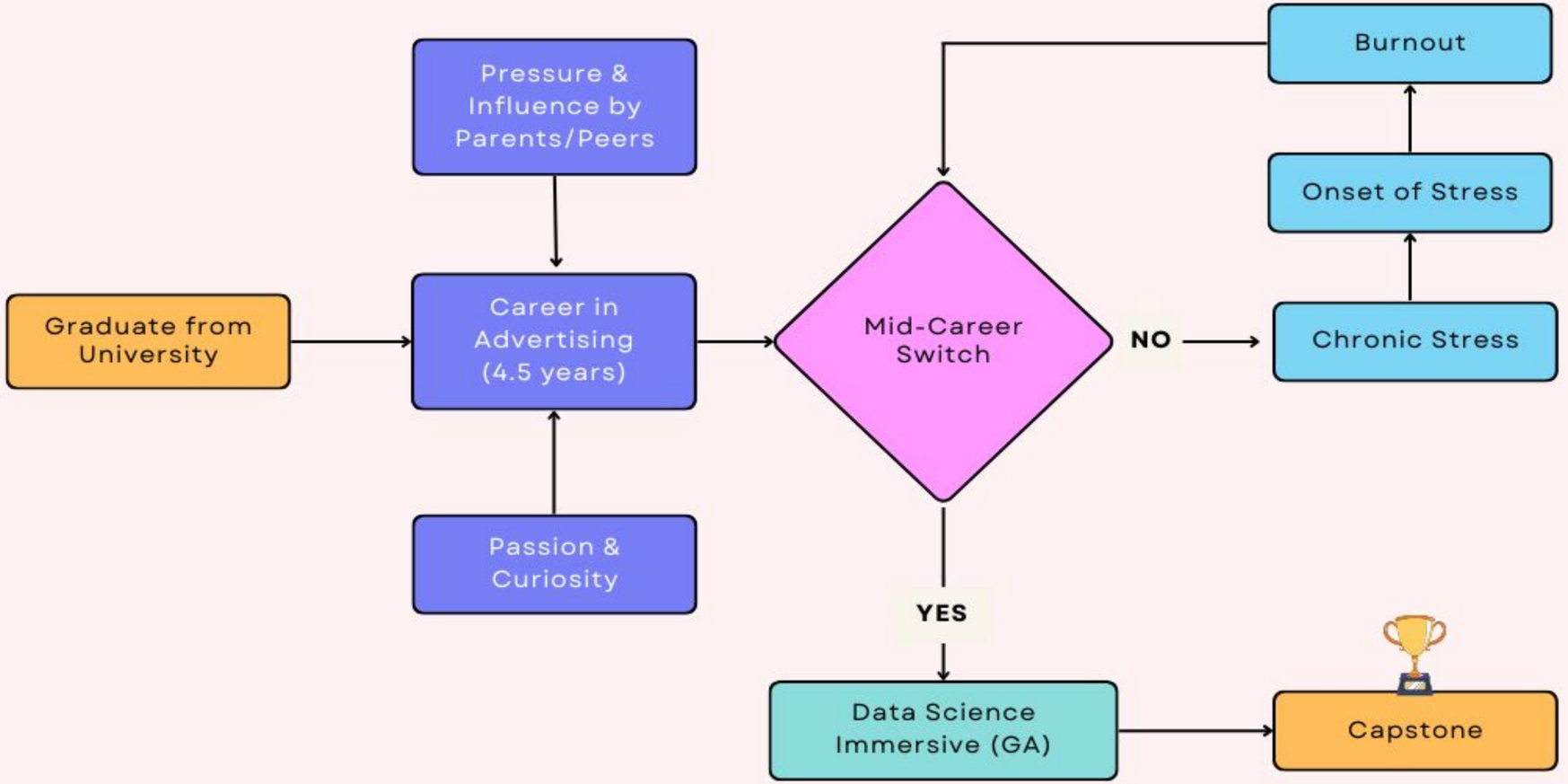
MY DSI JOURNEY



FeelFlow AI

Decoding Emotions, Advancing Patient Support

Capstone Presentation - DSI-SG-42

14th May 2024

Amoz Kuang

AGENDA

01 - Introduction

02 - Context + Problem Statement

03 - Methodology

04 - Insights

05 - Predictions

06 - Conclusion

07 - Demo

01 INTRODUCTION

SPEECH EMOTION RECOGNITION (SER)?

Speech Emotion Recognition (SER) in therapy refers to the use of technology to analyse and identify emotional states from a person's audial characteristics during speech. With the help of machine learning, this technology processes verbal cues such as tone, pitch, tempo, and volume can help to classify emotions such as anger, disgust, fear, happiness, neutral or sadness.

02 CONTEXT + PROBLEM STATEMENT

CONTEXT

Singapore

Prevalence of poor mental health increasing in Singapore; young adults have highest proportion at 25.3%

More Singapore residents are, however, willing to seek help, particularly from informal support networks, according to a survey by MOH.

More people in Singapore were willing to seek help for mental health issues in 2022. (Photo: iStock/Chaay_Tee)

Natasha Ganesan

27 Sep 2023 07:25PM
(Updated: 28 Sep 2023 03:42PM)

SINGAPORE: Mental health has worsened in Singapore but more are willing to seek help, a survey conducted by the Ministry of Health (MOH) has found.

The ministry's National Population Health Survey 2022, which was released on Wednesday (Sep 27), tracked the health, risk factors and lifestyle practices of Singapore residents aged 18 to 74, from July 2021 to June 2022.

Related Topics

MOH | mental health

ADVERTISEMENT

Data was collected from around 8,000 adults through self-reported household interviews and about 9,000 adults through health examinations.

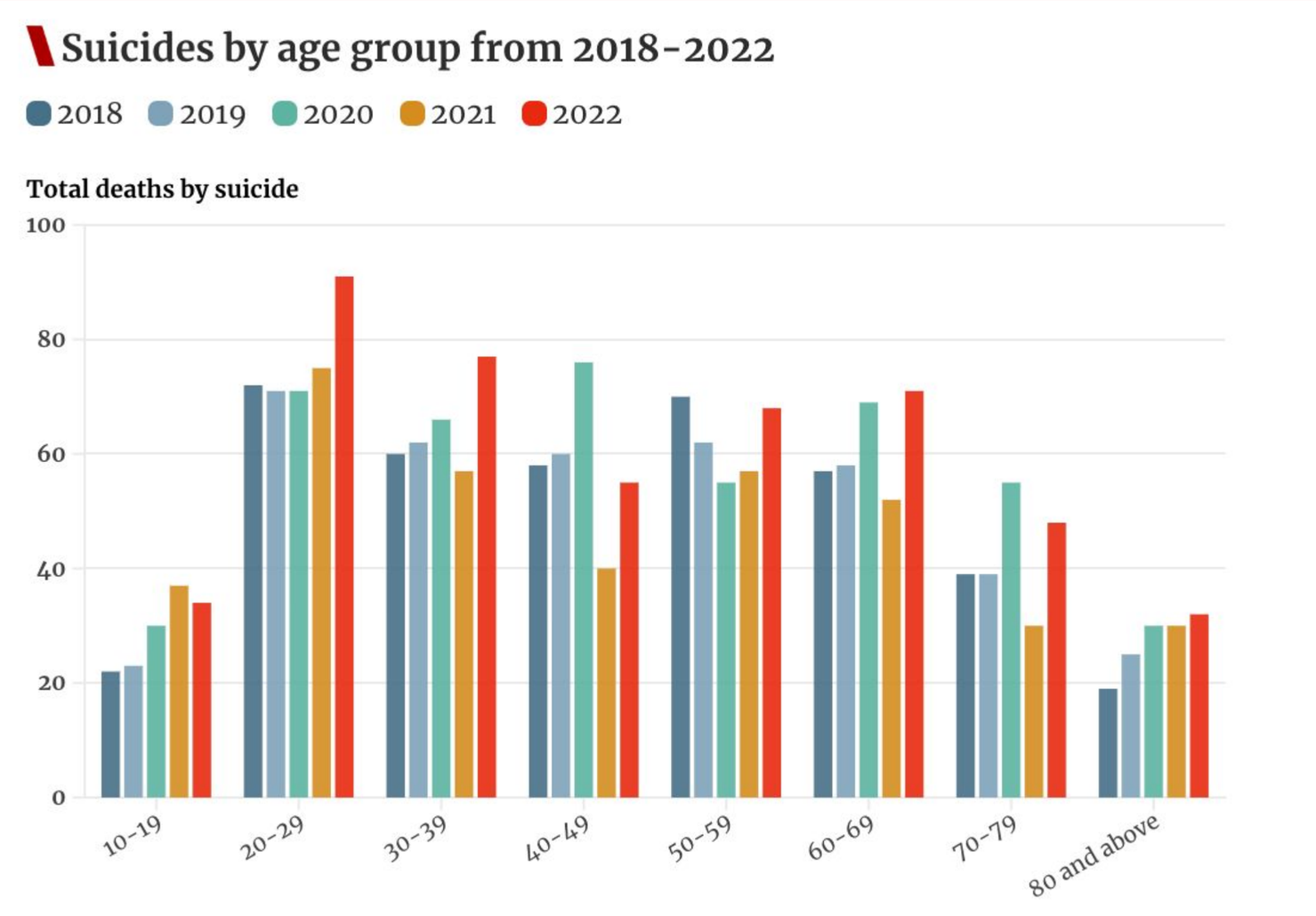
The survey findings showed that the prevalence of poor mental health increased from 13.4 per cent in 2020 to 17 per cent in 2022.

Younger adults aged 18 to 29 had the highest proportion of poor mental health at 25.3 per cent. The prevalence for other age groups was much lower, ranging from 10.5 per cent for those aged 60 to 74 to 19.4 per cent in those aged 30 to 39 years, results showed.

There were also more females (18.6 per cent) with poor mental health compared to males (15.2 per cent), according to the survey.

Sources: CNA, 2020, CNA, 2023

EFFECTS OF MENTAL HEALTH



Sources: CNA, 2023

CNA Insider

‘With school counsellors, it’s really hit-or-miss’: Behind the challenge of safeguarding student mental health

There is a disparity in student experiences when it comes to school counselling. (Photo: iStock)

Goh Chiew Tong
@ChiewTongCNA

Christy Yip
@ChristyYipCNA

01 Aug 2021 06:30AM
(Updated: 06 Sep 2021 01:51PM)

- Common complaints from students were that counsellors belittled problems or ‘snitched’ on them, but some had positive experiences.
- Teachers said that having one or two school counsellors is insufficient and does not offer choices to students.
- Teachers to receive enhanced training in mental health literacy, but some questioned whether the ‘burden’ should fall solely on them.
- Beyond improving ratios and the quality of care, an attitudinal shift may be needed.

** denotes name changed to protect the person’s identity*

SINGAPORE: When Jane* opened up to her secondary school counsellor about the cold wars and troubles she was having with some close friends, she was told that “life’s like that”.

Related Topics

Education | mental health | CNA Insider

"I felt ignored, like (school counselling is hopeless) even if I bring up my problems,"

"Jane* opened up to her secondary school counsellor about the cold wars and troubles she was having with some close friends, she was told that ‘life’s like that’."

Sources: CNA, 2021

THERE IS STILL DEMAND…

About 56.6 per cent of Singapore residents were willing to seek help from health professionals in 2022 - slightly lower than 58.3 per cent in 2021 but higher than 47.8 per cent in 2019.

Among the age groups, Singapore residents aged 60 to 74 years (48.1 per cent) were the least willing to seek help from healthcare professionals in 2022, while those aged 30 to 39 years (62 per cent) were the most willing.

Meanwhile, the proportion of residents willing to seek help from informal support networks such as friends and family rose to 79.7 per cent in 2022, up from 69.1 per cent in 2021 and 74.5 per cent in 2019.

Younger adults aged 18 to 29 years were most willing to seek help from these networks (88.1 per cent), while older adults aged 60 to 74 years were least willing (68.4 per cent).

"The earliest appointment they could give me was two months later. But I felt I needed help now."

CNA Insider

More youths seeking help with mental health - but finding it isn't always easy

Breaking the silence, youths share candidly with CNA Insider about their struggles with depression, self-harm and trauma – and about persevering on the sometimes-bumpy road to finding help and healing.

They struggled but managed to find help and people who care.

Neo Chai Chin

Goh Chiew Tong
@ChiewTongCNA

Christy Yip
@ChristyYipCNA

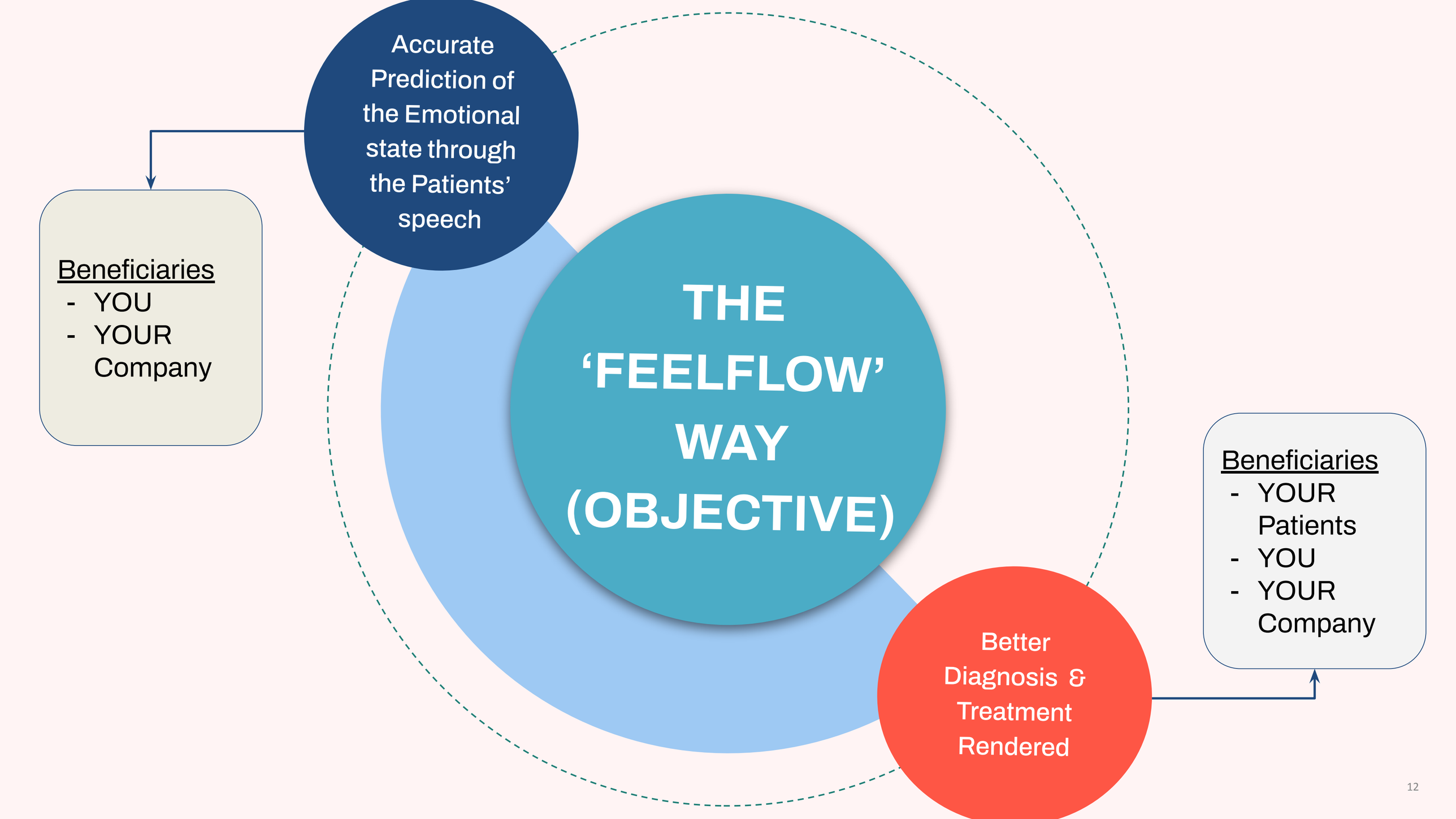
Ryan Tan Bing Yang

01 May 2022 06:15AM
(Updated: 02 May 2022 03:26PM)

Sources: CNA, 2022, CNA, 2023

PROBLEM STATEMENT

"Where discerning people’s emotion can sometimes be an unnerving guessing game, how can clinicians use **Speech Emotion Recognition** technology to accurately assess patients' emotional well-being, thereby improving diagnosis and treatment outcomes?"


Accurate Prediction of the Emotional state through the Patients' speech
Beneficiaries
- YOU
- YOUR Company
THE 'FEELFLOW' WAY (OBJECTIVE)
Better Diagnosis & Treatment Rendered
Beneficiaries
- YOUR Patients
- YOU
- YOUR Company

PERSONA

Thomas Chew, 55,
Senior Counsellor @ Samaritans of Singapore

Attitudes & Behaviours

- Recent portfolio focus on GenZ and millennials, facing issues like work/school/relationship-related burnout, anxiety, depression.

Pain points

- Does not have a kid. Has been out of touch with youth lingo, behaviours, knacks, and feels like it might be hard to reach his audience.

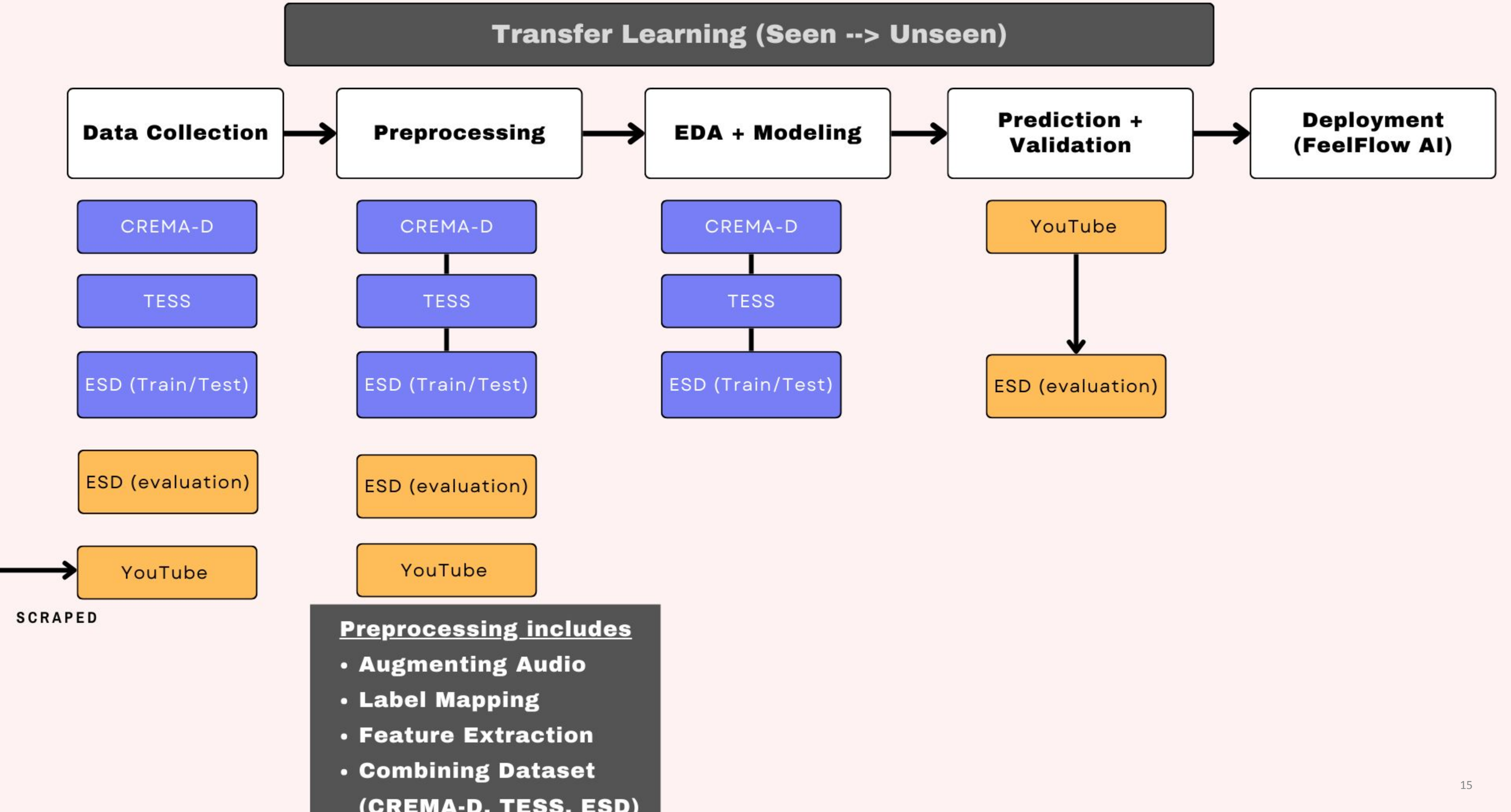
Scenario

- He heard that a teen that he recently counseled has displayed signs of depression, although during the sessions he seem perfectly normal

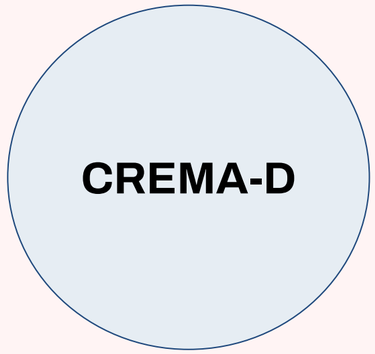
Motivation

- Embraces technology as a hobby, keeping up with the latest tech gadgets and news
- Is passionate about AI and tech in the MediTech space

03 METHODOLOGY


Transfer Learning (Seen --> Unseen)
Data Collection
Preprocessing
EDA + Modeling
Prediction + Validation
Deployment (FeelFlow AI)
CREMA-D
TESS
ESD (Train/Test)
ESD (evaluation)
YouTube
SCRAPED
CREMA-D
TESS
ESD (Train/Test)
ESD (evaluation)
YouTube
CREMA-D
TESS
ESD (Train/Test)
YouTube
ESD (evaluation)
Preprocessing includes
• Augmenting Audio
• Label Mapping
• Feature Extraction
• Combining Dataset (CREMA-D, TESS, ESD)

ABOUT THE DATA (SEEN)



Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D)

Data includes:
- 7,442 audio files
- 91 actors
- 6 emotions

(neutral, happy, angry, sad, disgust, and fear)

Toronto Emotion Speech Set (TESS)

Data includes 200 target words:
- 2800 audio files
- 7 emotions

(neutral, happy, angry, sad, disgust, fear, and pleasant surprise)

Emotion Speech Dataset (ESD)

Data includes 350 parallel utterances:
- 10 native Mandarin speakers, 10 English speakers
- 5 emotions

(neutral, happy, angry, sad, and surprised)
- Data is split into **training, test**, and evaluation

ABOUT THE DATA (UNSEEN)

Emotion Speech Dataset (ESD)

Data includes 350 parallel utterances:

- 10 native Mandarin speakers, 10 English speakers
- 5 emotions

(neutral, happy, angry, sad, and surprised)

