

# SPEECH EMOTION RECOGNITION

## ABSTRACT

The last few years have witnessed the capacity of machines to recognize and react to human emotions become increasingly important in **Human-Computer Interaction (HCI)**. Speech-emitted emotions convey important information that, if properly understood, can make intelligent systems more responsive and empathetic. This project outlines the development and design of a real-time **Speech Emotion Recognition (SER)** system that recognizes a speaker's emotional state from audio signals based on deep learning and audio signal processing methods.

The core idea behind this system is to use the **mel spectrogram** representation of speech—a time-frequency representation that retains salient speech features—and use a **Convolutional Neural Network (CNN)** model to predict the underlying emotion. The system has been trained on a heterogeneous, merged set of four publicly released emotion datasets: RAVDESS, CREMA-D, TESS, SAVEE. The datasets were pre-processed and augmented to improve model robustness and provide balanced class representation across seven standardized emotion classes: “Happy”, “Sad”, “Angry”, “Calm”, “Fearful”, “Disgust”, and “Surprised”.

The trained CNN model is incorporated within a Flask-based web application to allow users to record their voice within the browser. When the voice is recorded, the voice is automatically translated into a spectrogram image, which is used as input to the model to return the corresponding emotion prediction, which represents **Real-time Emotion Detection**. For better user experience, the application returns emotion-specific contextual feedback rather than raw labels like "Happy" or "Sad.". This system shows an end-to-end pipeline—audio acquisition to prediction and visualization—and points towards the potential of emotion-aware applications in virtual assistants, remote learning, interactive voice response, and mental health support software. Its simplicity, real-time, and interactive nature make it a desirable prototype for emotionally intelligent systems in the future.

**Keywords** : Speech Emotion Recognition (SER) ; Real-Time Emotion Detection ; Convolutional Neural Network (CNN) ; Mel Spectrogram ; Human-Computer Interaction (HCI) .

SAHITHI DULLA (22071A6714)

MISHA MANMAI GANJI (22071A6716)

CHARAN SAI GUNDU (22071A6719)