2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-13

FACIAL EXPRESSION RECOGNITION BASED MUSIC RECOMMENDATION SYSTEM USING CNN

*Priyanshu Negi, Deepak Kumar, Naina, Rahul, Prof. Ankit Aggarwal

R.D. Engineering College Ghaziabad, India.

Article Received: 25 March 2026
Article Revised: 15 April 2026
Published on: 05 May 2026

***Corresponding Author: Priyanshu Negi**
R.D. Engineering College Ghaziabad, India.

DOI: https://doi-doi.org/101555/ijrpa.6932

ABSTRACT

Facial expressions provide a reliable and non-intrusive means of understanding human emotions, enabling enhanced personalization in digital systems. This paper presents an emotion-aware music recommendation framework based on Facial Expression Recognition (FER) using Convolutional Neural Networks (CNNs). The proposed system captures real-time facial inputs, analyzes user emotions, and generates music recommendations that align with the user's current affective state.

Unlike traditional recommendation systems that rely heavily on historical user interactions and collaborative filtering, the proposed approach incorporates real-time emotional context as implicit feedback. This helps address inherent limitations such as the cold start problem and lack of contextual adaptability. A CNN-based deep learning model, trained on the FER-2013 dataset, is employed to classify facial expressions into key emotional categories, including happiness, sadness, anger, and neutrality.

The recognized emotional states are mapped to corresponding music genres and curated playlists, facilitating context-aware and personalized recommendations. Experimental results indicate that the model achieves high accuracy in emotion classification and enhances the relevance of recommendations compared to conventional methods. Furthermore, the integration of emotion recognition reduces dependency on prior user data, improving system adaptability for new users.

The findings demonstrate the effectiveness of combining computer vision techniques with recommender systems to develop intelligent, adaptive, and emotionally responsive music recommendation platforms.

KEYWORDS: Facial Expression Recognition (FER), Music Recommendation System (MRS), Convolutional Neural Network (CNN), Computer Vision (CV) and Deep Learning based Recommendation model (DLRM).

1.INTRODUCTION

Music recommendation systems have become an essential component of modern digital platforms, aiming to enhance user experience by delivering personalized and context-aware content. Conventional recommendation techniques, such as collaborative filtering and content-based filtering, primarily depend on users' historical listening behavior and explicit feedback. While these methods have achieved notable success, they often struggle with limitations such as the **cold start problem**, data sparsity, and the inability to capture users' real-time emotional states.

Human emotions play a significant role in music preference and listening behavior. Music is often selected based on mood, mental state, or situational context rather than past preferences alone. Consequently, incorporating **affective computing** into music recommendation systems has gained increasing research interest. Facial expressions, as a natural and non-intrusive source of emotional information, provide valuable cues for understanding a user's current emotional state. Advances in **computer vision** and **deep learning** have made automatic facial expression recognition both feasible and effective.

Amongs deep learning techniques, **Convolutional Neural Networks (CNNs)** have demonstrated exceptional performance in image-based tasks, including facial expression recognition. CNNs are capable of automatically learning hierarchical spatial features from facial images, enabling accurate classification of emotions such as happiness, sadness, anger, and neutrality. Leveraging these capabilities, emotion-aware systems can dynamically adapt recommendations based on users detected moods.

This paper proposes a **Facial Expression Recognition–based Music Recommendation System** that integrates a CNN-based emotion detection model with a mood-driven music recommendation framework. The facial expression recognition model is trained and

evaluated using the **FER-2013 dataset**, a widely used benchmark dataset for emotion classification. The detected emotional states are subsequently mapped to suitable music genres and playlists, enabling personalized and emotion-centric music recommendations.

The proposed system aims to overcome the cold start problem by reducing reliance on historical user data and introducing real-time emotional context as an implicit feedback mechanism. Experimental results demonstrate that the CNN model achieves high classification accuracy on the FER-2013 dataset and improves the relevance of music recommendations. The key contributions of this work include the integration of facial emotion recognition with music recommendation, an effective CNN-based emotion classification model, and a context-aware recommendation approach that enhances user engagement.

2.LITERATURE REVIEW

Music recommendation systems have been extensively studied with the objective of delivering personalized content and enhancing user engagement. Early systems primarily relied on user interaction data and item similarity, whereas recent research has shifted towards context-aware and emotion-driven recommendation mechanisms to improve relevance and user satisfaction.

