FACIAL EMOTION RECOGNITION USING TRANSFER LEARNING WITH MOBILENETV2

A project report submitted in partial fulfilment of the requirements for the degree of

MASTER OF COMPUTER APPLICATION (MCA)
OF
TEZPUR UNIVERSITY

2025

Submitted by

ASHIS SORAHIA
Roll No. CSM23052

Guided by

Dr ROSY SHARMA

Associate professor

TEZPUR UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
TEZPUR UNIVERSITY
TEZPUR — 784028
ASSAM

DECLARATION

I affirm that the project work entitled, "*Facial Emotion Recognition Using Transfer Learning With MobilenetV2*" submitted to the Department of Computer Science & Engineering at Tezpur University, was authored solely by me and has not been presented to any other institution for the purpose of obtaining any other degree.

Place: Tezpur (Ashis Sorahia)

Date : 01 June 2025 CSM23052

ACKNOWLEDGEMENT

I express my sincere appreciation to my project mentor**, Dr Rosy Sarmah** who is an Associate Professor in the Department of Computer Science and Engineering at Tezpur University, for granting me the opportunity to work under his guidance and offering extensive support throughout the project.

I also extend my gratitude to the Head of the Department of Computer Science and Engineering (CSE) and the entire faculty of the CSE Department, Tezpur University for their invaluable guidance and cooperation during the project. I would like to acknowledge and thank all individuals who, in various capacities, have provided assistance, whether directly or indirectly, with their skills and contributions.

Place: Tezpur (Ashis Sorahia)

Date : 01 June 2025 CSM23052

TEZPUR UNIVERSITY

CERTIFICATE

This is to certify that the project report entitled **Facial Emotion Recognition Using Transfer Learning With MobilenetV2** , submitted to the Department of Computer Science and Engineering, Tezpur University, in partial fulfilment for the award of the degree of **Master of Computer Application** is a record of work carried out by **Mr Ashis sorahia**, Roll No. **CSM23052**, under my supervision and guidance

All help received by him from various sources has been duly acknowledged.

No part of this report has been submitted elsewhere for award of any other degree.

Place : Tezpur
Date :

Dr Rosy Sarmah
Associate Professor
Supervisor

TEZPUR UNIVERSITY

CERTIFICATE

This serves to confirm that the project report, titled, "**Facial Emotion Recognition Using Transfer Learning With MobilenetV2**" has been submitted by Ashis Sorahia , identified by roll number CSM23052. Ashis Sorahia has satisfactorily completed the project requirements, fulfilling the necessary conditions and adhering to the regulations for the conferral of Master of Computer Application degree in Computer Science & Engineering at Tezpur University.

This achievement was attained during the academic session 2023-2025

To the best of my understanding, the content presented in the project report has not been presented for the attainment of any Degree or Diploma at any other university or institute.

Place : Tezpur
Date :01 June 2025

(Sarat Sarahia)
Professor & Head
Dept. of Computer Sc. & Engg.

DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
TEZPUR UNIVERSITY
TEZPUR – 784028, ASSAM, INDIA

CERTIFICATE

This dissertation entitled **Facial Emotion Recognition Using Transfer Learning With MobilenetV2** submitted by **Ashis Sorahia** in partial fulfilment of the requirements for the degree of Master of Computer Application (MCA) of Tezpur University has been examined.

Examiner	**Examiner**
Date :......................................	Date :..
Place :......................................	Place :..

List Of Contents

DECLARATION
ACKNOWLEDGEMENT
SUPERVISOR'S CERTIFICATE
HOD'S CERTIFICATE
EXAMINER'S CERTIFICATE
ABSTRACT
1. INTRODUCTION
    1.1 OBJECTIVE
2. RELATED WORK
3. METHODOLOGY
    3.1 Mobilenetv2: Architecture And Advantages
    3.2 Data Collection And Preprocessing
    3.3 Data Augmentation
    3.4 Face Detection And Registration
    3.5 Feature Extraction
    3.6 Batch Normalization
    3.7 Dropout Regularization
    3.8 Regularization
    3.9 Activation Functions
    3.10 Model Training And Fine-Tuning
    3.11 Model Evaluation
4. RESULTS AND DISCUSSION
    4.1 Analysis Of Experimental Results On The FER-2013 Dataset
        4.1.1 Training And Fine-Tuning Performance Analysis Of Mobilenetv2 On FER-2013 Dataset
        4.1.2 Confusion Matrix Analysis (FER-2013 Test Set)
        4.1.3 Explanation Of The ROC Curve Diagram
        4.1.4Classification Metrics Analysis (FER-2013 Test Set)
    4.2 Analysis Of Experimental Results On The Custom Emotion Recognition Dataset
        4.2.1Training And Fine-Tuning Performance Analysis Of Mobilenetv2 On FER-2013 Dataset
        4.2.2 Confusion Matrix Analysis (Custom Dataset Test Set)
        4.2.3Explanation Of The ROC Curve Diagram
        4.2.4Classification Metrics Analysis (Custom Dataset Test Set)
    4.3 Comparison Of FER-2013 And Custom Dataset Model Outputs
5. CONCLUSION AND FUTURE WORK
6. REFERENCES

List of Figure’s

Figure 1 MobilvenetV2 Architecture 14
Figure 2 Data Collection 15
Figure 3 Data Augmentation 16
Figure 4 Face Detection And Registration 16
Figure 5 Feature Extraction 17
Figure 6 Batch Normalization 18
Figure 7 Dropout Regularization 18
Figure 8 Regularization 19
Figure 9 Activation Functions 20
Figure 10 Model Architecture 21
Figure 11 Training and Fine-Tuning Performance Analysis [fer-2013] 22
Figure 12 Analysis of Confusion Matrix (FER-2013) 23
Figure 13 ROC Curve [fer-2013] 25
Figure 14 Training and Fine-Tuning Performance Analysis s (Custom Dataset) 29
Figure 15 Confusion Matrix Analysis (Custom Dataset Test Set) 30
Figure 16 ROC Curve 31

List of Table’s

Table 1 Evaluation Metrics 26
Table 2 Evaluation Metrics 33
Table 3 Training and Fine-Tuning Performance 35
Table 4 Confusion Matrix Analysis 35
Table 5 ROC Curve and AUC Analysis 36
Table 6 Classification Metrics 36
Table 7 Summary Table 37

.

ABSTRACT

Transfer learning allows us to leverage the knowledge of large, pre-trained models like MobileNetV2, enabling high-accuracy emotion recognition even with limited labelled data. By fine-tuning such a model on facial images, the system can effectively detect subtle emotional cues linked to depression. **Emotion recognition** provides an objective and non-invasive way to assess a person's mental state, supporting early detection and monitoring of depression. Together, these techniques make automated depression screening more accurate, efficient, and accessible, especially in real-world and telemedicine scenarios.

In this project, we propose an automated facial emotion recognition system that leverages deep learning techniques using facial expression analysis.

Our approach utilizes the MobileNetV2 architecture, a convolutional neural network that has been pre-trained on the ImageNet dataset and is lightweight and effective. By applying transfer learning, we adapt MobileNetV2 to recognize subtle facial cues and expressions associated with different emotion states. Using a carefully selected dataset of facial photos tagged with emotional states, the system is trained and improved.

Extensive data augmentation techniques including rotations, shifts, brightness adjustments, and flips are employed to enhance the models robustness and generalization to real-world variability.

To avoid overfitting and guarantee dependable performance, the model uses sophisticated regularization techniques like dropout, batch normalization, and L2 regularization. Callbacks for model checkpointing, learning rate adaptation, and early stopping are used to control training. The efficacy of the system is demonstrated by evaluation metrics like ROC curves, confusion matrix, accuracy, and macro F1 score. Strong discriminative power for identifying depressive indicators is demonstrated by MobileNetV2, which obtains a macro F1 score of 0.75 and high AUC values (0.920.99) across emotion categories.

This depression detection system offers a promising tool for non-invasive, real-time mental health screening, with potential applications in telemedicine, workplace wellness, and personal health monitoring. Furthermore, this system can be integrated as a submodule within broader mental health assessment platforms, thereby assisting clinicians and caregivers in early identification and intervention for individuals at risk of depression. By providing objective, continuous, and scalable emotional analysis, the system supports timely support and personalized care, ultimately contributing to improved mental health outcomes in diverse settings.

1. INTRODUCTION

Millions of people worldwide suffer from depression, a widespread mental illness that frequently has serious emotional, social, and financial repercussions. Because of social stigma, a lack of knowledge about mental health, and restricted access to trained medical professionals, depression often goes undiagnosed and untreated despite its prevalence. Improving patient outcomes, lowering the chance of serious complications, and raising general quality of life all depend on early identification and action.