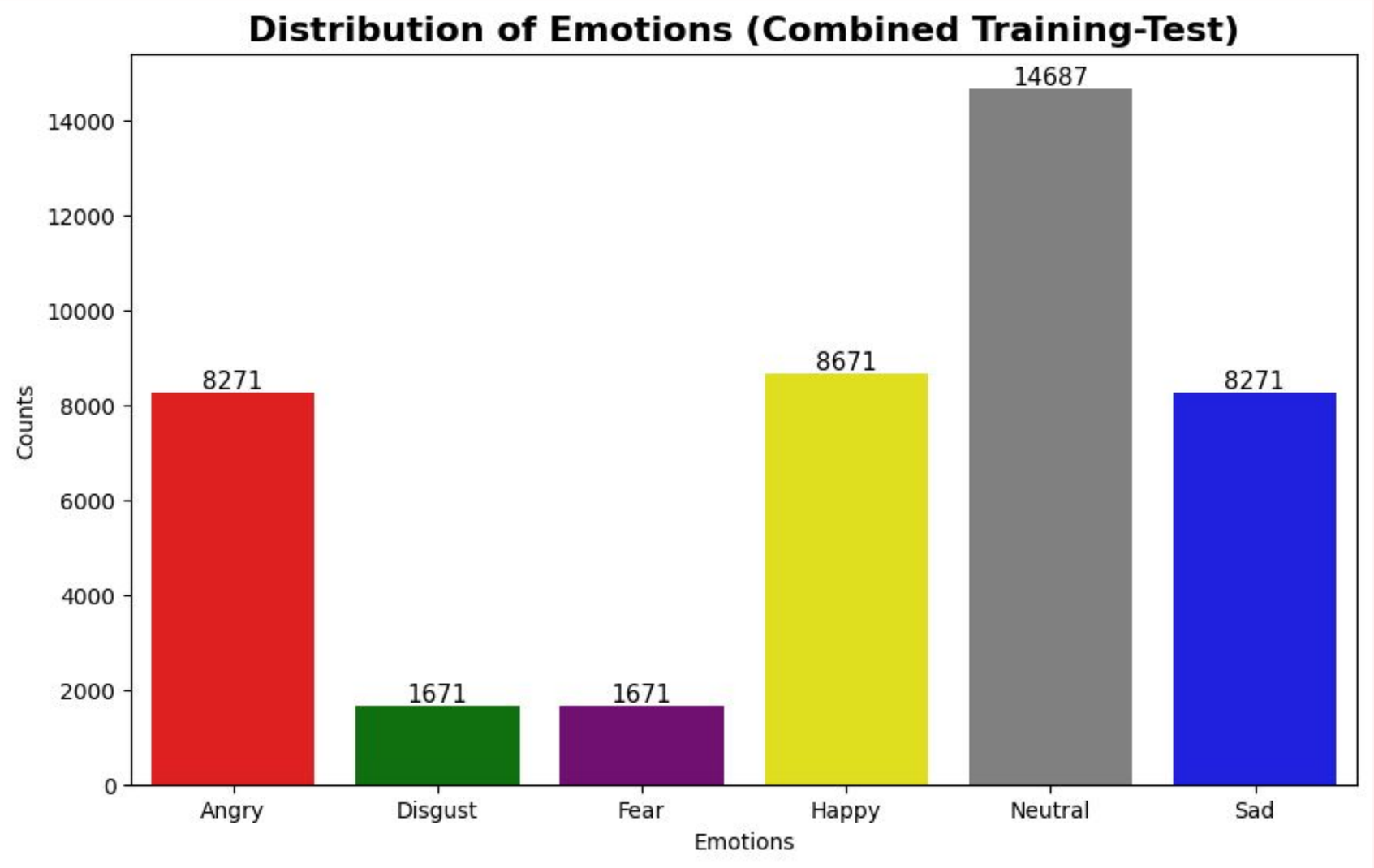
- Data is split into training, test, and **evaluation**

YouTube Data

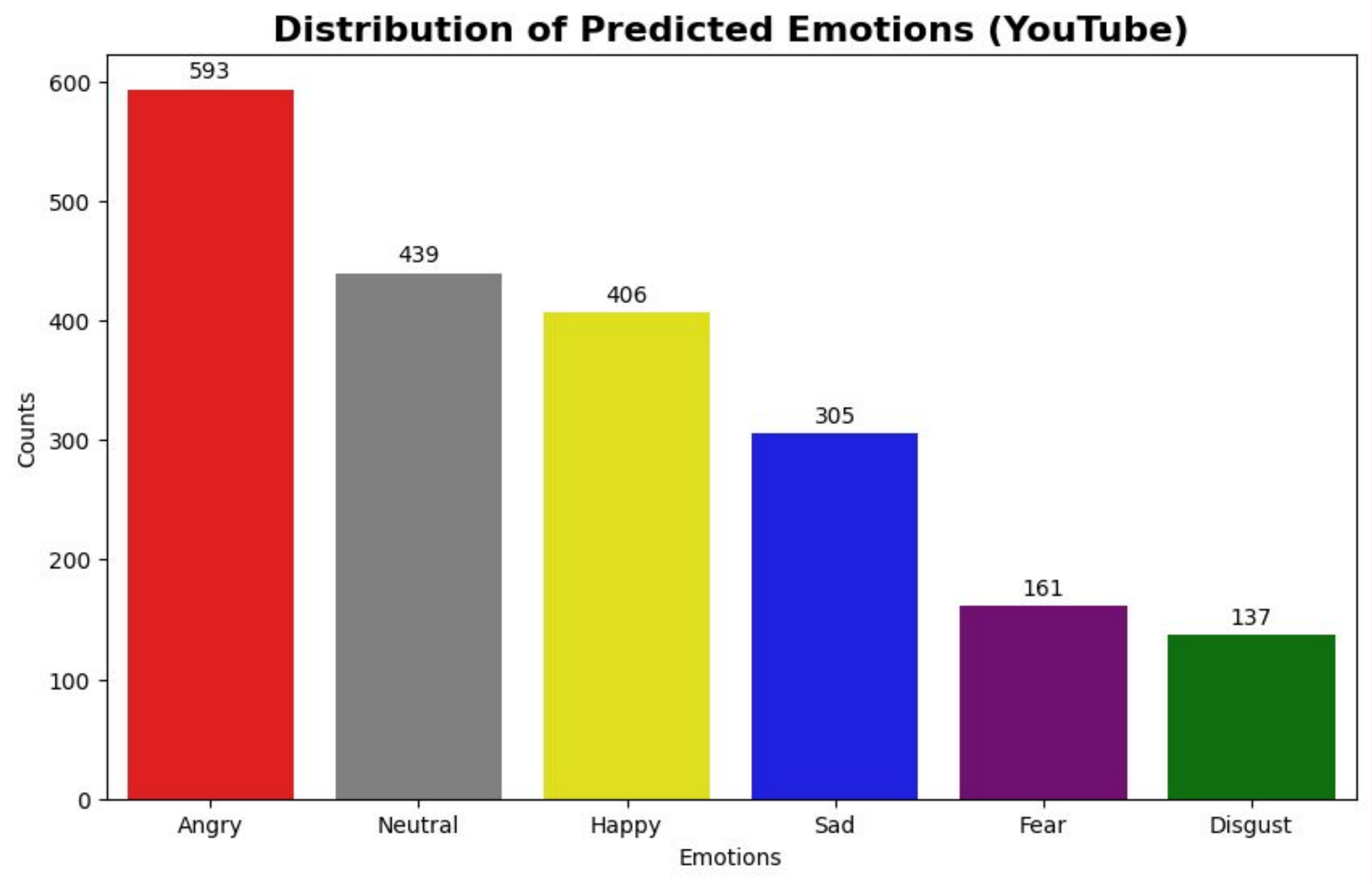
(scraped and unlabeled from 8 different sources - Podcasts, Docuseries, Documentaries etc.)

- Links found in Notebook

04 INSIGHTS

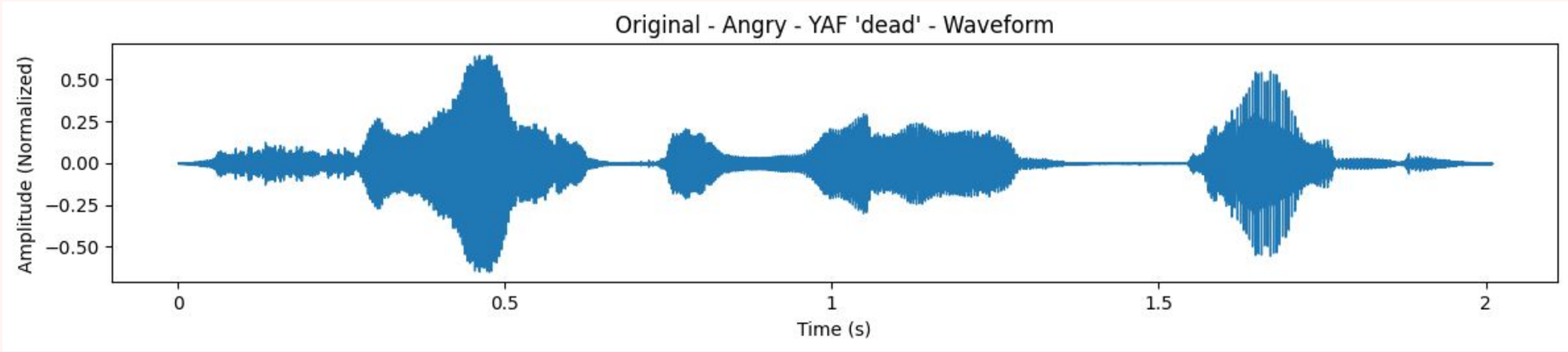


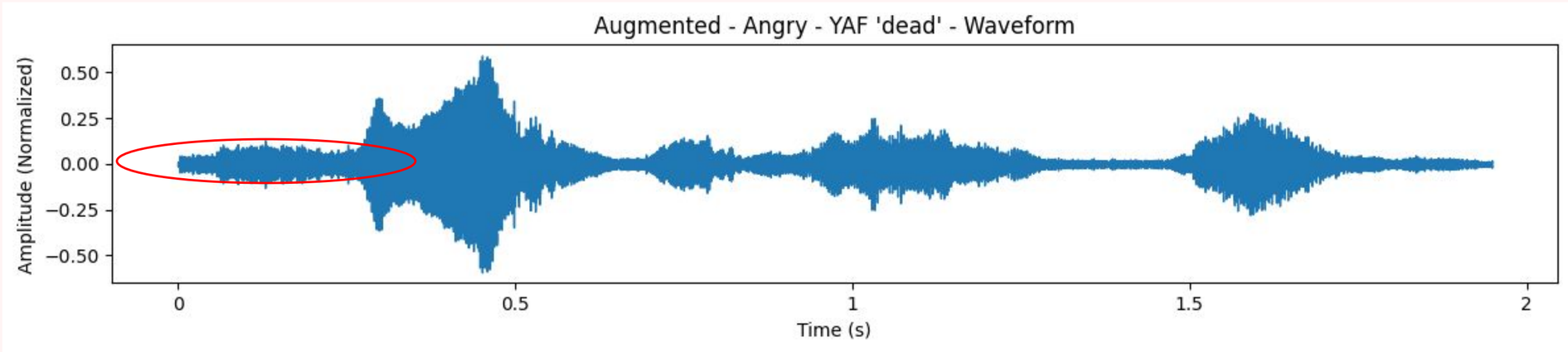
- ‘Disgust’ and ‘Fear’ are the minority class in the training-test dataset
- Rationale:
  - ‘Surprise’ emotion in ESD - mapped as ‘Neutral’ (~6.6k)



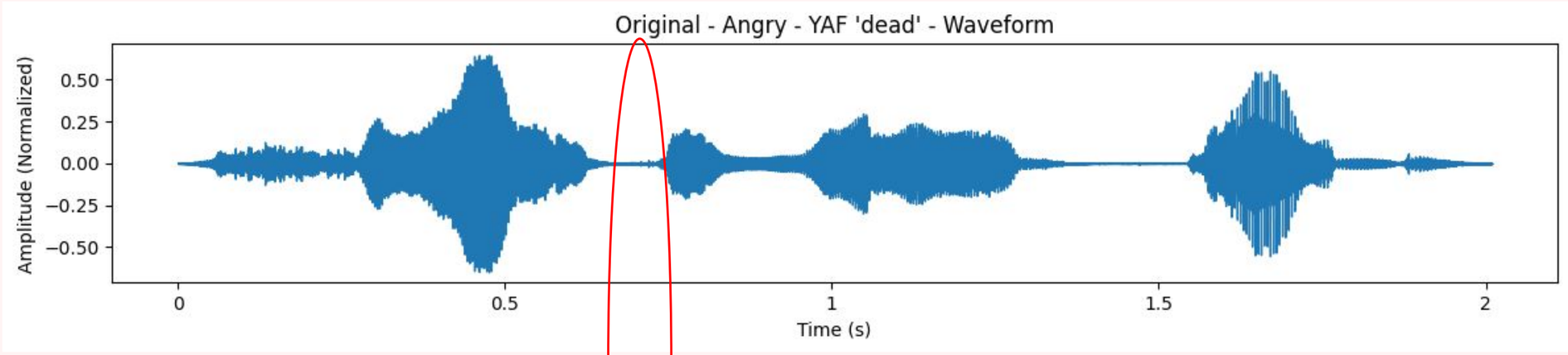
- Expected to have ‘Fear’ and ‘Disgust’ as the least distributed predicted emotions due to the imbalance trained and tested dataset - even after more class weights importance
- Overall, more balanced prediction than the train-test data
- Further validation is required from ESD (eval) dataset

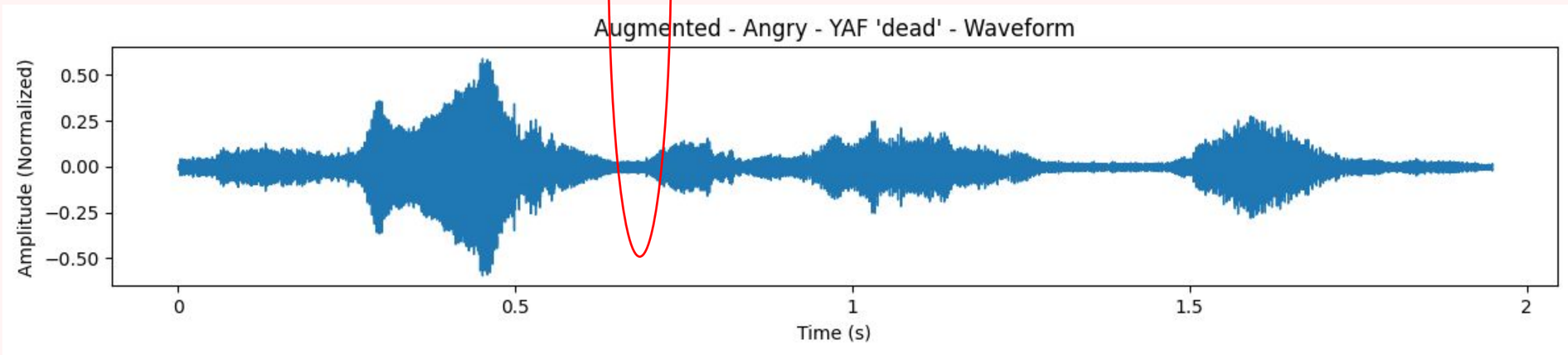
EFFECTS OF AUGMENTATION - ANGRY (NOISE ADDITION)



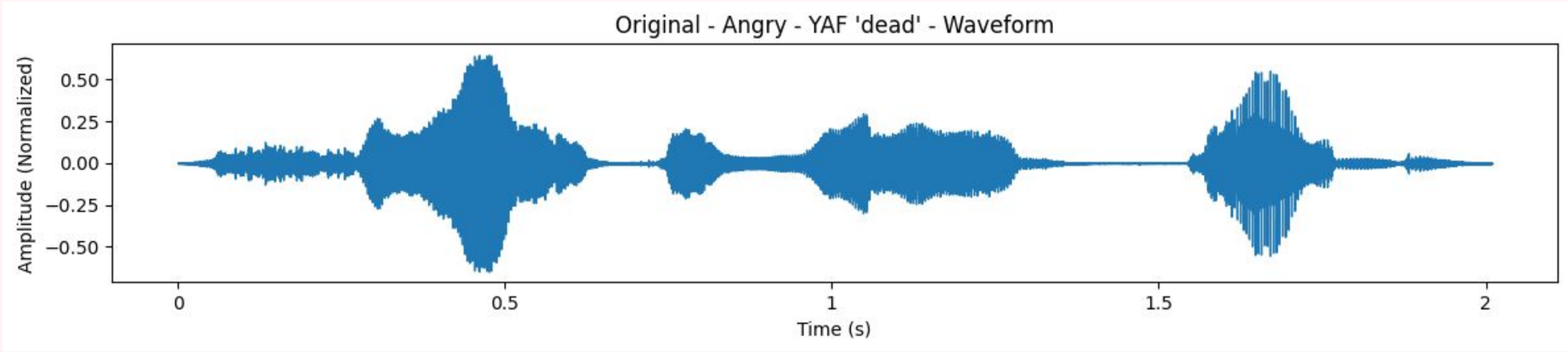


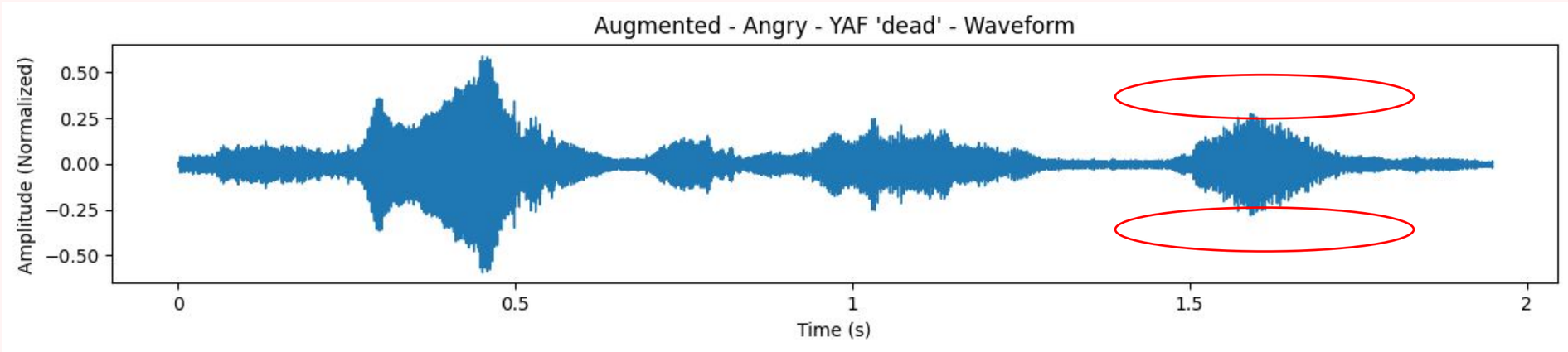
EFFECTS OF AUGMENTATION - ANGRY (TIME SHIFT)



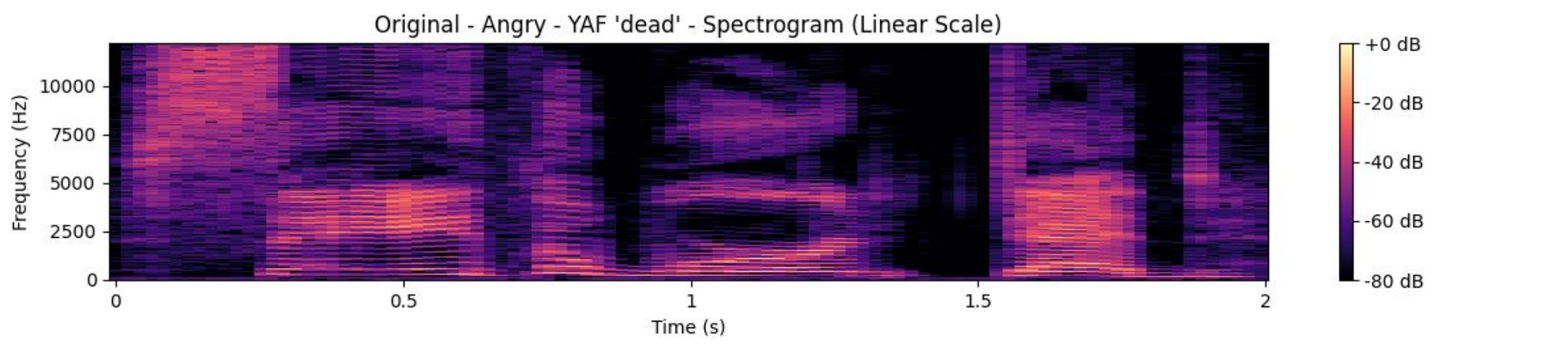


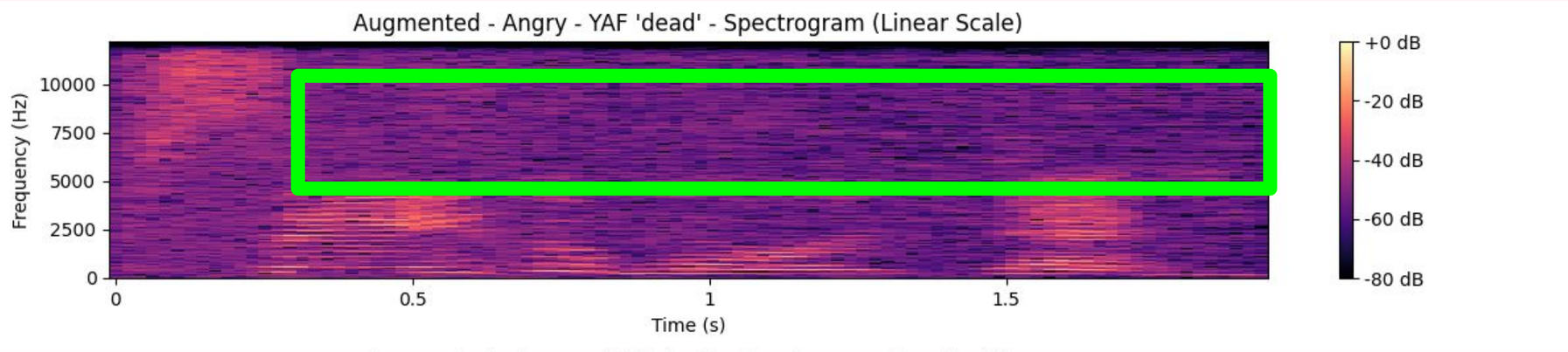
EFFECTS OF AUGMENTATION - ANGRY (PITCH SHIFTING)





