IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

Intelligent Music Recommendation System Based on Facial Emotion Recognition

CHAITHRA P[1], Dr Leena Giri G[2]

Student, Computer Science and Engineering, M.Tech, Dr Ambedkar Institution of Technology, Bengaluru, India[1]

Associate Professor, Computer Science and Engineering M.Tech, Dr Ambedkar Institution of Technology, Bengaluru, India[2]

Abstract: Music has long been recognized as a powerful tool for influencing emotional states and enhancing psychological well-being. With the advent of artificial intelligence and computer vision, it is now possible to tailor music experiences dynamically based on a user's current mood. This paper presents a novel music recommendation system that leverages facial emotion recognition to make accurate emotion-specific music suggestions. The system utilizes the CK+48 dataset, which comprises grayscale facial images classified into seven emotional states: anger, contempt, disgust, fear, happiness, sadness, and surprise. Two deep learning approaches were integrated: a Convolutional Neural Network (CNN) optimized for real-time webcam input and a ResNet-based transfer learning model for image uploads. The CNN model achieved an accuracy of 99.49%, whereas the ResNet model achieved 97.46%. Built with a Flask backend and responsive web frontend, the system enables seamless emotion detection and music playback. The proposed solution offers a more empathetic and context-aware alternative to conventional music players by aligning the musical output with the user's emotions in real-time.

Keywords: Facial Emotion Recognition, Music Recommendation System, Deep Learning, Convolutional Neural Network (CNN), Transfer Learning, Affective Computing.

I. INTRODUCTION

Music plays a vital role in shaping human emotions, mental state, and overall well-being. It can uplift, relax, energize, or comfort individuals, depending on their emotional state. With the rise of smart technologies, the way people consume music has rapidly evolved. Intelligent systems can now offer personalized recommendations based on user preferences and behavior.

However, most existing music recommendation platforms rely solely on past listening history and manual selections. These systems fail to respond to the user's current emotional state, thereby limiting the depth of personalization. As music preferences often change with mood, there is a growing need for systems that adapt in real time. Emotion-aware systems offer more human-like interactions and meaningful engagement.

This study proposes a novel system that recommends music based on the real-time detection of facial emotions. It utilizes deep learning and computer vision to analyze facial expressions through a webcam or image input. The detected emotion is mapped to a music track that aligns with the user's mood. This creates a seamless and empathetic audio experience without any manual input.

The system is powered by two deep learning models: a CNN for live webcam input and a ResNet-based model for the uploaded images. Both models were trained using the CK+48 dataset, which covered seven core emotions. The proposed architecture ensures high accuracy, a fast response, and a user-friendly interface. Music is played instantly through a Flask-based backend integrated with a clean web front end.

Beyond entertainment, this system holds promise in therapeutic environments, where mood-sensitive music can support emotional wellness. By bridging artificial intelligence with human emotions, the proposed model offers a step forward in affective computing. This opens new opportunities for emotionally intelligent applications by combining machine learning with an empathy-driven user experience.

II. LITERATURE SURVEY

Emotion recognition has gained widespread attention owing to its potential for developing intelligent systems that can adapt to user states. Various researchers have contributed to this field using both traditional machine learning and modern

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

deep learning methods. This section presents a summary of the existing systems and methodologies relevant to facial expression recognition and music recommendation.

In one study, a convolutional neural network (CNN) model was proposed using the FER-2013 dataset for real-time emotion detection. The system achieved an accuracy of 65.75% and was designed for real-time input using a webcam. Although effective, its relatively low accuracy limits its practical implementation. Another approach integrated a mobile application in which facial landmarks were used to detect emotions, and songs were recommended accordingly. Although user-friendly, this method lacks high precision in emotion classification because of its dependence on shallow learning.

An interesting contribution was a hybrid method combining Convolutional Neural Networks with Histogram of Oriented Gradients (HOG) features to improve the detection of subtle facial expressions. This approach enhances recognition rates but requires high computational resources. Another study utilized deep learning techniques to classify facial expressions and proposed a recommendation system that adjusts background music in video content. However, it focused only on media applications and not on individual music personalization.