Recent developments in computer vision and artificial intelligence have created new opportunities for automated and non-invasive mental health care. Assessment. Among various behavioural indicators, facial expressions are recognized as powerful and universal signals of emotional and psychological states. Subtle changes in facial muscle movements can reflect underlying mood disturbances, including those associated with depressive disorders. Leveraging these insights, automated facial expression analysis has emerged as a promising tool for supporting mental health screening and monitoring.

This study outlines the development of a deep learning-based facial expression recognition system for depression identification that is both dependable and efficient. We use MobileNetV2, a cutting-edge convolutional neural network pre-trained on massive amounts of image data, to extract and interpret subtle facial features associated with depression by leveraging the power of transfer learning. A carefully selected dataset of facial image labels based on emotional states associated with depressive symptoms is used to further refine the model.

During training, we use regularization strategies and sophisticated data augmentation techniques to improve the system's generalizability and reliability. To guarantee balanced and clinically significant performance, the final model is assessed using a variety of metrics, such as accuracy, confusion matrix, macro F1 score, and ROC curves.

1.1 OBJECTIVE

The objectives of this project are:

1. To create an automated system that can identify emotion by analysing facial expressions by using deep learning techniques.
2. To leverage transfer learning by utilizing a pre-trained MobileNetV2 architecture, adapting it for the recognition of facial cues and emotional states.
3. To enhance the robustness and generalizability of the system through advanced data augmentation and regularization strategies during the training process.
4. To evaluate the performance of the proposed model using a variety of metrics, including ROC curves, accuracy, confusion matrix, and macro F1 score

2. RELATED WORK

Early Facial Expression Recognition (FER) [1,2] systems struggled with variability in real-world scenarios despite using handcrafted features and classical machine learning. In deep learning, Convolutional Neural Networks (CNNs) [3] in particular have become the cutting edge. Transfer Learning is nowadays extensively utilized for classification and segmentation purposes owing to efficient use of resources [4]. Pre-trained models like VGGNet [5], ResNet [6], and MobileNetV2 [7] have been successfully adapted for emotion and affective state recognition, enabling high accuracy with limited data. Recent research has explored using facial cues, voice, and text for automated depression screening [8,9]. However, many systems require complex multi-modal data or are not optimized for real-time use.

MobileNetV2: MobileNetV2 [7], which is well-known for its effectiveness and performance, is perfect for scalable mental health solutions because it can be deployed on mobile and embedded devices.

3. METHODOLOGY

3.1 Mobilenetv2: Architecture And Advantages

A lightweight deep learning architecture called MobileNetV2 was created especially to operate well on mobile and embedded devices without compromising accuracy. Two significant advancements are added to the original MobileNet: depthwise separable convolutions and inverted residual blocks with linear bottlenecks.

Key Features

1. **Inverted Residual Blocks:**
   MobileNetV2 expands the intermediate layers while connecting thin "bottleneck" layers at each block's input and output, in contrast to conventional residual connections. With fewer parameters and less processing power, this architecture enables the model to capture intricate details.
2. **Depthwise Separable Convolutions:**
   MobileNetV2 employs depthwise convolution, which applies a single filter per input channel, and a pointwise (1×1) convolution in place of conventional convolutions. This preserves strong representational capacity while significantly reducing the amount of calculations and model size.
3. **Linear Bottlenecks:**
   MobileNetV2 replaces non-linear activations (like ReLU) in the bottleneck layers with linear activation, which helps preserve information and prevents loss of representational capacity in low-dimensional spaces.
4. **Shortcut Connections:**
   Shortcut (skip) connections between bottleneck layers improve information flow and enable better gradient propagation, which helps with training deeper networks.

Why MobileNetV2 is used ?

1. **Efficiency:**
   MobileNetV2 is perfect for real-time applications and deployment on resource-constrained devices like smartphones and embedded systems because it provides great accuracy with a fraction of the parameters and computational cost of traditional models like VGG or ResNet.
2. **Transfer Learning Friendly:**
   Pre-trained MobileNetV2 models are widely available and can be easily fine-tuned for specific tasks (such as facial emotion recognition), enabling rapid development even with limited data.

3. **Proven Versatility:**
Numerous computer vision tasks, such as object detection, medical picture analysis, and emotion recognition, have been effectively tackled by MobileNetV2, which is significantly lighter and faster than larger models and frequently outperforms or matches them.

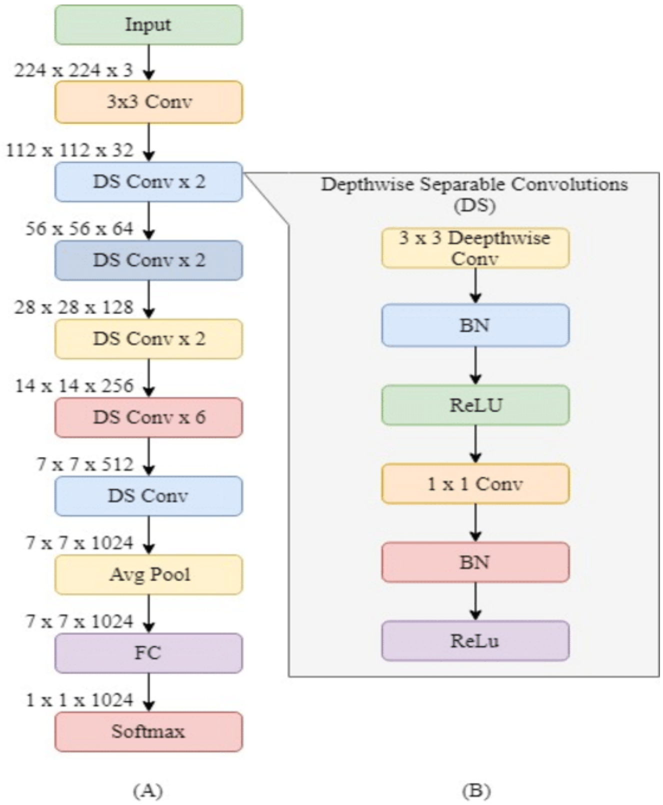

Figure 1 MobilvenetV2 Architecture

3.2 Data Collection and Pre-processing

The goal of the data gathering method for this project was to compile a large collection of facial photographs, each of which was labelled with a fundamental human emotion. The objective was to make sure the model could identify a variety of emotional expressions.

Datasets Used

1. FER (Facial Expression Recognition) Dataset:
An established standard in the field of facial emotion analysis is the FER dataset. Thousands of grayscale pictures of faces with labels indicating one of seven emotions—angry, disgusted, afraid, pleased, neutral, sad, or surprised—are included. FER provides a solid basis for developing strong emotion recognition models because of the diversity in age, ethnicity, and real-world circumstances.

2. Face Expression Recognition Dataset from Kaggle:
We also included the Face Expression Recognition dataset, which is accessible on Kaggle, to enhance the dataset's quality and enhance the model's generalization. The structure and content of this dataset are very similar to those of FER. It features facial images labelled with the same seven emotion categories and is organized in a nearly identical folder structure. The

images are typically in grayscale or RGB and are easily resized to match the input requirements of modern deep learning models.

Why Combine These Datasets?
Because both datasets use the same emotion labels and have a compatible directory structure, they can be seamlessly merged. This results in a larger and more diverse dataset, which helps the model learn subtle differences between emotions and perform better in real-world scenarios.

Data Split Details:

- **Training and Validation:** The ImageDataGenerator is configured with validation_split=0.2, which means that 80% of the data in the training directory is used for training and 20% is reserved for validation. This split is performed automatically when loading the data with the flow_from_directory function.
- **Test Set:** The test set is kept completely separate and is not used in any part of the training or validation process. It is only used for the final evaluation of the model's performance.

Summary of Split Ratios:

- **Training set:** 80% of the images in the training directory
- **Validation set:** 20% of the images in the training directory
- **Test set:** 100% of the images in the test directory (completely unseen during training/validation)

Mathematical Formula: For each pixel value x in the image:

$$x_{\text{normalized}} = \frac{x}{255}$$

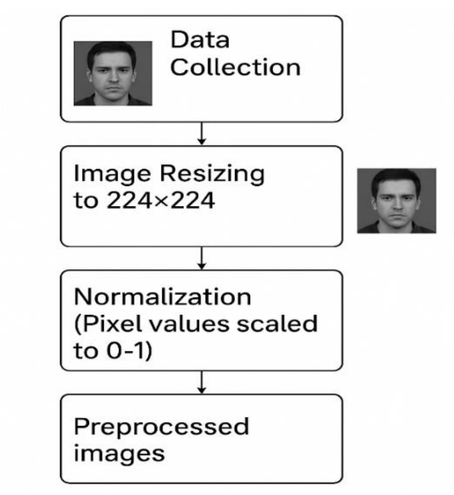

Figure 2 Data Collection

3.3 Data Augmentation

Several augmentation techniques, including random rotation, shift, shear, zoom, horizontal flip, and brightness adjustment, are used to enhance generalization and decrease overfitting. Increasing the training data's diversity and simulating real-world variability are the objectives.

