



# Personalized Music Playlists via Deep Learning Emotion Detection

M. Spoorthi<sup>1,\*</sup> Harshitha H. M.<sup>1</sup> Pooja R.<sup>1</sup> Anusha M. K.<sup>1</sup> Preethi R.<sup>1</sup>

<sup>1</sup> Department of Information Science and Engineering, Maharaja Institute of Technology, Thandavapura, Mysore, India

Emails: [mspoorthi03@gmail.com](mailto:mspoorthi03@gmail.com) · [harshitha@gmail.com](mailto:harshitha@gmail.com) · [poojarchnagar8@gmail.com](mailto:poojarchnagar8@gmail.com) · [anu.anushakempraj@gmail.com](mailto:anu.anushakempraj@gmail.com) · [preethi4mn20is020@gmail.com](mailto:preethi4mn20is020@gmail.com)

Received: October 24, 2023 Revised: January 27, 2024 Accepted: April 28, 2024 ★ Corresponding author

## ABSTRACT

Music holds significant sway in enriching the lives of individuals, serving as a vital source of entertainment for enthusiasts and listeners alike. Moreover, it transcends mere amusement, often adopting a therapeutic role in people's lives. In the ever-evolving landscape of music and technology, this project emerges as a groundbreaking endeavor driven by the profound impact music holds on individuals' lives. Leveraging technological advancements in music players, such as playback control and genre classification, our focus is on revolutionizing playlist creation. Instead of the laborious manual curation of playlists, we introduce automation based on users' emotional states, identified through real-time facial expression analysis via a camera. The human face, a rich source of mood indicators, becomes the key input for our system. By directly extracting emotional cues from facial expressions, the project aims to swiftly deduce the user's emotional state, crafting a tailored playlist without the need for time-consuming manual efforts. Implemented through deep learning using the VGG16 model, the system ensures intricate emotion recognition from image input. Python, OpenCV, and Keras facilitate seamless video processing and deep learning functionalities, complemented by a music player library for smooth playback control. This amalgamation of computer vision and deep learning delivers an interactive music player that dynamically selects tracks aligned with users' real-time emotional expressions, offering a personalized and immersive musical experience.

**Keywords:** Visual Geometry Group (VGG16) ▪ Computer Vision ▪ Emotion Detection ▪ Deep Learning

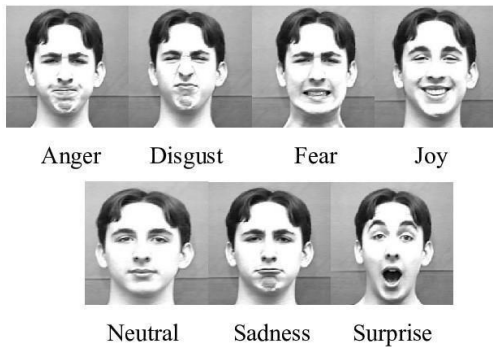
## 1. INTRODUCTION

The field of emotion-based music recommendation has witnessed significant advancements, driven by growing interest in leveraging human emotions to enhance personalized experiences. Understanding and interpreting human emotions, particularly through facial expressions, plays a pivotal role in designing interactive systems such as emotion-based music players. Emotion detection systems offer applications across education, healthcare, and entertainment. In recent years, facial expression recognition (FER) has emerged as a prominent technique for detecting human emotions. By

analyzing facial cues and expressions, FER systems can infer users' emotional states and tailor experiences accordingly.

This paper explores the development of an emotion-based music player system that integrates FER algorithms, specifically convolutional neural networks (CNNs) and Local Binary Patterns Histograms (LBPH), to accurately identify users' emotions from facial images. Unlike conventional approaches that often rely solely on pre-trained models or singular techniques, the system combines the strengths of CNNs and LBPH to achieve robust emotion detection. The work also leverages the efficiency and effectiveness of the VGG16 architecture for feature extraction and classification. By fine-tuning VGG16

on a custom dataset, the system improves accuracy in recognizing emotional states and recommends music tracks that resonate with users' detected emotions.



