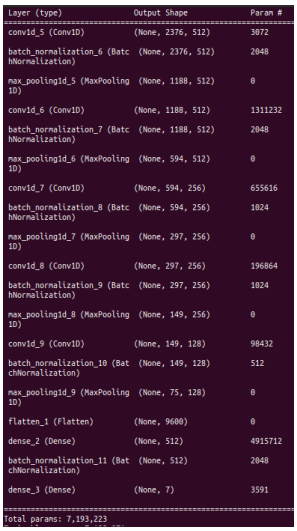
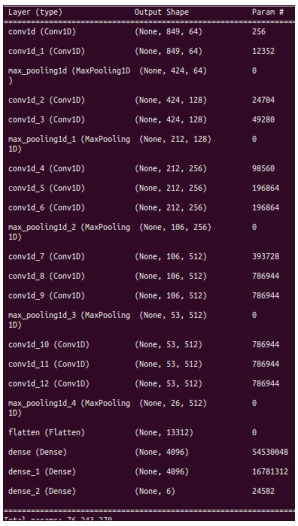
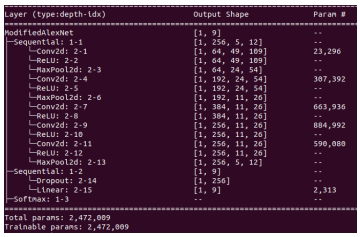
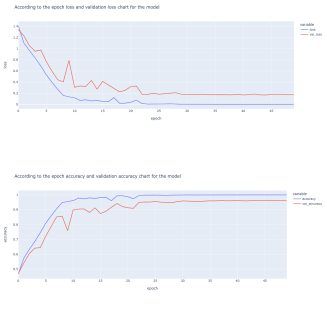
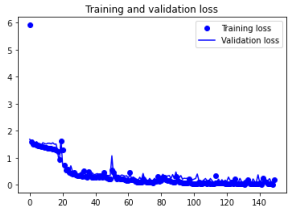
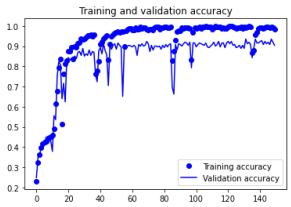
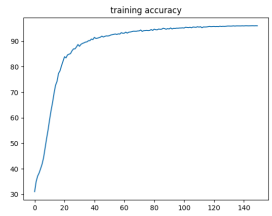
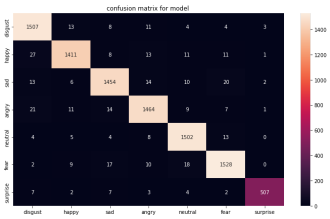
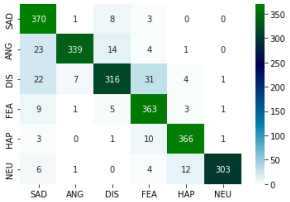
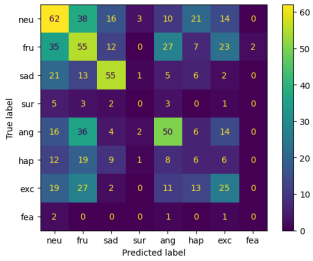
Model Report

Model Name	**Four Dataset Model**	**One Dimentional Model**	**Spectrogram Model**
Links to Notebooks	https://www.kaggle.com/code/ketakipatawa/speech-emotion-recognition-with-cnn	https://www.kaggle.com/code/mariamgharib/speech-emotion-recognition-cnn-1d-model	https://github.com/aris-ai/Audio-and-text-based-emotion-recognition/blob/master/G
Training Datasets* **	Ravdess, Savee, Tess, Crema-D	Crema-D	Iemocap
Preprocess	Noise adding, streching, shifting and pitching for augmentation	Noise reducing, length adjusting	Add delta spectrogram to get a 3d spectrogram
Input Features	1. Zero crossing rate (zcr) 2. Root-mean-square (rms) 3. Mel-frequency cepstral coefficient (mfcc)	1. Zero Crossing rate 2. *Energy* 3. Mel-frequency cepstral coefficient 4. *Spectral Centroid* 5. *Spectral Bandwidth* 6. *Spectral Flatness* 7. *Spectral Rolloff maximum frequencies* 8. *Spectral Rolloff minimum frequencies*	3D Spectrogram image
Input Sample Number	48648	7442	6738
Input Labels	1. angry 2. disgust 3. fear 4. happy	1. angry 2. disgust 3. fear 4. happy	1. neutral 2. sad 3. surprised 4. angry

	5. neutral 6. sad 7. *surprise*	5. neutral 6. sad	5. happy 6. excited 7. fear 8. disgusted
Model architecture			
Training epoch number	50	150	150
Training / Validation Results	 loss: 8.5794e-04 - accuracy: 0.9997 - val_loss: 0.1734 - val_accuracy: 0.9625	  train accuracy: 99.757480 62133789 val accuracy: 98.08428883 552551	  train accuracy: 96.074 train loss: 0.0054
Test Results (On training datasets)			