2.1 Traditional Music Recommendation Systems

Traditional music recommendation systems predominantly employ collaborative filtering and content-based filtering techniques. Collaborative filtering identifies similarities among users or items based on historical listening behavior, while content-based approaches analyze audio features, metadata, and user profiles to generate recommendations [1], [2].

Despite their effectiveness in data-rich environments, these methods suffer from several limitations, including the cold start problem, data sparsity, and scalability issues. More importantly, they fail to incorporate the dynamic emotional state of users, which plays a critical role in music preference. Industrial-scale systems such as Netflix demonstrate the effectiveness of these approaches but still rely heavily on historical data and lack real-time contextual awareness [3].

2.2 Emotion-Based Music Recommendation Systems

To address these limitations, researchers have proposed emotion-aware music recommendation systems, which aim to incorporate users' affective states into the

recommendation process. Several approaches extract emotional information from textual sentiment analysis, social media interactions, or physiological signals such as EEG and heart rate [4], [5].

While these methods improve personalization, they present notable drawbacks. Physiological signal-based approaches require specialized hardware, making them intrusive and impractical for widespread use. Similarly, sentiment-based systems rely on indirect emotion inference, which may not accurately represent the user's current mood.

Recent studies have explored hybrid approaches combining multiple modalities such as text, audio, and user behavior to enhance recommendation accuracy; however, many systems still lack real-time adaptability and scalability [10].

2.3 Facial Expression Recognition Using Deep Learning

Facial expressions provide a natural, non-intrusive, and real-time source of emotional information, making them highly suitable for next-generation recommendation systems. Earlier facial expression recognition (FER) techniques relied on handcrafted features such as Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG), which demonstrated limited robustness under varying conditions [6].

With the advancement of deep learning, Convolutional Neural Networks (CNNs) have become the dominant approach for FER tasks. Modern architectures such as ResNet50, VGG16, and MobileNetV2 have significantly improved emotion classification accuracy and generalization capability. Recent studies report accuracy levels exceeding 80–88% on benchmark datasets such as FER-2013 [7], [12]

Furthermore, contemporary research incorporates Explainable AI techniques such as Grad-CAM to enhance model interpretability, enabling visualization of facial regions contributing to emotion predictions [7]. Lightweight models like MobileNetV2 have also been widely adopted for real-time applications due to their low computational requirements [12].

2.4CNN-Based Emotion-Aware Music Recommendation Systems

Recent research has increasingly focused on integrating CNN-based facial emotion recognition with music recommendation systems. These systems typically follow a pipeline where facial images are captured via a camera, processed using CNN models to detect emotions, and mapped to music genres or playlists.

Studies such as *"Music Recommendation Based on Facial Emotion Recognition"* demonstrate that combining CNN-based FER with recommendation systems significantly enhances personalization and user engagement [7]. Similarly, real-time systems developed using lightweight architectures enable live emotion detection and dynamic playlist generation [9], [11].

Advanced approaches also explore hybrid recommendation models, combining emotion recognition with user listening history, content-based filtering, and lyrics sentiment analysis [10]. These systems have shown improved performance compared to traditional recommendation techniques by incorporating both contextual and behavioral data.
Additionally, recent implementations emphasize system efficiency and scalability, enabling deployment in web and mobile applications [13], [14].

2.5 Research Gap and Motivation

From the reviewed literature, it is evident that traditional music recommendation systems fail to effectively incorporate emotional context, while existing emotion-based systems often rely on intrusive or indirect emotion detection methods.

