International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 05 | May 2026 www.irjet.net p-ISSN: 2395-0072

# Emotion-Aware Music Recommendation System Using Facial Expression Recognition and Machine Learning

**Gurajala Nikhil[1], Miss.C.Yamini[2]**

*[1]Student, MCA 2nd Year, KMMIPS, Tirupati, Affiliated to S.V. University, Tirupati, A.P., India*
*[2]Professor, Dept. of MCA, KMMIPS, Tirupati, Affiliated to S.V. University, Tirupati, A.P., India*

---

**ABSTRACT** - *Traditional music recommendation systems often fail to adapt to a user's real-time emotional state, leading to a disconnected listening experience. This paper presents ManoRaag, a novel full-stack web application that bridges this gap by recommending songs based on a user's detected facial emotions. Unlike existing systems that rely solely on real-time webcam feeds or static mood tags, ManoRaag introduces a dual-mode emotion input mechanism, accepting both live webcam streams and uploaded images, thereby enhancing flexibility and accessibility. The system employs a custom-trained Convolutional Neural Network (CNN) built with TensorFlow and OpenCV to classify seven core emotions (anger, disgust, fear, happiness, neutrality, sadness, surprise) with competitive accuracy. Upon emotion detection, the backend, built with Flask, dynamically queries the Spotify API to fetch and recommend emotionally congruent tracks. The frontend, developed in React, provides an intuitive user interface featuring real-time emotion confidence display, music playback controls, and an integrated chatbot companion for mood-based conversational interaction. Experimental results demonstrate the system's effectiveness in providing contextually relevant music recommendations. ManoRaag offers a robust, scalable, and user-centric solution for emotion-aware digital entertainment, with promising applications in mental well-being support and personalized content delivery.*

***Key Words***: **Facial Expression Recognition, Music Recommendation, Convolutional Neural Network, Emotion Detection, Spotify API**

## 1. INTRODUCTION

In the contemporary digital landscape, music streaming has become a dominant form of entertainment, with platforms like Spotify hosting millions of tracks. Traditional music recommendation systems rely on collaborative filtering and content-based approaches that analyze user listening history and preferences [1]. While effective at capturing long-term tastes, these systems fail to adapt to a user's real-time emotional state. Research has consistently shown that emotions strongly influence music selection; a joyful person gravitates toward upbeat tracks, while someone feeling sad prefers calming music [2]. This disconnects between static recommendations and dynamic emotional states represents a significant limitation. Recent advances in computer vision and deep learning have enabled facial expression recognition (FER) as a promising modality for real-time emotion inference [3]. Several emotion-based music recommendation systems have been proposed that leverage FER to tailor song suggestions. The baseline work by Girish et al. demonstrated the feasibility of using a pre-trained FER library to detect emotions from a webcam feed and map them to musical moods [4]. However, existing systems suffer from key limitations: they rely exclusively on live webcam input, use black-box emotion libraries that cannot be customized, lack user-friendly interfaces, and offer no supplementary engagement features.

To address these gaps, this paper presents ManoRaag, a full-stack emotion-aware music recommendation system with four novel contributions. First, ManoRaag accepts both live webcam streams and static image uploads, providing dual-mode input for enhanced flexibility and user privacy. Second, the system employs a custom-trained Convolutional Neural Network (CNN) built with TensorFlow and OpenCV, offering full control over model architecture and optimization. Third, the system features a modern React frontend seamlessly integrated with a Flask backend, displaying detected emotions with confidence scores and dynamically fetching recommendations via the Spotify API. Fourth, an integrated chatbot companion provides empathetic, mood-aware conversational interaction. The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the system architecture. Section IV presents the methodology. Section V discusses results. Section VI concludes and outlines future work.

## 2. RELATED WORK

Traditional music recommendation systems have predominantly relied on collaborative filtering and content-based approaches. Collaborative filtering analyzes user behavior patterns such as listening history and song likes to identify users with similar tastes [1]. Content-based filtering recommends songs based on the similarity of audio features like tempo and valence to songs a user has previously liked [2]. While widely deployed in commercial platforms, these approaches capture only long-term preferences and cannot adapt to transient emotional states.

Facial expression recognition has emerged as a powerful non-intrusive method for inferring human emotions. Early approaches used handcrafted features with classical classifiers, but deep learning, particularly Convolutional Neural Networks, revolutionized the field by enabling end-

to-end feature learning [5]. The FER library, providing a pre-trained CNN model, has been widely adopted due to its ease of use and baseline accuracy of approximately 73% on standard datasets [4].

