Speech Emotion Recognition

Utilizing Speech Emotion Recognition to Enhance Human-Computer Interaction

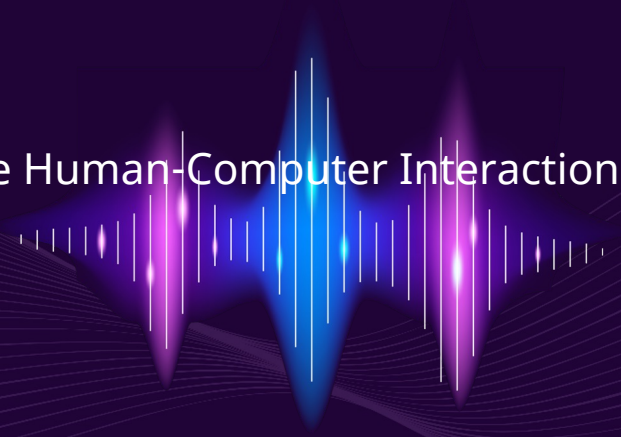
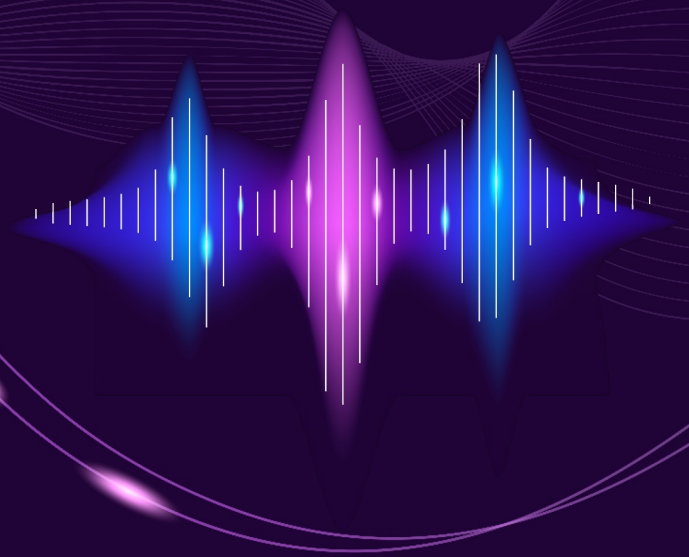
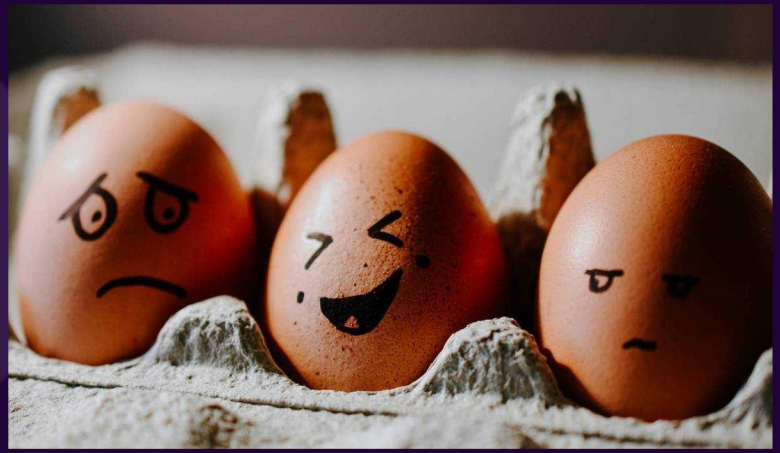


Table of contents



01 Introduction

- Human speech carries complex information beyond the spoken words.
- Utilizing those non-linguistic voice characteristics such as pitch and tone to identify underlying emotions.



Problem Statement

- Near-Real-time prediction of emotions from audio input.
- Aiming for accurate emotion labels based on audio signal analysis.
- The increasing need is driven by the rising demand for advanced communication technologies.
- We are classifying 8 different emotions: Surprise, Angry, Calm, Disgust, Sad, Fear, Neutral, Happy.

