Grinrey Conference Proceedings
ICMESR 2026, India, 04-05 May 2026
Volume 01 (2026) 001504
https://doi.org/10.55084/gcp/001504

EMOSIC: Emotion-Based Music Player Using Haar Cascade and Deep CNN

Aswin Animon[1], Nisy John Panicker[1,*], Antony A Chirayath[1], Aravind S Menon[1] and Arun Prasad M[1]

[1]Department of Computer Science and Engineering, Albertian Institute of Science and Technology, Kerala, India

Email: nisyjohn13@gmail.com ([*]Corresponding Author)

Received : 17 Jun 2026; Accepted : 08 Aug 2026; Published : 08 Aug 2026.

Abstract. This paper is aimed at enhancing the user experience of listening to music by incorporating emotion detection through facial recognition technology. Facial expressions are used to detect the user's emotional state and the system recommends a playlist matching their mood. After each song, the emotion detection process is iterated to make sure that the next song played serves the user's current emotional state. This is made possible by coining emotion detection algorithms and machine learning techniques. The user's facial expressions are captured using a camera which is connected to the system, and the image is processed to identify the emotion using deep learning models. The system then matches the detected emotion with a pre-defined playlist that matches the user's mood and plays a random song from that playlist. This creates a personalized music experience for the user. The proposed paper aims to provide users a unique and interactive music-listening experience. The system can accurately determine the user's emotional state and provide them with music that matches their mood. The repetition of the emotion detection process after each song ensures that the playlist remains relevant to the user's emotional state, creating a more personalized music experience.

Keywords: *Functional API, FER-2013, EMOSIC, Emotions*

1. INTRODUCTION

Emotion recognition is the process of identifying human emotion. It is related to the field of artificial intelligence for the purpose of automating various processes that are comparatively harder to perform manually. To make effective automated decisions best suited to a person, it is important to understand the state of mind based on their emotions. Humans have the ability to express and identify emotions. The computer identifies human emotions either through sensors or by image analysis. In our daily and professional life, we interact with many people directly or indirectly where it is necessary for people to be aware of their present emotions toward the person with whom they are interacting. Human emotions are generally classified as: surprise, fear, anger, happiness, sadness, disgust, and neutral.

Studies had proof that music is a super healer. However, most people face the difficulty of selecting songs that match with their current emotions. Many will randomly pick the songs available in the song folder and play them with a music player which mostly does not match their present emotion. Since the facial expression is closely associated with the emotion, it is mostly an involuntary action. It is nearly impossible for an individual to insulate himself from expressing emotions. Though there are systems to identify facial motion, existing methods do not incorporate recommending music by automatically fetching the human emotions; and some use additional hardware which incurs extra cost.

 ISBN: 978-81-996627-3-5

Facial expression analysis includes both detection and identification of facial expression [2]. The three approaches included in the automatic facial expression analysis (AFEA) are face acquisition, facial data extraction and facial expression recognition as shown in Fig.1.

