M Velmurugan , 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Music Recommendation System Using Facial Expressions

M Velmurugan[1], Danthaluri Tharuni[2], Maramreddy udaya sree[3]
Department of Computer Science & Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

Abstract- Music has a strong influence on how we feel and on how we experience our world. As AI and ML are rapidly developing, new possibilities are opening for making technology more personalized and emotionally intelligent.The project, entitled "Music Recommendation System Using Facial Expressions," seeks to create a music player that can recognize the emotional disposition of its listener and provide recommendations for songs that correspond to their emotional states.A webcam records the face of the user in real time. Open CV and CNN algorithms analyze various emotions of the user such as happiness, sadness, anger, surprise, or having a neutral expression. Once such an expression has been identified, the system selects music from an existing database that comprises songs for various moods, suggesting songs that fit the user's mood at any particular moment. A simple interface has also been offered for listening to these suggested songs.This kind of technology is quite different from the usual recommendations that depend solely on past listening patterns and selections. This technology offers a more intuitive and emotional link with the selection of music. This technology can prove beneficial in the entertainment sector itself and in providing health assistance. Preliminary testing of this technology has proven quite effective in terms of emotion recognition and popularity of the recommendations.Overall, this project aims to connect human emotions with smart digital systems by using computer vision, deep learning, and affective computing for a more personalized music experience.

Keywords: Facial Expression Recognition, Music Recommendation, Deep Learning, Emotion Detection, CNN, Hybrid Filtering.

I. INTRODUCTION

Music plays a very important part in our daily lives. It can cheer us up, soothe our mind after a tiring day, or simply allow us to go through those emotions which we find hard to express. People are getting more and more demanding with the arrival of streaming platforms such as Spotify, YouTube Music, and Apple Music. The days of one-size-fits-all playlists are gone. What they really want is music that understands them - namely their feelings at a certain time. Most of the recommendation systems, however, still rely on listening history, ratings, or trends. These methods fail to realize that the most important factor is the listener's emotional state at the present time.

We cannot help but show our emotions on our faces. If anyone is happy, sad, or even indifferent, subtle changes in their facial expressions usually have a lot to say. This is the reason why facial emotional recognition is becoming a very efficient means of personalizing the users' digital experience. The progress made in computer vision and deep learning has led to the fact that nowadays machines can understand these less obvious emotional signals almost as humans do.

Project presented in the report, "Music Recommendation System Using Facial Expressions," is an innovative concept aimed at creating a smarter recommendation process which responses directly to the user's mood. The user's face is dynamically grabbed by a webcam and OpenCV is exploited to detect and extract facial features. The expression on the image is then converted by a Convolutional Neural Network (CNN) into one of four or more emotions and labels such as joy, sadness, anger, or neutrality are assigned to the picture. Consequently, the system picks up songs from a database that have been categorized by different emotions and thus, tracks can be in line with user's current mood.

With this method, music can be tailored to users almost in the same way humans act when listening

to music. One does not have to look for the right playlist from thousands of options or trust old preferences as the system is able to respond to the user's instantly thus providing music which really connects with user's mentality. Such a system is not just pleasing as far as entertainment is concerned; there are numerous places focusing on emotional well-being where it can bring a positive change. The device, for example, might be used to support mood regulation during therapy sessions or as a component of smart home gadgets that vary media content depending on the user's emotional state.

In the end, this endeavor is like an effort to find a middle ground between human feelings and technology. With the help of AI, facial expression recognition, and tailor-made media delivery, this platform is just an example of technology adjusting to human nature instead of people having to adapt themselves. As computing ability to be aware of and handle human emotions keeps on developing, such an experience might be a usual digital interaction sooner rather than later.

II. METHODOLOGY

The Music Recommendation System Using Facial Expressions is a System which working process can be explained as five stages that are simply and closely connected. Initially, the face of a user is recorded through a camera. Beforehand, the image is prepared for the next stage so that the most relevant facial features can be extracted and be of great help for the later stages of the analysis. Then the system determines the feeling with the help of a machine learning model that has already been trained.

Once the emotion is recognized, it is linked to a particular mood category. Finally, the system selects and suggests songs that best match how the user is feeling at that moment.

Together, these steps allow the system to take a real-time facial expression and turn it into a meaningful and personalized music recommendation. The overall flow of the system is illustrated in Figure 3.1.

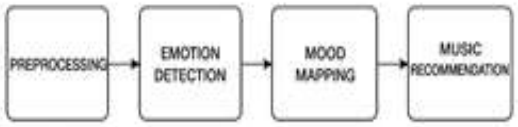

Figure 2.1 shows the overall workflow of the system how it processes.