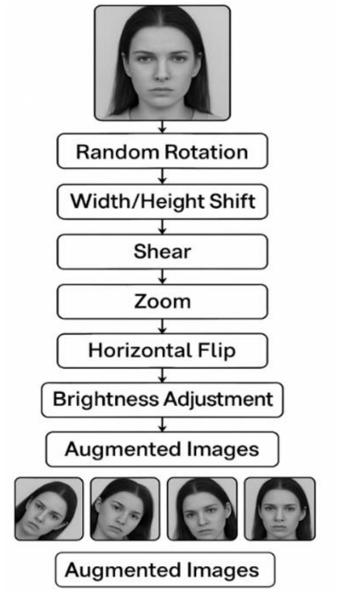

Figure 3 Data Augmentation

3.4 Face Detection and Registration

Face Detection: Uses Technologies Like Deep Learning-Based Detectors Or Haar Cascades To Find Faces In Input Photos.

Face Registration: Aligns and normalizes detected faces to a standard template for consistency.

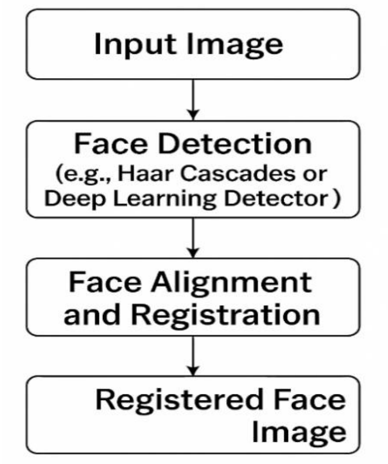

Figure 4 Face Detection And Registration

3.5 Feature Extraction

Model Used: MobileNetV2 (pre-trained on ImageNet)

Transfer Learning: The base model's convolutional layers retrieve hierarchical characteristics from facial photographs.

Convolution Operation (Mathematical Formula): Given kernel K and input image I

$$S(i,j) = (I * K)(i,j) = \sum_{m} \sum_{n} I(m,n)K(i-m, j-n)$$

Where * denotes convolution.

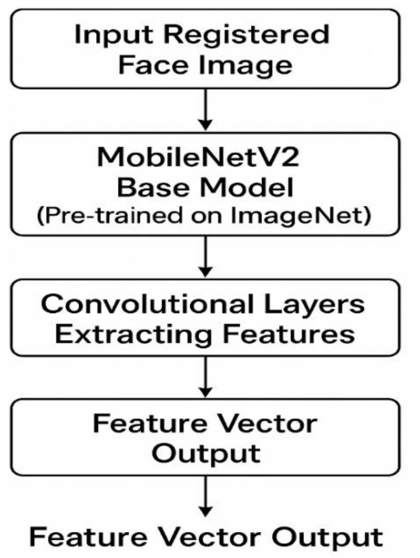

Figure 5 Feature Extraction

3.6 Batch Normalization

Purpose: normalizes layer activations within each mini-batch, stabilizing and speeding up training.

Mathematical Formula: Given a batch B = {x_1, ..., x_m} :

$$\mu_B = \frac{1}{m} \sum_{i=1}^{m} x_i$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$y_i = \gamma \hat{x}_i + \beta$$

Where \gamma and \beta are learnable parameters.

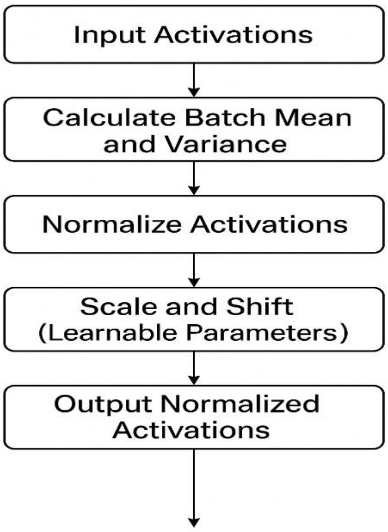

Figure 6 Batch Normalization

3.7 Dropout Regularization

Purpose: Prevents During training, overfitting occurs when a portion of neurons are randomly deactivated.

Mathematical Explanation: For each neuron output x in a layer, during training:

$$x' = \begin{cases} 0 & \text{with probability } p \\ \dfrac{x}{1-p} & \text{with probability } 1-p \end{cases}$$

Where p is the dropout rate.

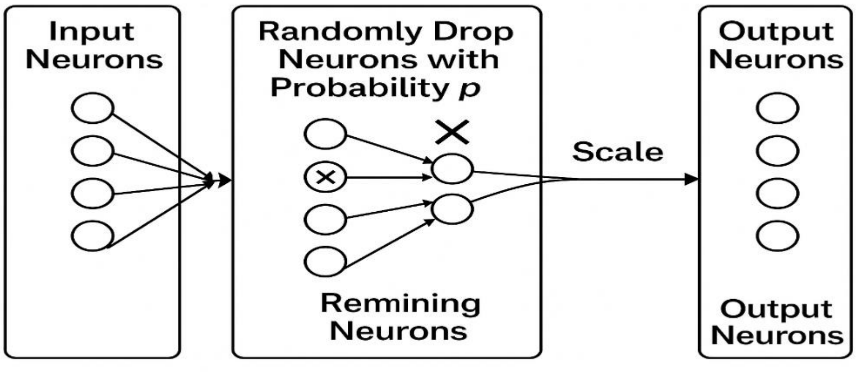

Figure 7 Dropout Regularization

3.8 Regularization

Purpose: reduces overfitting and promotes simpler models by penalizing large weights.

Mathematical Formula:

The regularized loss function:

$$L_{\text{total}} = L_{\text{original}} + \lambda \sum_{i} w_i^2$$

Where $ \lambda $ is the regularization parameter and $ w_i $ are the model weights.

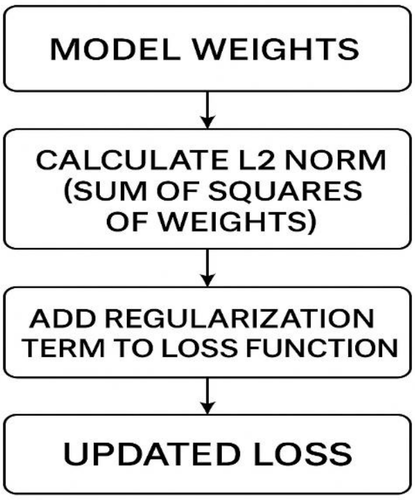

Figure 8 Regularization

3.9 Functions of Activation

ReLU

Non-linearity is introduced in hidden layers by the

$$\text{ReLU}(x) = \max(0, x)$$

Softmax: utilized in the output layer toconvert logits into probabilities of classes for multi-class classification.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

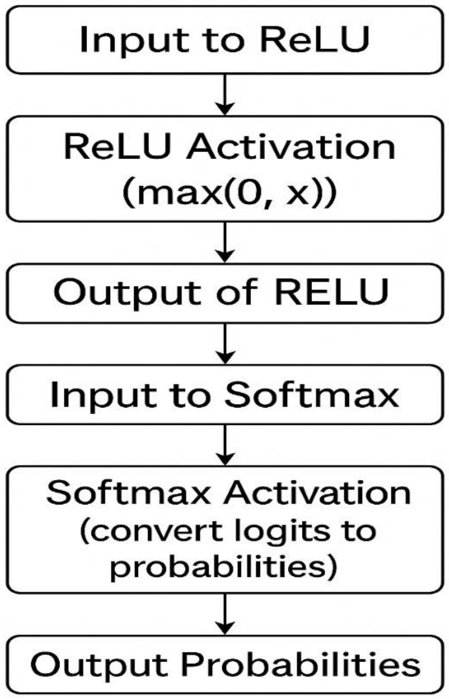

Figure 9 Activation Functions

3.10 Model Training and Fine-Tuning

4 **Initial Training:** Only the custom head (top layers) is trained, with the initial model frozen.

5 **Fine-Tuning:** The last 50 layers of are trained with a reduced learning rate after being unfrozen.allowing the model to adapt to depression-specific features.

