iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XII Month of publication: December 2025

DOI: https://doi.org/10.22214/ijraset.2025.75155

Survey Paper on Music Recommendation System using Facial Recognition and Deep Learning

Namrata Gawande[1], Prithviraj Rathod[2], Irfan Tadvi[3], Om Takale[4], Navraj Yadav[5]
Department of Computer Engineering, PCCOE, Pune, India

Abstract: This study introduces an emotion-aware music recommendation system that adapts song suggestions in real time based on a user's facial expressions. The proposed framework integrates deep learning techniques for emotion recognition with machine learning algorithms for personalized music selection. Facial emotions are identified using a lightweight MobileNetV2-based convolutional neural network (CNN), ensuring fast and ac- curate detection. The recommendation process employs content- based filtering to align users' emotional states with suitable music tracks. To maintain user privacy, the system implements secure data handling techniques, including encryption and authentication. Overall, the platform delivers responsive, mood-driven music recommendations with high accuracy and minimal latency, making it suitable for entertainment, emotional therapy, and adaptive music applications.

I. INTRODUCTION

Music has long been recognized as a powerful medium for expressing and influencing human emotions. With rapid progress in artificial intelligence (AI) and deep learning, mod- ern recommendation systems have transitioned from static, genre-based playlists to intelligent, emotion-aware platforms. Conventional recommendation methods typically depend on user preferences, play history, or collaborative filtering approaches. However, these systems often fail to reflect users' instantaneous emotional states, limiting personalization in real time.

To address this gap, the present study introduces a real- time music recommendation model that interprets a user's mood through facial expressions and suggests songs that align with the detected emotion. The system integrates facial recognition, deep learning, and secure data handling into a unified framework to ensure both accuracy and privacy. A lightweight convolutional neural network (CNN), based on the MobileNetV2 architecture, is employed for efficient facial emotion recognition. This design achieves high performance even on devices with restricted computational capacity.

After detecting emotional categories such as happiness, sad- ness, anger, fear, disgust, surprise, and neutrality, the system uses content-based filtering to map each emotion to suitable tracks or playlists. Additionally, data security is reinforced through encryption, user authentication, and access control mechanisms. The ultimate objective is to develop a scalable and responsive application that provides a smooth, real-time interaction experience while safeguarding user data.

II. LITERATURE SURVEY

Recent developments in deep learning and computer vision have greatly advanced emotion-aware music recommendation systems. Numerous researchers have explored diverse methods that integrate facial expression analysis with intelligent recommendation models to improve personalization, engagement, and user satisfaction.

Mariappan et al. [1] introduced FaceFetch, one of the earliest multimedia content recommendation systems driven by facial emotions. Their approach combined facial recognition using ProASM and SVM classifiers to enable real-time emotion-based content retrieval. Florence and Uma [2] further enhanced this concept by designing a webcam-assisted emo- tion detection and music suggestion framework utilizing image segmentation techniques, demonstrating potential benefits in stress management and emotion-aware entertainment.

Gadagkar et al. [3] implemented a MobileNetV2-based CNN capable of classifying emotions such as happy, sad, angry, and neutral, and integrated it with the Spotify API for automated playlist creation. Building on this work, Nalini and Pinninti [4] optimized the emotion-recognition pipeline to reduce computational complexity and latency, particularly for edge deployment.

Prasad et al. [5] developed a MATLAB-based emotion recognition system that produced song recommendations ac- cording to detected moods, achieving notable accuracy under controlled conditions. Joshi et al. [6] expanded on this by using OpenCV and CNN for real-time facial expression detection and music suggestion through a lightweight mobile interface, enhancing portability and responsiveness.

Maddala et al. [8] designed a CNN model that recognized seven fundamental emotions and synchronized recommenda- tions with song rhythm and mood. Mahir et al. [9] emphasized dataset expansion and TensorFlow-based training to im- prove classification accuracy and scalability, while Shaik and Bhutada [10] demonstrated that ResNet architectures offered better real-time stability than standard CNN models. Pathak et al. [11] explored Music Emotion Recognition (MER) by analyzing spectral and rhythmic features to establish a correlation between audio signals and user emotion, improv- ing personalization accuracy. Deore [12] employed a hybrid CNN–LSTM structure to enhance temporal understanding of expressions and refine emotion-driven recommendations. Vijayalakshmi et al. [13] developed an AI-assisted system that leveraged facial features for personalized music suggestions, yielding high user satisfaction levels.

Bagadi [14] proposed an end-to-end, deep learning-based real-time emotion detection and playlist generation model that offered seamless user interaction. Tsouvalas et al. [15] addressed data security by presenting a semi-supervised feder- ated learning framework for emotion recognition from speech, maintaining privacy during distributed model training.