Facial Image Acquisition

The first step of the system is to capture the user's face in real time using a webcam or any connected camera. As the camera keeps streaming video frames, the software looks for the presence of a face using techniques like the Haar Cascade Classifier or MTCNN, provided through the OpenCV library. When a clear and stable frame is detected where the face is visible without too much blur or movement that image is selected for further processing.

This stage is essential because good lighting and a clean facial image help the system analyze expressions more accurately in the later steps.

Preprocessing

Before the facial image is analyzed by the CNN model, it goes through a preprocessing phase to improve its quality. During this stage many different operations are performed in order to have the image visually sharp and with the most relevant facial features well-highlighted. Such improvements allow the model to be more precise in the recognition of the emotional states.

- **Grayscale Conversion:** The RGB image is converted into grayscale to reduce computational complexity.
- **Resizing:** The image is resized to a standard dimension (48×48 pixels as per the FER-2013 dataset).

- **Normalization:** Pixel values are normalized to the range [0,1] for better CNN convergence.
- **Region of Interest (ROI):** The face area is cropped, and unnecessary background elements are removed.

The result shows that the model is working well since it has good accuracy and low validation loss. The implication of this is that the model has the capability of recognizing the emotions correctly even for a new set of data that was not previously used.

Emotion Detection Using CNN

This is the most critical step involved in the entire process. When the image is cleaned and preprocessed, the image is then passed through the convolutional neural network that observes the user's facial expression and derives the emotion that is being conveyed through the image. In the CNN model, various layers are stacked on each other, with each layer performing different tasks for the derivation of the minute details of the facial expression.

- **Convolutional Layers:** Extract spatial features such as edges, contours, and expressions.
- **ReLU Layers:** Introduce non-linearity for better learning capability.
- **Pooling Layers:** Reduce spatial dimensions to minimize overfitting.
- **Fully Connected Layers:** Combine the extracted features and predict probabilities for each emotion class.
- **Softmax Output Layer:** Classifies the input into one of seven emotion categories — Happy, Sad, Angry, Fear, Disgust, Surprise, or Neutral.

More than 35,000 facial images labeled with one of the seven emotions were compiled into the FER-2013 dataset, which serves as the data source for the model to learn. As a means of assessing the performance of the model after training, metrics such as accuracy and loss can provide insight into whether or not it will produce the intended results when applied to real-world scenarios.

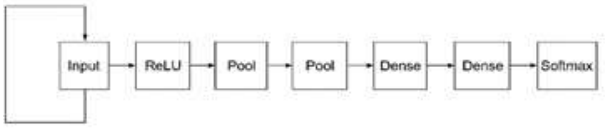

Figure 2.2 shows the overall layout of the CNN model. It begins with the input image, which moves through several convolution and pooling layers that extract important facial features. After that, the data flows into dense layers, and finally to a Softmax layer that determines which emotion is being expressed.

Mood Mapping

Once the emotion is detected, the system maps it to a corresponding music mood category.

The mapping is based on psychological associations between emotions and musical features such as tempo, rhythm, and energy.

Table 2.1: Emotion-to-Mood Mapping Table

Detected Emotion	Mapped Mood	Music Type
Angry	Intense	Rock, high tempo tracks
Fear	Soothing	Soft instrumental
Disgust	Neutral	Balanced tempo tracks
Surprise	Excited	Pop, upbeat songs
Sad	Calm	Slow, acoustic melodies
Neutral	Relaxed	Soft background tracks
Happy	Energetic	Fast beat, major scale songs

III. SYSTEM DESIGN

The Music Recommendation System Using Facial Expressions uses computer vision and recommendation algorithms to create personalized music suggestions based on the user's emotional state. The system has a modular structure with four key layers: Input Acquisition, Emotion Detection, Mood Mapping and Recommendation, and Output Interface.

The Input Acquisition Layer acquires the most recent facial pictures or video streams from the webcam or a mobile camera. Early works like face detection, excluding the unnecessary background, resizing, and normalizing the image for uniformity are done through OpenCV which the system uses. By using only the significant facial features, the model has less work to do and thus it can provide better results. To accelerate the system and at the same time keep user privacy intact, lightweight deep-learning models may also be implemented on the local device instead of the cloud processing, thus the device does not have to be dependent on the cloud entirely.

Next, the Emotion Detection Layer comes into play. This layer is powered by a Convolutional Neural Network (CNN) trained on the FER-2013 dataset to understand the emotional expressions. The model can detect seven emotions that are usually expressed — happiness, sadness, anger, fear, disgust, surprise, and neutral state. In the CNN's process of accurately identifying individuals' emotions, facial characteristics are examined in detail to assign a corresponding emotion label to the correct individual. By training with more refined datasets such as RAF-DB and AffectNet, the CNN can better recognize emotions in authentic contexts.