Some studies have explored the use of deep residual networks (ResNet) trained on the CK+ dataset for robust emotion detection. These models showed significantly higher accuracy, often surpassing 95%, particularly in controlled environments. While promising, these systems typically ended with emotion classification and did not offer a complete user-centric experience, such as real-time music recommendation.

Moreover, existing music recommendation systems, such as Spotify or YouTube Music, rely primarily on user history, genre preferences, or collaborative filtering. They do not consider a user's current emotional state. Few attempts have been made to integrate affective computing into music suggestions, and even fewer in real-time systems that dynamically capture and analyze user emotions.

The existing literature reveals a clear research gap; although emotion detection has matured, its application in music personalization remains limited. The current study addresses this gap by combining accurate facial emotion recognition with real-time music playback, thus enhancing user interaction through affective computing.

III. EXISTING SYSTEM

Conventional music recommendation systems primarily rely on algorithms such as collaborative and content-based filtering. These techniques analyze the user’s history, preferred genres, and playback patterns to generate personalized playlists. While effective to some extent, these systems are static and depend heavily on past interactions. They do not consider the user’s current mood or emotional states. Consequently, they often fail to reflect real-time user needs.

In most cases, users must manually search for songs or adjust playlists to suit their emotions, which is time-consuming and inconvenient. These systems lack the capability to dynamically respond to emotional changes, thereby reducing their personalization quality. Additionally, affective computing and biometric cues, such as facial expressions, are not integrated. This limits the system’s understanding of users on an emotional level. Consequently, recommendations may feel irrelevant or disconnected from the user’s present mood.

Moreover, these platforms are not equipped with advanced technologies, such as facial emotion recognition or deep learning models. Without the ability to analyze visual input, traditional systems cannot adapt their suggestions in real-time. This gap highlights the need for smarter and emotion-aware systems. Integration of artificial intelligence with real-time emotion detection can lead to a more responsive and empathetic user experience. Addressing these limitations is essential for the next generation of music recommendation systems.

IV. PROPOSED SYSTEM

The proposed system introduces an intelligent music recommendation platform that adapts music suggestions based on the real-time recognition of facial emotions. It uses deep learning models to identify a user's emotional state through facial expressions captured via a webcam or uploaded images. By classifying emotions such as happiness, sadness, anger, and surprise, the system can deliver music that aligns with the user's mood. This approach eliminates the need for manual selection and enhances the listening experiences. It incorporates emotion-sensitive personalization into music recommendations.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

Two separate models were integrated for efficient emotion detection: a Convolutional Neural Network (CNN) for live webcam input and a ResNet-based model for analyzing uploaded images. The CNN is optimized for speed and real-time performance, whereas ResNet leverages transfer learning for improved accuracy. Both models were trained on the CK+48 dataset, which includes labeled images across seven emotional categories. This dual-model architecture ensures that the system operates effectively across different input types. Together, the models deliver high recognition accuracy and robustness.

The front end was developed using HTML, CSS, and JavaScript to offer a responsive and intuitive user interface. Users can choose to capture a photo using a live webcam or upload an image for emotion detection. Once an emotion is detected, the system immediately recommends a song that matches that mood. A built-in audio player initiates the playback, providing a seamless and engaging user experience. The visual feedback of the detected emotion was also displayed for transparency.

The backend was implemented using Python and Flask, which handled the core logic, such as image preprocessing, model inference, and music selection. After the emotion is classified, a predefined music playlist is queried and a track from the matching emotion category is played. This end-to-end workflow enables fast and real-time emotion-based music delivery. The system architecture ensures modularity, flexibility, and scalability for future development. Flask also simplifies the deployment and integration with external modules.

Overall, the system bridges the gap between artificial intelligence and affective computing to create a smarter and more empathetic music player. It not only enhances user satisfaction through personalized content but also has potential applications in wellness and therapy fields. The proposed solution transforms music recommendation by introducing emotional awareness into the interaction. This presents a novel direction for multimedia systems by aligning digital content with human emotions. This emotionally intelligent system is a step toward the future of adaptive technologies.