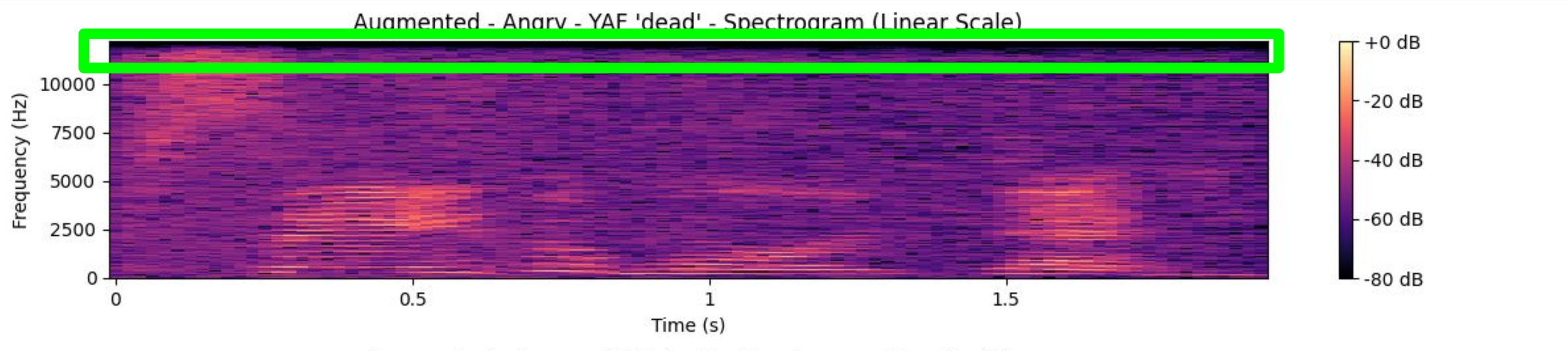
LINEAR SPECTROGRAM - ANGRY





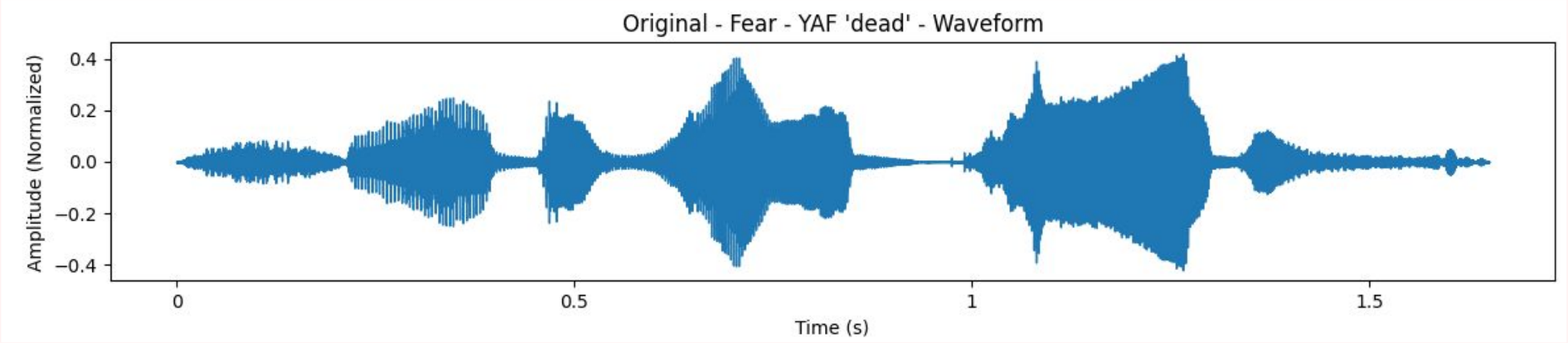
- “Blendedness” is referred to as smoothening of the pitch
- This makes the model less “sensitive” to spikes in volume and/or pitch especially when someone is angry

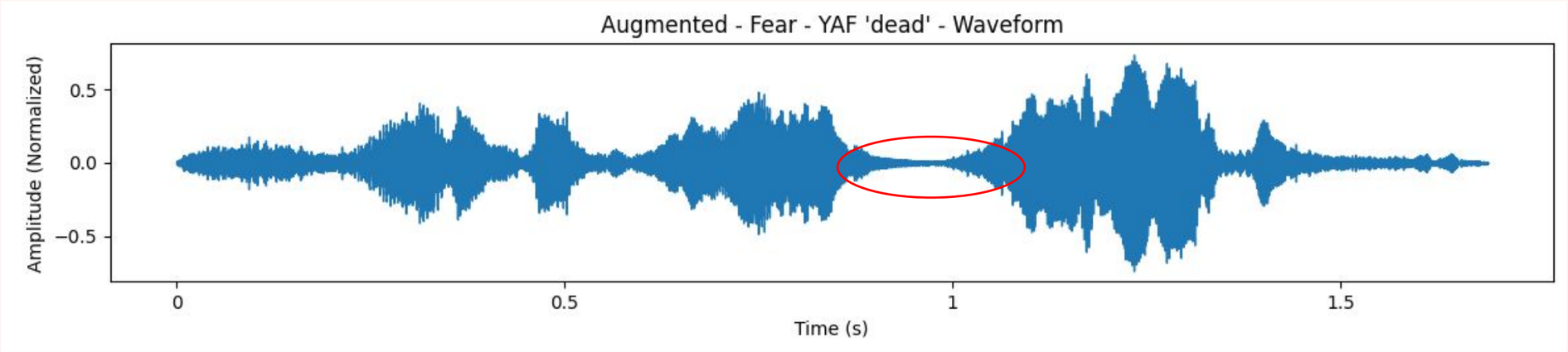
LINEAR SPECTROGRAM - ANGRY



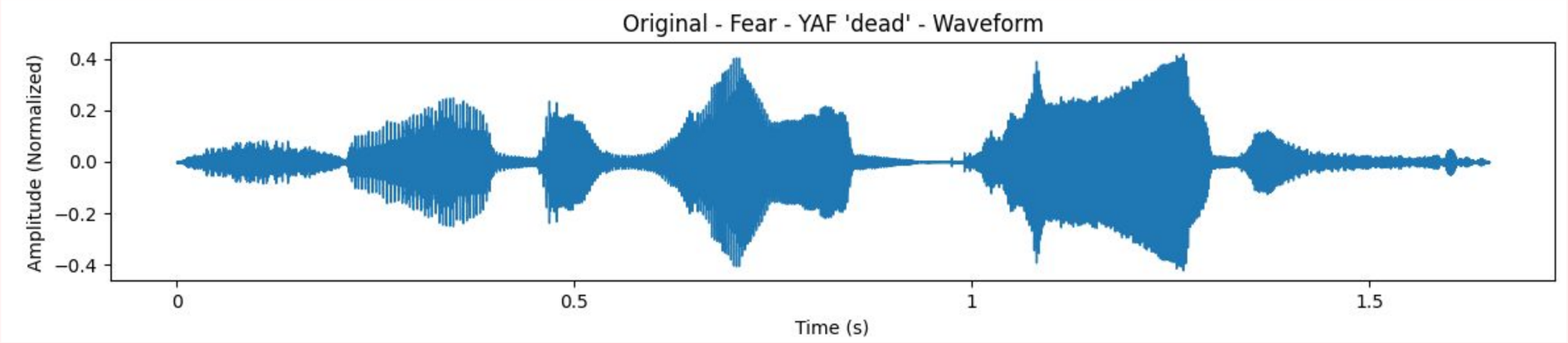
- Padding:
  - Consistency in duration of each audio data
  - Model Prep - especially for CNN + RNN-type models
  - Allows for real-time processing (buffering the audio stream)

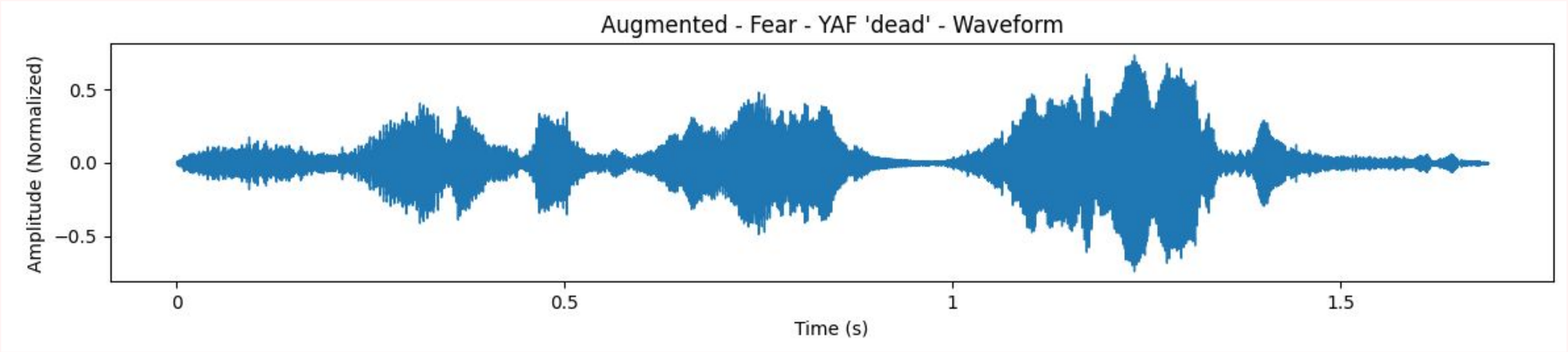
EFFECTS OF AUGMENTATION - FEAR (NOISE ADDITION)



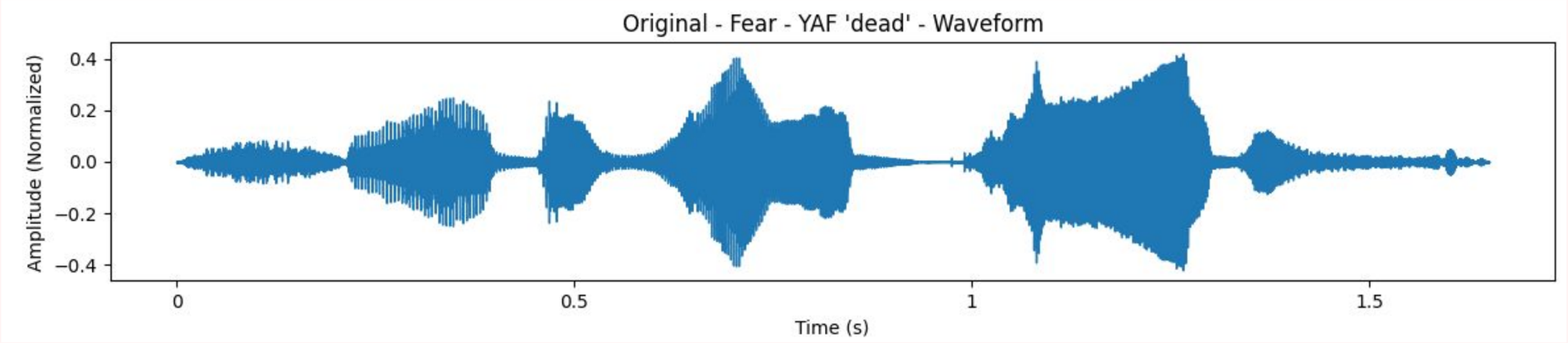


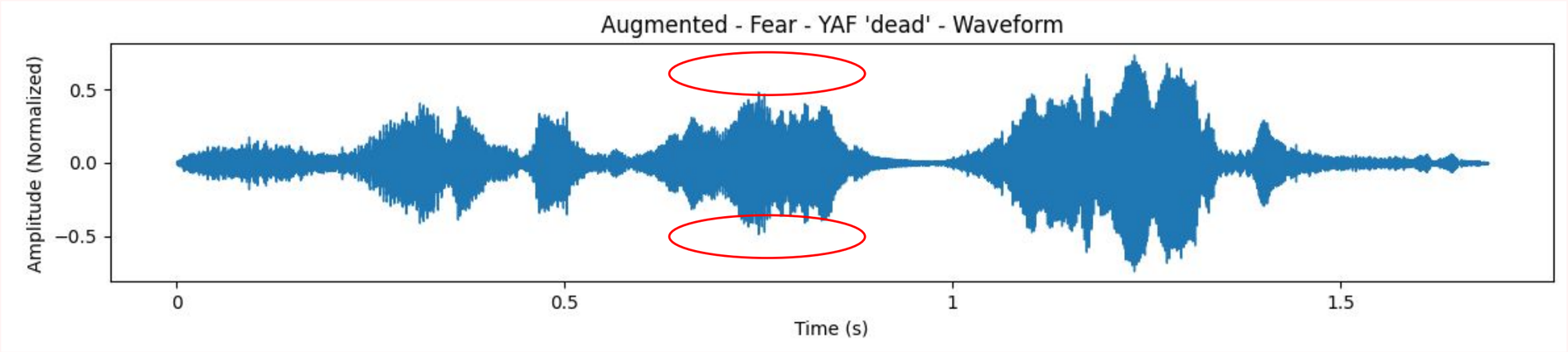
EFFECTS OF AUGMENTATION - FEAR (TIME SHIFT/STRETCH)



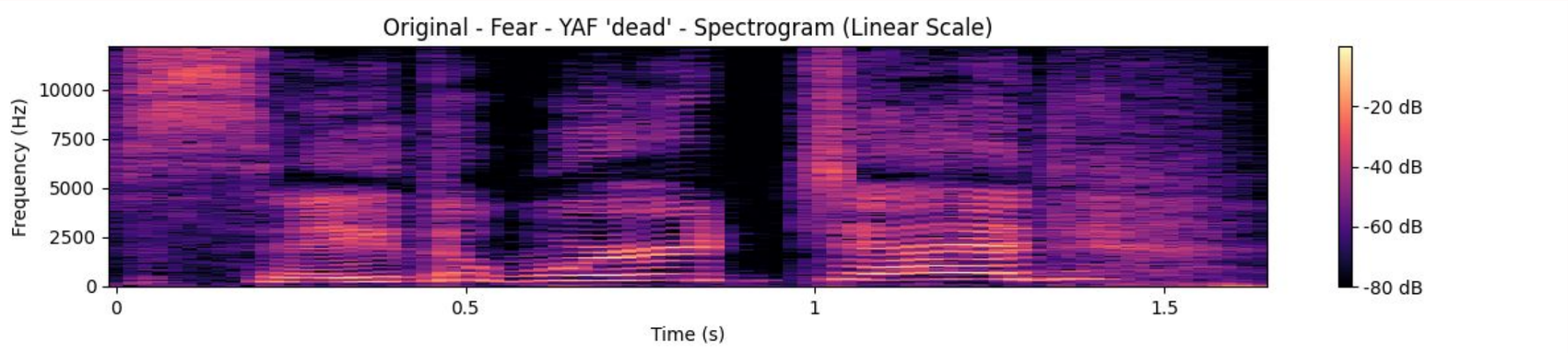


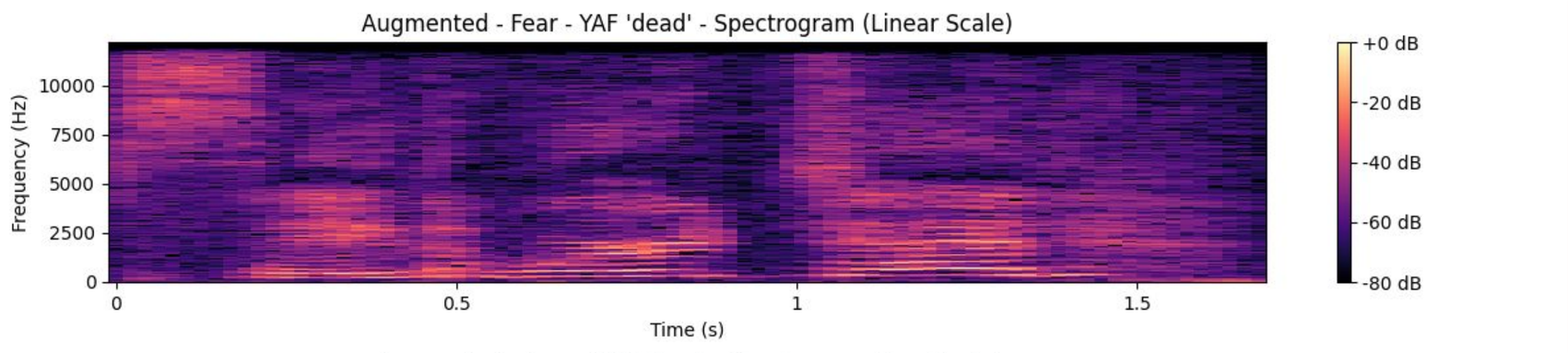
EFFECTS OF AUGMENTATION - FEAR (PITCH SHIFT)





LINEAR SPECTROGRAM - FEAR





- Augmentation here has little to no effect as compared to ‘Angry’ and ‘Disgust’ emotions.

05 PREDICTIONS

MODEL EVALUATION

Model Name	Train Accuracy	Validation Accuracy	Processing Time (sec)
Random Forest Classifier	1.0000	0.6167	33
Multilayer Perceptron Classifier	0.7019	0.6465	4038
LSTM + 1D CNN	0.5832	0.5412	1260
LSTM + 2D CNN	0.4007	0.4159	46
(Modified) WaveNet	0.6539	0.6168	1016

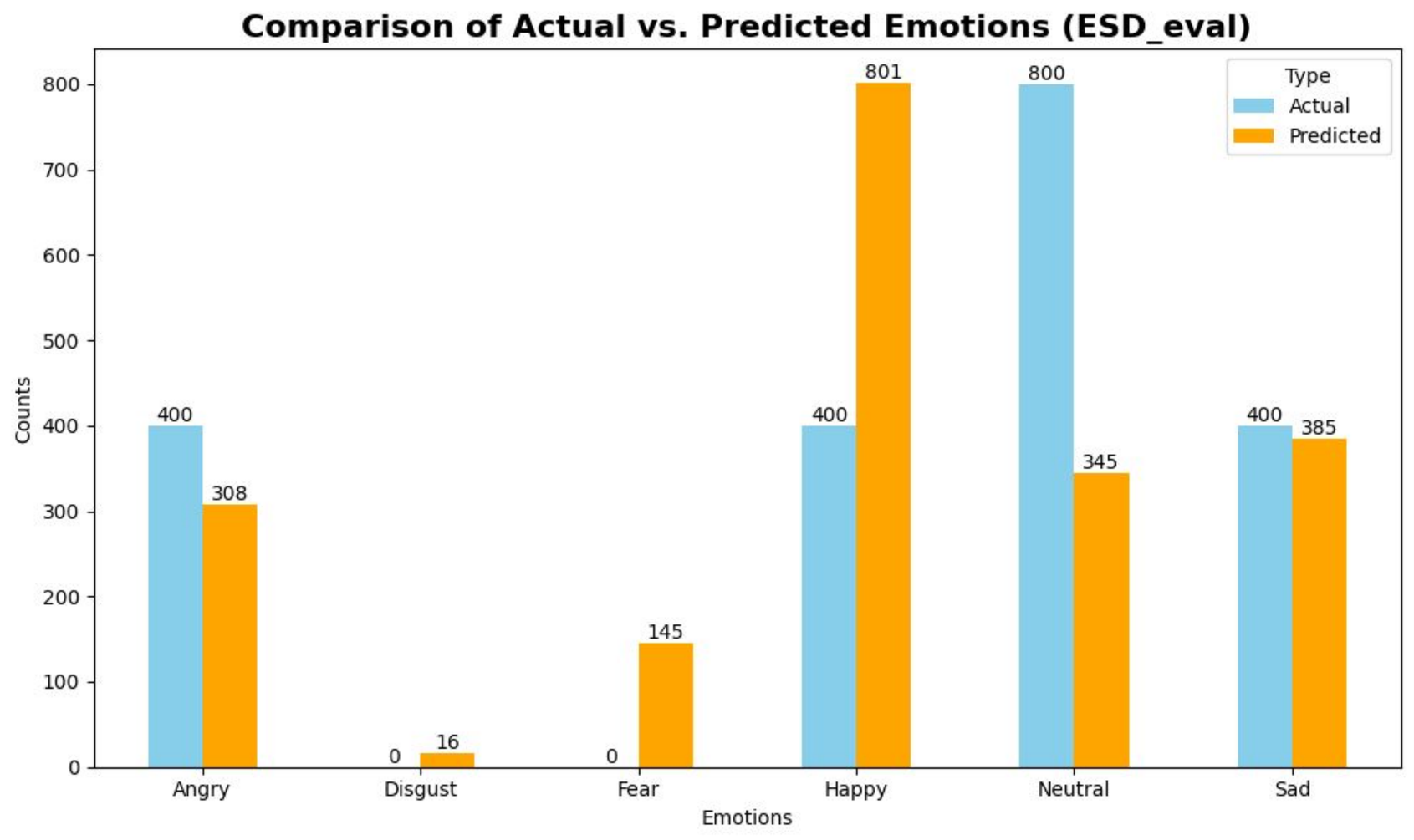
Analysis

- WaveNet as the champion model, no overfit, nor underfit. Generalises well.
- Allows for training of real-time data - which the study aims to explore in the next Phase(s)

DIFFERENCE IN WAVENET MODELS USED

Parameters / Models	Standard WaveNet	Modified WaveNet
Dilated Convolution	Different Temporal Resolution (1, 2, 4, 8 etc.)	[1, 2, 4, 8, 16, 32] Maximum Expansion for broader audio contexts
Skip Connections	SAME	
Activation & Normalisation	Gated Activation (tanh + sigmoid)	ReLU (more simple for quicker training time)
Output Processing	ReLU > 1x1 Con Layer > ReLU > 1x1 Con Layer	Skip Connection > ReLU > 1x1 Con Layer
Loss Functions & Optimisation	Softmax + Logits	Sparse Categorical Crossentropy (with Adam optimiser)
Training Configuration	Depends on application, not explicit	Dropout (0.2), Early Stopping (=5) with Custom Class weights

VALIDATION (ESD_EVAL)



- Negative Emotions
  - Overprediction on ‘Disgust’ and ‘Fear’, which may unnecessarily trigger attention required.
  - Slight underprediction for ‘Angry’ and ‘Sad’, may require more training of data, and clinicians to administer personal discernment
- ‘Happy’ - overprediction, which may delivered false reassurance to the clinicians.

06 CONCLUSION

COST BENEFIT ANALYSIS (MICRO)

Thomas Chew

Cost

- No monetary costs to him

Benefit

- Sessions per Day: ~ 5 sessions a day.
- Daily Savings: 20% time savings = 1 extra session/day (+$150-200)
  - Annual Savings: Assuming 200 working days/year, the additional revenue or value from enhanced efficiency could range from $30,000 to $40,000 annually.

COST BENEFIT ANALYSIS (MICRO)

Thomas Chew

Cost

- No monetary costs to him

Benefit

- Sessions per Day: ~ 5 sessions a day.
- Daily Savings: 20% time savings = 1 extra session/day (+$150-200)
  - Annual Savings: Assuming 200 working days/year, the additional revenue or value from enhanced efficiency could range from $30,000 to $40,000 annually.

PATIENT #A

Cost

- $150-$200/session

Benefit

- Estimated Treatment Duration: 20 sessions/year
  - Savings: 10% of total cost, equaling $300 - $400 per treatment cycle.

