JSS SCIENCE AND TECHNOLOGY, SJCE ,
Mysuru.
Department of computer science and engineering

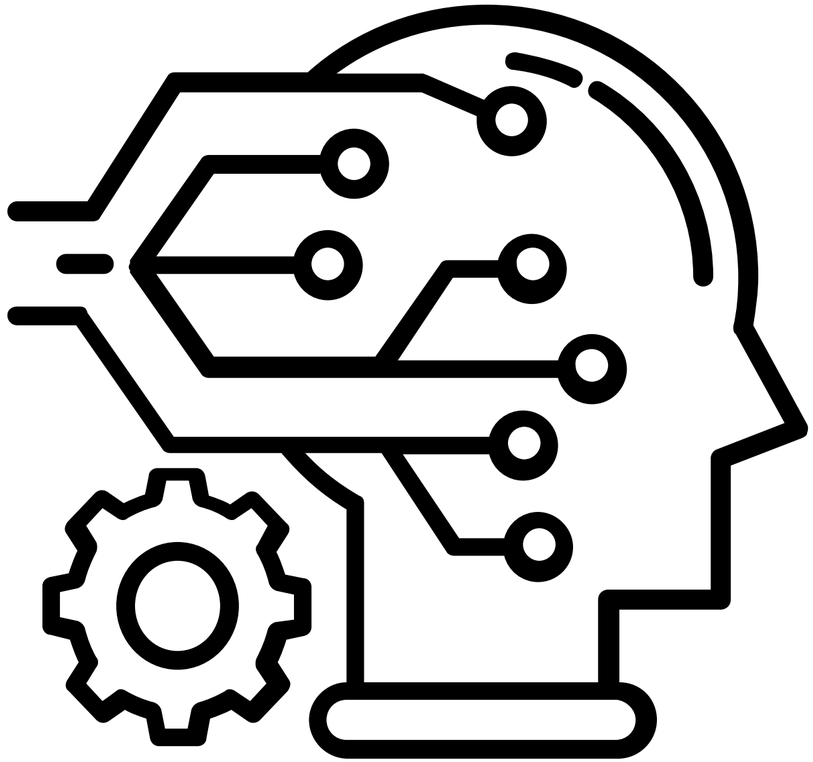
SPEECH EMOTION RECOGNITION

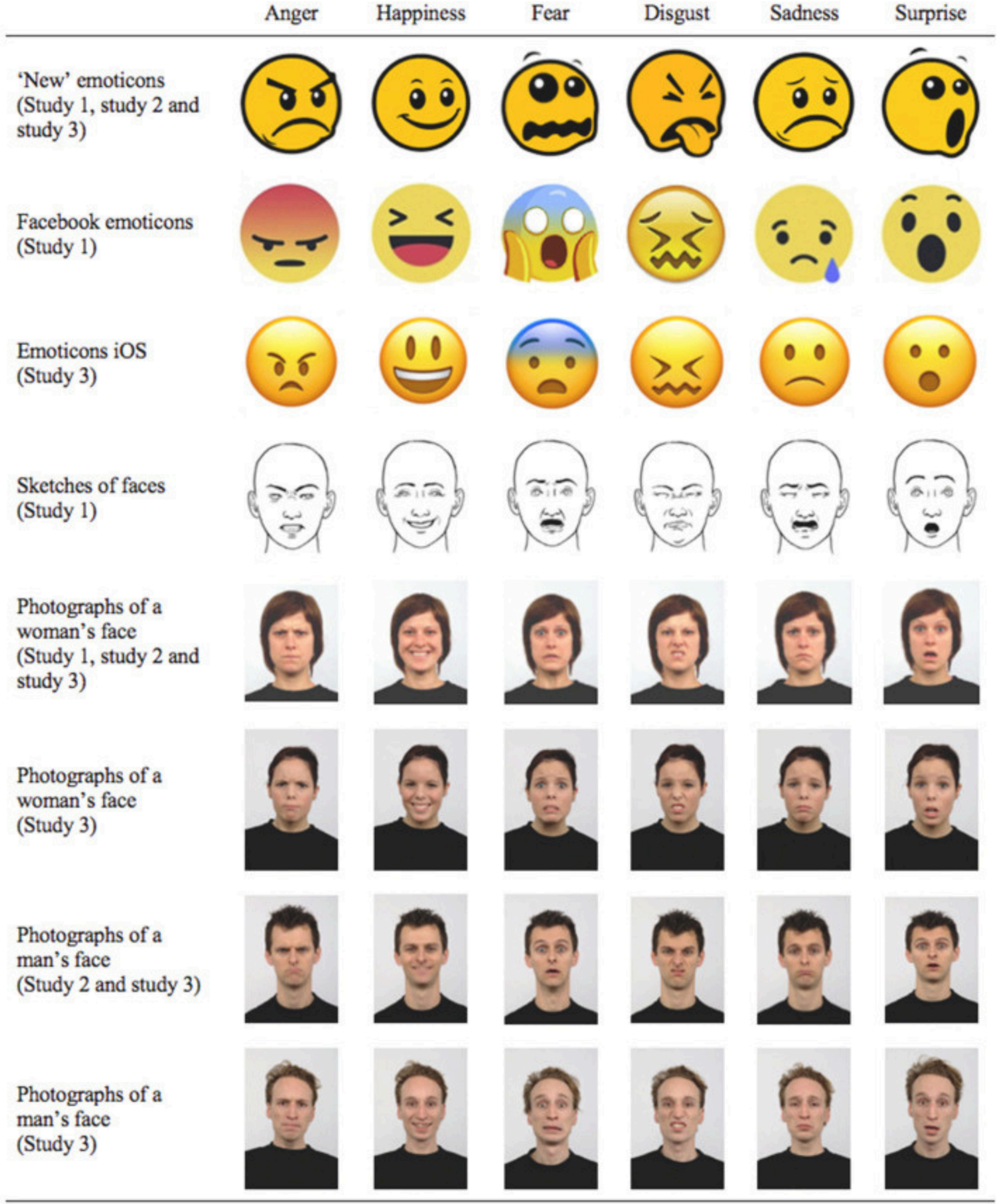
Project work by :