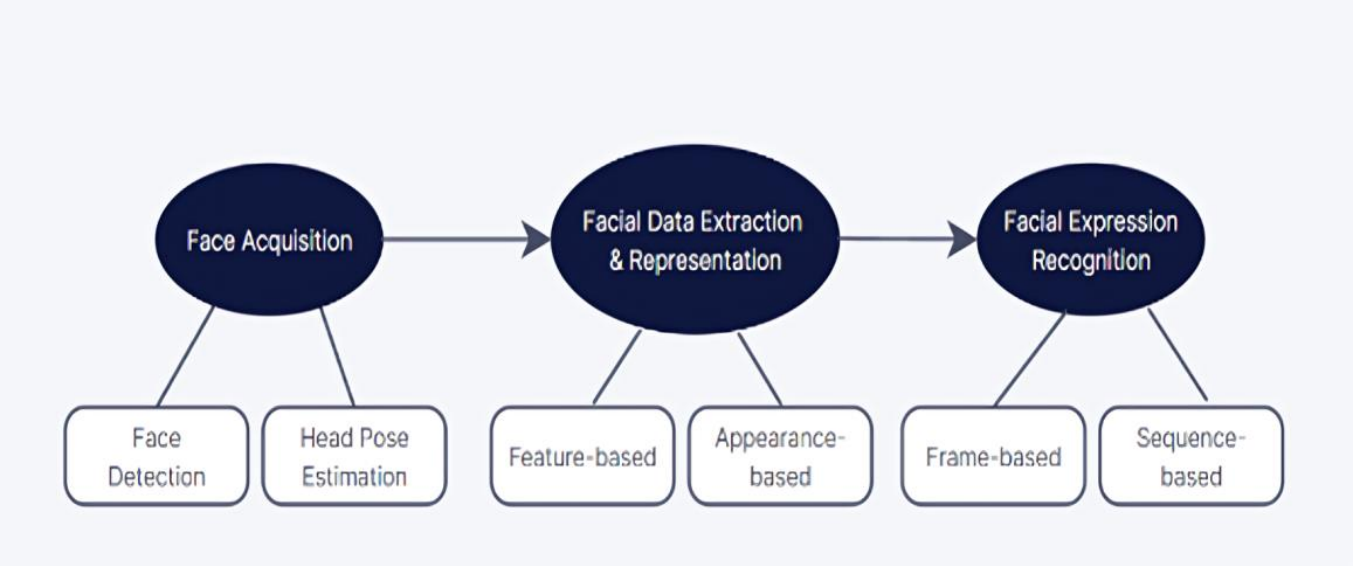

Fig. 1. Basic structure of AFEA

The "Emotion Based Music Player" is a software aimed to develop a real time facial emotion recognition system that detects the emotion of an individual using Haar cascade and classifies them using Deep Convolutional Neural Network. These emotions are automatically mapped to curated playlists and the music player will play songs that match the current emotion of the individual. The system focuses on the analysis of the facial expression only which does not include the head or facial movement.

2. RELATED WORKS

In 1994, Antoinette L. Bouhuys et al. studied the relationship between individuals' facial expressions after hearing sad music. The study proved that music can actually influence individuals' emotions [1]. Daniel T. Bishop, Costas I. Karageorghis, and Georgios Loizou presented their research on manipulating young tennis players' emotional states through music. The research signals that participants will often choose to listen to music in order to elicit various emotional states, increased arousal, and visual and auditory imagery [2]. In 2001, Matthew Montague Lavy developed four premises regarding music lovers and their relationship to music - music is heard as sound, a human utterance, a context and as a narrative [3].

Many deep learning techniques have been developed for facial and music emotion recognition separately. Yang and Chen [4] proposed a ranking-based approach for retrieving music corresponding to emotion recognition but without real time capability. Zhang et al. [5] combined CNN with edge detection for face recognition; Yin and Zhang [6] applied an improved activation function for image recommendation, not extended to facial emotion. Wang et al. [7] designed a CNN-based system for emotional big data achieving high accuracy but limited to an autism-care context.

Recent works have extended these directions: Gorasiya et al. [8] achieved high facial emotion accuracy using CNN but without continuous re-detection; Nguyen et al. [9] fused facial and musical emotion recognition through a rule-based mapping; and Bottu and Ragavan [10] extended facial CNN-based recognition with transformer-based lyrics sentiment analysis. The summary of these computational approaches is provided in Table 1.

