
Page 01
EE216-MACHINE LEARNING FOR SIGNAL PROCESSING
Group 12
Pathak Krish 230002049
Bhuva Vasudha - 230002017

PROBLEM DESCRIPTION

The problem addressed is the automatic detection and classification of human facial emotions from static images. The objective is to accurately recognize and categorize facial expressions into one of seven distinct emotional categories: angry, disgust, fear, happy, sad, surprise, and neutral. This task plays a crucial role in advancing natural human-computer interaction systems, where understanding non-verbal cues like facial expressions is essential for creating more intuitive, responsive, and empathetic technologies that can better interpret and react to human emotions.

krish
230002049

NOVELTY OF OUR WORK

1.CNN feature extraction: Instead of handcrafted feature extraction(LBP,HOG,SIFT), deep features are automatically learned from raw images.

2. Lightweight CNN architecture :

- Faster training
- Lower memory usage

3. Use of efficient preprocessing: Incorporation of techniques such as Haar-cascade classifiers and K-means clustering during preprocessing to better segment facial features before classification.

4.Real-world data retraining strategy enhances the robustness of the system for practical applications.

krish
230002049

DESCRIPTION OF THE ARCHITECTURE/MODEL

The deep learning model we implemented is based on a Sequential Convolutional Neural Network (CNN).

- Sequential model with Conv2D layers (first with 32 filters, second with 64 filters,and third with 128 filters).
- Kernel size of (3x3) and activation function 'ReLU'.
- Flatten layer to convert feature maps to 1D vectors.
- Dense layers at the output.
- Output layer with Softmax activation to predict 7 emotion classes.

Vasudha
230002017

DESCRIPTION OF THE ARCHITECTURE/MODEL

Detailed Architecture:

- Input: Grayscale face image (48x48 pixels, single channel).
- Layer 1:
  - Conv2D layer with 32 filters, kernel size (3x3), activation ReLU.(for edges, textures)
- Layer 2:
  Conv2D layer with 64 filters, kernel size (3x3), activation ReLU.(for shapes, corners, curves)
- Layer 3:
  - Conv2D layer with 128 filters, kernel size (3x3), activation ReLU.
  (like eyes, mouth shapes, facial parts)
- Layer 4:
  - fourth Conv2D(128) sees high-level features: entire face expressions!
- Layer 5:
  - Flatten Layer.
- Layer 6:
  - Dense Layer.
  - Purpose: High-level representation learning.
- Layer 7:
  - Output Layer with 7 neurons (one for each emotion class) and Softmax activation to predict probability distribution.

Vasudha
230002017

ITS UNIQUE FEATURES

Unique Features:

- Shallow but efficient network, avoiding deep architectures.
- ReLU activations after each Conv2D layer for faster training.
- Adam Optimizer dynamically adjusts learning rates.
- Softmax classifier ensures multi-class output.
- Potential use of Dropout layers suggested to reduce overfitting (can be added easily).

Vasudha
230002017

BASIC WORKING OF CODE

- Convert the image to grayscale, resize it to 48×48, normalize pixel values, and reshape it to match the model input.
- Pass the processed image into the model to predict the emotion.
- The model outputs class probabilities; the class with the highest probability is selected.
- Map the predicted class to its corresponding emotion label (e.g., happiness, sadness).
- Plot the input image using matplotlib and display the predicted emotion as the title

Vasudha
230002017

IMPLEMENTATION DETAILS

- Dataset Used: Kaggle facial emotion recognition dataset, processed into 48x48 grayscale images (dataset used: FER2013).
- Training/Validation/Test Splits:
  - ~34,488 images used for training.
  - ~1,250 images used for testing.
- Network Configuration:
  - Optimizer: Adam
  - Loss function: Categorical Crossentropy
  - Epochs: 60
  - Batch size: (standard Keras default, 32 or 64)

krish
230002049

CHALLENGES FACED:

