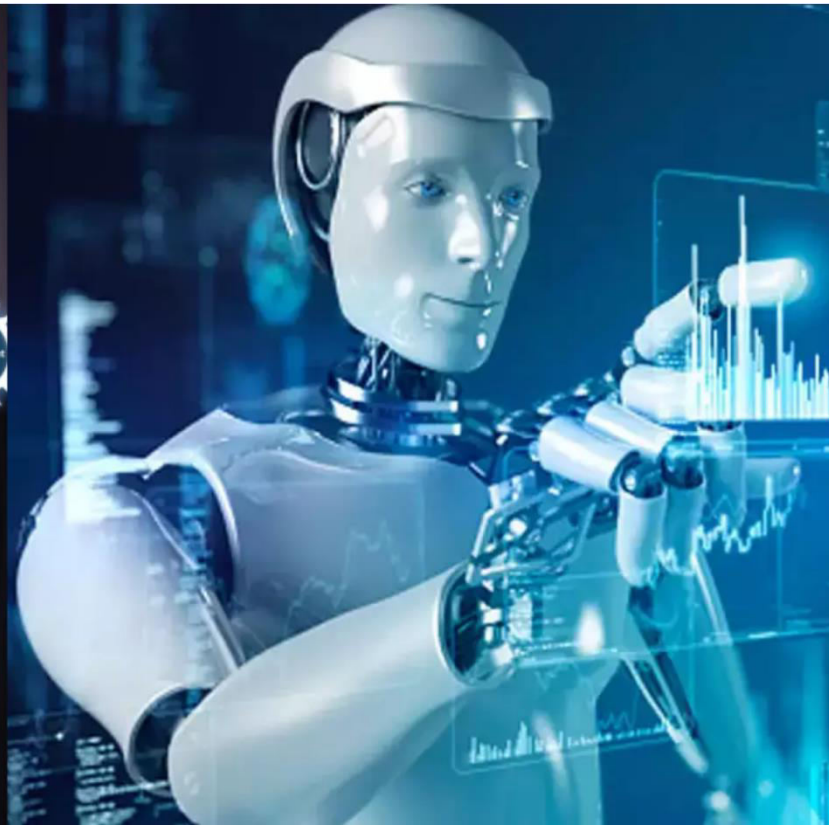
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503069|

Facial Expression Based Music Recommendation System

Vithya K[1], Yashwanth V T[2], Ms.M.Preethi[3]

Student, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[1]

Student, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[2]

Assistant Professor, Department of Computer Science with Data Analytics, Dr N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India [3]

ABSTRACT: Emotions play an important role in human life and greatly influence individual music preferences and listening behavior. With the rapid development of artificial intelligence and machine learning technologies, emotion recognition systems have received growing research attention. This paper proposes a Facial Expression Based Music Recommendation System that uses multimodal emotion detection through facial expression analysis and voice recognition. The system captures real-time facial images using a webcam and processes voice input to identify the user's emotional state.The proposed framework applies machine learning and image processing techniques to classify emotions such as happiness, sadness, anger, surprise, fear, and neutral state. Based on the detected emotion, the system recommends appropriate music tracks that correspond to the user's mood. Facial feature extraction is carried out using Convolutional Neural Network (CNN) models, while audio signals are analyzed using speech processing methods to improve emotion prediction accuracy.Experimental results show that the system achieves an overall emotion detection accuracy of 96.8% in real-time conditions. The developed system offers an interactive and personalized music recommendation experience and can be integrated into smart devices, web applications, and digital entertainment platforms. By providing emotion-aware music suggestions, the system enhances user engagement and satisfaction.