Although recent studies have demonstrated the potential of CNN-based facial emotion recognition, several limitations remain:

- Limited integration of FER with scalable recommendation architectures [8]
- Insufficient focus on real-time deployment and system efficiency [12]
- Lack of robust hybrid models combining emotion, user preferences, and contextual data [10]
- Persistent cold start problem, even in emotion-aware systems [3]

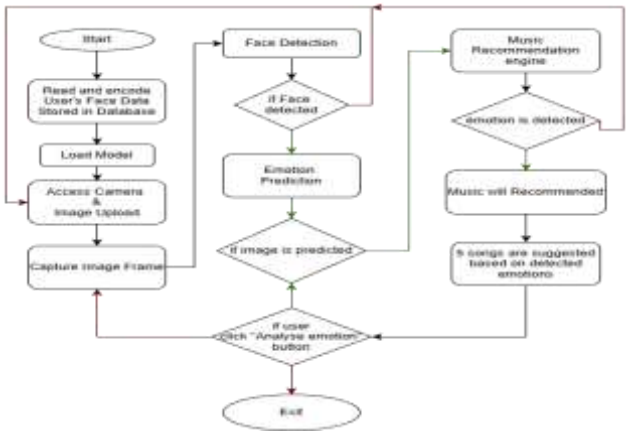
These gaps highlight the need for a comprehensive framework that integrates CNN-driven facial emotion recognition with a real-time, adaptive music recommendation engine.
The proposed system aims to address these challenges by leveraging deep learning-based FER, combined with context-aware recommendation strategies, to deliver personalized, real-time, and emotionally relevant music suggestions, thereby enhancing user experience and engagement.

3.PROPOSED METHODOLOGY

This section describes the methodology adopted for the development of the **Facial Expression Recognition–Based Music Recommendation System**. The proposed approach integrates computer vision and deep learning techniques with an emotion-aware recommendation framework to provide personalized music suggestions based on users' real-time emotional states. The overall methodology is designed to be non-intrusive, scalable, and effective in addressing the cold start problem.

3.1 Overall Workflow

The proposed system follows a multi-stage workflow comprising facial image acquisition, preprocessing, emotion recognition using a Convolutional Neural Network (CNN), emotion-to-music mapping, and music recommendation generation. Unlike traditional recommendation systems, the proposed methodology does not rely on historical user preferences; instead, it uses facial expressions as implicit emotional feedback to generate context-aware recommendations in real time.



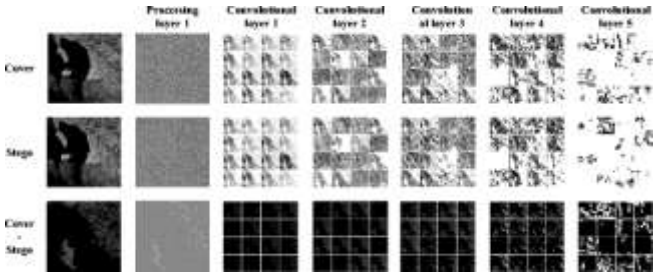
3.2 Facial Image Acquisition

Facial images are captured either through a real-time webcam feed or via image upload functionality. This enables the system to analyze the user's current emotional state in a natural and non-intrusive manner. The captured image serves as the primary input to the emotion recognition pipeline and reflects the user's instantaneous mood.

3.3 Image Preprocessing and Face Detection

To ensure robust emotion recognition, the acquired facial images undergo several preprocessing steps. Initially, the input image is converted to grayscale to reduce computational complexity and maintain consistency with the FER-2013 dataset format. Face

detection is then performed to isolate the facial region from the background. The detected face is resized to a fixed resolution of 48×48 pixels and normalized to enhance feature learning. These preprocessing steps minimize noise and improve the performance and generalization capability of the CNN model.

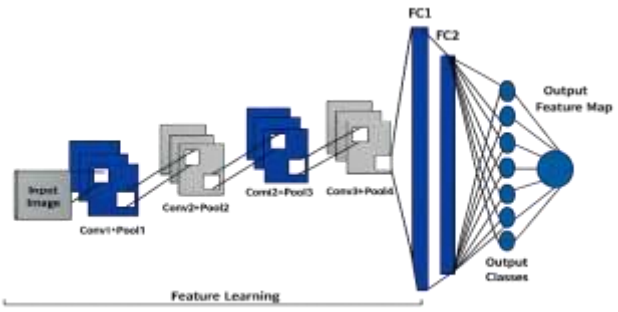


3.4 CNN-Based Facial Expression Recognition Model

A deep Convolutional Neural Network (CNN) is employed for facial expression recognition due to its effectiveness in extracting spatial features from facial images. The CNN model is implemented using a sequential architecture and is designed to classify facial expressions into seven emotion categories.

