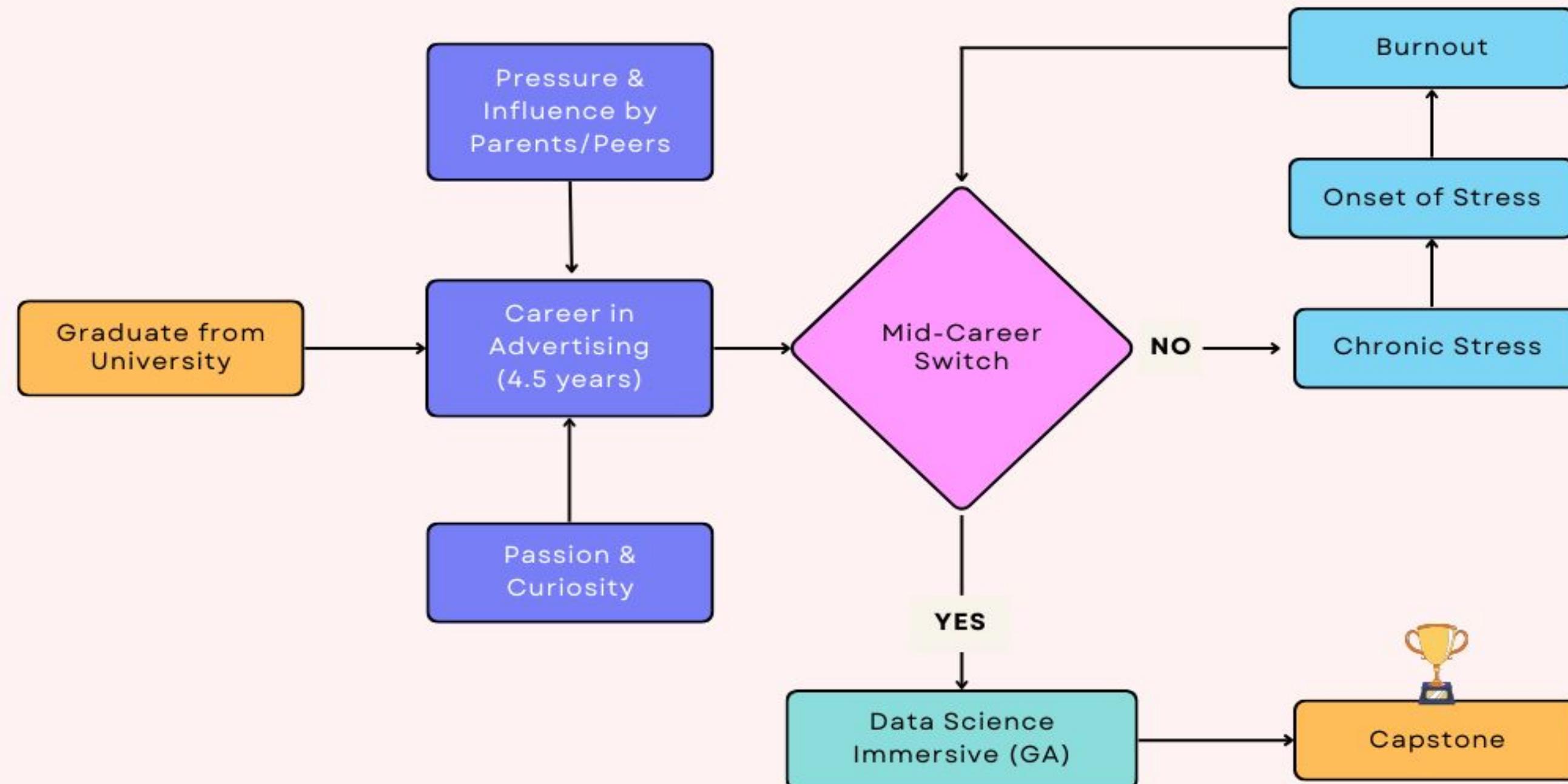
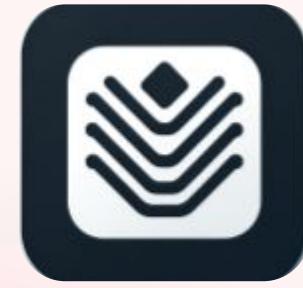


MY DSI JOURNEY





FeelFlow AI

Decoding Emotions, Advancing Patient Support

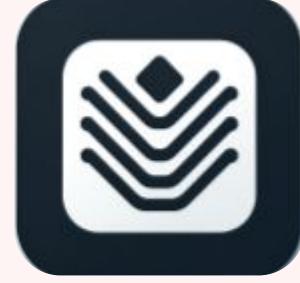
Capstone Presentation - DSI-SG-42

14th May 2024

Amoz Kuang

AGENDA

- 01 - Introduction
- 02 - Context + Problem Statement
- 03 - Methodology
- 04 - Insights
- 05 - Predictions
- 06 - Demo
- 07 - Conclusions

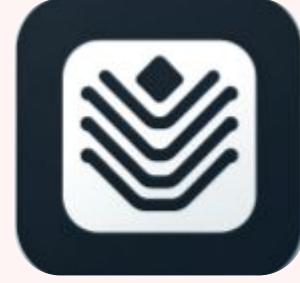


01 INTRODUCTION

SPEECH EMOTION RECOGNITION



Speech Emotion Recognition (SER) in therapy refers to the use of technology to analyze 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 like **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)

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.

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

Related Topics

MOH mental health

ADVERTISEMENT

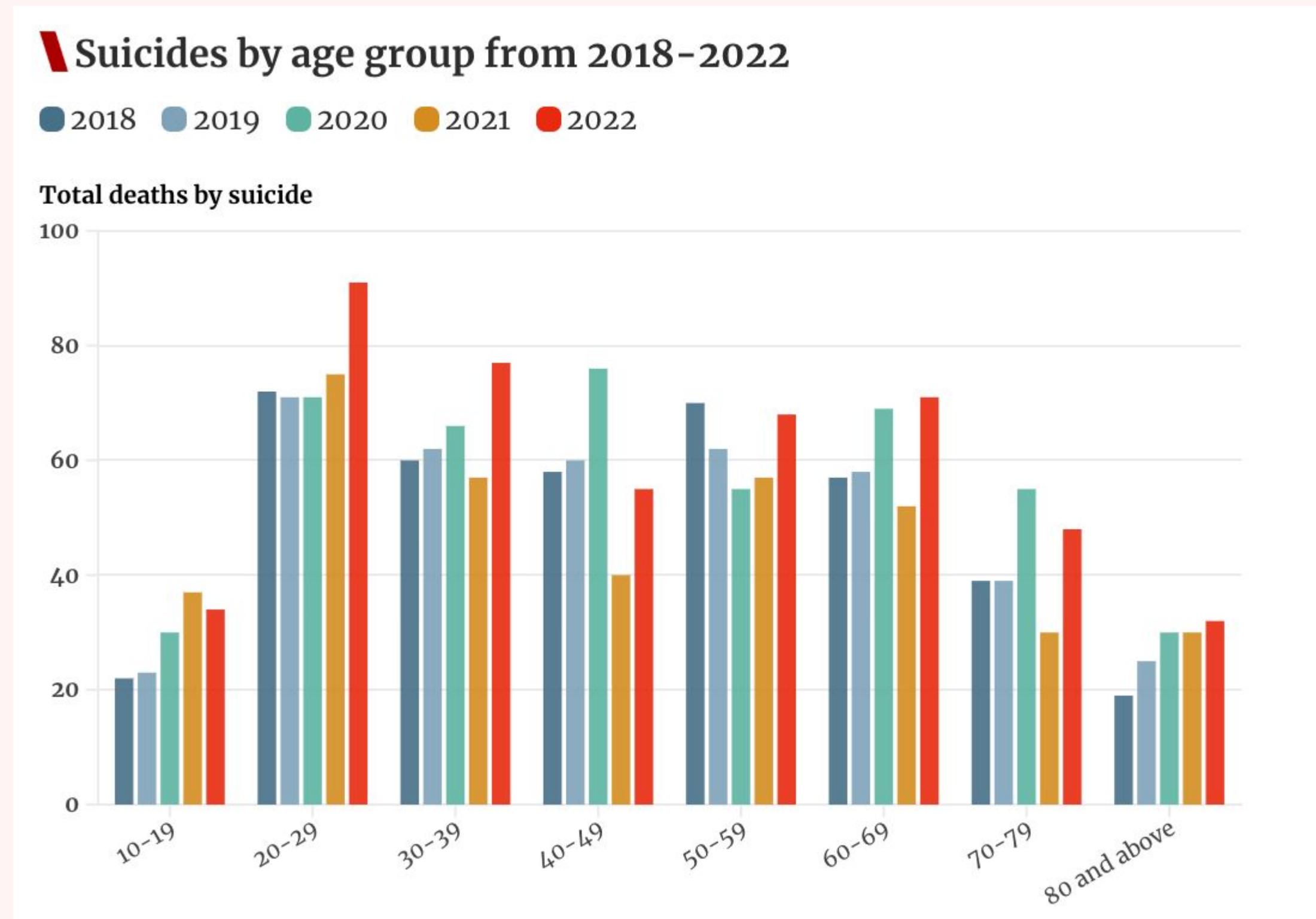
Data was collected from around 8,000 adults through self-reported household interviews and about 9,000 adults through health examinations.

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.

EFFECTS OF MENTAL HEALTH



'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)

- 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".

Goh Chiew Tong
@ChiewTongCNA

Christy Yip
@ChristyYipCNA

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

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"."

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
Christy Yip
Ryan Tan Bing Yang

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

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

THE 'FEELFLOW' WAY (OBJECTIVE)

Beneficiaries
- YOU
- YOUR
Company

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

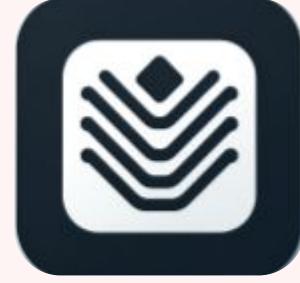
- Has been out of touch with youth lingo, behaviours, knacks, and feels like it might be hard to reach his audience.

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

PROJECT SCOPE (MULTI-PHASED)

WE'RE HERE!!!



FeelFlow AI [**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 package for Application

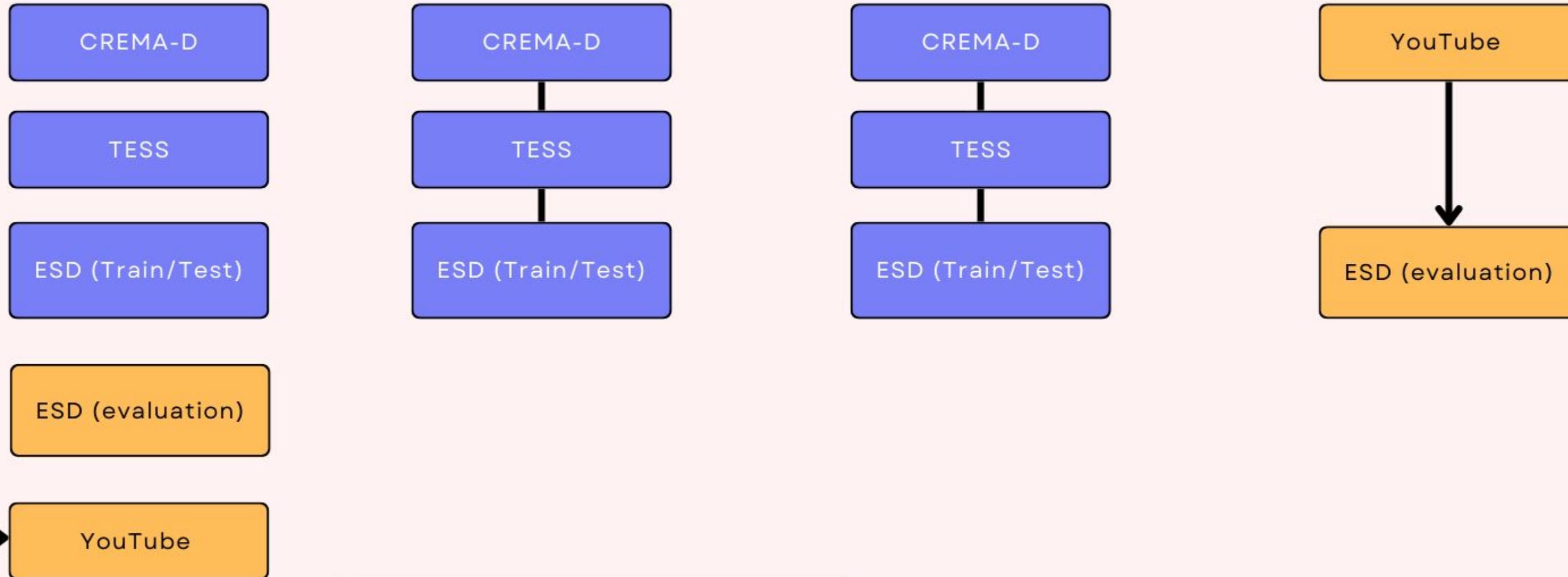
FeelFlow AI 2.0 [**BEYOND CAPSTONE **]

- 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.

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.

Transfer Learning (Seen --> Unseen)

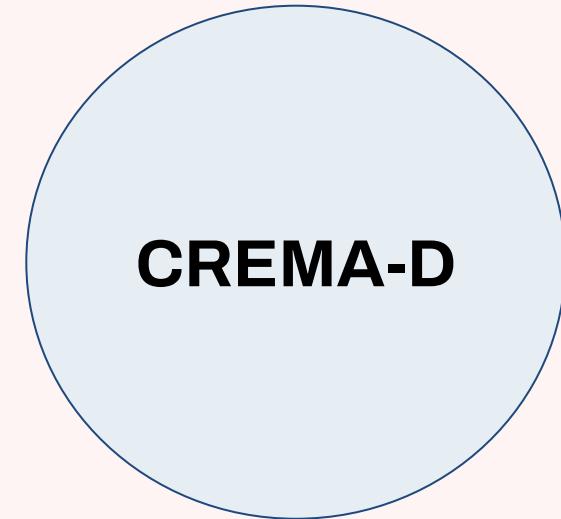


SCRAPED

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 (ES)

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

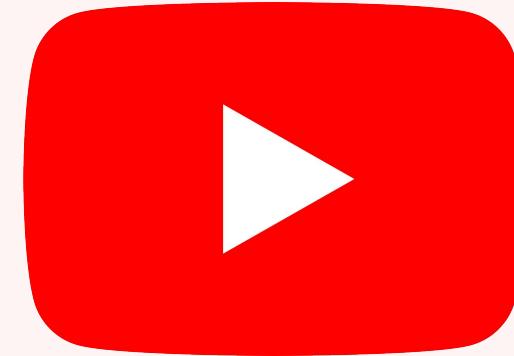
Data includes 350 parallel utterances:

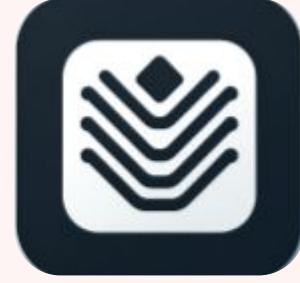
- 10 native Mandarin speakers, 10 English speakers
- 5 emotions
(neutral, happy, angry, sad, and surprised)
- Data is split into training, test, and **evaluation**

