© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718460

Emotion Based Music Recommendation System Using OpenCV and Machine Learning Technique

DR. M. SARASWATHI[1], Y. HARSHA VARDHAN[2], CHL PRAMAD[3]
[1] *Assistant Professor, Department of CSE, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya University, India*
[2,3] *UG Student, Department of CSE, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya University, India*

Abstract- ***Emotion plays a major role in influencing human music preferences. Traditional music recommendation systems mainly rely on listening history, ratings, and playlists, but they often fail to understand the user's real-time emotional state. This paper presents an Emotion Based Music Recommendation System using OpenCV, MediaPipe, Machine Learning, and Streamlit that detects user emotions through facial expressions and recommends suitable music accordingly. The system captures facial images through a webcam, extracts facial and hand landmarks using MediaPipe Holistic, and predicts emotions using a trained Convolutional Neural Network model. Based on the predicted emotion and user preferences such as language and singer, the system automatically recommends songs using YouTube automation and local music playback. The proposed system demonstrates the practical application of effective computing and intelligent recommendation systems in entertainment platforms.***

Keywords: Recommendation System, Machine Learning, Neural Network, OpenCV, MediaPipe.

I. INTRODUCTION

Music has a strong connection with human emotions and mental state. Most existing music recommendation systems rely on previous listening history, user ratings, playlists, or trending songs. However, these systems do not consider the user's current emotional condition while generating recommendations. As emotions greatly influence music preferences, static recommendation systems often fail to provide emotionally relevant suggestions.To overcome this limitation, the proposed Emotion Based Music Recommendation System uses facial expression analysis to detect emotions in real time and recommend suitable songs automatically. The system integrates OpenCV for image processing, MediaPipe for landmark extraction, TensorFlow/Keras for deep learning-based emotion classification, and Streamlit foraninteractiveuserinterface.

The project combines computer vision, machine learning, and web automation into a single lightweight application capable of running on normal systems without high-end GPU requirements. The system supports both online playback using YouTube automation and offline playback using locally stored songs.

II. SCOPE

This paper focuses on the design and implementation of an intelligent emotion-aware music recommendation system. The scope includes facial image acquisition using webcam input, landmark extraction using MediaPipe Holistic Landmarker, feature vector generation, CNN-based emotion classification, and music recommendation based on the predicted emotional state. The system also supports user preferences such as language and singer selection to improve personalization.

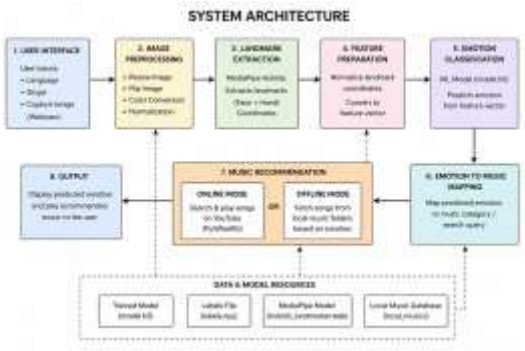
The project demonstrates the practical use of affective computing in entertainment systems and highlights how machine learning can improve user interaction and engagement. The system is designed as a lightweight prototype capable of real-time execution without expensive computational hardware.

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718460

III. METHODOLOGY

The development of the proposed system followed multiple stages including data collection, preprocessing, model training, implementation, and testing.

Process Workflow



The system architecture of the Emotion based music recommendation outlines a seamless pipeline where raw user input slowly transforms into a song that fits the user's mood. It starts at the User Interface, where the user provides simple inputs like language, preferred singer, and captures an image through the webcam. This is the only place where the user interacts directly with the system, and everything that follows is automated processing.

The captured image then moves into the Image Preprocessing stage. Here, the system cleans and standardizes the image by resizing it, flipping it for correct orientation, and converting color formats. This step is important because machine learning models are sensitive to inconsistencies. If you feed messy input, you get unreliable output, so this stage ensures the image is uniform and ready for analysis.

Next, the system performs Landmark Extraction using MediaPipe. Instead of analyzing the entire image pixel by pixel, it picks only important points like positions on the face and hands. These points act like a skeleton or blueprint of the expression. This approach is efficient because it reduces unnecessary data while preserving the essential structure needed to understand emotions.

