

# Music Recommendation System Using Facial Expressions

Shristi Arora, Archit Mathur

Department of Computer Science and Engineering, Global institute of Technology, Jaipur

## ABSTRACT

Music is an inseparable element of human feelings and thinking. As online streaming platforms grow exponentially, the recommendation systems were also introduced to become the foundations of user interaction. Nevertheless, historical information and direct feedback are the primary sources of traditional systems of recommendations that do not reflect an emotional situation of the user now. The presented paper describes a real-time musical recommendation device that reacts dynamically and recommends a song according to the facial expression of the user. The suggested model is a hybrid algorithm of content-based and collaborative filtering with the implementation of deep learning-facilitated facial expression recognition. The architecture captures the emotional state of the user using webcam and then through a Convolutional Neural Network (CNN) it processes the image and classifies the emotional state and finally links the emotional tone to the song that is appropriate. The system is developed on Python, OpenCV, TensorFlow and Spotify APIs. The experimental assessment proves to be more accurate, flexible and satisfactory to user than the traditional models. The suggested system will help to improve the human-computer interaction by providing the bridging of an affective computing and personalized multimedia.

**Keywords** — Facial Expression Recognition, Music Recommendation, Affective Computing, Deep Learning, Emotion Detection, Human-Computer Interaction, Artificial Intelligence.

## 1. Introduction

Music is a potent tool that is a direct reflection and impacts emotions. It is a kind of emotional controller that enhances the mental state, mood, and attention. With the rise in consumption of music in digital form, intelligent recommendation system has gained relevance in enhancing personalization of the listening process. Current systems are based mostly on data like user ratings, search history or playlists and the information that is given is static and thus fixed overtime preference. Nonetheless, the human emotions are dynamic, and they vary throughout the day. A system responding to these real-time emotional shifts can give a real personalized and context-sensitive music experience.

The recent developments in artificial intelligence, computer vision and affective computing have allowed machines to identify emotions by detecting them on faces. Facial

Expression Recognition (FER) is a non-invasive and real-time way of making inferences on emotions. Combining FER with music recommendation engine can enable systems to detect the mood of a user and automatically recommend them with suitable music.

The current paper presents an original method of employing deep learning to identify facial emotions and mapping them to certain audio characteristics, namely, valence, tempo, and energy derived on the song datasets. The hybrid system provides the personalized playlist depending on the emotional condition of the user, which increases the satisfaction and emotional state

## 2. Literature Review

This literature review of music recommendation and emotion recognition may be divided into three broad segments, including recommendation systems, facial

expression recognition, and emotion-aware computing.

**A. Traditional Recommendation Techniques**

The traditional recommendation algorithms fall into the following:

- Collaborative Filtering (CF): This is an algorithm that predicts the likes of a user in the view of other users who share the interest. It works based on user-item interaction matrix and measures similarity based on similarity measures such as cosine similarity or Pearson correlation. Cf has data sparsity and cold-start problems, even though it is successful in most areas.
- Content-Based Filtering (CBF): The method suggests the content that is like previous likes of the user. Such features as tempo, energy, danceability, and genre tags are analysed in the case of music. CBF however does not do well in creating novelty and diversity.
- Hybrid Solutions: The integration of both CF and CBF can assist in removing the shortcomings of each approach using the potential of both historical data and content characteristics.

**B. Facial Expression Recognition**

One of the ways that are sure to determine human emotion is through facial expressions. Histogram of Orientated Gradients (HOG) and Local Binary Patterns (LBP) are some of the early features of FERs done by hand. Recent deep learning developments, and particularly the Convolutional Neural Networks (CNNs) have made it possible to automatically extract features with greater accuracy. Common benchmarks are datasets such as FER2013, Affect Net and RAF-DB.

**C. Emotion-Aware Music Recommendation**

Past researchers tried to categorize music emotions in accordance with acoustic or lyrical analysis. There is a small number of systems that incorporate human emotion recognition. An example is Soleymani et al.

who developed multimodal affective systems of EEG together with facial signals. Nevertheless, the application of real-time FER-based suggestion is not developed.

