భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Speech and Text Emotion Detection using Different Approaches

Course: AI5100

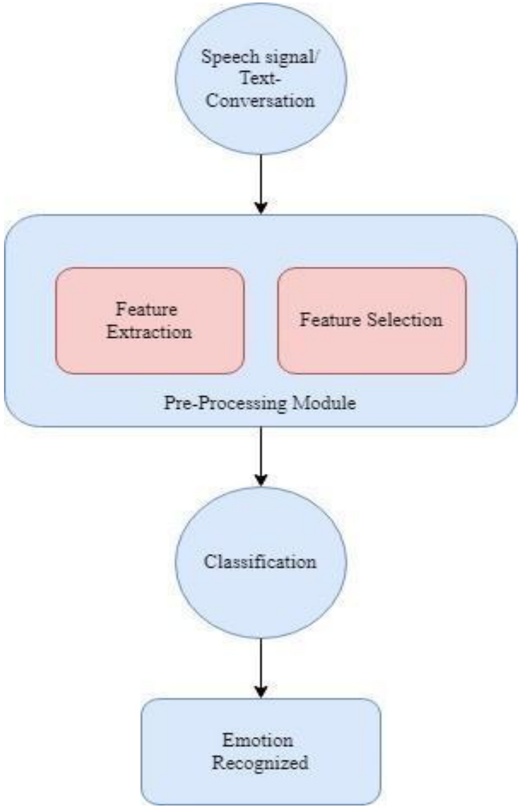
Team members:
Shashank Jerri (AI21MTECH11003)
Aman Ladkat (AI21MTECH14011)
Varshita Sharma(AI21MTECH14009)
Pratik Shetty (AI21MTECH12005)
Maddula Sai Sunamdha Harinhi (AI21MTECH14002)

Instructor:
Dr. Sumohana S. Channappayya
Professor
IIT Hyderabad

Contents

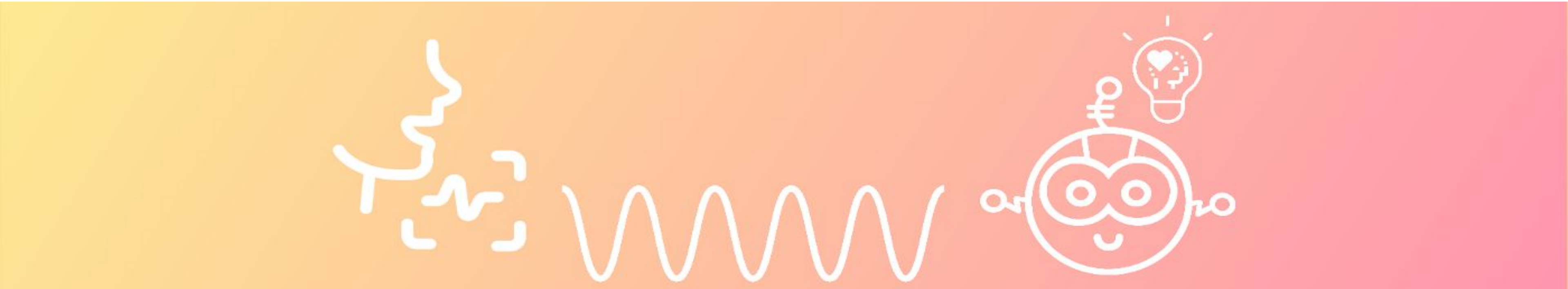
Problem Statement

- Detecting and recognizing human emotion is a big challenge in computer vision and artificial intelligence.
- The main aim of our project is to develop a robust system which can detect as well as recognize human emotions from provided information.
- Information provided is in the form of speech or text.



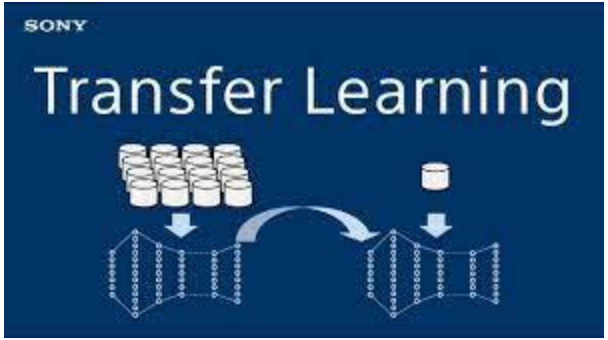
Motivation?