After extracting these points, the system moves to Feature Preparation. The raw coordinates are converted into a structured numerical format called a feature vector. During this step, normalization is applied so that the model does not get confused by where the face appears on the screen or how big it looks. This makes the system more stable and consistent across different users and environments.

The processed feature vector is then passed to the Emotion Classification stage, where the trained machine learning model (model.h5) predicts the user's emotion. It does not just give a direct answer but calculates probabilities for each possible emotion and selects the one with the highest confidence. This is the decision-making core of the system.

Once the emotion is identified, it moves to Emotion to Music Mapping. Here, the detected emotion is translated into a meaningful music context. For example, a "happy" emotion might be converted into an "upbeat pop" query. This phase is crucial because the model understands emotions, not songs, so this mapping bridges that gap.

The system then enters the Music Recommendation stage, which works in two modes. In the online mode, it uses YouTube to search and play songs automatically based on the generated query. In the offline mode, it retrieves songs from local folders organized by emotion and selects one randomly. This dual approach ensures that the system works even without internet access.

Finally, the result is delivered through the Output stage, where the user sees the detected emotion and hears the recommended music. This completes the full cycle from input to output, creating an interactive experience.

Supporting all these steps in the background is the Data and Model Resources section. This includes the trained machine learning model, the labels file that defines emotion categories, the MediaPipe model used for landmark detection, and the local music database. These components act as the backbone of the system, enabling each stage to function properly.

Overall, the diagram represents a well-structured pipeline where each block has a specific responsibility, and data flows step by step from raw input to meaningful output. The strength of this

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718460

architecture lies in its modular design, allowing each component to be improved or replaced independently without affecting the entire system.

3.1 MODULE DESCRIPTION

The proposed Emotion Based Music Recommendation System contains multiple integrated modules working together to provide real-time emotion-aware recommendations. Initially, facial emotion datasets containing different human emotional expressions were collected and organized into emotion categories such as Happy, Sad, Angry, Neutral, Surprise, and Rock. MediaPipe Holistic Landmarker was used to extract facial and hand landmarks from the images. The extracted landmark coordinates were normalized and converted into numerical feature vectors.

A Convolutional Neural Network (CNN) model was trained using supervised learning techniques. During training, the model learned emotional patterns from labeled landmark feature vectors. The trained model was saved in the form of model.h5 while the corresponding emotion labels were stored in labels.npy.

For implementation, Streamlit was used to create the web interface. OpenCV handled image processing tasks such as resizing, flipping, and color conversion. Based on the predicted emotion and user-selected preferences, the system generated music recommendations through YouTube automation and local music playback.

The user first enters preferred language and singer details through the Streamlit interface. After entering the details, the webcam captures the user's facial image. OpenCV processes the captured frame and converts it into the required format for landmarkextraction.

MediaPipe Holistic Landmarker detects face and hand landmarks from the image. These landmarks represent important facial and hand coordinate points that help identify emotional expressions. The coordinates are normalized and converted into feature vectors.

The generated feature vector is passed into the trained CNN model. The model predicts the emotional class with the highest probability score. The detected emotion is then mapped to predefined music categories. For example, Happy emotion is mapped to upbeat songs, while Sad emotion is mapped to slow acoustic music.

The system then generates a search query using the emotion, language, and singer preference. PyWhatKit is used to automatically open YouTube recommendations. The system also supports local playback by selecting random songs from local emotion-based folders.

IV. RESULTS AND DISCUSSION

The developed prototype successfully detected emotions in real time and generated suitable music recommendations according to the user's emotional state. The Streamlit interface provided smooth user interaction and webcam functionality. MediaPipe Holistic accurately extracted facial and hand landmarks, while the CNN model predicted emotions with satisfactory performance under normal lighting conditions.

The system successfully recommended different categories of music based on emotions such as Happy, Sad, Angry, Neutral, Surprise, and Rock. Online playback through YouTube automation and offline playback through local music folders both functioned correctly during testing.

The landmark-based approach reduced computational complexity compared to raw image-based CNN methods while maintaining efficient real-time performance.