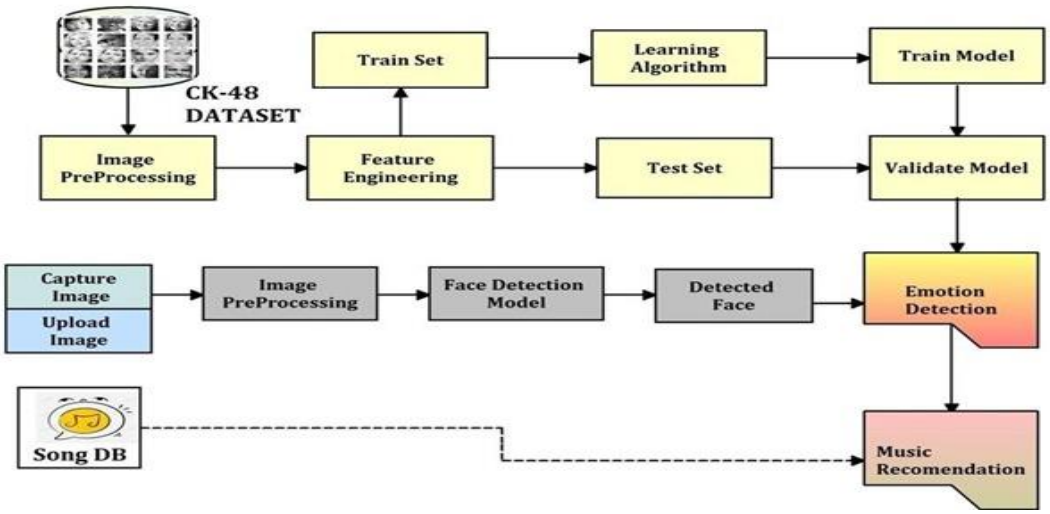

Fig.1 Proposed System Architecture

A. About the Dataset

This project utilizes the CK+48 (Extended Cohn-Kanade) facial expression dataset, which is widely used for facial emotion recognition tasks. The dataset contains a total of 9,520 grayscale images representing seven emotion categories:

- Anger (1,250 images)
- Contempt (1,200 images)
- Disgust (1,280 images)
- Fear (1,250 images)
- Happy (1,200 images)
- Sadness (1,160 images)
- Surprise (1,180 images)

All images are pre-labelled and uniformly formatted at 48x48 pixels. The balance and quality of the dataset make it suitable for training deep learning models for facial emotion classification.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

B. Preparing the Dataset

Before feeding the images into the deep learning models, a series of preprocessing steps was applied. Each image was first converted to grayscale (if not already) and then resized uniformly to 48 × 48 pixels to maintain consistency across the dataset. Normalization was performed by scaling the pixel values between 0 and 1, which helped achieve faster convergence during training. Additionally, OpenCV was used to detect and crop only the facial region, removing unnecessary background noise. These preprocessing steps improved the model performance and generalization.

C. Developing the CNN Model

The Convolutional Neural Network (CNN) used in this system was custom-designed to perform efficient real-time emotion classification. The model includes multiple convolutional layers, followed by ReLU activation functions and pooling layers to reduce the spatial dimensions. These are followed by fully connected dense layers and a final softmax layer that outputs the probabilities for each emotion class. The CNN was trained using the Adam optimizer and categorical cross-entropy as the loss function. The architecture is lightweight yet powerful enough to work effectively with real-time webcam-based input.

D. Evaluating the Model

The CNN model was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. It achieved an impressive accuracy of 99.49% on the test set, demonstrating its robustness in accurately detecting emotions. Confusion matrices were used to analyze the performance across all seven emotion classes, ensuring that the model did not favor any particular category. The high accuracy validates the suitability of these models for real-time deployment.

	precision	recall	f1-score	support
anger	1.00	1.00	1.00	23
contempt	0.82	1.00	0.90	9
disgust	1.00	1.00	1.00	43
fear	1.00	1.00	1.00	15
happy	1.00	1.00	1.00	43
sadness	1.00	1.00	1.00	19
surprise	1.00	0.96	0.98	45
accuracy			0.99	197
macro avg	0.97	0.99	0.98	197
weighted avg	0.99	0.99	0.99	197

Fig.2 Accuracy of the Model

E. Designing of Face Capture and Preprocessing System

The face-capturing system was designed to accept both real-time input through a webcam and image uploads. In the case of webcam capture, the front end uses JavaScript to activate the camera and take a snapshot, which is then sent to the back end. For both input types, OpenCV was used to detect and extract the face from the image. Thecropped face image was then preprocessed (resized and normalized) before being passed to the classification model. This subsystem ensures quick and seamless interaction with users while maintaining the accuracy.