YouTube Data

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

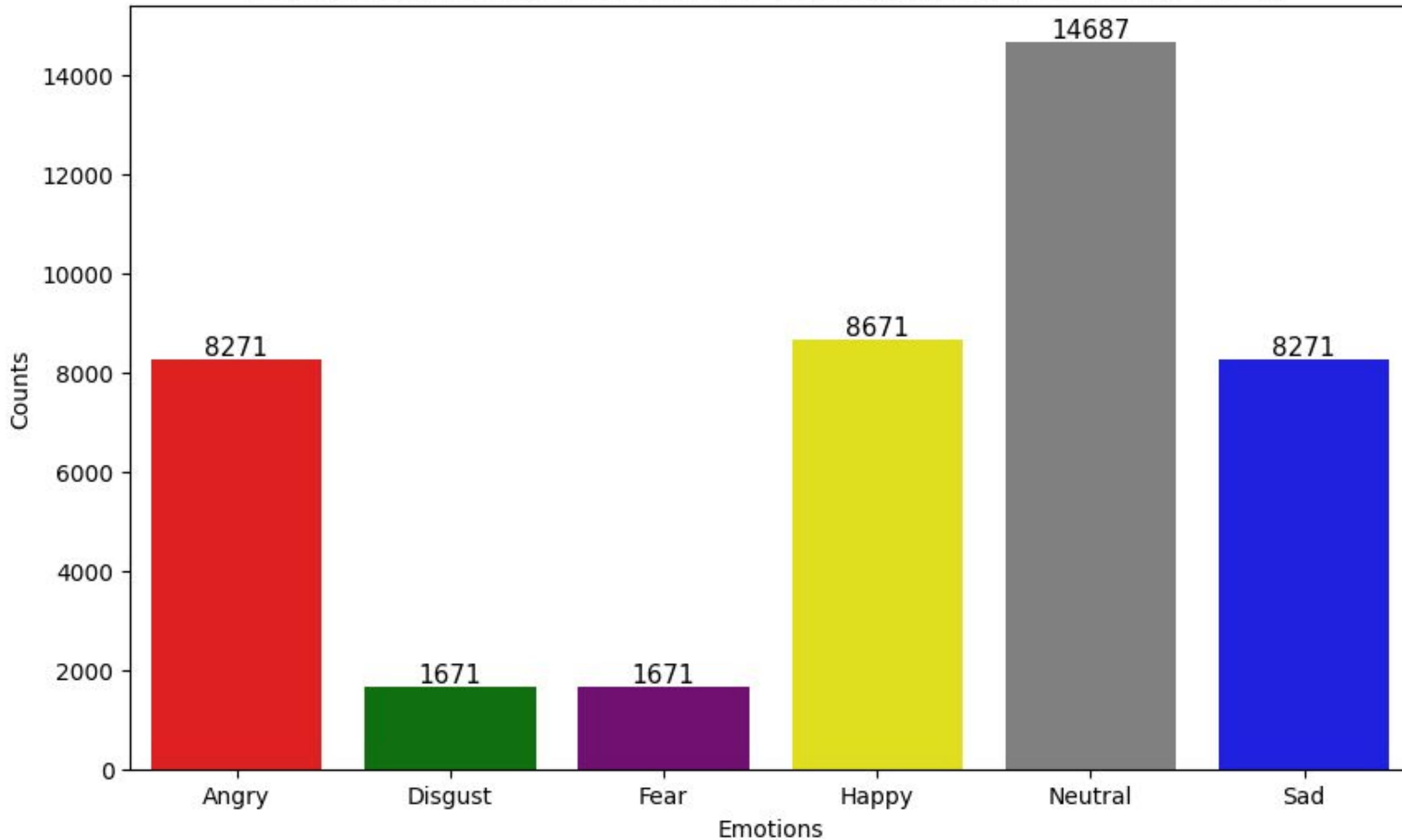
- Links found in Notebook





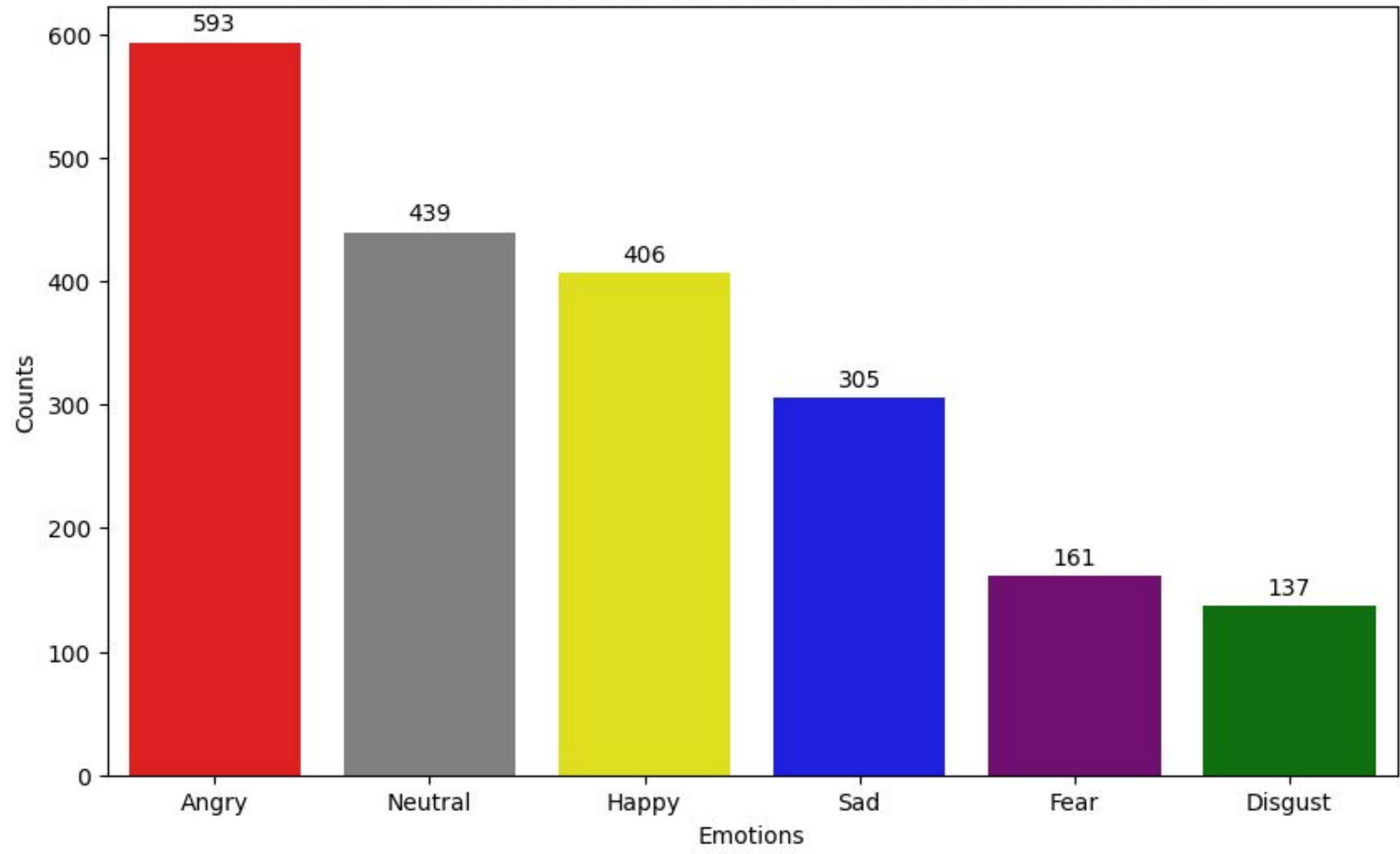
04 INSIGHTS

Distribution of Emotions (Combined Training-Test)



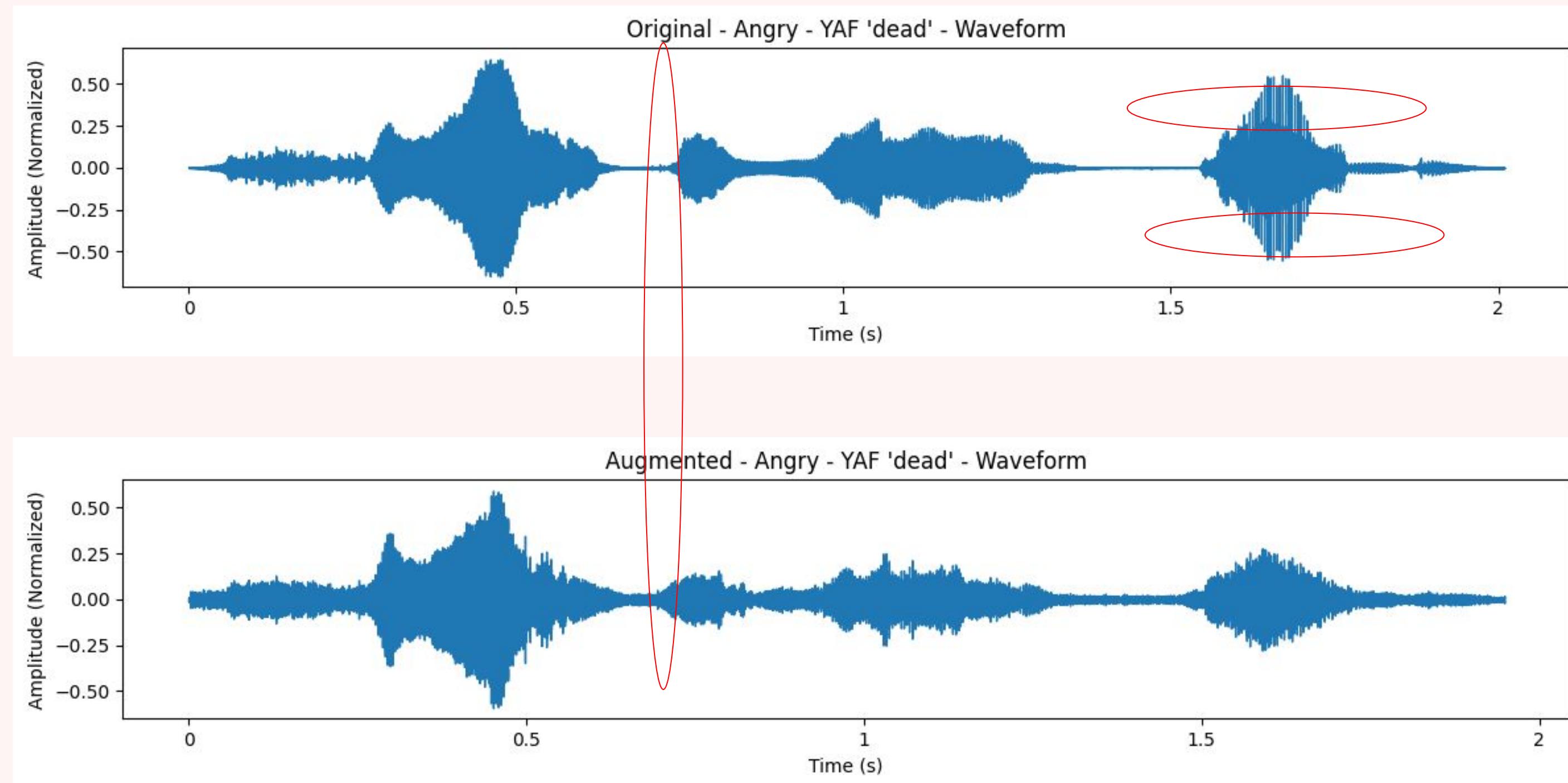
- ‘Disgust’ and ‘Fear’ are the minority class in the training-test dataset
- Rationale:
 - ‘Surprise’ emotion in ESD - mapped as ‘Neutral’ (~6.6k)

Distribution of Predicted Emotions (YouTube)

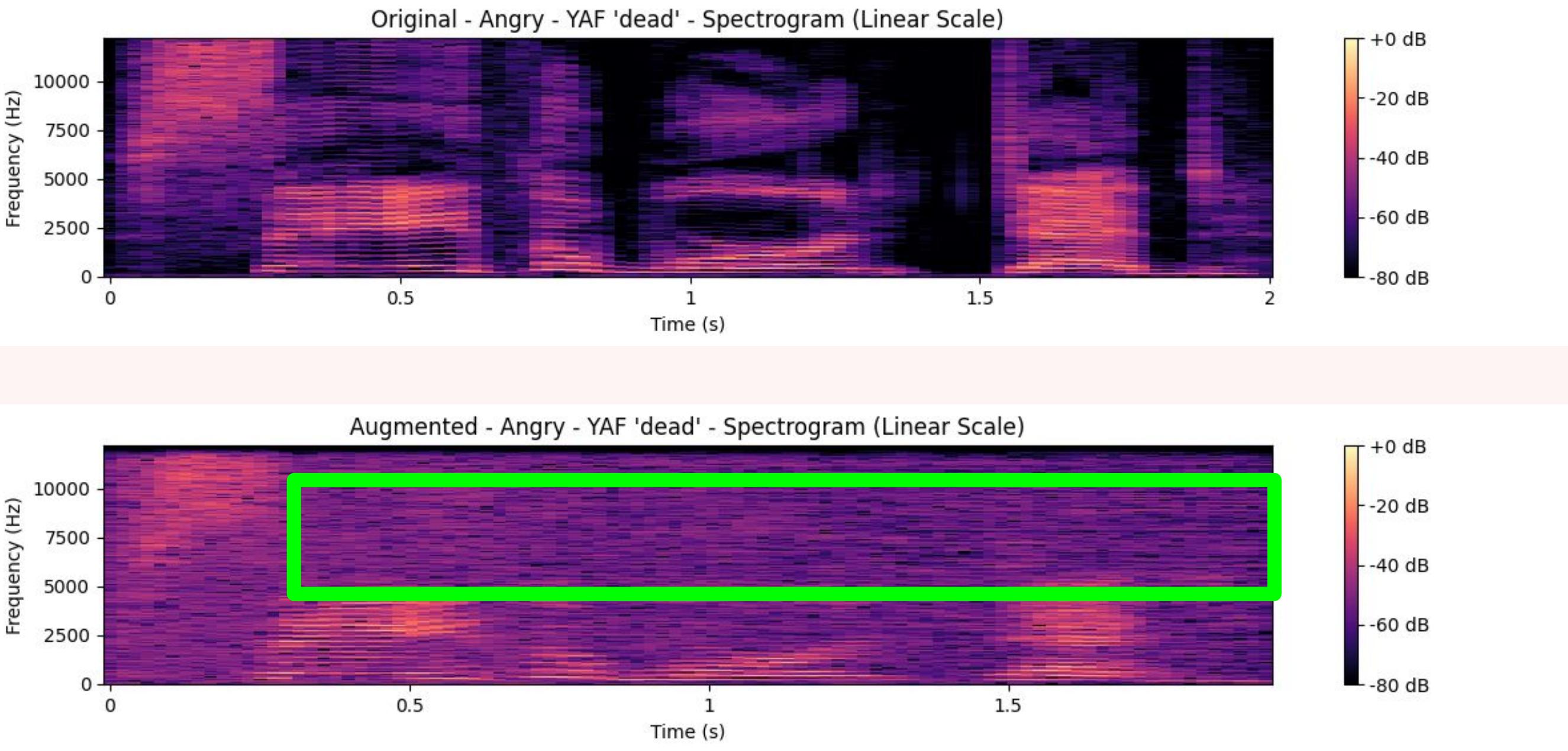


- Expected to have 'Fear' and 'Disgust' as the least distributed predicted emotions due to the imbalance trained and tested dataset - even after more class weights importance
- Overall, more balanced prediction than the train-test data
- Further validation is required from ESD (eval) dataset