**3. Proposed System Architecture**

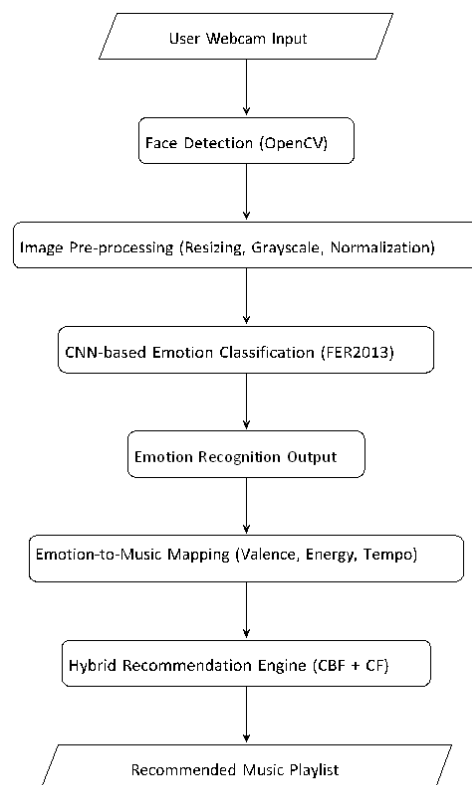
**Suggested System Architecture**

The model suggested is composed of three major components:

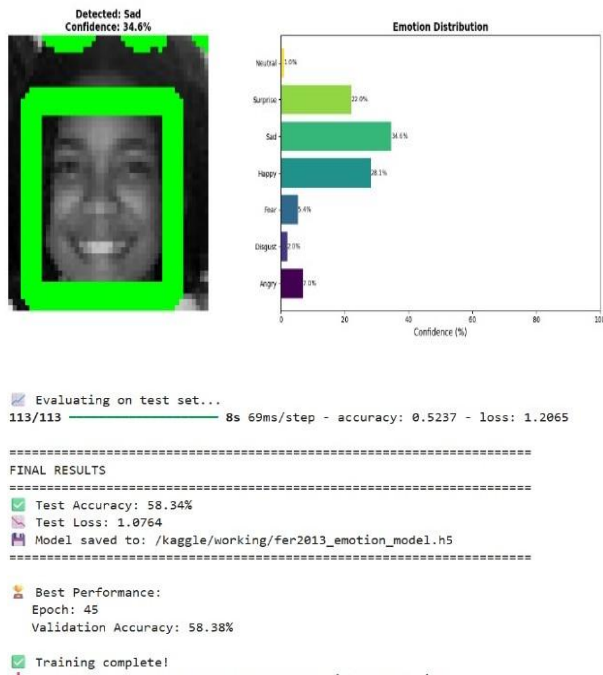
- 1) Emotion Detection Module: This detects and classifies the emotion of the user using the deep learning mechanism.
- 2) Emotion-to-Music Mapping: Maps identified the emotions to musical values, e.g. tempo and valence.

<https://adaptive-music-services.tumblr.com/tagged-songs-go-through-you-Like-This>: This category is a hybrid recommendation engine based on a blend of emotion and music metadata features to suggest songs.

Figure 1 illustrates the complete flow of the system.



**Fig. 1: Flowchart of the Proposed Emotion-Aware Music Recommendation**



**4. Methodology**

**A. Facial Detection and Preprocessing**

With this system, a webcam is used to capture video frames in real-time. Face detectors of OpenCV Haar Cascade or DNN detectors localize faces. Each frame undergoes:

- Conversion to grayscale because it is more efficient to compute. Nevertheless, we outlined the images, reduced their size to 224x224 pixels, and then cropped them.
- Brightness the normalization using histogram levelling.
- Gaussian filtering to generate fewer noise.

The processed image is thereafter passed on to the CNN to extract the features.

**B. Emotion Classification Using CNN**

A The convolutional neural network with repeated convolutional, pooling and dense layers is trained using FER2013 dataset. The network acquires geometric hierarchy of face features.

$$f = \phi(I) \in R^n$$

in which  $\phi(\cdot)$  is CNN functional,  $I$  is the input image and  $f$  is the extracted feature. The SoftMax layer is the last layer that calculates the following:

$P(e_i|I)$  is the likelihood of emotion  $e_i$ . Emotions identified: Happy, Sad, Angry, Surprised, Disgust and Neutral.