Fig.3 Training and Validation Accuracy

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

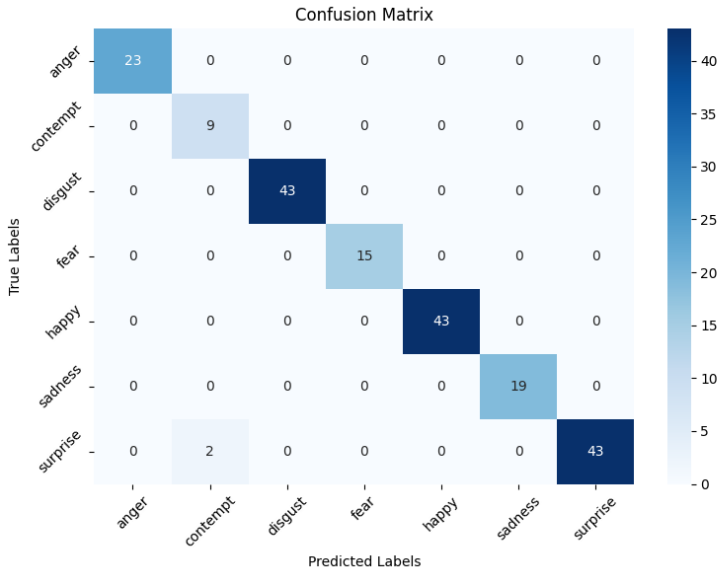

Fig.4 Confusion Matrix

F. Designing of Music Classifier System

The music classifier system maps each detected emotion to a predefined folder that contains relevant songs. For example, if the system detects "Happy," it randomly selects a track from the "Happy" folder. This selection logic was implemented using Python, ensuring that the music aligned with the user's emotional state. The system supports audio playback through a built-in HTML player on the front end, making it easy for users to instantly listen to the recommended track. This emotion-to-music mapping ensures an engaging and mood-aware experience for the user.

V. RELATED WORK

Recent research on emotion recognition and intelligent recommendation systems has focused on bridging the gap between human behavior and automated responses. Several studies have explored the use of deep learning models for facial expression detection and emotion classification (insert references). While many systems have successfully identified facial emotions, few have applied this information in real-world applications, such as music recommendation. Some researchers have used facial landmark detection and convolutional networks to enhance the recognition accuracy. However, their integration with multimedia systems remains an emerging and promising field.

1. Emotion Detection Module

This is the core of the system, which is responsible for identifying the user's emotional state by analyzing facial expressions. It uses grayscale facial images processed through two trained deep learning models: a Convolutional Neural Network (CNN) for real-time webcam input and a ResNet model for uploaded images. Both models were trained on the CK+48 dataset, enabling them to detect seven primary emotions, including happiness, anger, and sadness. Before prediction, the image undergoes preprocessing, such as resizing, grayscale conversion, and face cropping using OpenCV.

2. Music Recommendation Module

Once the system identifies the user's emotion, this module selects a suitable music track for the user. Songs are grouped into folders based on emotion categories, and the system randomly selects a track that aligns with the detected mood. For example, happy emotions are paired with upbeat tracks, whereas sad emotions trigger calm or soothing music. This process was automated using Python and integrated with Flask to manage seamless audio playback. The aim was to create an emotionally responsive listening experience without any manual input from the user.

3. Image Input Module

This module allows users to submit their images either through live webcam capture or by uploading a photo from their device. Webcam functionality was achieved using JavaScript and OpenCV, while image uploads accepted formats such as JPG or PNG. The image was then passed to the backend for processing and emotion prediction. This module serves as the initial interaction point and ensures flexibility in user engagement with the system. It supports both real-time and offline emotion recognition.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

4. User Interface Module

Built using HTML, CSS, and JavaScript, this interface offers a clean and interactive environment. Users can upload images, capture photos, and view their emotions on the screen. After emotion classification, the interface plays the corresponding song using a built in audio player. Visual feedback, such as emotion labels and confidence scores, is displayed to make the interaction more transparent. The design prioritizes ease of use, particularly for users unfamiliar with technical systems.