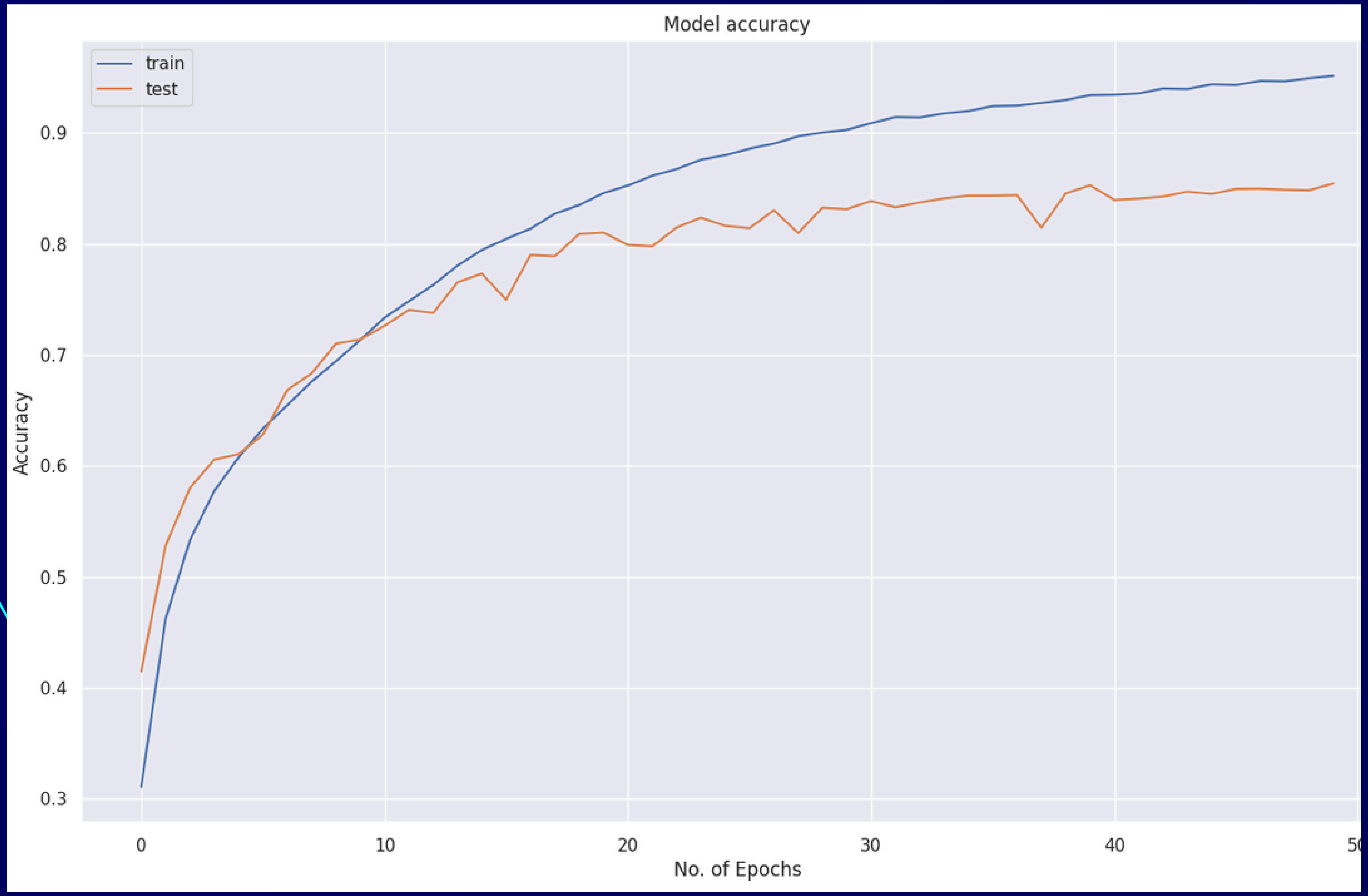
- Variability in Expressions: Different individuals express emotions differently, making generalization difficult.
- Environmental Factors: Variations in lighting and head poses impact facial recognition accuracy.
- Subtlety of Emotions: Some emotions are very subtle and hard to distinguish visually.
- Age and Gender Variations: Differences based on demographic factors affect how emotions are portrayed.

