AuraMood: Emotion Classification in Audios

Auramood is a project focused on emotion classification in audio using deep learning techniques.

NIDHI BAJAJ
AGRIMA JAIN
Group 16

Project Overview

1 Objective

Develop a system to classify emotions in speech data using deep learning.

2 Emotions

Classify audio into categories: happy, sad, anger, fear, surprise, and neutral.

3 Applications

Enhance human-computer interaction, customer service, and mental health monitoring.

Methodology

1

Data Collection

Use datasets like RAVDESS or EMO-DB containing labelled emotional speech samples.

2

Preprocessing

Remove noise, normalize signals, and extract features like MFCCs.

3

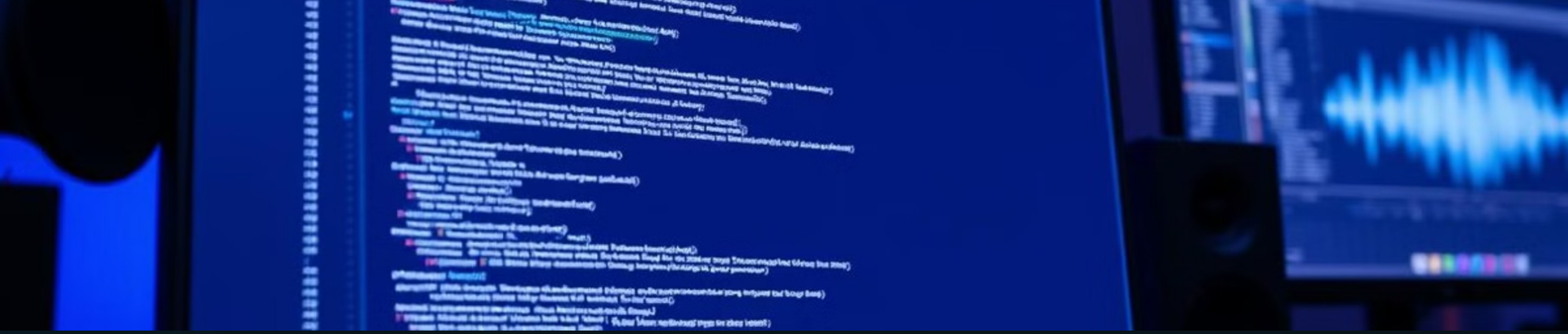
Model Development

Implement CNN architecture for pattern recognition in audio features.

4

Evaluation

Assess model performance using metrics like accuracy and F1-score.



Technical Implementation

Python

Primary programming language for development.

TensorFlow

Deep learning framework for model building and training.