EFFECTS OF AUGMENTATION - ANGRY

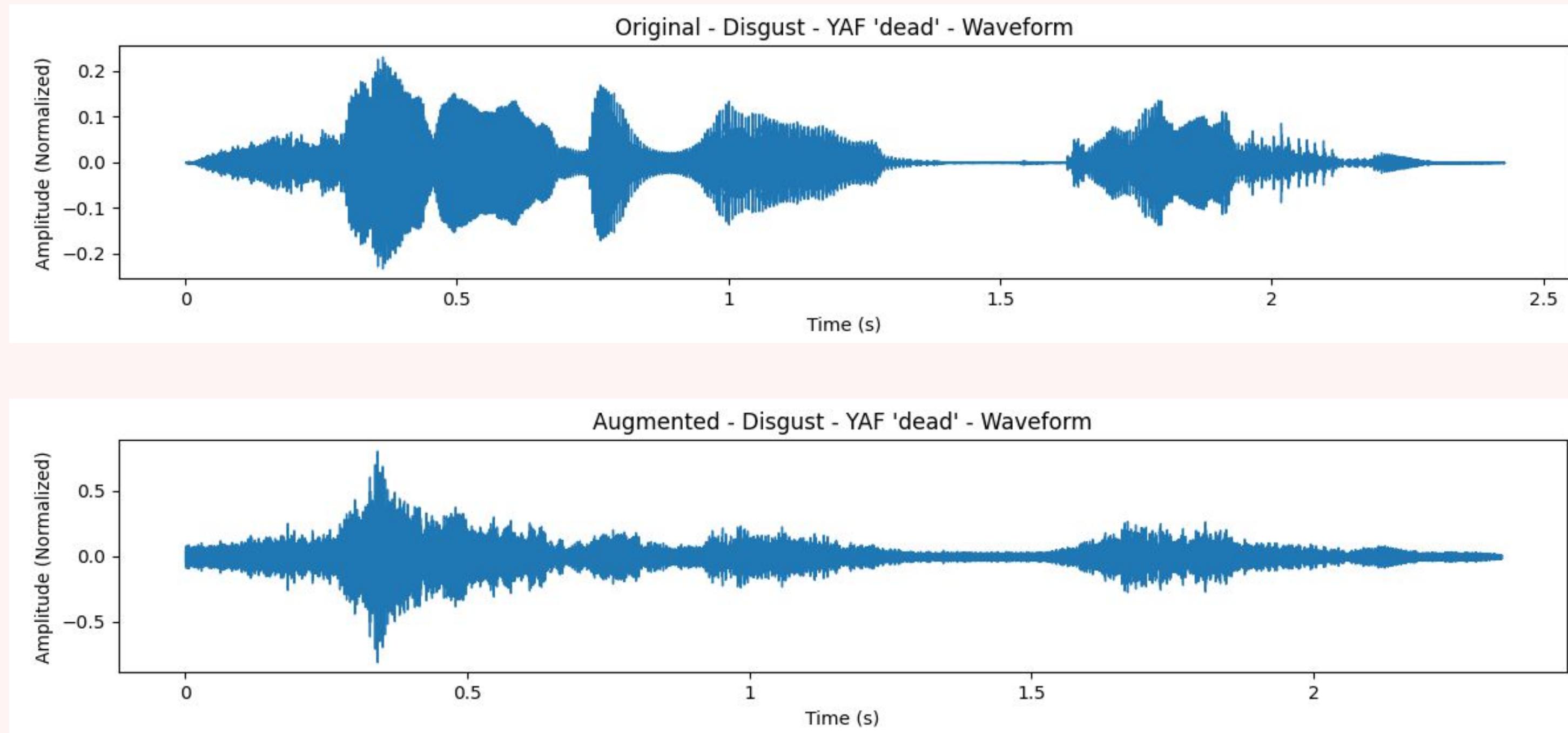


LINEAR SPECTROGRAM - ANGRY

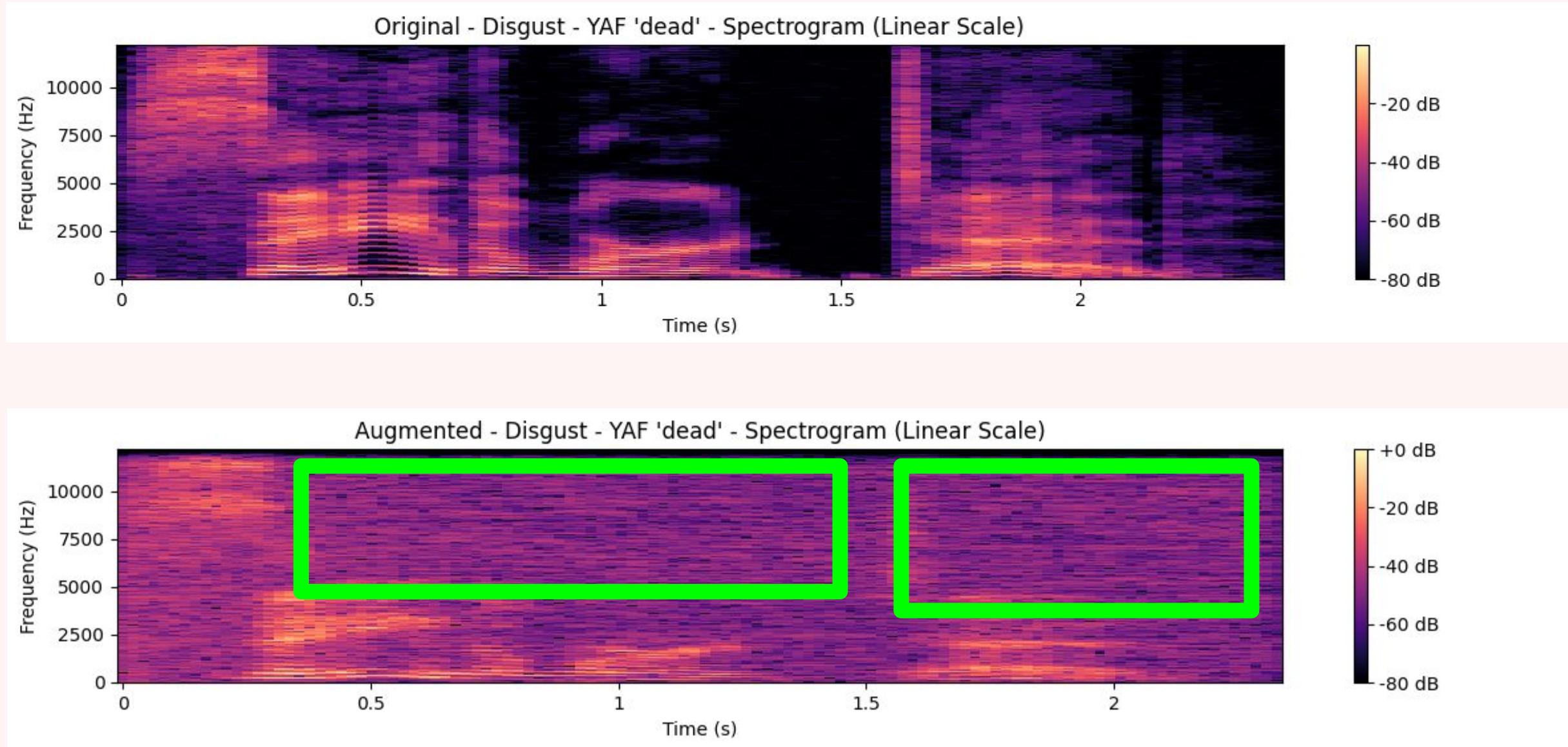


- “Blendedness” is referred to as smoothening
- This makes the model less “sensitive” to spikes in volume and/or pitch especially when someone is angry

EFFECTS OF AUGMENTATION - DISGUST

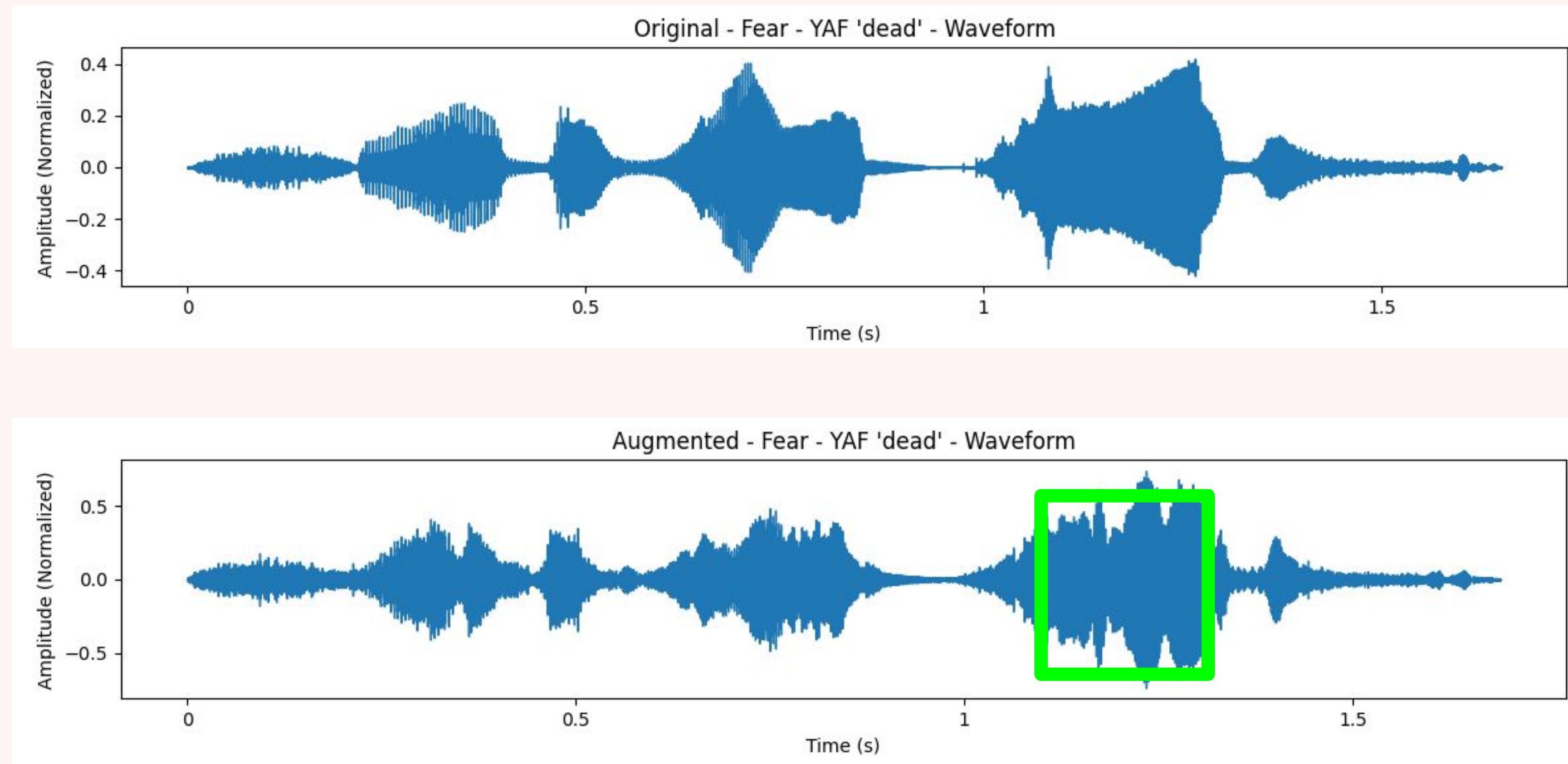


LINEAR SPECTROGRAM - DISGUST

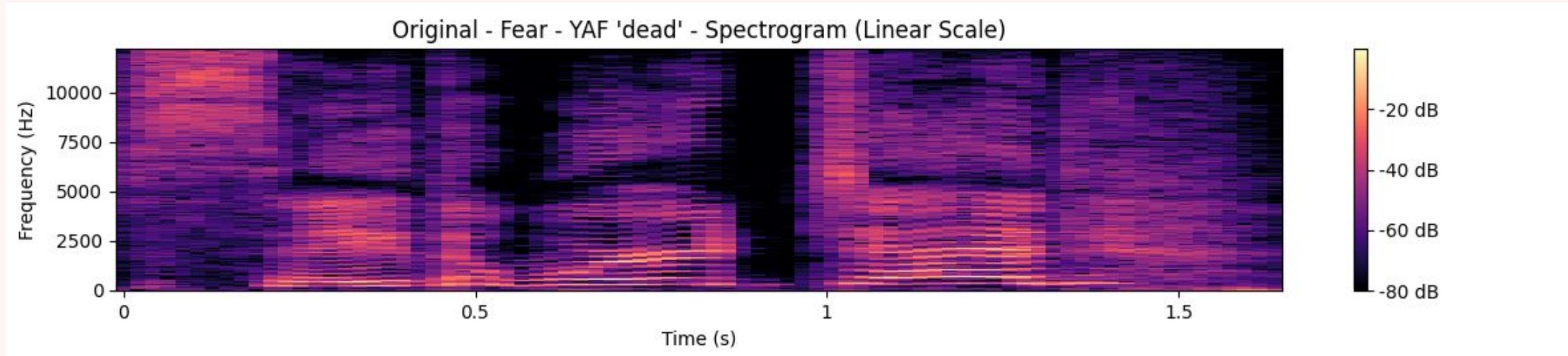


- The intensity of the Augmentation effect here is similar to the one experienced in ‘Angry’

