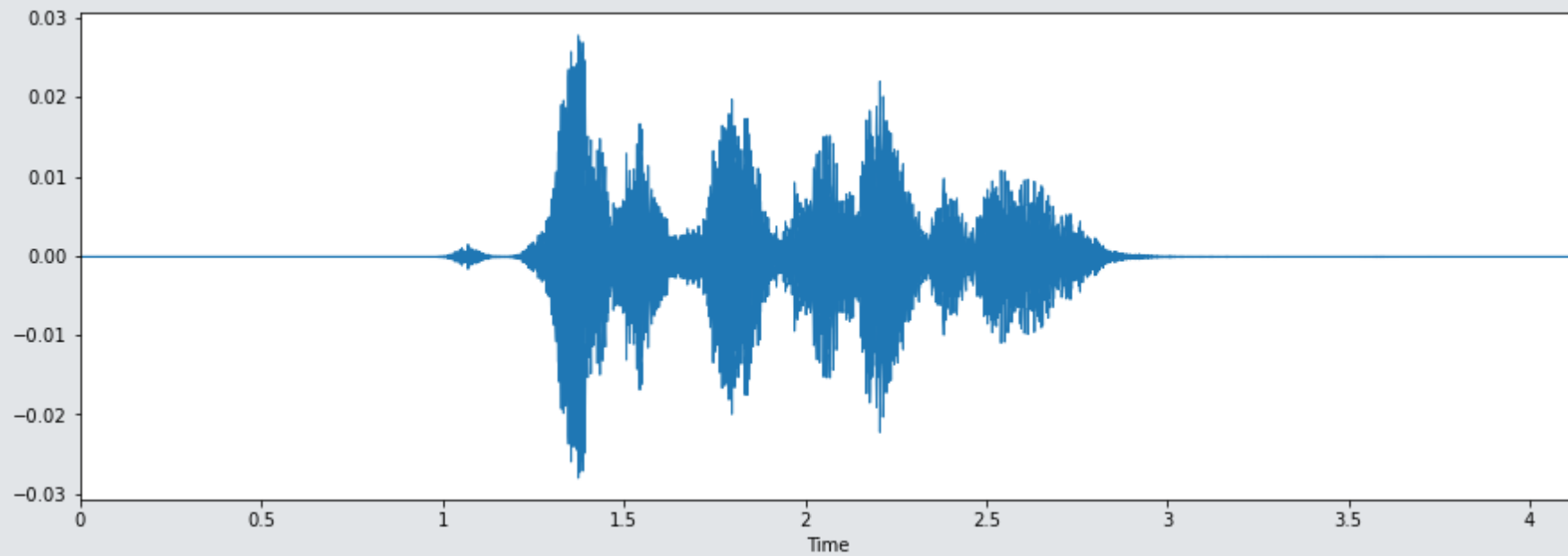


# Noise

---

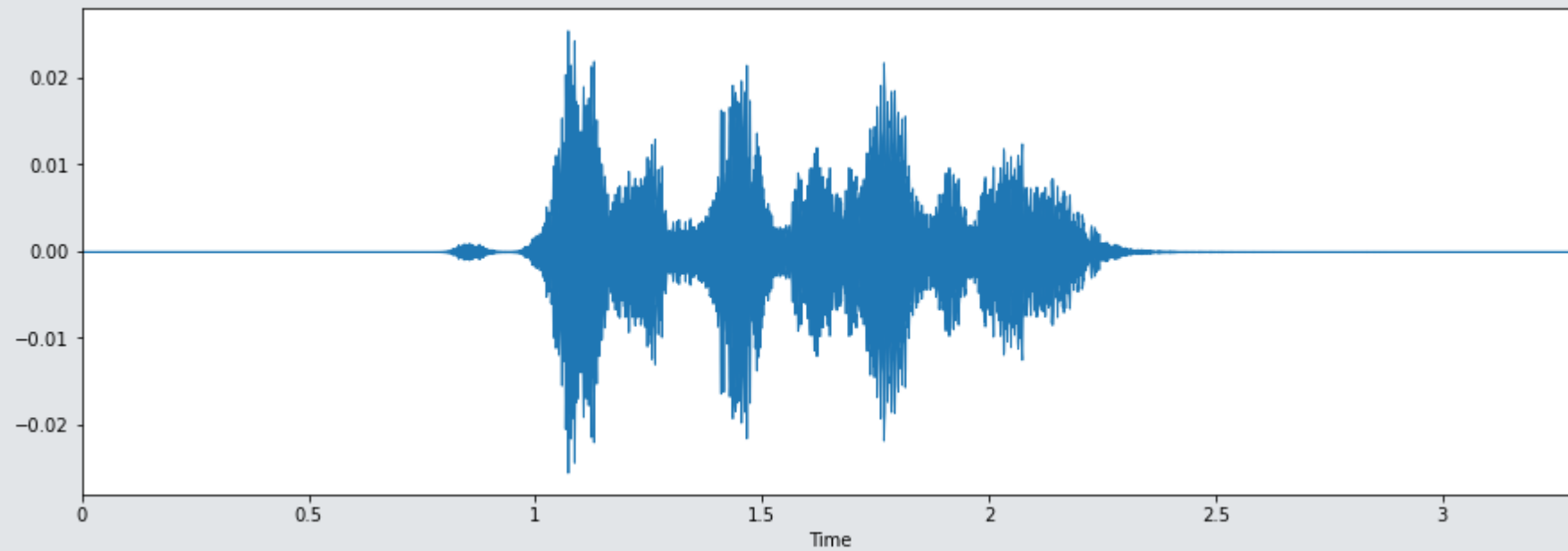


