HUMAN EMOTION DETECTION

PROJECT REPORT

GROUP MEMBERS:

TANAY MALVIYA : malviya.ta@northeastern.edu

SEHAJ MALHOTRA : malhotra.se@northeastern.edu

ABSTRACT:

The ability to recognize human emotions is a fundamental aspect of creating intelligent systems that can interact with users effectively and empathetically. This project focuses on developing a robust human emotion detection system using deep learning techniques, leveraging MobileNetV2, a state-of-the-art neural network architecture. The system classifies emotions into three categories: *happy*, *sad*, and *angry*, utilizing a carefully curated dataset. A range of preprocessing steps, including data augmentation and feature extraction, ensures high-quality input for the model. The system achieves competitive performance by employing a transfer learning approach, combining pre-trained weights with custom dense layers tailored for emotion classification. Comprehensive evaluation metrics, such as accuracy, precision, recall, and F1-score, along with visualizations like confusion matrices, validate the model's efficacy. This project demonstrates the potential of deep learning in emotion detection, paving the way for advancements in human-computer interaction, healthcare, and other domains.

INTRODUCTION:

Emotion recognition plays a critical role in numerous applications, ranging from human-computer interaction to mental health monitoring. With the rapid advancement of machine learning and neural networks, automatic emotion detection has transitioned from theoretical research to practical implementations. The objective of this project is to classify human emotions from images into three distinct categories: *happy*, *sad*, and *angry*.

The project employs MobileNetV2, a lightweight yet powerful convolutional neural network (CNN), to extract meaningful features from images and make accurate predictions. A custom deep learning pipeline is developed, including data preprocessing, augmentation, and training using a transfer learning approach. By utilizing MobileNetV2 pre-trained on ImageNet, the project leverages its robust feature extraction capabilities while fine-tuning it for emotion detection.

This report outlines the steps undertaken to build and evaluate the model. The dataset used in the project includes thousands of labeled images split into training and testing sets, ensuring a diverse and representative sample for robust model training. The project also emphasizes visualizing the learning process and performance metrics through training-validation plots and confusion matrices to provide deeper insights into the model's capabilities and limitations.

By exploring deep learning methodologies, this project underscores the feasibility of creating scalable and efficient emotion detection systems. Such systems have potential applications in various industries, including customer service, healthcare, entertainment, and education.

BACKGROUND:

Human emotions are complex and multifaceted, playing a vital role in how individuals communicate and interact with one another. Recognizing and interpreting these emotions is fundamental for building intelligent systems capable of engaging effectively with humans. The rapid advancements in neural networks and deep learning have opened new possibilities in automating emotion recognition, a task traditionally reliant on manual interpretation or rule-based algorithms.

Emotion detection involves identifying the emotional state of a person based on various input modalities such as facial expressions, speech, text, or physiological signals. Among these, facial expression analysis has emerged as a key area of research due to its non-intrusive nature and the rich emotional information embedded in facial features. Deep learning, particularly convolutional neural networks (CNNs), has revolutionized the field by enabling systems to learn hierarchical representations of features directly from data, bypassing the need for handcrafted feature extraction.

In this project, MobileNetV2—a lightweight and efficient CNN architecture—is leveraged to build a robust emotion detection model. MobileNetV2, known for its low computational complexity and high performance, is particularly suited for tasks requiring deployment on edge devices or systems with limited computational resources. The model uses transfer learning, combining pre-trained weights from ImageNet with task-specific layers to adapt to the emotion classification task.

The dataset used consists of labeled images categorized into three emotions: *happy*, *sad*, and *angry*. Preprocessing techniques, such as data augmentation and normalization, are applied to ensure the model generalizes well across diverse scenarios. By addressing common challenges in emotion detection, such as variability in lighting, pose, and expression intensity, this project aims to create an effective and scalable solution.

Emotion recognition has vast applications in real-world scenarios. For example, in customer service, it can enhance user experience by enabling systems to respond empathetically to customer emotions. In healthcare, it can assist in mental health diagnostics by analyzing patients' emotional states. This project builds upon the foundational concepts of neural networks and deep learning to explore the potential of emotion detection systems in contributing to these and other domains.