krish
230002049

OVERVIEW OF OBTAINED RESULTS

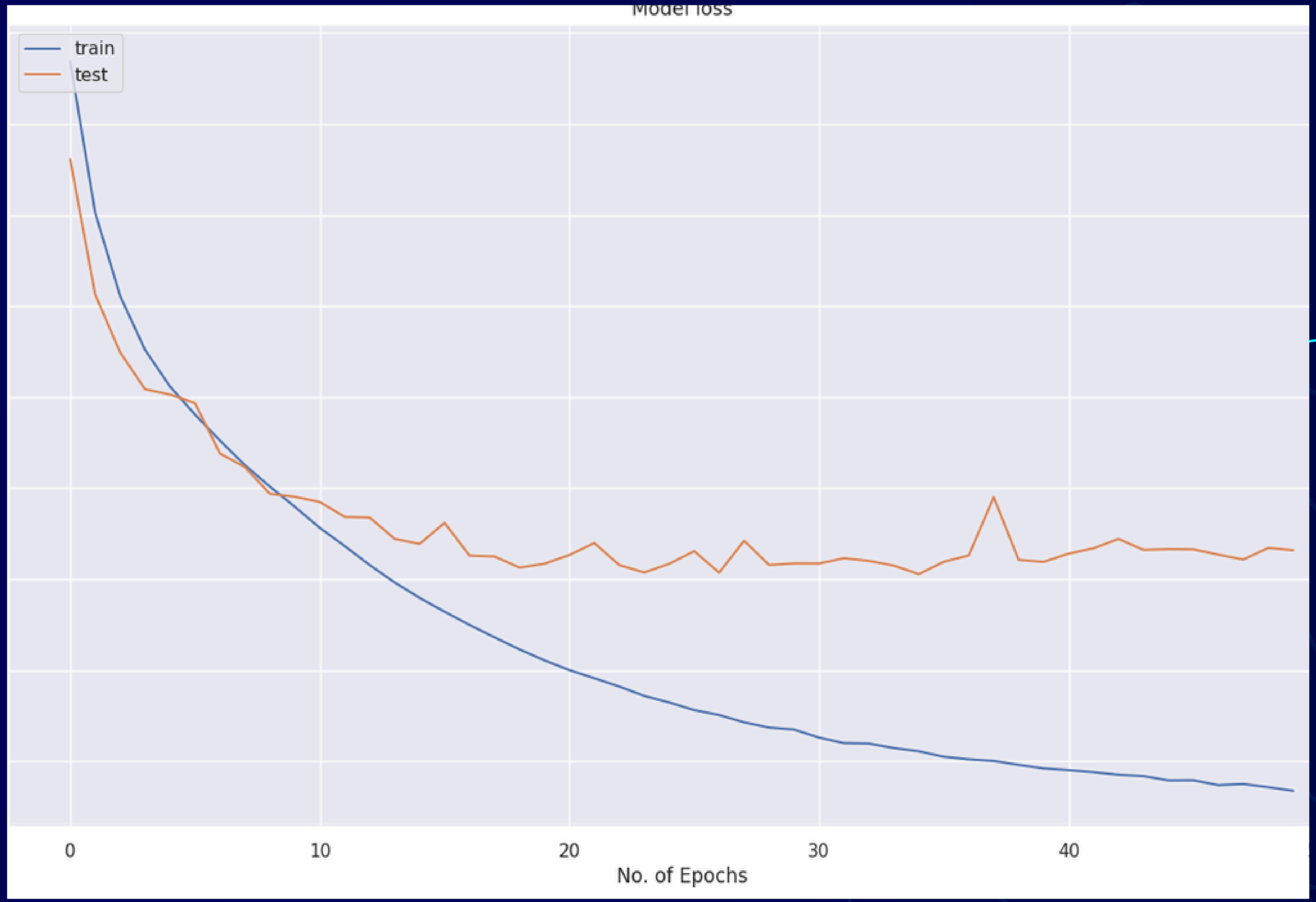
Results:

Achieved ~95% accuracy (training dataset),~85%(testing dataset).