Neethushree GN-01JCE21CS131
Bhoomika ks-01JCE21CS019
Nandan KR-01JST21CS089

UNDER THE GUIDENCE
Prof. Rakshitha R

CONTENTS





SPEECH EMOTION RECOGNITION USING LIBROSA

INTRODUCTION

Speech Emotion Recognition involves the analysis of vocal cues to identify and interpret the emotional states conveyed by a speaker. By leveraging sophisticated machine learning algorithms, we can train systems to detect emotions such as happiness, sadness, anger, and fear from speech patterns. This process involves extracting features from audio signals, such as pitch, tone, and rhythm, and using these features to build predictive models.

LITERATURE REVIEW

1]Speech Emotion Recognition using Neural Network and MLP Classifier(Jerry Joy,Aparna Kannan,Shreya Ram,Rama)
MLP Classifier
5 features extracted-MFCC,Contrast,Mel Spectrograph Frequency,Chroma and Tonnetz
Accuracy 70.28%
[2]Voice Emotion Recognition using CNN and Decision Tree(Navya Damodar,Vani H Y,Anusuya M A.)
Decision Tree,CNN
MFCCs extracted
Accuracy 72% CNN,63% Decision Tree
[3]Speech Emotion Recognition System With Librosa[IEEE]
(V.Siva Nagaraju,P.Ashok Babu,Rajeev Ratna Vallabhuni)
Decision Tree,CNN,MLP Classifier,librosa
MFCCs extracted
Accuracy 81% CNN,72%

PROPOSED SYSTEM

In the project MFCC has been used as the feature for classifying the speech data into various emotion categories employing Deep neural networks. The usage of the Neural Networks provides us the advantage of classifying many different types of emotions in a variable length of audio signal in a real time environment.
This technique manages to establish a good balance between computational volume and performance accuracy of the real-time processes.

OBJECTIVES

The objectives of a Speech Emotion Recognition (SER) project typically revolve around understanding and analyzing the emotional content conveyed through speech.