Rani et al. [16] introduced a modified CNN architecture that minimized false predictions in facial emotion classification, improving reliability. Selvi and Aakash [17] presented Emo- Tune, an advanced emotion detection and recommendation system built on MobileNetV3, showcasing the benefits of lightweight models for embedded devices. Adru and John- son [18] proposed a multimodal fusion approach combining facial and textual emotion cues for more accurate and context- aware music selection.

Parashakthi and Savithri [19] implemented a dynamic recommendation model that updated playlists based on evolving emotional states, while Mishra et al. [20] developed a deep neural network to detect subtle emotions and associate them with audio mood metrics. Rajesh et al. [21] presented an automated, end-to-end facial recognition-based song recom- mendation model, reflecting the growing adoption of emotion- driven automation in music systems.

Nguyen [22] proposed an emotion-based song recommender integrating convolutional networks with fuzzy logic, offering fine-tuned playlist emotional balance. Ashwini et al. [23] utilized YOLO v11 for efficient face detection and emotion classification, achieving faster inference on embedded plat- forms.

In parallel, several studies have addressed the inclusion of cybersecurity mechanisms within emotion-aware AI systems. Patel et al. [24] introduced secure data handling techniques, including encryption and protected workflows, for multimedia applications. Likewise, Sharma et al. [25] explored privacy-preserving deep learning frameworks that safeguard user data without compromising inference speed.

Collectively, these studies indicate that integrating facial emotion recognition with deep learning-based recommenda- tion and security measures significantly enhances personalization, system reliability, and user trust. This foundation supports the development of the proposed system, which employs MobileNetV2 for emotion detection, content-based filtering for song mapping, and encryption to maintain data confidentiality.

III. METHODOLOGY

The proposed emotion-aware music recommendation frame- work is organized into three major components, as illustrated below:

1) Facial Emotion Recognition using MobileNetV2: The system captures real-time facial images through a we- bcam. Each frame is preprocessed (resized, normal- ized, and filtered) before being fed into a trained Mo- bileNetV2 model to classify emotions such as happiness, sadness, anger, and others.
2) Music Recommendation through Content-Based Fil- tering: After identifying the user’s emotional state, the recommendation engine retrieves songs with closely matching mood characteristics derived from metadata such as tempo, rhythm, and lyrical sentiment.
3) Secure API and GUI Integration: All interactions be- tween the client and server are encrypted using AES-256 to ensure data confidentiality. A Flask-based backend coordinates emotion recognition and music recommen- dation, while a minimal web or mobile interface presents the output in real time.

A. Facial Emotion Recognition

The facial emotion detection module leverages a fine-tuned MobileNetV2 convolutional neural network (CNN) trained on benchmark datasets such as FER2013 and RAF-DB. Mo- bileNetV2 was selected for its superior trade-off between accuracy and computational efficiency, making it ideal for low- latency, real-time applications.

$$E = f(I) = \arg \max_{e_i \in Emotions} P(e_i|I) \qquad (1)$$

Here, I denotes the input image frame, and e_i represents the most probable emotion class predicted by the model. The system continuously processes live frames, ensuring seamless and adaptive emotion tracking.

B. Music Recommendation

The music recommendation module maps detected emotions to relevant songs using content-based filtering. Each track in the database is represented by a multidimensional feature vector containing parameters such as tempo, genre, and mood descriptors. Cosine similarity is used to measure the closeness between the user's current emotion vector and each song's feature vector:

$$Sim(a, b) = \frac{a \cdot b}{||a|| \ ||b||} \qquad (2)$$

A higher similarity value indicates a stronger correlation between the identified emotion and the song's emotional tone, allowing the system to generate an optimal playlist.

C. Cybersecurity Integration

To protect user data and ensure secure system communication, multiple cybersecurity layers are incorporated:

- AES-256 encryption is employed to secure data transmis- sion.
- JWT-based authentication validates user sessions and prevents unauthorized access.
- Role-based access control defines permissions and en- forces secure user operations.

These measures collectively maintain data confidentiality and protect the system against common security threats.

D. System Architecture

The complete architecture of the proposed model comprises four interconnected layers:

1) Frontend: Developed using React.js to handle webcam input and dynamically display the recommended songs.
2) Backend: Implemented with Flask/Django REST APIs that manage emotion detection, song retrieval, and sys- tem logic.
3) Database: A PostgreSQL database stores user profiles, emotional states, and detailed song metadata for efficient querying.
4) Cloud Infrastructure: The system is deployed on AWS to ensure scalability, with GPU acceleration for faster CNN inference and parallel data processing.

