Speech Emotion Recongition

DSI-21 Lloyd Lau

Table of Contents

Introduction

Human emotions can be detected and analysed in numerous ways:

- Tonal properties
- Facial expression
- Body gesture.

Business enterprises harness technological advancement powered by speech to:

- Harvest
- Forecast
- Evaluate

Problem Statement

An actionable business intelligence tool for call centres:

- Improve value of customer relationships
- Identify with precision customer’s needs
- Improve quality of interactions between agents / customers
- Improve functionality and prioritizing
- Ultimately to predict emotions from speeches

Table of Contents

Emotion Data

Kaggle

- Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)
  - 1440 Audio clips
  - 24 actors
- Crowd Sourced Emotional Multimodal Actors Dataset (CREMA-D)
  - 7,442 Audio clips
  - 91 actors

Emotion Representation

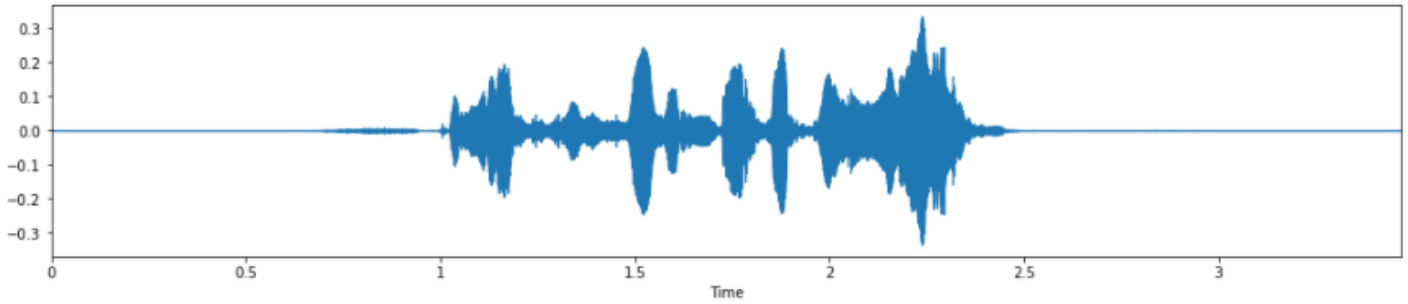
Categorical

- Sad
- Disgust
- Fear
- Happy
- Angry
- Neutral

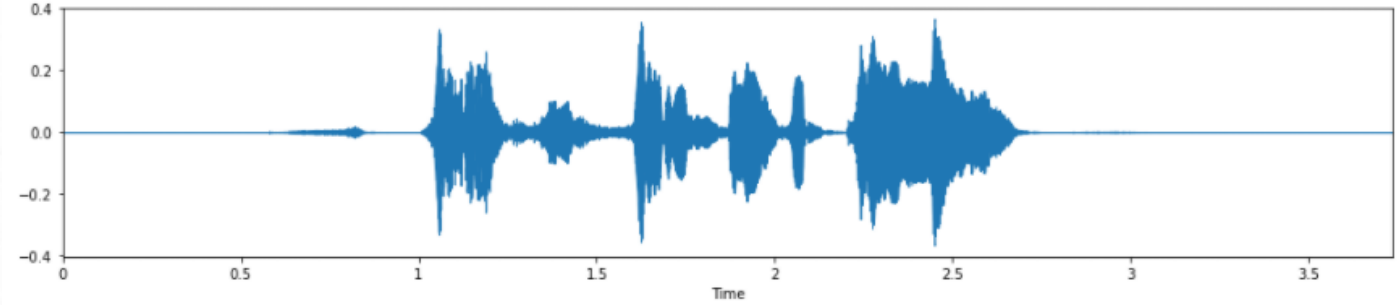
male_sad	767
male_disgust	767
male_fear	767
male_happy	767
male_angry	767
male_neutral	719
female_sad	696
female_angry	696
female_fear	696
female_disgust	696
female_happy	696
female_neutral	656