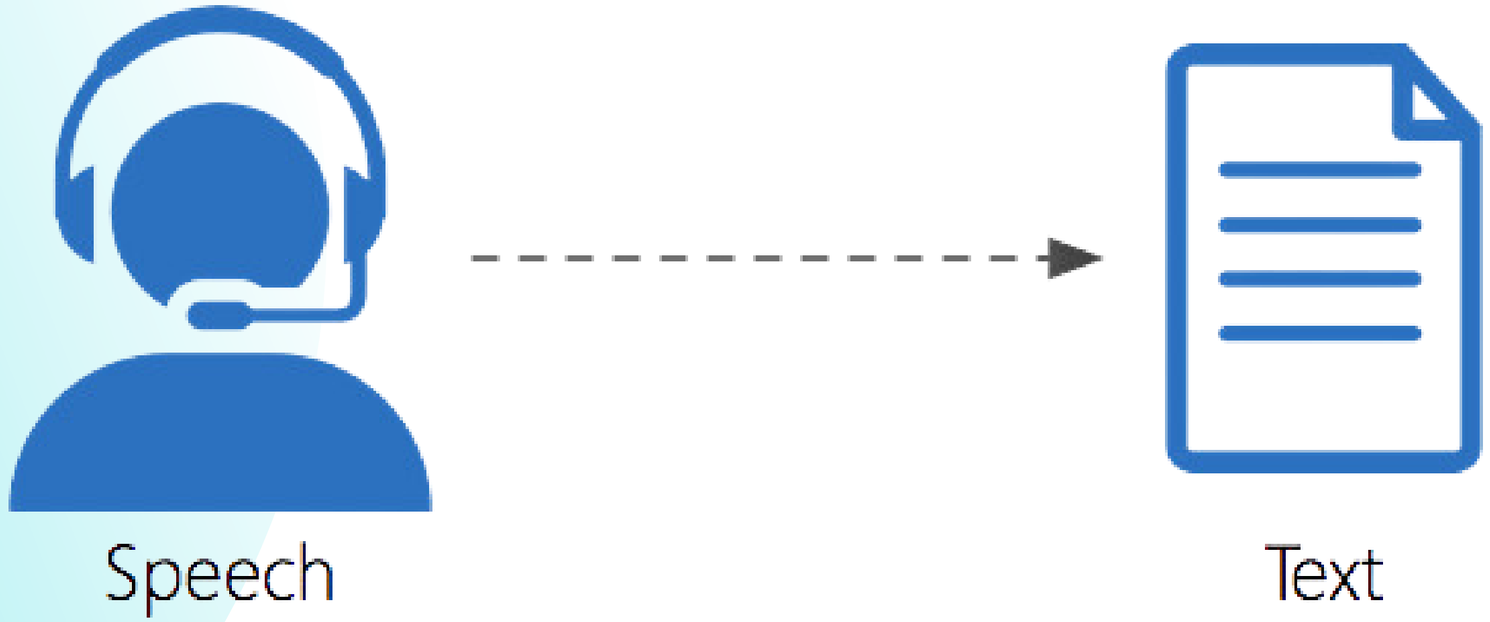
Other objectives of our project are:

- Emotion Classification
- Feature Extraction
- Cross-culture Validity
- Robustness

DESIGN

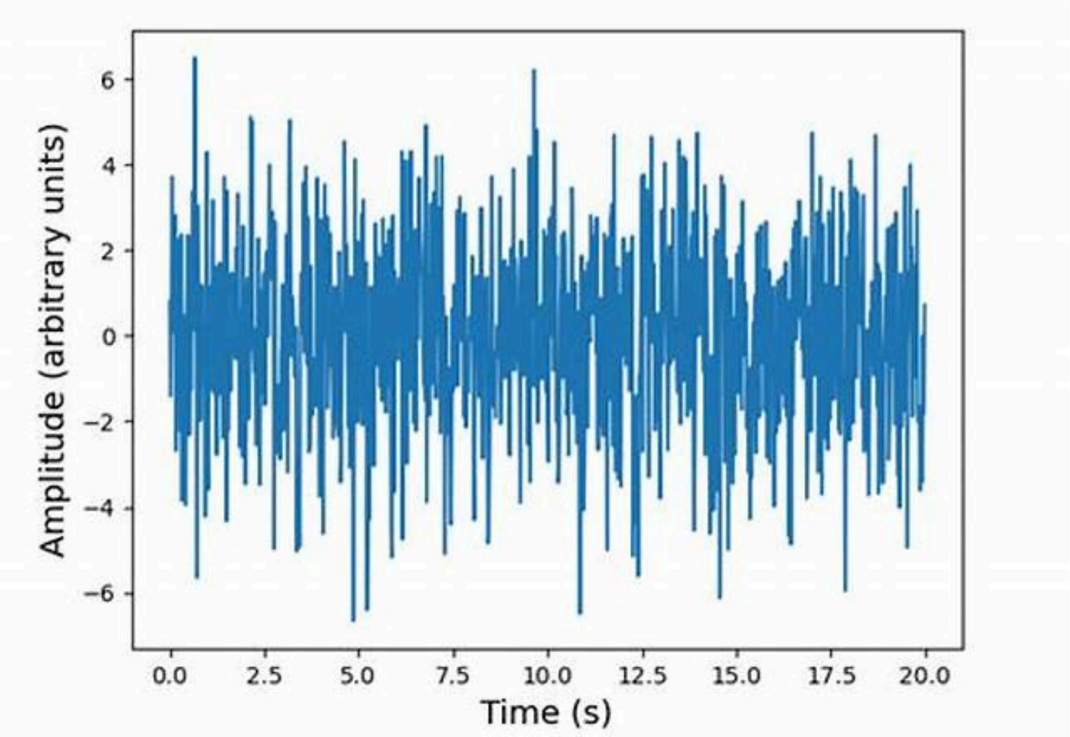
SPEECH TO TEXT TRANSLATION USING SPEECH EMOTION RECOGNITION API

Initially we tested the audio by translating it back into the text mode using the Speech Emotion API to know what the audio is all about.