# Stretch



# Pitch

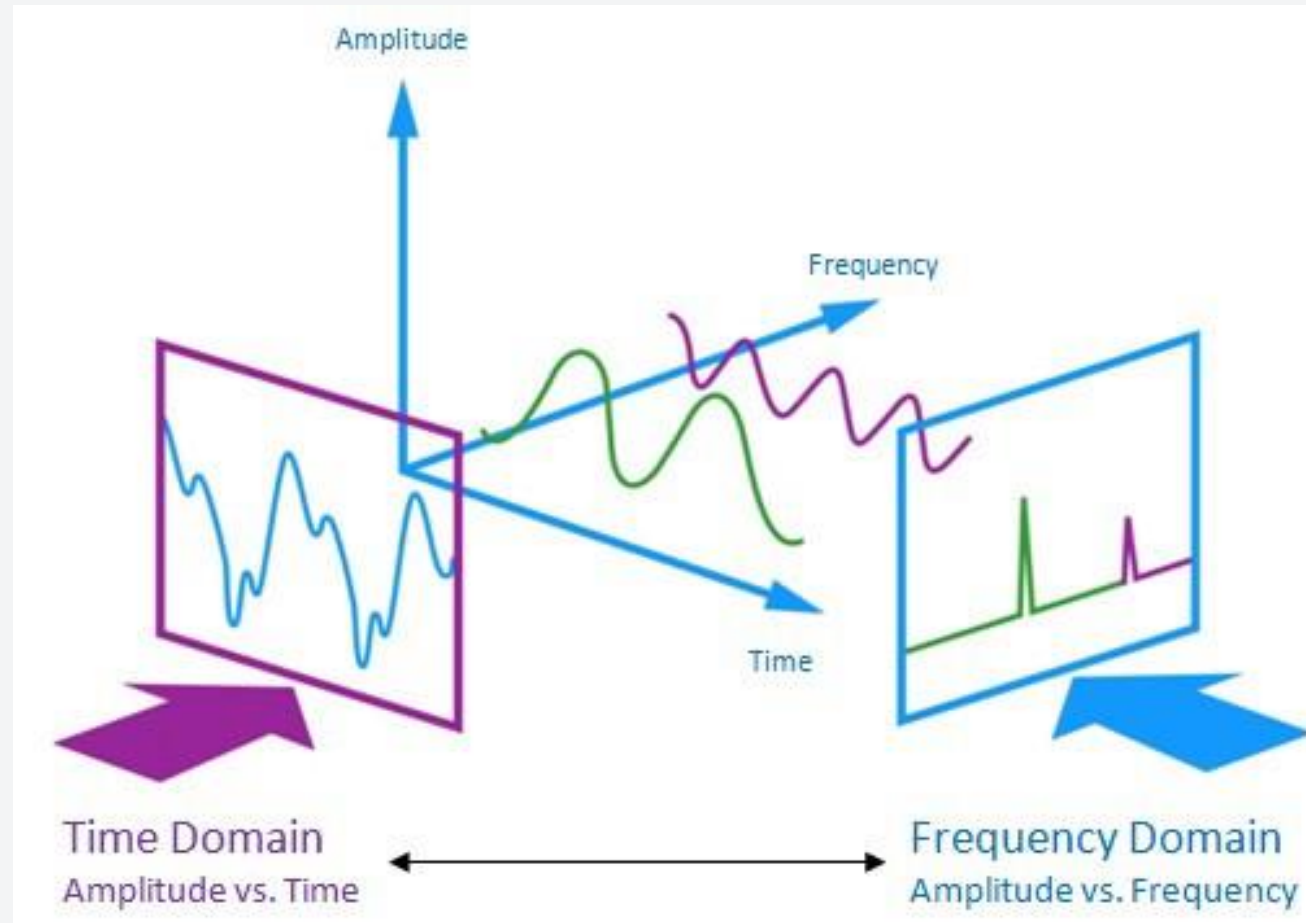
---







# **Feature Extraction**

