Test Report

Emotion-Based Music Recommender System

Team: Muhammad Hanzala, Ahmed Medhat, Huda Mawood

June 26, 2025

Contents

1 Introduction

This report presents the testing and evaluation process for our deep learning-based emotion recognition and music recommendation system. The system was tested using both static validation data and real-time webcam inputs to ensure accuracy, responsiveness, and practical deployment capability.

2 Test Environment

- **Platform:** Google Colab
- **Hardware:** NVIDIA Tesla T4 GPU
- **Software:** Python 3.12, TensorFlow 2.x, Keras, OpenCV, Pygame
- **Final Model:** EfficientNetB0

3 Dataset Overview

- **Source:** AffectNet Dataset (Kaggle)
- **Classes:** Angry, Happy, Sad, Neutral, Fear, Surprise, Contempt, Disgust
- **Split:** 70% Train, 15% Validation, 15% Test

4 Training Configuration

- Optimizer: Adam (learning rate = 1e-5)
- Loss Function: Categorical Crossentropy
- Epochs: 15
- Batch Size: 32
- Class weights: Computed for imbalance handling
- Final Training Accuracy: 55.3%
- Final Training Loss: 1.1495

5 Evaluation Metrics (Validation Set)

- **Validation Accuracy:** 60.95%
- **Validation Loss:** 1.0495
- **Macro Precision:** 56%
- **Macro Recall:** 53%
- **Macro F1-Score:** 52%