ANALYSING AUDIO SIGNALS

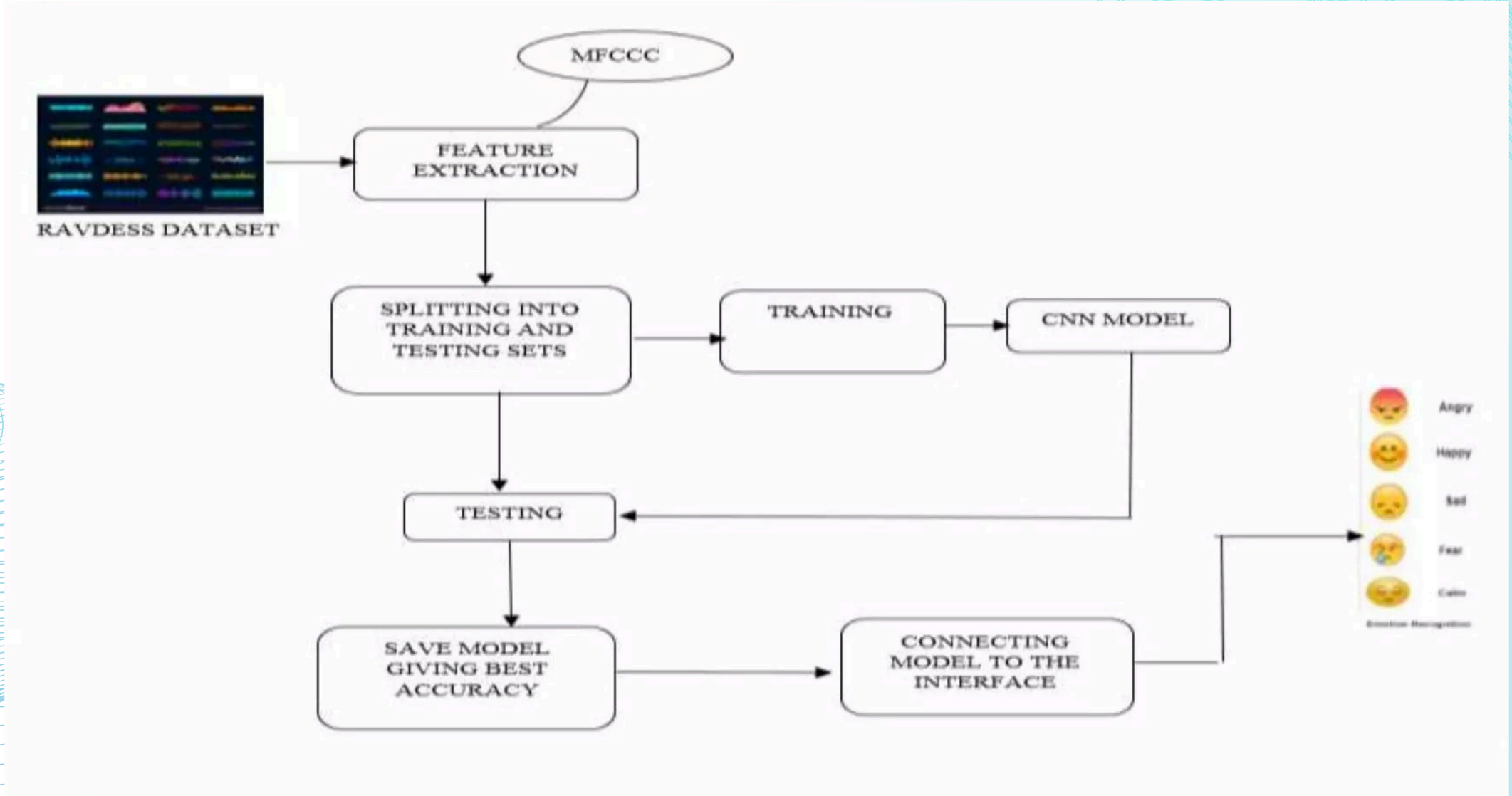
Second step is to test out the audio files by plotting out the waveform and spectrogram to see the sample audio files.