Table 1. Comparison of Related Emotion Recognition and Recommendation Techniques

Author(s)/ Year	Method/Technique	Dataset	Accuracy/ Results	Limitations
Yang & Chen, 2011 [3], [4]	RBF-ListNet ranking algorithm; valence-arousal Cartesian representation	Song-level dataset (ranked pairwise)	0.326 Gamma statistic for valence recognition	Addresses music emotion tagging, not facial-based detection; no real-time listener feedback loop
Zhang et al., 2019 [5]	CNN with edge detection, max pooling, Softmax classifier	FER-2013 + LFW	88.56% average recognition rate	Improves robustness but no integration with music
Yin & Zhang, 2020 [6]	CNN with improved LReLU-Softplus activation function; combines image + text data	Image libraries	88% overlap with human recommendation results	Focused on general image recommendation, not facial emotion or music-specific
Wang et al., 2021 [7]	CNN with two-stage strategy and joint supervision, camera/IoT-based facial capture	FER-2013, MCFER, CK+	95.89% accuracy on MCFER	Designed for autism-care context, not music recommendation; high accuracy but narrow application domain
Gorasiya et al., 2022 [8]	CNN, 7-category facial emotion classification	FER dataset	94% classification accuracy	High accuracy but no real-time re-detection loop; single-modality
Nguyen et al., 2024 [9]	ResNet18/VGG19 (facial), SVM/LDA (musical emotion), rule-based fusion	FER-2013, CK+, Spotify mood dataset	72.67–96.97% (facial), 86.96% (music, SVM)	Dual-model pipeline increases complexity; no continuous mood-tracking after playback
Bottu & Ragavan, 2025 [10]	CNN (4 conv layers) for FER + transformer-based lyrics sentiment fusion	FER-2013, curated 1,000-song lyrics dataset	64.93% test accuracy (facial); 85.71–88.45% (music,majority voting)	Large train-val gap (91.78% train vs. 64.73% val.) indicates overfitting; no real-time re-detection loop

3. DATASET

3.1 FER 2013 (Facial Expression Recognition 2013 Dataset)

Facial Expression Recognition 2013 (FER-2013) dataset was prepared in Challenges in Representation Learning: Facial Expression Recognition Challenge, hosted by Kaggle. The FER-2013 database has seven facial expression categories - angry, disgust, fear, happy, sad, surprise, and neutral. The training set consists of 28,709 images, validation set and test set of 3,589 images each. All images in this dataset are grayscale with 48×48 pixels, thus corresponding to faces with various poses and illumination, where several faces are covered by hand, hair, and scarves as shown in Fig.2.

Fig.2. FER 2013 Dataset Sample [11]

Before model training, the following pre-processing steps were performed. Each pixel was reconstructed to serve as valid input for the convolutional neural network. Using a label encoder, the seven emotion labels were encoded numerically and later converted to one-hot encoded vectors. After splitting the dataset into training and validation sets, normalization and data augmentation were done to reduce overfitting and improve model generalization.

4. METHODOLOGY

EMOSIC is an innovative paper that aims to revolutionize the way users experience and interact with music. Facial recognition, emotion detection algorithms, and machine learning are combined to create a personalized and emotionally immersive music-listening environment.

The recommendation system combines Haar Cascade detector and Deep Convolutional Network for face localization and emotion classification. The manual song selection process is avoided to deliver mood responsive music listening experience to the user.

4.1 Haar Cascade Frontal Face Detection Algorithm

In EMOSIC, the Haar Cascade frontal face detection algorithm is used as part of the facial recognition and emotion detection process. The algorithm works by analyzing patterns and features of the face, such as edges, corners, and textures.

The algorithm operates by moving a detection window over the image, applying the classifiers at different positions and scales. At each step, the algorithm evaluates whether the current region of the image matches the pattern of a face. If a match is found, the algorithm identifies the region as a potential face and continues to refine the detection by applying subsequent classifiers. False positives are minimized by adjusting the threshold values and incorporating a series of cascaded classifiers.

4.2 Deep Convolutional Neural Networks (DCNN)

In this paper, DCNNs are the main component of the emotion detection process. DCNNs are a class of neural networks specifically designed for analyzing visual data, such as images or video frames, and they have demonstrated exceptional performance in various computer vision tasks [7]. Convolutional layers apply a series of filters to extract spatial features from the input data, capturing patterns and structures at different levels of abstraction. Fully connected layers aggregate these extracted features and perform classification or regression tasks.

When a user interacts with the EMOSIC system, a camera captures their facial expressions. These facial expressions are then processed by the DCNN model to predict the emotional state of the user which is done by assigning a probability distribution across the different emotion categories, indicating the likelihood of each emotion being expressed.

4.3 Music Selection Mechanism

A personalized playlist based on the user's emotional state is created by the system. It considers the detected emotions from the user's facial expressions and matches them with a collection of songs that evoke similar emotions.