Use Cases

Human-Computer Interaction

Improve user engagement and satisfaction by responding empathetically to emotional cues.

Customer Service

Analyze customer calls to identify emotional states and provide appropriate support.

Mental Health Monitoring

Detect signs of distress or emotional dysregulation in speech patterns.

Educational Technology

Assess student engagement and understanding by detecting emotions during lectures or discussions.

Working Principle

1. Gathering Data

RAVDESS, TESS, CREMA-D

2. Pre-Processing

Data augmentation involves creating synthetic data samples by introducing small perturbations to the original training set.

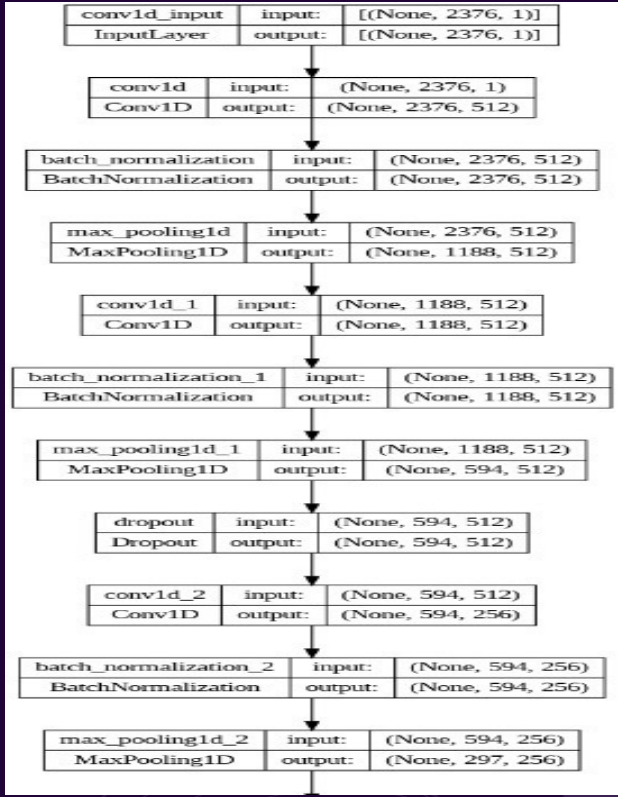
3. Feature Extraction

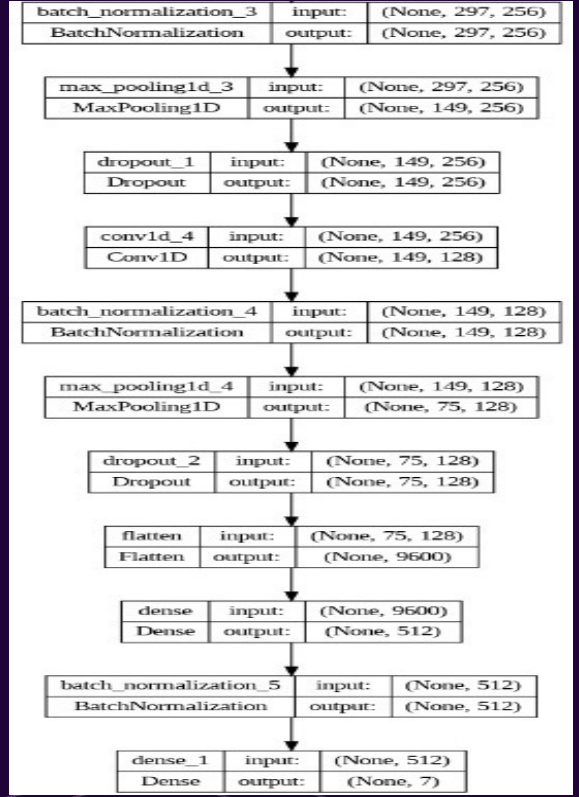
Features such as Zero-Crossing Rate (ZCR), Root Mean Square Error (RMSE), and Mel-frequency Cepstral Coefficients (MFCC) are extracted from the recorded audio using Librosa