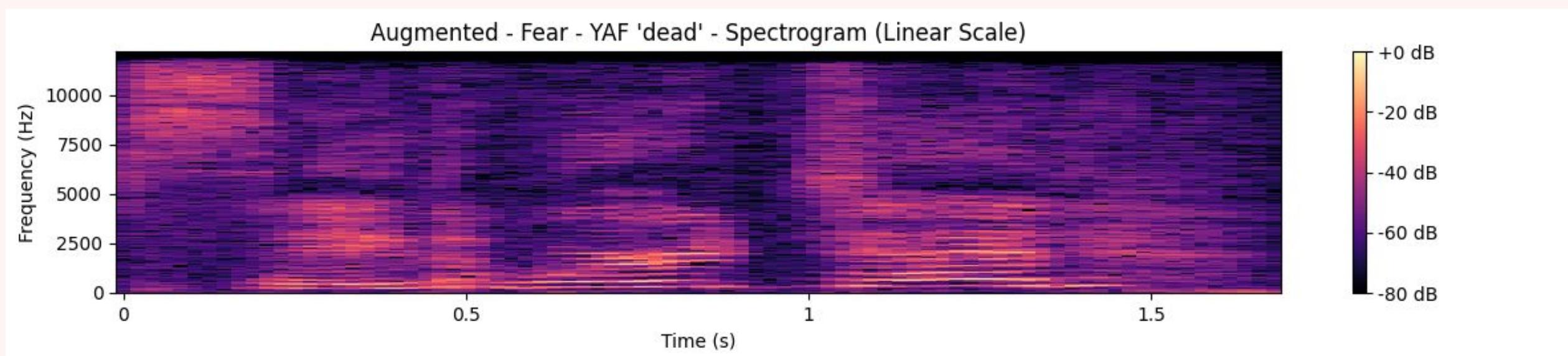
EFFECTS OF AUGMENTATION - FEAR



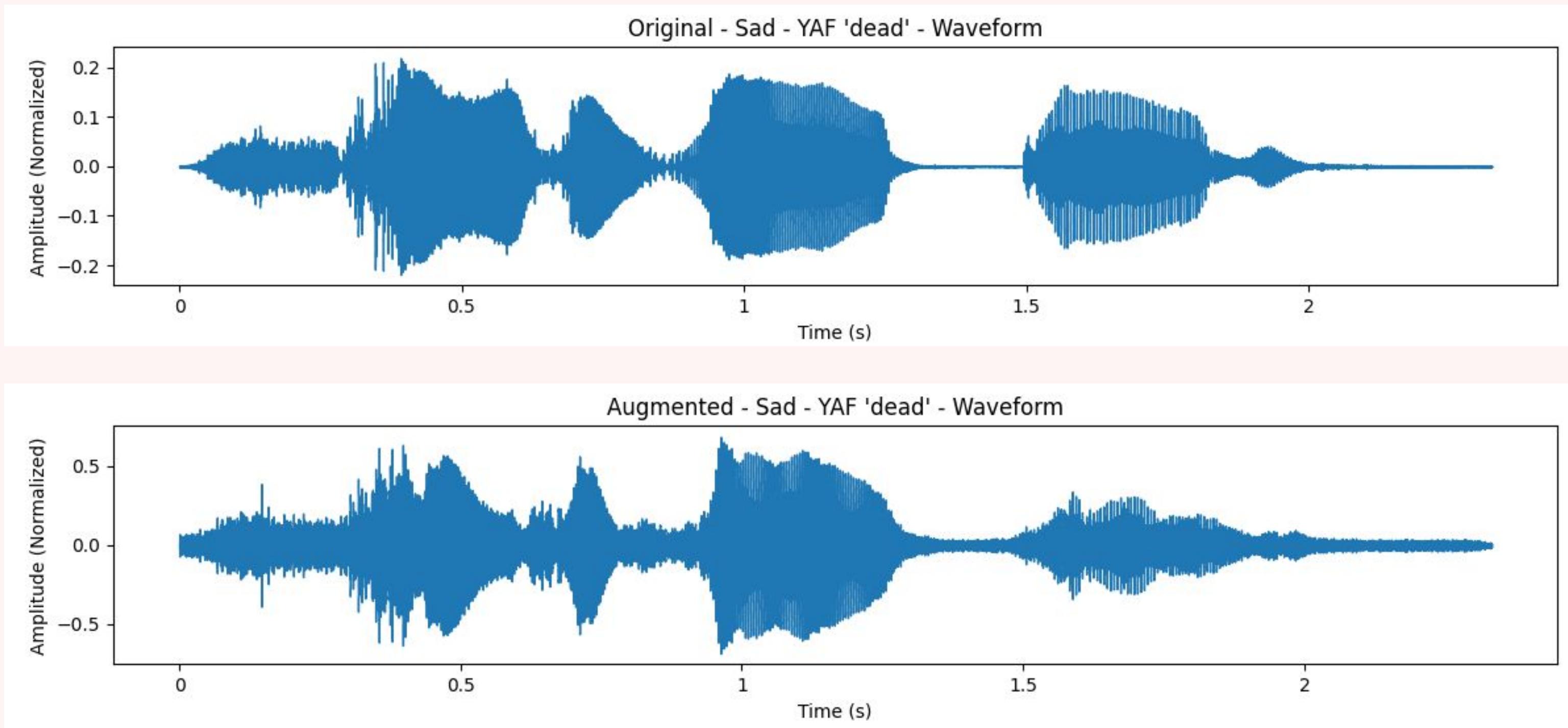
LINEAR SPECTROGRAM - FEAR



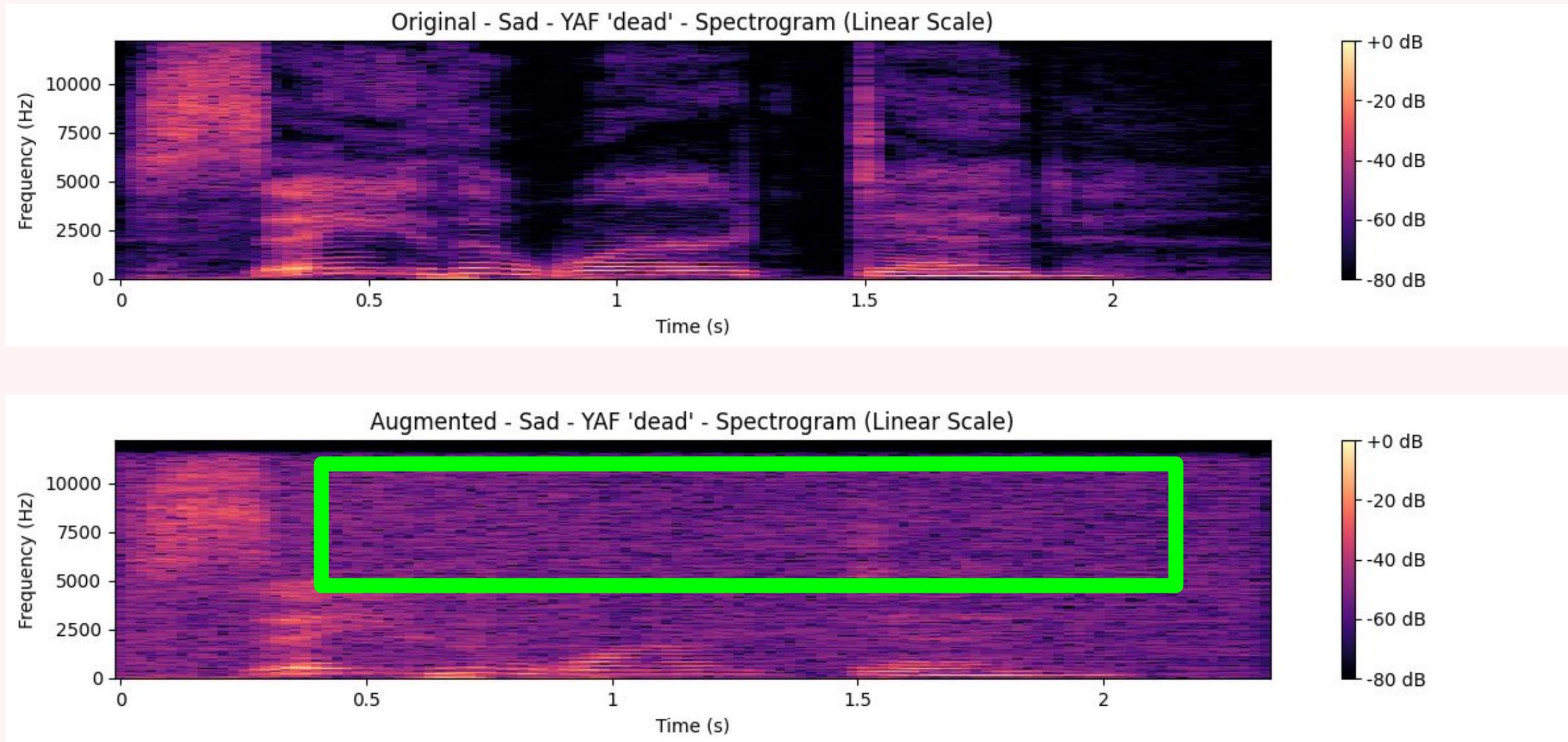
- Augmentation here has little to no effect as compared to 'Angry' and 'Disgust' emotions.



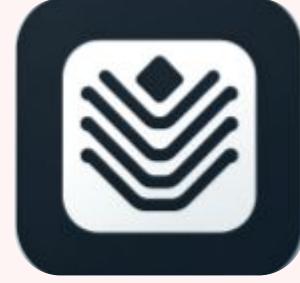
EFFECTS OF AUGMENTATION - SAD



LINEAR SPECTROGRAM - SAD



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



05 PREDICTIONS

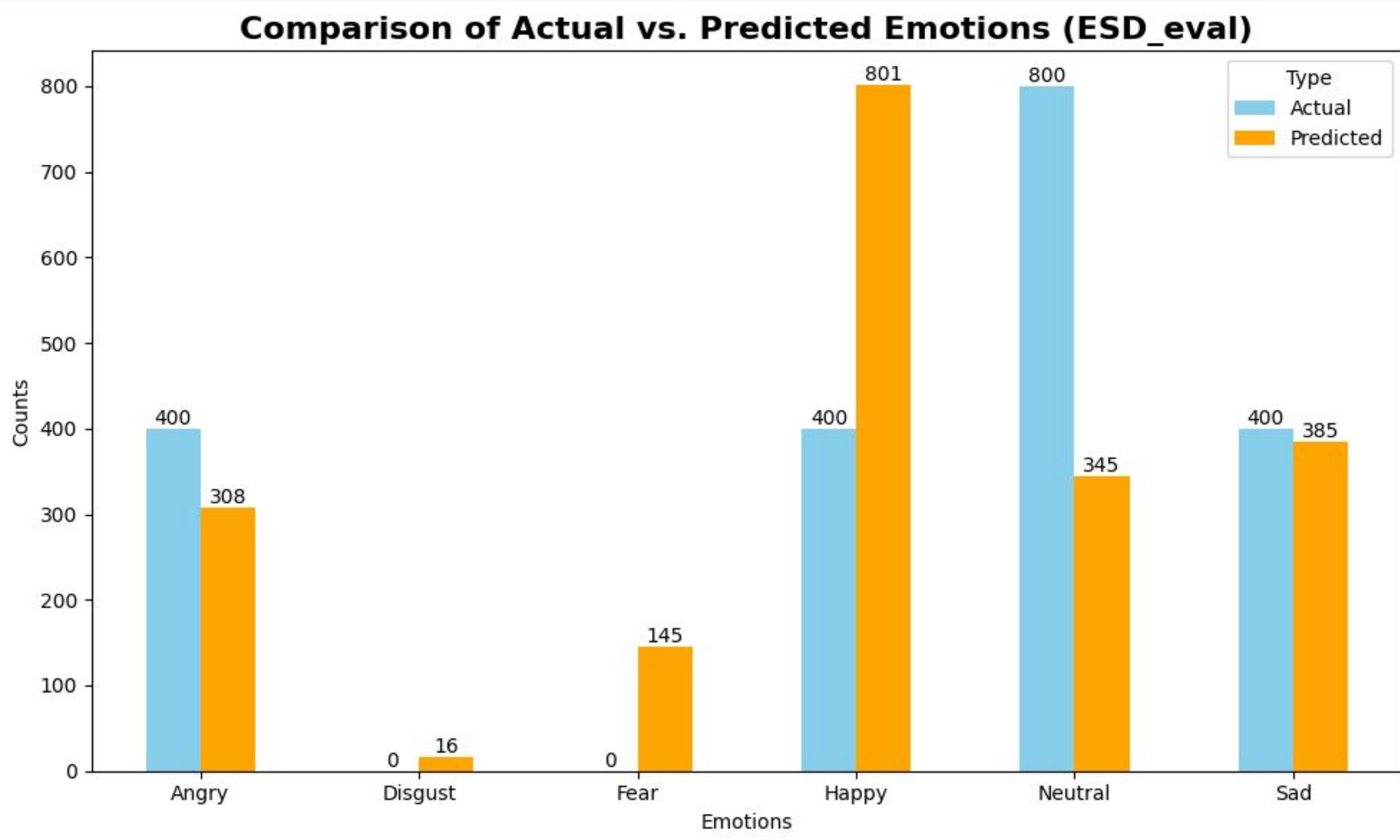
MODEL EVALUATION

| 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 |
| 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

VALIDATION (ESD_EVAL)

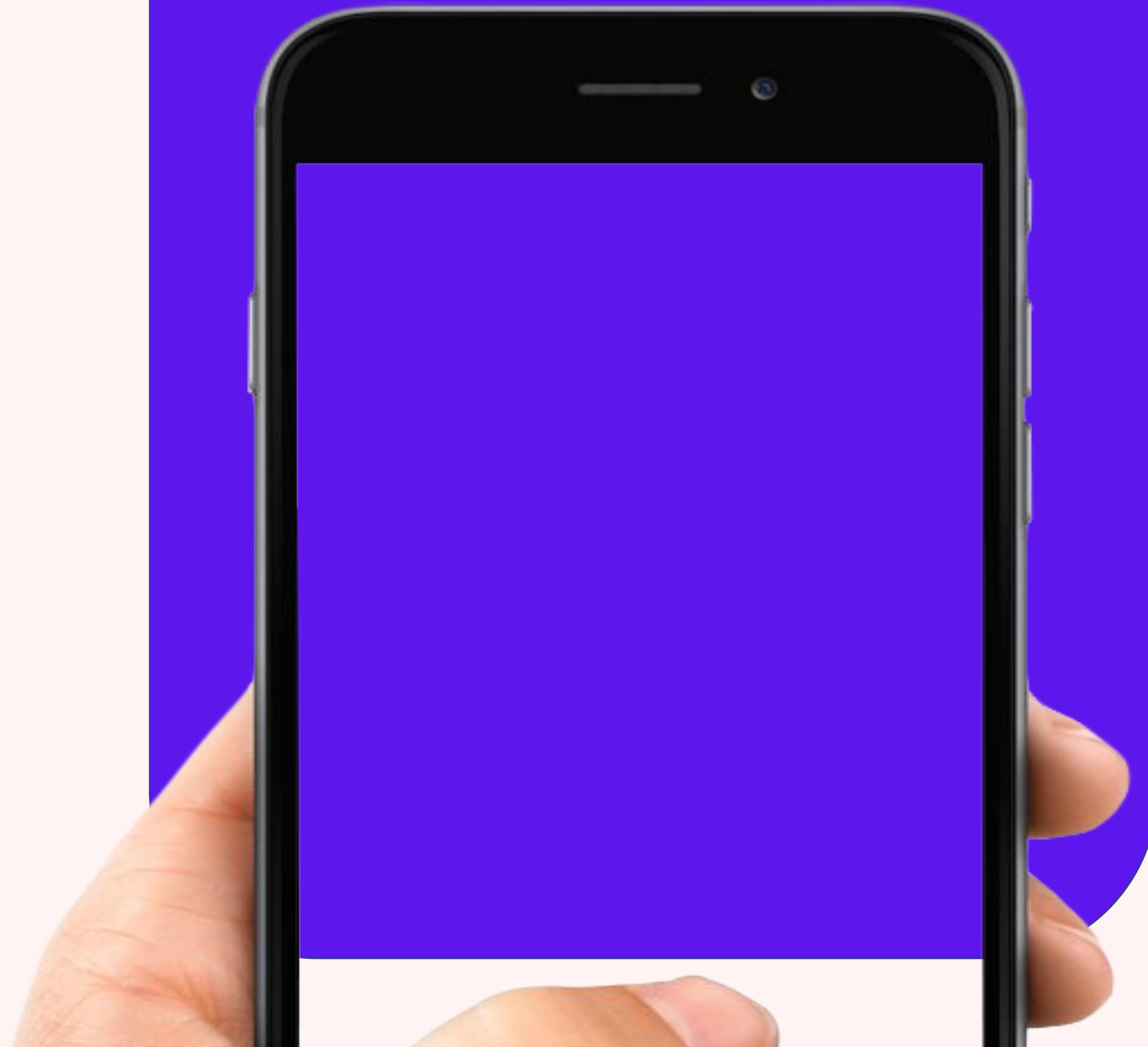


- Negative Emotions
 - Overprediction on 'Disgust' and 'Fear', which may unnecessarily trigger attention required.
 - Slight underprediction for, may require more training of data, and clinicians to administer personal discernment
- 'Happy' - overprediction, which may delivered false reassurance to the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



06 DEMO

SCAN HERE





07 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.

**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)



Cost

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



Cost

- $\sim \$435$ million (out of $\$16.68$ billion budget)

Benefit

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

Benefit

- Assuming 1% savings of allocated $\$435$ million budget to mental health (subsidies etc.)
 - $1\% \times \$16.68 \text{ billion} = \4.4 million

SUMMARY

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



Limitations of Study

Data Collection

- Culturally Nuanced (Other languages, beyond English)
- Generational Nuanced (Lingos used by GenZs)
- Does not account for accent

Emotional Overlap (e.g. Disappointment = Anger + Sadness?)

False Negative rates might consider metrics beyond Accuracy

Voice might not be able to sense contextual cues or things like sarcasm



Recommendations for Future

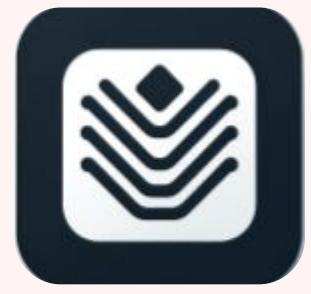
Collaborate with local Tertiary Institutions to collect and integrate data

Redefine emotional labels, and consider a wider but definitive spectrum

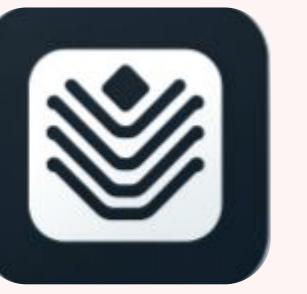
False Negative/Miss Rates
Emotional Detection Rate)

Incorporate other stimulus:

- Natural Language Processing of Text (Transcriptions)
- Facial Recognition (able to use EDR)