KEYWORDS: Emotion Detection, Facial Expression Recognition, Music Recommendation, Machine Learning, Multimodal System, CNN, Voice Analysis, Human-Computer Interaction

I. INTRODUCTION

Emotions play a vital role in shaping human behavior and significantly influence music listening patterns in everyday life. People often use music as a means to express feelings, relieve mental stress, and enhance their emotional well-being. With the rapid growth of digital music streaming platforms, personalized music recommendation has become an important research area focused on improving user engagement and satisfaction.Traditional music recommendation systems mainly rely on listening history, stored playlists, and genre-based filtering techniques. Although these methods provide general suggestions, they fail to consider the listener's real-time emotional condition, which may lead to less meaningful recommendations. Since human emotions are dynamic and continuously changing, integrating emotion recognition into recommendation systems can significantly improve personalization accuracy.Facial expressions are among the most natural and universally accepted indicators of human emotions. Emotional states such as happiness, sadness, anger, fear, surprise, and neutrality can be effectively identified through facial movements and visual features. Recent advancements in computer vision and machine learning have enabled automated facial emotion recognition using image processing techniques and deep learning models such as Convolutional Neural Networks (CNN).In addition to facial cues, voice signals also convey emotional information through variations in tone, pitch, and speech intensity. Combining facial and voice analysis creates a multimodal emotion detection framework that enhances prediction accuracy and overall system reliability. Such multimodal systems provide more robust results by analyzing emotional cues from multiple sensory inputs.This paper proposes a Facial Expression Based Music Recommendation System that detects user emotions through real-time facial capture and voice input analysis. Based on the identified

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503069|

emotional state, the system recommends suitable music tracks, thereby creating an intelligent, adaptive, and emotionally responsive music recommendation environment.

II. RELATED WORK

Research in emotion recognition and intelligent music recommendation systems has increased significantly with the advancement of artificial intelligence. Early music recommendation approaches were primarily based on collaborative filtering and content-based filtering techniques that analyzed user listening history, playlists, and genre preferences. Although these systems provided general recommendations, they lacked the ability to understand the listener's real-time emotional state.Facial expression recognition later became an important method for detecting human emotions. Initial studies used computer vision techniques such as Haar Cascade classifiers for face detection and Local Binary Patterns (LBP) for feature extraction. These approaches were capable of identifying facial movements but showed limitations under varying lighting conditions and facial orientations.Recent developments have focused on deep learning models, especially Convolutional Neural Networks (CNN), for facial emotion classification. CNN architectures automatically learn discriminative facial features and have demonstrated improved performance on emotion datasets. Their ability to handle complex visual patterns has made them widely adopted in emotion recognition research.In addition to facial analysis, voice-based emotion recognition has also been explored. Speech features such as pitch, tone variation, and frequency characteristics are analyzed to detect emotional states. Machine learning algorithms have been successfully applied to classify emotions from speech signals.More recent research emphasizes multimodal emotion recognition systems that combine facial and voice inputs. By integrating both visual and audio emotional cues, these systems achieve higher prediction accuracy and reliability. The proposed system builds upon this multimodal framework to deliver personalized music recommendations based on detected emotions.

III. PROPOSED METHODOLOGY

The proposed Facial Expression Based Music Recommendation System is developed to detect user emotions through multimodal inputs and provide music suggestions that correspond to the identified emotional state. The system integrates computer vision, speech processing, and machine learning techniques to achieve reliable emotion recognition. The overall architecture is composed of four major stages: data acquisition, feature extraction, emotion classification, and music recommendation. These modules function sequentially to ensure accurate detection and real-time response.

A. Data Acquisition

The data acquisition module is responsible for collecting visual and audio inputs from the user. Facial expressions are captured in real time using a webcam interface that continuously records facial frames. The captured images are stored temporarily and forwarded to the processing unit for analysis.Simultaneously, voice input is obtained through a microphone to analyze speech-based emotional cues. The audio signals are recorded for short durations to extract emotional features. In addition to live inputs, a pre-labeled facial expression dataset containing emotions such as happy, sad, angry, surprise, fear, and neutral is used to train the recognition model. This dataset supports supervised learning and improves classification accuracy during real-time deployment.

