Speech-Based Sentiment Analysis

Vivek Pandey (M23CSA541) [1] Abhishek Kumar Singh (M23CSA503) [1]

[1]Indian Institute of Technology Jodhpur

Introduction

Automatic detection of emotions and sentiments from speech is an emerging area of research. This project develops a system using CNNs to classify speech recordings into both emotional states and sentiment categories.

Sentiment classification can offer broader emotional trends (e.g., positivity/negativity) that are critical for customer satisfaction analysis or mental health diagnostics.

Applications include virtual assistants, call centers, and psychological diagnostics.

Key Contributions

- Unified emotion and sentiment classification in a single pipeline.
- Leveraged diverse datasets to improve generalization.
- Achieved competitive accuracy with a lightweight CNN model.

Concept Diagram

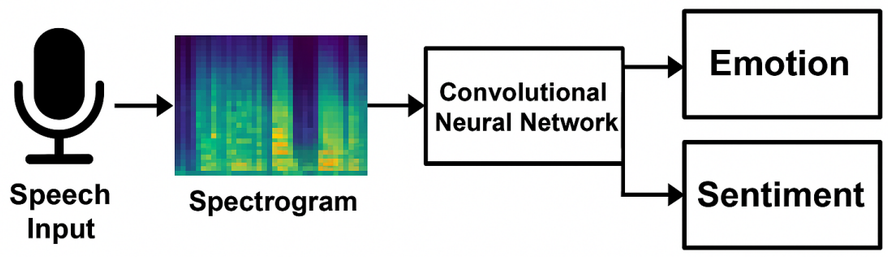

Figure 1. Workflow from Speech Input to Emotion/Sentiment Output

Databases Used

Publicly available datasets employed in this project:

- **RAVDESS** Ryerson Audio-Visual Database
- **TESS** Toronto Emotional Speech Set
- **CREMA-D** Crowd-sourced Emotional Dataset

Problems in Existing Approaches

- Focus only on emotion, overlooking sentiment.
- Poor generalization across datasets.
- Rare use of combined datasets for robust learning.

Preprocessing and Feature Extraction

- Collected and organized RAVDESS, TESS, and CREMA-D datasets.
- Extracted MFCC features using `librosa`.
- Normalized features and padded audio signals.

CNN Model Architecture

- Two CNNs trained:
  - Emotion Classifier: 6 classes happy, sad, angry, fear, neutral, disgust
  - Sentiment Classifier: 3 classes positive, neutral, negative
- Evaluated using accuracy, confusion matrix, and classification report.

Results and Analysis

Emotion Classification

- Accuracy: **83.5%**
- Architecture: 2-layer CNN on MFCC features

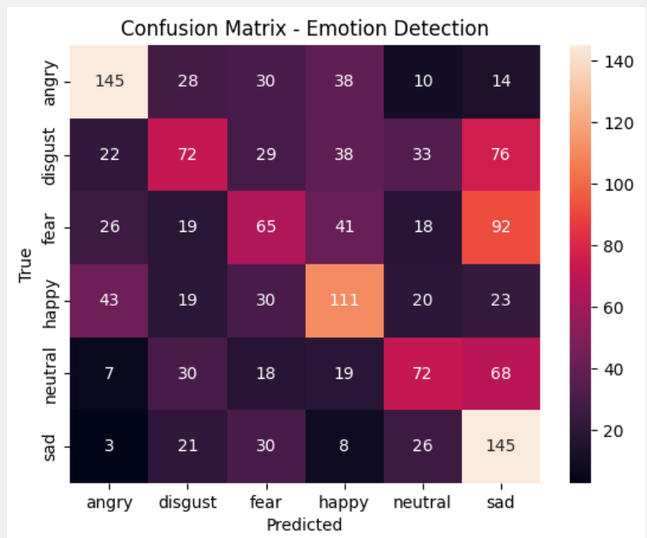

Figure 2. Confusion Matrix: Emotion Classification

Sentiment Classification

- Accuracy: **88.7%**
- Sentiment mapping: e.g., happy → positive, sad → negative

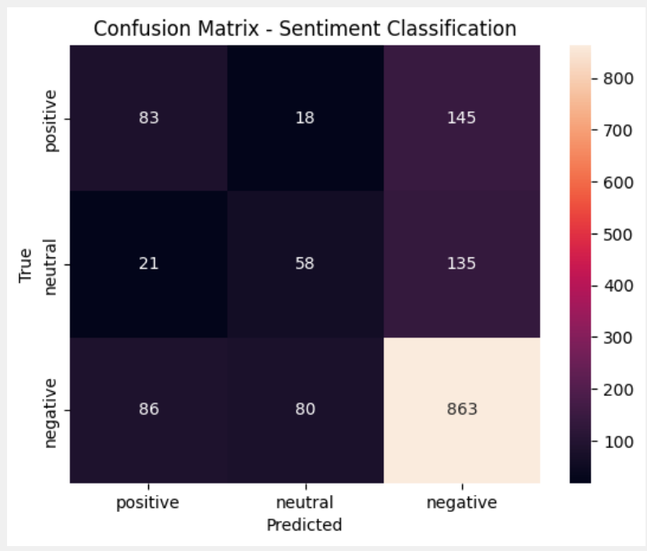

Figure 3. Confusion Matrix: Sentiment Classification

Applications

- Real-time emotion tracking in customer support calls.
- Sentiment analytics for social audio platforms.
- Mental health support tools for mood monitoring.

Conclusion and Future Work

CNN-based models are effective for speech sentiment and emotion detection. Future enhancements could include:

- Data augmentation with background noise.
- Adding LSTM or attention-based models.
- Deployment on real-time or mobile platforms.

References

[1] Crema-d dataset on kaggle.
`https://www.kaggle.com/datasets/ejlok1/cremad`.
Accessed: 2025-04-11.

[2] Keras documentation.
`https://keras.io`.
Accessed: 2025-04-11.

[3] Tess dataset on kaggle.
`https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess`.
Accessed: 2025-04-11.

[4] Steven R. Livingstone and Frank A. Russo.
The ryerson audio-visual database of emotional speech and song (ravdess).
arXiv preprint arXiv:1804.01083, 2018.

[5] Brian McFee et al.
librosa: Audio and music signal analysis in python, 2015.
Available at: https://librosa.org.

m23csa503@iitj.ac.in Department of Computer Science, IIT Jodhpur — April 2025 m23csa541@iitj.ac.in