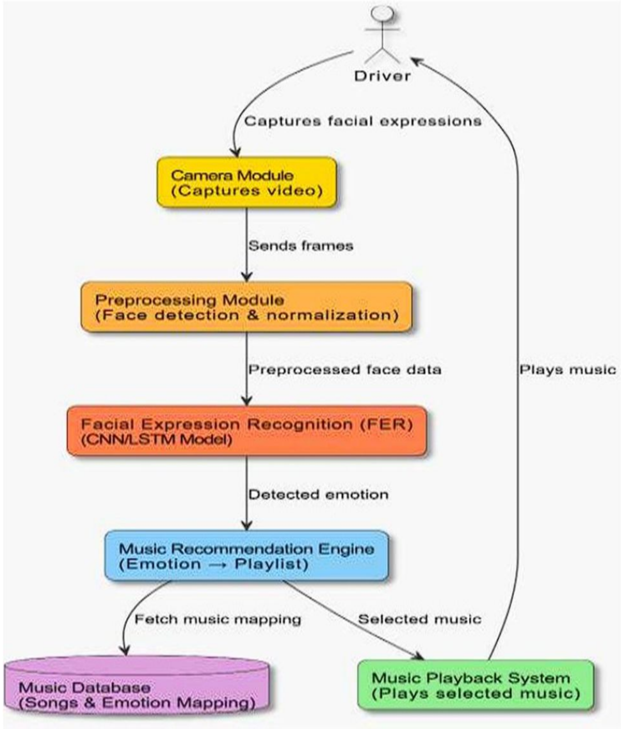


Fig. 1: Overview of the Proposed System Architecture

IV. RESULTS

A. Data Collection and Preprocessing

The proposed system was trained and evaluated using the FER2013 dataset, which contains approximately 35,887 grayscale facial images of size 48×48 pixels. These images are categorized into seven emotion labels: happy, sad, angry, fear, disgust, surprise, and neutral. To improve model generalization and prevent overfitting, several data augmentation techniques such as random rotation, horizontal flipping, and scaling were applied. Each image was normalized to a pixel intensity range between 0 and 1, ensuring consistent illumination and contrast across the dataset. The augmented dataset thus provided better diversity and robustness for emotion recognition under varying facial expressions and lighting conditions.

B. Training the MobileNetV2 Model

The emotion recognition model was built upon the Mo- bileNetV2 architecture, which was initially pretrained on the ImageNet dataset. Transfer learning was then applied to fine- tune the network for facial emotion recognition. MobileNetV2 utilizes depthwise separable convolutions, significantly reduc- ing the number of parameters and computational load while retaining strong feature extraction performance.

The convolutional transformation can be expressed as:

$$y = \sigma(W_f * x + b) \qquad (3)$$

where W_f denotes the learned convolutional filter weights, x is the input feature map, b represents the bias term, and σ is the nonlinear activation function. Through iterative fine- tuning, the model achieved stable convergence with improved emotion classification accuracy and reduced latency, making it suitable for real-time applications.

C. Music Dataset and Recommendation Engine

For the music recommendation module, a curated dataset was created containing metadata and acoustic features such as tempo, rhythm intensity, energy level, and valence. These attributes were extracted using the Spotify Web API and the Librosa Python library. Each song is represented as a multi-dimensional feature vector, and content-based filtering is employed to match the detected user emotion with an appropriate group of songs. For instance, when the detected mood is "happy," the system prioritizes songs with higher valence and faster tempo, whereas "sad" moods correspond to lower-energy, slower tracks. This approach allows the recommendation engine to adapt dynamically to emotional variations while maintaining high personalization accuracy.

D. Graphical User Interface

A lightweight and interactive web application was devel- oped using the Flask framework to demonstrate the function- ality of the proposed system. The interface includes:

1) Real-time emotion recognition through webcam input.
2) Immediate generation of music recommendations aligned with the detected emotion.
3) User feedback options that enable system learning and continuous improvement of future recommendations.

The GUI emphasizes simplicity and responsiveness, pro- viding smooth user interaction and low latency during real- time inference. The complete setup ensures that users receive emotionally adaptive music suggestions with minimal computational overhead.

V. COMPARATIVE STUDY

To assess the effectiveness of the proposed emotion recognition model, a comparative study was carried out using multiple convolutional neural network (CNN) architectures on the FER2013 dataset. The evaluated models included VGG16, ResNet50, and the optimized MobileNetV2-based network. Each model was trained using identical preprocessing tech- niques, dataset partitions, and hyperparameters to ensure a fair and unbiased comparison. The assessment focused on key per- formance metrics such as accuracy, precision, recall, F1-score, and average system latency, which collectively determine the suitability of each model for real-time emotion recognition and music recommendation applications.