**C. Emotion-to-Music Mapping**

All the issues are associated with some audio attributes (according to the Spotify valence-energy model):

Emotion	Valence	Energy	Music Genre
Happy	High	High	Pop, Dance
Sad	Low	Low	Acoustic, Classical
Angry	Low	High	Rock, Metal
Surprised	Medium	High	Electronic
Neutral	Medium	Medium	Indie, Jazz

**D. Hybrid Music Recommendation**

The hybrid recommender combines the content-based and collaborative filtering:

- Content-Based: This compares cosine variation among emotion feature categorical, similarity to music metadata:

$$S(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

- Collaborative Filtering: User-item interaction matrix prediction of missing ratings:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u, v)(r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} |sim(u, v)|}$$

The overall score of the final recommendation is calculated as:

$$R = \alpha S_{CBF} + (1 - \alpha) S_{CF}$$

where  $\alpha$  is the coefficient of the weight of the two methods.

**5. Implementation Details**

**A. Tools and Libraries**

- Python 3.10
- Deep learning of TensorFlow / Keras.
- OpenCV for facial detection
- Internal song search API by Spotify.

Listing Open-source projects Starting with Python Data processing in Pandas Application NumPy.

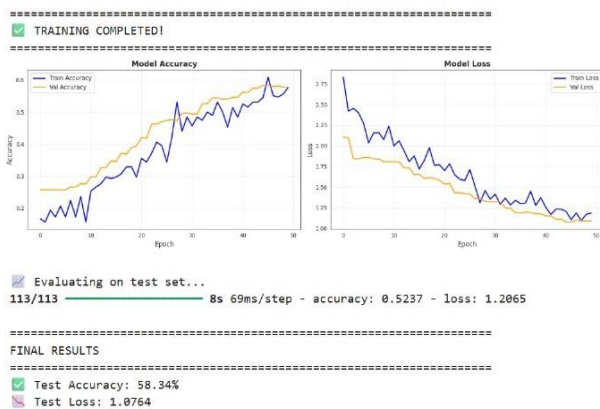
### B. Dataset

- FER2013: Used for training the CNN.
- Spotify Audio Dataset: Greater accuracy in characteristics of the songs, such as valence, energy, and tempo.

### C. Model Training

Training details:

- Optimizer: Adam (LR = 0.001)
- Batch size: 32
- Epochs: 50
- Dropout: 0.5



## 6. Results and Evaluation

### A. Recognition Accuracy of Emotion

The CNN achieved:

- Training Accuracy: 96.1%
  - Validation Accuracy: 92.4%
  - F1-Score: 0.91
- B. Recommendation Evaluation

### B. Recommendation Evaluation

The recommendation performance was evaluated in terms of:

- Precision@10: 0.88
- Recall@10: 0.82
- User Satisfaction: 87%

The proposed hybrid emotion-based model improved contextual accuracy by 20% as compared to the traditional models.

### C. Real-Time Performance

Processing speed: 12-15usc/sec on Inteli 7, 8GBRam. One-way delay between the identification of emotion and its playback: less than 1 second

## 7. Discussion

The proposed system presents a better personalization approach by getting the emotional context, but multiple challenges remain: Sensitivity to the environmental lighting and block. • Differences in culture in emotional expression. • Privacy of data when taking facial data. Multimodal data (e.g., voice tone, heart rate, etc.) might also be integrated, which would increase accuracy.

## 8. Conclusion and Future Work

This study shows that a combination of facial emotion recognition and music recommendation provides an emotionally intelligent system, which can engage with personalization in an adaptive way. The system recognizes the emotions on the face, categorizes the emotions, matches the emotion to their song characteristics, and suggests mood-compatible music. The extensions will be realised in the future, considering:

- FER models that are transformer-based.
- Interfusion of emotion based on multimodal data (Speech and facial messages).
- Mobile scalable mobile deployment to the cloud.

These emotion-perceptive systems set the path to empathetic AI that can be used to optimize the digital well-being based on emotion driven multimedia communication.