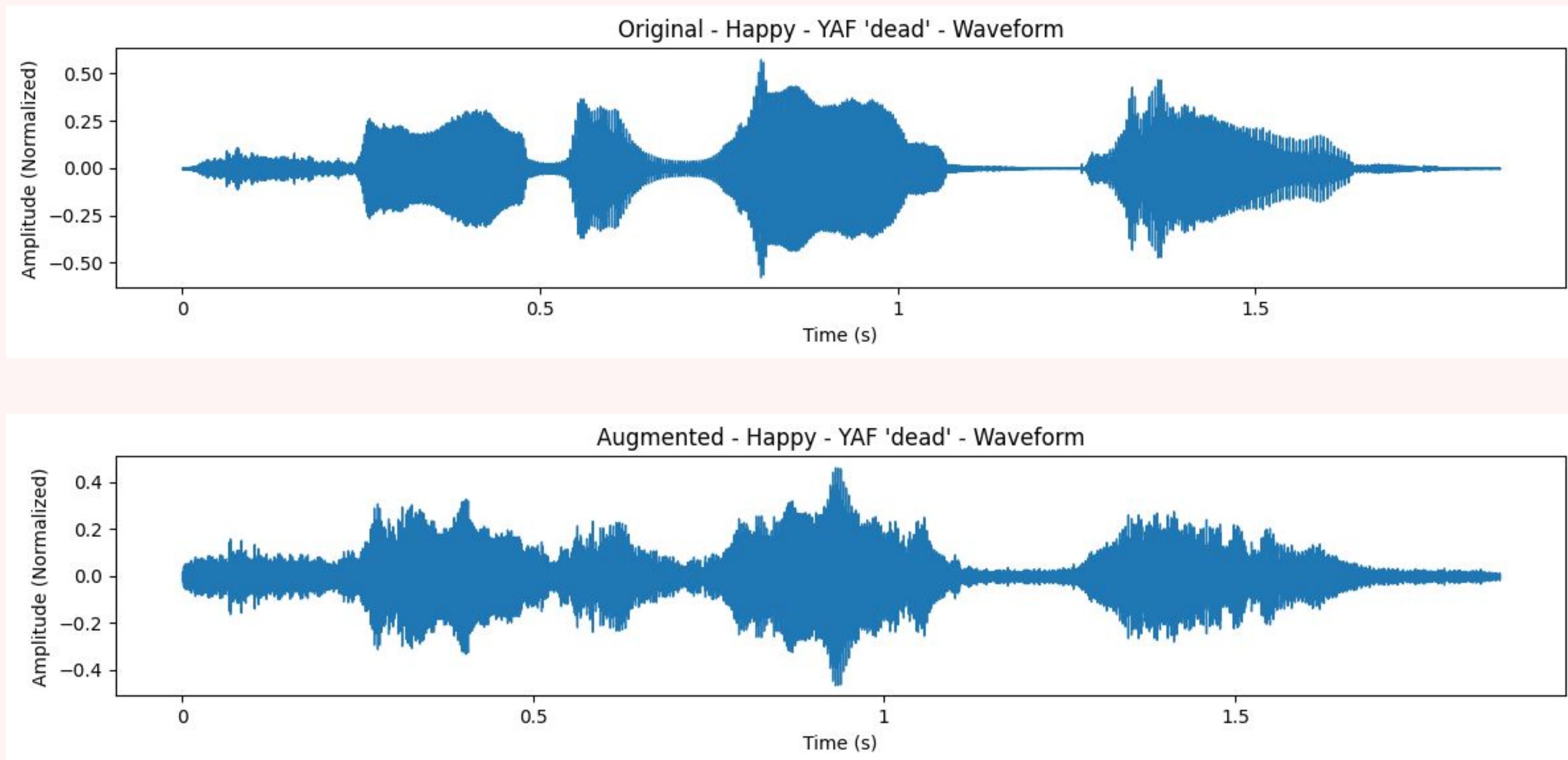


Thank You

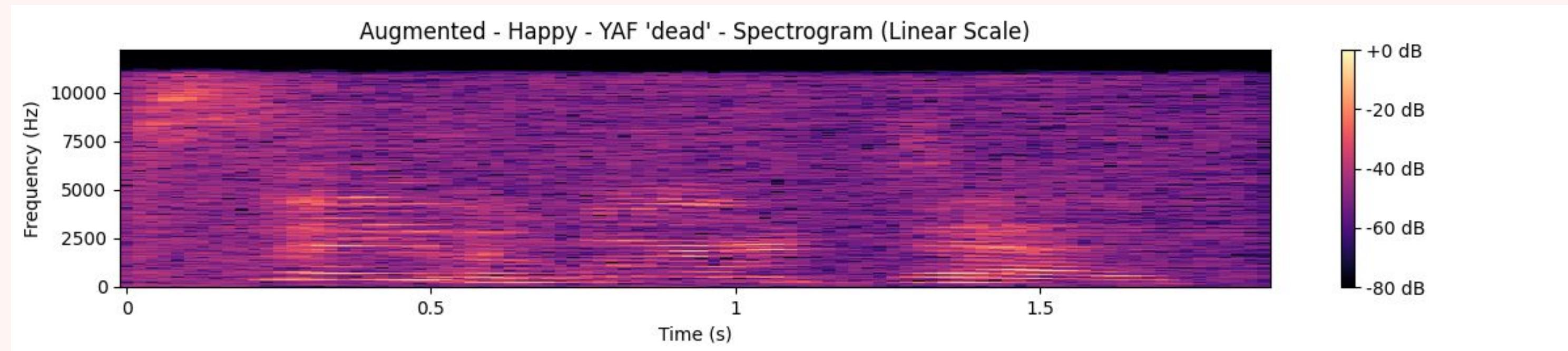
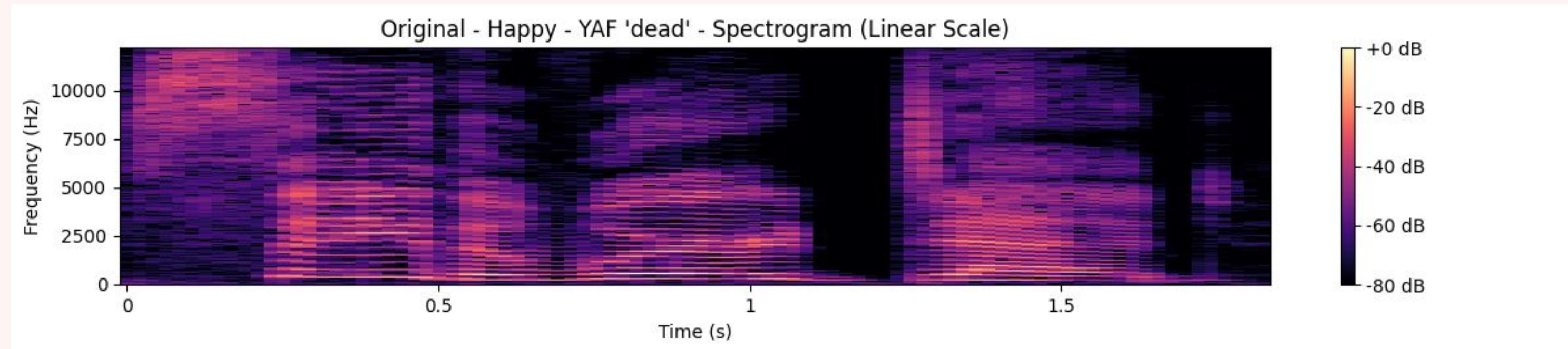


APPENDIX

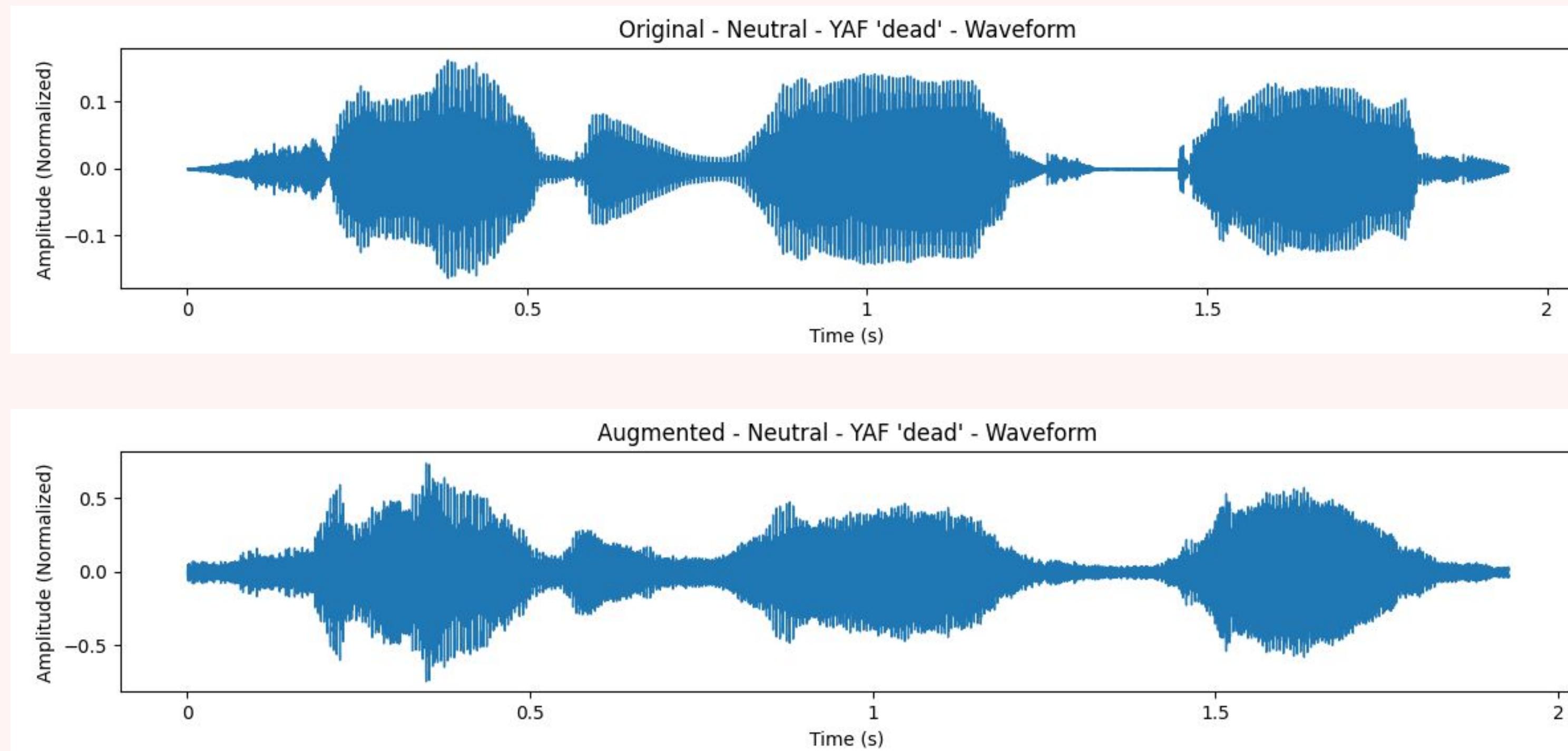
EFFECTS OF AUGMENTATION - HAPPY



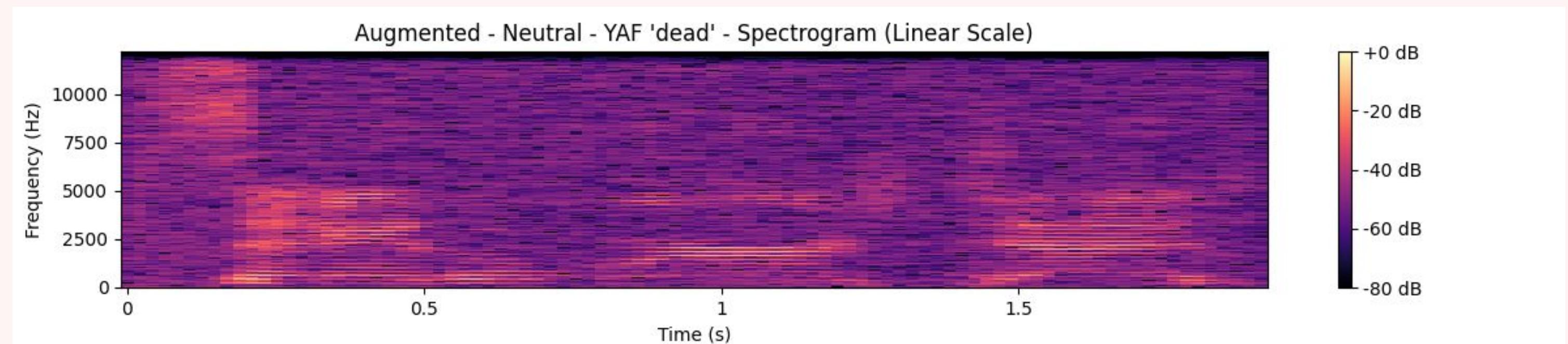
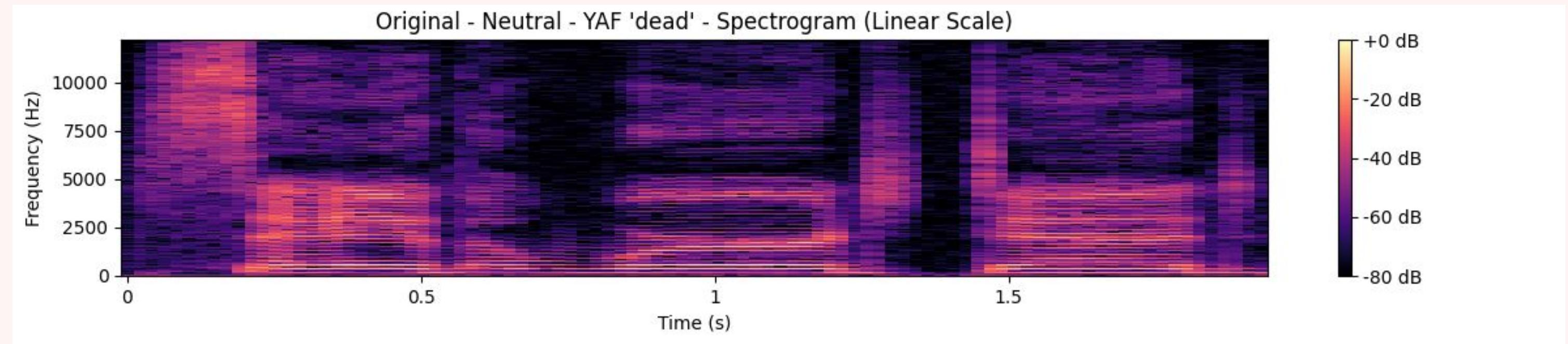
LINEAR SPECTROGRAM - HAPPY