- Emotion recognition provides benefits to many institutions and aspects of life.
- It is useful and important for security, healthcare purposes and robotic applications.
- This system can also be used to detect and recognize racial differences in emotion recognition.
- In the automobile industry, car manufacturers use AI to help them understand human emotions.



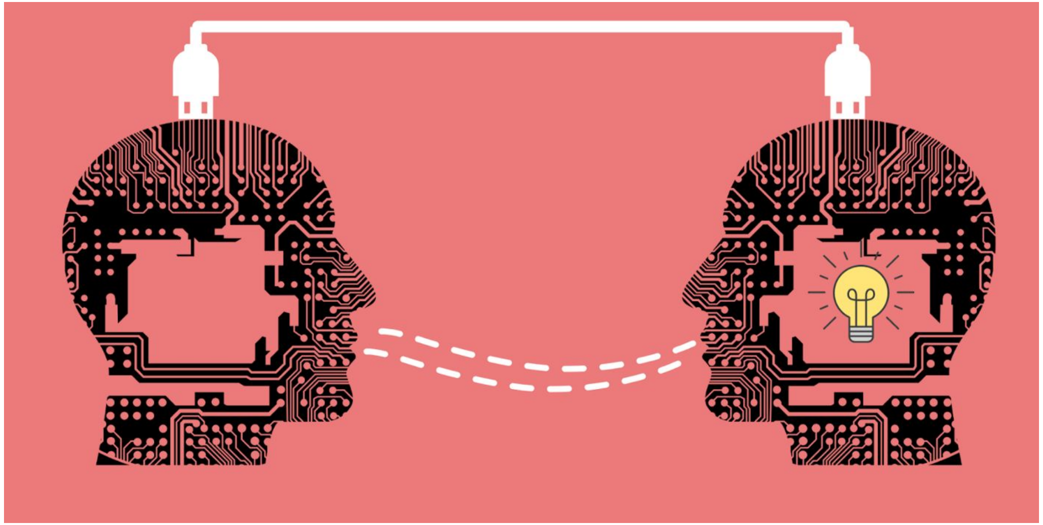
Transfer Learning

- Transfer Learning is a machine learning method where we reuse a pre-trained model as the starting point for a model on a new task.
- It helps us utilize knowledge from previously learned task and apply it newer (related) ones.



Why Transfer Learning?

- Quicker Development
- Less Data
- Better Results



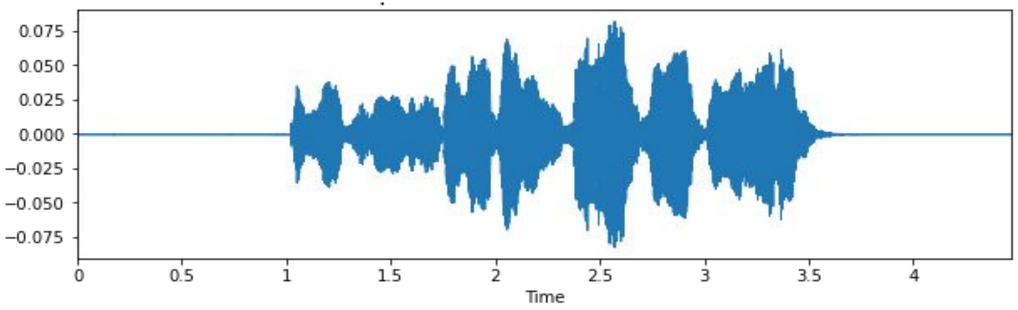
Methodology

Speech Emotion Recognition (using CNN)