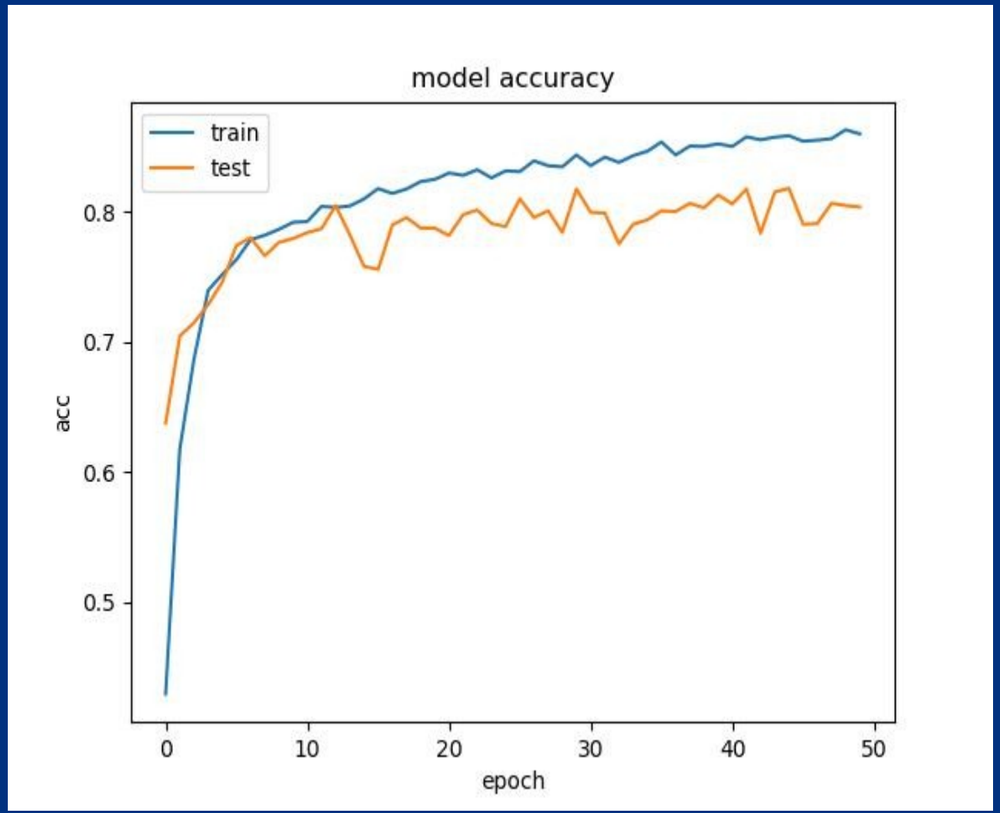
Librosa

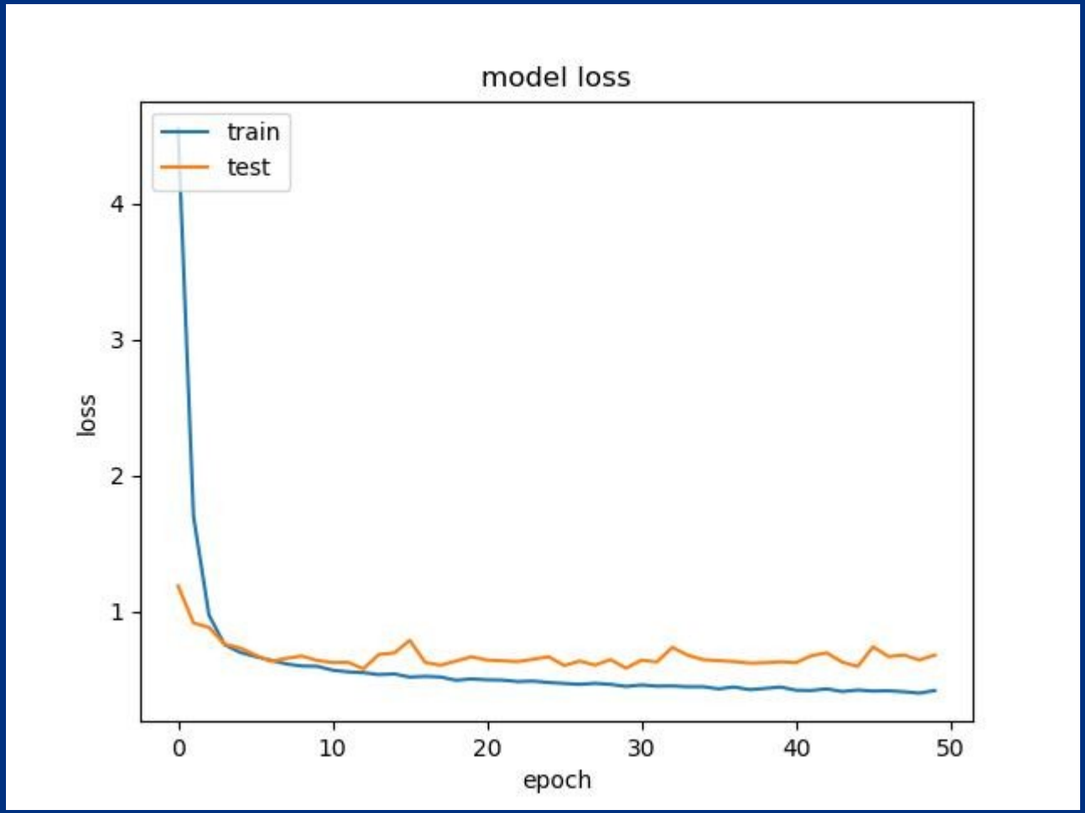
Library for audio processing and feature extraction.

Matplotlib

Visualization tool for plotting graphs and matrices.

Model Accuracy And Loss





Evaluation Metrics & Results

	precision	recall	f1-score	support
0	0.90	0.84	0.87	192
1	0.62	0.80	0.70	123
2	0.88	0.72	0.79	264
3	0.90	0.69	0.78	275
4	0.75	0.94	0.83	252
5	0.80	0.83	0.82	241
6	0.73	0.85	0.79	197
7	0.86	0.78	0.82	190
accuracy			0.80	1734
macro avg	0.81	0.81	0.80	1734
weighted avg	0.82	0.80	0.80	1734

```
[[161  14   0   2   5   2   6   2]
 [  2  99   4   1   1   4  11   1]
 [  4  23 189   1  24   8   7   8]
 [  5  18   7 190  16  23  13   3]
 [  1   1   2   2 237   1   6   2]
 [  4   2   3  12  11 201   2   6]
 [  0   2   4   1  14   6 167   3]
 [  2   1   6   1  10   5  16 149]]
```

Model Summary

Layer (type)	Output Shape	Param #
conv1d_3 (Conv1D)	(None, 40, 128)	768
activation_4 (Activation)	(None, 40, 128)	0
dropout_3 (Dropout)	(None, 40, 128)	0
max_pooling1d_2 (MaxPooling1	(None, 5, 128)	0
conv1d_4 (Conv1D)	(None, 5, 128)	82048
activation_5 (Activation)	(None, 5, 128)	0
dropout_4 (Dropout)	(None, 5, 128)	0
flatten_2 (Flatten)	(None, 640)	0
dense_2 (Dense)	(None, 8)	5128
activation_6 (Activation)	(None, 8)	0

Total params: 87,944
Trainable params: 87,944
Non-trainable params: 0

Challenges Faced

Data Imbalance

Some emotional categories were overrepresented, requiring mitigation techniques.