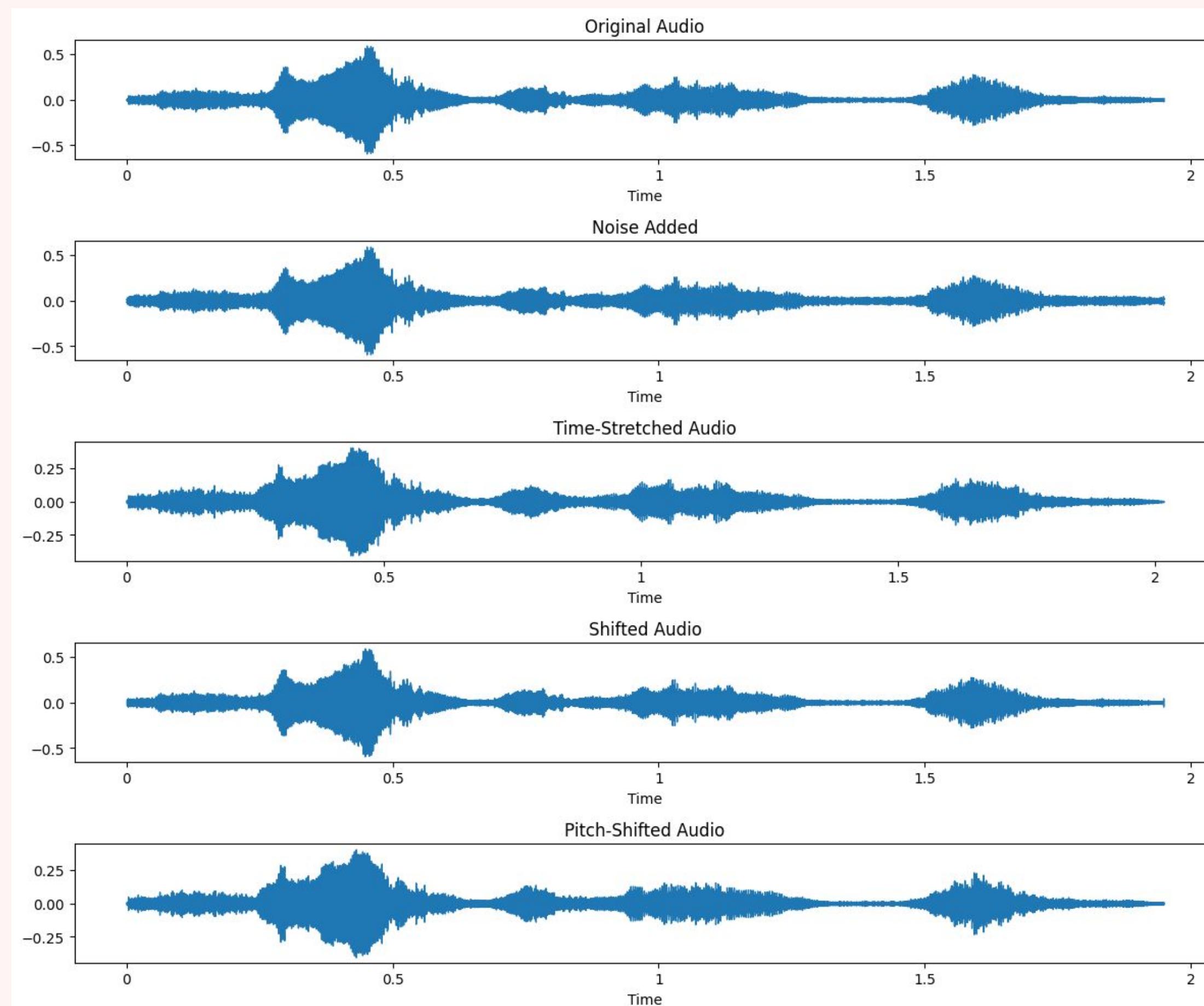
EFFECTS OF AUGMENTATION - NEUTRAL

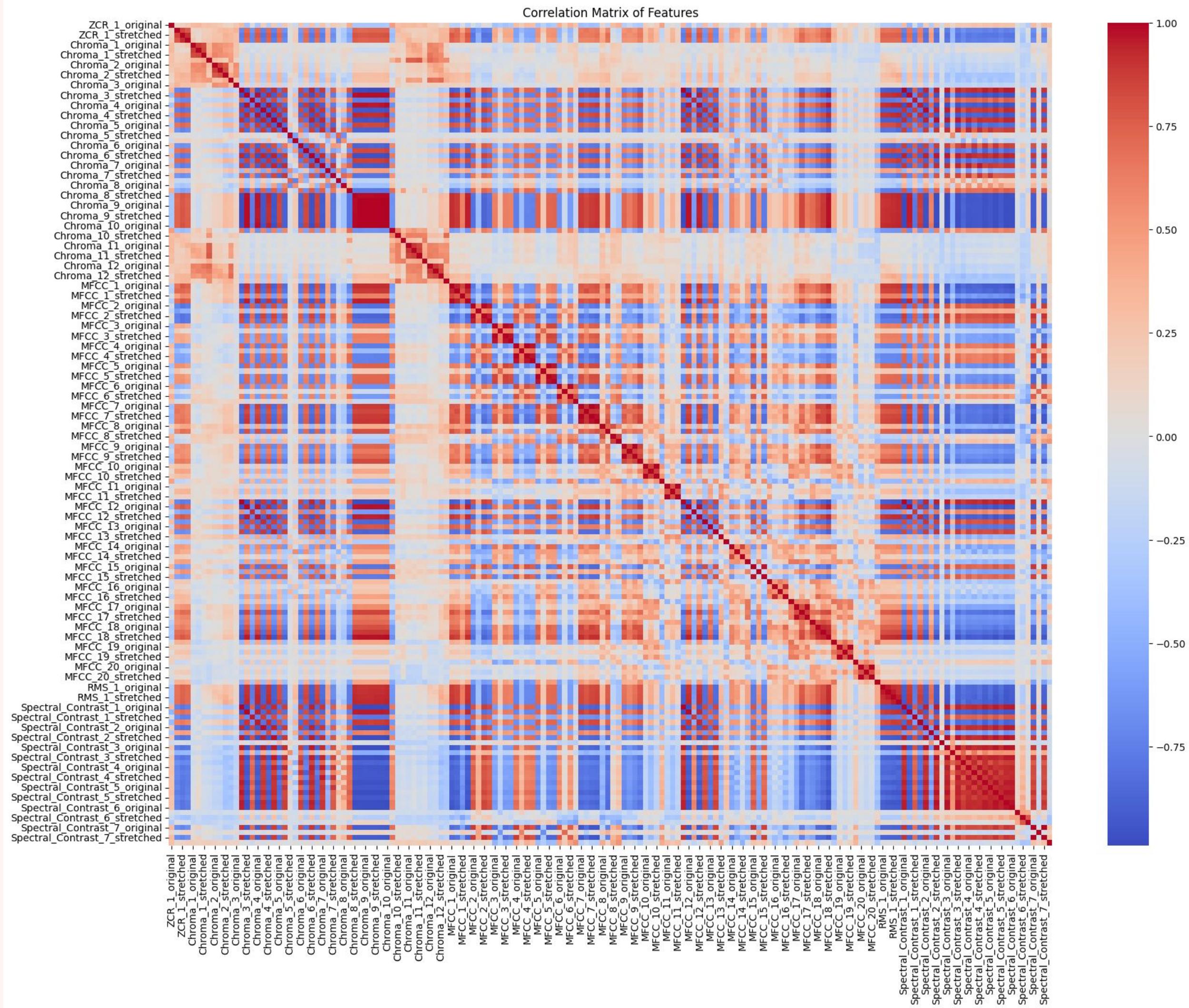


LINEAR SPECTROGRAM - NEUTRAL



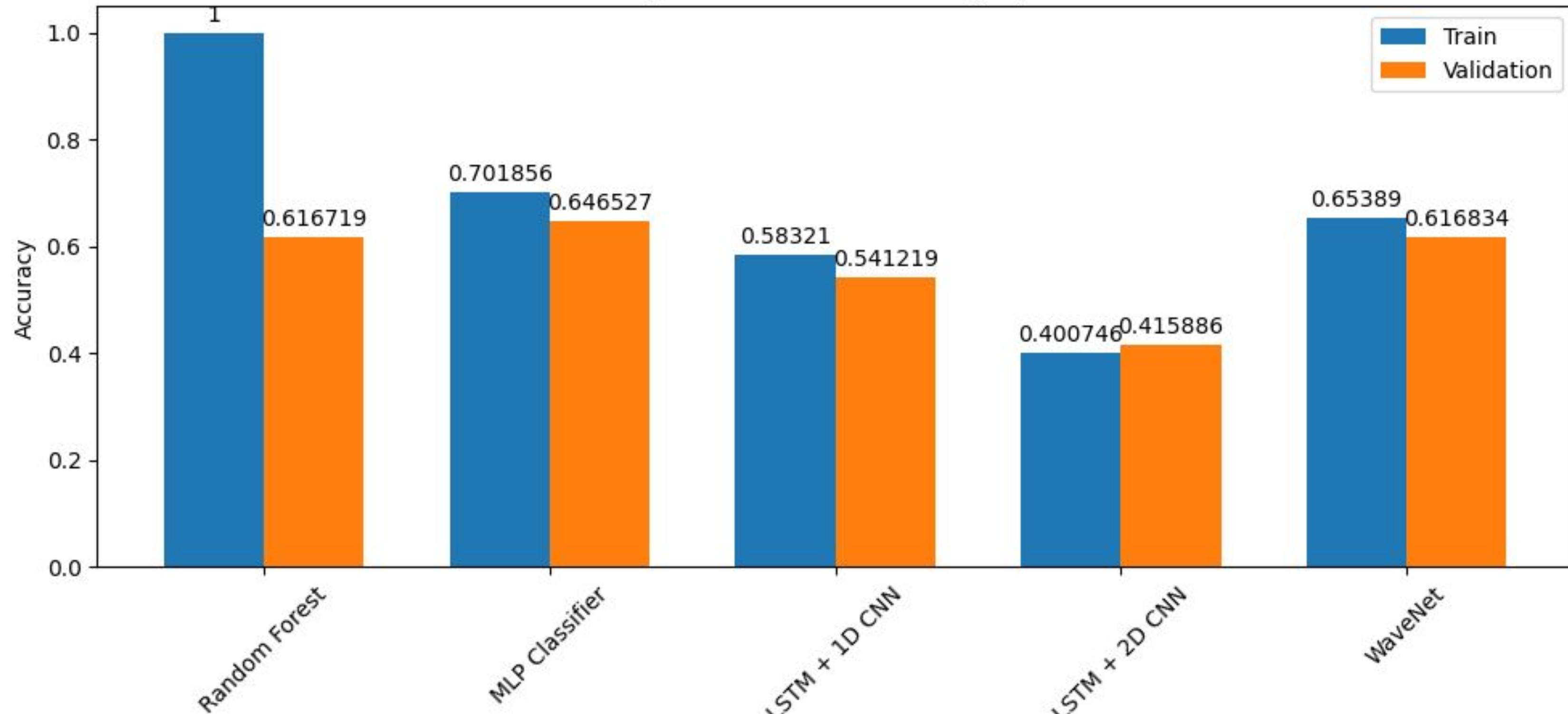
TYPES OF AUGMENTATION VS. ORIGINAL



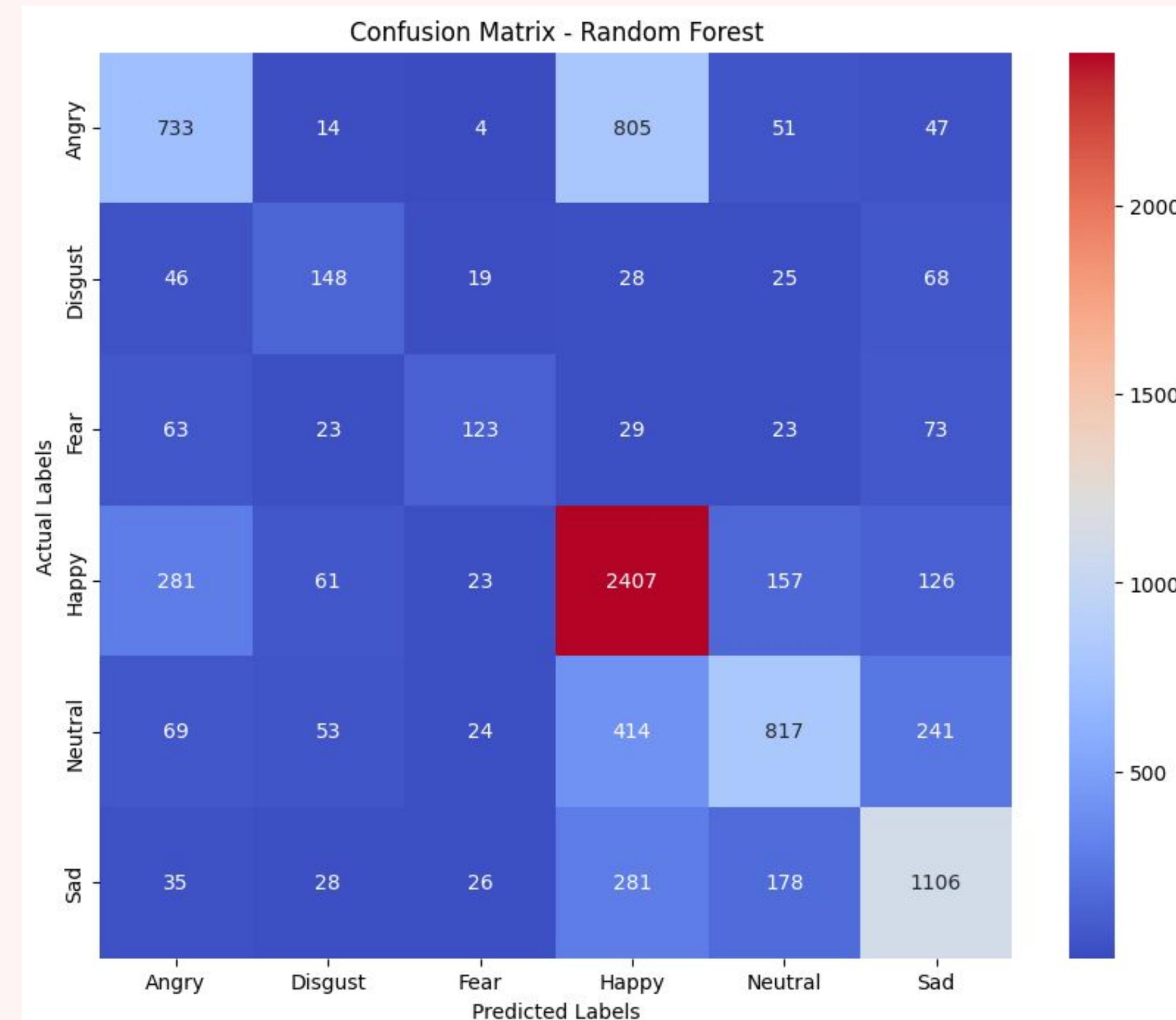


MODEL SCORE COMPARISON

Training and Validation Accuracy by Model

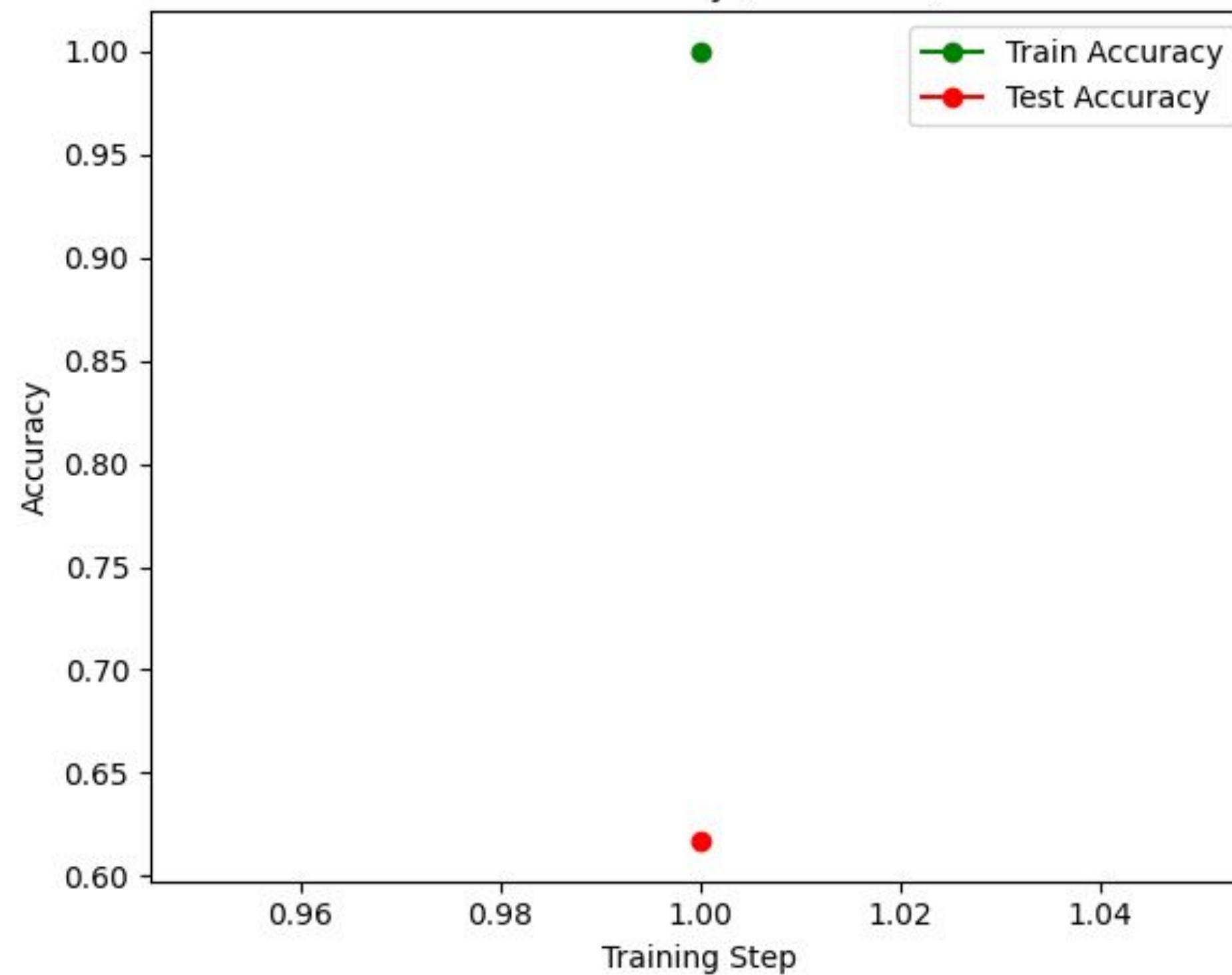


RANDOM FOREST CLASSIFIER

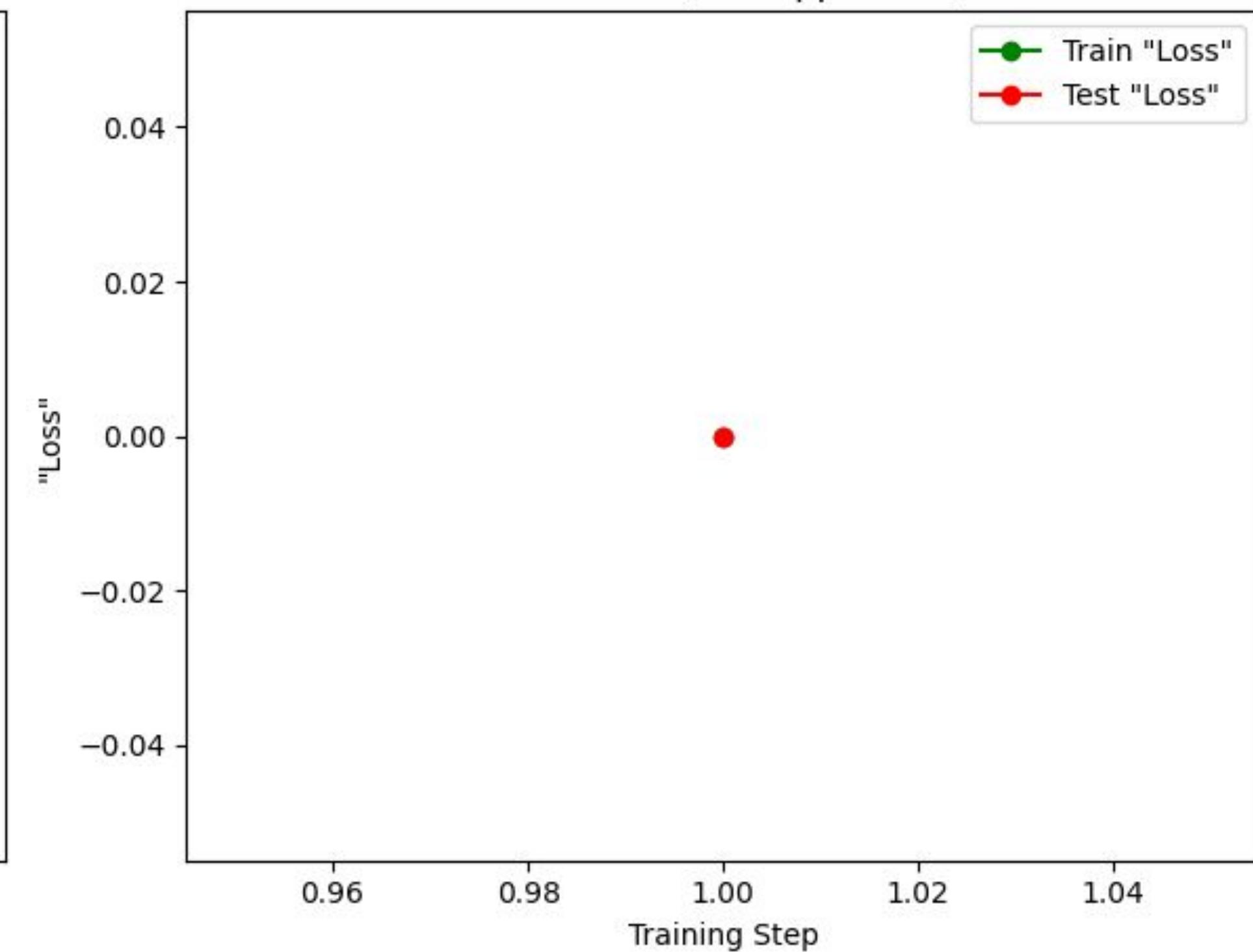


RANDOM FOREST CLASSIFIER

Model Accuracy (Simulated)