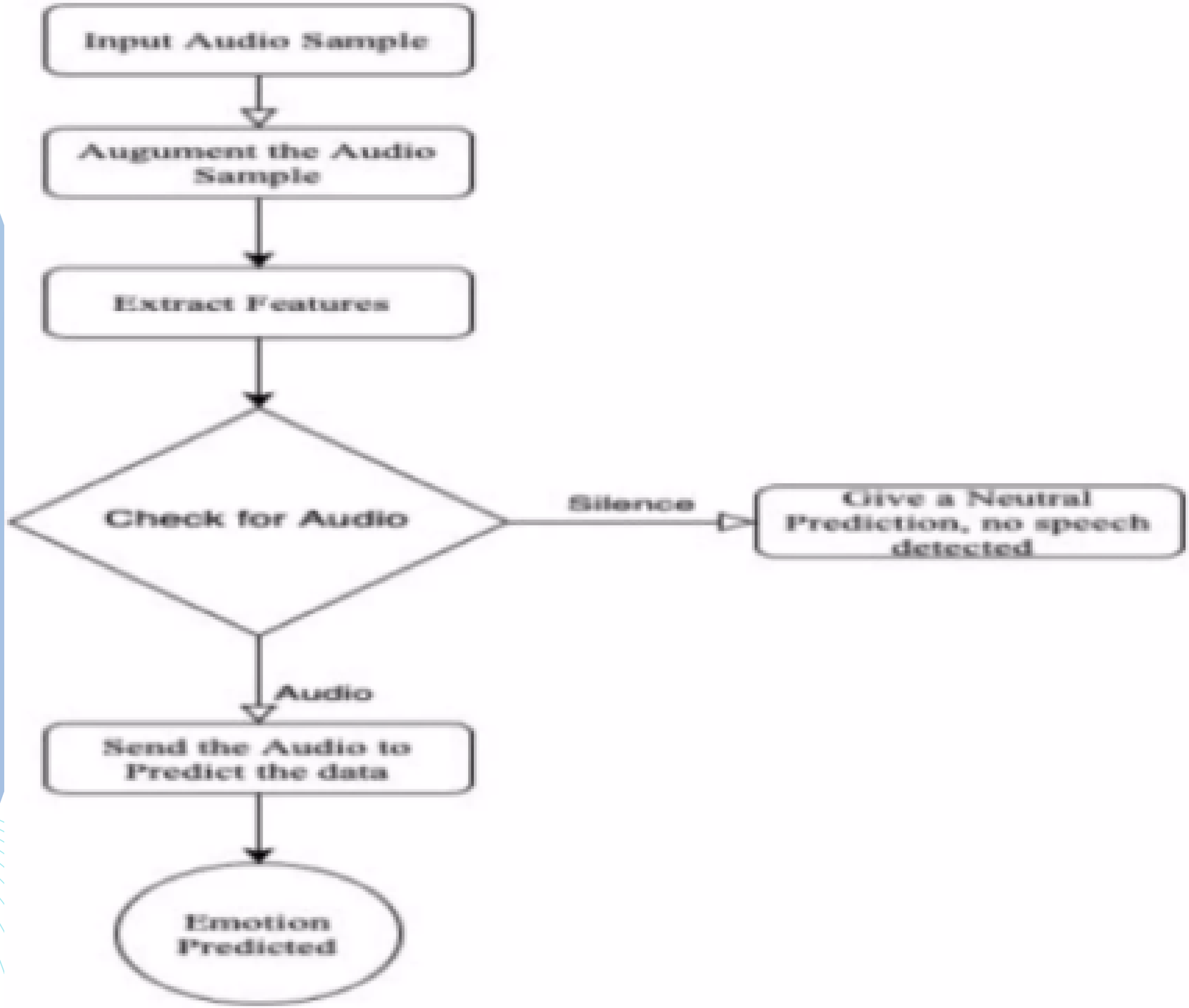
MASKING AND CLEANING OF AUDIO FILES

Next Step is to clean the audio files by lowering down the sample rate and removing the unwanted noise around the raw audio via MASKING

