Speech Emotion Recognition Using Deep Learning

Introduction

This project implements a deep learning model to recognize emotions from speech audio. It uses Mel-Frequency Cepstral Coefficients (MFCCs) for feature extraction and a hybrid model of Convolutional Neural Networks (CNN) and Bidirectional LSTM layers for classification.

Dataset: RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song)

Emotions classified: Neutral, Calm, Happy, Sad, Angry, Fearful, Disgust, Surprised

Model Overview

Model Architecture:

- Conv1D layer with 64 filters
- MaxPooling1D and Dropout
- Bidirectional LSTM with 64 units
- Dense layers for classification
- Softmax activation for multi-class output
- Categorical Crossentropy loss and Adam optimizer

Instructions

To run the project:

1. Install required packages:

   pip install numpy librosa matplotlib seaborn scikit-learn keras tensorflow

2. Download and extract the RAVDESS dataset into a folder named 'RAVDESS'.

3. Run the Python script to train the model and evaluate performance.

4. The plot shows accuracy and loss over training epochs.

Speech Emotion Recognition Using Deep Learning

Training Accuracy and Loss