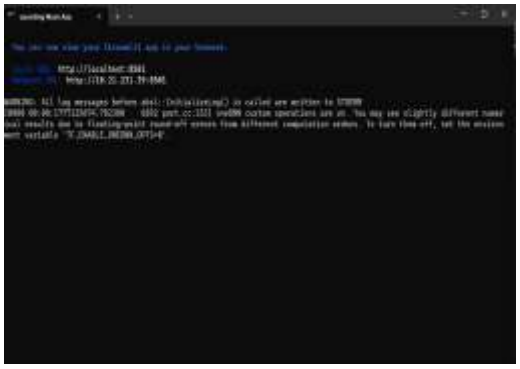


Fig 2 Launching the streamlit app

© MAY 2026 | IRE Journals | Volume 9 Issue 11 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I11-1718460

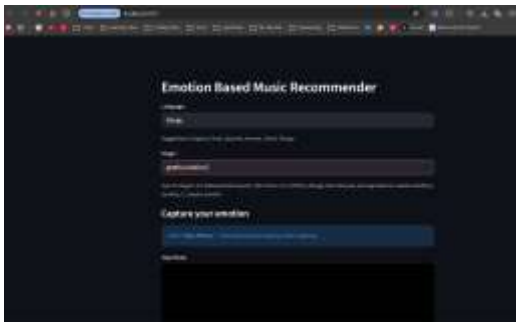


Fig 3 Main page

Fig 4: Recommends song based on emotion

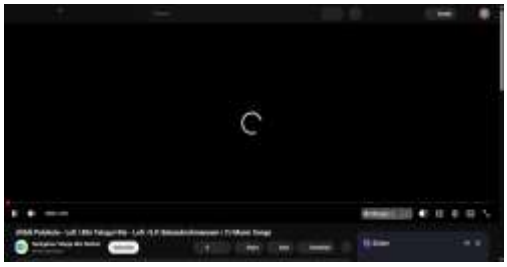


Fig 5: Redirects you to youtube based on emotion from local player

The developed prototype successfully detected emotions in real time and generated suitable music recommendations according to the user's emotional state. The Streamlit interface provided smooth user interaction and webcam functionality. MediaPipe Holistic accurately extracted facial and hand landmarks, while the CNN model predicted emotions with satisfactory performance under normal lighting conditions.

The system successfully recommended different categories of music based on emotions such as Happy, Sad, Angry, Neutral, Surprise, and Rock. Online playback through YouTube automation and offline playback through local music folders both functioned correctly during testing.

The landmark-based approach reduced computational complexity compared to raw image-based CNN methods while maintaining efficient real-time performance.

V. CONCLUSION & FUTURE SCOPE

The Emotion Based Music Recommendation System effectively combines computer vision, machine learning, and web technologies to provide real-time emotion-based music recommendations. By analyzing facial expressions, the system improves personalization and user experience in entertainment applications. Future enhancements may include integration with Spotify and JioSaavn APIs, multimodal emotion recognition using voice and text, continuous emotion tracking, mobile app deployment, and advanced deep learning models for better accuracy and performance.

REFERENCES

[1] A. Kumar and S. Patel, "Emotion-aware music recommendation using deep learning techniques," IEEE Access, vol. 11, pp. 105432–105445, Oct. 2023.

[2] R. Sharma, P. Gupta, and V. Singh, "Facial expression recognition using convolutional neural networks for real-time applications," IEEE Transactions on Affective Computing, vol. 14, no. 4, pp. 2256–2268, Dec. 2023.

[3] M. Chen, Y. Li, and H. Zhang, "Deep learning-based emotion recognition from facial landmarks," Pattern Recognition Letters, vol. 168, pp. 45–52, Jan. 2023.

[4] S. Roy and A. Banerjee, "Music recommendation systems based on emotional context: A survey," ACM Computing Surveys, vol. 55, no. 6, pp. 1–35, Nov. 2022.

[5] T. Nguyen and D. Lee, "Real-time facial emotion detection using lightweight CNN models," IEEE Sensors Journal, vol. 23, no. 5, pp. 6789–6798, Mar. 2023.