- Datasets used: TESS, RAVDESS and SAVEE

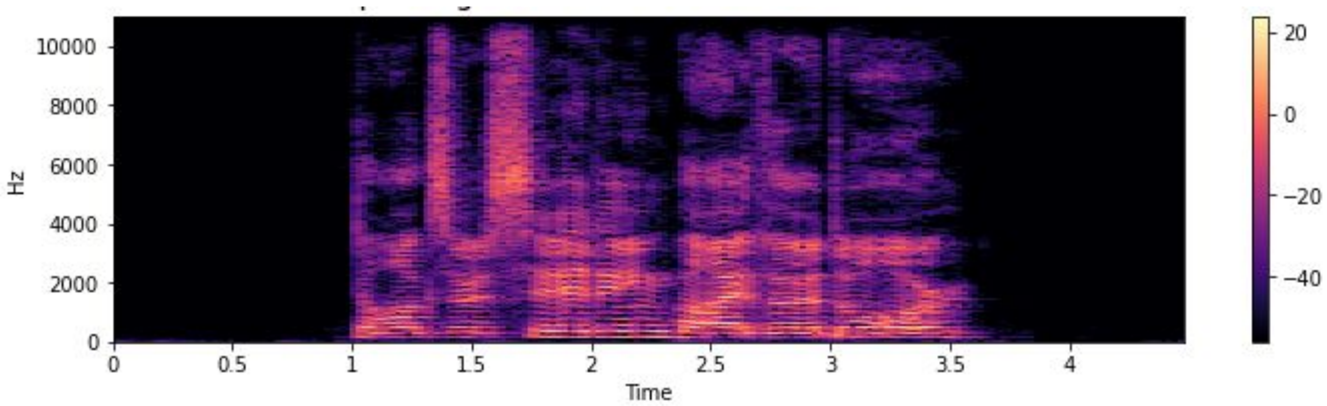
- Step 1: Extracted labels from each dataset separately.

- Step 2: Data Visualization
    - Amplitude envelope of speech signal



Methodology contd.

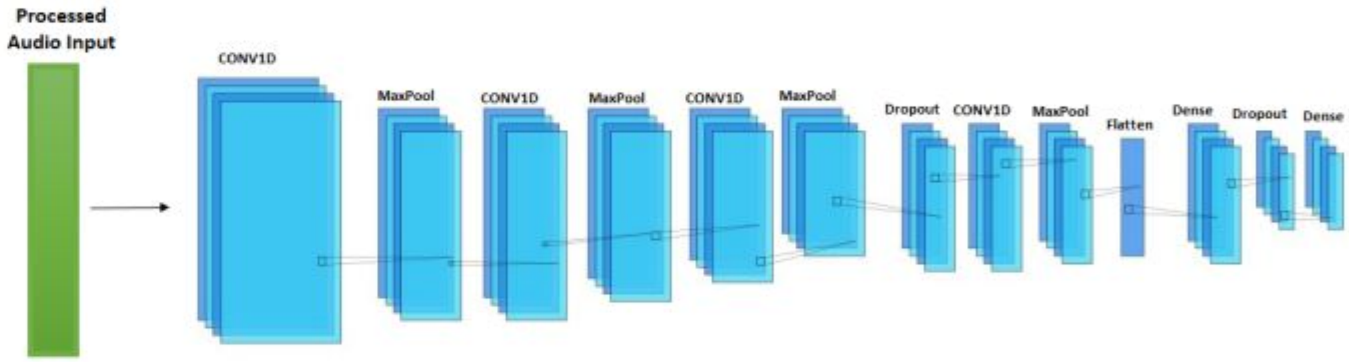
- Spectrogram of audio file



- Step 3: Data Augmentation
  - Added gaussian noise, pitch shift and signal stretching

Methodology contd.

- Step 4: Feature Extraction
  - Features extracted:
    - Zero-crossing rate
    - Chromagram
    - Mel Frequency Cepstral Coefficients (MFCC)
    - Root Mean Square value (RMS)
- Step 5: Training
  - Model used was a CNN with 1D convolution layers



Methodology contd.

