

# A Novel Approach to Music Recommendation via Facial Expression Analysis

**Mohammad Ilyas**  
Assistant Professor  
Moradabad Institute of  
Technology  
Moradabad, India  
[mohd.passion@gmail.com](mailto:mohd.passion@gmail.com)

**Danish Sharma** Computer  
Science and Engineering  
Moradabad Institute of  
Technology  
Moradabad, India  
[sharmadanish2003@gmail.com](mailto:sharmadanish2003@gmail.com)

**Mohd Amaan Khan**  
Computer Science and  
Engineering Moradabad  
Institute of Technology  
Moradabad, India  
[amaan2982@gmail.com](mailto:amaan2982@gmail.com)

**Gaurav Atri**  
Computer Science and  
Engineering  
Moradabad Institute of  
Technology  
Moradabad, India  
[atri41654@gmail.com](mailto:atri41654@gmail.com)

**Mohd Ubaid** Computer  
Science and Engineering  
Moradabad Institute of  
Technology Moradabad,  
India  
[mohdubaid.t@gmail.com](mailto:mohdubaid.t@gmail.com)

## **Abstract**

Facial expressions are a direct manifestation of human emotions and hold significant implications in enhancing human-computer interaction [1]. In this study, we present a novel approach to music recommendation that makes use of facial emotion detection. The system uses a Convolutional Neural Network (CNN) to classify facial expressions into one of seven emotion categories and maps these to corresponding music playlists. Using a dataset of 35,000 labelled images, the model achieved a training accuracy of 75%. This paper discusses methodology, training process, and system integration. Results reflect that the proposed approach is viable in real-time emotion-based music recommendation systems.

## **Keywords**

Facial Emotion Recognition, Music Recommendation System, Convolutional Neural Networks (CNN), TensorFlow, Machine Learning, Image Classification.

## **i. Introduction**

Music is a crucial component of human existence, impacting moods, lowering stress, and promoting wellness [2]. Most traditional music recommendation systems depend on listening history, user preferences, or manual entry but do not cater to a user's actual time emotional state. In our early work, we performed a survey of previous emotion-based music

recommendation projects and discovered that most methods employed sentiment analysis from text or physiological signals such as heart rate and EEG. Although these methods have demonstrated encouraging results, facial expression-based music recommendations are less explored and provide a more intuitive and easier method for emotion detection. Here, we suggest a 'facial emotion recognition-based music recommendation system' that makes real-time, personalized song recommendations.

The system works in two stages: emotion detection and music recommendation. A Convolutional Neural Network (CNN) [2] works on grayscale face images (48x48 pixels) to recognize emotions into seven categories: happiness, sadness, anger, surprise, fear, disgust, and neutrality. In contrast to current systems that involve user inputs or external sensors, our method involves real-time facial analysis to directly correlate emotions to pre-defined playlists for a seamless and automatic recommendation process. Our system is unique in that it does away with manual interactions and uses real-time facial recognition to adapt instantly to user emotions. Through the connection of emotional states to music suggestions, we increase the listening experience to be more dynamic and engaging. This paper continues to examine the methodology, system architecture, and viability of using this solution in real-world scenarios.

## ii. Literature Review

In the last couple of years, facial emotion recognition and music recommendation systems have become very popular in affective computing and human-computer

interaction. Numerous research activities and projects have attempted to apply this interdisciplinary area for improving user experience through real-time emotional state-based music recommendation. The advent of sophisticated deep learning methods, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [4], has enabled more precise emotion recognition from facial expressions, which in turn enhances the performance of such recommendation systems.

One such project on GitHub, "Music Recommendation Based on Facial Emotion Recognition" [5], uses CNNs to categorize user emotions into seven groups—happiness, sadness, anger, surprise, fear, disgust, and neutrality—accurately at a rate of about 70%. Depending on the emotion detected, the system creates a customized playlist that seeks to match or enhance the user's existing mood. This method guarantees an adaptive and interactive listening experience, rendering music choice more intuitive and user-oriented.

Another major contribution in this field is the research paper "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation" [6]. This paper introduces a cross-platform music player that combines facial emotion recognition with a recommendation system to recommend songs according to the mood of the user. Through the use of affective computing methods, the system responds dynamically to its music recommendations, thus enhancing the relevance of recommended songs. This paper brings out the possibility of emotion-aware systems to enhance user interaction to be more seamless and engaging.

Moreover, the research "Music Recommendation Based on Face Emotion Recognition" discusses the feasibility of using facial expression analysis to generate emotion-driven playlists. The study shows that when a negative emotion, like sadness or anger, is identified, the system suggests uplifting or calming songs to improve the emotional state of the user. This piece of work highlights the larger potential of combining emotion recognition with music recommendation, and how such systems might be useful for mental health purposes by reducing stress and enhancing mood stability.