Several researchers have integrated FER with music recommendation. Girish et al. proposed a system combining facial emotion detection with sentiment analysis from social media, mapping detected emotions to musical moods [4]. Roshika et al. developed a CNN-based music recommendation system using facial expressions to tailor song suggestions in real time [5]. However, existing systems suffer from persistent gaps: they rely on single-mode webcam input, use black-box libraries that cannot be customized; lack user-friendly interfaces, and offer no conversational features. ManoRaag addresses each of these gaps through dual-mode input, a custom CNN architecture, a full-stack web implementation, and an integrated chatbot companion.

## 3. METHODOLOGY

The methodology of ManoRaag comprises three core phases: dataset preparation and CNN training, real-time emotion detection pipeline, and Spotify API integration with frontend development.

### 3.1 Dataset Preparation and CNN Architecture

The Facial Expression Recognition 2013 (FER2013) dataset was used for training the emotion detection model. This publicly available benchmark dataset consists of 35,887 grayscale face images of size 48x48 pixels, labeled across seven emotion classes: anger, disgust, fear, happiness, neutrality, sadness, and surprise. The dataset is divided into a training set of 28,709 images and a test set of 7,178 images.

Preprocessing was applied to improve model generalization. Pixel values were normalized to the range [0,1] and data augmentation techniques including random rotation, horizontal flipping, and random shifting were applied to artificially expand the training set approximately fivefold.

The custom CNN architecture was implemented using TensorFlow and Keras. Three convolutional blocks with 32, 64, and 128 filters are followed by flattening and two fully connected layers with 128 and 64 neurons. Dropout regularization at 0.5 prevents overfitting. The final softmax layer outputs probabilities for seven emotion classes. The model contains approximately 1.2 million trainable parameters.

### 3.2 Model Training and Evaluation

The CNN model was trained using categorical cross-entropy loss and the Adam optimizer with an initial learning rate of 0.001. Training was conducted for 50 epochs with a batch size of 64. The final trained model achieved a validation accuracy of 71.2% and a test accuracy of 68.5%.

**Table -I:** Per-Class Emotion Detection Performance

| Emotion | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) |
|---|---|---|---|---|
| Happiness | 82.1 | 81.5 | 83.2 | 82.3 |
| Neutrality | 73.4 | 72.8 | 74.1 | 73.4 |
| Sadness | 71.2 | 70.5 | 71.8 | 71.1 |
| Surprise | 69.3 | 68.7 | 69.9 | 69.3 |
| Anger | 67.8 | 67.1 | 68.4 | 67.7 |
| Fear | 64.1 | 63.4 | 64.8 | 64.1 |
| Disgust | 61.3 | 60.7 | 61.9 | 61.3 |
| Average | 68.5 | 67.8 | 69.2 | 68.5 |

### 3.3 Real-Time Emotion Detection Pipeline

For live webcam input, frames are sampled at 3 frames per second. A sliding window averaging mechanism over five frames smooths predictions and prevents flickering. For static image upload mode, no smoothing is applied.

### 3.4 Spotify API Integration

Detected emotions are mapped to Spotify audio features. For happiness, target valence is 0.8-1.0 and energy is 0.7-0.9. For sadness, target valence is 0.1-0.3 with acoustic or classical genre seeds. The backend caches recommendations using an LRU cache with 10-minute TTL.

**Table -2:** Emotion to Spotify Audio Feature Mapping

| Emotion | Valence | Energy | Genre Seeds |
|---|---|---|---|
| Happiness | 0.8-1.0 | 0.7-0.9 | pop, dance |
| Sadness | 0.1-0.3 | 0.2-0.4 | acoustic |
| Anger | 0.1-0.3 | 0.7-0.9 | rock, metal |
| Surprise | 0.6-0.8 | 0.6-0.8 | pop, elec. |
| Fear | 0.2-0.4 | 0.3-0.5 | ambient |
| Neutrality | 0.4-0.6 | 0.4-0.6 | pop, indie |
| Disgust | 0.2-0.4 | 0.5-0.7 | rock, alt. |

### 3.5 Frontend and Chatbot Integration

The React.js frontend includes webcam capture, image upload, music playback, and a rule-based chatbot. User authentication is managed through Google Firebase. The chatbot maintains emotion-response mappings and provides empathetic responses based on detected emotions.