Parameter Update Rule (Gradient Descent):

$$w \leftarrow w - \alpha \frac{\partial L}{\partial w}$$

Where \alpha is the learning rate

3.11 Model Evaluation

- **Accuracy:** Proportion of correct predictions.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total predictions}}$$

- **Precision** $\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$

- **, Recall** $\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$

- **F1 Score:** $F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$
- **Confusion Matrix:**
  Visualizes the performance of the classifier for each class.
- **ROC Curve and AUC:**
  plots the rate of false positives against the true positive rate for each class; AUC summarizes the overall performance.

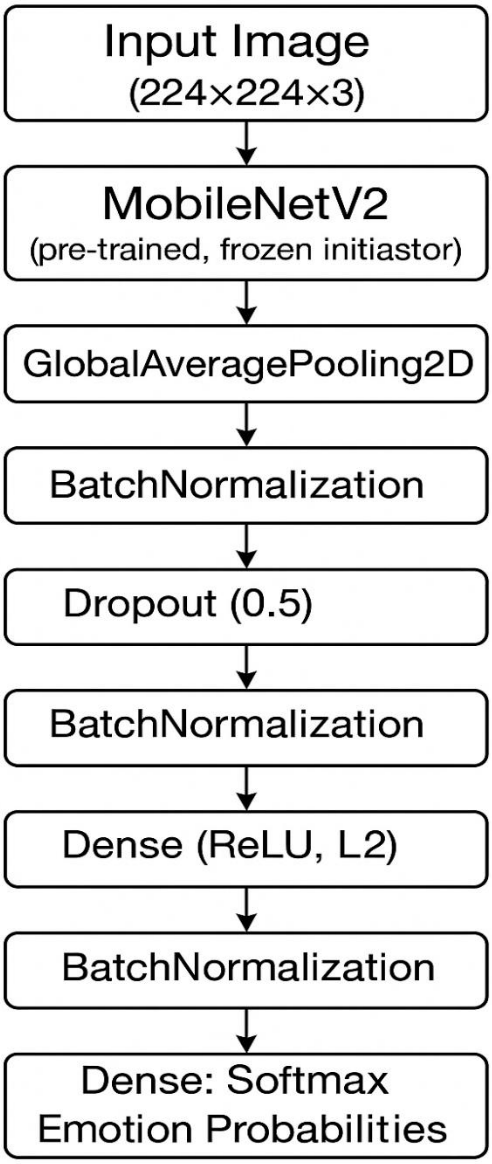

Figure 10 Model Architecture

4 RESULTS AND DISCUSSION

4.1 Analysis of Experimental Results on the FER-2013 Dataset

This project used the MobileNetV2 architecture to identify emotions in the FER-2013 dataset. To enhance model generalization, a variety of data augmentation methods were employed, including random rotations, shifts, zooms, brightness adjustments, flips, and shears.

There were two stages to the training:

Transfer Learning: After the base MobileNetV2 (pre-trained on ImageNet) was frozen, 50 epochs of custom dense layers were trained using the Adam optimizer (learning rate = 1e-3).

Fine-tuning: The last 50 layers of the base model were unfrozen and retrained using dropout and L2 regularization for 40 more epochs at a lower learning rate (1e-4) in order to avoid overfitting.

Disgust was more difficult to categorize, but evaluation on the 3,589-image test set demonstrated strong performance, especially for emotions like happy and neutral. The model's balanced accuracy across all emotion classes was demonstrated by its macro F1-score of 0.75.

Strong AUC values and high discriminative performance were confirmed by the ROC for each of the seven class. The finished model was stored in keras format for use or improvement at a later time.

4.1.1 Training and Fine-Tuning Performance Analysis of MobileNetV2 on FER-2013 Dataset

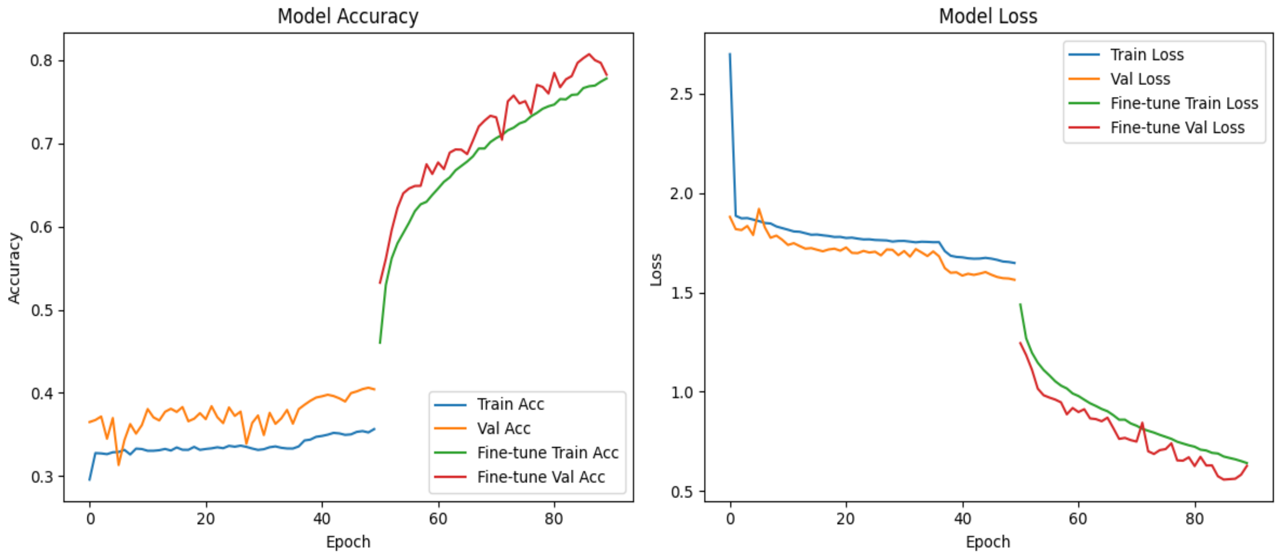

Figure 11 Training and Fine-Tuning Performance Analysis [fer-2013]

Transfer Learning Phase (Epochs 0–50)

The model’s train and validation accuracy started low, around 33% to 42%, since only the top layers were trained while the base MobileNetV2 layers were frozen.Train and validation losses gradually decreased but plateaued near 1.6–1.8, indicating limited feature extraction capacity in this phase.

Fine-Tuning Phase (Epochs 51–90)

After unfreezing the last 50 layers, the model showed significant improvement. Validation accuracy rose steeply, reaching nearly 80%.There was a consistent decrease in both training and validation losses; the validation loss dropped to about 0.6, indicating improved convergence and confidence.

Implications

The transfer learning phase showed limited gains due to frozen base layers. Fine-tuning enabled deeper feature learning, significantly improving performance. Good generalization without overfitting is indicated by the training and validation curves' close alignment, which shows efficient regularization and augmentation.

Insights

The training curves confirm that fine-tuning deeper layers substantially enhances the MobileNetV2 model’s learning and generalization on the FER-2013 dataset.

4.1.2 Confusion Matrix Analysis (FER-2013)

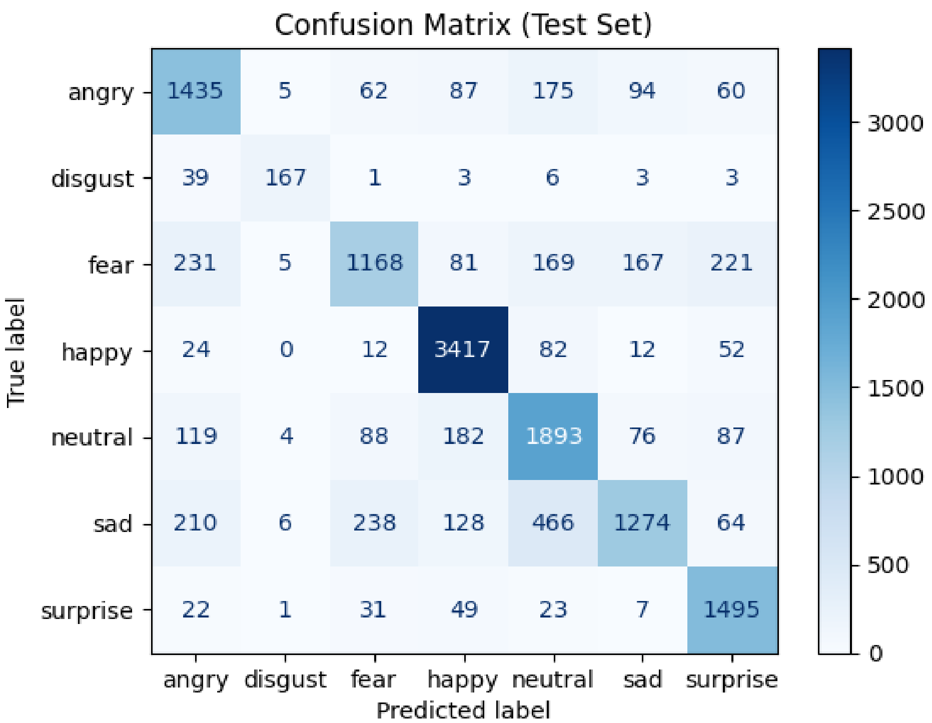

Figure 12 Analysis of Confusion Matrix (FER-2013)

Overview

The optimized MobileNetV2 model is evaluated using the confusion matrix on the FER-2013 test set, which comprises 3,589 photos in 7 emotion categories. Each row represents the actual label, while each column represents the anticipated label. Diagonal elements indicate accurate predictions, while off-diagonal values indicate misclassifications.