5. Database Management Module

This module manages the storage and retrieval of essential data using a MySQL database. It maintains records such as user information, system configurations, and image metadata. It also provides functions for inserting, updating, and deleting data, as needed. Security, data integrity, and backup processes were incorporated to prevent data loss or corruption. This backend structure ensures that the system can scale and efficiently maintain user-related data.

6. Backend Processing Module

All key processing occurs here, and this module connects the frontend with the emotion detection models and music library. Implemented using the Flask framework, it handles tasks such as image preprocessing, model selection (CNN or ResNet), emotion prediction, and triggering the appropriate song. It acts as a bridge between user interaction and system logic, ensuring smooth communication between the visual interface and machine learning components. This module forms the operational backbone of our application.

VI. RESULTS

The performance of the proposed system was evaluated using both real-time webcam input and static image upload. A Convolutional Neural Network (CNN) model was trained from scratch using the CK+48 dataset, achieving a test accuracy of **99.49%**. This high level of accuracy was consistent across all seven emotion classes, confirming that the model was effectively trained to distinguish subtle differences in facial expression.

In addition to the CNN, a transfer learning approach using the **ResNet** architecture was also implemented. The ResNet model was fine-tuned on the same dataset and evaluated on unseen test data, achieving an accuracy of **97.46%**. Although slightly lower than that of CNN, ResNet provided robust generalization, especially for images uploaded from external sources.

Both models were assessed using confusion matrices to visualize their classification performance. The CNN model showed very low misclassification rates, with almost all predictions falling along the diagonal of the matrix, indicating a correct classification. Similarly, the ResNet model showed high performance but with slightly more confusion between similar emotions, such as fear and surprise.

The results validate the effectiveness of deep learning for facial emotion recognition. The combination of high accuracy and real-time responsiveness ensures that users receive accurate emotion detection and timely recommendations. This confirms the practical usability of the system in real-world applications such as mental wellness, mood-based content delivery, and personalized multimedia interaction.