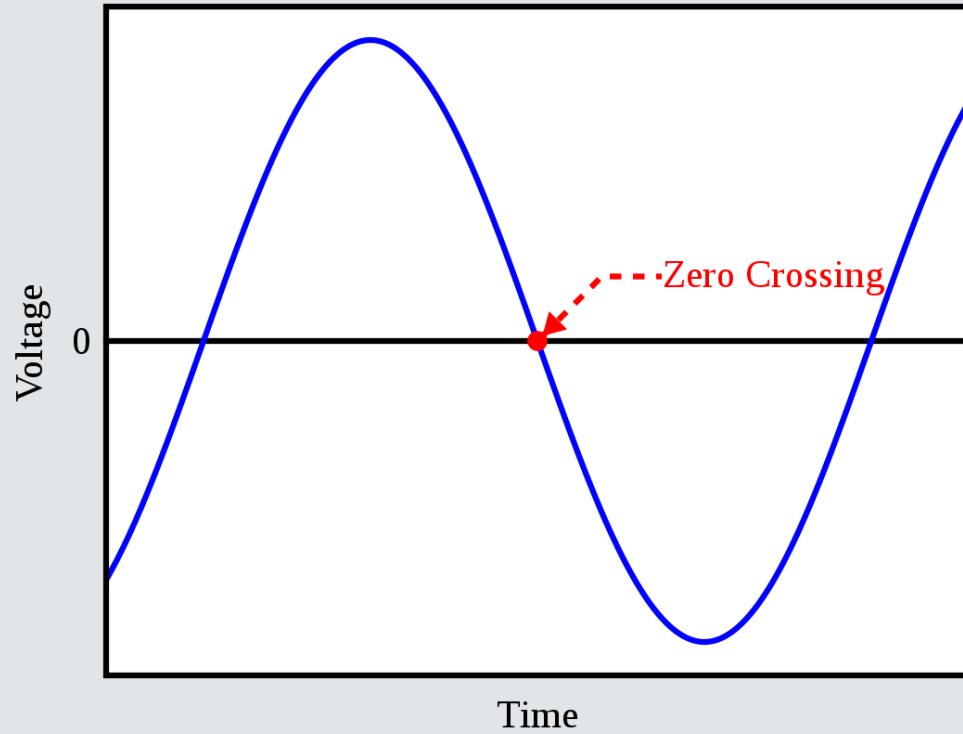


# Feature Extraction

---

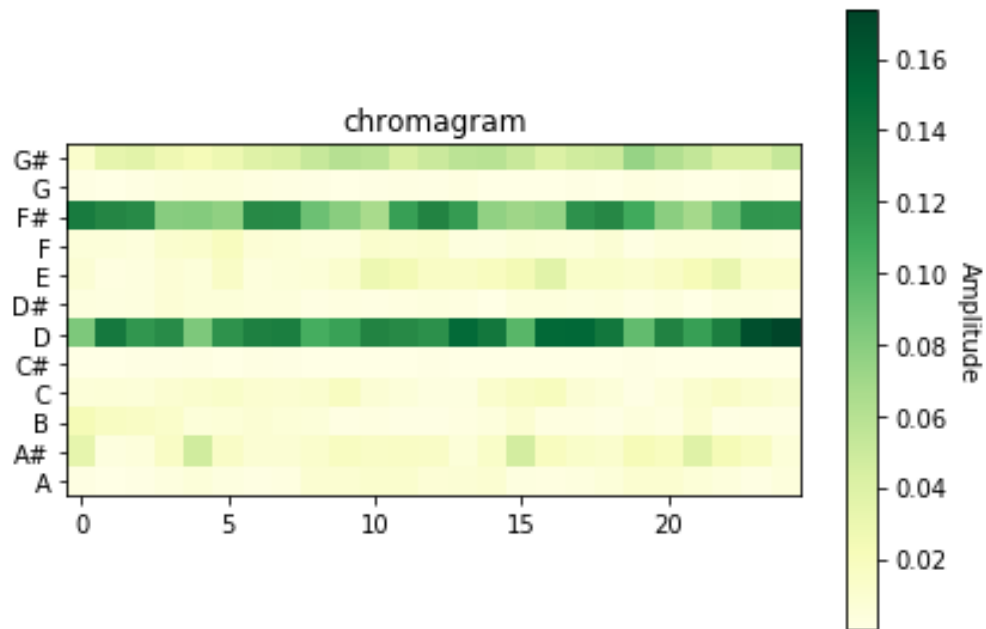
- Root Mean Square
- Zero Crossing Rate
- Chroma STFT
- MFCC
- Mel Spectrogram