**Figure 1.** Seven basic human emotions.

### 1.1 Motivation

The motivation behind crafting personalized music playlists through emotion detection using deep learning, specifically leveraging VGG16, is rooted in the effort to revolutionize the music listening experience. The project seeks to integrate deep learning models capable of deciphering real-time emotions from facial expressions. The underlying drive is to create a sophisticated system that detects the user's face and interprets emotional state with nuance. Combining the technical strength of VGG16 in image classification with the emotional resonance of personalized playlists enables a dynamic and immersive journey through music, driven by the fusion of artificial intelligence and emotion-aware entertainment.

### 1.2 Objective

The primary objectives of this research are:

1. To implement robust emotion detection algorithms within the music player to accurately analyze users' facial expressions and understand emotional states.
2. To develop a user-friendly interface that integrates facial emotion detection, automatically generates personalized playlists tailored to mood, and allows users to customize the sensitivity of emotion detection features.
3. To refine and optimize the emotion detection model through machine learning, ensuring adaptability to diverse facial expressions and dynamically adjusting song recommendations based on real-time emotional changes during playback.

## 2. RELATED WORKS

Chaohui Lv et al. [1] reviewed methods and algorithms for image signal processing, biomedical engineering, and informatics, including segmentation, feature extraction, and classification. Mehmet Bilal Er and Ibrahim Berkan Aydilek [2] developed music emotion recognition using chroma spectrograms and deep visual features. Harsha Vijay Bodhe [3] explored music recommendation based on facial expression recognition and deep learning, mapping detected expressions to corresponding music tracks.

Atabansi et al. [4] investigated transfer learning with VGG16 for near-infrared facial expression recognition and reported strong accuracy on the Oulu-CASIA NIR dataset. Indumathi et al. [5] proposed a real-time emotion-based music player using CNN architectures and transfer learning models. Chankuptarat et al. [6] presented an emotion-based music player that suggests songs using heart-rate or facial-image analysis.

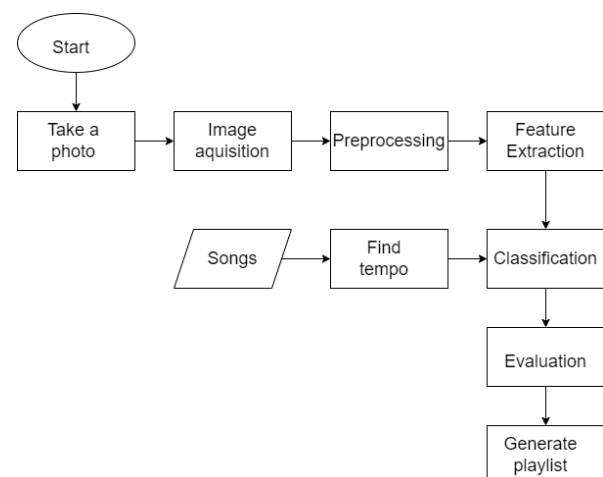
Other works have examined affective gaming, CNN-based real-time emotion detection, and machine-learning music players that select songs according to detected emotions [7, 4, 8]. Mood-based and smart-home systems have also used facial expression detection and playlist selection to synchronize music with user states [9, 10, 11]. Several related systems employ Fisher Face, Haar cascade, EMO algorithms, APIs, and facial-recognition pipelines to generate personalized playlists based on moods [12, 13, 14, 15, 16, 17, 18].

## 3. PROPOSED METHODOLOGY

The proposed system integrates live image acquisition, preprocessing, feature extraction, fine-tuning, emotion classification, and music recommendation. A camera captures the user's facial image, preprocessing removes irrelevant information, and VGG16 extracts discriminative facial features. A classifier predicts the emotion class, and a recommendation engine retrieves songs that match the detected mood.