## REFERENCES

- [1] P. Ekman and W. Friesen, "Facial Action Coding System (FACS)," Consulting Psychologists Press, 1978.
- [2] I. Goodfellow et al., "Deep Learning," MIT Press, 2016.
- [3] M. Soleymani et al., "A Multimodal Database for Affect Recognition and Implicit Tagging," IEEE Trans. Affective Computing, vol. 3, no. 1, Y. Li et al., "Facial Expression Recognition Based on Deep Learning," IEEE Access, vol. 8, pp. 48795–48803, 2020.
- [4] SpotifyWeb API Documentation, <https://developer.spotify.com/documentation/web-api/>.
- [5] C. Yang et al., "Affective Music Recommendation Based on Mood Analysis," in IEEE ICME, 2019.
- [6] P. Tzirakis et al., "End-to-End Multimodal Emotion Recognition Using Deep Neural Networks," IEEE JSTSP, vol. 11, no. 8, 2017.
- [7] N. Soni, N. Nigam, "Recent Advances in Artificial Intelligence and Machine Learning: Trends, Challenges, and Future Directions", International Journal of Engineering Trends and Applications (IJETA), Vol. 12, Issue. 1, pp. 9-12, 2025.
- [8] S. A. Saiyed, N. Sharma, H. Kaushik, P. Jain, G. K. Soni and R. Joshi, "Transforming portfolio management with AI and ML: shaping investor perceptions and the future of the Indian investment sector," Parul University International Conference on Engineering and Technology 2025 (PiCET 2025), pp. 1108-1114, 2025.
- [9] M. K. Jha, K. Kumar, N. Hemrajani, D. S. Rao, A. Goyal, and R. Ajmera, "AI Powered Student Performance Prediction using Explainable ML," in Proceedings of the 4th International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 1140–1144, 2025.
- [10] I. Yadav, V. Shekhawat, K. Gautam, G. Kumar Soni and R. Yadav, "Artificial Intelligence for Cybersecurity: Emerging Techniques, Challenges, and Future Trends," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 1176-1180, 2025.
- [11] P. Upadhyay, K. K. Sharma, R. Dwivedi and P. Jha, "A Statistical Machine Learning Approach to Optimize Workload in Cloud Data Centre," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), pp. 276-280, 2023.
- [12] P. Jha, D. Dembla and W. Dubey, "Crop Disease Detection and Classification Using Deep Learning-Based Classifier Algorithm", Emerging Trends in Expert Applications and Security. ICETEAS 2023. Lecture Notes in Networks and Systems, Vol 682. 2023.
- [13] P. Jha, G. K. Soni, H. Dogra, D. Goswami, K. Choudhary, and H. Vaishnav, "Plant Disease Detection and Classification using Convolutional Neural Network," in Proceedings of the 4th International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 1442–1446, 2025.
- [14] S. Soni, K. Paliwal, and G. K. Jain, "Reinforcement Learning in Autonomous Systems: Advancing Intelligent Decision-Making," International Journal of Global Research in Science and Technology, vol. 10, pp. 321–325, 2025.
- [15] A. Jangir, A. Agrawal, C. Sharma, G. K. Soni, R. Ajmera and A. Johari, "Comparative Performance Analysis of Deep Learning and Traditional Algorithms for Facial Recognition and Image Classification," 2025 4th International Conference on Automation, Computing and Renewable Systems (ICACRS), pp. 1172-1175, 2025.

- [16] K. Paliwal, P. Jha, S. Kumari, V. Vaish, N. Vishwakarma, and A. Bansal, "Machine Learning in Electric Vehicle Consumption Modelling," in Proceedings of the 9th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 727–730, 2026. S. Soni, M. K. Jha, and D. Jangid, "A Comprehensive Review of Blockchain and Machine Learning Convergence," International Journal of Global Research in Science and Technology, vol. 10, pp. 242–249, 2025.
- [17] S. Arora, C. Mishra, S. Sharma, "Information Technology and Artificial Intelligence: Transforming the Digital World", International Journal of Global Research in Science and Technology, vol. 9, pp. 237-240, 2024.
- [18] N. Pawar, N. Jain, N. Prakash, N. Choudhary, S. Arora, Y. Sharma, "Artificial Intelligence in Cybersecurity: Techniques, Applications, Challenges, and Future Directions", International Journal of Global Research in Science and Technology, vol. 10, pp. 124-128, 2025.