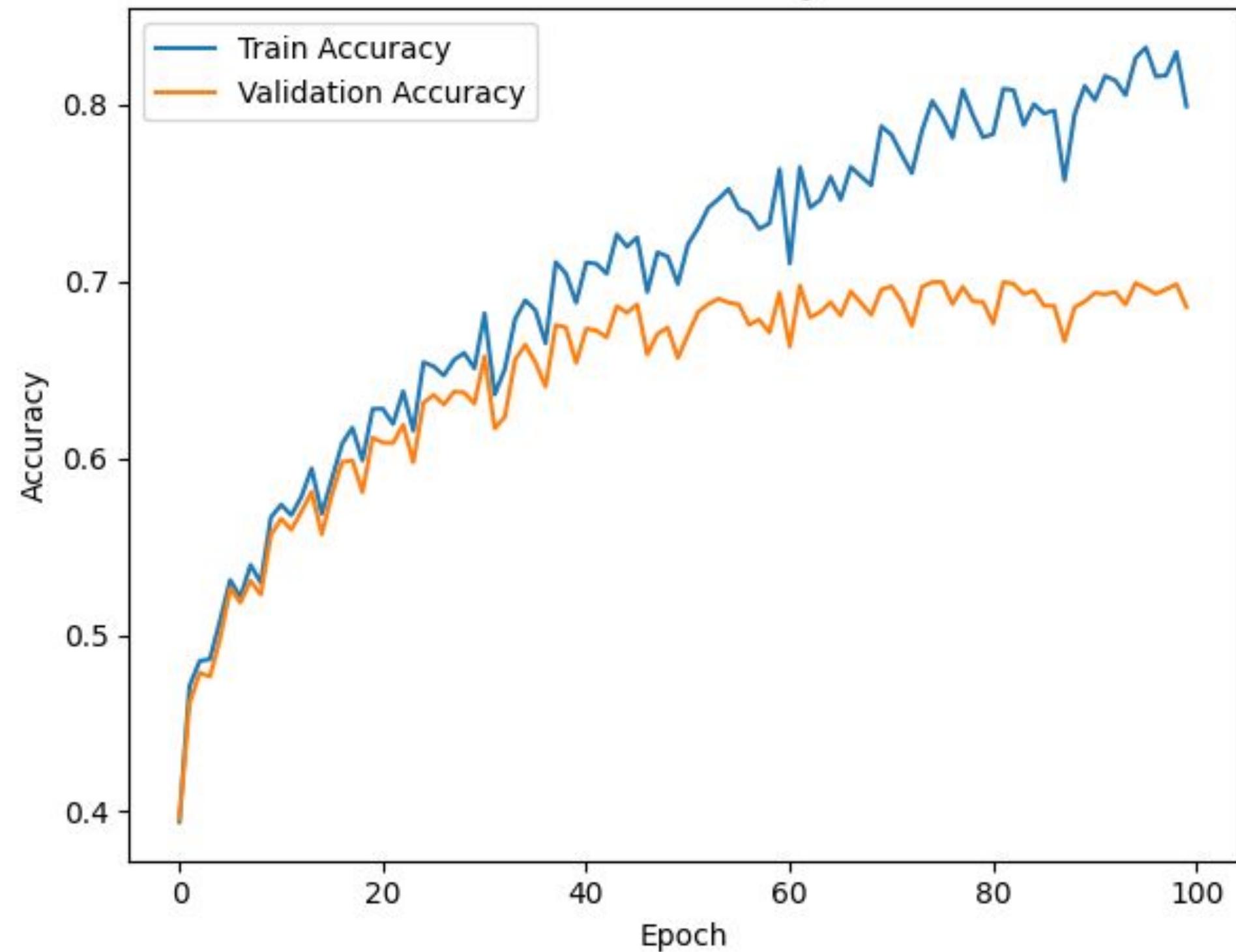
Model Loss (Not Applicable)



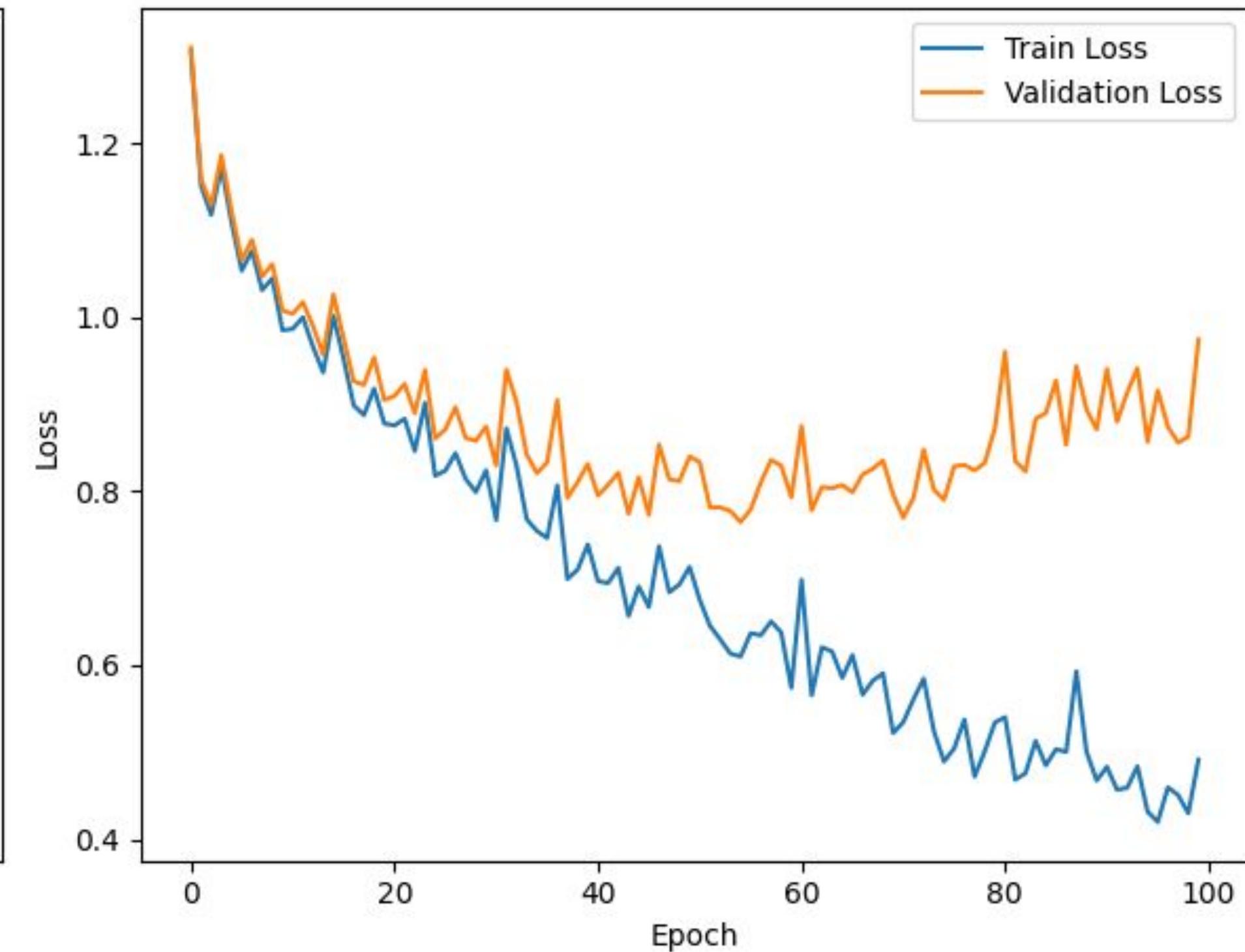
Note: Random Forest models cannot churn out Model Accuracy and Loss curve because they don't run by layers

