### **Abstract:**

This project focuses on **Speech Emotion Recognition (SER)** using a **1D Convolutional Neural Network (Conv1D)**. The dataset consists of six emotions: *sad, happy, disgust, neutral, angry (modified), and fear*. Audio files are preprocessed using **Librosa** for feature extraction, including **MFCC, chroma, and mel spectrograms**. Data augmentation techniques such as **noise addition** and **time shifting** are applied to enhance robustness.

A custom **Conv1D model** is designed with multiple convolutional and pooling layers, followed by dropout and fully connected layers. The dataset is split into **training and validation sets**, and the model is trained using **CrossEntropyLoss** with **Adam optimizer**. The training process is monitored for **loss and accuracy**.

It processes the input through a series of convolutional, ReLU activation, max-pooling, and dropout layers to extract features at multiple levels.

The goal is to develop a robust deep-learning model capable of accurately classifying emotions from speech signals, which can be applied in **human-computer interaction, AI-driven assistants, and mental health monitoring**.

**📌 Key Features**

* **Dataset:** Custom dataset with 50 samples per emotion
* **Preprocessing:** Feature extraction using **Librosa** (MFCCs, Chroma, Mel Spectrogram)
* **Data Augmentation:** Noise addition and time shifting for robustness
* **Model Architecture:** Conv1D-based deep learning model with multiple convolutional and pooling layers
* **Activation Function:** Uses **ReLU** for non-linearity and better feature extraction
* **Training:** Optimized using **Adam** with **CrossEntropyLoss**
* **Applications:** Human-computer interaction, AI-driven assistants, mental health monitoring