### 3.1 Image Acquisition

The system first acquires facial images from a webcam. Face detection algorithms such as Haar cascades and LBPH localize the face in real time. Background subtraction separates the user's face from the surrounding scene by analyzing pixel-intensity differences between frames. Image enhancement, including brightness, contrast, sharpness adjustment, noise reduction, and color correction, improves visual clarity. Resolution and frame-rate control balance detailed capture with computational efficiency.



**Figure 2.** Proposed methodology.

### 3.2 Preprocessing

Preprocessing prepares the acquired images for analysis. Background subtraction removes irrelevant visual content and isolates the region of interest. Image cropping retains the facial region only, reducing computational complexity.

Image resizing standardizes the cropped face to  $224 \times 224$  pixels, matching the input requirements of VGG16. Normalization adjusts pixel values to a common scale, mitigating variations in lighting and color intensity. Data augmentation applies transformations such as rotation, flipping, and scaling, increasing training diversity. Quality-control steps remove low-quality or blurry images to improve reliability.

### 3.3 Feature Extraction

Feature extraction transforms preprocessed images into informative representations of facial expressions. CNNs learn hierarchical representations directly from pixels through convolutional, pooling, and fully connected layers. VGG16, pre-trained on ImageNet, is used as a feature extractor and fine-tuned for emotion detection. Feature maps generated by convolutional layers capture edges, textures, and higher-level facial structures such as eyes, nose, and mouth. Pooling layers reduce spatial dimensions while retaining key information, and ReLU activation introduces non-linearity for learning complex expression patterns.

### 3.4 Fine-Tuning and Classification

Fine-tuning adjusts the parameters of the pre-trained VGG16 network to adapt it to facial expression recognition. The weights of selected layers are updated during training so that the model learns subtle patterns essential for emotion classification. Dense layers are added on top of the convolutional base to map extracted features to emotion classes. Dropout layers reduce overfitting, and a softmax classifier computes the probability of each class. The predicted emotion is selected as the class with the highest probability:

$$P(y = i | \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad \hat{y} = \arg \max_i P(y = i | \mathbf{x}), \quad (1)$$

where  $K$  is the number of emotion classes and  $z_j$  is the score for class  $i$ .

### 3.5 Recommendation

The recommendation engine generates personalized song recommendations based on the detected emotion. Content-based filtering analyzes song features such as tempo, key, genre, mood, lyrical sentiment, and instrumentation. Collaborative filtering leverages preferences of users with similar emotional profiles. Hybrid approaches combine content-based and collaborative strategies to increase recommendation relevance and diversity. A music database stores songs tagged by emotional content and acts as the foundation for mapping emotion predictions to playlists.

## 4. PROPOSED ALGORITHM

### 4.1 Feature Extraction Using LBPH

The LBPH procedure begins by taking an input facial image and converting it to grayscale. Local binary patterns are calculated for each pixel by comparing the center pixel with its neighbors. The image is divided into regions, histograms are computed for each region, and all regional histograms are concatenated to form the final LBPH feature vector.

**Step 1:** Input a facial image and convert it to grayscale.

**Step 2:** For each pixel, compare neighboring intensities with the center pixel to produce a local binary code.

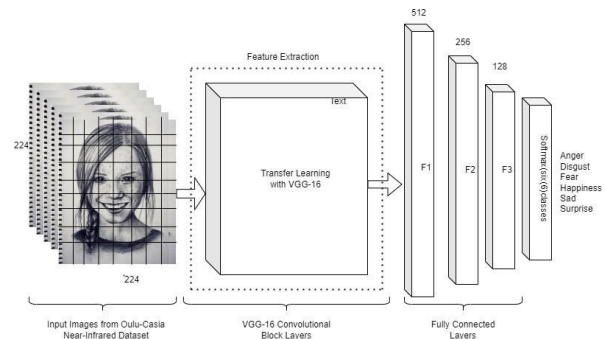
**Step 3:** Divide the image into spatial regions and compute a histogram of LBP codes for each region.