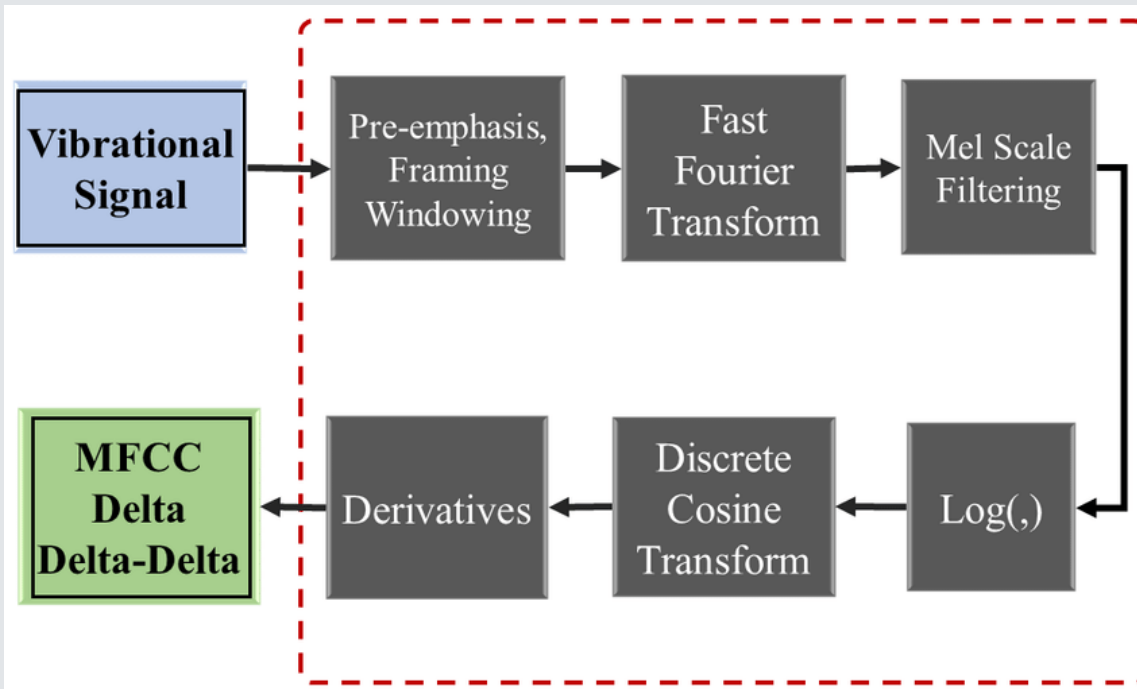
# Zero Crossing Rate

The rate of sign-changes of the signal during the duration of a particular frame.



# Chroma STFT

A 12-element representation of the spectral energy where the bins represent the 12 equal-tempered pitch classes of western-type music.

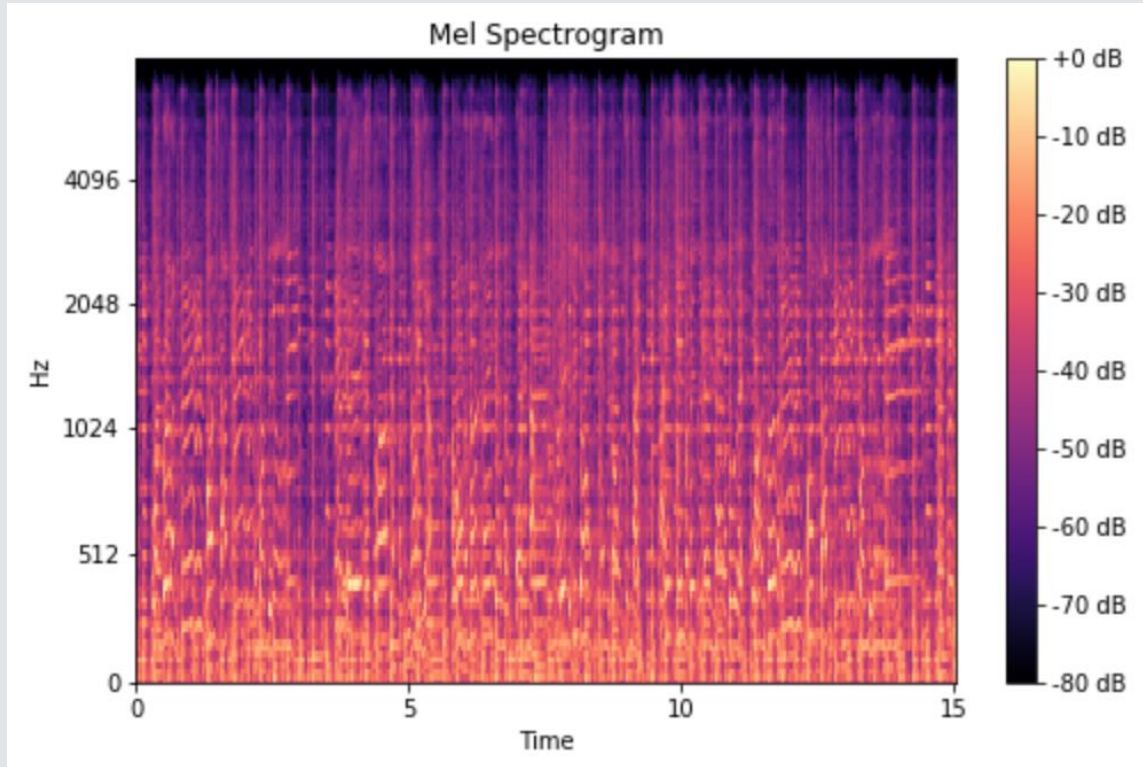


# MFCC

It scales the frequency in order to match more closely what the human ear can hear (humans are better at identifying small changes in speech at lower frequencies). It helps better represent phonemes (p, b, d..).