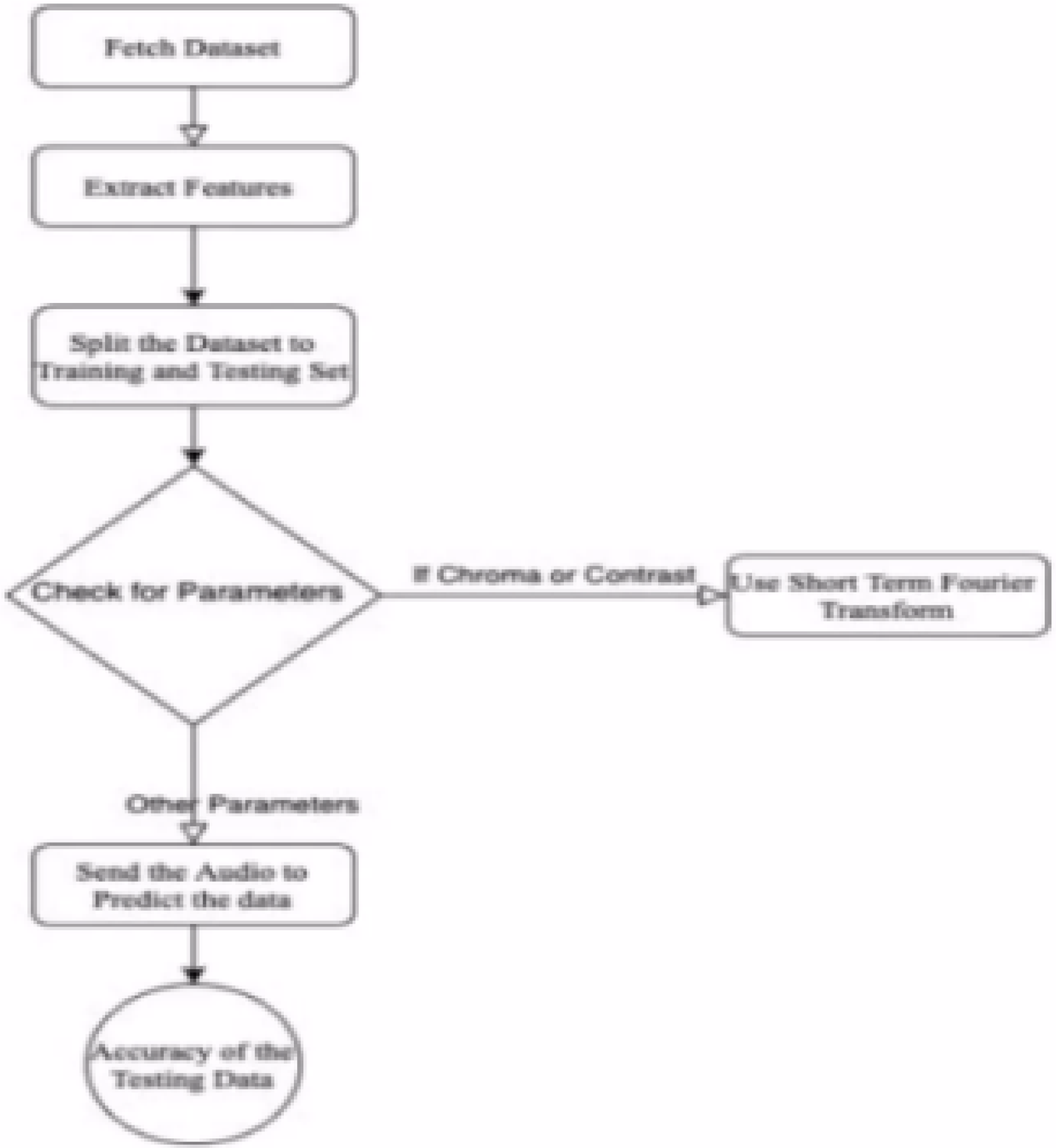
SYSTEM BLOCK DIAGRAM



TRAINING PROCESS WORKFLOW



TESTING PROCESS WORKFLOW



DATASET

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

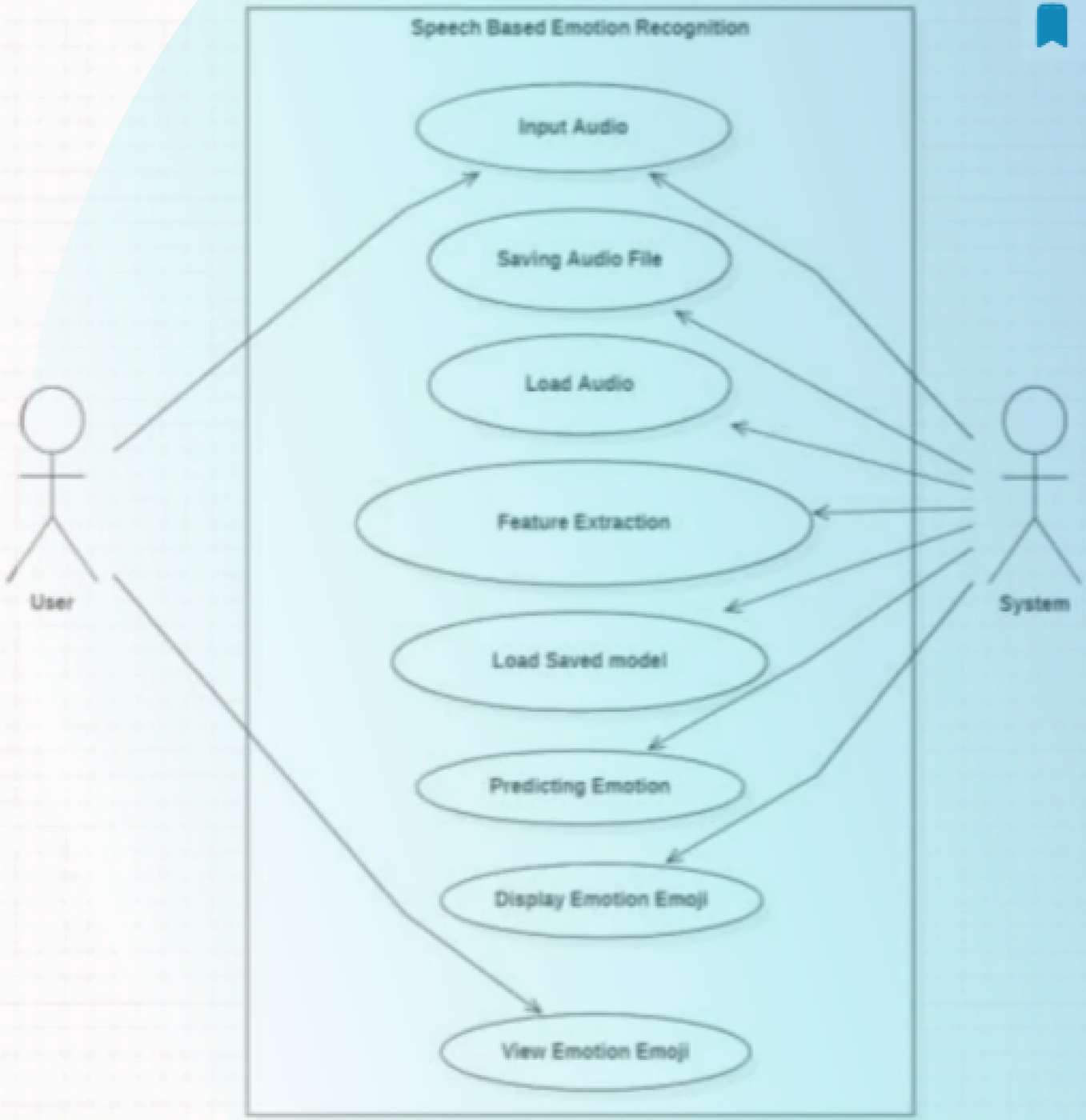
- **RAVDESS** dataset has recordings of 24 actors, 12 male actors and 12 female actors, the actors are numbered from 01 to 24 in North American accent.
- All emotional expressions are uttered at two levels of intensity: normal and strong, except for the 'neutral' emotion, it is produced only in normal intensity. Thus, the portion of the RAVDESS, that we use contains 60 trials for each of the 24 actors, thus making it 1440 files in total.

LIBRARIES USED

Librosa Library: Librosa is a Python library for analysing audio and music. It provides many tools for processing and analysing audio signals, including the capacity to load audio files, glean information from audio signals, and see spectrograms and waveforms[3]. Speech processing, sound event recognition, and music information retrieval are some typical use cases for Librosa.

Pandas Library: Pandas can organise disorganised data sets, making them readable and useful