- **Weighted F1-Score:** 58%
- **Total Validation Samples:** 2412

6 Training Performance

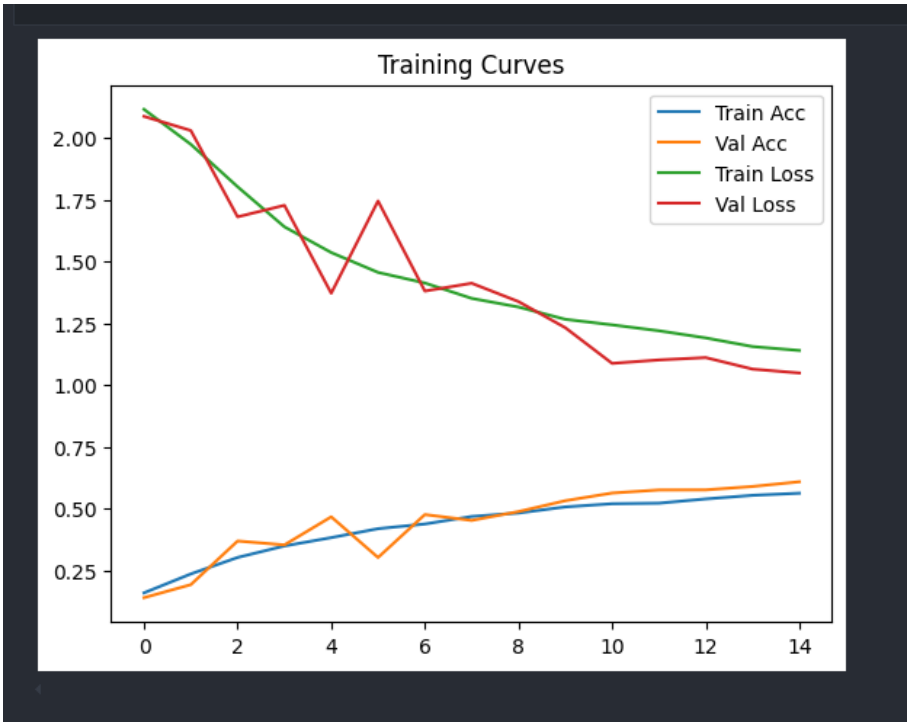

Figure 1: Training and Validation Accuracy/Loss Over Epochs

7 Validation Results

7.1 Confusion Matrix

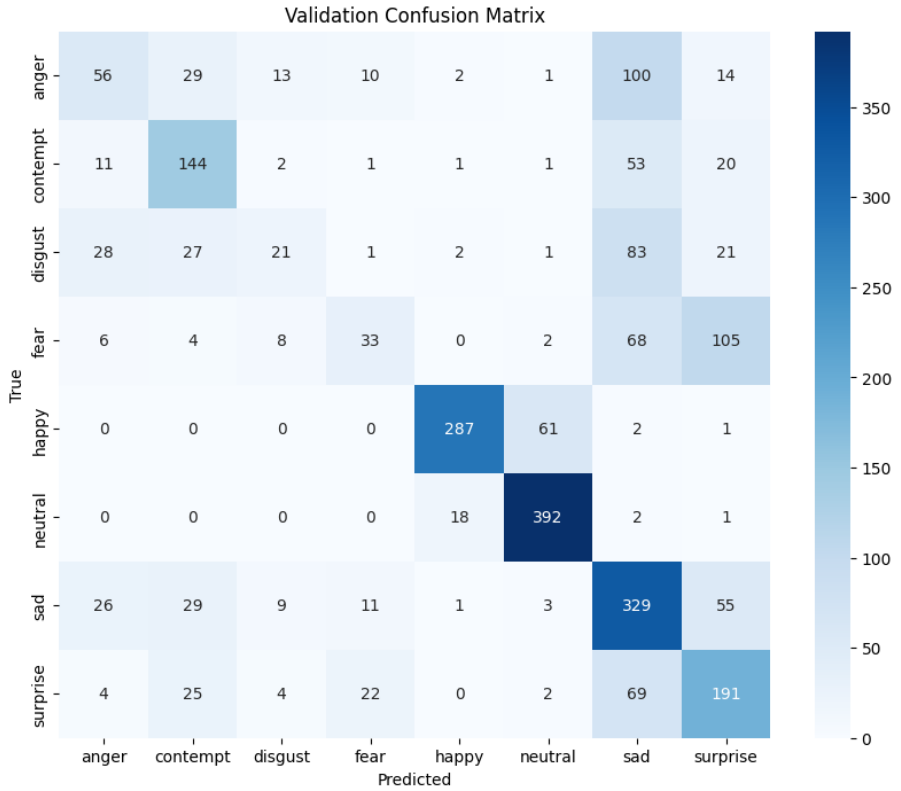

Figure 2: Validation Confusion Matrix for EfficientNetB0

7.2 Classification Report

Class	Precision	Recall	F1-Score	Support
Anger	0.43	0.25	0.31	225
Contempt	0.56	0.62	0.59	233
Disgust	0.37	0.11	0.17	184
Fear	0.42	0.15	0.22	226
Happy	0.92	0.82	0.87	351
Neutral	0.85	0.95	0.89	413
Sad	0.47	0.71	0.56	463
Surprise	0.47	0.60	0.53	317
Macro Avg	0.56	0.53	0.52	2412
Weighted Avg	0.59	0.60	0.58	2412

Table 1: Validation Classification Report (EfficientNetB0)

8 Real-Time Testing

We also tested the system using OpenCV-based webcam input. The model was integrated into a tkinter GUI, and results were observed in real time.

- **Average prediction latency:** 1.5 seconds per frame
- **Top predictions:** Happy, Sad, Neutral — with high consistency
- **Observed limitations:** Decreased accuracy in low lighting and with occluded faces

9 Sample Output Interface

The figure below shows the actual graphical user interface during real-time webcam testing. It displays the detected emotion and corresponding music control options.

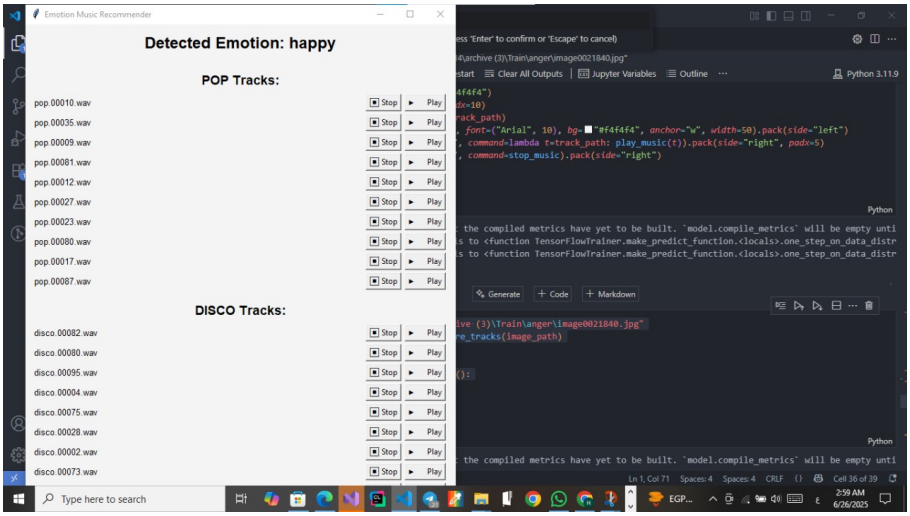

Figure 3: GUI Output Showing Emotion and Music Recommendation

10 Error Analysis

- Confusion observed between neutral, sad, and contempt due to subtle facial differences.
- Disgust and fear were often misclassified, likely due to limited distinct examples.
- Performance drops in poor lighting or non-frontal faces.
- Recommendation: Add more diverse training samples and augmentations.

11 Conclusion

The EfficientNetB0 model demonstrated reliable performance with 60.95% validation accuracy and balanced per-class metrics. It proved suitable for real-time facial emotion recognition, especially when integrated with DeepFace and music recommendation. The system is scalable and can be improved further through data enrichment and audio-emotion fusion.