**Step 4:** Concatenate regional histograms to create the LBPH descriptor.

**Step 5:** Pass the descriptor to the classifier or combine it with deep features for emotion recognition.

### 4.2 Feature Extraction Using VGG16

The VGG16-based algorithm takes preprocessed facial images resized to  $224 \times 224$  pixels. Images are normalized and passed through frozen convolutional blocks of the VGG16 network. The extracted feature maps are flattened or pooled into feature vectors. These vectors are passed to newly added dense layers and classified with softmax into seven emotions: happiness, sadness, surprise, anger, disgust, neutral, and fear.



**Figure 3.** Proposed system architecture using VGG16 feature extraction and softmax classification.

## 5. RESULTS AND DISCUSSION

### 5.1 Dataset

The FER2013 dataset consists of grayscale images portraying facial expressions labeled with one of seven emotions: happy, disgust, sadness, neutral, anger, surprise, and fear. It contains 35,887 images divided into training, validation, and test sets. The dataset is used to train CNN models for emotion-based music player systems. After training, the model classifies facial expressions into emotion categories and enables the system to recommend music tailored to the detected emotional state, such as upbeat songs for happiness and calming melodies for sadness.

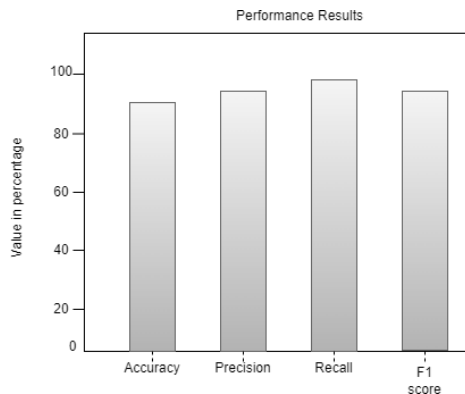
### 5.2 Experimental Setup

The experimental setup uses a computer with sufficient processing power, a webcam for image capture, speakers or headphones for audio output, and a stable internet connection to access the Spotify API. The software environment includes a Python runtime, OpenCV for image processing, TensorFlow or PyTorch for CNN implementation, and Pandas for preprocessing. GPU acceleration can improve processing speed, and additional libraries may support multimodal inputs or voice analysis.

### 5.3 Model Evaluation

The VGG16 model was trained over 400 epochs, with each epoch processing batches of 478 images. Hyperparameters

were adjusted to optimize performance, and training was conducted on Google Colab. The model was evaluated using accuracy, recall, precision, and F1-score.



**Figure 4.** Proposed CNN-VGG16 performance metrics.

## 5.4 Result

The proposed CNN-VGG16 model performs well, with high recall and precision and a balanced F1-score. These results indicate effective identification of positive instances while minimizing false positives. Table 1 compares the proposed method with existing algorithms.

**Table 1.** Comparison of Existing and Proposed Systems

Algorithm	Accuracy	Precision	Recall	F1-Score
CNN	57.1%	57.2%	57.1%	57.1%
KNN	74.2%	73.9%	73.9%	75.1%
Haar Cascade	71.8%	70.9%	70.9%	71.1%
VGG16 (proposed)	85.8%	84.0%	85.0%	85.2%

## 6. CONCLUSION AND FUTURE WORK

### 6.1 Conclusion

The fusion of the VGG16 model for facial recognition with the LBPH algorithm for expression-based music playback creates a potent system that leverages computer vision to enrich user experiences. Deep learning, increasingly prominent across fields such as finance and medicine, finds significant utility in this application. The system's ability to detect human emotions from live video inputs has implications for communication, interaction, behavioral research, and medical rehabilitation. Facial images provide non-invasive emotion detection with swift and effective results. The integration of neural networks achieved strong real-time emotion-recognition performance, underscoring the efficacy of the VGG16-LBPH approach for personalized and emotionally attuned music experiences.