MLP CLASSIFIER

Model Accuracy



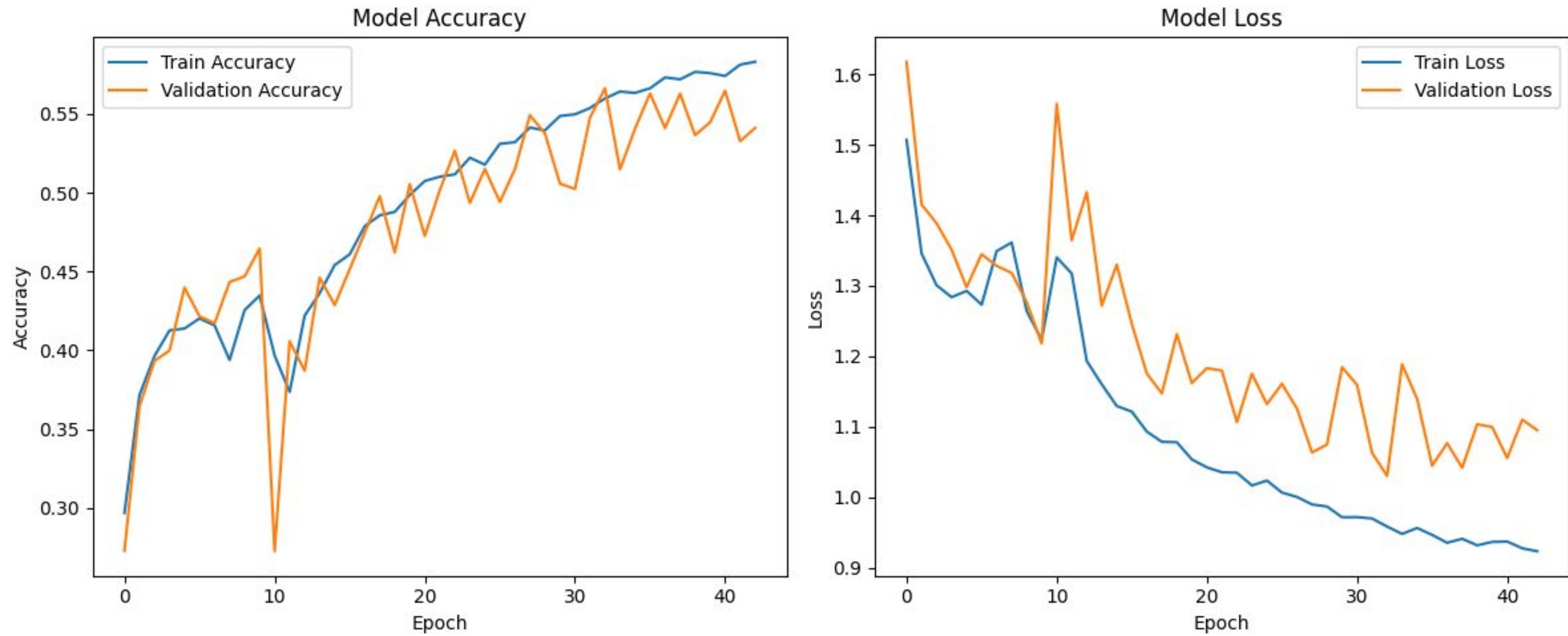
Model Loss



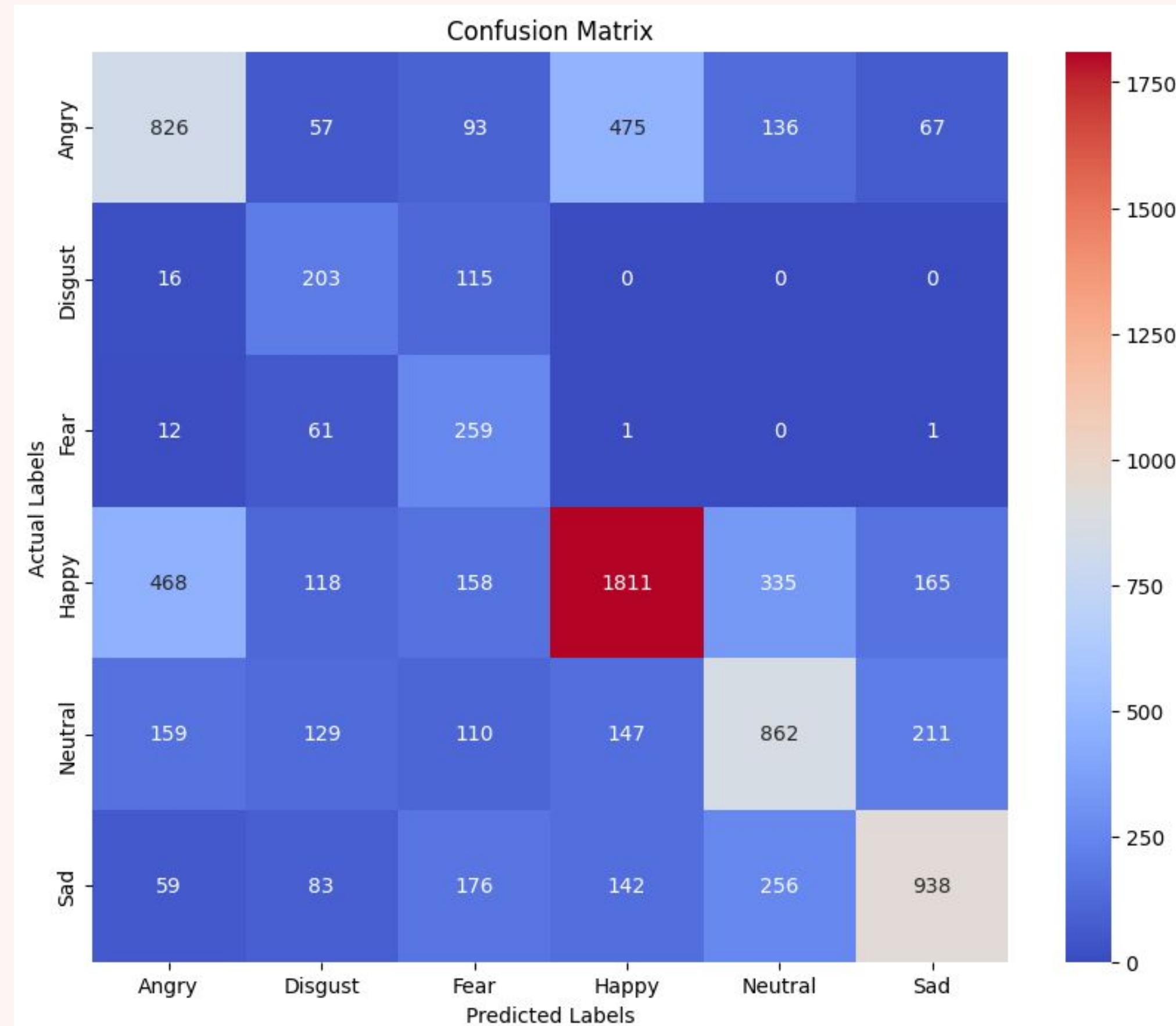
MLP CLASSIFIER



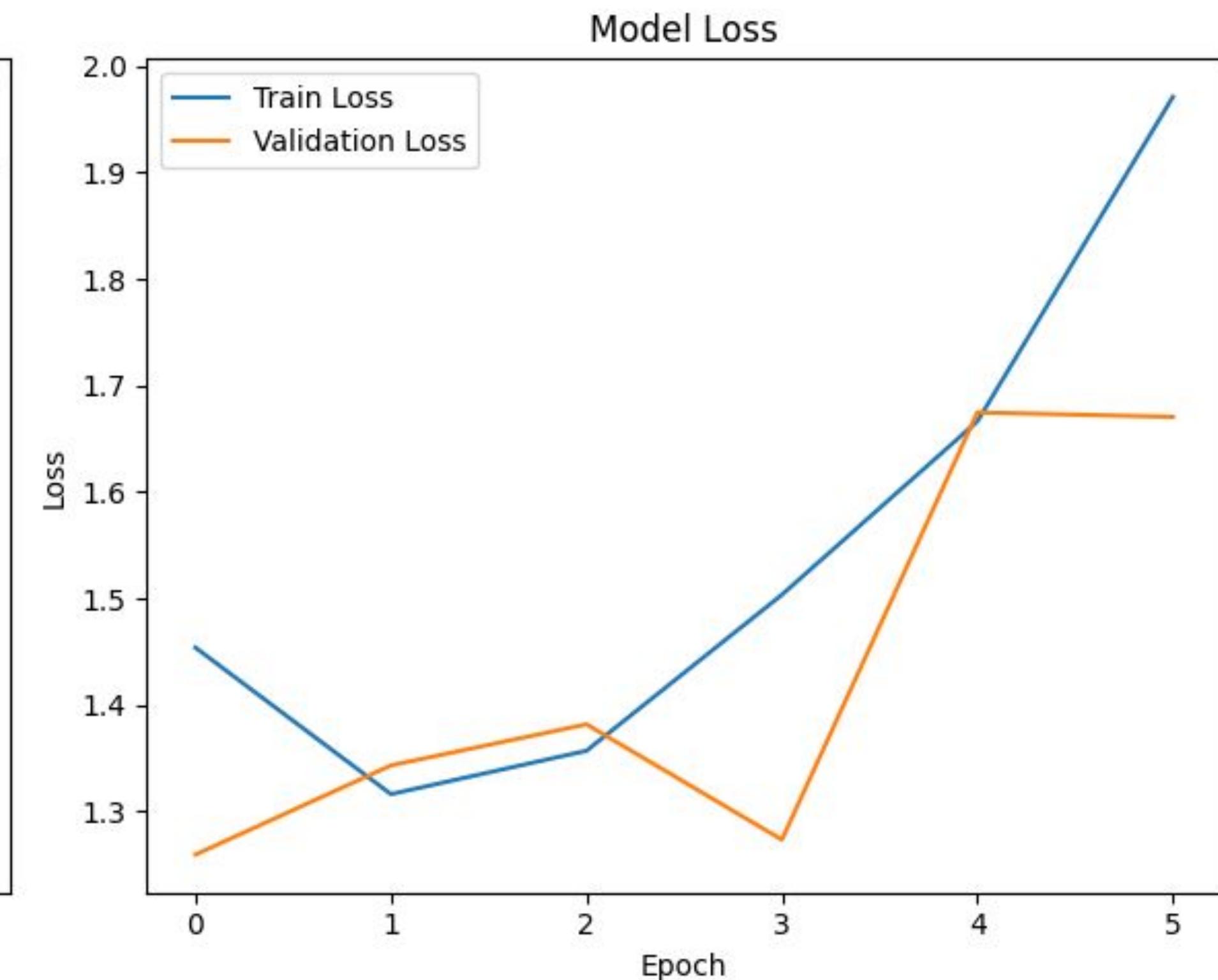
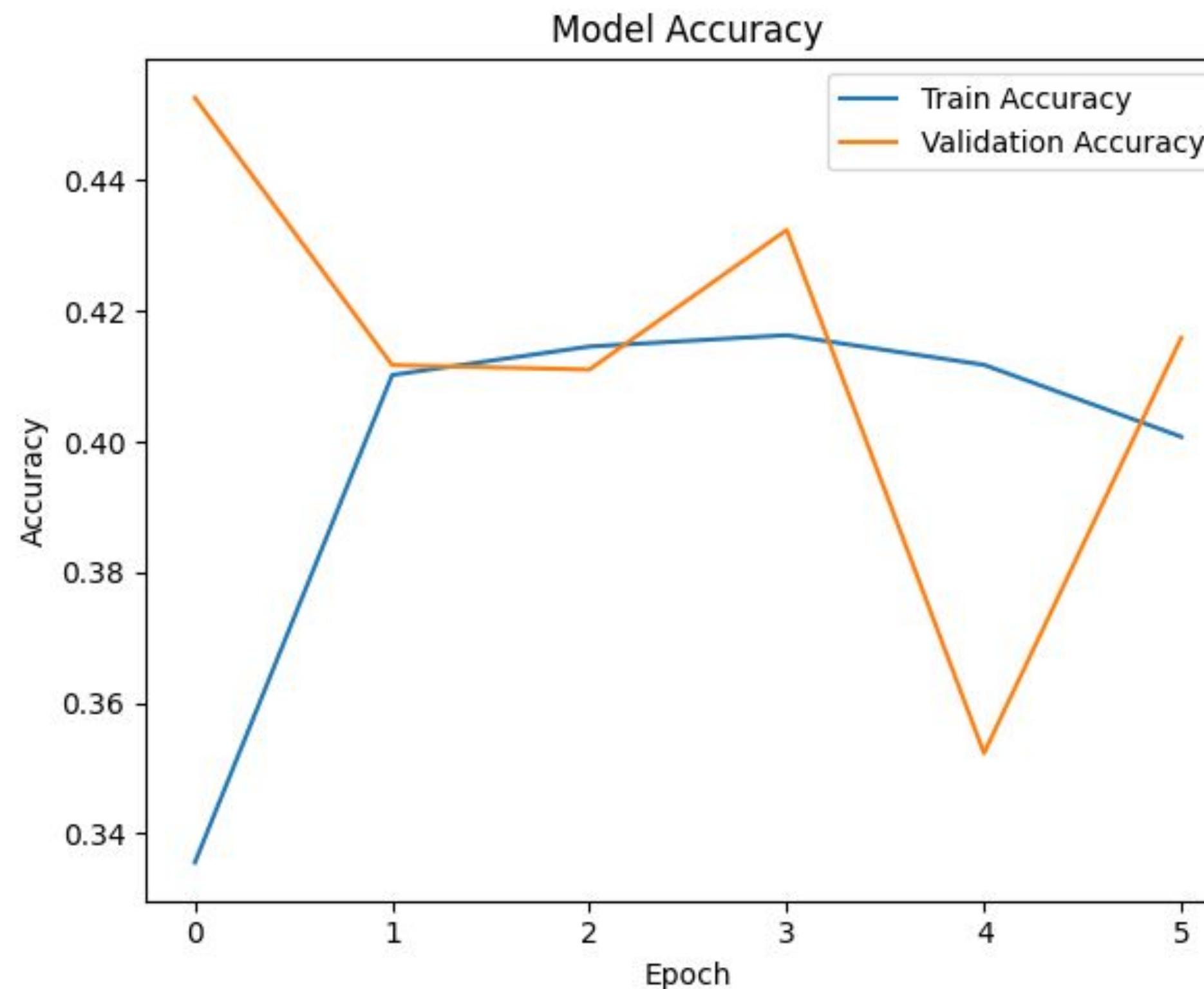
LSTM + 1D CNN



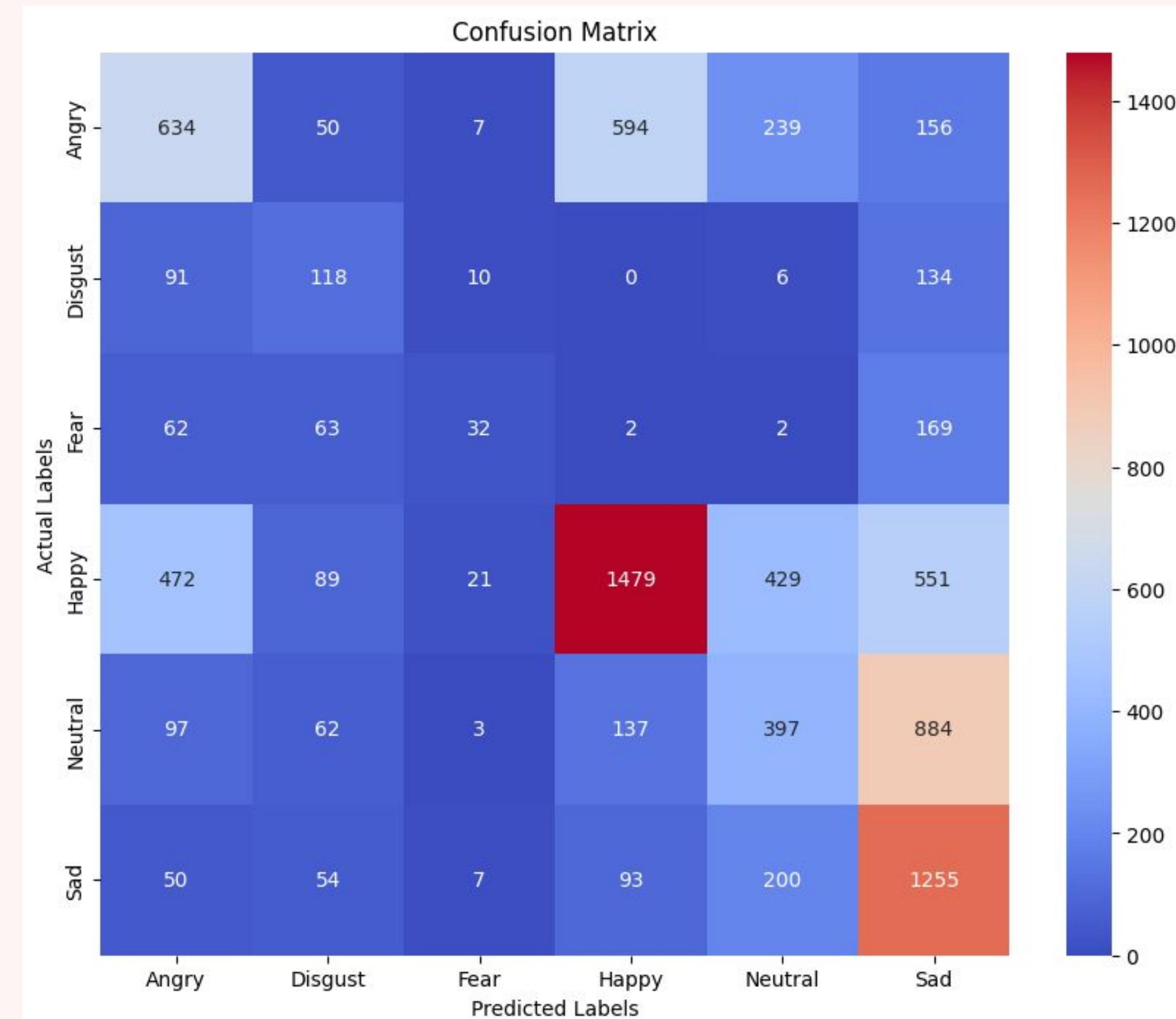
LSTM + 1D CNN



LSTM + 2D CNN

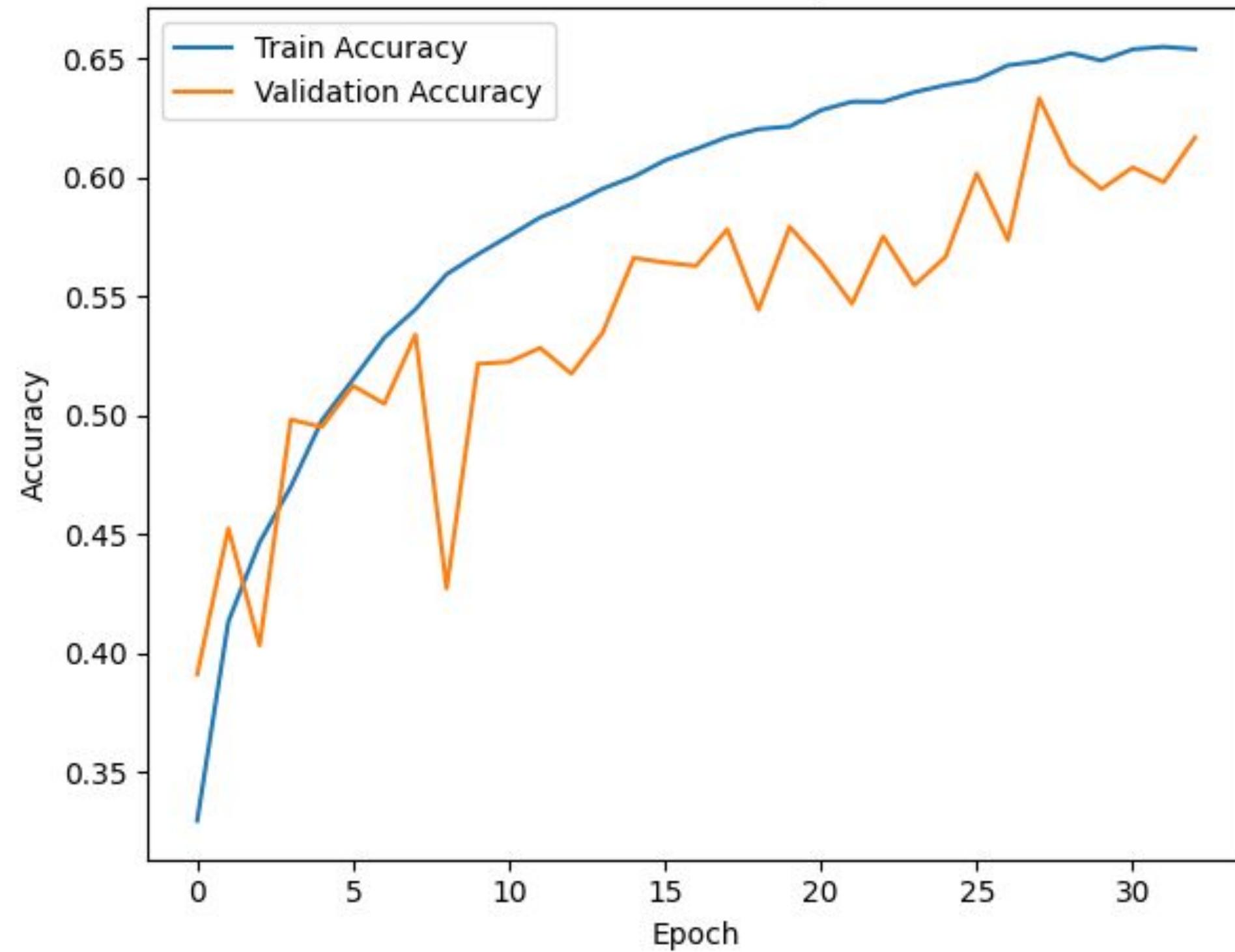


LSTM + 2D CNN

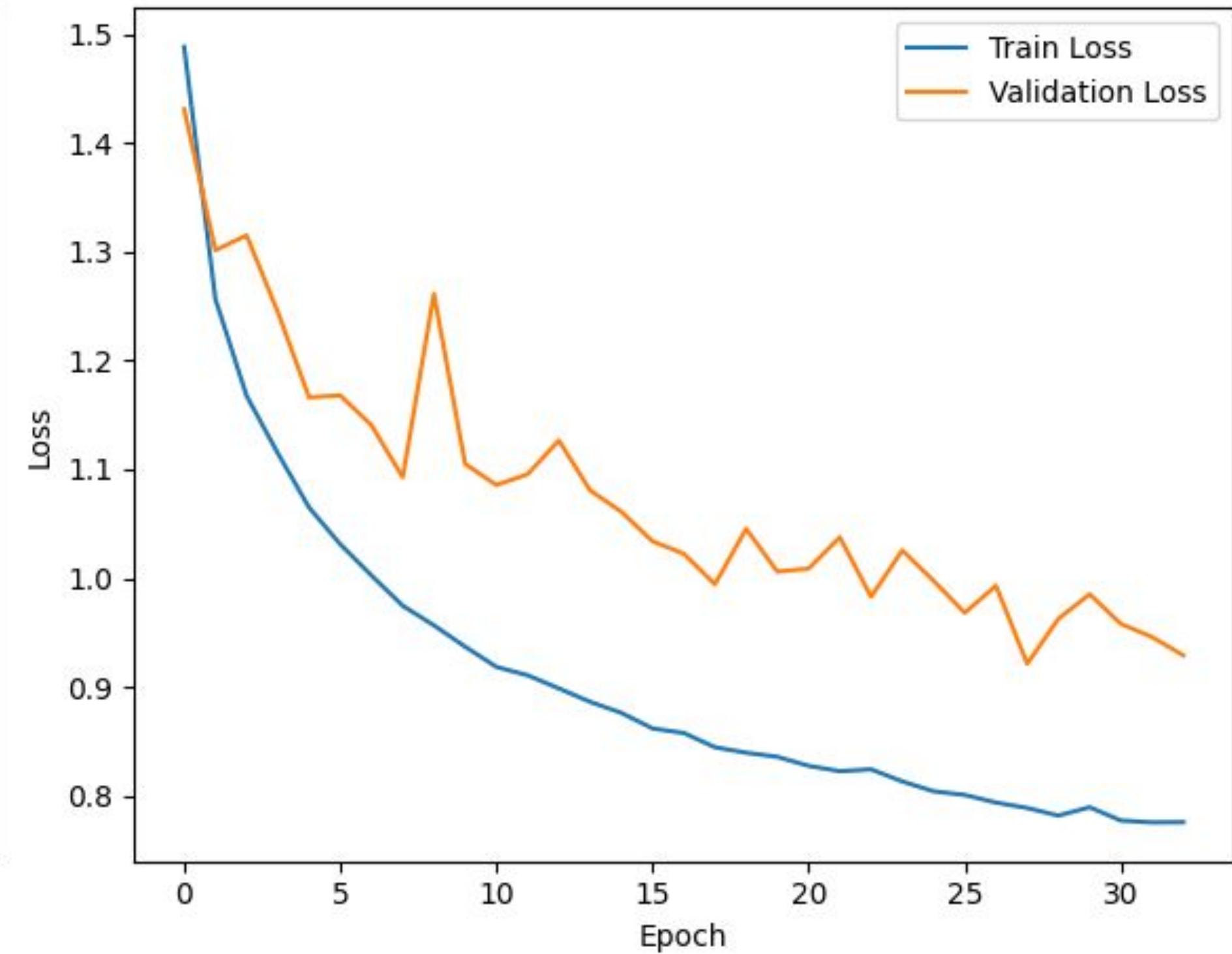


WAVENET (CHAMPION MODEL)

Model Accuracy



Model Loss



WAVENET (CHAMPION MODEL)