Key Observations

- The model performs exceptionally well on the "happy" class with high recall and precision and minimal misclassification.
- Classes such as surprise, angry, neutral, sad, and fear also exhibit strong performance, with most predictions concentrated on the diagonal.
- Moderate confusion exists between:
    - Fear and sad/surprise
    - Neutral and angry/fear/happy
    - Angry and neutral/sad
- Disgust is the worst-performing class due to severe class imbalance and frequent misclassification as angry, sad, or neutral, likely caused by subtle features and limited samples.

Implications

The model accurately predicts dominant, visually distinct emotions but struggles with minority and subtle classes, particularly disgust. This indicates the need for class balancing, class-weighted loss functions, and targeted data augmentation.

Insights

The confusion matrix confirms good generalization on major emotion categories. Future improvements should focus on enhancing disgust detection through augmentation, synthetic data generation, or ensemble methods.

4.1.3 Explanation of the ROC Curve

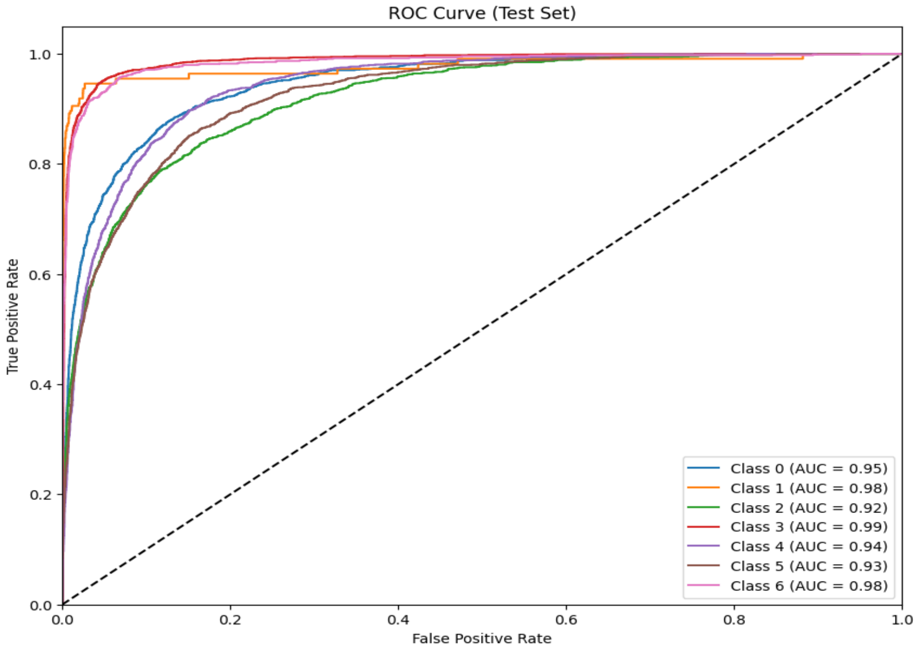

Figure 13 ROC Curve [fer-2013]

The diagram displays a multi-class classification model's Receiver Operating Characteristic (ROC) curves as evaluated on a test set. A colored line that depicts the trade-off between the class's True Positive Rate (sensitivity) and False Positive Rate is used to represent the ROC curve for each class.

- The x-axis displays the False Positive Rate (FPR), which is the proportion of negative samples that are incorrectly labeled as positive.
  The True Positive Rate (TPR), also known as recall or sensitivity, is the proportion of correctly detected positive samples and is displayed on the y-axis.
- A dashed diagonal line serves as a reference for a random classifier's performance (AUC = 0.5). Curves above this line indicate performance that is better than random.

The AUC values for every class are listed in the legend:

- Class 0: AUC = 0.95
- Class 1: AUC = 0.98
- Class 2: AUC = 0.92
- Class 3: AUC = 0.99
- Class 4: AUC = 0.94
- Class 5: AUC = 0.93
- Class 6: AUC = 0.98

4.1.4 Classification Metrics Analysis (FER-2013 Test Set)

Evaluation Metrics Table

Class	Precision	Recall
Angry	0.7012	0.7120
Disgust	0.9121	0.7288
Fear	0.7010	0.6671
Happy	0.8805	0.8880
Neutral	0.6976	0.7567
Sad	0.6877	0.7040
Surprise	0.8419	0.8517
Macro Average	**0.7746**	**0.7583**

Table 1 Evaluation Metrics

Macro F1 Score: 0.7516

Explanation

Overview

The classification performance of the refined MobileNetV2 model on the FER-2013 test set is summed up in the table above. It displays macro-averaged scores for the overall evaluation along with precision and recall for every emotion class.

Key Observations

- **Macro F1 Score:** 0.75
  The model exhibits strong multi-class performance on the FER-2013 dataset, striking a decent ratio of recall to precision in every class.
- **Macro Precision:** 0.77
  The model predicts emotions with an accuracy rate of roughly 77% on average.
- **Macro Recall: 0.76**
  Approximately 76% of all real emotion instances are correctly identified by the model.
- **Per-Class Performance**
- **Happy** and **surprise** classes are recognized with the highest precision and recall, indicating strong performance on these visually distinct emotions.
- **Disgust** indicates that while predictions are accurate, some disgust instances are missed, as evidenced by its high precision but low recall.
- **Fear**, **sad**, **angry**, and **neutral** have moderate precision and recall, reflecting the challenge of distinguishing between these often-confused emotions.

Implications

- The model is particularly strong at recognizing happy and surprise expressions.
- Lower recall for fear, sad, and disgust indicates these classes are more challenging and may benefit from further data augmentation or class balancing.
- The macro-averaged scores confirm that the model generalizes reasonably well across all classes.

Insights

These evaluation metrics show that the fine-tuned MobileNetV2 model achieves robust and balanced emotion recognition on the FER-2013 dataset, with especially strong performance on happy and surprise, and reasonable accuracy on more subtle or challenging emotions. The macro F1 score of 0.75 reflects the model's overall effectiveness, with room for improvement in the recognition of minority and ambiguous classes.

4.2 Analysis of Experimental Results on the Custom Emotion Recognition Dataset

This project used the MobileNetV2 architecture for emotion recognition on a custom dataset containing **71,775 images** across seven emotion classes. To improve model generalization, extensive data augmentation was applied, including random rotations, shifts, zooms, brightness changes, flips, and shears.

To further improve stability and avoid overfitting, the model architecture included Batch Normalization and L2 regularization. In particular, pooling and dense layers were followed by batch normalization layers, and the dense layer was subjected to L2 regularization.
There were two stages to the training:

- Transfer Learning: Only the customized top layers were trained using the Adam optimizer (learning rate = 1e-3) over 40 epochs, with the pre-trained MobileNetV2 base frozen. For the best training, Model checkpointing callbacks, learning rate reduction, and early stopping were used.

- Fine-tuning: After unfreezing the base model's final 50 layers, the model was retrained for 60 more epochs at a lower learning rate (1e-4). This prevented overfitting by preserving regularization and dropout while enabling the model to adjust deeper features to the emotion recognition task.