Emotion Representation

Happy

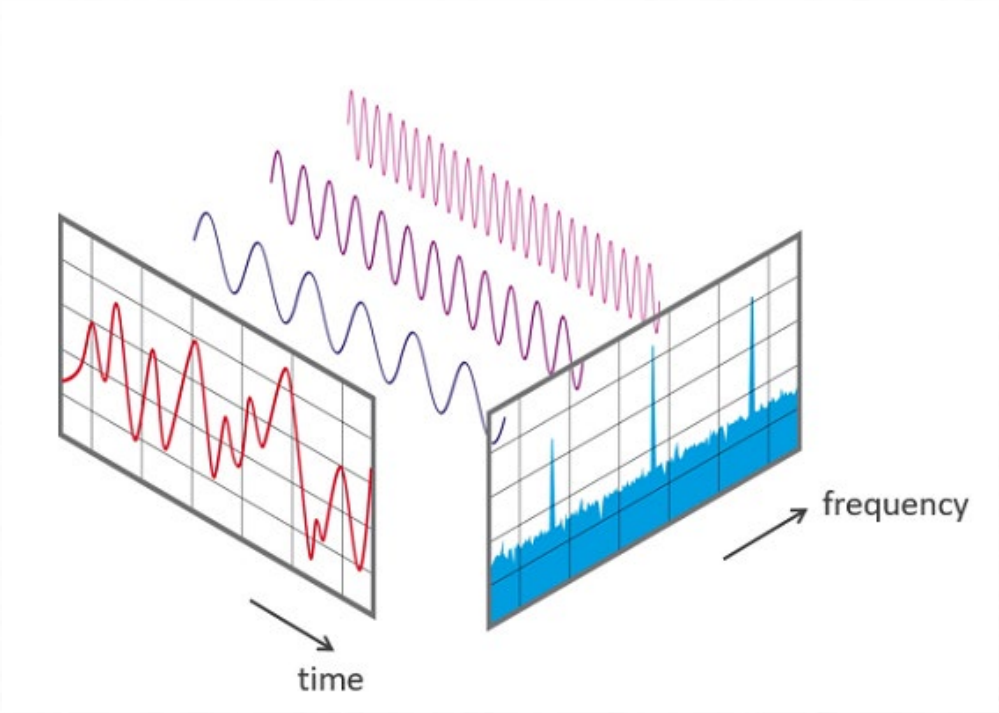


Fearful

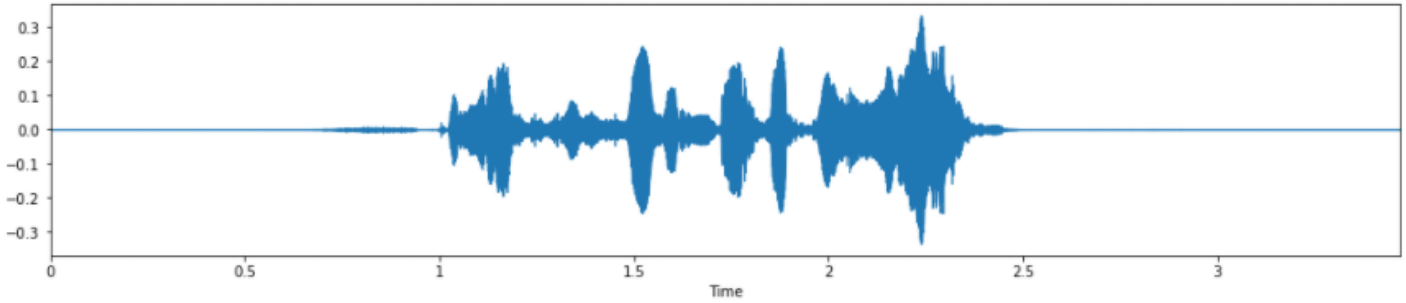


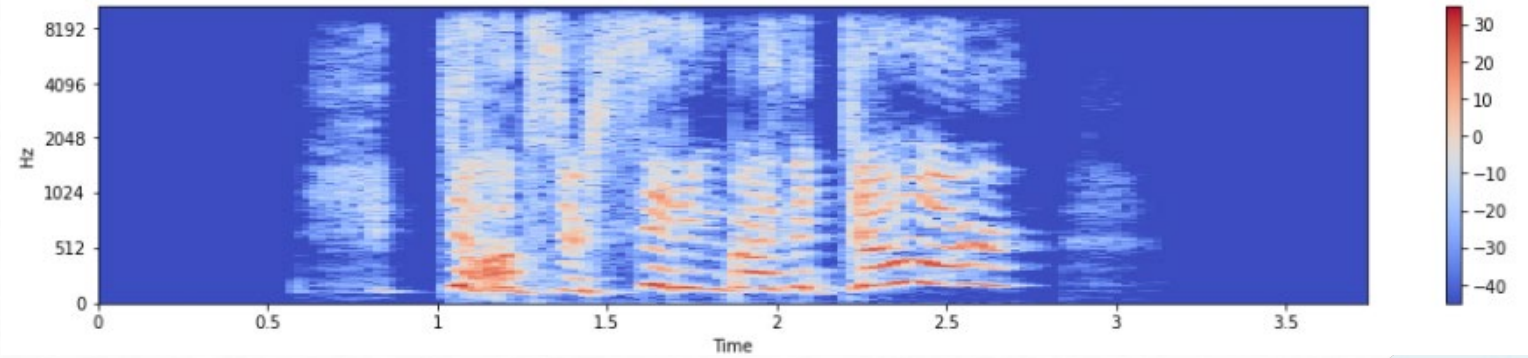
“Dogs are sitting by the Door”

Audio Representation



Audio Representation





Audio Representation

Mel-frequency Cepstrum (MFC)

- Short-term power spectrum of a sound by transforming the audio signal through a series of steps to mimic the human hearing.

Mel-Frequency Cepstral Coefficients (MFCC):

- Coefficients which capture the envelope of the short time power spectrum.