COST BENEFIT ANALYSIS (MACRO)

THERAPY CLINIC(S)

Cost

- Implementation costs: ≥ $50k
- Cost of Maintenance + Analysing: $10-50k

Benefit

- Average Monthly Patient Load: 300 patients.
- Additional Patients: 30 patients (10%)
  - Increased Revenue: 30 patients x $175 (average session cost) x 20 sessions = $105k

Sources:

CNA, 2024, MOF, 2023

COST BENEFIT ANALYSIS (MACRO)

THERAPY CLINIC(S)

Cost

- Implementation costs: ≥ $50k
- Cost of Maintenance + Analysing: $10-50k

Benefit

- Average Monthly Patient Load: 300 patients.
- Additional Patients: 30 patients (10%)
  - Increased Revenue: 30 patients x $175 (average session cost) x 20 sessions = $105k

Cost

- ~$435 million (out of $16.68 billion budget)

Benefit

- Assuming 1% savings of allocated $435 million budget to mental health (subsidies etc.)
  - 1% x $16.68 billion = $4.4 million

Sources:

CNA, 2024, MOF, 2023

SUMMARY

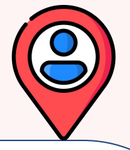
- WaveNet model classified 6 Emotions accurately 6 out of 10 times (0.65)
  - Allows lesser reliance on clinician’s self-discernment to assess patients’ moods/emotional state

SUMMARY

- WaveNet model classified 6 Emotions accurately 6 out of 10 times (0.65)
  - Allows lesser reliance on clinician’s self-discernment to assess patients’ moods/emotional state
- Proof-of-Concept (FeelFlow 1.0) worked
  - Recognition of emotions
  - Transcribing of words spoken

PROJECT SCOPE (MULTI-PHASED APPROACH)

WE'RE HERE!!!



FeelFlow AI 1.0
[CAPSTONE**]**

- Transfer Learning to the YouTube dataset

(which emulates a real Therapy setting)

- Classify audio into 6 emotional labels using the Modified WaveNet
- Using OpenAI's Whisper package for Application

FeelFlow AI 2.0
[BEYOND CAPSTONE **]**

- Transcription allows for us to conduct sentiment analysis to build a more robust model that classifies emotions in not just through audible speech, but also in the nuances meaning of the words, sentences verbalised.

[*Duration: 2-3 months*]

FeelFlow AI 3.0
[BEYOND CAPSTONE **]**

- Incorporate the use of facial recognition and body language. Mood and emotions are not just expressed in speech but also through the physical body as well.

[*Duration: 3-4 months*]

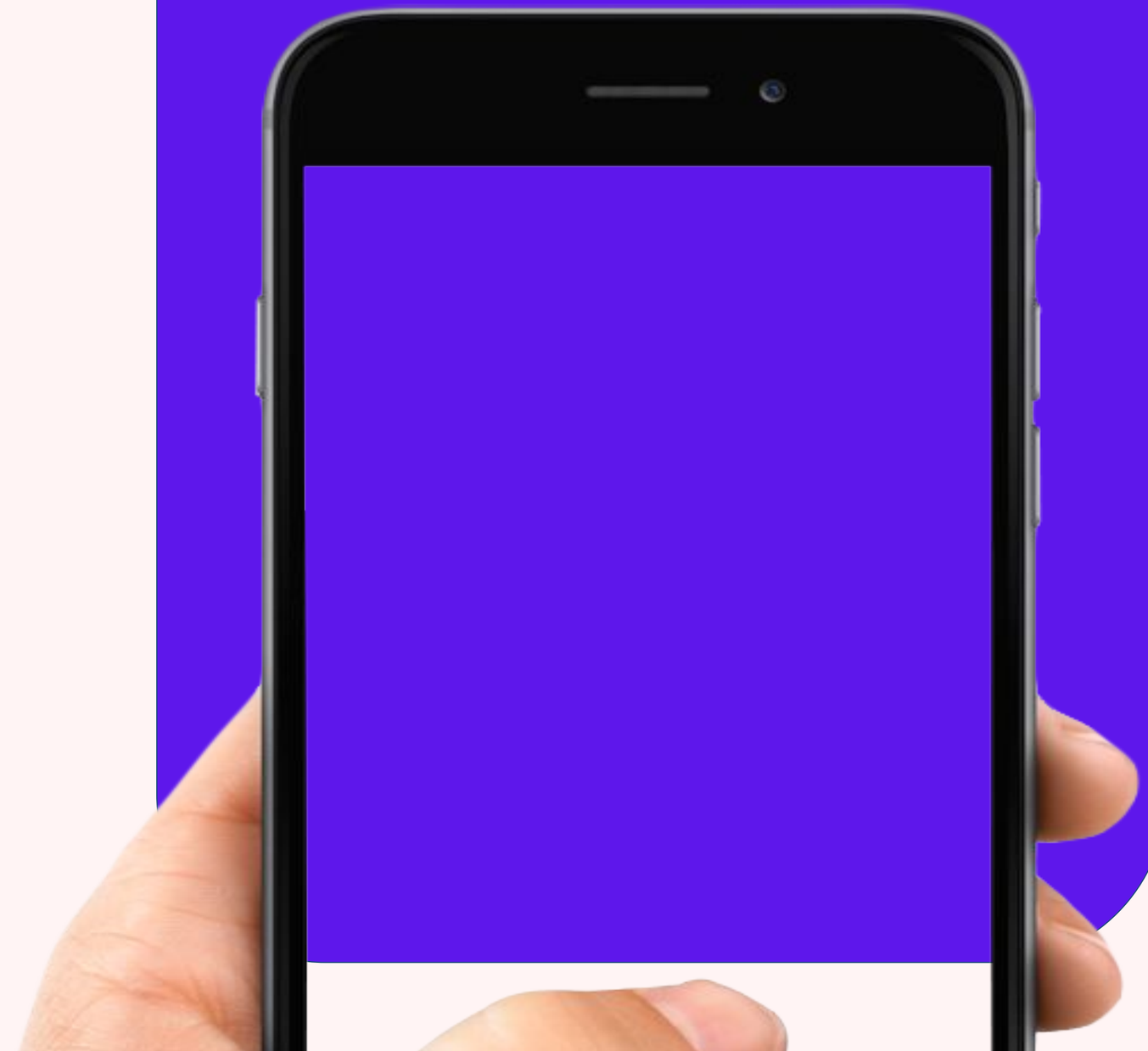
⚠ **Limitations of Study**	**Recommendations for Future**
Data Collection ● Culturally Nuanced (Other languages, beyond English) ● Generational Nuanced (Lingos used by GenZs) ● Does not account for accent (Singapore = multicultural?)	Collaborate with local Tertiary Institutions to collect and integrate data. [Note: Patient consent/PDPA has to be sought at point of collection]

⚠ Limitations of Study	Recommendations for Future
Data Collection ● Culturally Nuanced (Other languages, beyond English) ● Generational Nuanced (Lingos used by GenZs) ● Does not account for accent (Singapore = multicultural?)	Collaborate with local Tertiary Institutions to collect and integrate data. [Note: Patient consent/PDPA has to be sought at point of collection]
Limited defined Emotional Labels related to Mental Health (Sad = Depressed?)	Redefine emotional labels, and consider a wider but definitive spectrum [Note: Be sure not to overlap emotions - e.g. Disappointment = Anger + Sadness]

⚠ Limitations of Study	Recommendations for Future
Data Collection ● Culturally Nuanced (Other languages, beyond English) ● Generational Nuanced (Lingos used by GenZs) ● Does not account for accent (Singapore = multicultural?)	Collaborate with local Tertiary Institutions to collect and integrate data. [Note: Patient consent/PDPA has to be sought at point of collection]
Limited defined Emotional Labels related to Mental Health (Sad = Depressed?)	Redefine emotional labels, and consider a wider but definitive spectrum [Note: Be sure not to overlap emotions - e.g. Disappointment = Anger + Sadness]
Voice might not be able to sense contextual cues or things like sarcasm, lying etc.	Incorporate other stimulus: - Natural Language Processing - Semantics of Text (FeelFlow 2.0) - Facial Recognition (FeelFlow 3.0) [calculated by Emotion Detection Rate] Collaboration with AI SG - Speech Labs (already trained on local accent and speech semantics)

07 DEMO

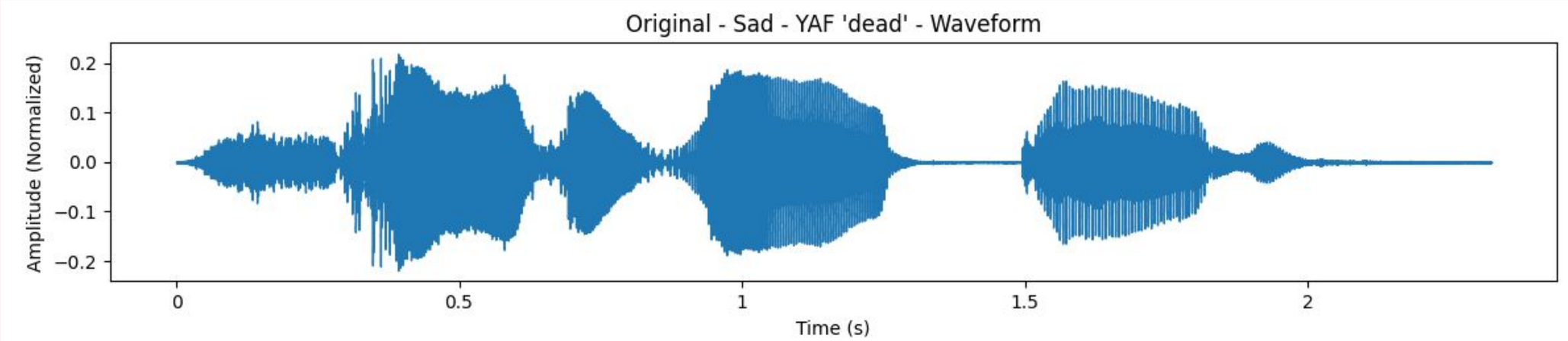
SCAN HERE

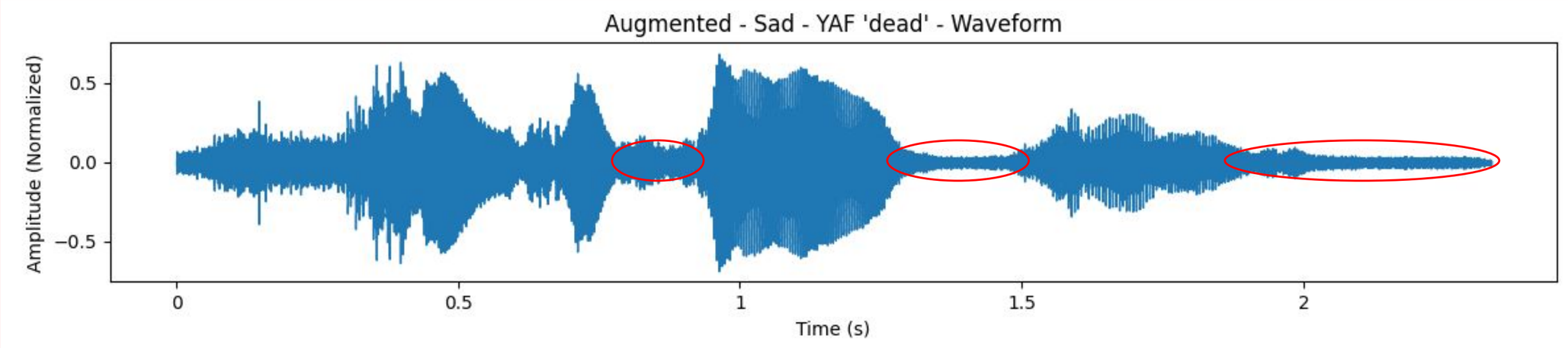


Thank You

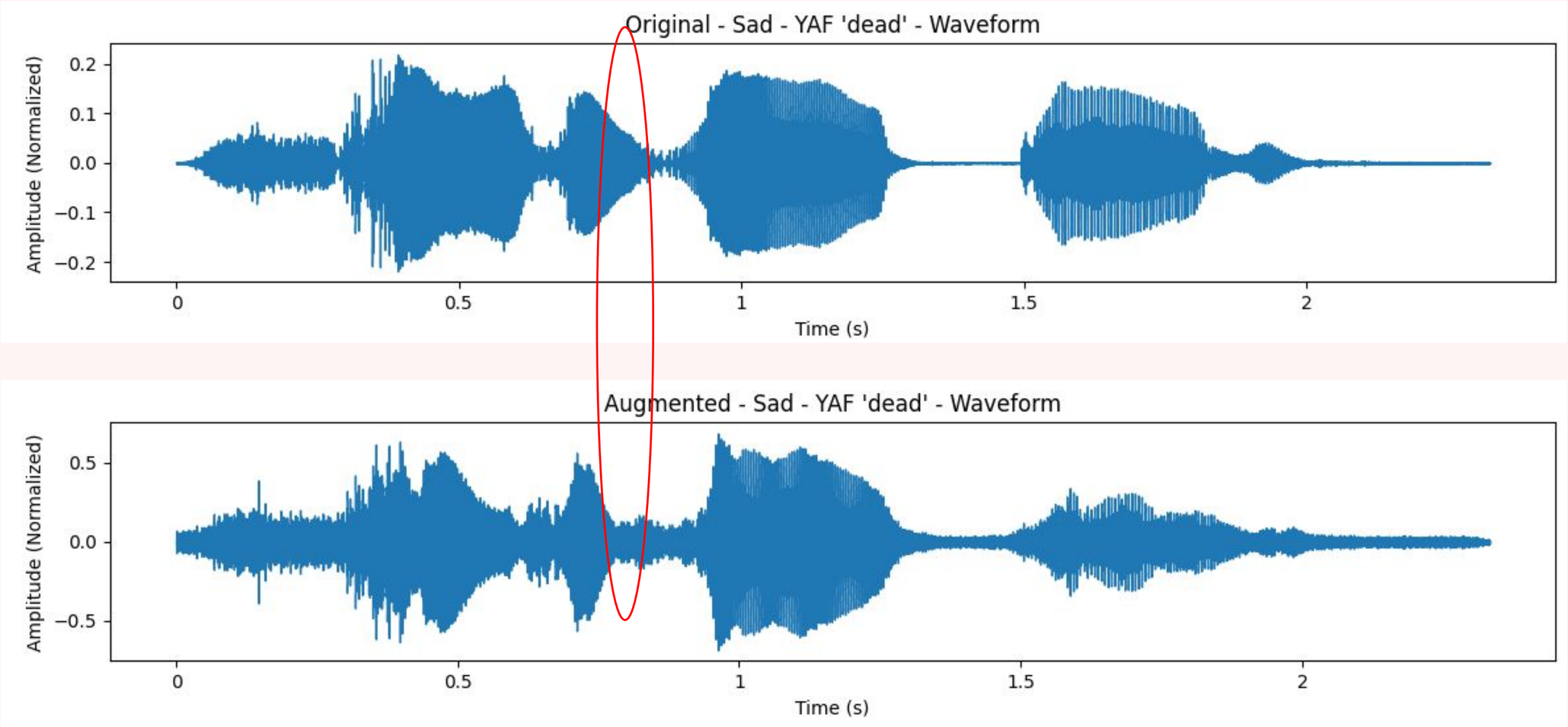
APPENDIX

EFFECTS OF AUGMENTATION - SAD (NOISE ADDITION)

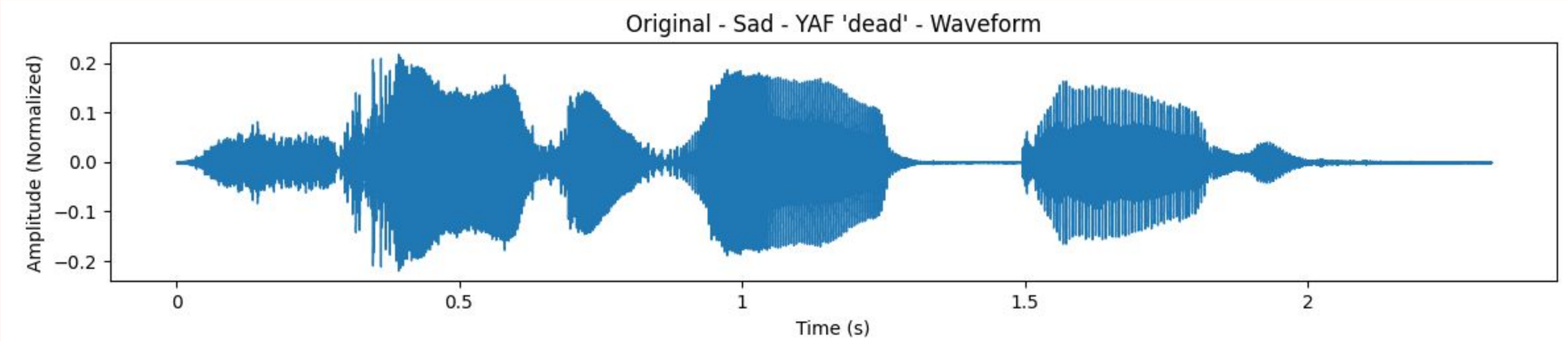


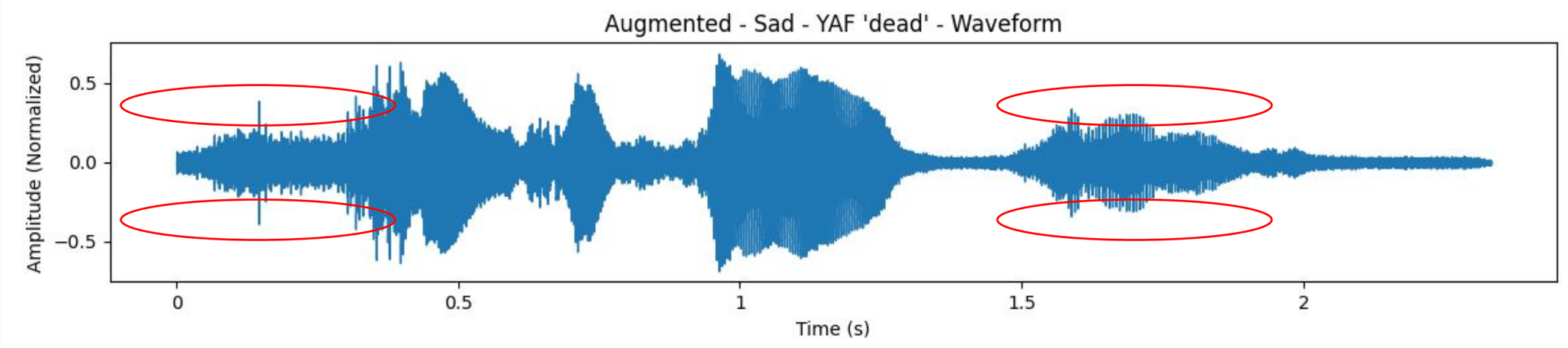


EFFECTS OF AUGMENTATION - SAD (TIME SHIFT/STRETCH)

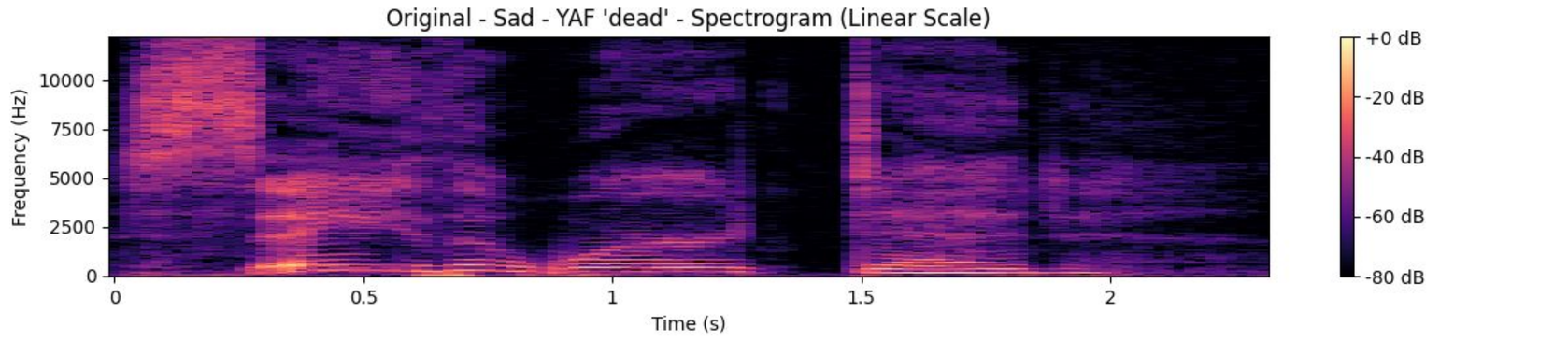


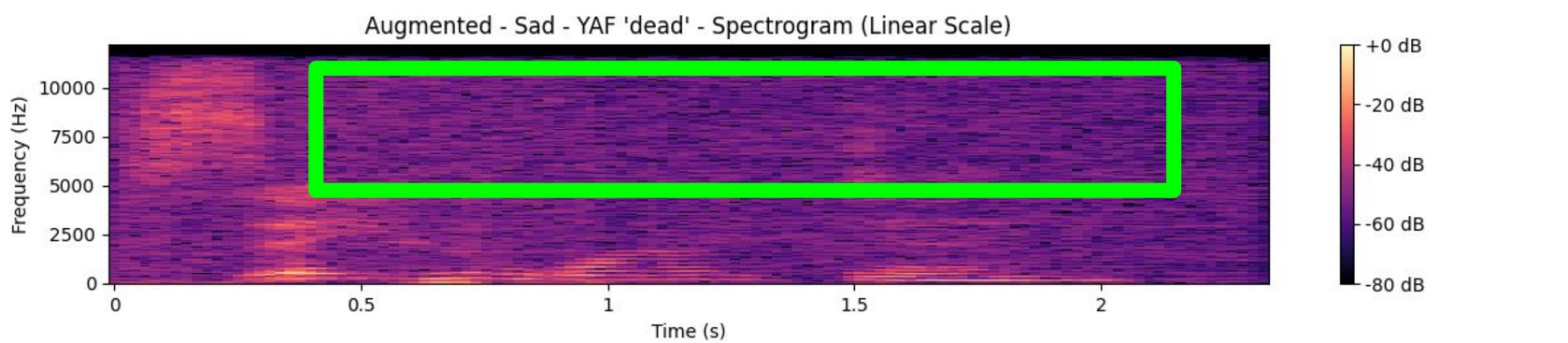
EFFECTS OF AUGMENTATION - SAD (PITCH SHIFT)