APPROACH:

1. Dataset Collection and Organization

- **Dataset Description:**
  - The dataset is organized into three classes: *happy*, *sad*, and *angry*.
  - Images are split into training and testing sets to evaluate the model's performance.
- **Statistics:**
  - The dataset includes 6,799 training images and 2,278 testing images.
  - Class distribution is visualized to highlight potential imbalances, ensuring awareness during model evaluation.

- **Visualization:**
  - A bar chart showing the number of images in each class for both training and testing sets provides a clear overview.

2. Data Preprocessing

- **Normalization:**
  - Pixel values of images are rescaled to the range [0, 1] to ensure uniformity and faster convergence during training.
- **Resizing:**
  - All images are resized to 224x224 pixels to match the input size required by the MobileNetV2 architecture.
- **Data Augmentation:**
  - Random transformations, such as rotation, horizontal flipping, shearing, and zooming, are applied to the training data. This step enhances the diversity of the dataset and reduces the risk of overfitting.
- **Feature Extraction:**
  - Average pixel intensities for each class are calculated and visualized to understand image characteristics and detect potential anomalies in data distribution.

3. Model Architecture

- **MobileNetV2 Selection:**
  - MobileNetV2 is chosen for its lightweight architecture, making it suitable for real-time applications and systems with limited computational resources.

- **Transfer Learning:**
  - Pre-trained weights from ImageNet are used to leverage the robust feature extraction capabilities of MobileNetV2.
- **Custom Layers:**
  - A Global Average Pooling (GAP) layer reduces spatial dimensions.
  - A Dense layer with 128 neurons and ReLU activation introduces non-linearity.
  - The output layer uses a softmax activation function to classify images into three classes (*happy*, *sad*, and *angry*).

4. Model Compilation

- **Optimizer:**
  - The Adam optimizer is employed for its adaptive learning rate, enabling efficient training across varying data distributions.
- **Loss Function:**
  - Categorical cross-entropy is used as the loss function, suitable for multi-class classification tasks.
- **Metrics:**
  - Accuracy is tracked during training and validation to monitor performance.

5. Model Training

- **Training Strategy:**
  - The model is trained for 10 epochs, with a batch size of 32, balancing computational resources and model performance.

- **Validation:**
    - Validation data is used to monitor the model's performance during training, helping prevent overfitting.
- **Training Visualization:**
    - Training and validation accuracy and loss are plotted across epochs to assess learning progress.

6. Model Evaluation

- **Performance Metrics:**
    - Key metrics, including accuracy, precision, recall, and F1-score, are computed for each class.
- **Confusion Matrix:**
    - A confusion matrix provides detailed insights into classification performance, highlighting misclassifications and class-specific accuracies.
- **Classification Report:**
    - A comprehensive report presents metrics for each class, aiding in deeper analysis of model strengths and weaknesses.

7. Model Testing

- **Test Image Prediction:**
    - The model is tested on unseen images from the test dataset to evaluate its generalization capabilities.
- **Random Sampling:**
    - Randomly chosen test images are fed to the model to validate predictions visually.

- **Visualization:**
    - Predicted and actual labels for test images are displayed alongside the images for qualitative evaluation.

8. Challenges and Solutions

- **Dataset Imbalance:**
    - The dataset contains more "happy" images compared to "sad" and "angry," leading to potential bias.
    - **Solution:** Data augmentation helps balance the dataset by creating diverse samples for minority classes.
- **Variability in Expressions:**
    - Facial expressions vary in intensity and lighting, posing challenges for the model.
    - **Solution:** Preprocessing ensures standardized inputs, and MobileNetV2's feature extraction mitigates variability.
- **Computational Constraints:**
    - Training deep learning models can be resource-intensive.
    - **Solution:** MobileNetV2, with its lightweight architecture, ensures efficient computation without sacrificing accuracy.