USE CASE DIAGRAM



TECH STACK USED

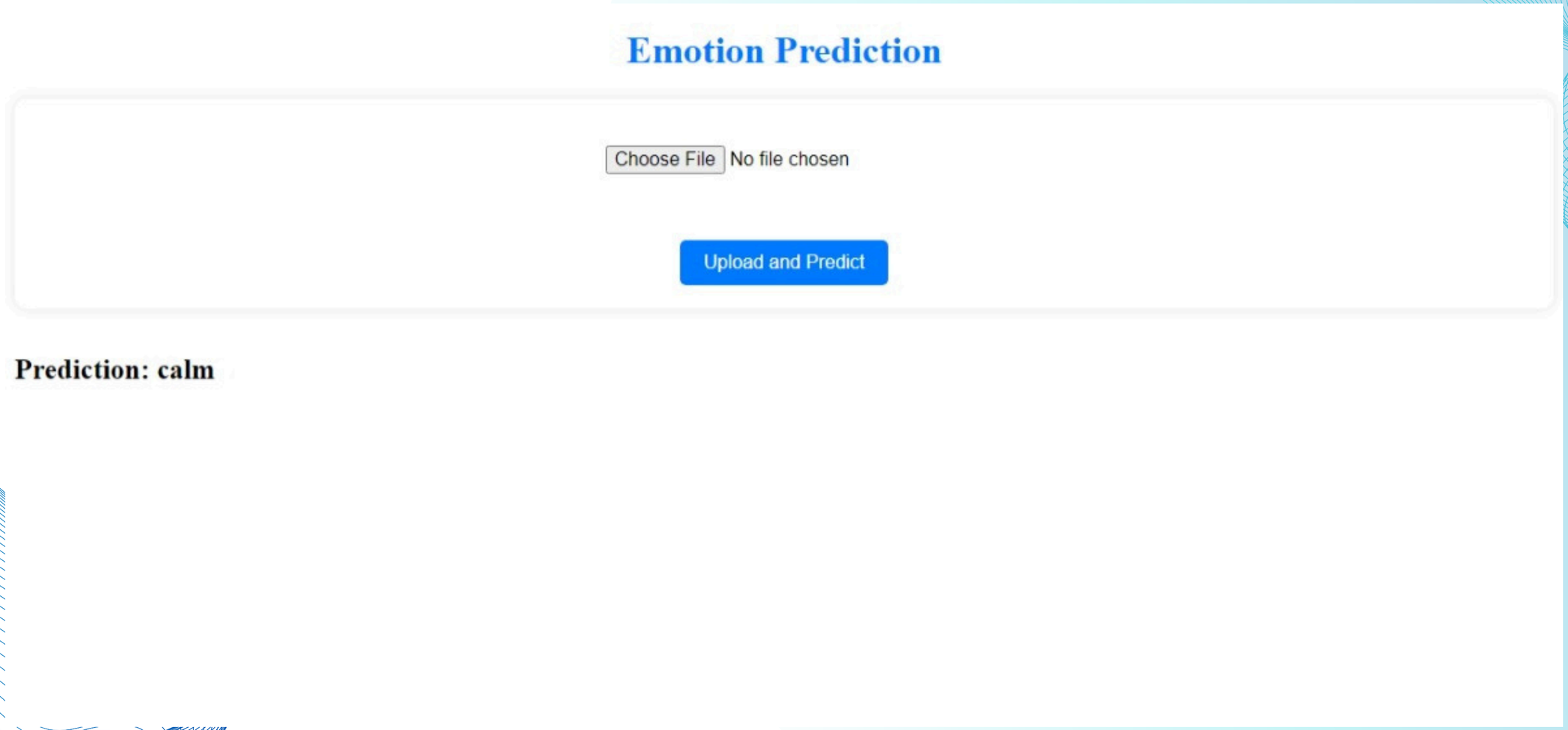
FRONT-END :
HTML
CSS
FLASK:

Flask
web development,
one drop at a time

BACK-END :
PYTHON
GOOGLE COLLAB

DESIGN AND IMPLEMENTATION

UI INTERFACE



TEST CASES AND RESULTS

LIVE DEMO

Prediction Results :

Now we can record the voice of any user via microphone and can then save it as .wav file . So that later we can load our model to predict the result out of that particular .wav file

FUTURE SCOPE

- HUMAN-COMPUTER INTERACTION
- MENTAL HEALTH MONITORING
- MARKET RESEARCH AND CUSTOMER FEEDBACK ANALYSIS
- EDUCATION AND LEARNING
- HEALTHCARE

LIMITATIONS

- Limited expressiveness
- Variable length audio files are not understandable
- Since being pre-trained , cannot predict the unknown Voice and emotion
- Data quality and Noise
- Overfitting and Generalisation

CONCLUSION

In conclusion, Speech Emotion Recognition holds immense potential in various fields including psychology, human-computer interaction, and healthcare.

In our project , we are working towards making the system nearest to accqurate. We are training the model to detect various emotions like sad, happy, anger, and frustration .

We are creating a webpage to take in the input voice and produce the the output emotion by interfacing the backend Machine Learning model with datasets to the front end HTML and JavaScript and enhancing with CSS styling.

TEAM

NEETHUSHREE G N
01JCE21CS131

NANDAN K R
01JST21CS089

BHOOMIKA K S
01JCE21CS019

THANK YOU