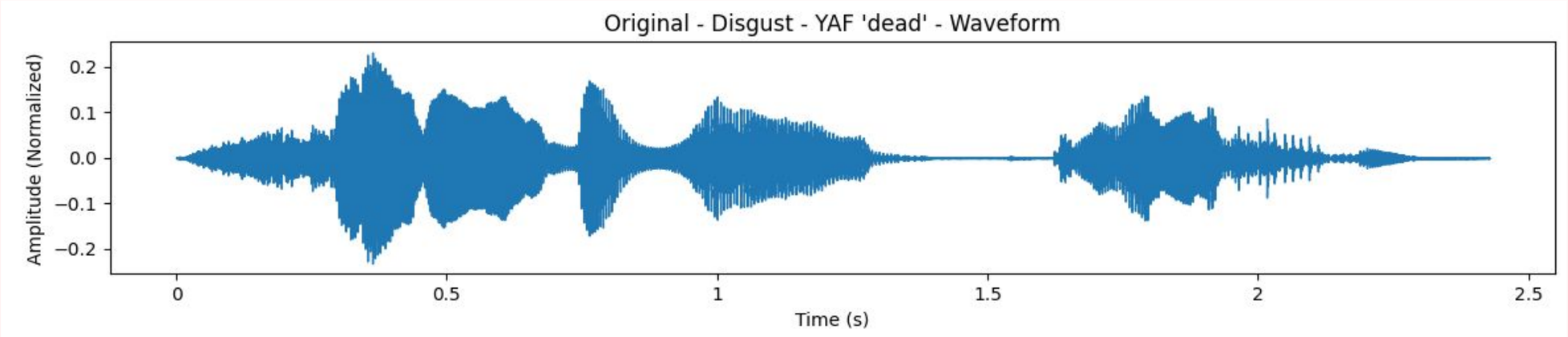
LINEAR SPECTROGRAM - SAD

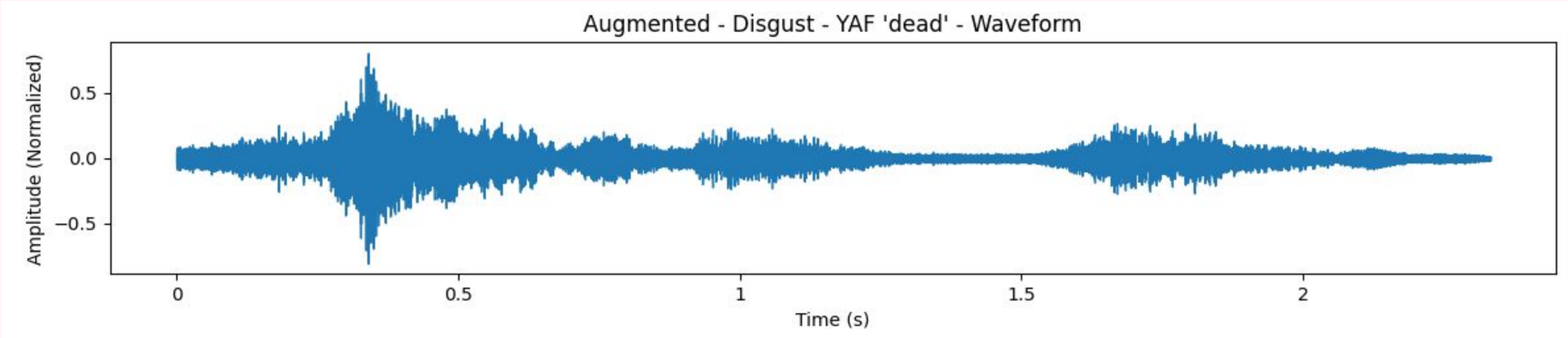




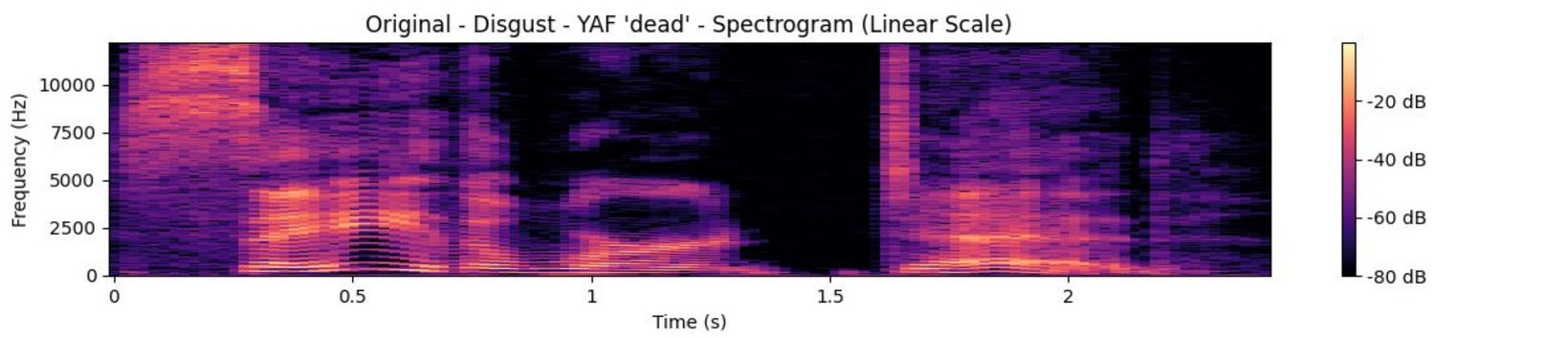
- Sadness is usually monotonous and sombre in nature, hence high pitch is an anomaly.
- Augmenting here has removed the unnecessary pitch

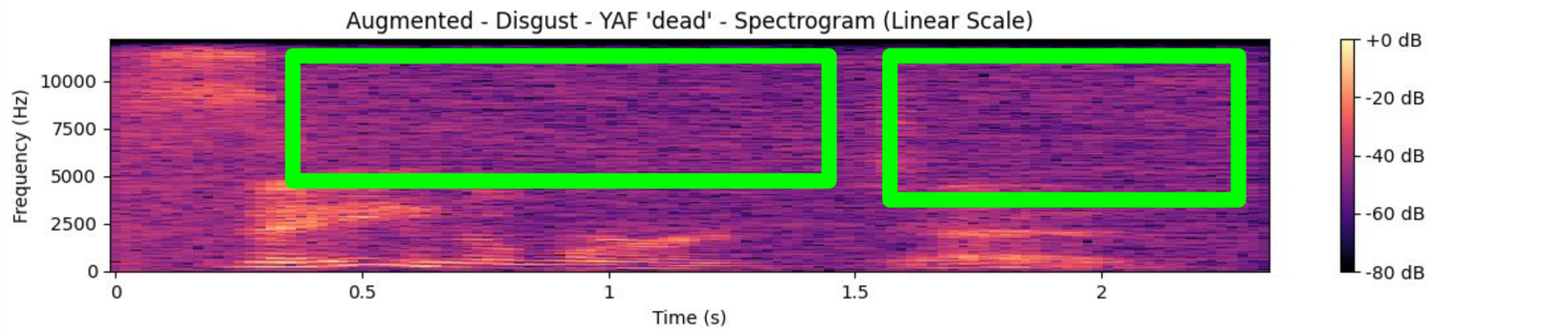
EFFECTS OF AUGMENTATION - DISGUST





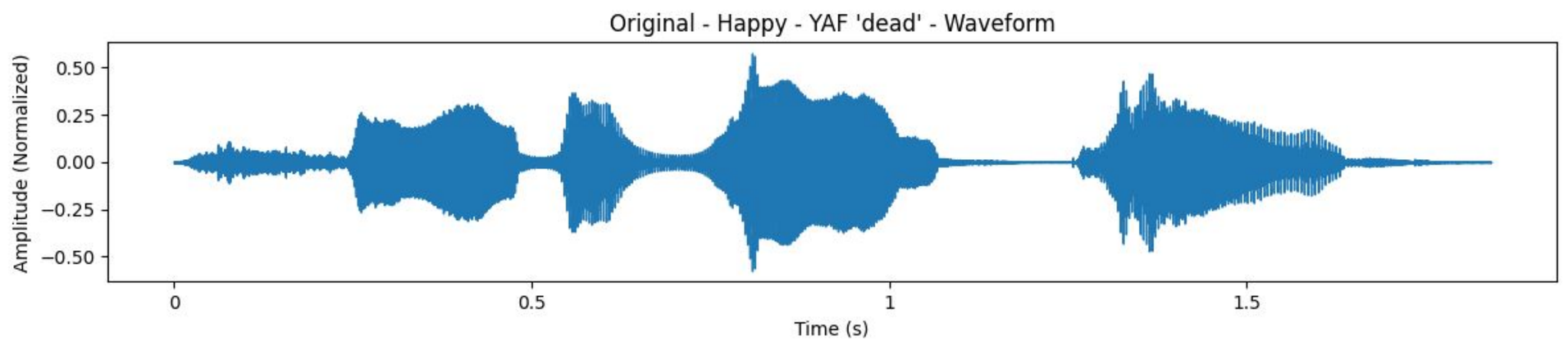
LINEAR SPECTROGRAM - DISGUST

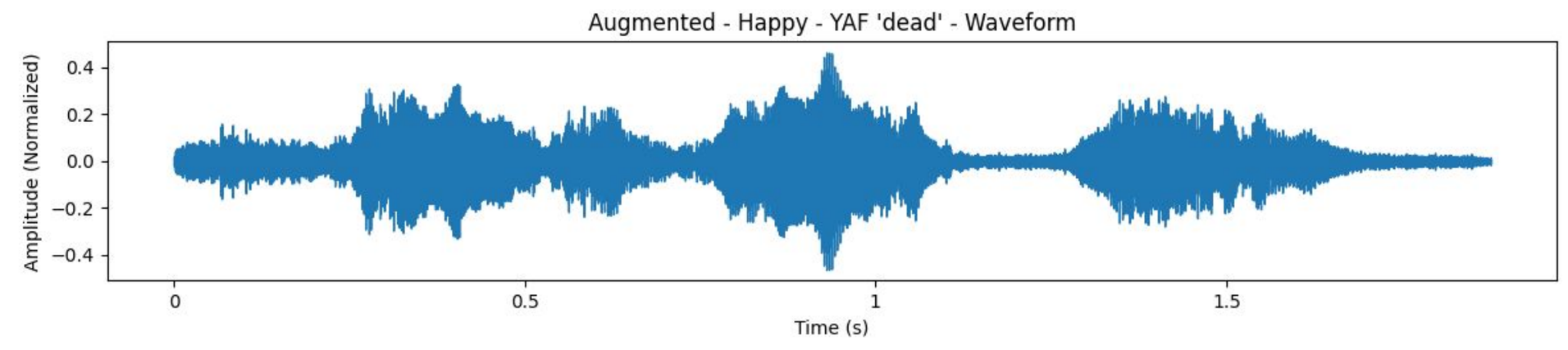




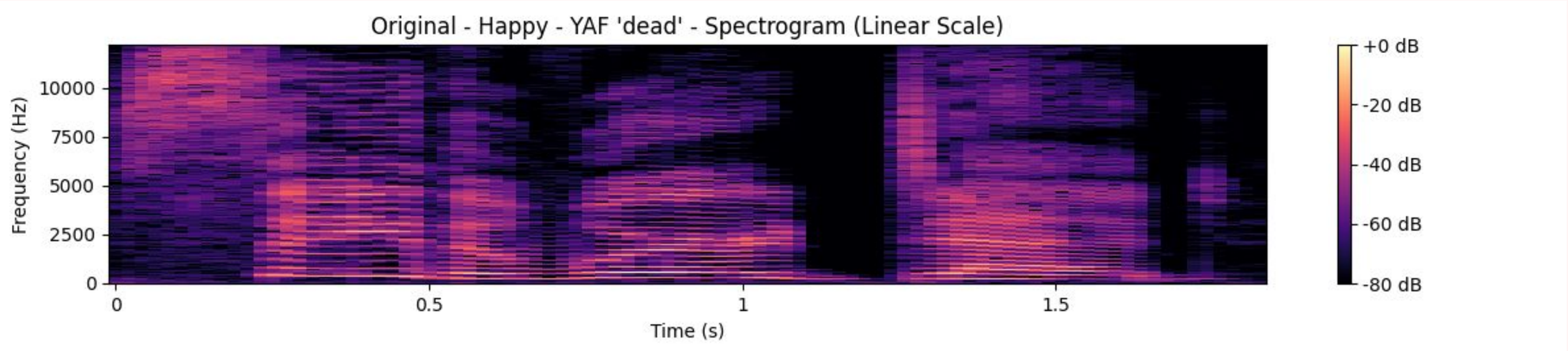
- The intensity of the Augmentation effect here is similar to the one experienced in ‘Angry’