The Mood Mapping and Recommendation Layer then converts these emotions into musical moods. For instance, if someone appears happy, the system will select songs that are uplifting or fast-paced; if someone appears sad, the system will pick songs that relax or calm them down.

The hybrid approach deployed by the system not only looks at song features to recommend better (such as tempo and emotional tone), but also by the user's previous interactions. The learning process is facilitated by music libraries such as the Spotify Tracks Dataset or the Million Song Dataset.

The Output Interface Layer is, in fact, the last element in the architecture, communicating the findings back to the user. Through a user-friendly web interface, which is made with React.js and linked to a FastAPI backend via RESTful APIs, the recommendations are displayed. The interface shows the emotion detected, the mood mapped, and the list of recommended songs. Users are permitted to like, skip, or replay tracks, besides which these engagements are made use of by the system to constantly refine its recommendations in accordance with the listener's preferences.

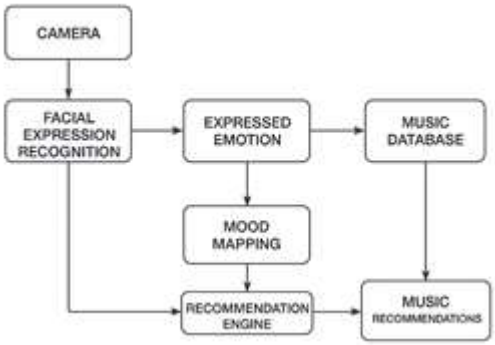

Figure 3.1 illustrates the total layout of the music recommendation system using facial expressions. The chart shows the flow starting from capturing the user's face, then emotion detection and mood interpretation. After recognizing the emotion, the recommendation engine picks appropriate songs from the music database. So, the system provides personalized music recommendations that correspond to the user's present emotional condition.

IV. RESULT AND ANALYSIS

The facial emotion recognition system was effectively created and trained with a selected Convolutional Neural Network (CNN) architecture. The training data set consisted of seven different emotional categories. Several preprocessing steps were taken to prepare the images for the model that would help the system to generalize better.

Among these were normalizing the images and introducing variation through data augmentation techniques like horizontal flipping, zooming, and slight shearing. The adjustments made to the model improved its ability to identify feelings based on various facial expressions from multiple positions and perspectives.

In order to avoid excessive fitting during training, the authors used techniques like Early Stopping and Model Checkpointing. These two features allowed them to stop training at the appropriate moment and keep the instance of the model that performed best on validation data.The outcomes were consistent — both training and validation accuracy were improving from one epoch to another, thus the model was getting trained on the emotional cues of the images.

The plots obtained after the training showed that the metrics related to loss had a clear decreasing pattern with the corresponding increase in classification performance. Therefore the CNN architecture that entails several convolution and pooling layers in addition to dropout was effective in extracting the most useful spatial features from the images while at the same time being quite resistant to overfitting. The dropout layer in particular was instrumental in the model's ability to maintain good performance under the test with new images.

Besides this, during the real-time experimental deployment, the system behaved stably. It could work efficiently with images taken from a live webcam feed, then categorize them into one of the seven different emotions that the model was capable of recognizing, and finally, make music suggestions based on the predictions. The seamless integration between the recognition model and the recommendation module is thus evidenced by the fact that the predictions were utilized for triggering context-aware music recommendations.

Overall, the experimental analysis confirms that the proposed CNN-based recognition system successfully supports emotion-aware music recommendation tasks by detecting facial expressions and triggering appropriate responses. The outcome establishes a strong foundation for multimedia applications integrating affective computing, user behavior sensing, and personalized entertainment technologies.

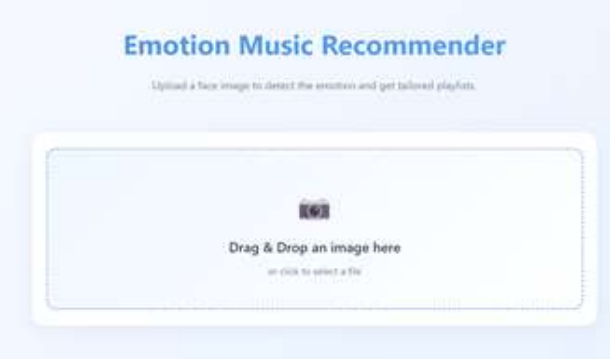

Figure 4.1 shows the frontend of the website where we can drag and drop the image to predict the mood and to recommend the music related to the mood.