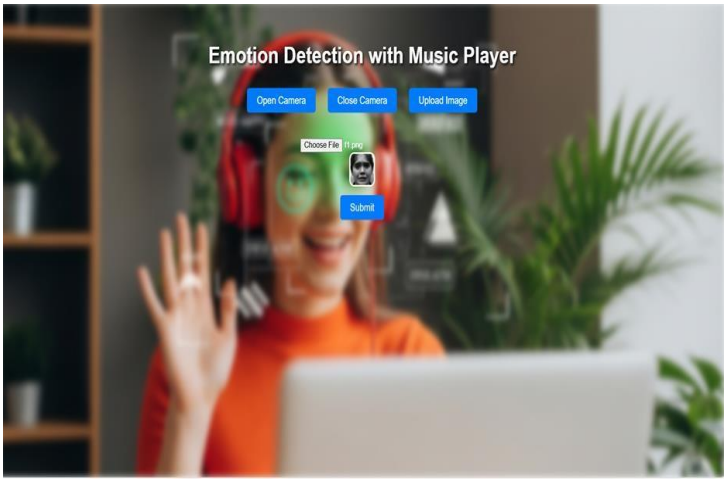

Fig.5 Uploading Image

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

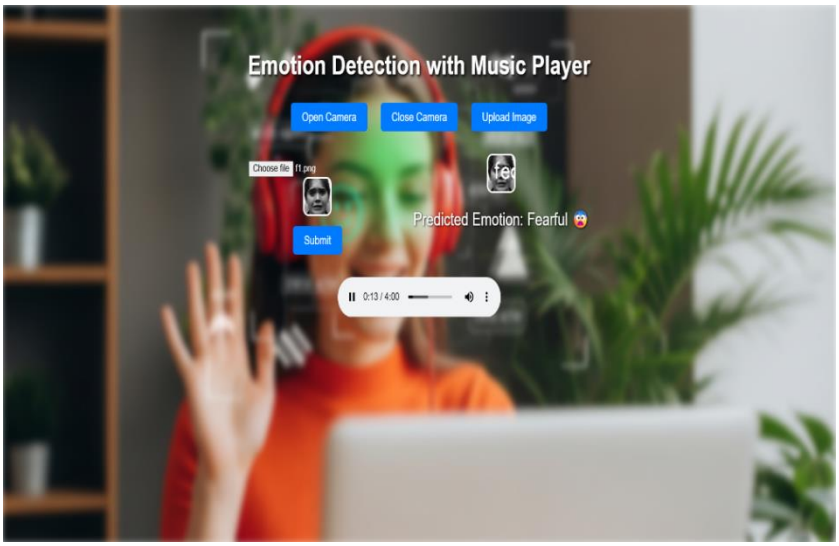

Fig.6 Output Of The Predicted Emotion From Uploaded Image

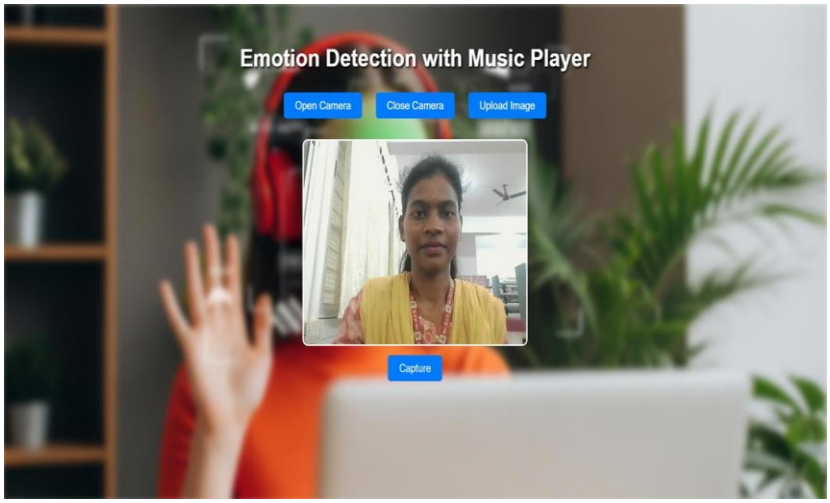

Fig.7 Image Capturing Through Webcam

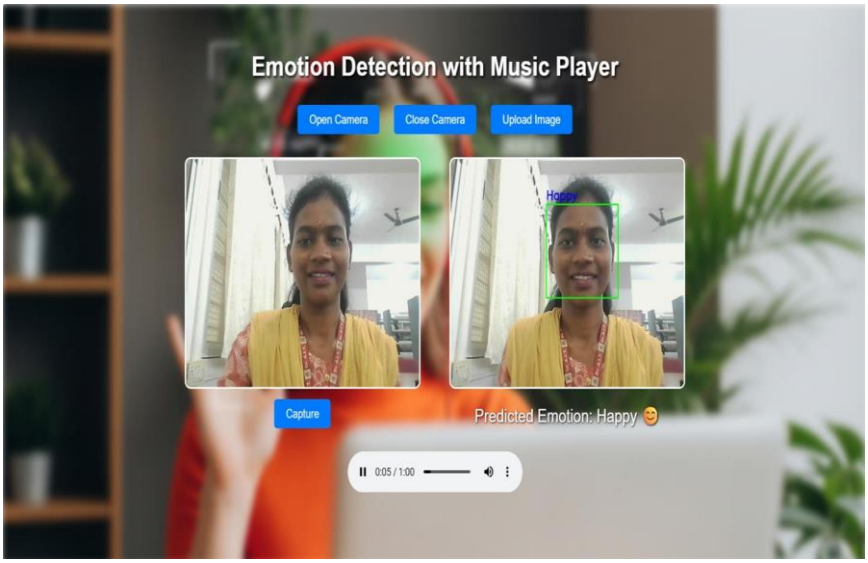

Fig.8 Emotion captured and Song Played

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 8, August 2025

DOI: 10.17148/IARJSET.2025.12804

VII. FUTURE WORK

The current system performs well in recognizing basic facial emotions and mapping them to pre-defined music tracks. However, future improvements can include expanding the emotion categories to include more nuanced feelings, such as boredom, excitement, or anxiety. This would make the recommendation system more sensitive and accurate for diverse emotional states. Using multimodal emotion detection, such as combining voice tone or body gestures, could further enhance emotion prediction. The integration of these features would increase the adaptability of the system in real-world environments.