Audio Representation

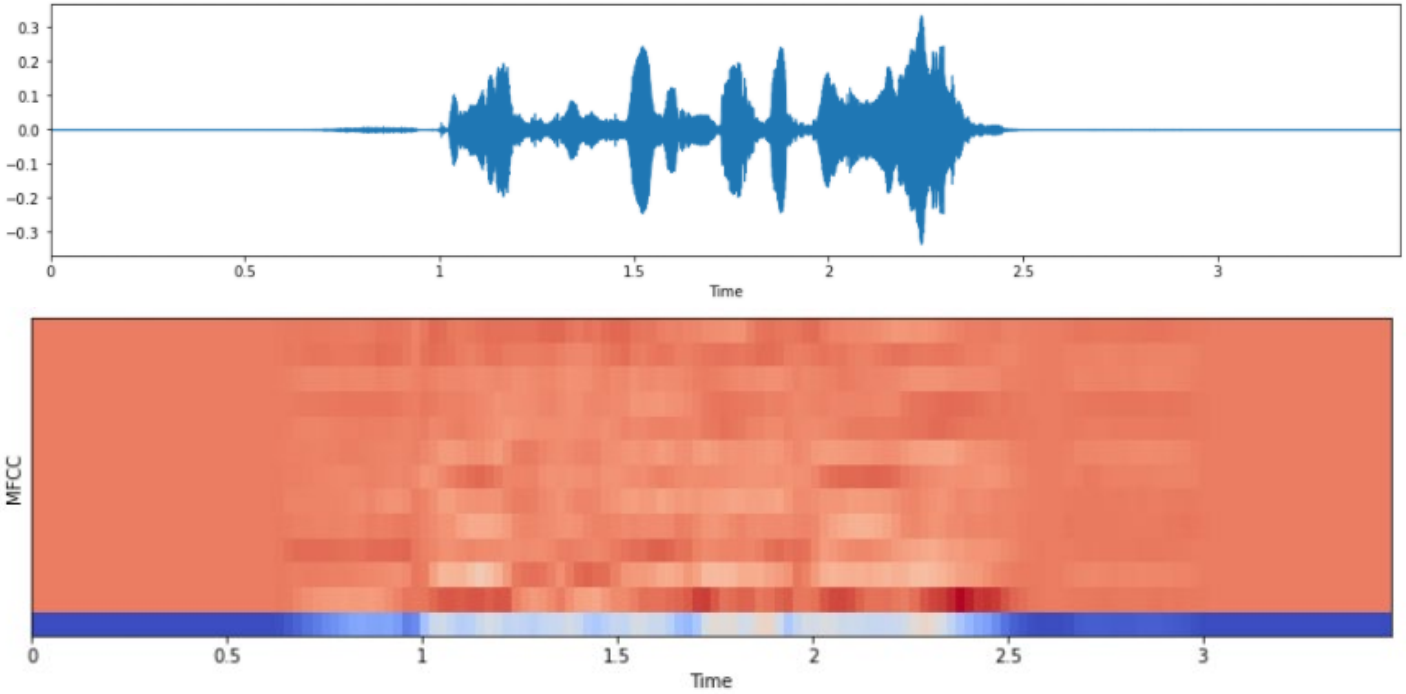
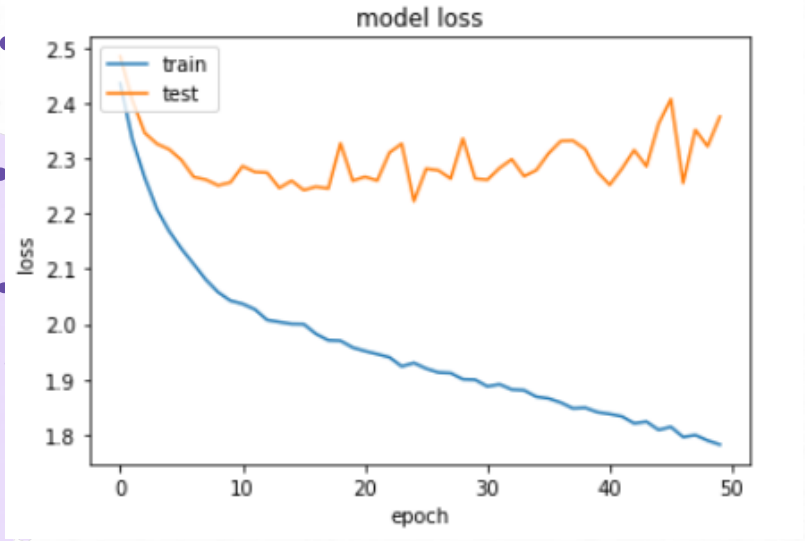
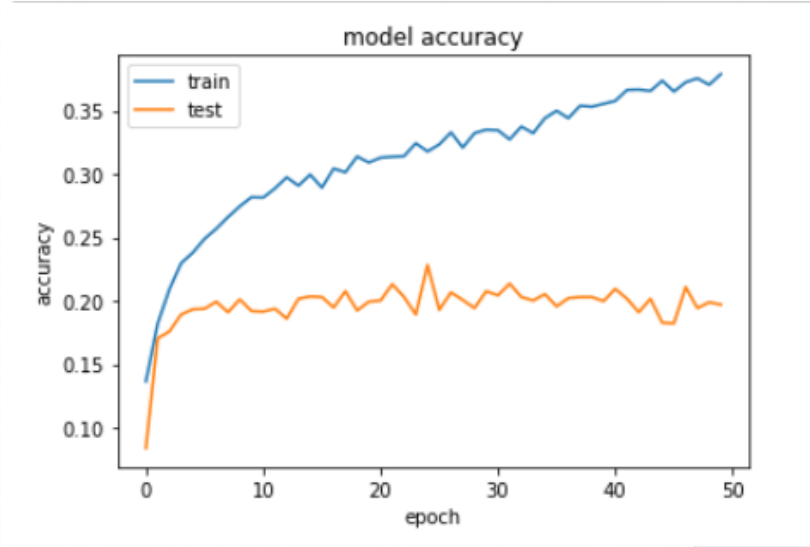