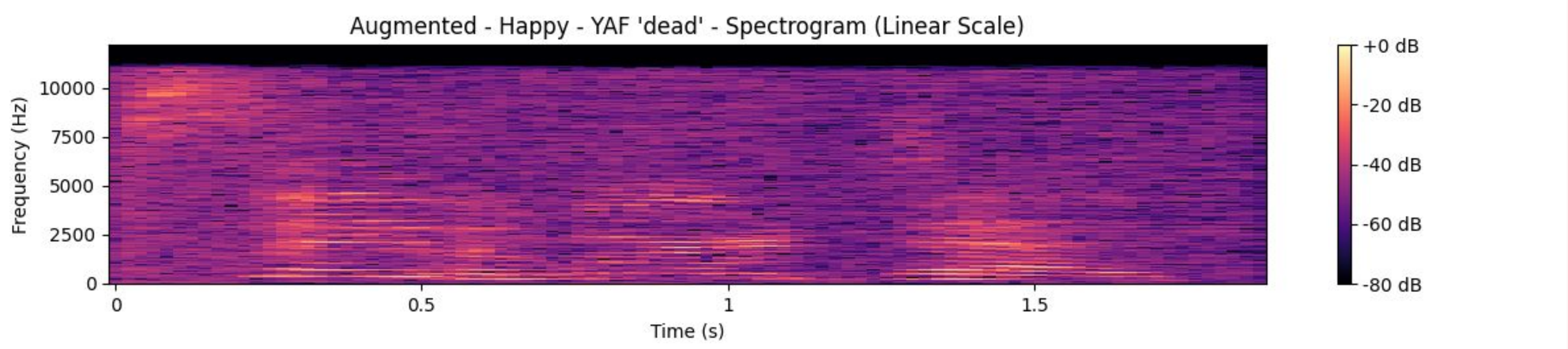
EFFECTS OF AUGMENTATION - HAPPY



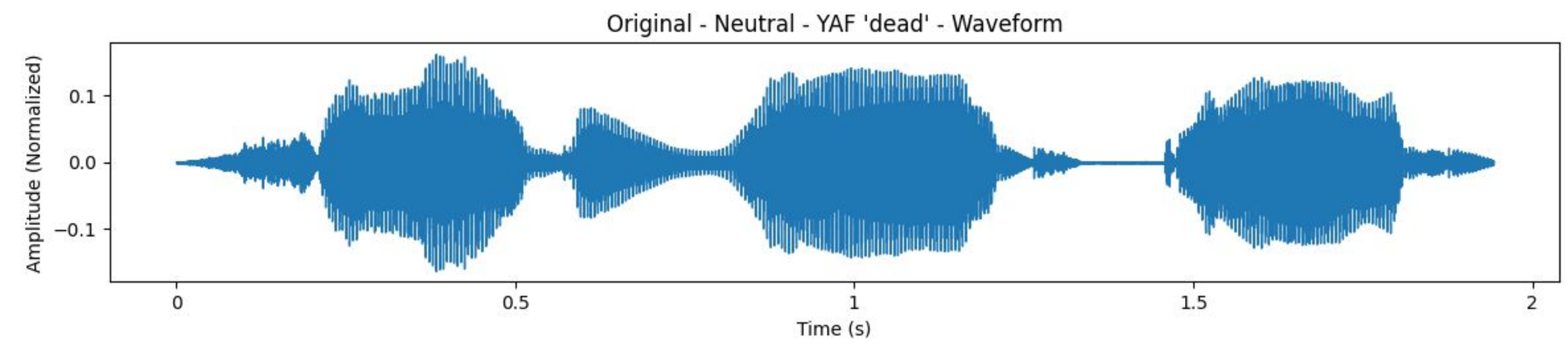


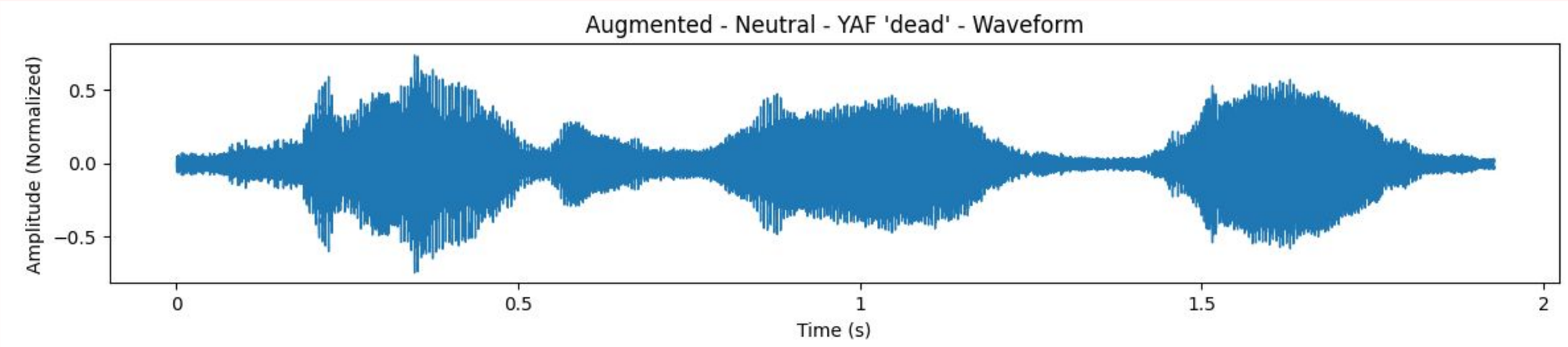
LINEAR SPECTROGRAM - HAPPY



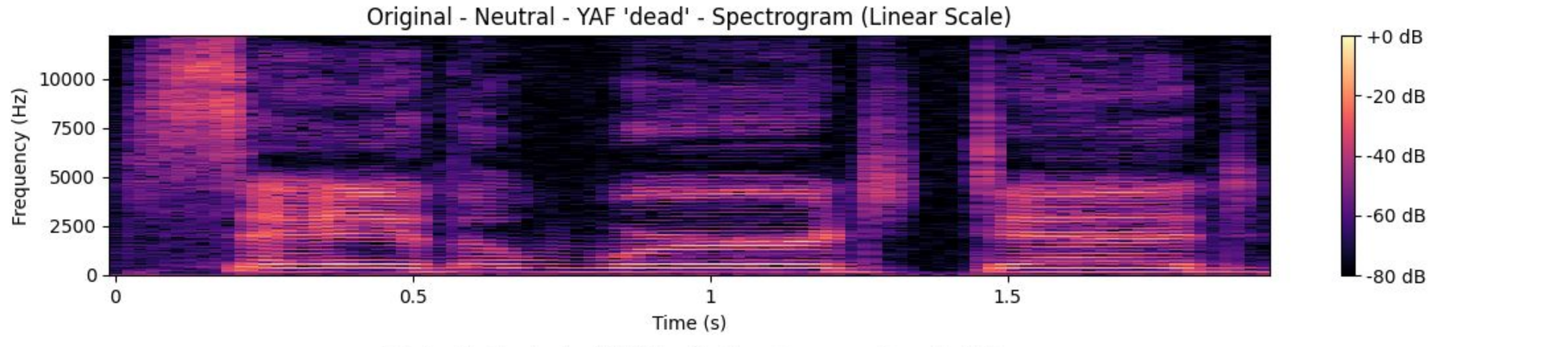


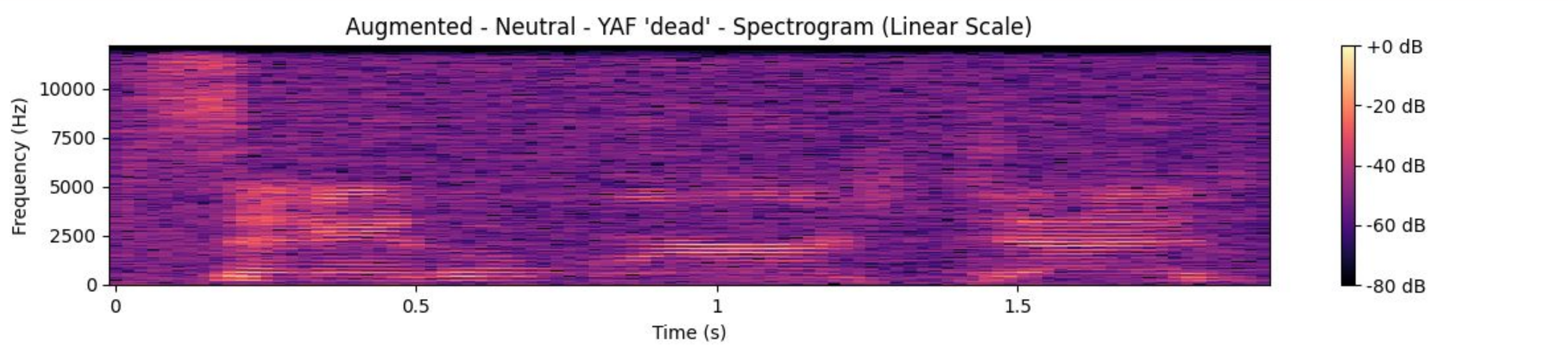
EFFECTS OF AUGMENTATION - NEUTRAL



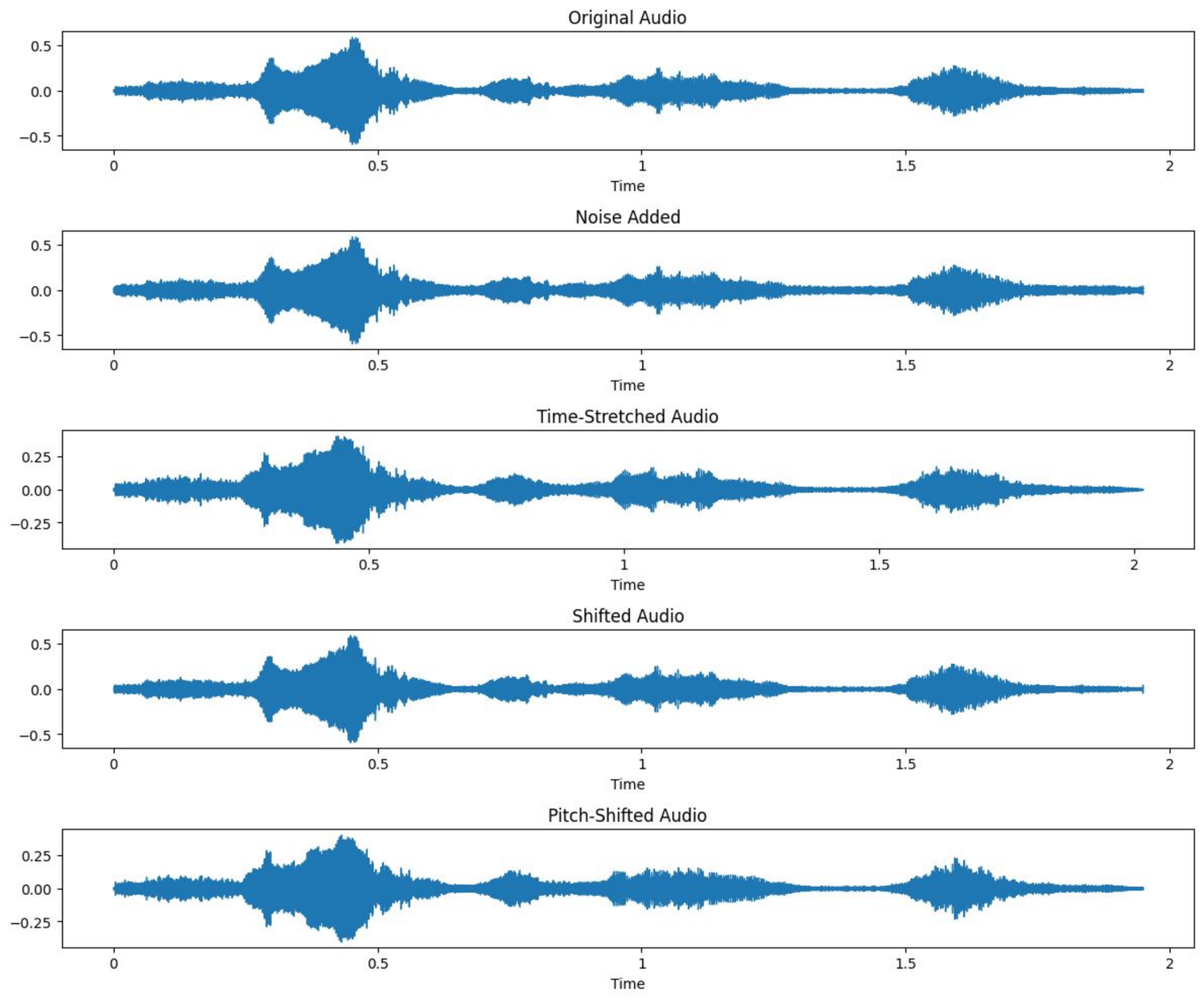


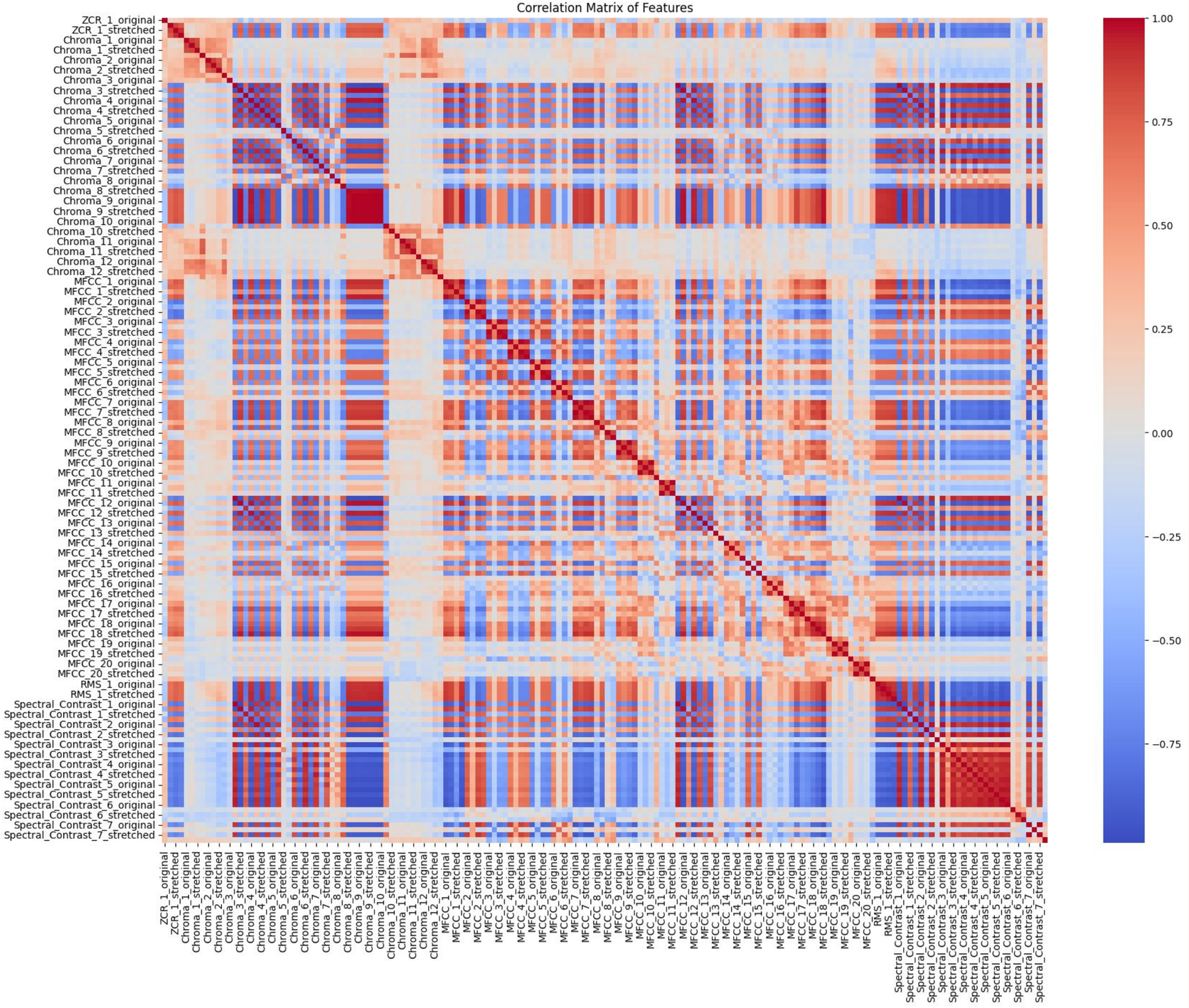
LINEAR SPECTROGRAM - NEUTRAL



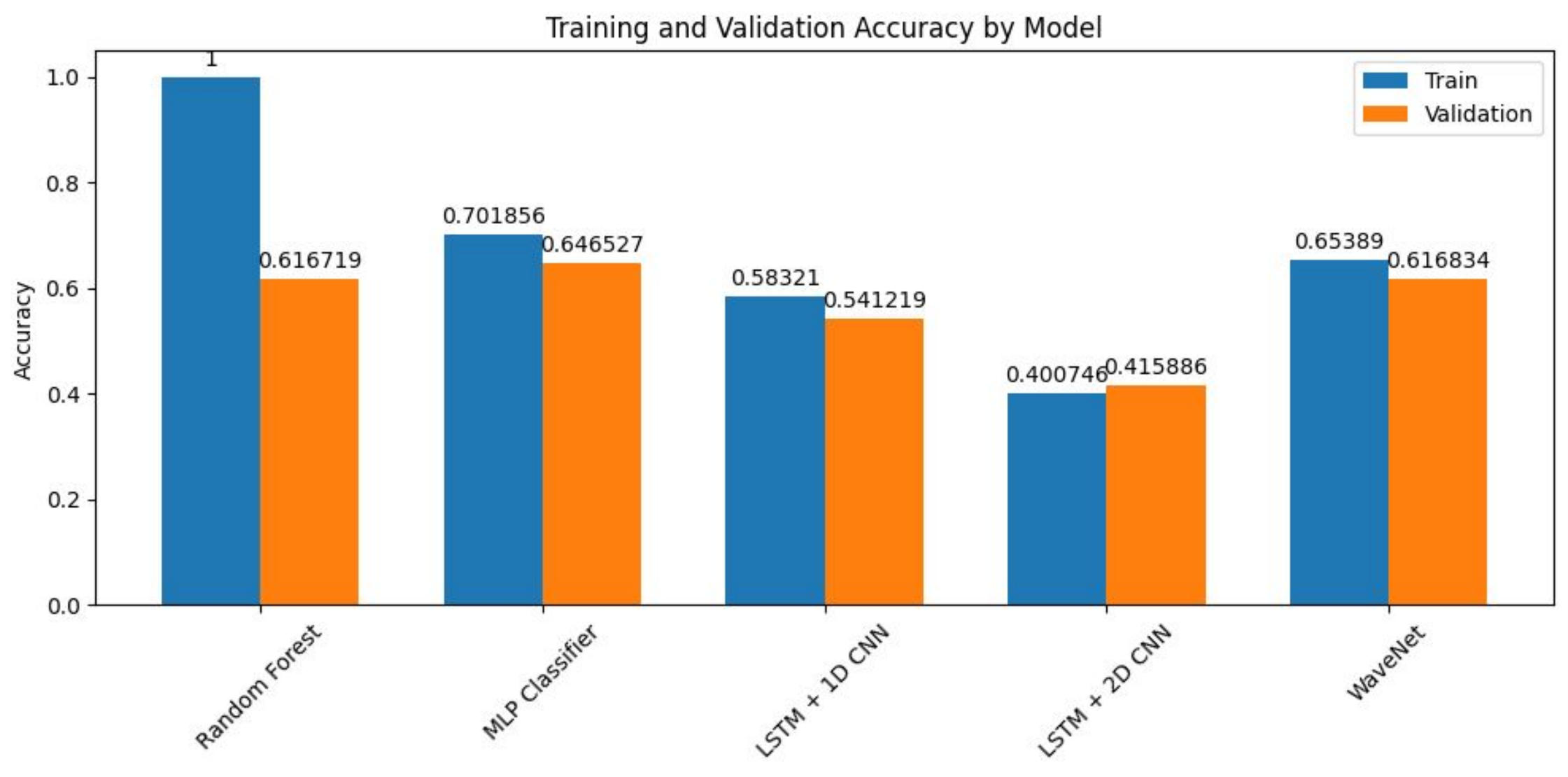


TYPES OF AUGMENTATION VS. ORIGINAL

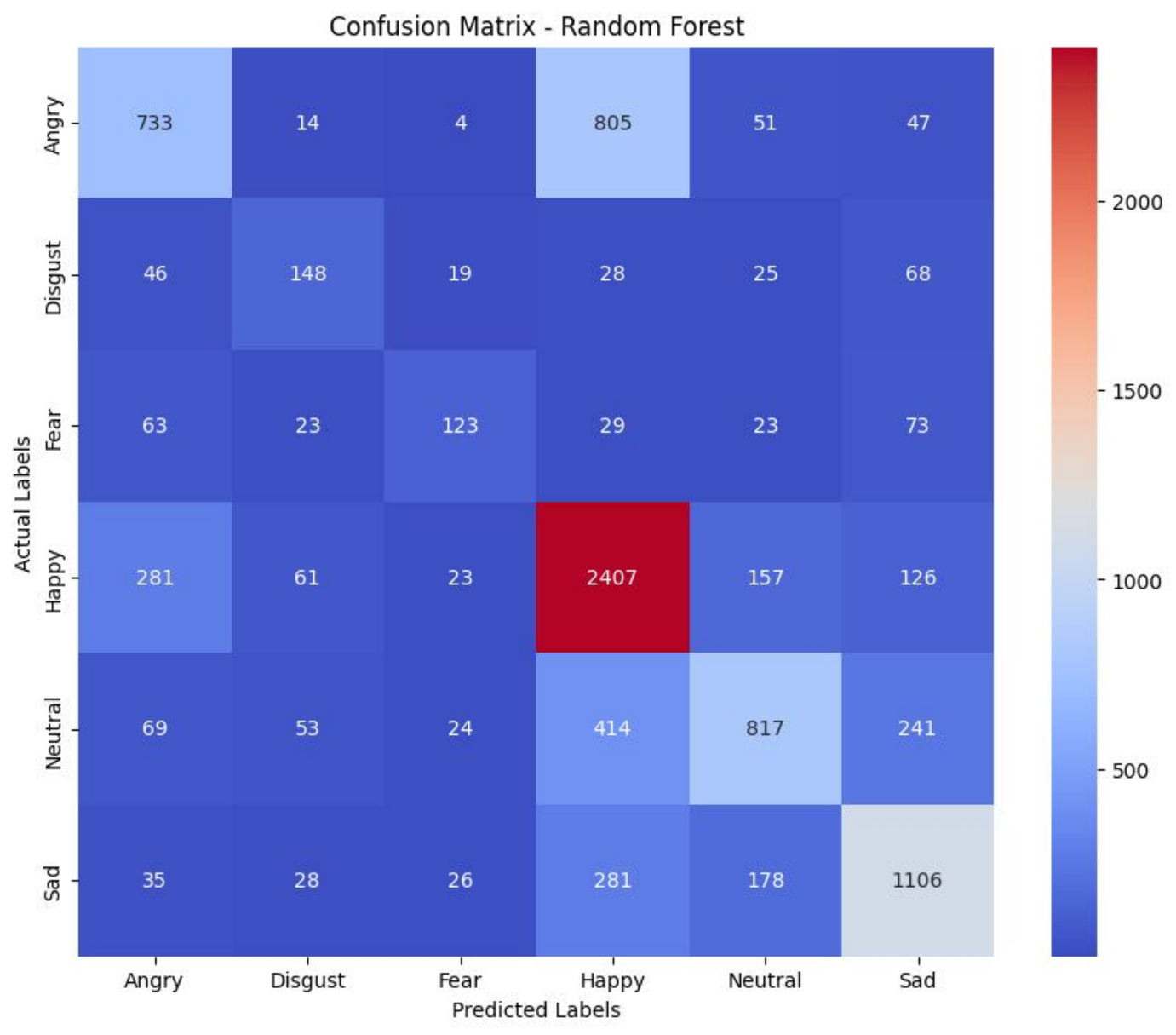



Correlation Matrix of Features
ZCR_1_original
ZCR_1_stretched
Chroma_1_original
Chroma_1_stretched
Chroma_2_original
Chroma_2_stretched
Chroma_3_original
Chroma_3_stretched
Chroma_4_original
Chroma_4_stretched
Chroma_5_original
Chroma_5_stretched
Chroma_6_original
Chroma_6_stretched
Chroma_7_original
Chroma_7_stretched
Chroma_8_original
Chroma_8_stretched
Chroma_9_original
Chroma_9_stretched
Chroma_10_original
Chroma_10_stretched
Chroma_11_original
Chroma_11_stretched
Chroma_12_original
Chroma_12_stretched
MFCC_1_original
MFCC_1_stretched
MFCC_2_original
MFCC_2_stretched
MFCC_3_original
MFCC_3_stretched
MFCC_4_original
MFCC_4_stretched
MFCC_5_original
MFCC_5_stretched
MFCC_6_original
MFCC_6_stretched
MFCC_7_original
MFCC_7_stretched
MFCC_8_original
MFCC_8_stretched
MFCC_9_original
MFCC_9_stretched
MFCC_10_original
MFCC_10_stretched
MFCC_11_original
MFCC_11_stretched
MFCC_12_original
MFCC_12_stretched
MFCC_13_original
MFCC_13_stretched
MFCC_14_original
MFCC_14_stretched
MFCC_15_original
MFCC_15_stretched
MFCC_16_original
MFCC_16_stretched
MFCC_17_original
MFCC_17_stretched
MFCC_18_original
MFCC_18_stretched
MFCC_19_original
MFCC_19_stretched
MFCC_20_original
MFCC_20_stretched
RMS_1_original
RMS_1_stretched
Spectral_Contrast_1_original
Spectral_Contrast_1_stretched
Spectral_Contrast_2_original
Spectral_Contrast_2_stretched
Spectral_Contrast_3_original
Spectral_Contrast_3_stretched
Spectral_Contrast_4_original
Spectral_Contrast_4_stretched
Spectral_Contrast_5_original
Spectral_Contrast_5_stretched
Spectral_Contrast_6_original
Spectral_Contrast_6_stretched
Spectral_Contrast_7_original
Spectral_Contrast_7_stretched
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