Model accuracy vs no. of epochs

(Blue line represent training data and Red line represents testing data accuracy)



Model loss vs no. of epochs

(Blue line represent training data and Red line represents testing data accuracy)

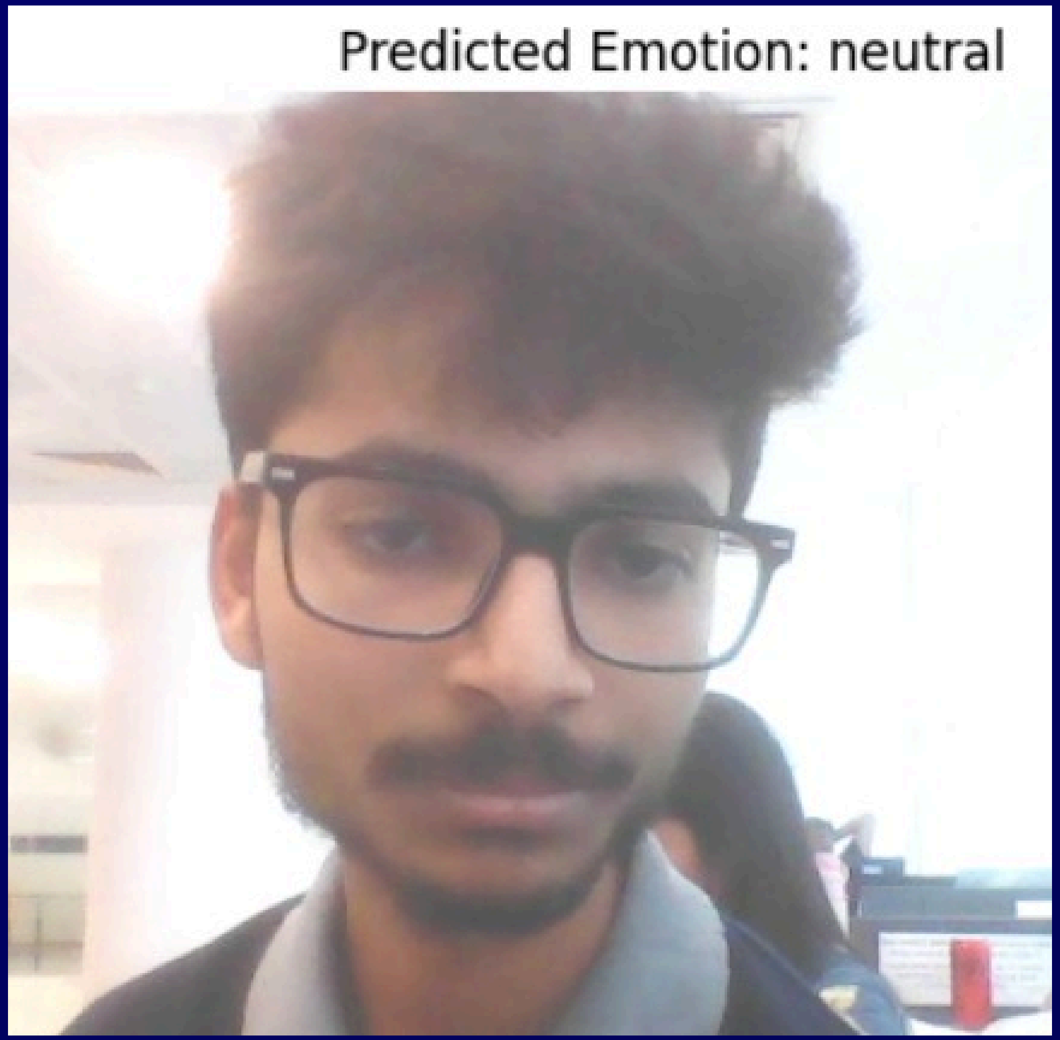
krish
230002049

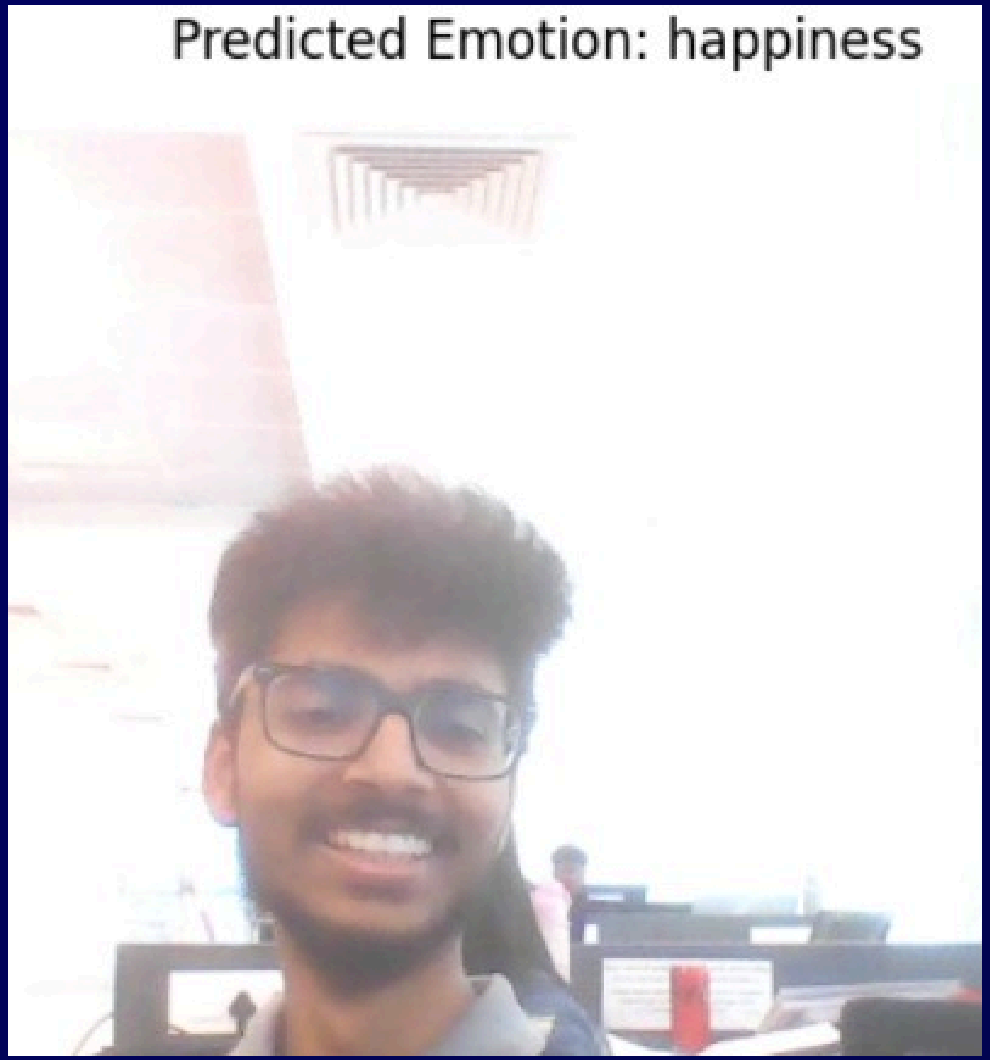
COUNTERINTUITIVE INFERENCE AND SCOPE FOR IMPROVEMENT

