Emotion-Based Music Recommender System

Using Deep Learning and Audio Integration

Team Members:
Muhammad Hanzala
Ahmed Medhat
Huda Mawood

June 26, 2025

Contents

1 Project Overview **3**

2 Model Selection Process **3**
2.1 ResNet50 3
2.2 Xception (Pretrained on FER2013) 3
2.3 ResNet (Pretrained on FER2013) 3
2.4 EfficientNetB0 Model 3
2.5 EfficientNetB3 Model 4

3 Dataset Structure **4**

4 Data Preprocessing **4**

5 Model Architecture **4**

6 Training Configuration **5**

7 Model Evaluation **5**

8 Emotion Prediction from Image **5**

9 Music Recommendation System **5**

10 Graphical User Interface (GUI) **6**

11 Real-time Webcam Detection **6**

12 Libraries Used **6**

13 Conclusion **7**

14 References **7**

1 Project Overview

This system detects emotions from human facial expressions using a deep learning model, then plays music based on the predicted emotion. The goal is to provide a mood-aware recommendation experience that connects emotion recognition with audio response.

2 Model Selection Process

We tested several deep learning models with different datasets and configurations. Below is a summary of our experimentation process.

2.1 ResNet50

- Model: ResNet50.h5
- Dataset: Custom (841 images)
- Source: Preprocessed dataset from Roboflow: https://universe.roboflow.com/hipster-pi5hv/custom-workflow-object-detection-kuaeb
- Problem: Overfitting due to limited data. Validation accuracy was poor.

2.2 Xception (Pretrained on FER2013)

- Model: Xception.h5
- Dataset: FER2013
- Source: https://www.kaggle.com/datasets/msambare/fer2013
- Issue: Poor generalization despite using data augmentation and optimizer tuning.

2.3 ResNet (Pretrained on FER2013)

- Model: resnet-fer2013.h5
- Dataset: FER2013
- Observation: Continued misclassifications, especially with similar expressions like sad and neutral.

2.4 EfficientNetB0 Model

- Model: EfficientNetB0.h5
- Dataset: AffectNet
- Source: https://www.kaggle.com/datasets/mstjebashazida/affectnet
- Performance: Most stable predictions
- Selected as final model

2.5 EfficientNetB3 Model

- Model: EfficientNetB3.h5
- Dataset: Translated and structured Roboflow dataset (Spanish to English)
- Source: https://universe.roboflow.com/reconocimiento-facial-irbfo/emociones-qbf5i
- Result: Handled more classes with fine-tuned precision, especially in expressions with subtle variations.

3 Dataset Structure

Our primary training dataset is organized by emotion class:

```
dataset
 Train
     happy
     sad
     angry
     neutral
     fear
     surprise
     contempt
     disgust
```

4 Data Preprocessing

We used ImageDataGenerator with the following techniques:

- Rescaling images to [0, 1]
- Rotation, zoom, shifts
- Horizontal flips
- Validation split: 15

5 Model Architecture

- Base: EfficientNetB0 (include_top=True)
- GlobalAveragePooling2D
- Dropout (0.5 and 0.3)
- Dense(128, ReLU)
- Final Dense layer with Softmax

6 Training Configuration

- Optimizer: Adam (learning rate = 1e-5)
- Loss: Categorical Crossentropy
- Epochs: 15
- Batch Size: 32
- Class weights computed for imbalance handling

7 Model Evaluation

- Accuracy and loss plotted per epoch
- Confusion matrix for classification analysis
- Precision, recall, and F1-score

8 Emotion Prediction from Image

- Loads and preprocesses an image
- Predicts emotion using CNN (state-of-the-art technique)
- Maps emotion to music genres

9 Music Recommendation System

Each predicted emotion is mapped to a specific music genre, based on psychological and emotional associations intended to either reinforce or regulate the listener's emotional state.

Emotion	Genres
Anger	Rock, Metal
Contempt	Classical, Jazz
Disgust	Metal, Rock
Fear	Classical
Happy	Pop, Disco
Neutral	Hip-Hop, Reggae
Sad	Blues, Jazz
Surprise	Pop, Disco

Rationale Behind Genre Classification:

- **Anger:** Rock and Metal offer strong beats and intense energy, helping users release tension and channel their emotions constructively.
- **Contempt:** Classical and Jazz promote inner calm and sophistication, encouraging thoughtful reflection and emotional grounding.
- **Disgust:** Metal and Rock reflect the raw intensity of disgust, allowing listeners to process these feelings through heavy rhythm and bold sounds.
- **Fear:** Classical music provides a calming and structured environment that may help reduce anxiety and soothe fearful emotions.
- **Happy:** Pop and Disco enhance happiness with upbeat tempos, lively rhythms, and cheerful melodies that keep the mood elevated.
- **Neutral:** Hip-Hop and Reggae add light engagement and rhythm, helping maintain a balanced mood while subtly energizing the listener.
- **Sad:** Blues and Jazz are emotionally expressive genres that help in understanding, validating, and releasing feelings of sadness.
- **Surprise:** Pop and Disco match the high arousal state of surprise with vibrant and energetic musical experiences.

10 Graphical User Interface (GUI)

Built using tkinter, the GUI:

- Displays predicted emotion
- Shows recommended music tracks
- Provides play/stop buttons for each track

11 Real-time Webcam Detection

- Captures image using OpenCV
- Feeds image directly into prediction pipeline
- Displays GUI with results

12 Libraries Used

- TensorFlow, Keras
- OpenCV
- Numpy, Matplotlib, Seaborn
- Tkinter
- Pygame

13 Conclusion

We successfully built an emotion-aware system that connects facial expression recognition with personalized music. Despite challenges in expression similarity and data limitations, the system achieved promising results with EfficientNetB0.

After extensive research and comparison, we concluded that **DeepFace** is one of the most effective tools for facial emotion recognition. It integrates multiple pretrained models (e.g., VGG-Face, Facenet, OpenFace) and supports real-time emotion analysis with high-level abstraction. DeepFace's `analyze()` function provided accurate predictions and required minimal configuration, which made it highly suitable for integration with our application.

- **Ease of use:** Simple API and fast integration.
- **Real-time capability:** Supports webcam and live video inputs.
- **Multiple features:** Emotion, age, gender, and race prediction.
- **Pretrained models:** Enables robust performance without requiring large custom datasets.

As a result, DeepFace was integrated in our system to enhance accuracy, reduce training requirements, and support practical real-time usage.

14 References

- **DeepFace Library:** https://github.com/serengil/deepface
- **EfficientNetB0 Dataset (AffectNet via Kaggle):** https://www.kaggle.com/datasets/mstjebashazida/affectnet
- **FER2013 Dataset (Used with Xception and ResNet):** https://www.kaggle.com/datasets/msambare/fer2013
- **Roboflow Dataset for ResNet50:** https://universe.roboflow.com/hipster-pi5hv/custom-workflow-object-detection-kuaeb
- **Roboflow Dataset for EfficientNetB3 (Translated):** https://universe.roboflow.com/reconocimiento-facial-irbfo/emociones-qbf5i
- **EfficientNet Paper:** https://arxiv.org/abs/1905.11946
- **Xception Paper:** https://arxiv.org/abs/1610.02357
- **OpenCV Haar Cascades:** https://github.com/opencv/opencv/tree/master/data/haarcascades
- **TensorFlow:** https://www.tensorflow.org/
- **Keras:** https://keras.io/
- **Pygame Documentation:** https://www.pygame.org/docs/