Table of Contents

Baseline Model

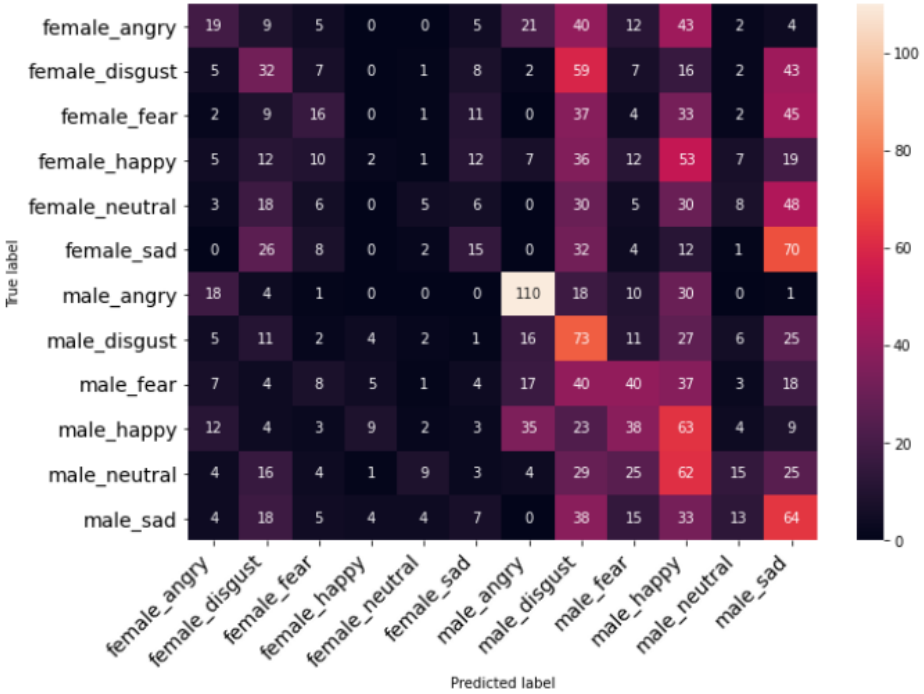
Convolutional neural networks (CNN)





Baseline Model

Confusion Matrix



Accuracy: 21%

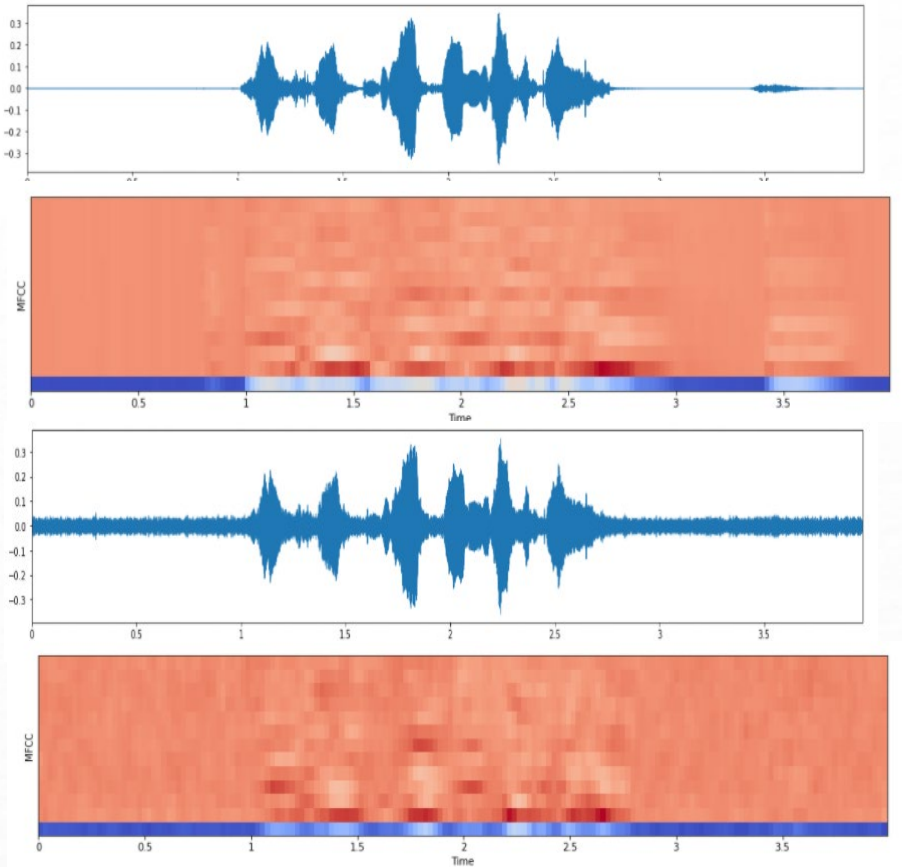
	precision	recall	f1-score	support
female_angry	0.23	0.12	0.16	160
female_disgust	0.20	0.18	0.19	182
female_fear	0.21	0.10	0.14	160
female_happy	0.08	0.01	0.02	176
female_neutral	0.18	0.03	0.05	159
female_sad	0.20	0.09	0.12	170
male_angry	0.52	0.57	0.54	192
male_disgust	0.16	0.40	0.23	183
male_fear	0.22	0.22	0.22	184
male_happy	0.14	0.31	0.20	205
male_neutral	0.24	0.08	0.12	197
male_sad	0.17	0.31	0.22	205
accuracy			0.21	2173
macro avg	0.21	0.20	0.18	2173
weighted avg	0.21	0.21	0.19	2173