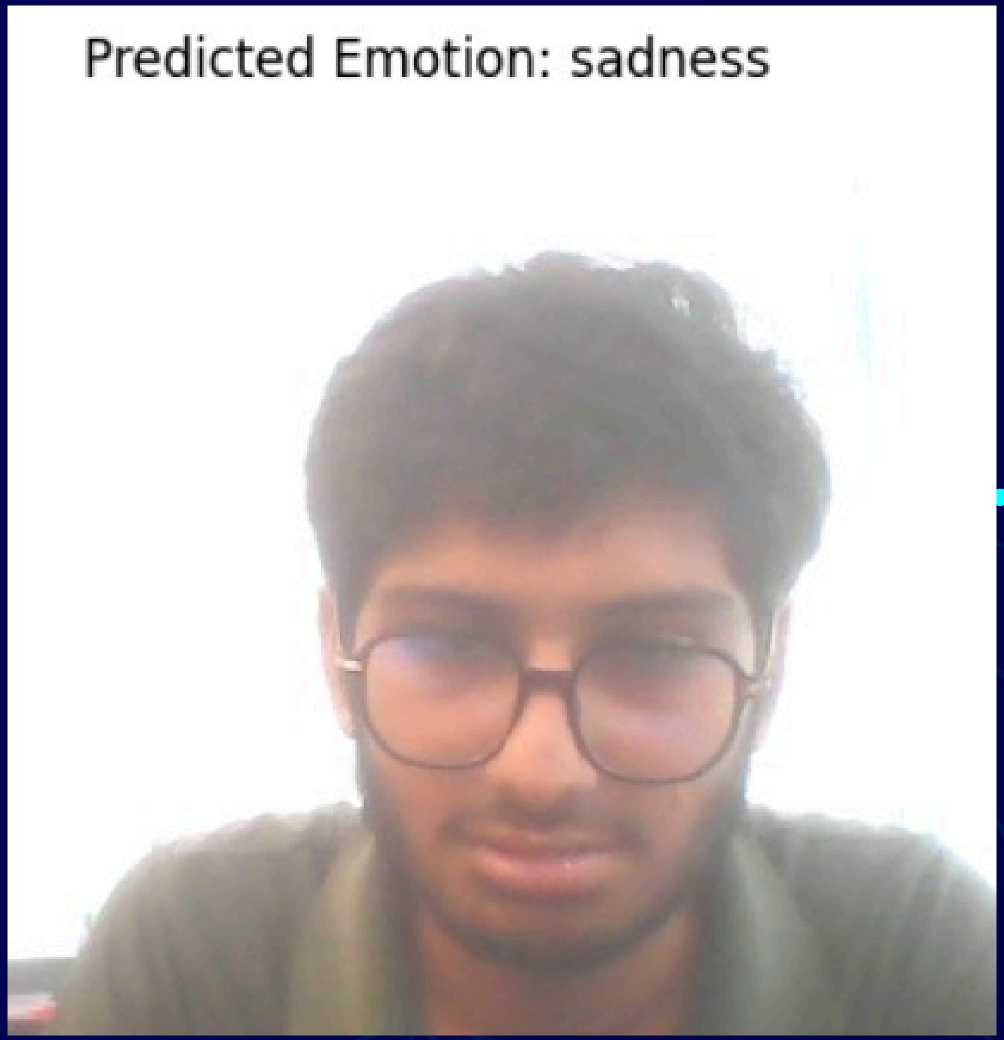
- Counterintuitive Inference:
  - Even with a shallow CNN architecture, very high accuracy was achieved on a relatively complex task, showing that lightweight models can sometimes outperform expectations if the preprocessing and feature extraction are strong.
- Scope for Improvement:
  - Use deeper CNNs or ensemble models (e.g., ResNet-50 fine-tuning) for potentially even better accuracy.
  - Introduce more real-world variations in training (like lighting, occlusions) to enhance robustness.
  - Explore RNNs or LSTMs for capturing temporal sequences if moving to video-based emotion recognition.

Vasudha
230002017

OVERVIEW OF OBTAINED RESULTS







THANK YOU