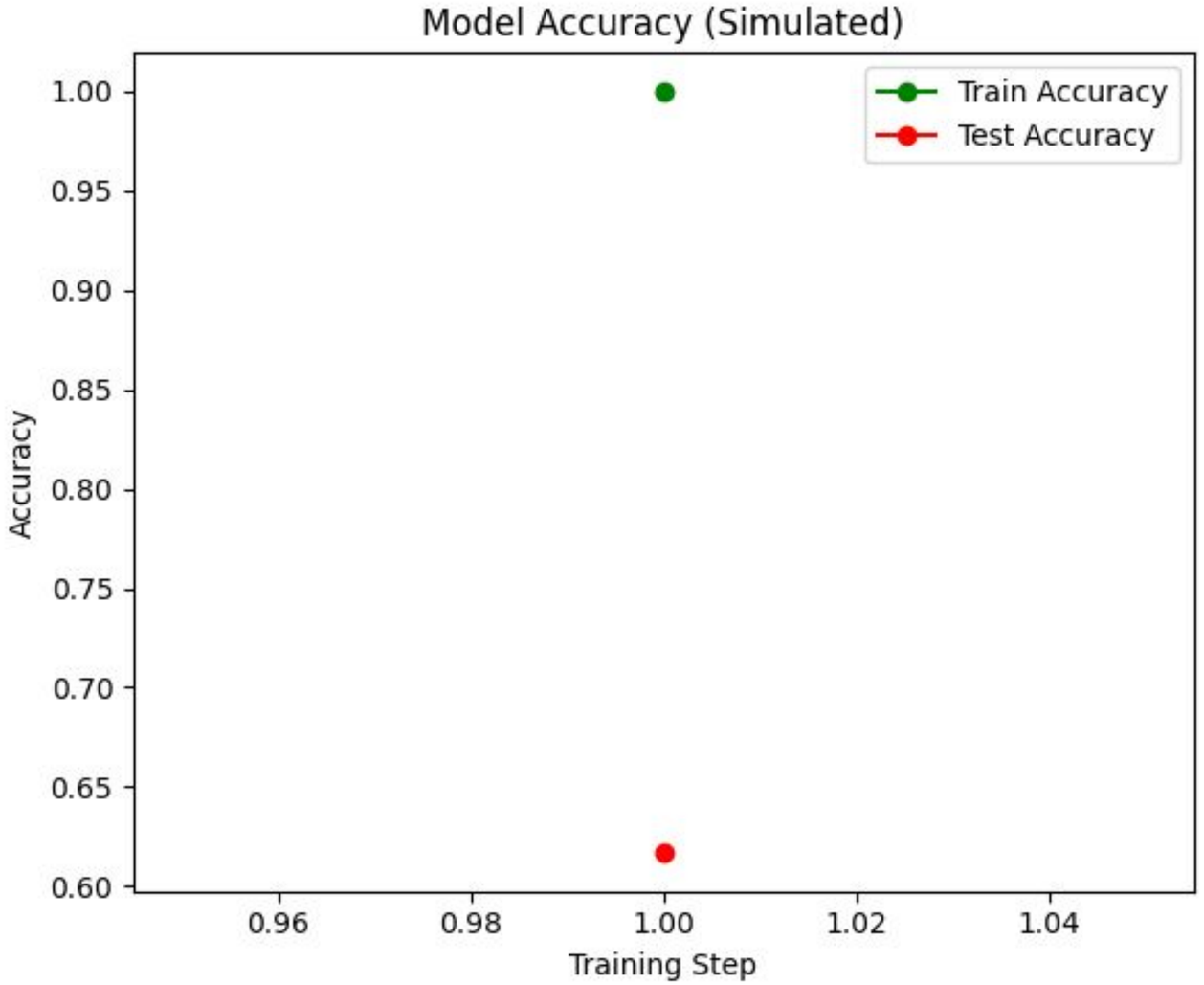
MODEL SCORE COMPARISON

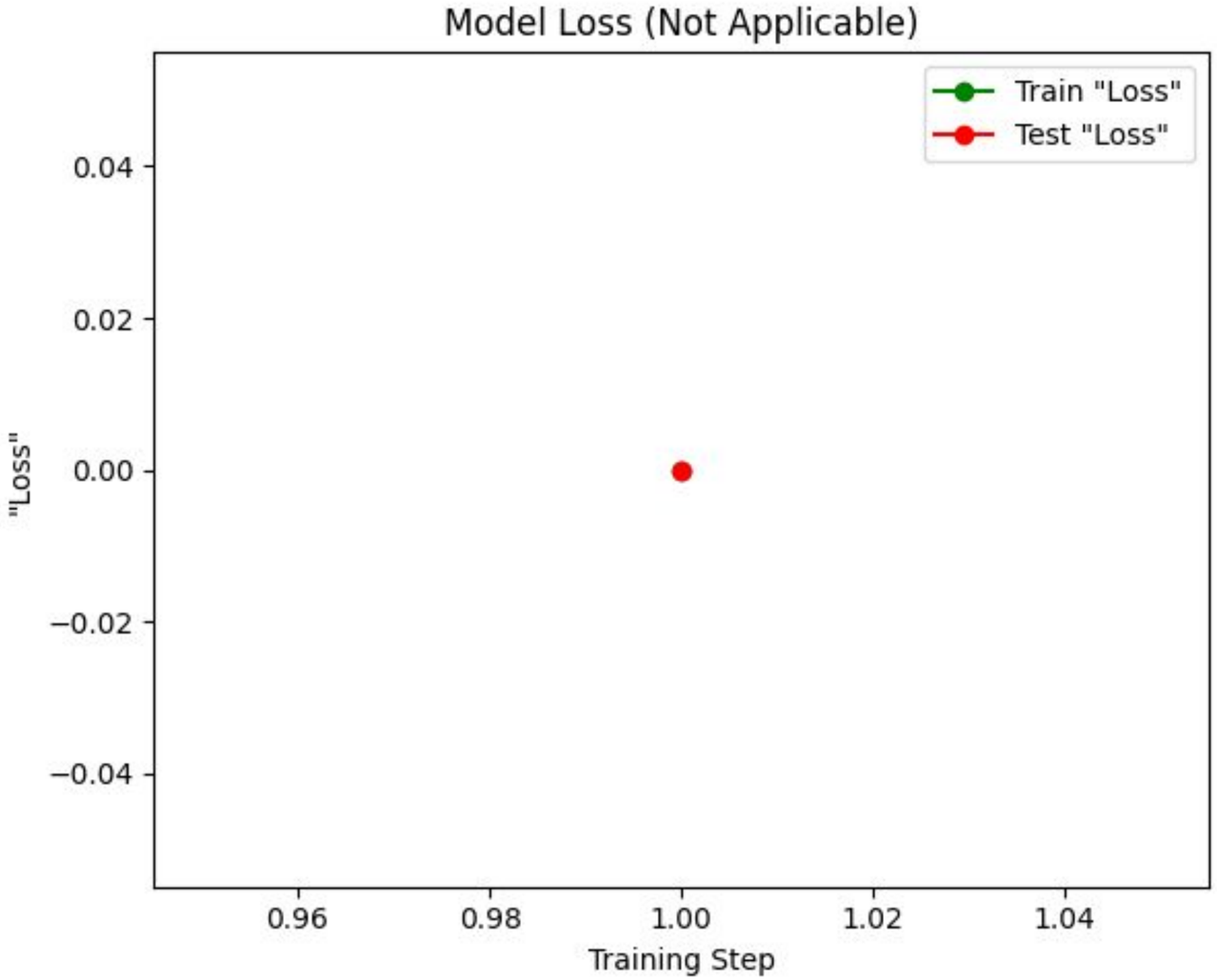


RANDOM FOREST CLASSIFIER



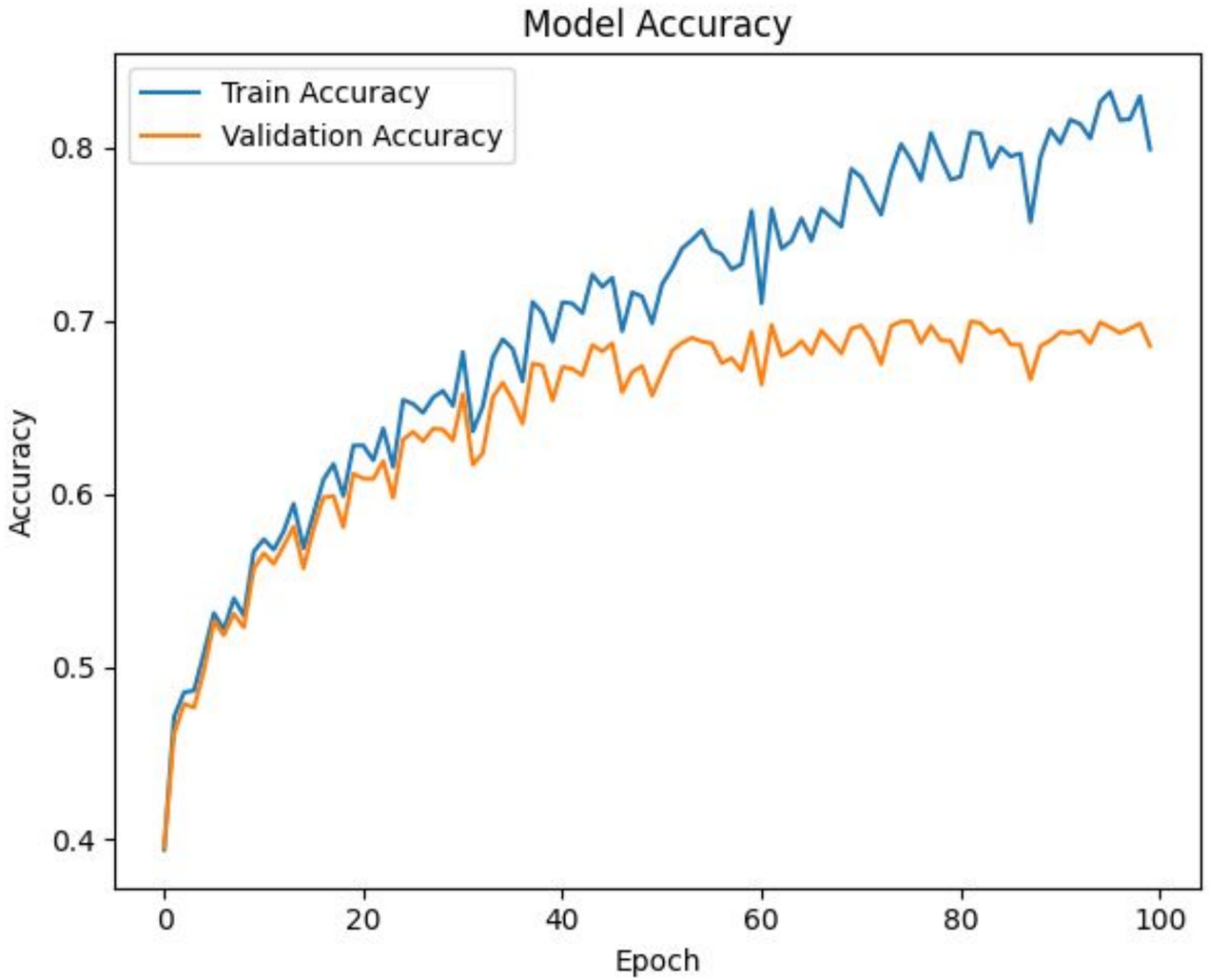
RANDOM FOREST CLASSIFIER

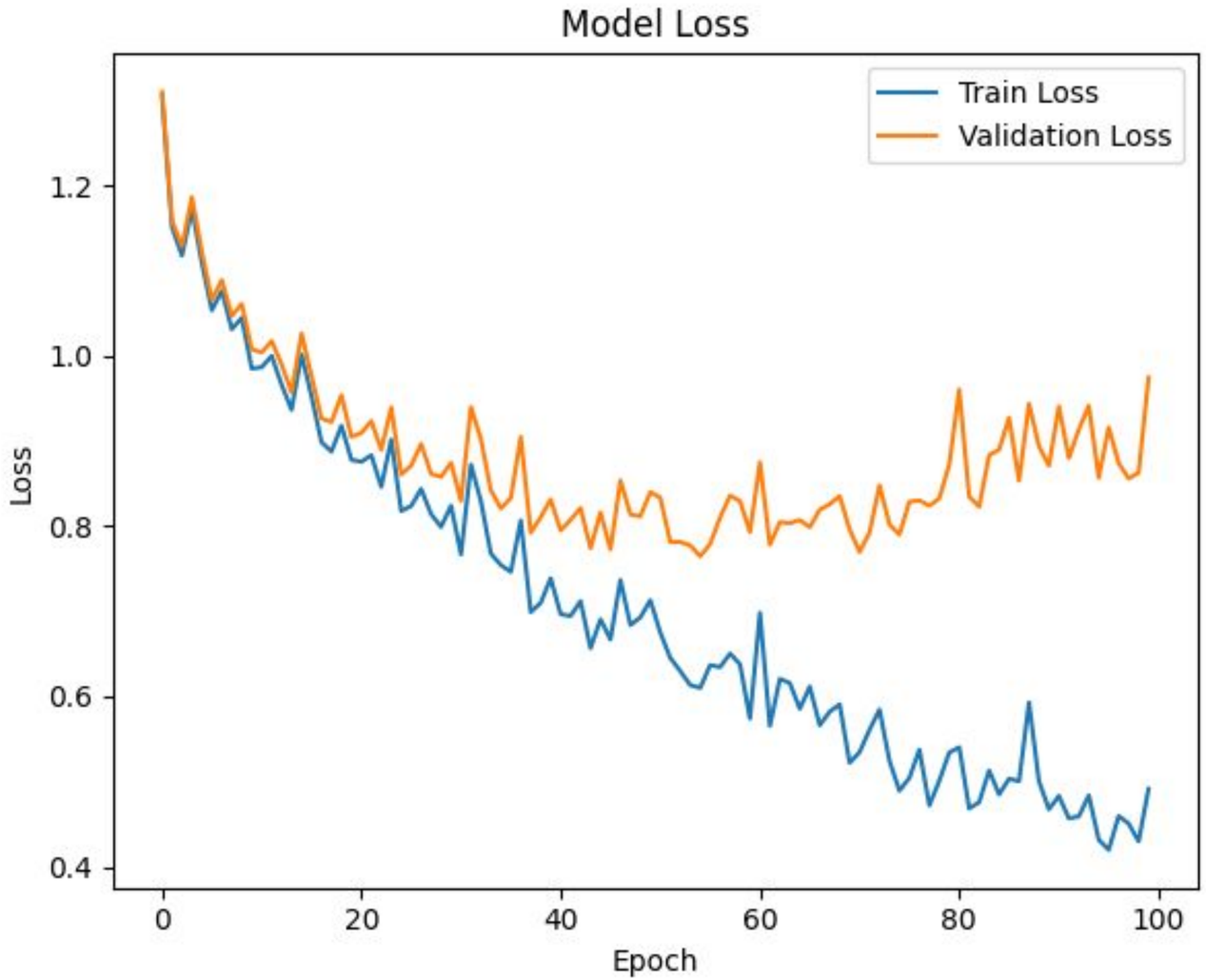




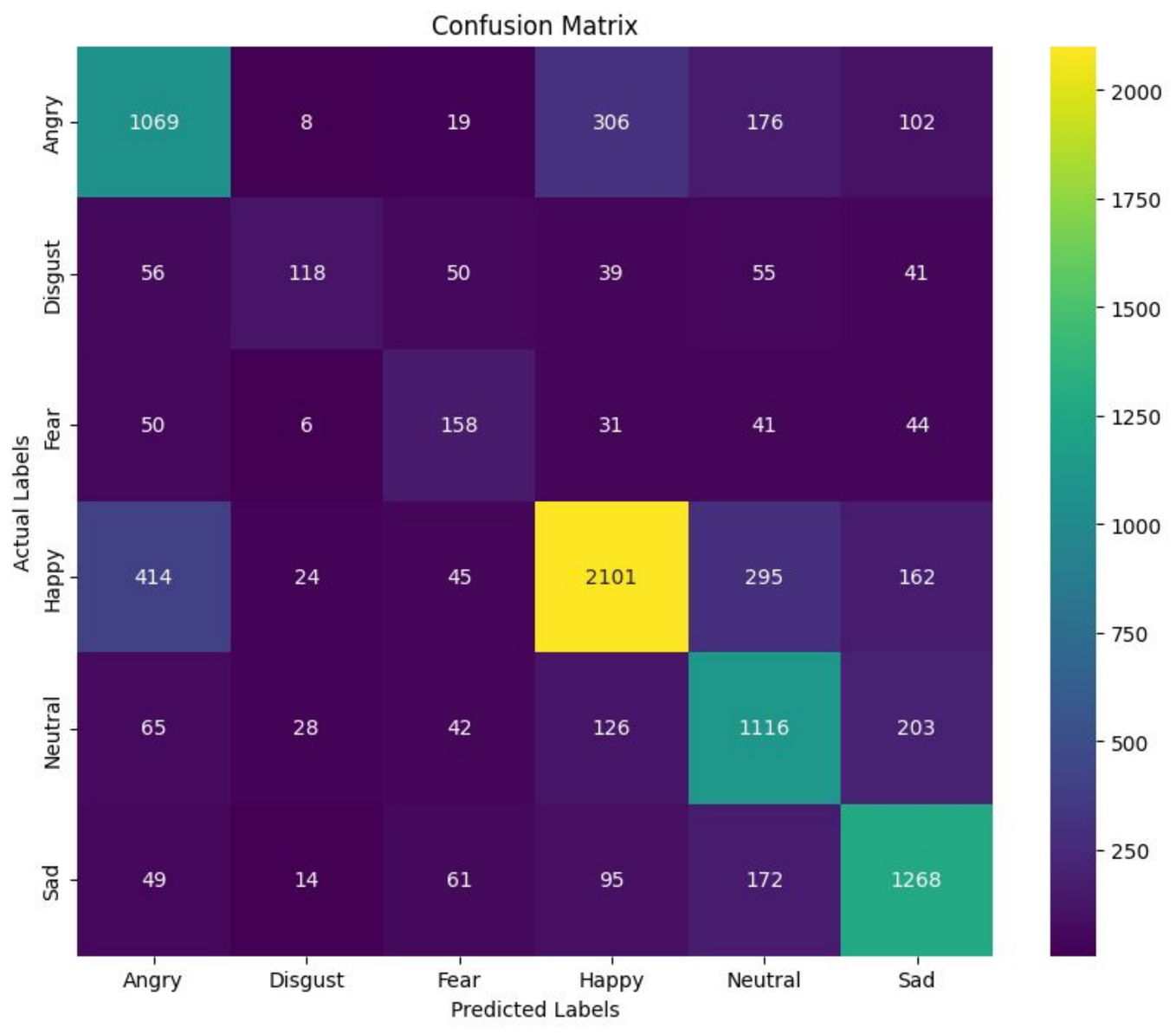
Note: Random Forest models cannot churn out Model Accuracy and Loss curve because they don't run by layers

MLP CLASSIFIER

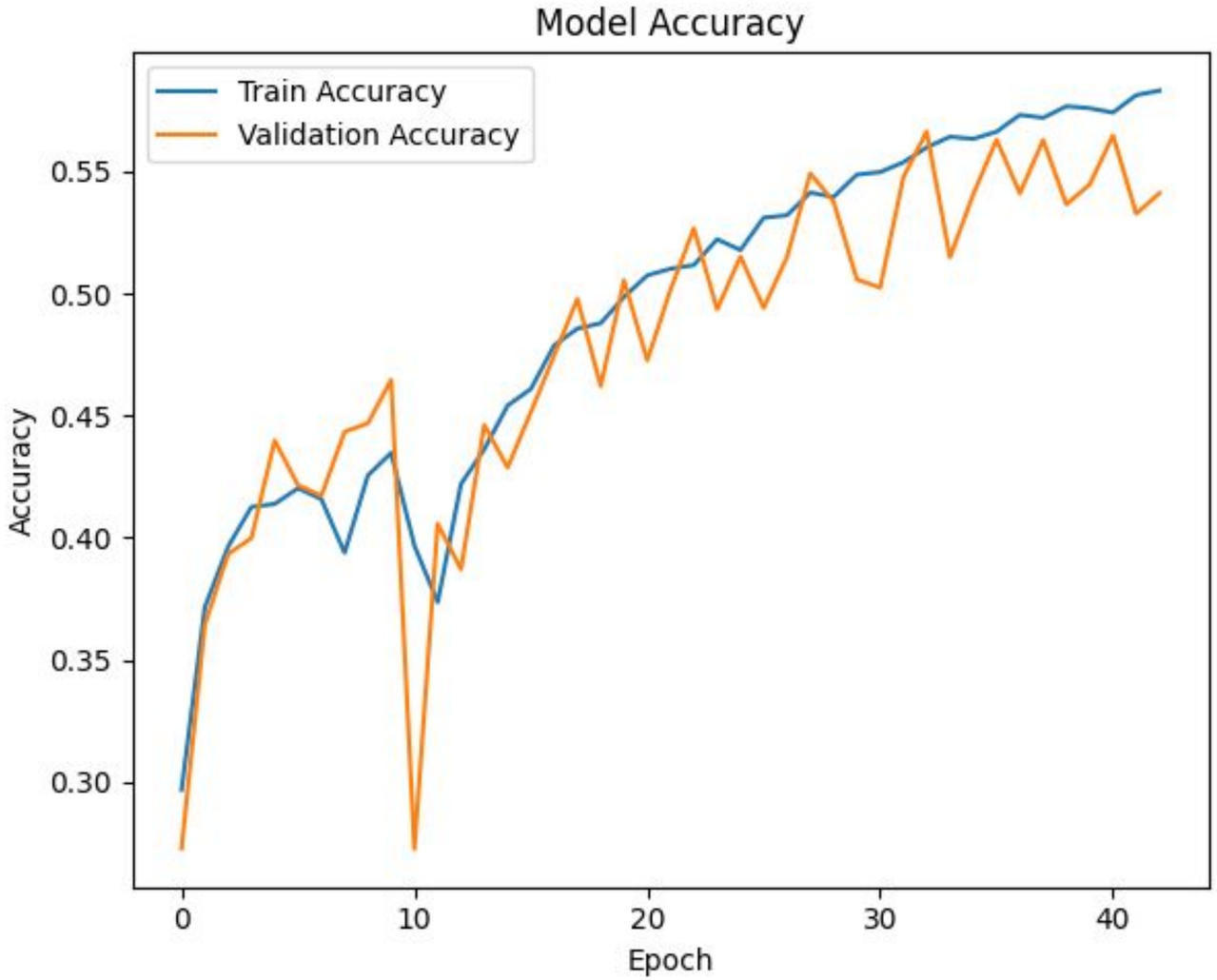


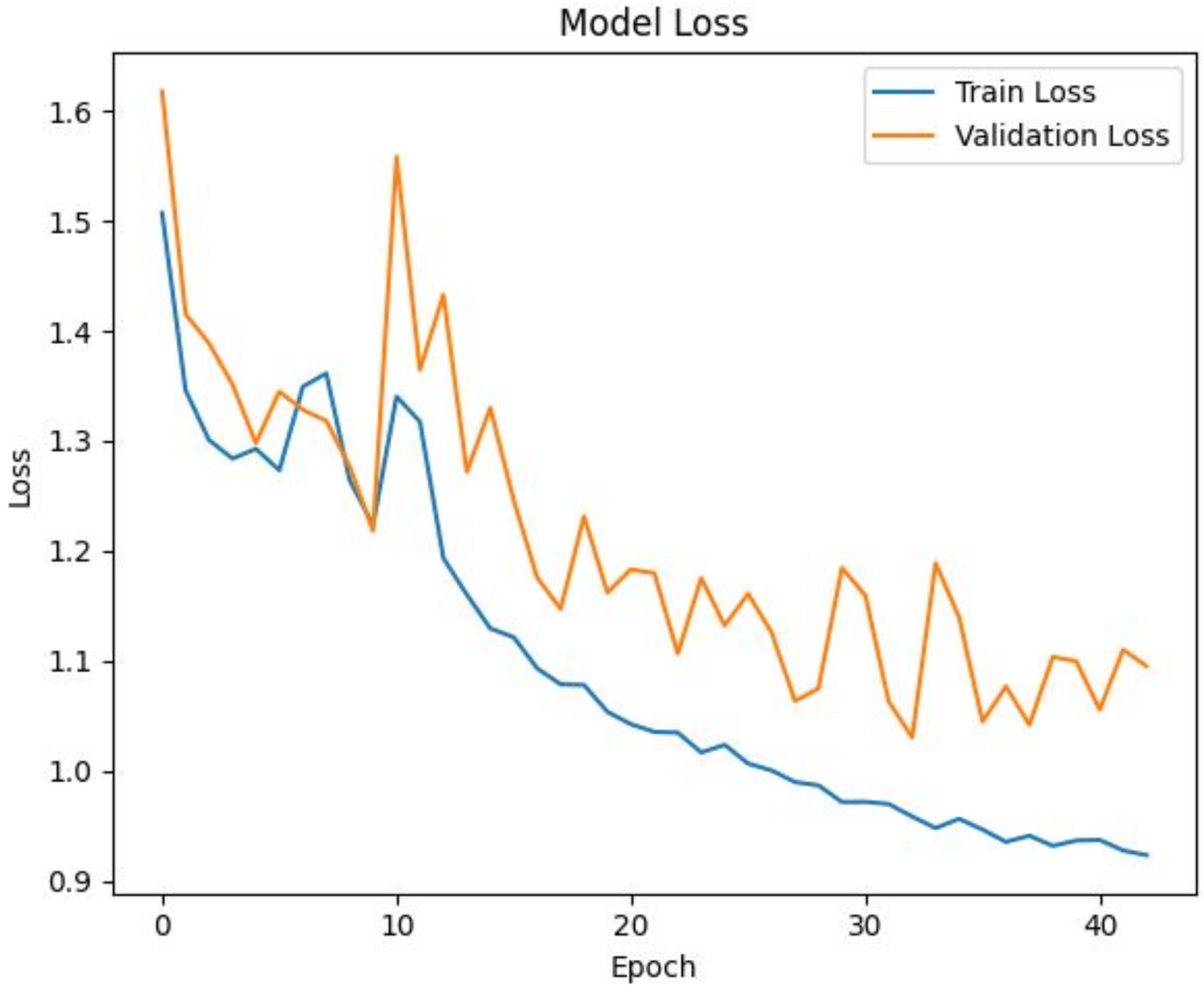


MLP CLASSIFIER

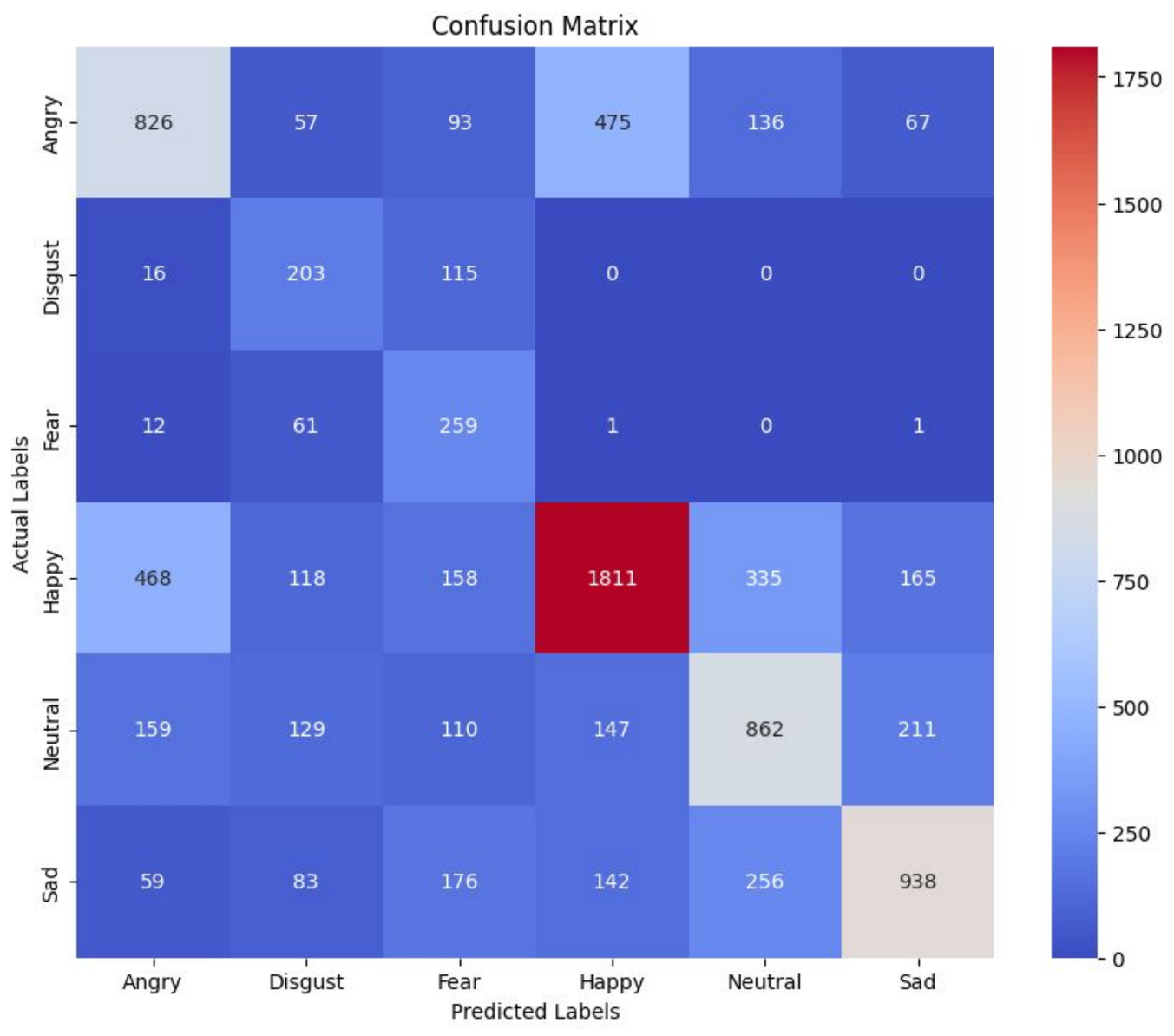


LSTM + 1D CNN

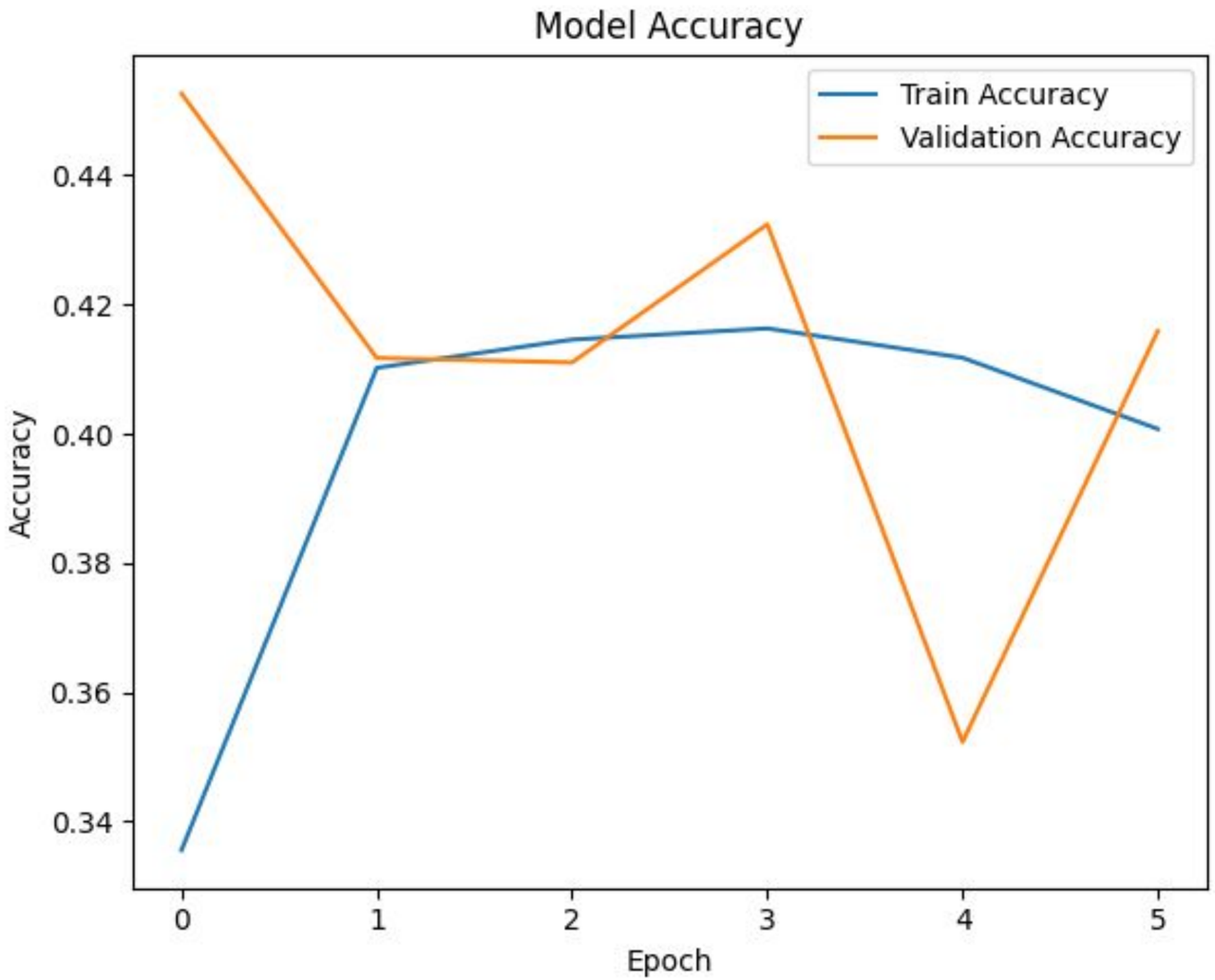

Model Accuracy
Train Accuracy
Validation Accuracy
Accuracy
0.55
0.50
0.45
0.40
0.35
0.30
0
10
20
30
40
Epoch