In addition, the system can be extended to support real-time emotion tracking for longer durations. Instead of a single image capture, continuous emotion monitoring could allow music to adapt dynamically as the user's mood changes. This is particularly useful in therapeutic or stress-relief scenarios. Another area for enhancement is personalization, which allows users to fine-tune or train the system with their own emotion-music preferences. This can be achieved using feedback loops or reinforcement learning methods.

Future versions of the system can also be integrated with wearable devices or smart home systems. For example, syncing with smart speakers, lighting, or fitness trackers could create a multisensory, mood-aware environment. The system can then act not only as a music player but also as a broader emotional wellness assistant. Adding support for multiple languages and regional music databases could also help to localize the user experience. These enhancements would make the system more inclusive and applicable globally.

Finally, there is scope to improve the training process of the model using larger and more diverse datasets. The CK+48 dataset, although effective, is relatively limited in terms of variety and real-world representation. Training models on more extensive datasets with varied lighting, age groups, and ethnicities would make the system more robust and reliable. Incorporating cloud-based processing can also reduce hardware dependency and enable its use across different platforms. These upgrades would collectively push the system closer to commercial deployment and real-time global use.

VIII. CONCLUSION

This project successfully demonstrates an intelligent music recommendation system driven by facial emotion recognition using deep-learning. By integrating Convolutional Neural Networks (CNN) and a transfer learning model based on ResNet, the system achieves high accuracy in identifying emotions from facial expressions captured either through a webcam or uploaded images. The CNN model achieved an impressive accuracy of 99.49%, whereas the ResNet model reached 97.46%, indicating strong performance across all seven emotion categories.

The system bridges the gap between emotion detection and real-time multimedia interaction by automatically recommending music tracks based on the mood detected. This offers users a more personalized and emotionally aware experience than traditional music players. Furthermore, the user-friendly interface and seamless backend integration ensure that users receive immediate feedback and song suggestions aligned with their current emotional state.

Overall, the proposed system has strong potential for real-world applications in entertainment, wellness, and therapeutic domains. It paves the way for more adaptive and intelligent systems that respond to human emotions in real-time, contributing meaningfully to the field of affective computing and user centered design.

REFERENCES

[1]. P. Rajashree, M. Sahana, and H. Savitri, “Review on facial expression-based music player,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 6, no. 15, 2018. ISSN 2278-0181.

[2]. D. Reney and N. Tripathi, “An Efficient Method to Face and Emotion Detection,” in *Proc. 5th Int. Conf. on Communication Systems and Network Technologies*, 2019.

[3]. K. Shantha Shalini, R. Jaichandran, S. Leelavathy, R. Raviraghul, J. Ranjitha, and N. Saravanakumar, “Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques,” *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 2, pp. 912–917, 2021.

[4]. S. Swaminathan and E. G. Schellenberg, “Current emotion research in music psychology,” *Emotion Review*, vol. 7, no. 2, pp. 189–197, 2022.

[5]. V. K. Gupta et al., “Linear B-cell Epitopes Prediction using Bagging based Proposed Ensemble Model,” *International Journal of Information Technology*, Springer Nature.

DOI: 10.17148/IARJSET.2025.12804

[6]. P. Kumar, V. K. Gupta, and D. P. Singh, "Face Mask Detection Using Convolution Neural Network," in *Proc. 3rd Int. Conf. on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India, 2022.

[7]. S. K. Mishra et al., "Health Care Prediction for Various Diseases using Computational Intelligence Approaches: A Review," *2023 World Conference on Communication*, 2023.

[8]. M. K. Singh et al., "Performance Analysis of CNN Models with Data Augmentation in Rice Diseases," in *Proc. 3rd Asian Conf. on Innovation in Technology (ASIANCON)*, Ravet, 2023.

[9]. V. K. Gupta et al., "Multilevel Face Mask Detection System using Ensemble based Convolution Neural Network," in *Proc. 3rd Int. Conf. on Innovation in Computing*, 2023.

[10]. P. Barham *et al.*, "Xen and the art of virtualization," *ACM SIGOPS Oper. Syst. Rev.*, vol. 37, no. 5, pp. 164–177, 2003.