Data Augmentation

- Static noise
- Shift
- Stretch
- Pitch
- Dynamic change
- Speed and pitch

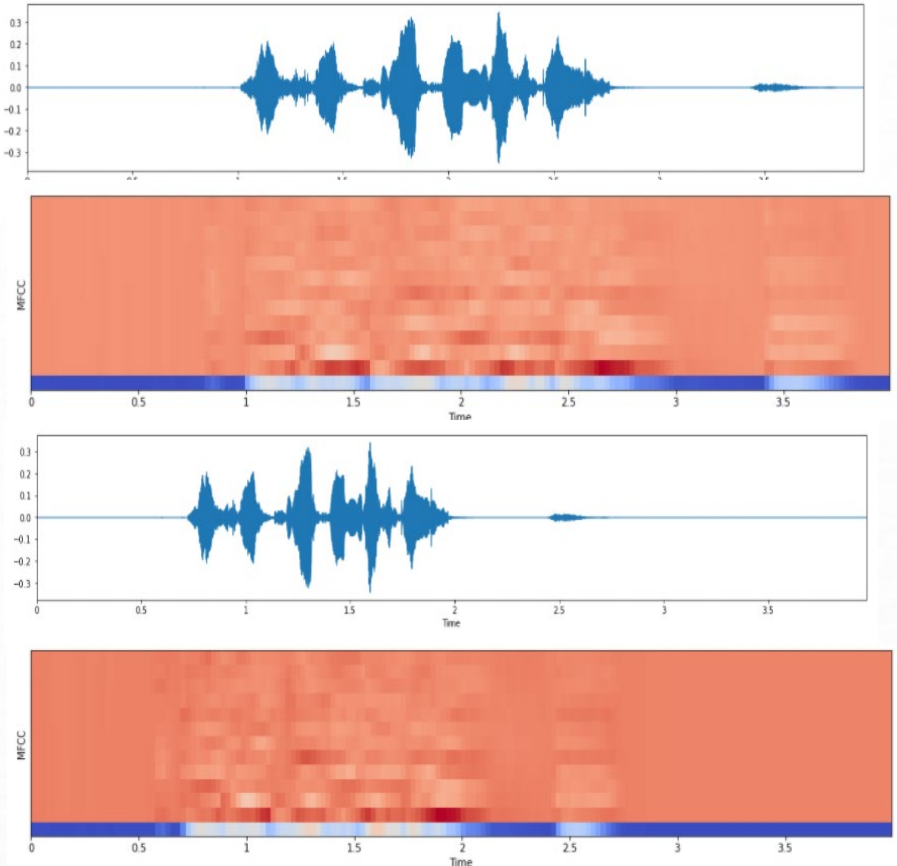
Data Augmentation

- Original Data
- Static Noise



Data Augmentation

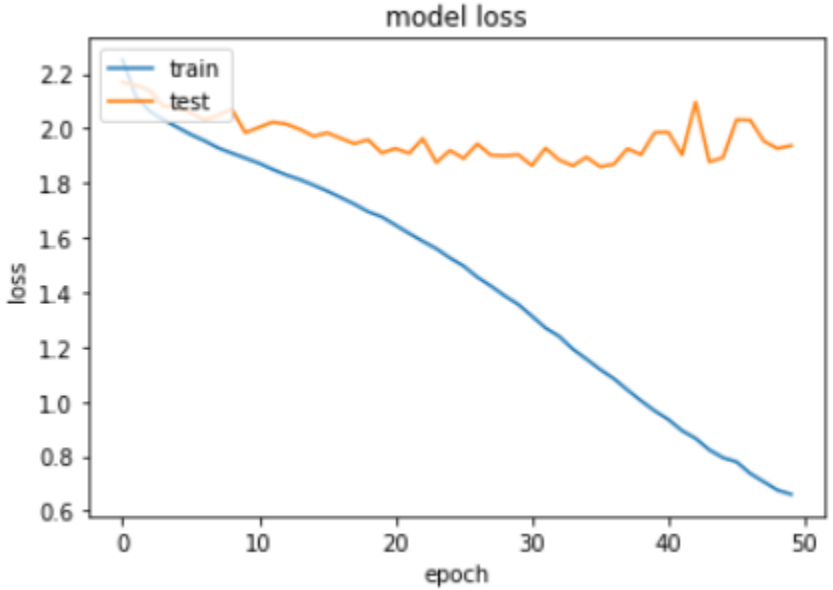
- Original Data

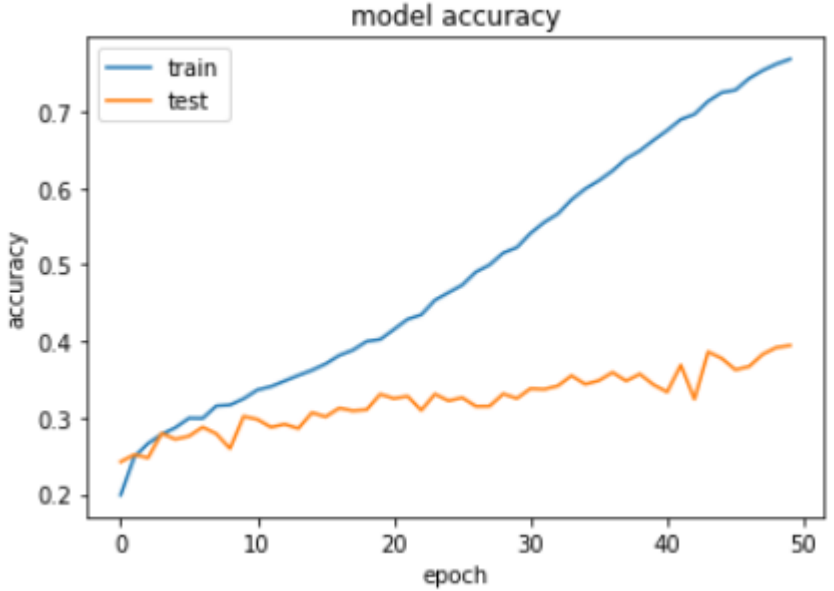


- Speed & Pitch

Modelling

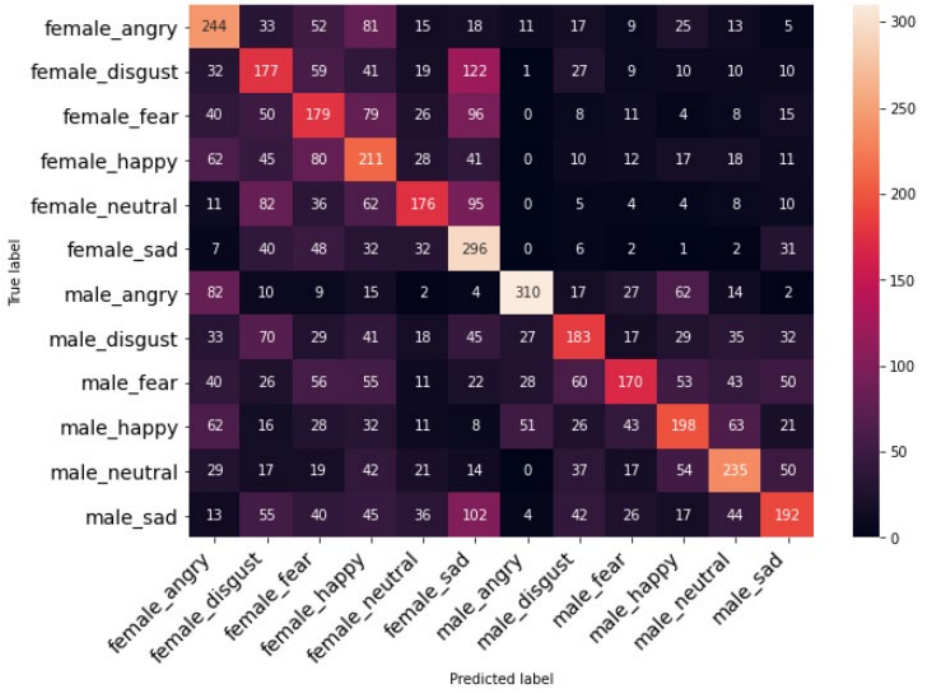
Convolutional neural networks (1D-CNN)