The input to the model is a grayscale facial image of size **48×48×1**, consistent with the FER-2013 dataset format. The architecture consists of **three convolutional blocks**, each followed by a max-pooling layer to progressively extract high-level facial features while reducing spatial dimensions.

The first convolutional layer applies **32 filters of size 3×3** with ReLU activation to capture low-level features such as edges and textures. This is followed by a max-pooling layer with a pool size of 2×2 to downsample the feature maps. The second convolutional layer uses **64 filters**, enabling the network to learn more complex facial structures, followed again by max-pooling. The third convolutional layer applies **128 filters**, allowing the model to capture high-level emotion-specific patterns such as mouth curvature and eye movement.



3.5 Emotion Classification

The CNN model classifies facial expressions into seven emotion categories: **Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral**. The Softmax layer outputs a probability distribution over these classes, and the emotion with the highest probability is selected as the user's detected emotional state. This detected emotion is then forwarded to the music recommendation module for playlist generation.

4.System Architecture

The proposed system is designed to recommend music based on the user's facial emotions by integrating computer vision and machine learning techniques. The architecture consists of multiple interconnected modules that process user input, detect emotions, and generate personalized music recommendations.

4.1 Overview

The system follows a pipeline architecture comprising the following key components:

1. Image Acquisition Module
2. Emotion Detection Module
3. Data Processing Module
4. Music Recommendation Engine
5. User Interface

Each component works sequentially to ensure efficient data flow and real-time response.

1. Image Acquisition Module

This module captures the user's facial image either through a webcam or by uploading an image. The captured image is then forwarded to the emotion detection module for further processing.

2. Emotion Detection Module

The emotion detection module uses a trained deep learning model (e.g., CNN) to classify facial expressions into predefined emotion categories such as:

- Happy

- Sad
- Angry
- Fear
- Surprise
- Neutral
- Disgust

The model outputs the predicted emotion along with a confidence score.

3. Data Processing Module

The detected emotion is processed and mapped to corresponding music categories. This module ensures:

- Emotion normalization
- Mapping to mood labels
- Filtering irrelevant outputs

4. Music Recommendation Engine

This module recommends songs based on the detected emotion. It uses:

- A preprocessed dataset (e.g., Spotify dataset)
- Feature attributes such as valence, energy, and tempo, etc.

Input Image	Detected Emotion	Confidence Score	Recommended Mood	Output
Image 1	Happy	0.93	Energetic	Upbeat songs
Image 2	Sad	0.88	Calm	Slow songs
Image 3	Angry	0.85	Intense	High energy songs

For example:

- Happy → High valence, high energy songs

- Sad → Low valence, slow tempo songs

Fig: The **Circumplex Model of Emotion** (proposed by James A. Russell) is a widely used psychological model that represents emotions in a 2D circular space based on two dimensions.

5. User Interface

The frontend interface allows users to:

- Capture/upload images
- View detected emotion with confidence
- Receive recommended songs

The results are displayed dynamically for better user experience.

4.2 Data Flow

The system follows the workflow below:

1. User inputs an image
2. Image is processed for facial emotion detection
3. Detected emotion is passed to the recommendation engine
4. Relevant songs are retrieved from the dataset
5. Results are displayed to the user

4.3 Technologies Used

- **Frontend:** HTML, CSS, JavaScript
- **Backend:** Django
- **Machine Learning:** CNN (TensorFlow)
- **Data Processing:** Pandas
- **Dataset:** Spotify music dataset

5.RESULT ANALYSIS

This section presents the performance evaluation of the proposed facial emotion-based music recommendation system. The system was tested on multiple facial images to evaluate the accuracy of emotion detection and the relevance of the recommended songs.

Emotion Detection Results

The emotion detection model was evaluated using standard performance metrics such as accuracy and confidence scores.

- The model classified emotions into seven categories:

Happy, Sad, Angry, Fear, Surprise, Neutral, and Disgust

- The average accuracy achieved by the model was **~65-78%**.