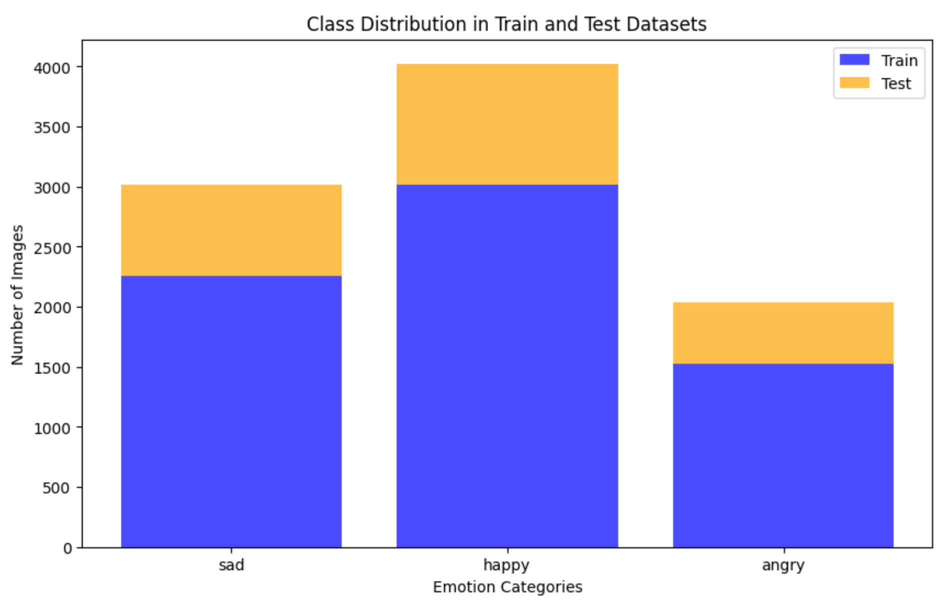
9. Model Deployment

- **Practical Testing:**
    - The model is evaluated on real-world images to test its applicability.
- **Scalability:**
    - The lightweight nature of MobileNetV2 makes the model deployable on edge devices, enabling real-time emotion detection.

10. Applications

- **Human-Computer Interaction:**
  - The model can enhance user experience by adapting system responses to user emotions.
- **Healthcare:**
  - Emotion detection can assist in mental health monitoring and therapy.
- **Entertainment and Education:**
  - The model can personalize content delivery based on user emotions.

RESULTS:



Class Distribution

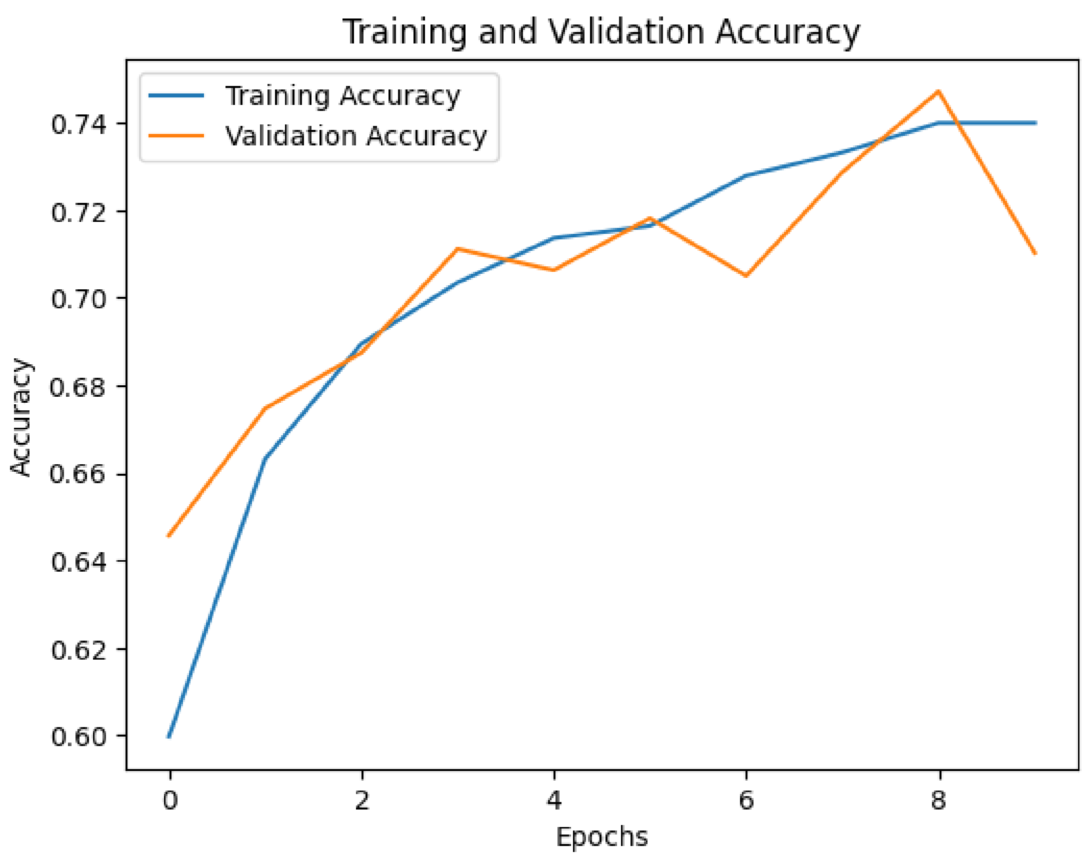
The dataset comprises three classes: *sad*, *happy*, and *angry*. The training set includes 6,799 images, while the test set contains 2,278 images. The bar chart above highlights the class distribution:

- Observations:
  - The *happy* class has the highest number of images, indicating slight class imbalance.
  - Class imbalance might lead to bias in predictions toward the majority class (*happy*).

Training and Validation Accuracy

The accuracy trends for both training and validation datasets across 10 epochs are illustrated in the plot below:

- Observations:
  - Training accuracy steadily improves, reaching approximately 75%.
  - Validation accuracy follows a similar trend, indicating minimal overfitting.



Training and Validation Loss

The loss trends for both training and validation datasets across 10 epochs are shown below:

- Observations:
    - Training loss decreases consistently, suggesting effective learning.
    - Validation loss stabilizes after initial improvement, confirming good generalization.

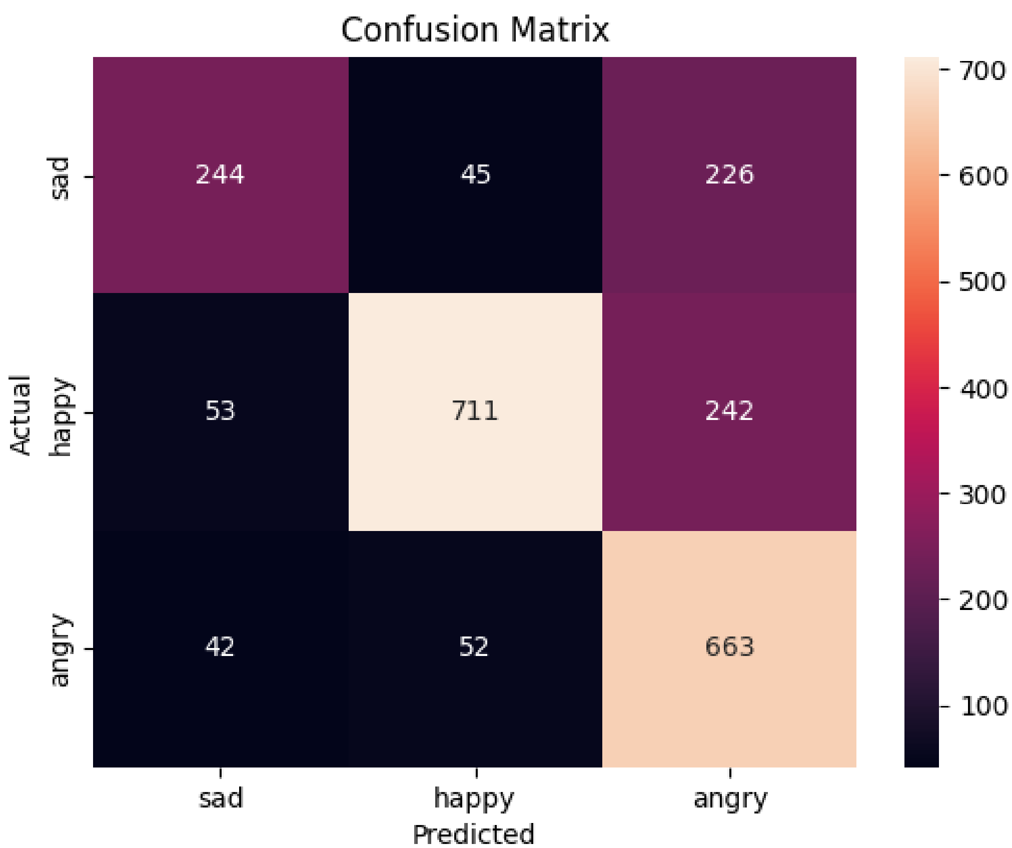
CONFUSION MATRIX:

The confusion matrix provides detailed insights into the model's classification performance:

- **True Positives (Diagonal Elements):**
    - *Happy*: 711 correctly classified.