Modelling

Confusion Matrix

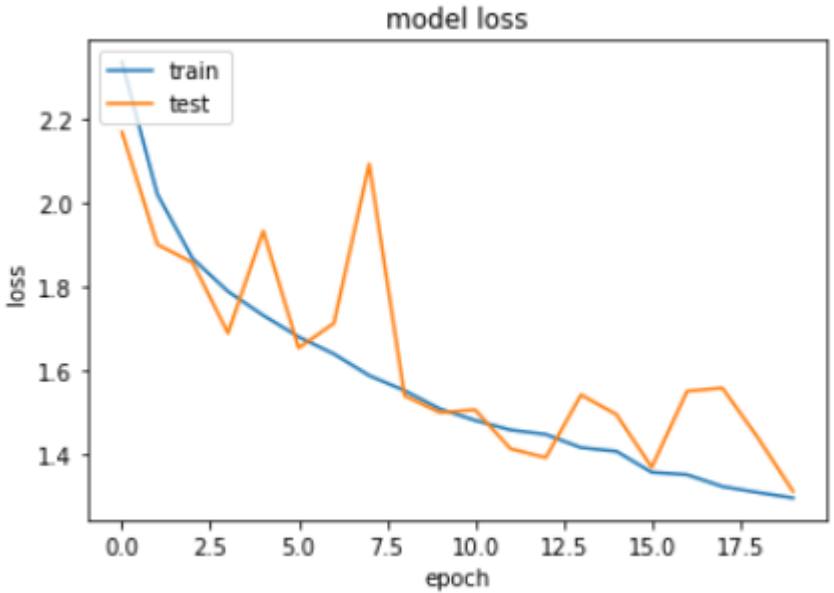


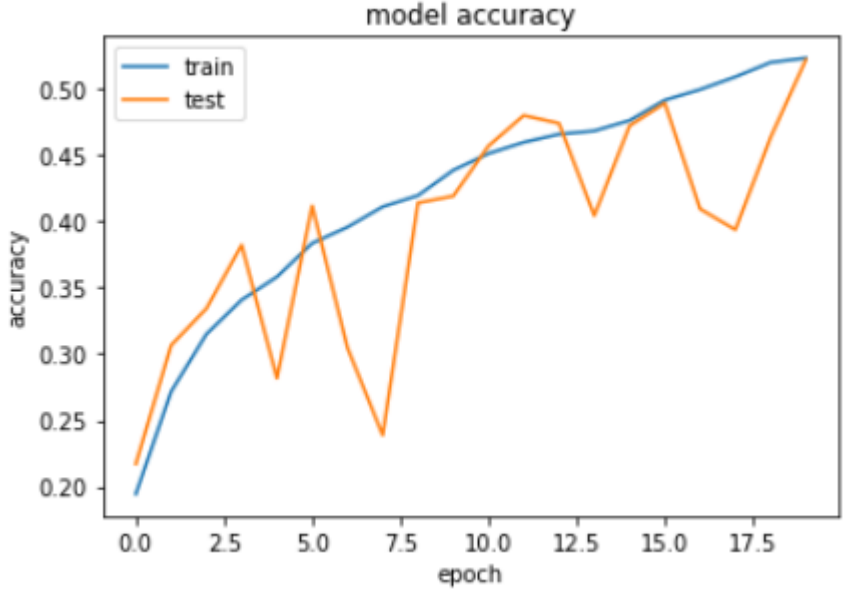
Accuracy: 39%

	precision	recall	f1-score	support
female_angry	0.37	0.47	0.41	523
female_disgust	0.29	0.34	0.31	517
female_fear	0.28	0.35	0.31	516
female_happy	0.29	0.39	0.33	535
female_neutral	0.45	0.36	0.40	493
female_sad	0.34	0.60	0.44	497
male_angry	0.72	0.56	0.63	554
male_disgust	0.42	0.33	0.37	559
male_fear	0.49	0.28	0.35	614
male_happy	0.42	0.35	0.38	559
male_neutral	0.48	0.44	0.46	535
male_sad	0.45	0.31	0.37	616
accuracy			0.39	6518
macro avg	0.42	0.40	0.40	6518
weighted avg	0.42	0.39	0.40	6518

Modelling

Convolutional neural networks (2D-CNN)

