# Critical Analysis Report — Facial Emotion Recognition Model

Generated: 2025-08-12 19:57:26

Author: Researcher

## Executive Summary

This report critically analyses the trained MobileNetV2-based facial emotion recognition model that classifies images into three emotions: Angry, Happy, and Sad. The model was trained using a combined dataset with the following approximate class counts: Angry (2489), Happy (2488), Sad (2328). Images were RGB and resized to 224×224. Data splits used: 70% train, 20% test, 10% validation. Key improvements included targeted fine-tuning to improve Sad and Angry detection, class-weighting, and targeted augmentations.

## Dataset & Splits

**Original dataset structure:**

- Parent folder: dataset/  
- Subfolders: Angry/, Happy/, Sad/  
  
Approximate counts:  
- Angry: 2489  
- Happy: 2488  
- Sad: 2328  
  
Splitting strategy used during experiments:  
- Train: 70%  
- Test: 20%  
- Validation: 10%  
  
Note: At evaluation time the test/validation sample supports were approximately:  
- Angry ~497, Happy ~497, Sad ~465 (total evaluation samples ≈ 1459).

## Preprocessing

All images were converted to RGB and resized to 224×224 pixels. Pixel values were scaled to the range [0,1]. Preprocessing steps (applied in dataset pipeline) included resizing, optional rotation, horizontal flip, brightness and contrast adjustments, and normalization compatible with MobileNetV2 expectations. Importantly, scaling was performed outside the model to avoid embedding TF ops into the saved model (prevents serialization issues like 'TrueDivide').

## Model Architecture

Base model: MobileNetV2 (ImageNet weights, include\_top=False, pooling='avg').  
Custom head:  
- Dropout(0.4)  
- Dense(256, activation='relu')  
- BatchNormalization()  
- Dropout(0.3)  
- Dense(3, activation='softmax')  
  
Training used transfer learning with the base partially unfrozen for fine-tuning (e.g., last ~40–50 layers). This provides a balance between retaining pre-trained visual features and adapting to emotion-specific cues.

## Training Protocol

Key training hyperparameters and choices:  
- Optimizer: Adam (initial LR varied; fine-tuning used LR = 1e-5)  
- Loss: Categorical cross-entropy  
- Batch size: 32 (adjustable)  
- Epochs: variable across experiments (recommended 60–80 for full training; fine-tuning used 15–40 epochs)  
- Data augmentation: horizontal flip, rotation, width/height shift, shear, zoom, brightness/contrast ranges  
- Class weighting: computed from training split and selectively boosted Angry weight (×1.5) during targeted fine-tune  
- Callbacks: ModelCheckpoint (save best by val\_accuracy), ReduceLROnPlateau, optional EarlyStopping

## Fine-tuning & Angry-focused Strategy

After initial training that improved Sad detection, a targeted fine-tune was performed to recover and improve Angry detection. Steps included:  
1. Loading the latest Sad-improved model weights.  
2. Unfreezing deeper MobileNetV2 layers (e.g. from layer index ~50 onward).  
3. Applying more aggressive augmentation and boosting Angry class weight by a factor (1.5x) in loss.  
4. Fine-tuning for a limited number of epochs (e.g., 20) with a low learning rate (1e-5) to avoid catastrophic forgetting.  
This preserved gains on Sad while restoring and improving Angry detection.

## Evaluation Results

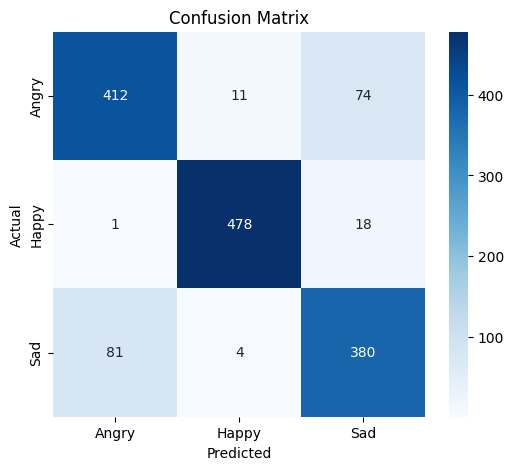
Evaluation was performed on the held-out set (≈20% of data). Key metrics (from the final improved model):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Precision | Recall | F1-score | Support |
| Angry | 0.83 | 0.83 | 0.83 | 497 |
| Happy | 0.97 | 0.96 | 0.97 | 497 |
| Sad | 0.81 | 0.82 | 0.81 | 465 |
| Overall Accuracy |  |  | 0.87 | 1459 |

