DeepEMO: Deep Learning for Speech Emotion Recognition

Presented by :

1805104 - Swarnali Saha

1805116 - Sanjida Islam Era

Base Paper Details

Paper Name: DeepEMO: Deep Learning for Speech Emotion Recognition.
Published Date : 9 September 2021(Preprinted repository is available on arXiv)
Resource Link: https://arxiv.org/pdf/2109.04081.pdf

Problem Definition
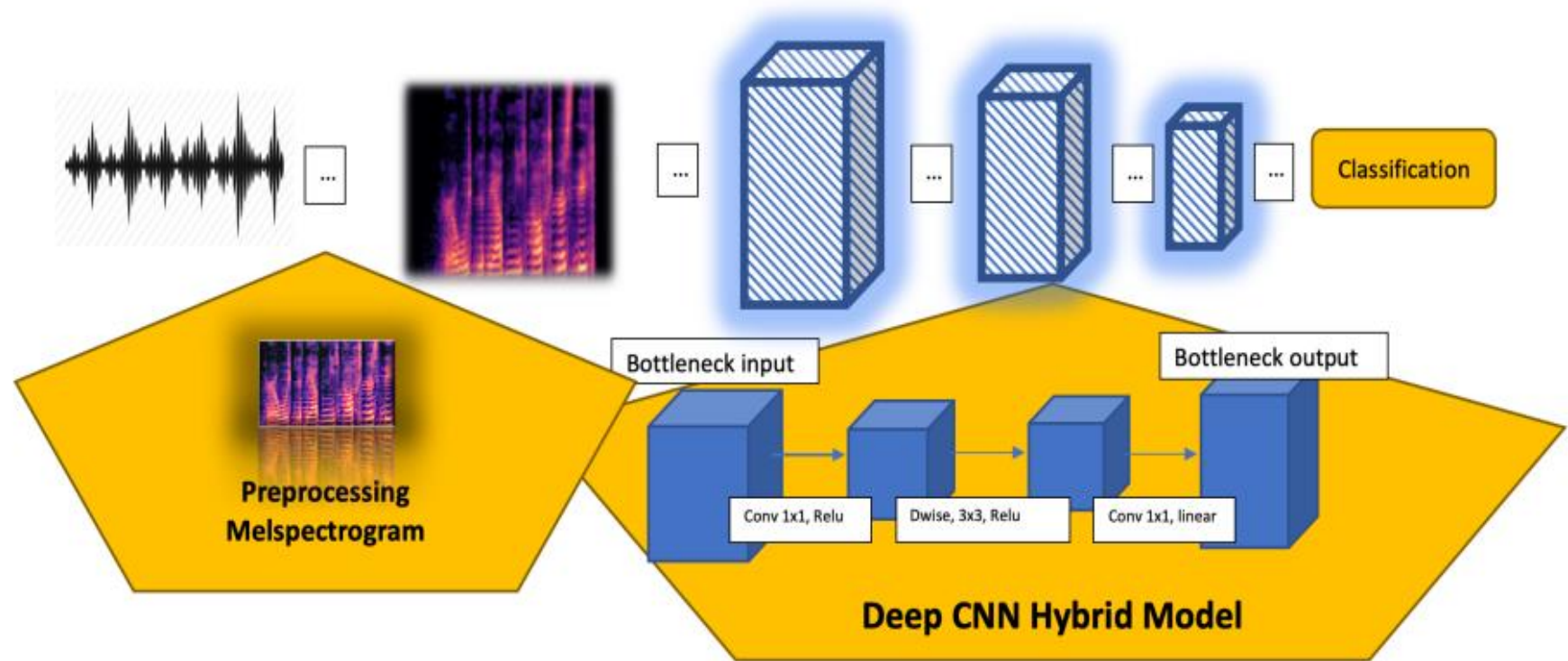
Input: Speech data in natural language.
Output: Accurate emotion recognition.