# Mel Spectrogram

In 1937, Stevens, Volkman, and Newman proposed a unit of pitch such that equal distances in pitch sounded equally distant to the listener. This is called the **mel scale**.



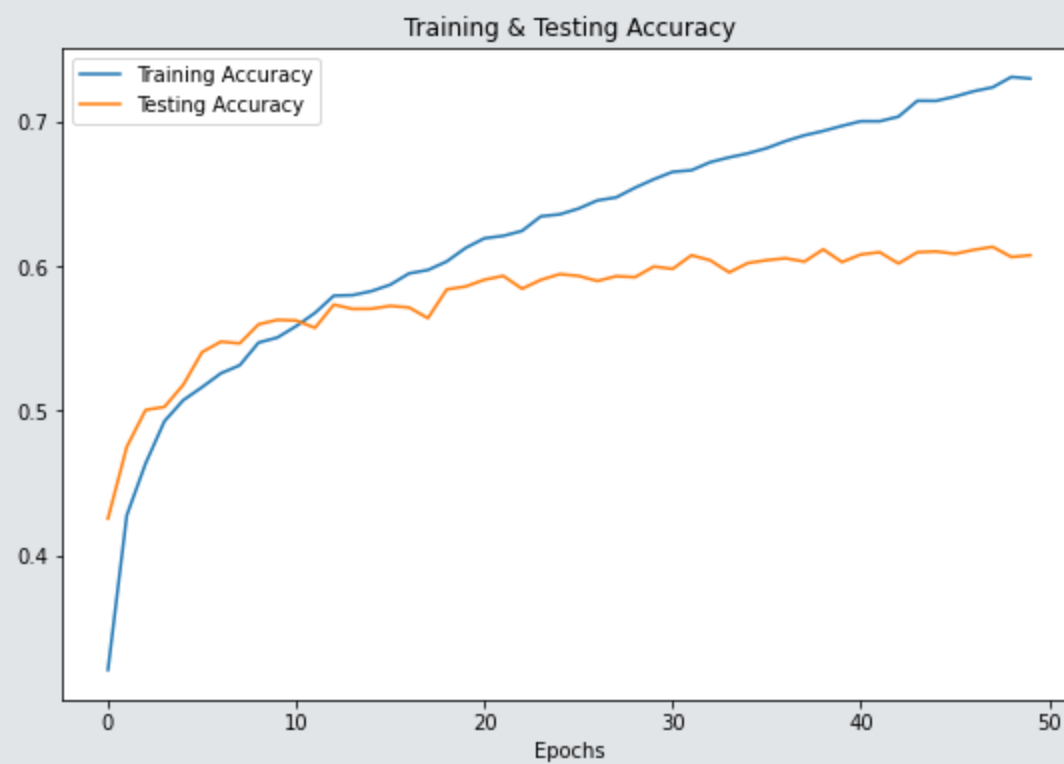
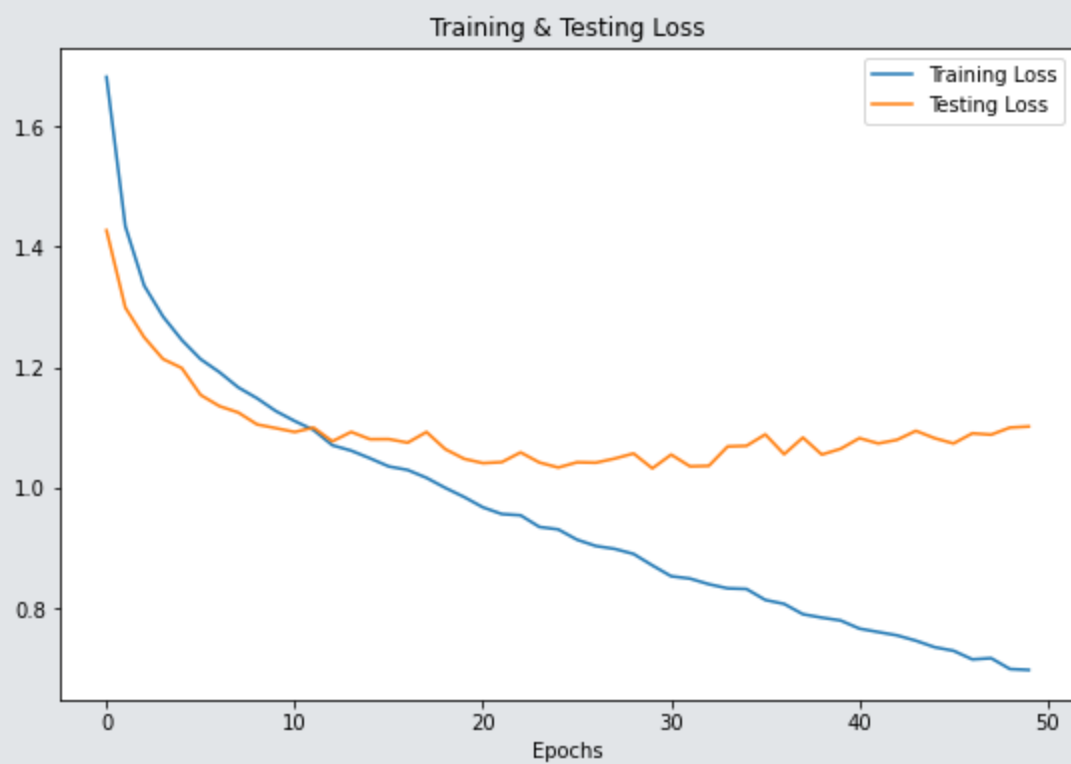
## 162 Features