Figure 4.2 shows the result of the mood in the given picture and how confidence the model is about the result. According to the result it recommends the songs and playlists to choose.

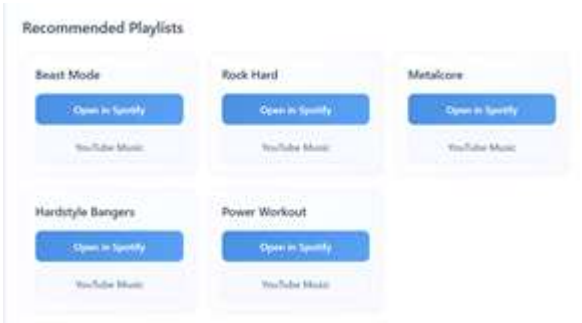

Figure 4.3 shows the recommendation of the playlists from them we can select any one of the playlist and we will be redirected to the Spotify page or YouTube music according to which we select.

V. CONCLUSION AND FUTURE SCOPE

The proposed Music Recommendation System Using Facial Expressions combines computer vision and machine learning to create an emotion-based music recommendation experience. It analyzes real-time facial expressions using a Convolutional Neural Network (CNN) to identify the user's emotional state. Then, it maps this state to a suitable set of songs that match the detected mood. This technology enhances user engagement and personalization vs. the traditional playlist music system. Emotion-aware computing has shown, through empirical evidence, to be quite accurate in the identification of emotional responses as well as in the retrieval of music, thus, it is a technology that can be very useful in the field of entertainment. Next steps could be to expand the emotion dataset, use more sophisticated deep learning models for better precision and also connect to real-time streaming that can create music recommendations on a user's mood.

REFERENCES

1. J. Chen, "CNN-Based Facial Emotion Recognition: Modeling and Evaluation," Engineering Applications of Artificial Intelligence, vol. 135, Art. no. 108815, 2024.
2. R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, "Induced Emotion-Based Music Recommendation through Reinforcement Learning," Applied Sciences, vol. 12, no. 21, Art. no. 11209, 2022.
3. F. Ghaffar, "Facial Emotions Recognition using Convolutional Neural Net," arXiv preprint arXiv:2001.01456, 2020.
4. S. Gupta, "A Survey on Music Recommendation Systems Using Emotion Detection," Cyber Tech Publications, 2023.
5. A. K. M. Monzurul Islam, "Exploring Convolutional Neural Networks for Facial Expression Recognition: A Comprehensive Survey," Global Mainstream Journal of Innovation in Engineering & Emerging Technology, vol. 3, no. 2, pp. 14–26, 2024.
6. E. Jing, Y. Liu, Y. Chai, S. Yu, L. Liu, Y. Jiang, and Y. Wang, "Emotion-aware Personalized Music Recommendation with a Heterogeneity-aware Deep Bayesian Network," arXiv preprint arXiv:2406.14090, 2024.
7. A. Mahadik, S. Milgir, J. Patel, V. B. Jagan, and V. Kavathekar, "Mood Based Music Recommendation System," International Journal of Engineering Research & Technology (IJERT), vol. 10, no. 06, Jun. 2021.
8. A. Petrescu, "Music Recommendations Based on User's Mood Using Convolutional Neural Networks," Studia Universitatis Babeș-Bolyai Informatica, 2022.
9. M. B. C. Rao et al., "Bi-Sampling Approach to Classify Music Mood leveraging Raga-Rasa Association in Indian Classical Music," arXiv preprint arXiv:2203.06583, 2022.
10. M. B. Sutar and A. Ambhaikar, "A Survey on Deep Facial Expression Recognition," Mathematical Statistician and Engineering Applications, vol. 72, no. 1, pp. 470–485, 2023.
11. P. A. Thasneem, Subina, and P. Santhi, "Facial Emotion Recognition using CNN," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), 2021.
12. S. M. Tiwari and M. S. Alam, "Facial Emotion Recognition," IJR Applied Science Engineering Technology, 2023.
13. S. Vinya Sri, M. Harshitha, A. Mahith Reddy, and B. Vyshali, "An Emotion based music recommendation system," IJCRT Research Journal, vol. 15, no. 2, pp. 50583–50591, 2025.
14. Y. Wang and A. Serrano Elcid, "Research on Personalized Music Recommendation Model Based on Personal Emotion and Collaborative Filtering Algorithm," in Proc. International Conference on Machine Learning & Intelligent Computing, PMLR, vol. 245, pp. 404–412, 2024.
15. Y. Wang and A. S. Elcid, "Research on Personalized Music Recommendation Model Based on Personal Emotion and Collaborative Filtering Algorithm," PMLR, vol. 245, pp. 404–412, 2024.