4. Data processing and Encoding

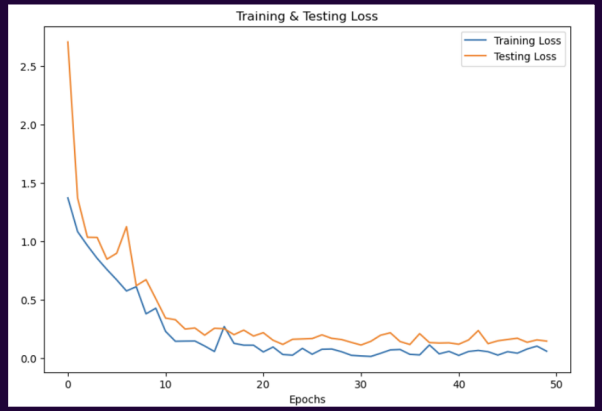
Feature vectors are processed, handling missing values, standardized, and emotion labels are one-hot encoded for model training.

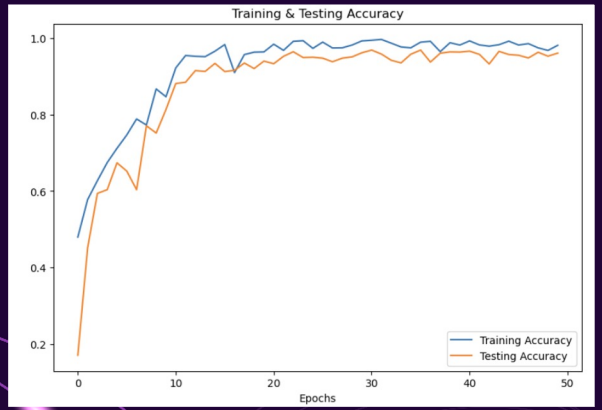
Build CNN Model





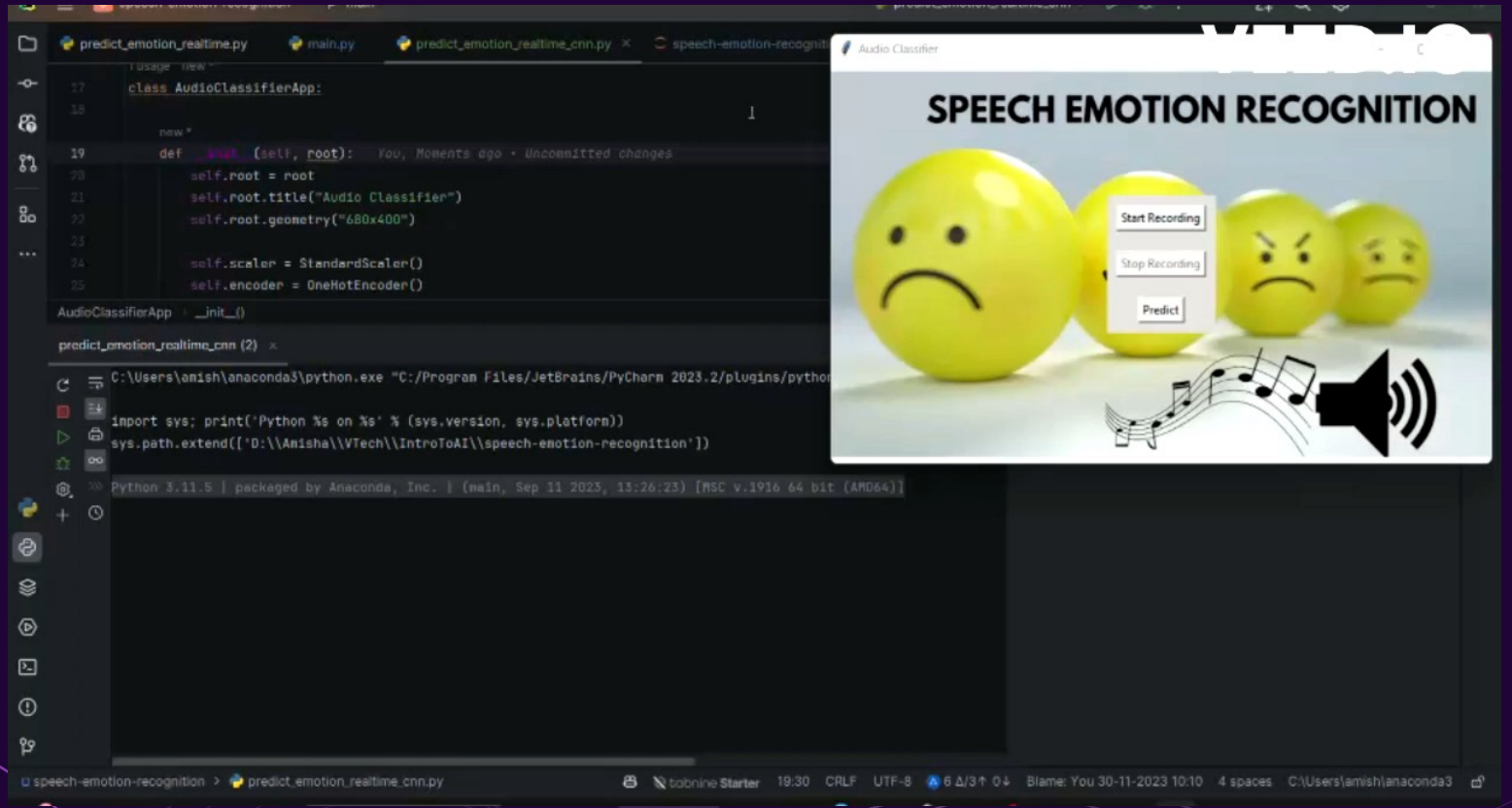
Results


Training & Testing Loss
Training Loss
Testing Loss
2.5
2.0
1.5
1.0
0.5
0.0
0
10
20
30
40
50
Epochs


Training & Testing Accuracy
1.0
0.8
0.6
0.4
0.2
Training Accuracy
Testing Accuracy
0
10
20
30
40
50
Epochs


Audio Classifier
SPEECH EMOTION RECOGNITION
Start Recording
Stop Recording
Predict

Demo


predict_emotion_realtime.py
main.py
predict_emotion_realtime_cnn.py
class AudioClassifierApp:
def __init__(self, root):
self.root = root
self.root.title("Audio Classifier")
self.root.geometry("680x400")
self.scaler = StandardScaler()
self.encoder = OneHotEncoder()
AudioClassifierApp > __init__()
predict_emotion_realtime_cnn (2)
C:\Users\amish\anaconda3\python.exe "C:/Program Files/JetBrains/PyCharm 2023.2/plugins/python
import sys; print('Python %s on %s' % (sys.version, sys.platform))
sys.path.extend(['D:\\Amisha\\VTech\\IntroToAI\\speech-emotion-recognition'])
Python 3.11.5 | packaged by Anaconda, Inc. | (main, Sep 11 2023, 13:26:23) [MSC v.1916 64 bit (AMD64)]
Audio Classifier
SPEECH EMOTION RECOGNITION
Start Recording
Stop Recording
Predict
speech-emotion-recognition > predict_emotion_realtime_cnn.py
tabnine Starter
19:30
CRLF
UTF-8
Blame: You 30-11-2023 10:10
4 spaces
C:\Users\amish\anaconda3

Lessons Learned

- Selection audio features are crucial for emotion detection through speech.
- Training the model with good quality data can help in improving accuracy drastically.
- Tweaking the Epochs to optimum level plays an important role in improving accuracy.

References

[1]https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio
[2]https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess
[3]https://jonathanbgn.com/speech/2020/10/31/emotion-recognition-transfer-learning-wav2vec.html
[4]https://github.com/CheyneyComputerScience/CREMA-D
[5]https://www.sciencedirect.com/science/article/abs/pii/S1746809420300501


Thank You
Q&A