Key Idea

The key idea of the paper is to employ deep transfer learning and efficient feature extraction for robust speech emotion recognition

Solution Overview

Classification

Bottleneck input

Bottleneck output

Preprocessing Melspectrogram

Conv 1x1, Relu

Dwise, 3x3, Relu

Conv 1x1, linear

Deep CNN Hybrid Model

Feature Extraction

Fast Fourier Transform (FFT)

transforms the audio from its original form into a format that shows what frequencies are present in the sound.

Mel Scale Generation

The mel scale is a way to measure frequency that matches how humans perceive pitch.

Spectrogram Generation

It represents the changes over time in a visual format (spectrogram) that can be used for further analysis in deep learning models.

Deep Transfer Learning

❑ It used the deep convolutional backbone model that is ResNet18 which consists of assemble of convolutional layers

❑ Cross Entropy Loss function (CE) and Adam optimize implemented for the model.

Code and Data Availability

Code

DeepEMO Codebase

Dataset

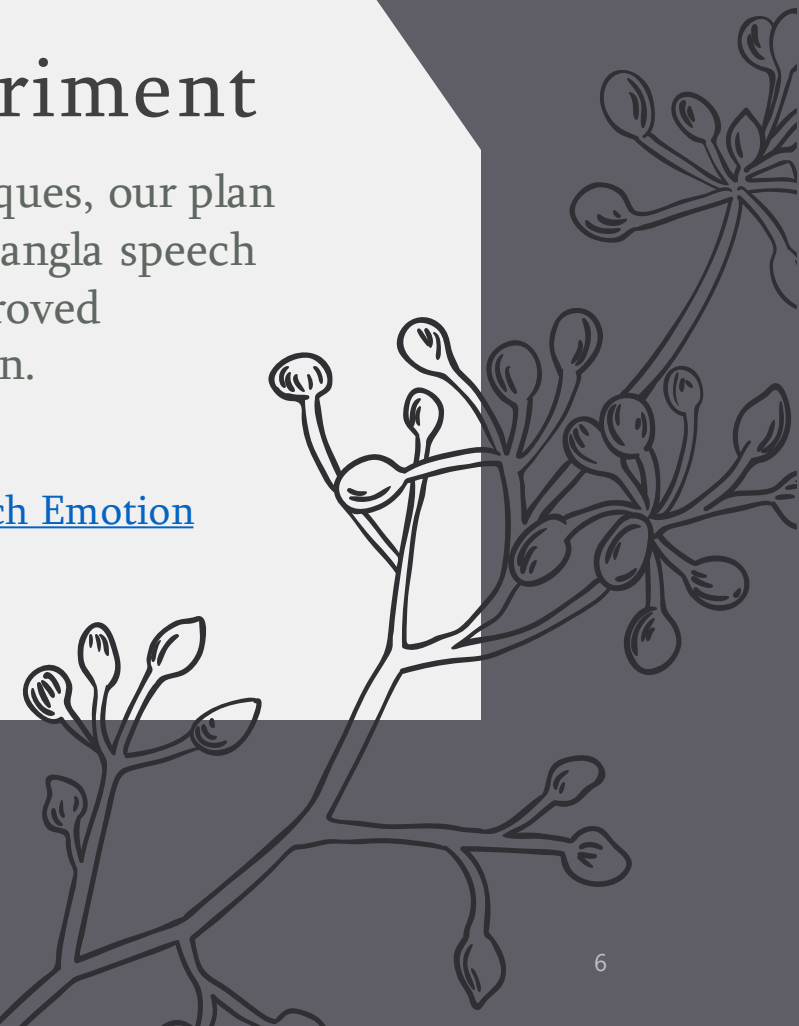RAVDESS Emotional speech audio Dataset

Proposed Experiment

Adapting DeepEMO's techniques, our plan is to apply deep learning to Bangla speech emotion recognition for improved interpretation and articulation.

Bangla Dataset Link:
[BASER-Bangla Adolescent Speech Emotion Recognition Dataset](#)

Thank You!