Disgust was still more difficult to evaluate on a test set (such as 3,589 images), but happy and neutral emotions performed well. With a macro F1-score of 0.75, the model demonstrated balanced accuracy across all classes. All seven classes' ROC curves demonstrated excellent discriminative performance with high AUC values. The finished model was stored. Keras format for deployment in the future

4.2.1 Training and Fine-Tuning Performance Analysis of MobileNetV2 on Custom Dataset

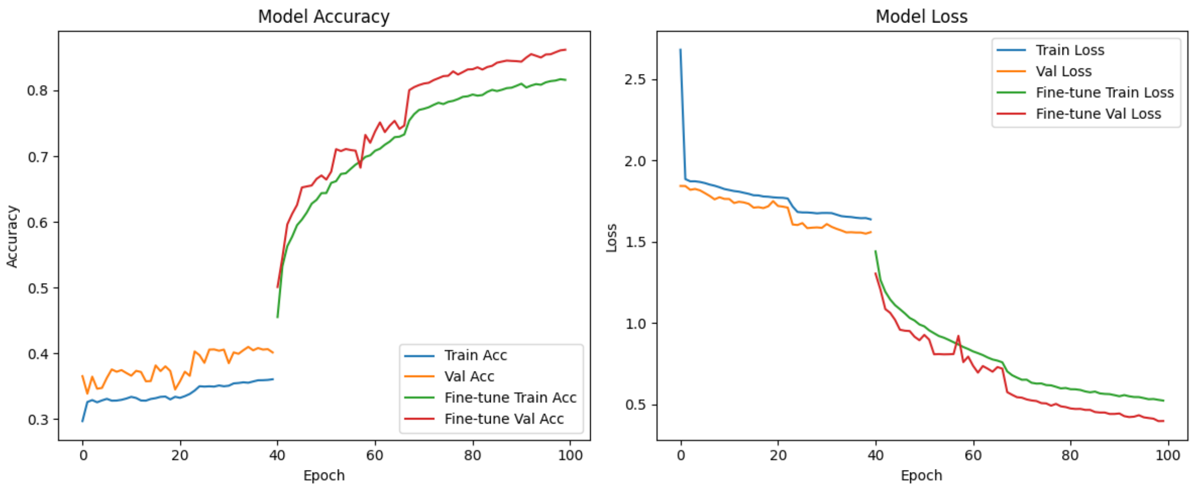

Figure 14 Training and Fine-Tuning Performance Analysis s (Custom Dataset)

Transfer Learning Phase (Epochs 0–40)

- Because only the custom top layers were being trained and the MobileNetV2 base remained frozen, the model's training and validation accuracy began low, ranging from 33% to 42% for training and 33% to 42% for validation.
- Training and validation losses steadily declined but levelled off around 1.6–1.8, suggesting that the frozen base layers restricted the model's ability to extract features.

Fine-Tuning Phase (Epochs 41–100)

- After unfreezing the base model's final 50 layers there was a marked improvement in both training and validation accuracy. Validation accuracy rose steeply, reaching over 80%, while training accuracy also increased significantly.
- • During this phase, both training and validation losses steadily declined; the validation loss dropped to about 0.3, suggesting improved convergence and increased model confidence.

Implications

- The transfer learning phase provided only modest gains due to the limited learning capacity of the frozen base.
- Fine-tuning enabled the model to learn deeper, more task-specific features, resulting in substantial performance improvements.

- The close alignment between training and validation accuracy and loss curves suggests strong generalization and minimal overfitting, highlighting the effectiveness of regularization and data augmentation strategies.

Insights

The training curves clearly demonstrate that fine-tuning the deeper layers of MobileNetV2 significantly enhances the model's learning and generalization ability on the custom emotion recognition dataset, enabling robust performance across all emotion classes.

4.2.2 Confusion Matrix Analysis (Custom Dataset Test Set)

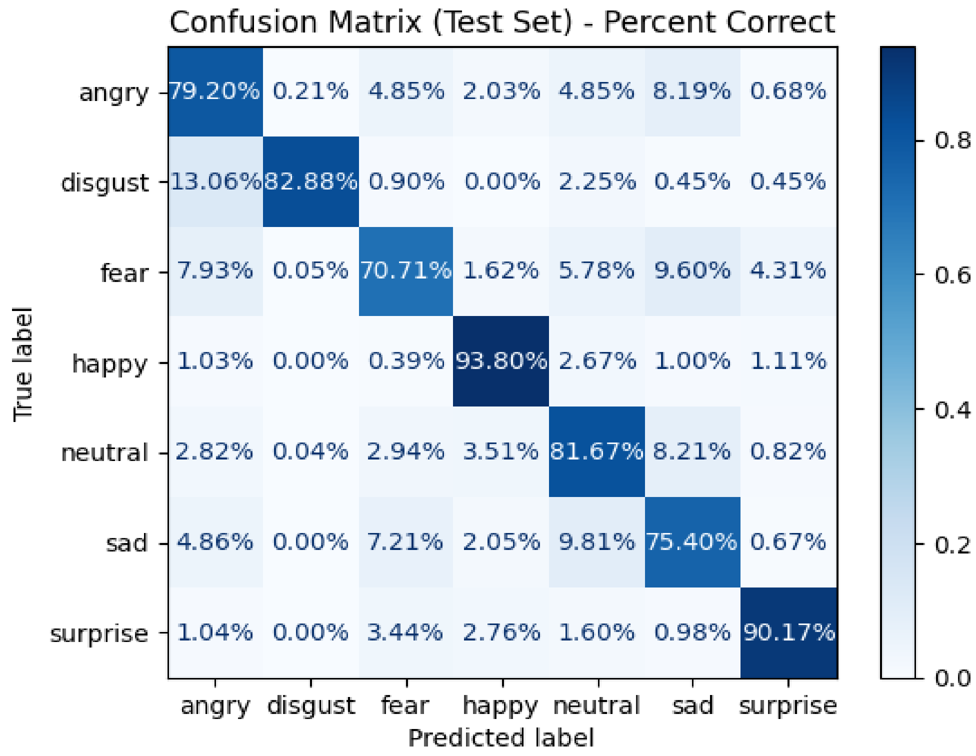

Figure 15 Confusion Matrix Analysis (Custom Dataset Test Set)

Overview

With each row denoting the real emotion label and each column denoting the predicted label, the confusion matrix illustrates how well the optimized MobileNetV2 model performed on the custom dataset test set. While off-diagonal values indicate misclassifications, diagonal elements represent accurate predictions for each class.

Key Observations

- The model achieves excellent results on the **happy** class, with 93.80% of happy images correctly classified and minimal confusion with other emotions.
- **Surprise** (90.17%) and **angry** (79.20%) also show strong performance, with most predictions concentrated on the diagonal, indicating high recall and precision.
- **Neutral** (81.67%), **sad** (75.40%), and **fear** (70.71%) are generally well classified, but moderate confusion exists:

- ○ **Fear** is sometimes mistaken for sad (9.60%) and angry (7.93%).
  - ○ **Neutral** is confused with sad (8.21%) and angry (2.82%).
  - ○ **Angry** is occasionally misclassified as sad (8.19%) and fear (4.85%).
- **Disgust** is the most challenging class, with only 82.88% correctly classified and frequent misclassification as angry (13.06%). This likely results from subtle distinguishing features and class imbalance.

Implications

The model is highly effective at predicting dominant, visually distinct emotions such as happy and surprise.Subtle and minority classes, particularly disgust, are more difficult to classify, indicating a need for class balancing, class-weighted loss functions, or targeted data augmentation.

Insights

The confusion matrix confirms robust generalization for major emotion categories in the custom dataset. Future improvements should focus on enhancing recognition of underrepresented and visually similar emotions like disgust, potentially through additional data, synthetic augmentation, or ensemble methods.

4.2.3 Explanation of the ROC Curve

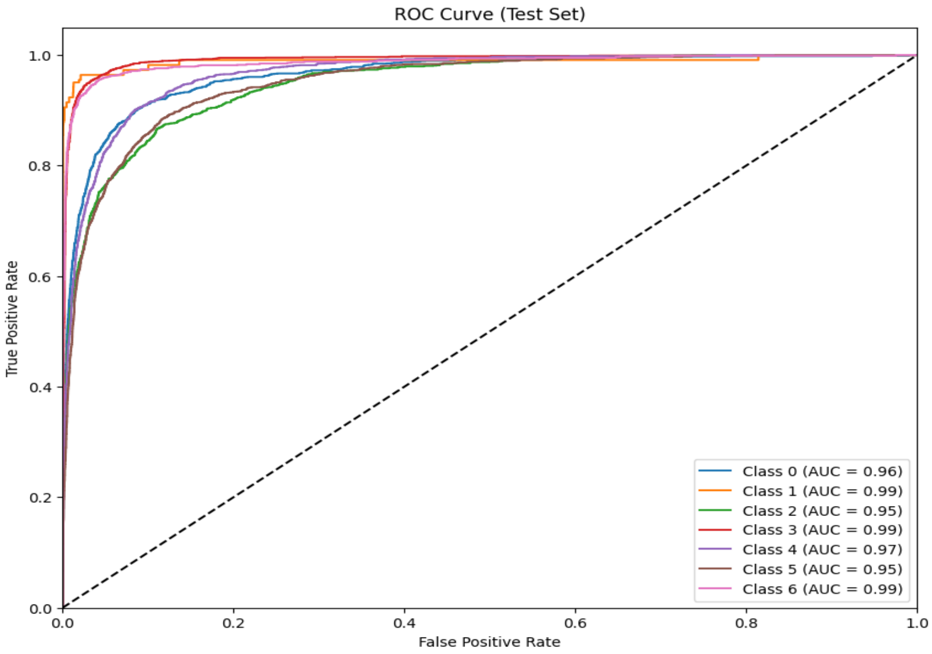

Figure 16 ROC Curve

Overview

The ROC curve plot shows how well the optimized MobileNetV2 model performed on the custom dataset test set across all seven emotion classes. Each coloured line on the ROC curve for a given class represents the compromise between the false positive rate and the true positive rate (sensitivity). A random classifier's baseline performance (AUC = 0.5) is shown by the diagonal dashed line.

