SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE

Department of AI and Machine Learning

SYMBIOSIS Golden Jubilee Celebrating 50 Years of Excellence

PROJECT TITLE : Implementation of Speech based emotion recognition on RAVDESS emotional speech Dataset

INTRODUCTION

Emotion can play an important role in decision making. Emotion can be detected from different physiological signal also. If emotion can be recognized properly from speech then a system can act accordingly. Identification of emotion can be done by extracting the features or different characteristics from the speech and training needed for a large number of speech database to make the system accurate.

MOTIVATION

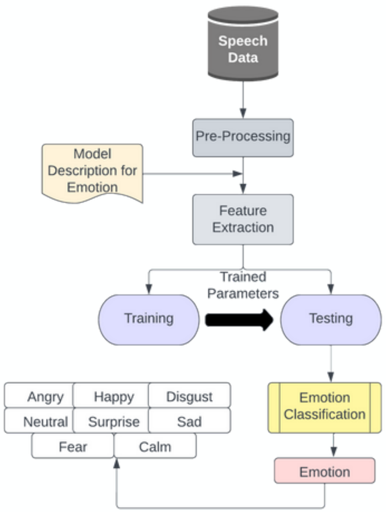
- Improving communication
- Mental health applications
- Marketing and advertising
- Entertainment

OBJECTIVES & AIMS:

- To build a model to recognize emotion from speech using the librosa and sklearn libraries and the RAVDESS dataset.
- To present a classification model of emotion elicited by speeches based on ML Classification based on acoustic features such as Mel Frequency Cepstral Coefficient (MFCC). The model has been trained to classify eight different emotions (calm, happy, fearful, disgust, angry, neutral, surprised, sad).
- To deploy the model on such a platform, such that user can be able to Test and interact with the Model .

METHODOLOGY:

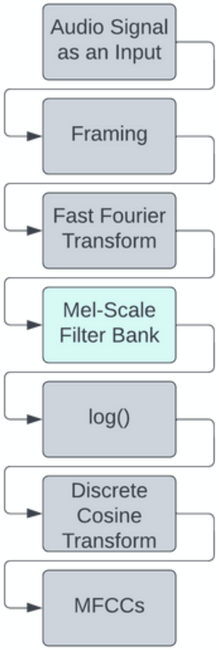
Work Flow Flowchart



Algorithm Used :-

- Decision Tree
- Random Forest
- Support Vector Machines (SVM)
- Naive Bayes
- K-Nearest Neighbors (KNN)
- Multilayer Perceptron (MLP)
- Convolutional Neural Network (CNN)
- Long Short Term Memory (LSTM) And Convolutional Neural Network (CNN)

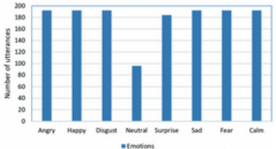
Feature Extraction - (MFCCs)[1]



Other Useful Feature Extraction Methods are -

- ZCR (Zero-Crossing Rate)
- Mel Spectrogram
- Time-domain features
- Frequency-domain features
- Chroma features
- Wavelet-based features
- Spectro-temporal features

DATASET : RAVDESS (Ryerson audio-visual database of emotional speech and song [2])



RAVDESS Data Count

Gender	Count	Trials per Actors	Total Count
Male	12	60	720
Female	12	60	720
Total			1440

RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) is a database of emotional speech and song. The RAVDESS dataset contains a total of 1440 files (1.15 GB) of audio, spoken and music. It is a multimodal and dynamic set of North American English facial and vocal expressions. The collection features 24 performers who deliver two lexically matched lines in a neutral North American accent. Neutral, calm, happy, sad, angry, terrified, disgusted, and astonished are the eight speech emotions. Each expression has two strength levels: strong and normal, with a neutral expression as an extra. All data is provided in three modes: audio-video, audio solo, and video only. In this investigation, only the audio files, comprising 1440 files (24 actors * 60 trials per actor), are used.

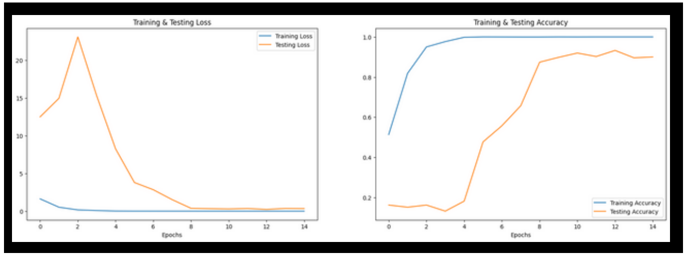
RESULT :

Comparative Analysis of the Algorithms :

Algorithm	Accuracy
Decision Tree	49%
SVM	89%
Naive Bayes	46%
KNN	77%
MLP	84%
CNN	93%
LSTM & CNN	82%

From here we can conclude that CNN we are achieving the highest accuracy of our model i.e 93%.

Because we are working with CNN, we need also examine the loss function graph.



REFERENCE:

[1] https://www.mdpi.com/1424-8220/22/6/2378

[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8898841/

CONCLUSION AND FUTURE SCOPE:

This is the first time, to the best of our knowledge, that we have used the RAVDESS dataset to apply Machine Learning approaches to our model. . In the future, we plan to look more into deep learning algorithm, . in such a way that our model will give us the highest possible accuracy .

GROUP MEMBERS : Dhrubo Bhattacharjee (21070126026) Devansh Tiwari (21070126025) Harshit Jain (21070126034)

PROJECT MENTOR NAME : Dr.Preksha Pareek