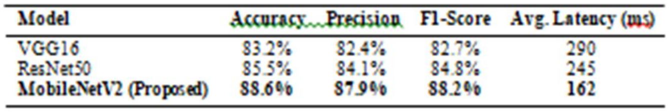
A. Performance Metrics

The following parameters were employed to evaluate model performance:

1) Accuracy: Reflects the proportion of correctly identified emotional states across all test samples.
2) Precision: Indicates how often predicted emotional labels were correct, measuring prediction reliability.

3) Recall: Represents the ability of the model to detect all relevant instances of a particular emotion.
4) F1-Score: The harmonic mean of precision and recall, providing a balanced view of classification performance.
5) Response Time: Denotes the average time taken by the model to process an image and produce the associated music recommendation.

TABLE I: Comparison of CNN Architectures on FER2013 Dataset



Model	Accuracy	Precision	F1-Score	Avg. Latency (ms)
VGG16	83.2%	82.4%	82.7%	290
ResNet50	85.5%	84.1%	84.8%	245
MobileNetV2 (Proposed)	**88.6%**	**87.9%**	**88.2%**	**162**

The results summarized in Table 1 reveal that the Mo- bileNetV2 model achieved the highest performance across all measured parameters. The proposed network attained an accuracy of 88.6% and an average latency of just 162 millisec- onds, outperforming both VGG16 and ResNet50. The superior performance of MobileNetV2 can be attributed to its use of depthwise separable convolutions and inverted residual blocks, which significantly reduce computational complexity while maintaining robust feature extraction. This efficient design makes it particularly well-suited for deployment on low-power and real-time devices.

In contrast, the VGG16 model, though widely recognized for its accuracy in general image classification, contains a substantially higher number of parameters. This leads to slower execution times and greater GPU memory consump- tion. ResNet50 introduces residual connections that enhance generalization but increases model depth, adding to its com- putational cost. MobileNetV2, however, strikes a balance by combining high accuracy with low inference latency, enabling smooth and efficient real-time performance.

The improved precision and F1-score achieved by the pro- posed model demonstrate its consistent ability to distinguish between closely related emotional states, such as neutral and sad expressions. Moreover, the reduced processing time enhances the system’s practicality in applications where immediate feedback is essential—such as adaptive media platforms, driver monitoring systems, and emotion-based therapy tools.

Overall, the comparative evaluation confirms that Mo- bileNetV2 provides the most effective compromise between classification accuracy, computational speed, and scalability. These characteristics make it a strong candidate for real-world emotion-aware systems operating on both edge devices and cloud-based infrastructures.

VI. CONCLUSION

This work presents a comprehensive emotion-aware music recommendation framework that seamlessly integrates facial expression recognition with intelligent music selection. The system leverages the efficiency of the MobileNetV2 architecture to identify user emotions such as happiness, sadness, anger, fear, disgust, surprise, and neutrality in real time with high precision and minimal computational overhead. Owing to its lightweight structure, MobileNetV2 enables smooth performance even on devices with constrained hardware, al- lowing for broad applicability across both web-based and mobile environments. The real-time webcam-based emotion detection module enables dynamic interaction, creating an adaptive music experience that responds instantly to user mood variations. The music recommendation engine employs a content- based filtering strategy that aligns detected emotions with corresponding song characteristics, including tempo, rhythm, and lyrical sentiment. This design ensures personalized and contextually relevant song suggestions that enhance user engagement and satisfaction. Security and privacy are prioritized through AES-256 encryption, authentication protocols, and access control mechanisms, protecting user data during trans- mission and interaction. A user-friendly graphical interface developed with Flask facilitates real-time detection, immediate playlist updates, and collection of user feedback for continuous model refinement. The outcomes of this research confirm thatzzz the proposed MobileNetV2-based approach achieves superior recognition accuracy and faster inference compared to conventional mod- els such as VGG16 and ResNet50. Beyond improved performance, the system demonstrates the practical fusion of deep learning, computer vision, and cybersecurity principles into a cohesive real-time solution. In the future, the framework can be expanded to include multimodal emotion sensing by integrating speech tone, physiological signals, or textual sentiment for a more holistic understanding of user affect. Further enhancements may involve reinforcement learning for adaptive playlist optimization and large-scale cloud deployment with GPU acceleration to support concurrent users efficiently. Overall, the study illustrates that emotion-aware recommendation systems have significant potential in applications such as personalized media streaming, mental health therapy, and intelligent in-car entertainment, marking a step forward in human-centered AI-driven personalization.