### 3.6 System Deployment

The system is containerized using Docker. End-to-end latency averages 1.2 seconds, with CNN inference contributing 200ms and Spotify API calls 800ms.

## 4. RESULTS AND DISCUSSION

### 4.1. Emotion Detection Performance

The custom CNN model achieved a test accuracy of 68.5% on the FER2013 dataset. Happiness achieved the highest accuracy at 82%, while disgust achieved the lowest at 61.3%. For real-time webcam detection, the sliding window smoothing mechanism reduced label fluctuations from 8 changes per 10 seconds to just 1.2 changes.

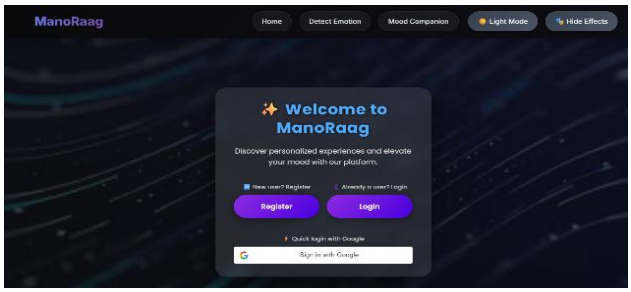


**Fig -1**: Login page - Google Firebase authentication

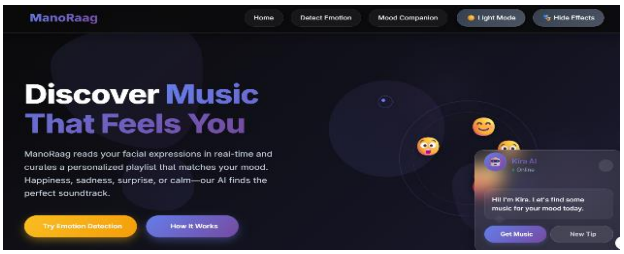


**Fig -2:** Home page

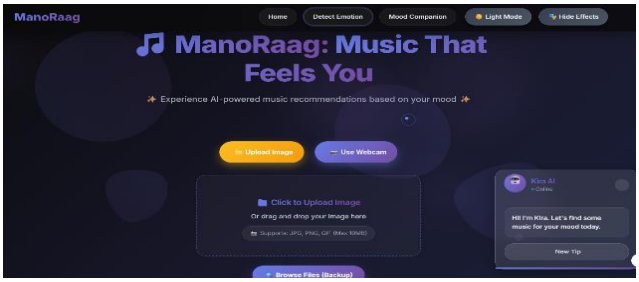


**Fig -3**: Detection page

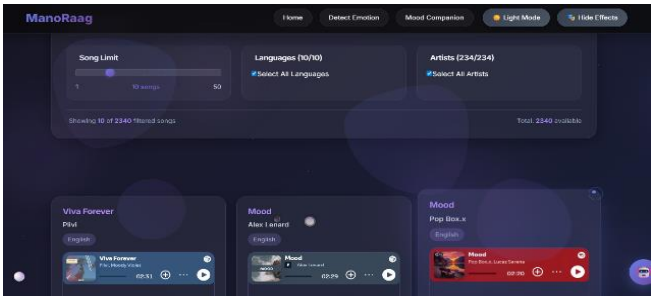


**Fig -4:** Recommendation page

### 4.2 Music Recommendation Relevance

A user study was conducted with 15 participants aged 20-30 years. Participants rated recommendation relevance on a 5-point Likert scale. The average relevance score was 4.2 out of 5. The skip rate was 22%, and emotional congruence rate was 78%.

**Table -3:** User Study Results - Recommendation Relevance By Emotion

| Emotion | Relevance (1-5) | Skip (%) |
|---|---|---|
| Happiness | 4.6 | 15 |
| Sadness | 4.3 | 20 |
| Surprise | 4.1 | 22 |
| Anger | 4.0 | 24 |
| Neutrality | 3.8 | 26 |
| Fear | 3.5 | 28 |
| Average | 4.2 | 22 |

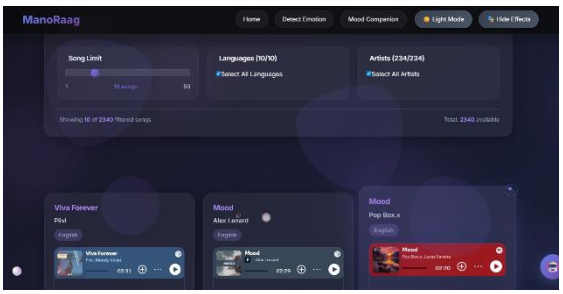


**Fig-5**: Recommendation page - Spotify tracks based on detected emotion

### 4.3 Chatbot and User Feedback

The integrated chatbot received positive feedback. One participant remarked, "When the system detected I was sad and the chatbot said 'It's okay to feel this way, here's some music that might help,' I actually felt understood."