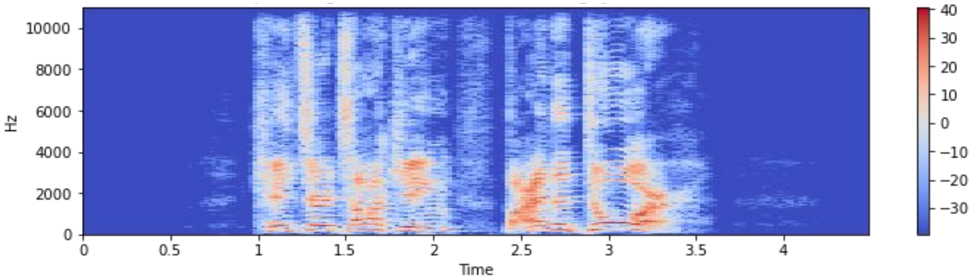
Speech Emotion Recognition (using Transfer Learning)

- Datasets used: TESS, RAVDESS AND SAVEE
- Step 1: Extracted labels from each of the datasets
- Step 2: Feature Extraction
  - Extracted 3D audio spectrogram
- Step 3: Training
  - Used pre-trained AlexNet model

Instance

Input audio

Feature extraction



Testing

ANGRY

Predicted Emotion

One of the extracted features: Spectrogram for input audio

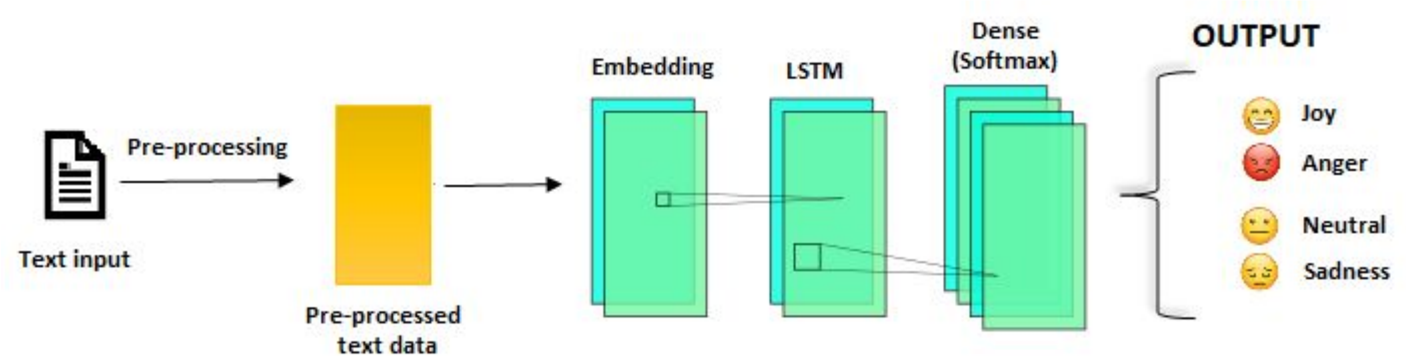
Methodology contd.

Text Emotion Recognition (using LSTM)

- Dataset used: IEMOCAP (text)
- Step 1 - Data Visualization
- Step 2 - Data pre-processing:
    - Label encoded target classes
    - Lower casing
    - Removal of Stop words
    - Stemming
    - Tokenization

Methodology contd.

- Step 3 - Model training:
    - Used an embedding, an LSTM and a dense layer.



Instance

Input text

“No we won't. It's pointless it's like a waiting up to see Santa Claus.”

Pre-process →

“pointless like wait see santa clau”

Predication →

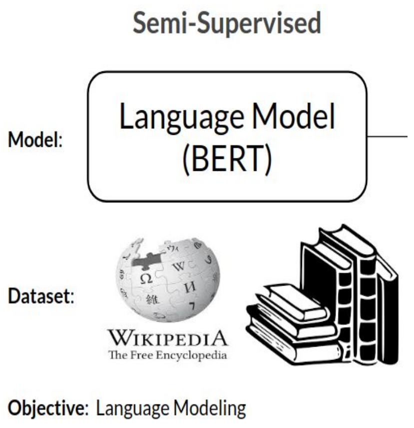
SAD

Predicted emotion

Methodology contd.