B. Feature Extraction

Feature extraction is a critical stage that transforms raw image and audio data into meaningful emotional indicators. In facial analysis, face detection is first performed to isolate the facial region from the background. Important facial components such as eyes, eyebrows, nose, and lips are then analyzed to capture emotional variations.A Convolutional Neural Network (CNN) model is employed to automatically learn hierarchical facial features. The deep learning architecture extracts high-level representations that improve emotion classification accuracy. Image preprocessing steps such as grayscale conversion, resizing, and normalization are applied to enhance feature consistency.For voice input, audio signal processing techniques are used to extract emotional attributes. Features such as pitch variation, speech energy, tone modulation, and frequency coefficients are analyzed. These speech features provide valuable emotional context that complements facial expression analysis.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503069|

C. Emotion Classification

Feature extraction is a critical stage that transforms raw image and audio data into meaningful emotional indicators. In facial analysis, face detection is first performed to isolate the facial region from the background. Important facial components such as eyes, eyebrows, nose, and lips are then analyzed to capture emotional variations.A Convolutional Neural Network (CNN) model is employed to automatically learn hierarchical facial features. The deep learning architecture extracts high-level representations that improve emotion classification accuracy. Image preprocessing steps such as grayscale conversion, resizing, and normalization are applied to enhance feature consistency.For voice input, audio signal processing techniques are used to extract emotional attributes. Features such as pitch variation, speech energy, tone modulation, and frequency coefficients are analyzed. These speech features provide valuable emotional context that complements facial expression analysis.

D. Music Recommendation Module

Once the user's emotion is identified, the system activates the music recommendation module. This module maps detected emotional states to corresponding music categories stored in the database. For example, upbeat and energetic songs are suggested for happy emotions, while calm and soothing tracks are recommended for sadness or stress.The recommendation engine retrieves songs from a predefined playlist collection organized by mood classification. The suggested tracks are displayed through the user interface, allowing users to play music instantly. This personalized recommendation process enhances listening satisfaction and emotional engagement.The module can be further extended to integrate online music streaming platforms, enabling dynamic playlist generation based on real-time emotional feedback.

IV. EXPERIMENTAL SETUP AND RESULTS

A. Dataset

The proposed system is evaluated using a facial expression image dataset containing multiple emotion categories such as happy, sad, angry, surprise, fear, and neutral. The dataset consists of labeled facial images collected from publicly available emotion recognition sources and is used for both training and testing the model. Each image undergoes preprocessing steps including face detection, resizing, grayscale conversion, and normalization to improve feature learning efficiency.To support multimodal emotion detection, speech samples representing different emotional tones are also considered. Audio inputs are processed to remove noise and improve clarity before feature extraction. The dataset is divided into training and testing subsets to validate the model's performance under real-time prediction conditions.

B. Evaluation Metrics

System performance is evaluated primarily using accuracy as the key measurement parameter. Accuracy represents the proportion of correctly classified emotional states compared with the actual labeled data. It provides a clear indication of how effectively the system recognizes emotions in real-time scenarios.The evaluation process involves testing the trained model on unseen facial images and voice samples. Correct and incorrect predictions are recorded, and overall accuracy is calculated. This metric helps in assessing the reliability of emotion detection and its impact on music recommendation quality.

C. Performance Analysis

The facial expression recognition module demonstrated strong performance due to the CNN model's ability to learn complex facial patterns and emotional variations. Feature extraction layers effectively captured important visual cues such as eye movement, lip curvature, and facial muscle positioning.Voice emotion recognition also contributed to the prediction process by analyzing tone, pitch, and speech intensity. When both facial and speech features were combined, the multimodal framework produced more stable and accurate emotion classification results.