# **Model Building**



Layer (type)	Output Shape	Param #
=====	=====	=====
conv1d_28 (Conv1D)	(None, 162, 256)	1536
max_pooling1d_28 (MaxPooling)	(None, 81, 256)	0
conv1d_29 (Conv1D)	(None, 81, 256)	327936
max_pooling1d_29 (MaxPooling)	(None, 41, 256)	0
conv1d_30 (Conv1D)	(None, 41, 128)	163968
max_pooling1d_30 (MaxPooling)	(None, 21, 128)	0
dropout_13 (Dropout)	(None, 21, 128)	0
conv1d_31 (Conv1D)	(None, 21, 64)	41024
max_pooling1d_31 (MaxPooling)	(None, 11, 64)	0
flatten_7 (Flatten)	(None, 704)	0
dense_13 (Dense)	(None, 32)	22560
dropout_14 (Dropout)	(None, 32)	0
dense_14 (Dense)	(None, 8)	264
=====	=====	=====



**Train Accuracy = 72.94%**

**Val Accuracy = 60.74%**

**F1 Score = 64%**

	Predicted Labels	Actual Labels
0	neutral	disgust
1	sad	sad
2	sad	sad
3	fear	disgust
4	happy	happy
5	sad	fear
6	disgust	sad
7	happy	happy
8	angry	happy
9	happy	happy

	precision	recall	f1-score
angry	0.78	0.69	0.73
calm	0.62	0.86	0.72
disgust	0.54	0.48	0.51
fear	0.63	0.51	0.57
happy	0.53	0.62	0.57
neutral	0.55	0.57	0.56
sad	0.58	0.68	0.62
surprise	0.85	0.79	0.82



# Conclusion

# Future Work

---

- **Explore other augmentation techniques**
- **Explore other feature extraction methods**
- **Incorporate real-time recognition**
- **Consider report generation**



**Thank You**