A number of other research projects have further developed this idea. For instance, in the paper "Affective Computing-Based Music Recommendation Using Deep Learning Techniques" [7] authors explain how emotion detection accuracy can be improved using deep learning models, especially CNN and Long Short-Term Memory (LSTM) networks [8]. The study highlights the need for large and properly annotated databases, like the FER-2013 [9] and AffectNet datasets [10], to train strong emotion recognition models. The research indicates that integrating CNNs with RNN-based architectures can drastically enhance the performance of emotion-driven music recommendation systems.

These series of works highlight the tremendous promise in bringing together facial emotion detection with music recommending systems. The advances in machine learning, especially affective computing, have opened up the way for more responsive and intuitive applications that respond to user emotions in real time. As research continues in this area, future systems are bound to use more modalities, including voice analysis and physiological signals (e.g., heart rate or EEG data), to improve the accuracy and effectiveness of

emotion-based music recommendations. The creation of these systems is potentially useful for a broad array of purposes, including entertainment and user interaction, as well as therapeutic and psychological interventions.

### iii. Problem Statement

Current music recommendation systems are weak in adapting to the real-time emotional state of the user. These systems largely rely on genres, listening history, or even manual input from the users, which may not reflect their immediate needs. Further, although there have been advancements in emotion detection technologies, their application in music recommendation remains underexplored. There is a lack of integrated systems that can seamlessly analyse facial expressions and provide personalized music suggestions. The challenge lies in creating a robust, scalable model that can classify emotions accurately while mapping them to appropriate music categories.

### iv. Proposed Solution

The proposed project introduces a facial detection-based music recommendation system. The solution relies on a CNN-based model to make facial emotion recognition possible. The model takes 48x48-pixel grayscale images and classifies the emotion of the face into one of seven categories: happiness, sadness, anger, surprise, fear, disgust, and neutrality. Once an emotion is sensed, it is mapped to a predetermined playlist of songs that represent the mood of the user. This system not only improves personalization but also enhances user engagement as it responds to real-time emotional cues.

## v. Methodology

The music recommendation system based on facial detection is a product with several components seamlessly integrated to recognize emotions from the face and recommend music tracks in response. It consists of four major components: the face detection module, the emotion database, the training process, and the music recommendation module. All of these components are vital to providing the system's accuracy and efficiency. The process (Figure 1) starts by taking the face image of the user, pre-processing, and emotion classification employing a Convolutional Neural Network (CNN) model. When the emotion is identified, it is translated to a pre-determined music genre, resulting in the suggestion of a suitable playlist that matches the emotional state of the user.

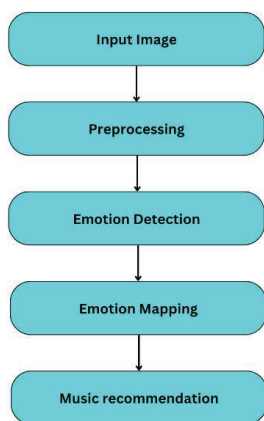


Figure 1: Flow Chart

The initial step in the system is face detection when the facial photo of the user is taken from a camera live. Alternatively, an image database can also act as an input source. The image taken would be of various qualities based on varying lighting effects and background points, so prior to

input in the emotion detector model, there is a requirement to optimize the image. Pre-processing here is an important operation as it serves to refine the image while minimizing unnecessary computational complexity. The image is initially converted to grayscale, discarding colour information but preserving necessary facial structures required for proper emotion recognition. It is then resized into a standard dimension of 48x48 pixels for compatibility with the CNN model. Secondly, pixel values are also rescaled between 0 to 1, normalizing the data to enhance training stability and convergence rate. Through this pre-processing, only the most pertinent facial features are extracted with minimal noise, thus making the system more effective for real-time emotion detection.

Subsequently, the pre-processed image is input into the emotion detection model, which utilizes a CNN architecture. The model learns to categorize facial expressions as one of seven pre-defined emotions: happiness, sadness, anger, surprise, fear, disgust, or neutrality. CNNs are well suited for this job since they are able to detect intricate facial features and patterns that differentiate between two different emotions. The architecture has several convolutional layers that extract low-level features like edges and textures, then deeper layers that identify more complex facial structures. The layers continually improve the features extracted, allowing the model to make very precise classifications. The CNN model also uses dropout layers to avoid overfitting by randomly disabling neurons while training to ensure that the system generalizes well to unknown images. The last layer of the model employs a SoftMax activation function [11] to provide probabilities for every emotion class so that the system can

identify the most probable emotion from the user's facial expression.