REFERENCES

[1] M. B. Mariappan, M. Suk, B. Prabhakaran, "FaceFetch: A User Emotion Driven Multimedia Content Recommendation System Based on Facial Expression Recognition," in 2012 IEEE International Symposium on Multimedia, 2012, pp. 84–87.

[2] S. M. Florence and M. Uma, "Emotional Detection and Music Recom- mendation System based on User Facial Expression," IOP Conf. Ser.: Mater. Sci. Eng., vol. 912, p. 062007, 2020.

[3] A. V. Gadagkar, S. Begum, S. Santhosh, and A. S. M. Ashwin, "Emo- tion Recognition and Music Recommendation System based on Facial Expression," in 2024 International Conference on Recent Advances in Science & Engineering Technology (ICAIT-2024), 2024.

[4] B. Nalini and C. Pinninti, "Efficient Facial Emotion Based Music Recommendation System," Yigkx.org.cn, 2024.

[5] M. Prasad, G. N. Swetha, and K. M. Riyaz Ali, "Creation of A Music Recommendation System using Facial Expression Recognition with MATLAB," IJISAE, 2024.

[6] M. Joshi, D. Khimasiya, and U. Limbachiya, "Emotional Detection and Music Recommendation System based on User Facial Expression," IARJSET, 2024.

[7] G. More, S. Gholap, S. Gayke, V. Hon, and S. Rokade, "Music Recommendation System Using Facial Emotion Gestures," IJRASET, 2024.

[8] J. B. Maddala et al., "Music Recommendation Based on Facial Expres- sions by using CNN," IJRASET, 2024.

[9] A. J. Mahir A et al., "Song Recommendation System Based on Facial Emotion," Goldn Cloud Publications, 2024.

[10] S. Shaik and S. Bhutada, "Deep Learning Approach for Expression- Based Songs Recommendation System," SpringerLink, 2024.

[11] P. Pathak, R. Arora, A. Gupta, and S. Abrol, "Music Emotion Recogni- tion for Intelligent and Efficient Recommendation Systems," Springer- Link, 2024.

[12] S. P. Deore, "Enriching Song Recommendation Through Facial Expres- sion Using Deep Learning," IIETA, 2022/2023.

[13] V. Vijayalakshmi, P. Shrivastav, and G. Thiyagarajan, "Facial Expression Based AI System for Personalized Music Recommendations," Atlantis Press, 2025.

[14] G. Bagadi, "Facial Emotion Detection and Music Recommendation using Deep Learning," ResearchGate, 2025.

[15] V. Tsouvalas, T. Ozcelebi, and N. Meratnia, "Privacy-preserving Speech Emotion Recognition through Semi-Supervised Federated Learning," in Proc. IEEE PerCom Workshops, 2022, pp. 128–133.

[16] A.ója Mabel Rani, M. S. Nivetha, N. M. Jothi Swaroopan, and K. Hari Kumar, "Face Emotion Based Music Recommendation System Using Modified CNN," in Proc. RMKMATE, 2023, pp. 1–6.

[17] A. Senthil Selvi and Aakash S, "EmoTune: Deep Emotion Detection and Music Recommendation System using MobileNetV3," in Proc. RMKMATE, 2023, pp. 1–6.

[18] S. L. Adru and S. Johnson, "Harmonizing Emotions: A Fusion of Facial Emotion Recognition and Music Recommendation System," in Proc. Confluence, 2024, pp. 268–273.

[19] M. Parashakthi and S. Savithri, "Facial Emotion Recognition-Based Mu- sic Recommendation System," International Journal of Health Sciences, 2022.

[20] N. Mishra, R. Gupta, and A. Raj, "Music Recommendation System by Analyzing Facial Emotions Using Deep Neural Network," SSRN, 2024.

[21] Rajesh B, Keerthana V, Narayana Darapaneni, and Anwesh Reddy P, "Music Recommendation Based on Facial Emotion Recognition," arXiv, 2024.

[22] H. Nguyen, "A Model for Song Recommendation Based on Facial Emotion Recognition," INASS, 2024.

[23] Ashwini et al., "Music Recommendation System Using YOLO v11 for Facial Expression," IJERT, 2025.

[24] M. Patel et al., "Secure AI Systems for Multimedia Applications," IEEE Transactions on Multimedia, 2022.

[25] N. Sharma et al., "Data Privacy in Deep Learning Systems," Journal of Cybersecurity, 2023.