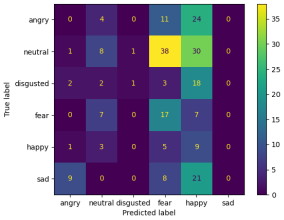

```
Test Loss: 0.164213314652442
93
Test Accuracy: 0.96330934762
95471
```

test accuracy
97.04433497536945

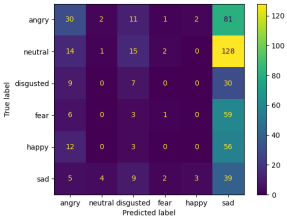
```
test accuracy: 33.59
```

Test Results on Emodb dataset *

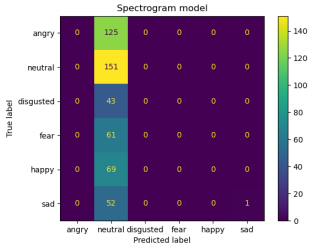
Accuracy: 6.54 %



Accuracy: 14.57 %

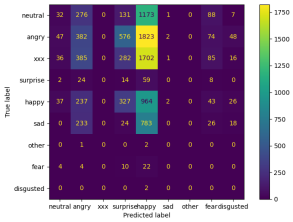


Accuracy: 28.41 %

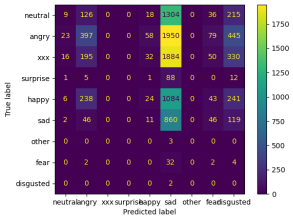


Test Results on Iemocap dataset *

Accuracy: 12.869807749775875 %



Accuracy: 13.86592290068732 %

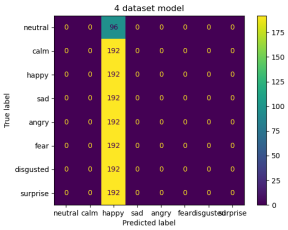


This model is already trained with IEMOCAP dataset, so no need for additional test

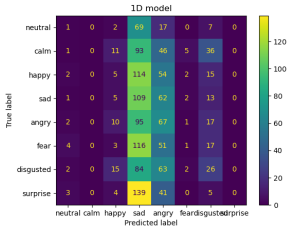
---

Test results on Ravdess dataset

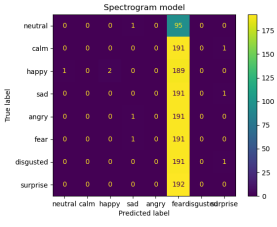
Accuracy: 13.33 %



Accuracy: 14.51 %



Accuracy: 13.40 %



***Notes on Datasets

Ravdess

- 60 trials per actor x 24 actors = 1440 samples
- Calm, happy, sad, angry, fearful, surprise, and disgust
- Normal, strong or neutral expression

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/03006498-f51e-4cc3-bbd6-1906685fe591/03-01-08-02-02-02-14.wav

SAVEE

- 120 trials per actor x 4 male actors = 480 samples
- Anger, disgust, fear, happiness, sadness , surprise and neutral

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/41c9f4e5-7715-4719-a479-28d8866b7077/KL_su15.wav

TESS

- 200 words x 7 emotions x 2 female actresses = 2800 samples
- Anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/9d329c77-1253-4193-bc6f-43dda165665d/YAF_white_ps.wav

Crema-D

- 91 male and female actors, 12 sentences → 7442 samples
- Anger, Disgust, Fear, Happy, Neutral, and Sad
- Low, Medium, High, and Unspecified intensity

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/040b9615-9e04-41d9-980e-adfab6b96da2/1015_TIE_SAD_XX.wav

EmoDB

- In German
- 10 actors (5 male, 5 female) → 535 samples
- Anger, boredom, anxiety, happiness, sadness, disgust and neutral.

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/2b02ec37-66bb-492b-926a-dcb327315884/16b10Wa.wav

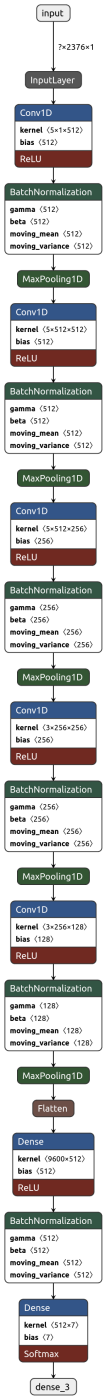
IEMOCAP

- 10 actors (5 male, 5 female) in scripted and improvised situations → 10039 samples
- anger, happiness, excitement, sadness, frustration, fear, surprise, other and neutral state

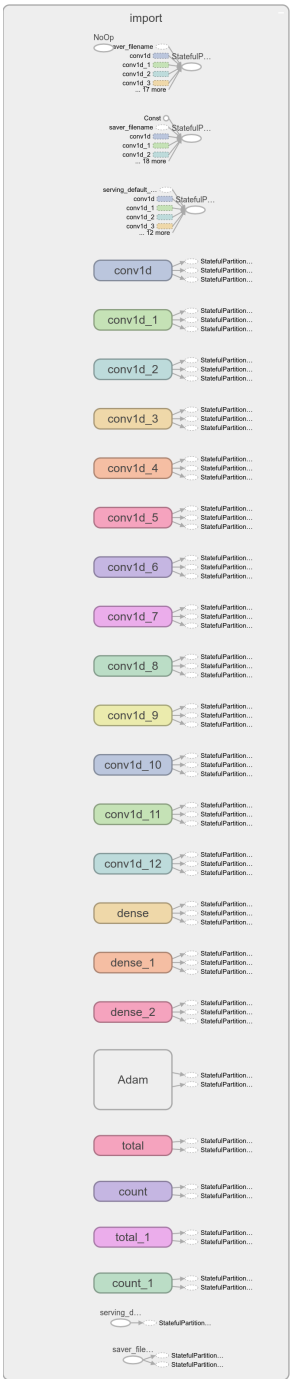
https://s3-us-west-2.amazonaws.com/secure.notion-static.com/76781a35-82df-420f-80f6-b2f17e9b5344/Ses01M_script01_1_M042.wav

Model block diagrams

Four Dataset Model



One Dimensional Model



Spectrogram Model