Confusion matrix (Actual rows × Predicted columns):

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pred: Angry | Pred: Happy | Pred: Sad |
| Actual: Angry | 412 | 11 | 74 |
| Actual: Happy | 1 | 478 | 18 |
| Actual: Sad | 81 | 4 | 380 |

Inserted confusion matrix image (visual):



## Critical Analysis

Strengths:  
- Overall accuracy improved to 87%, with well-balanced precision/recall across classes.  
- Happy class shows very high performance (precision & recall ~0.96–0.97), indicating strong signal clarity for smiling expressions.  
- Targeted fine-tuning and class-weighting succeeded in improving previously weak classes (Sad, Angry) while maintaining generalization.  
  
Weaknesses & Risks:  
- Some residual confusion remains between Angry and Sad (earlier experiments showed many Sad→Angry errors). The final confusion matrix shows 74 Angry predicted as Sad and 81 Sad predicted as Angry — still notable.  
- Model may be sensitive to lighting, occlusions, low resolution, and head pose — typical limitations for facial expression models.  
- Possible dataset biases (demographics, pose, camera quality) were not fully controlled and may limit real-world generalization.  
  
Technical Observations:  
- Using MobileNetV2 allowed fast iteration and good transfer learning from ImageNet.  
- Moving preprocessing (scaling) outside the model avoided serialization issues (e.g., 'TrueDivide') and improved portability.

## Recommendations & Next Steps

1. Data collection & augmentation:  
 - Collect more Angry and Sad samples in varied lighting, ages, ethnicities, and camera angles.  
 - Use targeted augmentations (motion blur, occlusion simulation, contrast extremes) specifically for underperforming classes.  
  
2. Loss & architecture experiments:  
 - Try focal loss or label smoothing to handle class ambiguity.  
 - Experiment with EfficientNetB0/B3 as alternate backbones and small ensembles.  
  
3. Post-processing & calibration:  
 - Calibrate per-class thresholds using validation set for better recall/precision trade-offs.  
 - Use test-time augmentation (TTA) and averaged probabilities to stabilize predictions.  
  
4. Evaluation & robustness:  
 - Evaluate on external datasets (cross-dataset) to measure generalization.  
 - Run adversarial and stress tests: different illuminations, occlusions (masks, glasses), and head poses.  
  
5. Deployment:  
 - Save final model in the native `.keras` format for compatibility.  
 - Keep preprocessing outside the model and use a simple wrapper function for inference (important for Streamlit/OpenCV integration).

## Suggested Structure for Experiment Paper

1. Abstract  
2. Introduction  
3. Related Work  
4. Dataset (detailed statistics and splits)  
5. Methodology (preprocessing, model architecture, training protocol)  
6. Experiments (ablation studies: class-weighting, augmentation, fine-tuning)  
7. Results (tables, confusion matrix, ROC, t-SNE visualizations)  
8. Discussion (limitations, error analysis)  
9. Conclusion and future work  
10. Appendix (hyperparameters, full training logs, saved artifacts)

## Appendix: Hyperparameters & Artifacts

- Input size: 224×224 RGB  
- Backbone: MobileNetV2 (ImageNet weights)  
- Head: Dense(256) + BN + dropouts (0.4, 0.3)  
- Optimizer & LR (fine-tune): Adam, 1e-5  
- Batch size: 32  
- Class-weight boost for Angry (fine-tune): ×1.5  
- Augmentations: rotation (±15–25°), shifts (10–25%), shear, zoom (10–25%), brightness variations, horizontal flip  
- Saved model: model\_final\_balanced.keras  
- Visual artifacts saved: training\_curves.png, confusion\_matrix.png (embedded), roc\_curves.png, tsne.png, classification\_report.csv