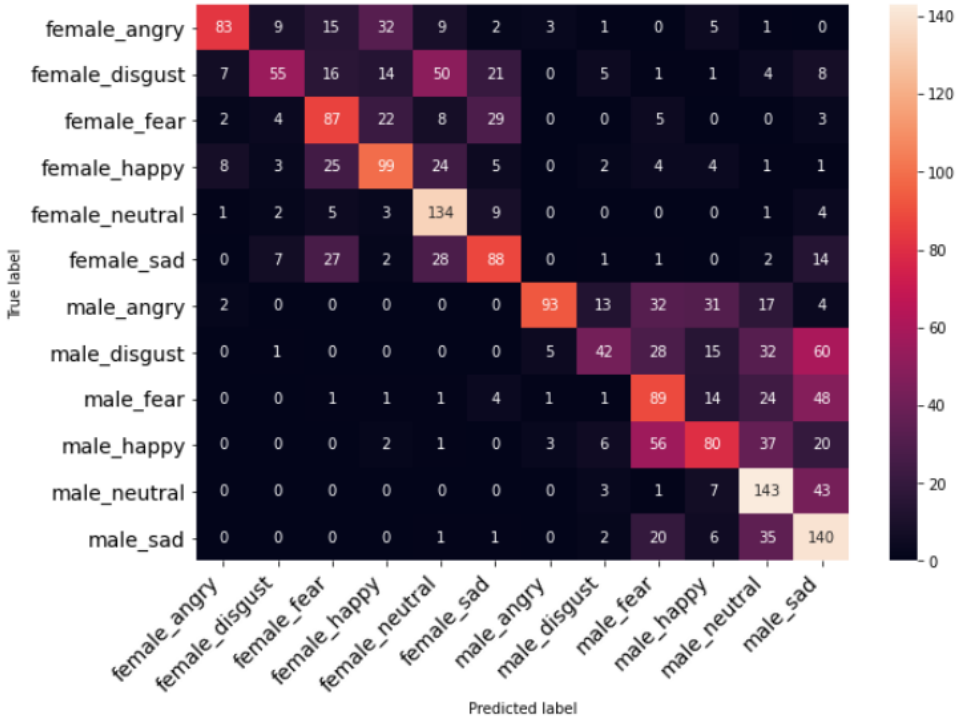




With Augmentation

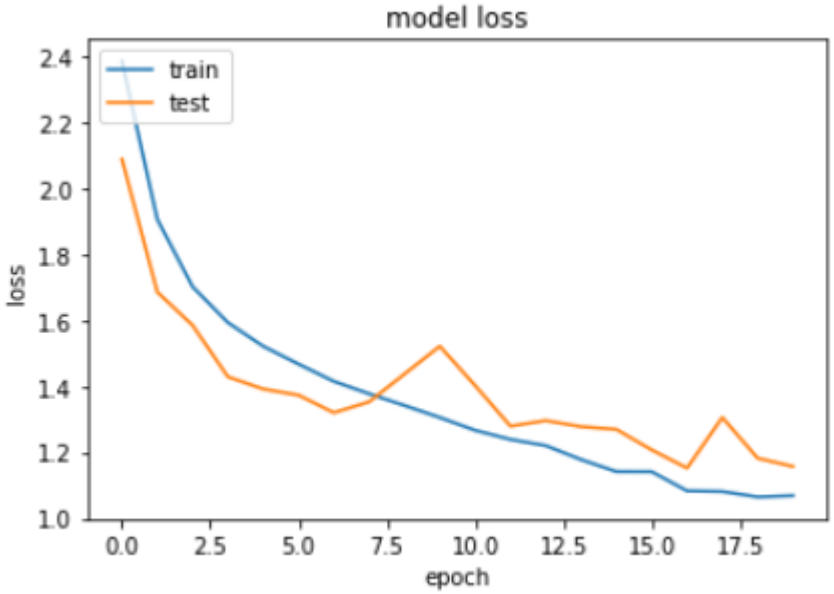
Modelling

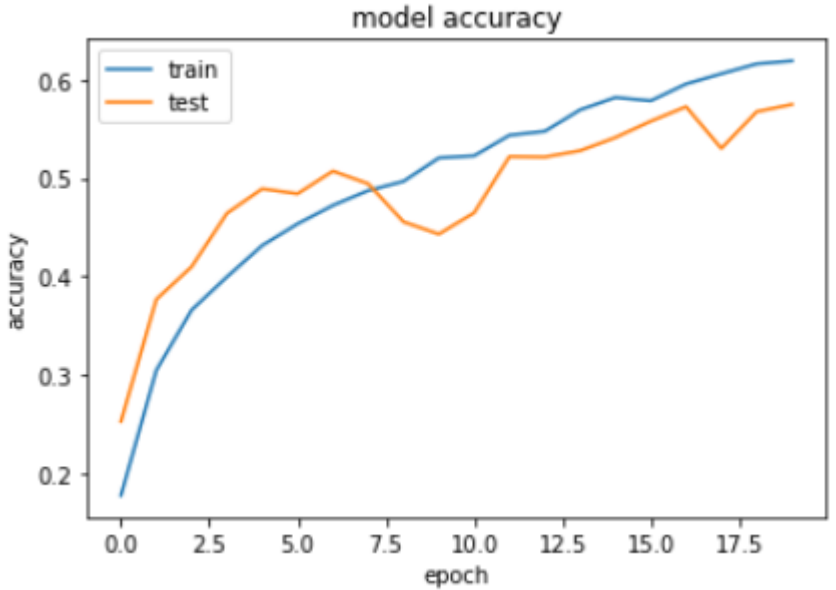
Accuracy: 53% with Augmentation



Modelling

Convolutional neural networks (2D-CNN)

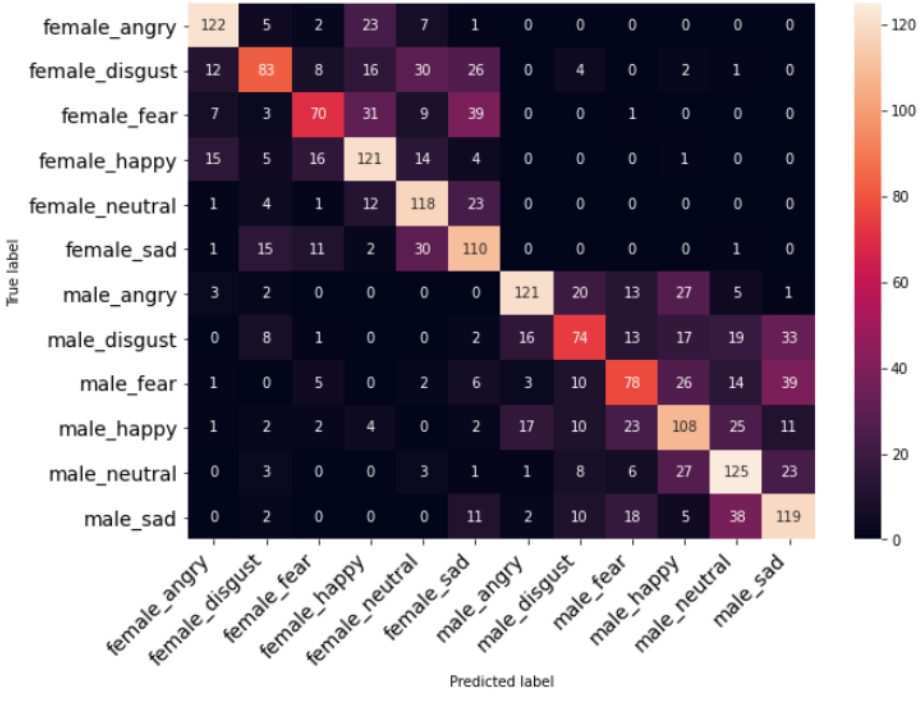




Without Augmentation

Modelling

Accuracy: 57% without Augmentation



Modelling

Model	Data Augmentation	Accuracy
Baseline		21%
CNN-1D	Yes	39%
CNN-2D (MFCC)	Yes	52.1%
CNN-2D (MFCC)	No	57.4%

Table of Contents

Conclusion

With this model developed, accuracy is strengthened and hence predictability will be enhanced.

Recommendations

Application can also be further developed to adapt to real time scenarios for emotive situations.

Further exploration with additional features to analyse other speech variations

THANKS

Do you have any questions?
addyouremail@freepik.com
+91 620 421 838
yourcompany.com