Once the user's emotions are recognized using Haar Cascade and DCNNs, the system maps these emotions to a pre-defined set of playlists that correspond to different emotional states. Each playlist contains a collection of songs that are known to evoke specific emotions. For instance, a playlist associated with happiness may include cheerful and upbeat songs, while sadness may contain more sad and introspective tracks.

Based on the detected emotions and the corresponding emotion-playlist mapping, the music selection mechanism recommends a playlist that matches the user's emotion, and a random song from that playlist is played so that each time the user gets a new feel.

4.4 User Interface

The graphical user interface of the system is developed using Django that covers various elements and functionalities allowing users to interact with the system and customize their music experience. Following is an overview of the user interface components:

- User Authentication: The user interface includes a login and registration system where users can create accounts and log in to access their personalized music profiles.
- Dashboard: The dashboard serves as the main interface where users can view and control their music-listening experience.
- Emotion Detection: By clicking this button an emotion detection window will appear and start to detect and a subsequent song is played.
- Playlist Management: The admin will be able to add or remove songs to the playlists according to the emotion.

5. RESULT & ANALYSIS

A registration page is available for new users and a login page for existing users in the EMOSIC. After successfully logging in, the user gets access to multiple modules like Emotion Detection, All Songs, and Playlists.

When the "Emotion Detection" module is selected, a detection window appears and detects the facial emotion of the user via webcam and appropriate song according to the emotion is played. The facial emotion is processed in real-time and the detected emotion playlist is chosen from which a random song is played. After the completion of the song, the emotion detection window will appear again and repeat the process and a new song according to the newly detected emotion is played. On clicking the "All Songs" module, the users can see and play all music that has been added to the system. In the "Playlists" module, every song categorized under each emotion can be viewed. The admin of the system can add music along with its details including name and emotion playlist to which it is to be designated.

5.1 Model Training Accuracy

When the Simple CNN model was trained with the FER-2013 dataset it showed an accuracy percentage of 70.64% and a validation accuracy of 69.02% after 86 epochs as shown in Fig.3.

5.2 Model Training Loss

In machine learning, the loss of a model refers to a measure of how well the model is performing in terms of its ability to predict the correct output. When the Simple CNN model was trained with the FER-2013 dataset it showed a loss percentage of 80.50% and a validation loss of 87.68% after 86 epochs as shown in Fig.4.

Fig.3. Accuracy Plot

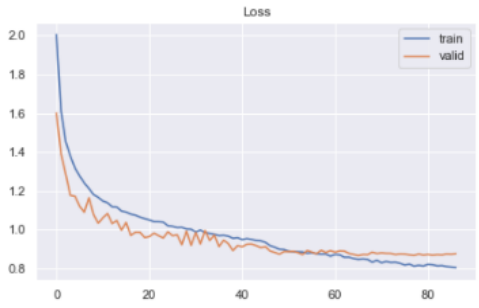

Fig.4. Loss Plot

5.3 Model Testing

The trained model is tested in real-time for the detection of 7 different emotions and the results are shown in Fig.5.

Fig.5. Emotion detection by EMOSIC

5.4 Discussion of Results

The model's performance on the FER2013 dataset with the encodings 0 - Anger, 1 - Disgust, 2 - Fear, 3 - Happiness, 4 - Sadness, 5 - Surprise, 6 - Neutral are shown in Fig.6. And Fig.7. The model achieved high accuracy for the emotion Happiness with 792 out of 899 samples correctly classified. Only 233 out of 512 Fear samples were correctly classified as it was confused with Sadness and Anger. Though the precision for Disgust was high (0.83), the recall was low (0.55) as they were wrongly classified as Anger.