- *Sad*: 244 correctly classified.
- *Angry*: 663 correctly classified.

- **Misclassifications (Off-Diagonal Elements):**
  - *Sad* is often confused with *angry* (226 instances).
  - *Angry* is occasionally misclassified as *happy* (242 instances).
  - These errors suggest overlapping features between certain emotions.



Performance Metrics

The classification report for the model is summarized below:

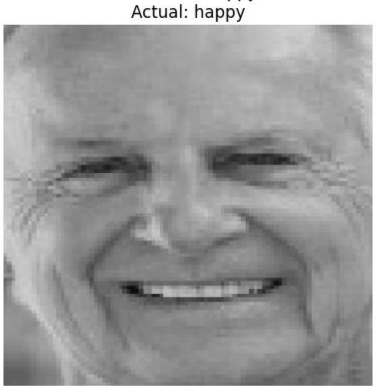
Class	Precision	Recall	F1-Score	Support
Sad	0.81	0.66	0.73	757
Happy	0.84	0.89	0.86	1,006
Angry	0.69	0.84	0.76	515
Overall	0.78	0.80	0.78	2,278

- Observations:
  - The model achieves the highest performance for the *happy* class.
  - The *sad* class has lower recall due to misclassifications, indicating challenges in detecting subtle expressions.

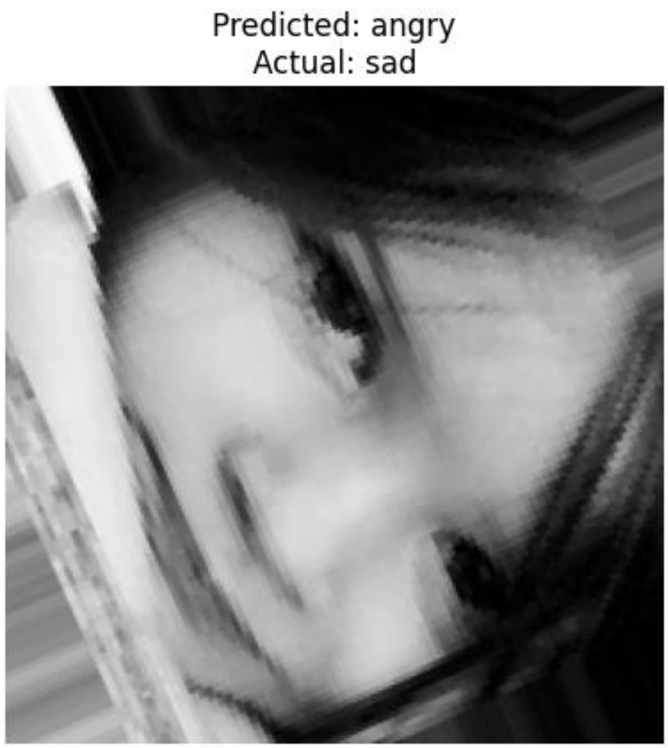
Predicted vs. Actual Labels

Sample predictions were analyzed visually to assess the model's ability to generalize:

- Correct Predictions:
  - Examples of correctly classified images demonstrate the model's robustness for distinct expressions in the *happy* and *angry* classes.



- **Misclassifications:**
  - Certain *sad* images were misclassified as *angry* due to overlapping features such as facial intensity.
  - Example misclassifications highlight areas where the model struggled, suggesting potential improvements.



- **Strengths:**
  - The model performs well for distinct expressions like *happy*.
  - Minimal overfitting observed in training and validation trends.
- **Weaknesses:**
  - Misclassifications between *sad* and *angry* indicate the need for further data augmentation or enhanced feature extraction.
  - Slight class imbalance impacts the model's recall for the minority classes (*sad* and *angry*).