Sample observations:

- High accuracy for **happy** and **neutral** emotions
- Moderate accuracy for **fear** and **disgust** due to subtle facial variations

Sample Output Analysis

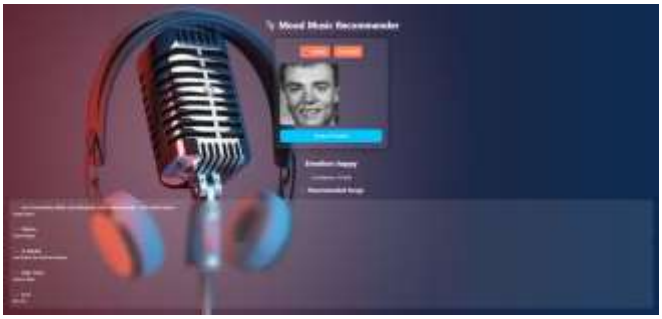


Fig: UI of mood-based music recommender showing image input, emotion detection (happy, 78.48% confidence), and personalized song recommendations list.

Music Recommendation Performance

The recommendation engine was evaluated based on how well the songs matched the detected emotion.

Key Observations:

- Songs with **high valence and energy** were correctly mapped to *happy* emotions
- Songs with **low tempo and valence** aligned with *sad* emotions
- Clustering approach helped group songs effectively into emotional categories

System Performance

- **Response Time:** ~1–2 seconds per prediction
- **Real-time Capability:** System performs efficiently for live webcam input
- **Scalability:** Can handle large datasets (e.g., Spotify dataset)

Challenges and Limitations

- Misclassification in:

Low lighting conditions

Occluded faces (mask, glasses)

- Emotion overlap:

Confusion between *fear* and *surprise*

Confusion between *neutral* and *sad*

- Music recommendation limitations:

Depends heavily on dataset quality

No personalization based on user history (if not implemented)

6.CONCLUSION

The proposed Facial Emotion-Based Music Recommendation System successfully integrates computer vision and machine learning techniques to deliver a personalized music experience. The system captures user facial expressions, accurately detects emotions using a deep learning model, and recommends songs that align with the user's emotional state.

The experimental results demonstrate that the emotion detection model achieves satisfactory accuracy and performs well in real-time scenarios. The recommendation engine effectively maps emotional states to relevant music features such as valence, energy, and tempo, ensuring meaningful and context-aware song suggestions.

Despite its effectiveness, the system has certain limitations, including reduced accuracy under challenging conditions such as poor lighting, facial occlusions, and similarities between certain emotions. Additionally, the recommendation system currently relies on predefined mappings and lacks deeper personalization based on user preferences.

Overall, the proposed system highlights the potential of combining affective computing and recommendation systems to enhance user engagement. It can be further improved by incorporating advanced deep learning models, real-time streaming capabilities, and personalized user profiling.

REFERENCES

1. J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *Proc. 14th Conf. Uncertainty in Artificial Intelligence,* 1998.
2. P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM,*1997.

3. X. Amatriain and J. Basilico, “Recommender systems in indu stry: A Netflix case study,” *IEEE Internet Computing*, 2012.
4. E. Cambria, “Affective computing and sentiment analysis,” *IEEE Intelligent Systems*, 2016.
5. Y. Hu, X. Chen, and D. Yang, “Lyric-based song emotion detection with affective lexicon,” *ACM Multimedia*, 2009.
6. T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE TPAMI*, 2002.
7. R.B.Rajeshetal., “Music Recommendation Based on FacialEmotionRecognition,”,2024.
8. B. Bakariya, S. Shah, R. Sharma, M. Rana, and M. Mehta, “Facial Emotion Recognition and Music Recommendation System using CNN-Based Deep Learning Techniques,” Evolutionary Intelligence,2024.
9. G.Arulselvietal.,“Music Recommendation System Based on Facial Expression,”International Research Journal of Advanced Engineering and Management (IRJAEM), 2025.
10. S. K. Yadav “Emotion-Based Music Recommendation System Integrating Facial Expression Recognition and Lyrics SentimentAnalysis,”, 2025.
11. A.Kumarand P.Singh, “Music Recommendation System by Analyzing Facial Expressions” SSRN Electronic Journal, 2024.
12. M. Sharma , “Emotion-Aware Music Recommendation using MobileNetV2 CNN,” International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2024.
13. R. Patel, “Music Recommendation Based on Facial Expressions using CNN,” International Journal of Engineering Trends and Technology (IJETT), 2024.
14. L. Wijaya, “Music Recommendation System Based on Facial Expression Analysis,”, Journal of Informatics and Visualization (JOIV), 2025.