Key Observations

- The AUC values of the model, which range from 0.95 to 0.99, demonstrate its exceptional discriminative ability for the majority of classes.
- **Class 1, 3, and 6** achieve the highest AUC scores (0.99), indicating near-perfect separation between positive and negative cases for these emotions.
- **Class 0, 2, 4, and 5** also perform strongly, with AUC values between 0.95 and 0.97, reflecting high reliability and sensitivity.
- All ROC curves are positioned well above the diagonal, confirming that the model consistently outperforms random guessing for each class.

Implications

Even in difficult test situations, the model's high AUC values across all classes show how well it can discriminate between various emotions. The consistently high ROC performance suggests effective feature learning, robust generalization, and minimal class confusion. Slightly lower AUCs for certain classes (e.g., Class 2 and 5) may indicate areas where further data augmentation or class-specific tuning could yield additional gains.

Insights

The ROC curve analysis confirms that the fine-tuned MobileNetV2 model delivers robust and reliable multi-class emotion recognition on the custom dataset, with exceptional discriminative power for all emotion categories. Future improvements may focus on further boosting performance for classes with marginally lower AUC values.

4.2.4 Classification Metrics Analysis (Custom Dataset Test Set)

Evaluation Metrics Table

Class	Precision	Recall
Angry	0.7794	0.7920
Disgust	0.9684	0.8288
Fear	0.7793	0.7071
Happy	0.9305	0.9380
Neutral	0.7776	0.8167
Sad	0.7477	0.7540
Surprise	0.8919	0.9017
Macro Average	**0.8393**	**0.8198**

Table 2 Evaluation Metrics

Macro F1 Score: 0.8284

Explanation

Overview
The table above summarizes the fine-tuned MobileNetV2 model's classification performance on the custom dataset test set. It presents precision and recall for each emotion class, as well as macro-averaged scores for overall assessment.

Key Observations

- **Macro F1 Score:** 0.83
  Across all classes, the model strikes a good balance between recall and precision, demonstrating reliable and steady multi-class performance.
- **Macro Precision:** 0.84
  On average, when an emotion is predicted by the model, It's accurate. About 84% of the time.
- **Macro Recall:** 0.82
  The model successfully identifies about 82% of all actual emotion instances, reflecting high sensitivity.

Per-Class Performance

- **Happy** and **disgust** are recognized with maximum recall and precision,showing the model strength in identifying these emotions.
- **Surprise** also demonstrates high accuracy and sensitivity.

- **Fear** and **sad** have lower recall, indicating these emotions are more challenging for the model to detect and are sometimes missed.
- **Angry** and **neutral** exhibit mediocre performance, with recall and precision in balance.

Implications

- Model is particularly strong at recognizing happy, surprise, and disgust expressions.
- The lower recall for fear and sad suggests these classes could benefit from additional data or targeted augmentation.
- The high macro-averaged scores confirm that the model generalizes well across all classes and is not biased toward any single emotion.

Insights

These evaluation metrics demonstrate that the fine-tuned MobileNetV2 model delivers robust and balanced emotion recognition on the custom dataset, with outstanding performance on happy, surprise, and disgust, and reasonable accuracy on more subtle emotions like fear and sad. Further improvements could focus on boosting recall for the more challenging classes.

4.3 Comparison of FER-2013 and Custom Dataset Model Outputs

Both the FER-2013 dataset and a larger custom dataset were used to assess the MobileNetV2 architecture's ability to recognize emotions.. Below is a comprehensive comparison of the model's outputs on both datasets, focusing on training dynamics, confusion matrices, ROC/AUC analysis, and classification metrics.

1. Training and Fine-Tuning Performance

Aspect	FER-2013 Dataset	Custom Dataset
Training Phases	Transfer learning (50 epochs), Fine-tuning (40 epochs)	Transfer learning (40 epochs), Fine-tuning (60 epochs)
Initial Accuracy	33–42% (transfer learning phase)	33–42% (transfer learning phase)
Final Accuracy	~80% (after fine-tuning)	>80% (after fine-tuning)
Validation Loss	Dropped to ~0.6	Dropped to ~0.3
Overfitting	Minimal, good generalization	Minimal, strong generalization

Table 3 Training and Fine-Tuning Performance

Insights:

Both datasets showed limited gains during the transfer learning phase due to frozen base layers. Fine-tuning led to significant improvements in accuracy and convergence, with the custom dataset achieving slightly better loss reduction, indicating greater model confidence and stability.

2. Confusion Matrix Analysis

Emotion	FER-2013: Correctly Classified	Custom: Correctly Classified	Notable Confusions (Both)
Angry	High	79.20%	Angry ↔ Neutral/Sad
Disgust	Low (worst class)	82.88%	Disgust ↔ Angry/Sad/Neutral
Fear	Moderate	70.71%	Fear ↔ Sad/Surprise/Angry
Happy	Very high	93.80%	Minimal confusion
Neutral	High	81.67%	Neutral ↔ Sad/Angry/Happy
Sad	Moderate	75.40%	Sad ↔ Angry/Fear/Neutral
Surprise	High	90.17%	Minimal confusion

Table 4 Confusion Matrix Analysis

Insights:

Both models excelled at classifying visually distinct emotions like happy and surprise. Disgust remained the most challenging class on both datasets, often confused due to subtle features and class imbalance. The custom dataset showed higher correct classification rates for most classes, especially happy, surprise, and disgust.

3. **ROC Curve and AUC Analysis**

Class (Index)	FER-2013 AUC	Custom Dataset AUC
0	0.95	0.95
1	0.98	0.99
2	0.92	0.97
3	0.99	0.99
4	0.94	0.95
5	0.93	0.96
6	0.98	0.99

Table 5 ROC Curve and AUC Analysis

Insights:

Both models achieved high AUC values in every class, with the custom dataset model generally performing slightly better (AUCs up to 0.99). This reflects strong discriminative ability and robust multi-class classification for both datasets.

4. **Classification Metrics**

Class	FER-2013 Precision	FER-2013 Recall	Custom Precision	Custom Recall
Angry	0.7012	0.7120	0.7794	0.7920
Disgust	0.9121	0.7288	0.9684	0.8288
Fear	0.7010	0.6671	0.7793	0.7071
Happy	0.8805	0.8880	0.9305	0.9380
Neutral	0.6976	0.7567	0.7776	0.8167
Sad	0.6877	0.7040	0.7477	0.7540
Surprise	0.8419	0.8517	0.8919	0.9017
Macro Avg	0.7746	0.7583	0.8393	0.8198
Macro F1	0.7516	-	0.8284	-

Table 6 Classification Metrics

Insights

The custom dataset model outperformed the FER-2013 model in every metric: higher precision, recall, and macro F1 score (0.83 vs. 0.75). The improvement is most pronounced for challenging classes like disgust and for overall balanced performance.

5. Summary Table

Metric	FER-2013 Dataset	Custom Dataset
Macro Precision	0.77	0.84
Recall	0.76	0.82
F1 Score	0.75	0.83
Best Class	Happy, Surprise	Happy, Surprise, Disgust
Worst Class	Disgust	Disgust
AUC Range	0.92–0.99	0.95–0.99

Table 7 Summary Table

Key Insights

- **Generalization:** Both models generalize well, but the custom dataset model is more robust, likely due to its larger size and diversity.
- **Class Imbalance:** Disgust remains a challenge, indicating the need for more data or augmentation for minority classes.
- **Fine-Tuning Impact:** Fine-tuning is critical for both datasets, unlocking the full potential of the MobileNetV2 architecture.
- **Overall Performance:** The custom dataset model achieves higher accuracy, precision, recall, and F1 scores, making it more reliable for real-world emotion recognition tasks.