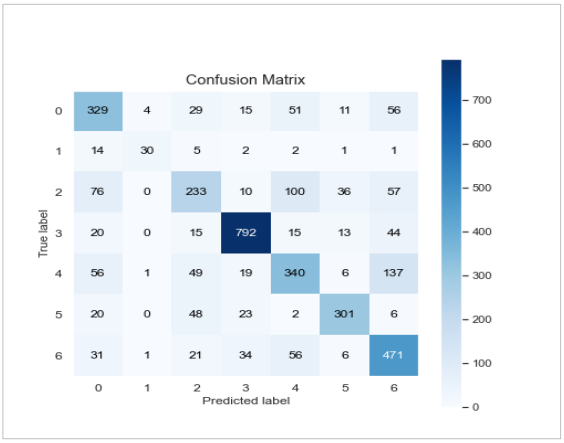

Fig.6. Confusion Matrix

	Precision	Recall	f1-score	Support
0	0.60	0.66	0.63	495
1	0.83	0.55	0.66	55
2	0.58	0.46	0.51	512
3	0.88	0.88	0.88	899
4	0.60	0.56	0.58	608
5	0.80	0.75	0.78	400
6	0.61	0.76	0.68	620
Avg/total	0.70	0.70	0.69	3589

Fig.7. Other Metrics

The training accuracy of 70.64% and validation accuracy of 69.02% shows a gap but it is relatively small when compared to other systems such as CNN and the lyrics-fusion system in [9]. The later methods had a training accuracy of 91.78% but a lower test accuracy of 64.93%. Overall, EMOSIC generalizes reliably for Happiness, Surprise, and Neutral emotions with improvements needed for Fear and Sadness as they are overlapping emotion categories.

6. CONCLUSION

The Emotion-based Music Player offers a unique and personalized music-listening experience by including facial recognition technology to detect the user's emotion. The playlist recommended by the system matches well with the user's mood as the facial expressions are accurately identified thereby creating an emotional connection with music. While benefiting from EMOSIC, users can use various streaming platforms in future so that the system gets access to the enormous music libraries. Additionally, a mobile application of this system can be developed to attract more users.

REFERENCES

[1] A. L. Bouhuys, G. M. Bloem, and T. G. Groothuis, "Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects," *Journal of Affective Disorders*, vol. 33, no. 4, pp. 215-226, 1995. [CrossRef] [Publisher]

[2] D. T. Bishop, C. I. Karageorghis, & G. Loizou, "A Grounded Theory of Young Tennis Players' Use of Music to Manipulate Emotional State," *Journal of Sport and Exercise Psychology*, vol. 29, no. 5, pp 584-607, Oct. 2007. [CrossRef] [Publisher]

[3] Lavy, M. Montague, "Emotion and the Experience of Listening to Music: A Framework for Empirical Research," 2001, Apollo - University of Cambridge Repository. [CrossRef] [Publisher]

[4] Y. H. Yang and H. H. Chen, "Ranking-Based Emotion Recognition for Music Organization and Retrieval," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 4, pp. 762-774, May 2011. [CrossRef] [Publisher]

[5] H. Zhang, A. Jolfaei and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," in IEEE Access, vol. 7, pp. 159081-159089, 2019. [CrossRef] [Publisher]

[6] P. Yin and L. Zhang, "Image Recommendation Algorithm Based on Deep Learning," in IEEE Access, vol. 8, pp. 132799-132807, 2020. [CrossRef] [Publisher]

[7] H. Wang, D. P. Tobón V., M. S. Hossain and A. E. Saddik, "Deep Learning (DL)-Enabled System for Emotional Big Data," in *IEEE Access*, vol. 9, pp. 116073-116082, 2021. [CrossRef] [Publisher]

[8] T. Gorasiya, A. Gore, D. Ingale and M. Trivedi, "Music Recommendation based on Facial Expression using Deep Learning," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1159-1165. [CrossRef] [Publisher]

[9] H. Nguyen, N. Tran, Y. Tran, D. Ly, A. Tran, A. Nguyen, and H. Vo, "A model for song recommendation based on facial emotion analysis and musical emotion," *International Journal of Intelligent Engineering & Systems*, vol. 17, no. 4, pp. 1028-1041, 2024. [CrossRef] [Publisher]

[10] V. S. G. S. Phaneendra Bottu and K. Ragavan, "Emotion-Based Music Recommendation System Integrating Facial Expression Recognition and Lyrics Sentiment Analysis," in *IEEE Access*, vol. 13, pp. 87740-87752, 2025. [CrossRef] [Publisher]

[11] FER-2013, Accessed on 30 June 2026. [Publisher]