The performance of the emotion detection system largely relies on the emotion database, which is used as the basis for training the CNN model. This dataset consists of an exhaustive set of 35,000 labelled facial images, each belonging to one of the seven emotion classes. A balanced dataset is important to avoid overrepresentation of any specific emotion, which would otherwise result in biased prediction. To make the dataset more diverse and resilient, the dataset has images of people belonging to different age groups, genders, and ethnicity, enabling the model to generalize excellently across different populations. Further, data augmentation methods like horizontal flipping, random rotation, and small shifts are used to synthetically augment the dataset. These methods bring variations in the training set, replicating actual conditions, and minimizing overfitting. In exposing the model to various facial expressions under varying conditions, the system becomes more accurate in detecting emotions.

In the training process, the CNN model goes through a highly rigorous process to learn complex patterns linked with various emotions. The training is conducted through several epochs, and the model manipulates its internal parameters to decrease classification errors with each passing epoch. The convolutional layers retain significant features, whereas max-pooling layers decrease spatial dimensions without losing significant information. Dense, fully connected layers combine the feature-extracted inputs with ReLU activation functions [12] introduced to provide non-linearity to enhance decision-making. The final SoftMax output produces probability distributions over the

seven emotion classes that identify the best label for the input image. The training of the model is strictly monitored on the basis of evaluation metrics such as accuracy, validation loss, and the weighted F1 measure, which measures the performance of the model when dealing with imbalanced classes. The model is trained for 50 epochs at a batch size of 64, improving gradually in accuracy while maintaining resistance to overfitting.

After the system identifies an emotion successfully, it moves to the music recommendation stage, where the identified emotion is translated into a related music category. The translation is based on established psychological research correlating emotions with musical tastes. For example, happiness is linked to active and uplifting music that strengthens positive emotions [13], whereas sadness is paired with slow and comforting melodies that give emotional solace. In the same way, anger is coupled with soothing and stress-reducing music to calm the user down, while fear is matched with comforting and peaceful music to provide a feeling of security. Surprise is complemented by dynamic and energetic tracks that capture the state of increased alertness, whereas disgust is balanced by neutral and inspirational songs to redirect the user's attention. Neutral emotions are suggested a general playlist that provides a balanced listening experience.

In order to provide a varied and stimulating music choice, the system combines with a music library that is carefully curated and includes a diverse collection of songs from various genres. The recommendation algorithm fetches tracks according to the identified emotion and displays a custom playlist to the user. Subsequent improvements might include the ability to use user preferences and

ratings to make recommendations more adaptive as time passes.

### vi. Result and discussion

The performance of the model is evaluated based on the confusion matrix (Figure 2), which shows class-wise accuracy and scope for improvement in emotion classification. The confusion matrix indicates how accurately the model separates different emotions. The best accuracy is seen in the "happy" emotion classification with 1468 samples being correctly classified, while emotions like "disgust" have high misclassification due to few samples and overlap with other emotions. The model also has fair success in emotion prediction for "neutral" and "sad" emotions but performs poorly in distinguishing between close emotions like "fear" and "surprise," resulting in some misclassification.

The graphs for training accuracy and loss (Figure 3) give insight into the process of learning of the model through several epochs. The training accuracy is steadily increasing, meaning that the model is learning well. The validation accuracy fluctuates, which can mean that the model could be sensitive to changes in the validation data. The training loss plot illustrates a gradual decrease, ensuring that the model is reducing errors with the passage of time. Yet, the validation loss plot indicates fluctuations, suggesting overfitting or inconsistent generalization performance on unseen data.



Figure 2: Confusion matrix

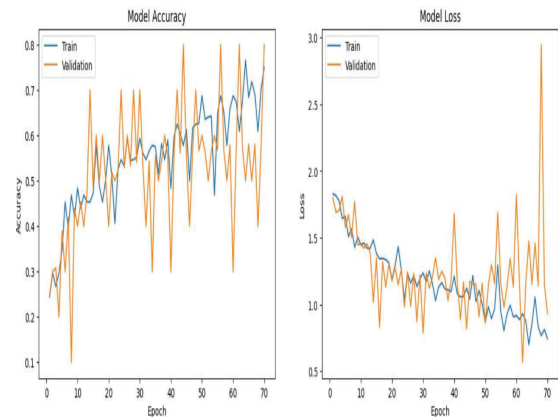


Figure 3: Accuracy and Loss

Real-time performance assessment was ensured by checking the model's capacity to identify and classify emotions in various situations. Figures 4, 5, and 6 present the model's performance in identifying emotions from human facial expressions. Figure 4 depicts the proper identification of a "sad" expression, whereas Figure 5 shows the model correctly detecting an "angry" expression. Likewise, Figure 6 illustrates the model properly detecting a "happy" emotion. All these outcomes support the model's usability in real-life applications and establish its efficiency in processing real-time facial inputs.