Text Emotion Recognition (using Transfer Learning)

- Dataset used : IEMOCAP (text)
- Pre-trained Model : BERT (**bert-based-uncased**)
- Step 1 - Input Formatting
  - Add Special Tokens
  - Fixed Sentence Length and Attention mask
- Step 2 - Tokenization using bert tokenizer
- Step 3 - Train using **BertForSequenceClassification**
- Step 4 - Use Adam optimiser to update model parameters



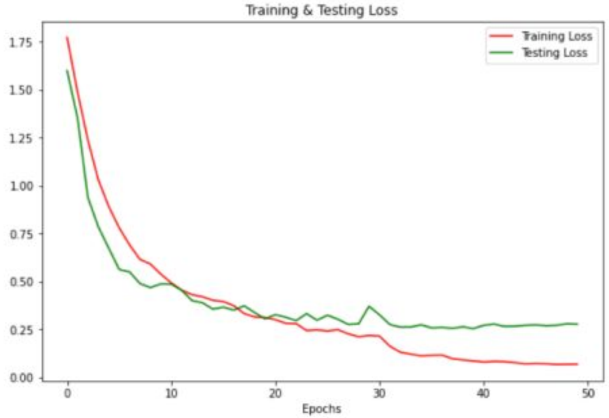
భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

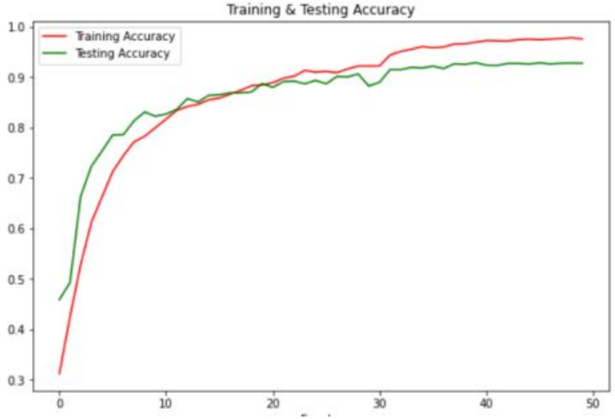
Results

- Speech Emotion Recognition

- Accuracy and Loss

	Training loss	Training accuracy	Testing loss	Testing accuracy
Approach1	0.0692	0.9751	0.2775	0.9270
Approach2	0.0425	0.9871	0.1469	0.9438





Results

- Text Emotion Recognition

- Accuracy and Loss

	Training loss	Training accuracy	Testing loss	Testing accuracy
Approach1	0.2893	0.8771	1.5738	0.6116
Approach2	0.1573	0.9042	1.236	0.8247

Conclusion

- Transfer learning approach for emotion detection and classification performs better than CNN for speech-emotion-recognition and LSTM for text-emotion recognition.

- Transfer learning has potential of leveraging multiple sources of emotion-specific speech/text data to improve emotion recognition performance.

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

References

- Z. Peng, Y. Lu, S. Pan, and Y. Liu, "Efficient Speech Emotion Recognition Using Multi-Scale CNN and Attention," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 3020-3024, doi: 10.1109/ICASSP39728.2021.9414286.
- Zhang, Yuanyuan & Du, Jun & Wang, Zirui & Zhang, Jianshu & Yanhui, tu. (2018). Attention Based Fully Convolutional Network for Speech Emotion Recognition. 1771-1775. 10.23919/APSIPA.2018.8659587.
- Giannoulis, Panagiotis, and Gerasimos Potamianos. "A hierarchical approach with feature selection for emotion recognition from speech." In LREC, pp. 1203-1206. 2012.
- Sharma, J. Jayapradha Soumya, and Yash Dugar. "Detection and recognition of human emotion using a neural network." Int J Appl Eng Res 13, no. 8 (2018): 6472-6477.
- Ghosal, Deepanway, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. "Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation." arXiv preprint arXiv:1908.11540 (2019).
- Guizzo, Eric, Tillman Weyde, and Jack Barnett Leveson. "Multi-time-scale convolution for emotion recognition from speech audio signals." In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6489-6493. IEEE, 2020.

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Group member contributions

Shashank (AI21MTECH11003) - Speech Emotion Recognition

Pratik (AI21MTECH12005) - Text Emotion Recognition

Aman (AI21MTECH14011) - Speech Emotion Recognition

Varshita (AI21MTECH14009) - Text Emotion Recognition

Maddula Sai Sunamdha Harinhi (AI21MTECH14002) - Text Emotion Recognition

THANK YOU!