SPEECH EMOTION CLASSIFICATION USING RAVDESS

INTRODUCTION

Goal: Build a system to classify emotions in speech using deep learning.
Dataset: RAVDESS Emotional Speech Dataset taken from Kaggle

APPROACH

1. Data Preparation:
   - Used the RAVDESS dataset containing labeled emotional speech.
   - Audio preprocessing: resampling to 16kHz, trimming silence, normalizing amplitude.
   - Extracted MFCC features (Mel-frequency cepstral coefficients), widely used for speech tasks.
2. Feature Representaion:
   - Each audio sample converted to an MFCC feature matrix (timesteps × n_mfcc).
   - Standardized/padded sequences to fixed length for RNN input.
3. Modeling:
   - Designed an LSTM-based Recurrent Neural Network (RNN) to capture temporal dependencies in speech.

MODELS

1. RNN (LSTM) Model
   Input: (200 timesteps, 40 MFCC features).
   Hidden layers: stacked LSTMs with dropout regularization.
   Dense output layer with softmax across emotion categories.
2. Label Encoder used to map numeric predictions back to human-readable labels.
3. Model saved as rnn_emotion_model.h5 and deployed in a Streamlit app (app.py).

RESULTS

1. The model achieved reasonable accuracy 85.6%

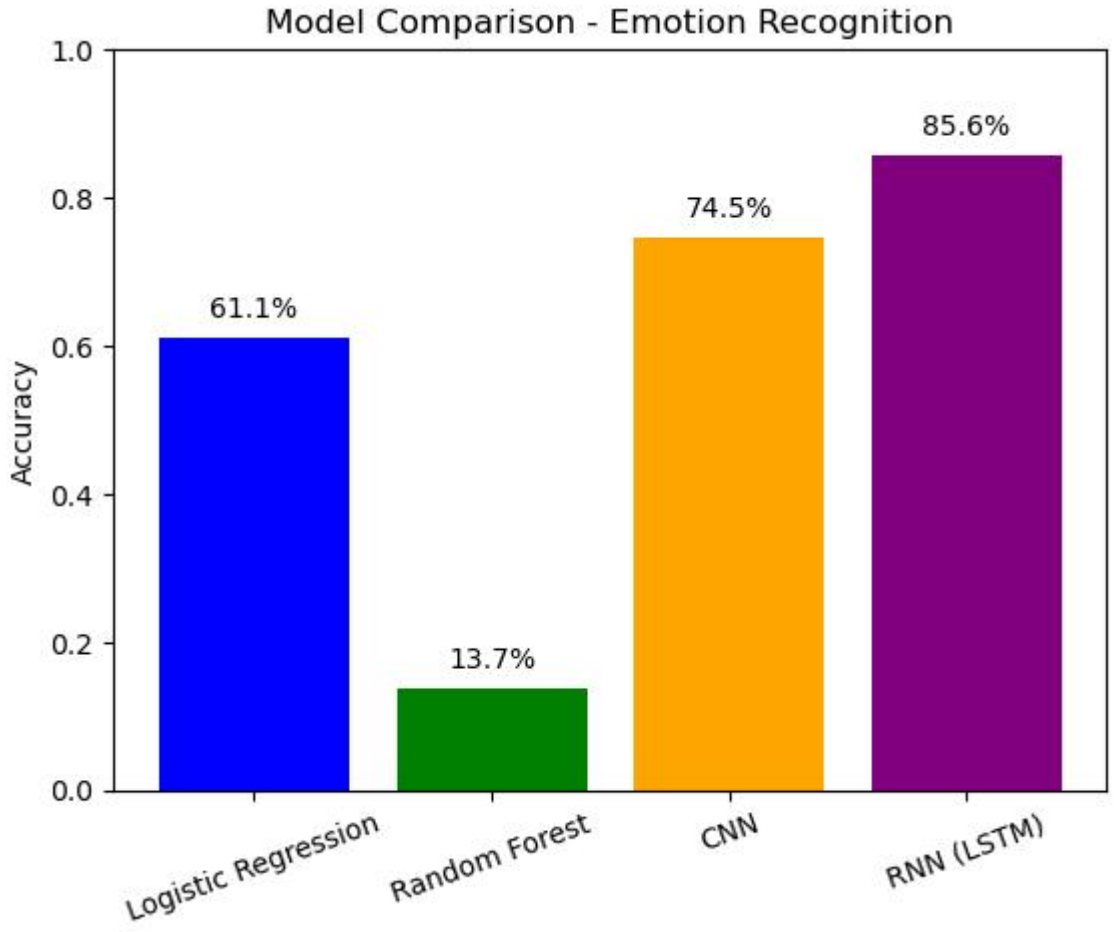
2. Demonstrated ability to distinguish common emotions such as neutral, happy, sad, angry, fear, disgust, surprise, calm.

3. The Streamlit app allows users to upload a .wav file, extracts MFCC features, runs the RNN model, and outputs:

Predicted Emotion

Probability distribution across classes (visualized as a bar chart).

COMPARISION OF DIFFERENT MODELS



CHALLENGES

1. Sequence Lengths: Audio clips were of different lengths, so padding/truncating was needed before feeding into the RNN.

2. Training Time: Model training with RNN took longer compared to simple ML models.

3. Deployment Setup: Integrating the trained model with Streamlit and handling uploaded files smoothly.