Feature Selection

Selecting relevant features from complex audio signals required iterative experimentation.

Overfitting

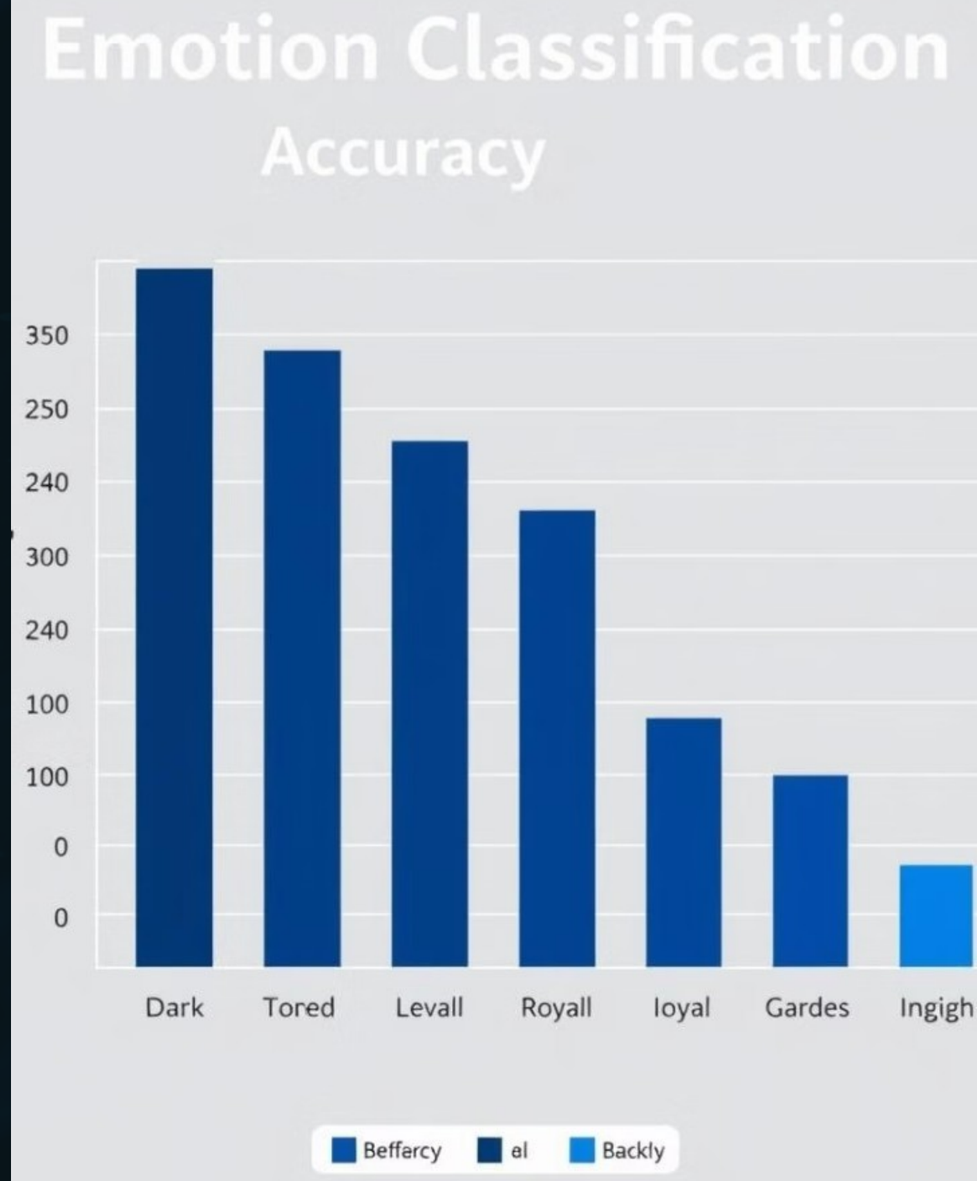
Initial model overfitting was addressed using dropout layers and early stopping.

Real-Time Processing

Optimizing for real-time emotion detection remains a future challenge.

Results and Performance

Accuracy	Over 85% on test dataset
Best Performance	Happy and sad emotions
Confusion	Between anger and fear



Conclusion

Effectiveness

CNN models proved effective in classifying emotions from speech data.

Applications

Potential for use in customer service, healthcare, and HCI.

Future Scope

Room for improvements in architecture and dataset expansion.

Future Work

1

Real-Time Detection

Enhance system to process live audio feeds for immediate emotion classification.

2

Multimodal Recognition

Combine audio with visual data to improve overall detection accuracy.

3

Model Optimization

Experiment with advanced models like LSTM or transformers for better performance.

4

Dataset Expansion

Incorporate diverse languages and accents to improve model robustness.

Thank You