### 6.2 Future Work

Improvements in deep learning techniques continue to enhance the accuracy and speed of facial recognition and emotion analysis, particularly on resource-limited devices. Future advancements may merge biometric modalities such as fingerprints or voice patterns with facial recognition to better understand user emotions. Such integration can support applications including therapy sessions, personalized music suggestions, augmented and virtual reality environments, and

drowsiness detection for safer travel. As these technologies evolve, ethical and privacy concerns surrounding sensitive biometric data must be addressed.

## FUNDING

This research received no external funding.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest.

## REFERENCES

- [1] C. Lv, S. Li, and L. Huang, "Music emotion recognition based on feature analysis," in *11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2018.
- [2] M. B. Er and I. B. Aydilek, "Music emotion recognition by using chroma spectrogram and deep visual features," *International Journal of Computational Intelligence Systems*, pp. 1622–1634, 2019.
- [3] H. V. Bodhe, "Movie and music recommendation system based on facial expression," *International Research Journal of Modernization in Engineering Technology and Science*, pp. 1014–1019, 2024.
- [4] C. C. Atabansi, T. Chen, R. Cao, and X. Xu, "Transfer learning technique with vgg-16 for near-infrared facial expression recognition," *Journal of Physics: Conference Series*, pp. 1–11, 2021.
- [5] I. S. K., S. K., and K. M. C., "Real time emotion based music player using cnn architectures," *EPRA International Journal of Multidisciplinary Research*, pp. 272–278, 2022.
- [6] K. Chankuptarat, R. Sriwatanaworachai, and S. Chotipant, "Emotion-based music player," in *IEEE Proceedings*, 2019.
- [7] P. Kalansooriya *et al.*, "Affective gaming and real-time emotion detection using eeg signals."
- [8] S. Deebika *et al.*, "Machine learning based music player using convolutional neural networks."
- [9] S. M. G. *et al.*, "Mood-based music system using machine learning techniques."
- [10] S. S. Mishra, K. B. Abgul, P. Patil, and U. Singh, "Mood recognition and playlist generator," *International Journal of Research Publication and Reviews*, pp. 2326–2330, 2023.
- [11] G. U. Vijaysinh, G. S. Shivaji, M. R. Ashok, and J. H. Shahaji, "Emotion based music recommendation system," *International Journal of Innovative Science and Research Technology*, pp. 542–545, 2022.
- [12] M. S. Vani and N. S. Divya, "Emotion based music recommendation system," *International Journal of Creative Research Thoughts*, pp. 697–702, 2022.

- 
- [13] P. Vishali and V. Narayani, "A survey on emotion-based music player," *International Research Journal of Engineering and Technology*, pp. 5888–5891, 2020.
- [14] M. Talele, Y. Gurnani, H. Rochani, M. Patil, and K. Soneja, "Smart music player using mood detection," *International Research Journal of Modernization in Engineering Technology and Science*, pp. 2201–2207, 2022.
- [15] S. Kalpande, R. Allampallewar, P. Vyavhare, A. Kinwad, and S. Sargaiyye, "Emotion based music player," *International Research Journal of Modernization in Engineering Technology and Science*, pp. 2752–2753, 2023.
- [16] G. P. P. Sutar, K. Kumari, R. Pandey, L. Malphedwar, and R. V. Kherde, "A preliminary report on emotion based api music player," *International Research Journal of Modernization in Engineering Technology and Science*, pp. 2716–2722, 2022.
- [17] M. Gurale, S. Kalpande, R. Allampallewar, P. Vyavhare, A. Kinwad, and S. Sargaiyye, "Emotion based music player: Enhancing music experiences," *International Research Journal of Modernization in Engineering Technology and Science*, pp. 7594–7598, 2023.
- [18] C. Mounika, S. Jagannathan, and V. M. Prasad, "Smart music player that is based on real-time facial expression in real time," *International Journal of Creative Research Thoughts*, pp. 451–457, 2022.