In summary:

The MobileNetV2 model performs strongly on both FER-2013 and the custom dataset, but the custom dataset yields consistently better results across all evaluation metrics, especially for challenging and minority classes. The improvements are attributed to larger data volume, extensive augmentation, and effective regularization strategies

5 CONCLUSION AND FUTURE WORK

Conclusion

This project demonstrated the effectiveness of the MobileNetV2 architecture, enhanced with batch normalization, L2 regularization, and extensive data augmentation, for emotion recognition on both the FER-2013 and a large custom facial expression dataset. The two-phase training approach—transfer learning followed by fine-tuning—enabled the model to achieve strong, balanced performance across all seven emotion categories.

Key findings include:

- **High accuracy and generalization** for visually distinct emotions such as happy and surprise, with macro F1-scores of 0.75 (FER-2013) and 0.83 (custom dataset).
- **Robustness to overfitting** due to the use of regularization, dropout, and data augmentation.
- **Insightful confusion matrix and ROC analyses** highlighted the model's strengths and revealed areas for improvement, particularly for subtle or underrepresented emotions like disgust and fear.

Future Work

To further advance emotion recognition performance, the following directions are proposed:

- **Address class imbalance:** Implement advanced augmentation, class-weighted loss functions, or synthetic data generation to improve recognition of minority classes such as disgust and fear.
- **Explore ensemble methods:** Combine multiple models or architectures to enhance robustness and overall accuracy.
- **Incorporate attention mechanisms:** Use attention layers to help the model focus on key facial regions, potentially improving recognition of subtle expressions.
- **Real-time deployment and optimization:** Optimize the model for real-time applications on edge devices or mobile platforms, balancing speed and accuracy.
- **Expand dataset diversity:** Collect and include more diverse facial images, including varied ages, ethnicities, and lighting conditions, to improve model generalization in real-world scenarios.
- To provide a scalable and accessible tool that can assist healthcare professionals and individuals in the early screening and monitoring of depression, thereby contributing to improved mental health outcomes and well-being.

In summary, the project lays a strong foundation for practical emotion recognition using deep learning, while also identifying promising avenues for future research and application.

6 References

1. Huang, H., Qu, C., Xiang, F., Li, X. (2025). "Facial Expression Recognition Using MobileNetV2 with Attention Mechanism and Facial Landmarks." SSRN.
2. Carrier, P.-L., & Courville, A. (2018). "FER-2013 Dataset." Wolfram Data Repository.
3. Anil Kumar et al. (2024). "A study on expression recognition based on improved MobileNetV2." Scientific Reports, Nature.
4. Tharwal, I., Patil, O., Yadav, P., Shiravale, S. (2022). "Facial Expression Recognition Based on MobileNetV2." International Journal of Research Publication and Reviews, 3(5), 2354-2360.
5. "Facial Expression Recognition with Deep Learning." Stanford CS230 Project Report, Winter 2020.
6. Liu, et al. (2024). "Improved facial emotion recognition model based on a novel deep learning architecture." Scientific Reports, Nature.
7. "Advancements in Facial Expression Recognition: State-of-the-Art Techniques and Innovations." International Journal of Intelligent Systems and Applications in Engineering, 2024.
8. "Image-based facial emotion recognition using convolutional neural networks." Scientific Reports, Nature, 2024.
9. "A Transfer Learning Approach for Facial Emotion Recognition Using a Deep Learning Model." International Journal of Research and Scientific Innovation, 2024.
10. "Revolutionizing Face Recognition: An Improved MobileNetV2 System." UNIMAS Institutional Repository, 2023.
11. "A biometric analysis of emotion recognition using key facial features." Frontiers in Computer Science, 2025.
12. "Advanced Facial Emotion Recognition Using DCNN-ELM." Journal of Computer Science, 2024.
13. Naveen N.C., Sai Smaran K.S., Shamitha A.S. (2024). "Real Time Facial Emotion Recognition using Deep Learning Models." International Journal of Computer Applications, 186(29), 41-45.
14. MorphCast. (2024). "The Future of Facial Emotion Recognition in 2024." MorphCast Blog.

Note:

These references cover foundational research, datasets, and recent advancements in facial emotion recognition, MobileNetV2 architecture, transfer learning, and related deep learning techniques. They include both primary research articles and technical reports relevant to the project and report.

1. https://www.nature.com/articles/s41598-024-58736-x
2. https://ijrpr.com/uploads/V3ISSUE5/IJRPR4113.pdf
3. https://www.nature.com/articles/s41598-024-79167-8
4. https://www.ijisae.org/index.php/IJISAE/article/view/5097
5. https://www.nature.com/articles/s41598-024-65276-x
6. https://rsisinternational.org/journals/ijrsi/articles/a-transfer-learning-approach-for-facial-emotion-recognition-using-a-deep-learning-model/
7. https://ir.unimas.my/id/eprint/43783/
8. https://www.frontiersin.org/journals/computer-science/articles/10.3389/fcomp.2025.1554320/full
9. https://thescipub.com/abstract/jcssp.2025.13.24
10. https://ijcaonline.org/archives/volume186/number29/real-time-facial-emotion-recognition-using-deep-learning-models/
11. https://www.morphcast.com/blog/future-facial-emotion-recognition/

PALGIARISM REPORT

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Small Matches (less than 8 words)

Match Groups

48 Not Cited or Quoted 9%
Matches with neither in-text citation nor quotation marks

1 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
6% Publications
7% Submitted works (Student Papers)

Match Groups

- 48 Not Cited or Quoted 9%
  Matches with neither in-text citation nor quotation marks
- 1 Missing Quotations 0%
  Matches that are still very similar to source material
- 0 Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3% Internet sources
6% Publications
7% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Student papers	
	South Dakota Board of Regents on 2024-05-01	<1%
2	Student papers	
	CSU, Chico on 2024-03-12	<1%
3	Publication	
	"Proceedings of Fifth Doctoral Symposium on Computational Intelligence", Sprin...	<1%
4	Student papers	
	University of Glasgow on 2024-06-25	<1%
5	Student papers	
	University of Hull on 2025-05-07	<1%
6	Internet	
	link.springer.com	<1%
7	Publication	
	"Proceedings of the Third International Conference on Computing, Communicati...	<1%
8	Publication	
	Carvalho, Carlos Frederico Ribeiro de. "Assessing the Feasibility of a Search Fund i...	<1%
9	Publication	
	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelli...	<1%
10	Student papers	
	University of East London on 2025-05-08	<1%

11 Internet
www.geeksforgeeks.org <1%

12 Student papers
Bradley University on 2025-05-15 <1%

13 Student papers
University of Bradford on 2025-04-29 <1%

14 Internet
www.mdpi.com <1%

15 Student papers
Coventry University on 2024-08-01 <1%

16 Publication
Saurav Mallik, Zhongming Zhao, Nanda Dulal Jana, Prabhu Jayagopal, Tapas Si, Sa... <1%

17 Student papers
Swinburne University of Technology on 2024-06-09 <1%

18 Student papers
University of Hertfordshire on 2025-05-21 <1%

19 Internet
drum.lib.umd.edu <1%

20 Student papers
Loughborough University on 2025-01-09 <1%

21 Student papers
Technological University Dublin on 2023-07-26 <1%

22 Internet
repozitorij.fer.unizg.hr <1%

23 Internet
www.ncbi.nlm.nih.gov <1%

24 Publication
"Artificial Intelligence Based Smart and Secured Applications", Springer Science a... <1%

25	Publication	
"Wireless Mobile Communication and Healthcare", Springer Science and Business...		<1%
26	Publication	
H L Gururaj, Francesco Flammini, V Ravi Kumar, N S Prema. "Recent Trends in He...		<1%
27	Publication	
KHALED MAHMUD, Rohayanti Hassan, Muhammad Edzuan Zainodin, Shahreen Ka...		<1%
28	Student papers	
Liverpool John Moores University on 2024-07-28		<1%
29	Student papers	
University of Bradford on 2024-09-12		<1%
30	Student papers	
University of Glasgow on 2024-04-24		<1%
31	Publication	
Winter, R.A.. "Liability insurance, joint tortfeasors and limited wealth", Internatio...		<1%
32	Internet	
meetingminds.qatar.cmu.edu		<1%
33	Internet	
www.ijraset.com		<1%
34	Publication	
"Proceedings of Third International Conference on Advances in Computer Engine...		<1%
35	Publication	
Arvind Dagur, Dhirendra Kumar Shukla, Nazarov Fayzullo Makhmadiyarovich, Ak...		<1%
36	Student papers	
Manchester Metropolitan University on 2024-10-04		<1%
37	Publication	
Saiyed Salim Sayeed, Hemant Kumar Sharma, Pramod Kumar Yadav, Brijesh Mish...		<1%
38	Student papers	
University of Basel on 2009-12-10		<1%