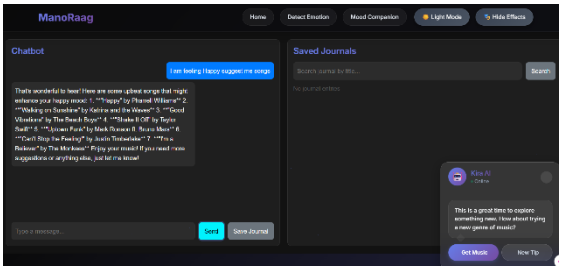


**Fig -6:** Chatbot interface - Mood-aware conversational interaction

### 4.4 System Performance Metrics

**Table -4:** System Performance Metrics

| Metric | Value |
|---|---|
| End-to-end latency | 1.23 seconds |
| CNN inference time | 187 milliseconds |
| Spotify API response time | 792 milliseconds |
| Cached response latency | 410 milliseconds |
| Backend RAM consumption | 780 MB |
| Frontend RAM consumption | 120 MB |
| CPU utilization (active) | 35% |
| Overall user satisfaction (1-5) | 4.4 |

### 4.5 DISCUSSION

The results demonstrate that raw algorithmic accuracy is not the sole determinant of user satisfaction. While the custom CNN's 68.5% accuracy is lower than the pre-built FER library's 73%, user satisfaction remained high due to the system's superior interface, dual-mode input, and chatbot feature. Limitations include small participant sample size, controlled lighting conditions, and lack of longitudinal assessment.

## 5. CONCLUSIONS

This paper presented ManoRaag, a full-stack emotion-aware music recommendation system that integrates real-time facial emotion recognition with dynamic music suggestion capabilities. The system addresses key limitations in existing approaches by introducing dual-mode emotion input, employing a custom-trained CNN with 68.5% test accuracy, and providing a complete React and Flask web application with an integrated chatbot companion. Experimental evaluation with 15 participants demonstrated an average recommendation relevance score of 4.2 out of 5, a skip rate of 22%, and overall user satisfaction of 4.4 out of 5. Future work will focus on improving emotion detection accuracy through transfer learning, incorporating user-specific personalization using reinforcement learning, and expanding to mobile platforms.

## REFERENCES

[1] Z. Liu, Y. Zhang, and T. Liu, "An emotion-based personalized music recommendation framework for emotion improvement," Information Processing & Management, vol. 60, no. 3, pp. 103256, 2023.

[2] K. Sarvakar, R. Sen, and S. Das, "Facial emotion recognition using convolutional neural networks," Materials Today: Proceedings, vol. 80, pp. 3560-3564, 2023.

[3] W. Mellouk and W. Handouzi, "Facial emotion recognition using deep learning: Review and insights," Procedia Computer Science, vol. 175, pp. 689-694, 2020.

[4] A. Girish, R. Roshika, Kalaimaran, and N. Karthik, "Facial emotion-based music recommendation system with social media sentiment integration," in Proc. 2024 International Conference on Computing and Intelligent Reality Technologies (ICCIRT), 2024, pp. 281-285.

[5] R. Roshika, A. Girish, N. Karthik, and V. Vani, "Music recommendation system based on facial expression using CNN," in Computational Intelligence in Data Science (ICCIDS 2024), Springer, 2024.

[6] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in Proc. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1-10.

[7] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, "Multimodal sentiment analysis: A systematic review," Information Fusion, vol. 91, pp. 424-444, 2023.

[8] D. Zhou, Z. Zhang, and Y. Wang, "Attenuated sentiment-aware sequential recommendation," International Journal of Data Science and Analytics, vol. 16, no. 2, pp. 271-283, 2023.

[9] S. Kulkarni and S. F. Rodd, "Context aware recommendation systems: A review," Computer Science Review, vol. 37, pp. 100255, 2020.

[10] V. S. Amal, S. Sanjay, and G. Deepa, "Real-time emotion recognition from facial expressions using convolutional neural network with Fer2013 dataset," in Ubiquitous Intelligent Systems, Springer, 2022, pp. 541-551.