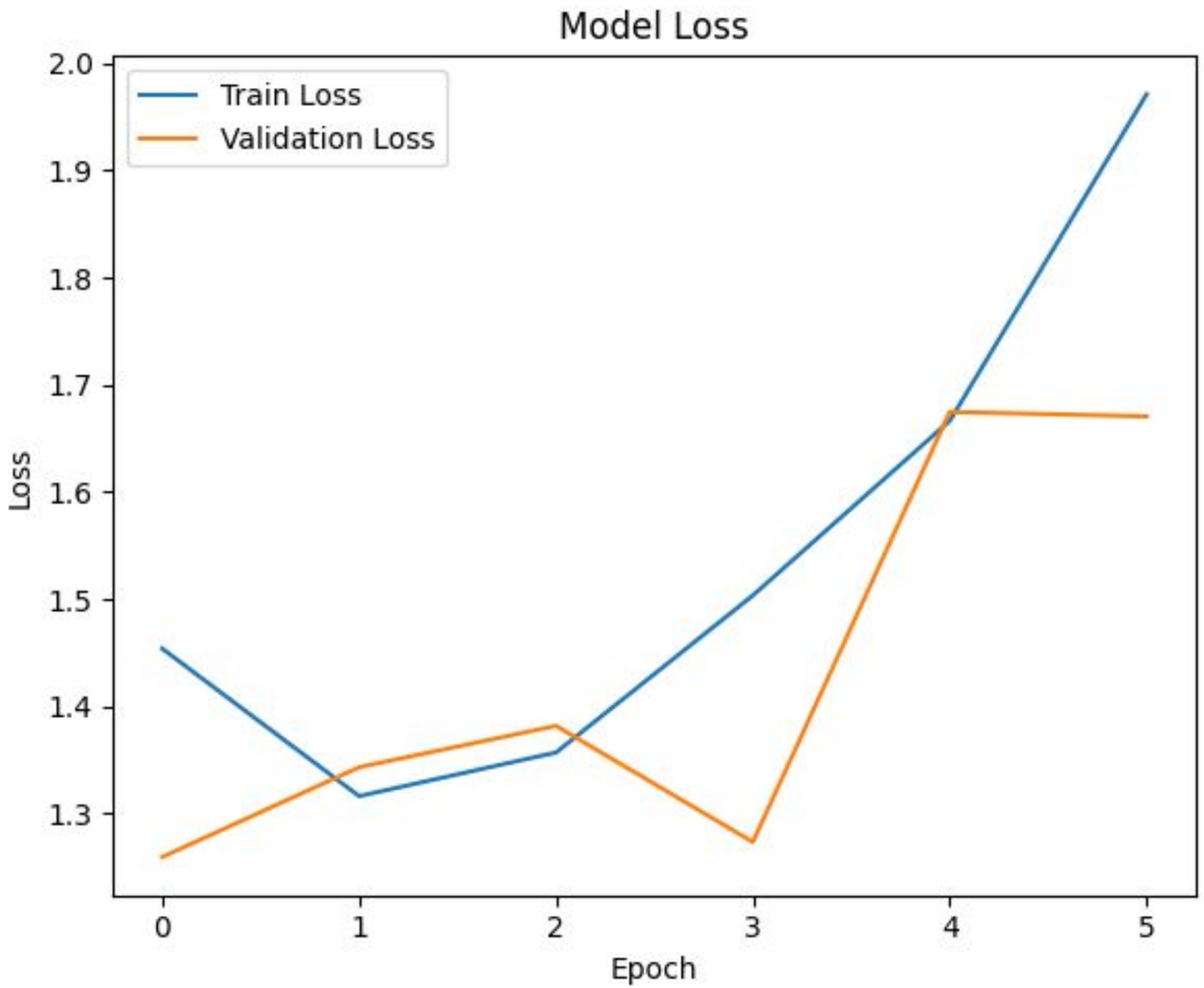

Model Loss
Train Loss
Validation Loss
Loss
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0
10
20
30
40
Epoch

LSTM + 1D CNN

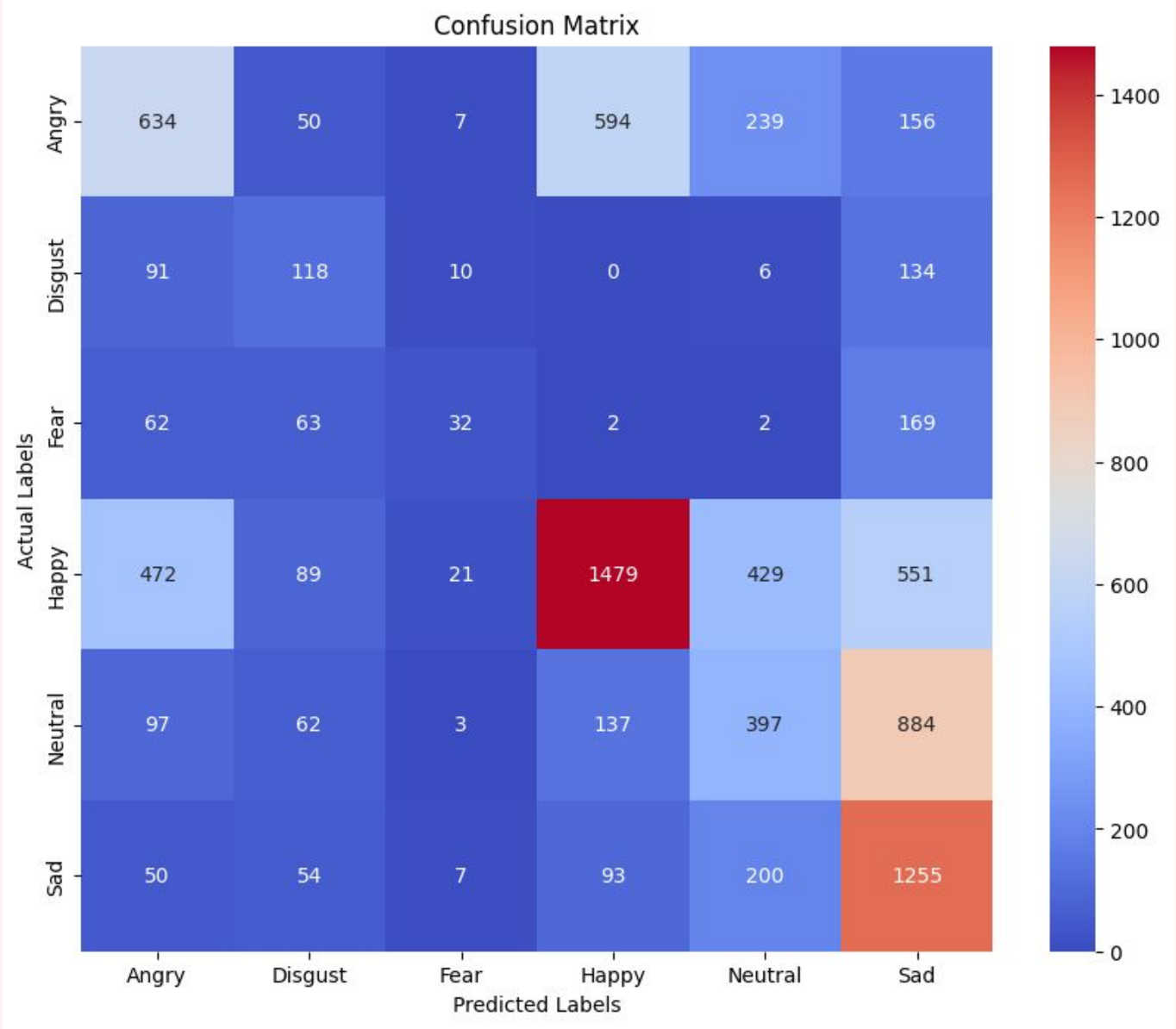

Confusion Matrix
Angry
Disgust
Fear
Happy
Neutral
Sad
826 57 93 475 136 67
16 203 115 0 0 0
12 61 259 1 0 1
468 118 158 1811 335 165
159 129 110 147 862 211
59 83 176 142 256 938
Actual Labels
Angry
Disgust
Fear
Happy
Neutral
Sad
Predicted Labels
1750
1500
1250
1000
750
500
250
0

LSTM + 2D CNN

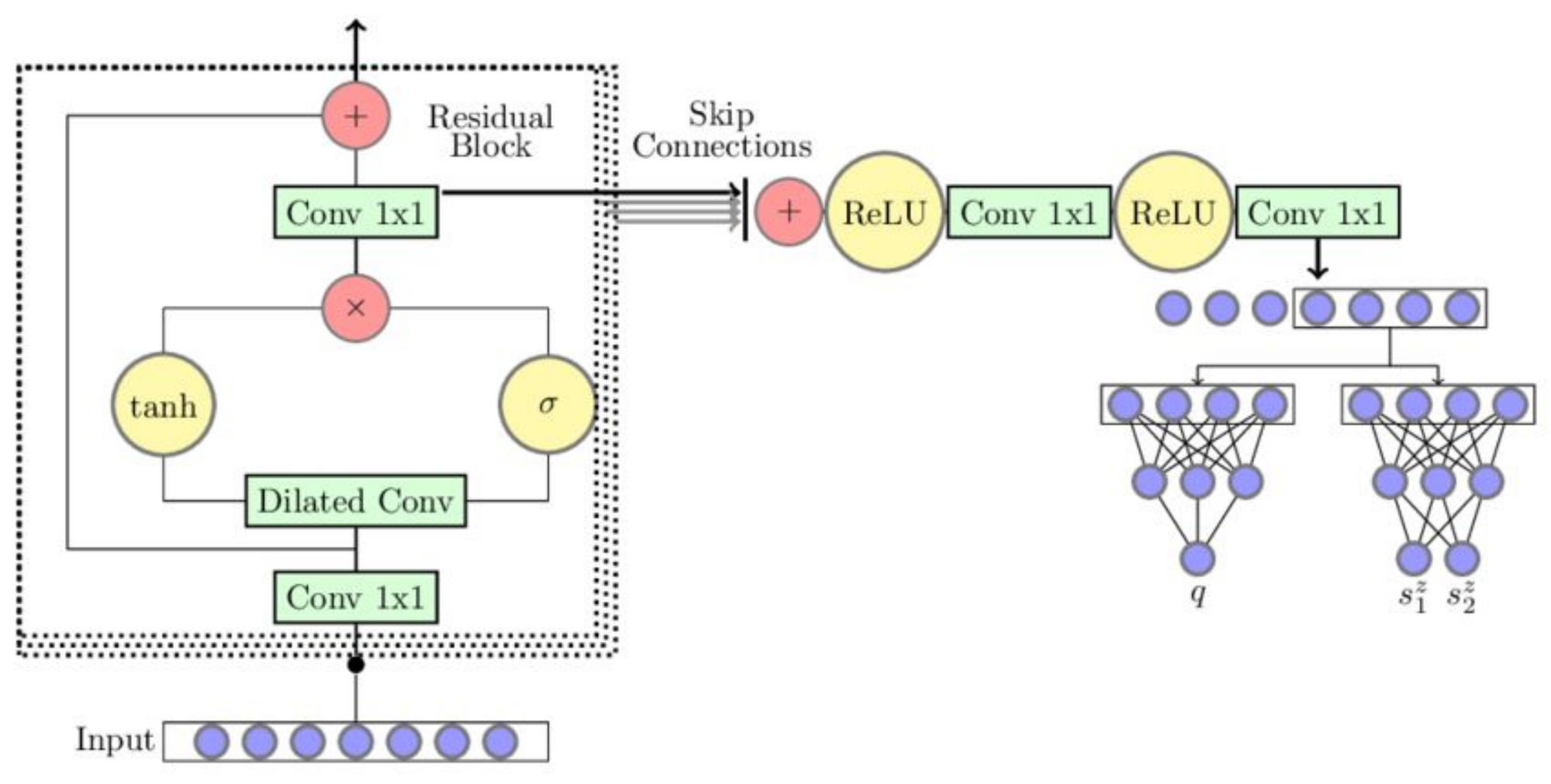

Model Accuracy
Train Accuracy
Validation Accuracy
Accuracy
0.44
0.42
0.40
0.38
0.36
0.34
0
1
2
3
4
5
Epoch


Model Loss
Train Loss
Validation Loss
Loss
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
0
1
2
3
4
5
Epoch

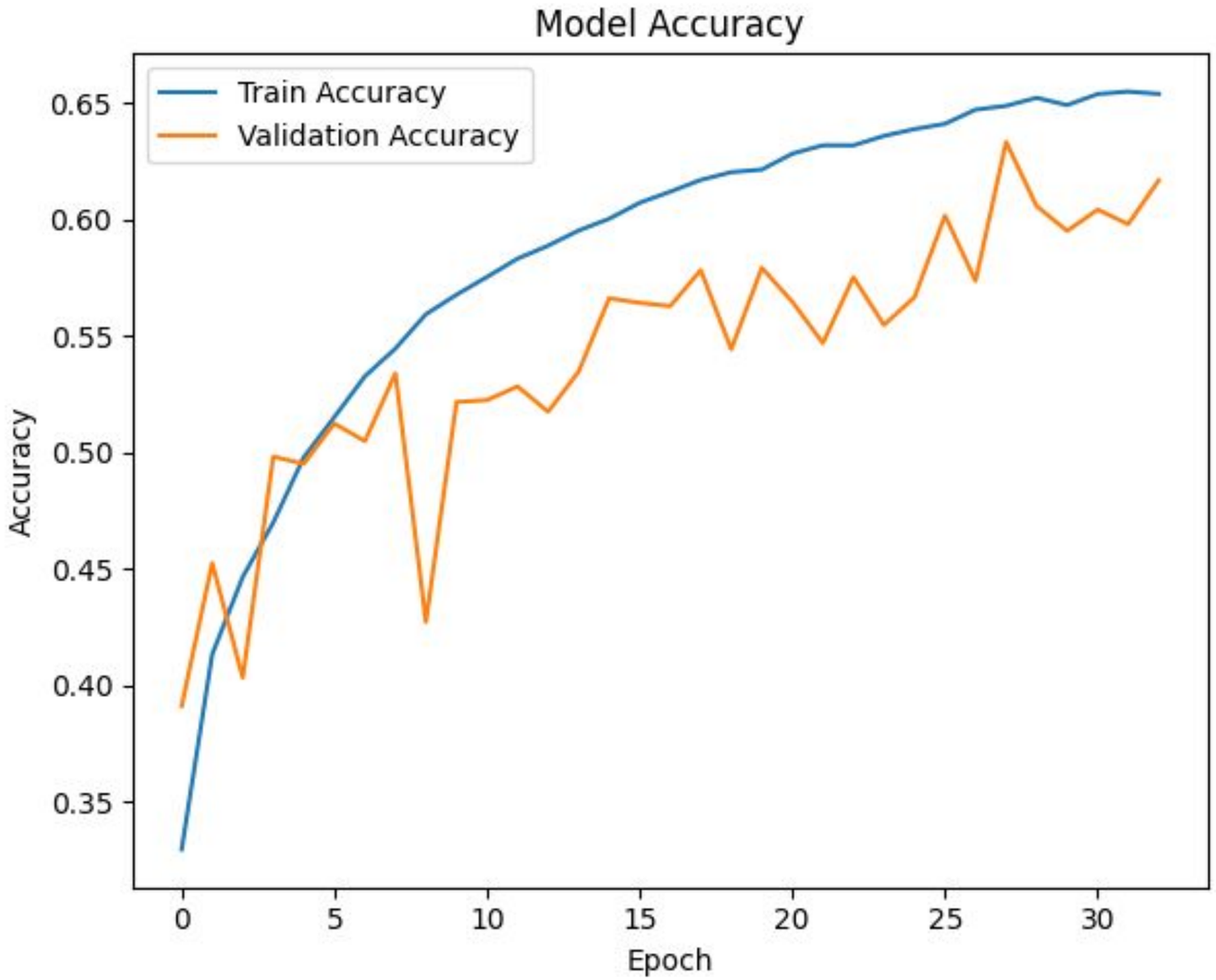
LSTM + 2D CNN

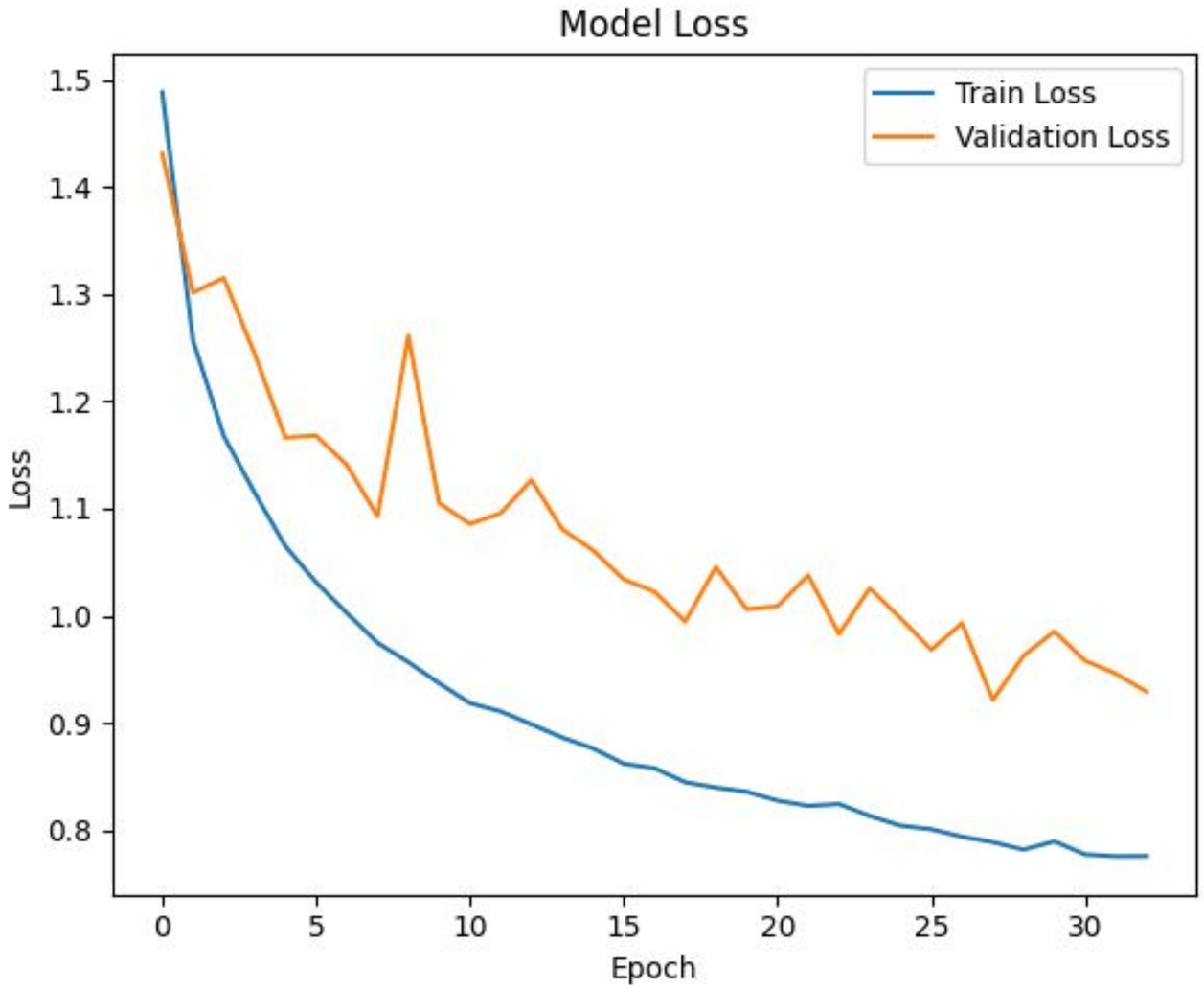

Confusion Matrix
Actual Labels
Angry
Disgust
Fear
Happy
Neutral
Sad
634 50 7 594 239 156
91 118 10 0 6 134
62 63 32 2 2 169
472 89 21 1479 429 551
97 62 3 137 397 884
50 54 7 93 200 1255
Angry
Disgust
Fear
Happy
Neutral
Sad
Predicted Labels
1400
1200
1000
800
600
400
200
0

WAVENET - MODEL ARCHITECTURE (CHAMPION MODEL)



WAVENET - ACCURACY/LOSS (CHAMPION MODEL)





WAVENET (CHAMPION MODEL)