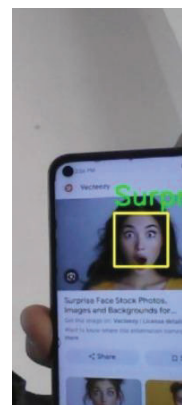


Figure 4

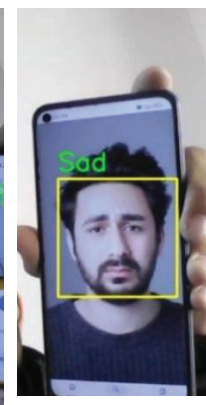


Figure 5

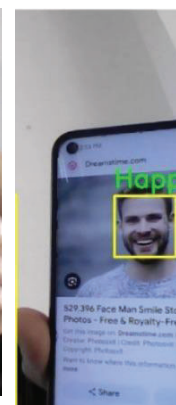


Figure 6

The integration of the emotion detection model with the music recommendation system was tested to ensure its operability. Figure 7 shows the graphical user interface (GUI) of the system developed, in which the emotion detected affects the choice of music tracks. The system is able to play songs that match the identified emotion effectively, showing a good connection between facial expression analysis and music playing. The interface supports playing, pausing, shuffling, and changing tracks, making the overall user experience better. The application of this system demonstrates the prospect of emotion-based personal content recommendation in entertainment usage.



Figure 7: Music Player

Generally, the model makes sensible predictions with satisfactory accuracy and has encouraging real-time response. There are still difficulties in differentiating similar emotions and alleviating validation loss oscillation. Future work can involve increasing the dataset, improving model structure, and utilizing advanced methods like attention mechanisms to sharpen classification accuracy and generalization capability.

## vii. Conclusion

The facial detection-based music recommendation system improves the user experience using real-time emotional detection along with music recommendations customized to users. Through smooth analysis of facial expressions, the system ensures that the music chosen suits the user's prevailing emotional state, creating a more engaging and immersive listening experience. Integrating deep learning methods with image processing techniques enables effective and efficient recognition of emotions, making the recommendation mechanism dynamic and extremely responsive. The research also highlights the efficiency of affective computing in enhancing human-computer interaction and showcases its potential for designing more intuitive, emotionally intelligent, adaptive technologies.

## viii. Future Scope

Future direction of this project involves some main areas for improvement. Firstly, increasing the dataset with richer facial expressions across different populations will enhance the model's accuracy and generality. Secondly, combining advanced deep learning algorithms like transformer-based models will possibly improve emotion recognition efficiency. Thirdly, real-time optimization via lightweight models will enable efficient deployment on mobile and embedded devices. In addition, the inclusion of user feedback mechanisms will further enhance the music recommendation process to provide more personalized and context-sensitive suggestions. Lastly, extending the system to other multimedia content, like video recommendations or ambient lighting control, could also enhance user experience and interaction.

## References

- [1] Karpouzis, K., & Kollias, S. (2008). Facial Animation and Affective Human-Computer Interaction. In B. Furht (Ed.), *Encyclopedia of Multimedia* (pp. 250–251)
- [2] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4), 541–551.
- [3] Brightwater Care Group. (2018, December 5). 6 reasons why music improves wellbeing.
- [4] Github Dhruba59, Music Recommendation based on Facial Emotion Recognition (2021)
- [5] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar (2017 International Conference on Wireless Communications, WiSPNET), Smart music player integrating facial emotion recognition and music mood recommendation
- [6] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536
- [7] Rushabh Chheda, Dhruv Bohara, Rishikesh Shetty, Siddharth Trivedi, Ruhina Karani (International Conference on Machine Learning and Data Engineering; *Procedia Computer Science* 218 2023, 383-392), Affective Computing-Based Music Recommendation Using Deep Learning Techniques
- [8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780
- [9] Goodfellow, I., Erhan, D., Carrier, P. L., Courville, A., Mirza, M., Hamner, B., & Bengio, Y. (2013). Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing* (pp. 117-124). Springer
- [10] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1), 18-31
- [11] Boltzmann, L. (1868). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte*, 58, 517–560
- [12] Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814.
- [13] Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20), 11818–11823 & Koelsch, S., Fritz, T., v. Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. *Human Brain Mapping*, 27(3), 239–250