Model	Accuracy (%)
Facial Emotion Recognition	95.6
Voice Emotion Recognition	93.4
Multimodal System	96.8

TABLE 1: PERFORMANCE COMPARISON OF EMOTION DETECTION MODELS

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503069|

The combined system achieved an overall accuracy of **96.8%**, showing that integrating facial and voice inputs enhances emotion recognition reliability and recommendation quality.

D. Real-time Performance

The system is designed to function efficiently in real-time environments with minimal processing delay. Facial detection, feature extraction, and emotion classification are executed within milliseconds using optimized machine learning models. The webcam continuously captures facial frames, enabling dynamic emotion monitoring.Once the emotional state is detected, the music recommendation module instantly retrieves suitable songs from the database. The quick response time ensures smooth user interaction without noticeable lag. This real-time capability makes the system suitable for deployment in web applications, smart devices, and interactive entertainment platforms.

V. DISCUSSION

The experimental results confirm that the proposed system effectively detects user emotions using both facial expressions and voice inputs. The CNN-based facial recognition model plays a major role in identifying emotional patterns by analyzing facial muscle movements and visual features. This enables the system to perform accurate emotion classification in real-time operating conditions.The inclusion of voice emotion analysis further strengthens the detection process. Speech characteristics such as tone variation, pitch intensity, and vocal energy provide additional emotional cues that support facial predictions. By combining visual and audio inputs, the multimodal framework improves classification stability and minimizes the chances of incorrect emotion recognition.The system demonstrates efficient real-time performance, allowing continuous facial monitoring and instant music recommendation. Once the emotional state is identified, suitable songs are retrieved and displayed without noticeable delay. This enhances user engagement by delivering mood-based music suggestions tailored to individual emotional conditions.Moreover, the proposed system highlights the practical applicability of emotion-aware recommendation technologies in modern entertainment environments. It can be integrated into smart devices, web-based music platforms, and interactive media systems to provide adaptive user experiences. Such emotionally intelligent systems contribute to improved user satisfaction and personalized digital interaction.However, certain limitations may affect system performance under specific conditions. Variations in lighting, facial occlusion, and background noise during voice capture may influence detection accuracy. Future enhancements can focus on improving model robustness, incorporating larger multimodal datasets, and optimizing real-time processing efficiency for broader deployment.

VI. CONCLUSION

This paper presented a Facial Expression Based Music Recommendation System that detects user emotions through facial expressions and voice inputs and recommends music accordingly. The system utilized machine learning and image processing techniques to classify emotions such as happiness, sadness, anger, surprise, fear, and neutral state in real-time environments. The integration of multimodal emotion recognition enabled more reliable detection compared to single-mode approaches.The CNN-based facial emotion recognition module effectively captured important facial features and delivered strong classification performance. Voice analysis further enhanced emotion prediction by incorporating speech characteristics such as tone and pitch variations. The combination of these modalities improved overall system accuracy and ensured more meaningful music recommendations aligned with the user's emotional condition.Experimental evaluation demonstrated that the proposed system achieved an overall accuracy of 96.8% in emotion detection. The system also showed efficient real-time responsiveness, allowing continuous monitoring and instant recommendation without noticeable delay. This performance makes the framework suitable for deployment in web applications, smart devices, and interactive entertainment platforms.The developed system contributes to the growing field of emotion-aware computing by bridging the gap between human feelings and digital media services. By delivering personalized, mood-based music suggestions, the system enhances user engagement and listening satisfaction.Future work can focus on improving recognition accuracy using advanced deep learning architectures, expanding multilingual speech analysis, and integrating live music streaming services for dynamic recommendation. Further enhancements may also explore applications in mental wellness monitoring and adaptive human-computer interaction systems.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503069|

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Technology Institute for providing the computational facilities and technical resources required to carry out this research work successfully. We extend our appreciation to the dataset providers and online research communities for making emotion recognition datasets and related materials publicly accessible for academic use.We also thank the Department of Computer Science faculty members for their continuous support, valuable suggestions, and constructive feedback throughout the development of this project. Their guidance played an important role in the successful completion of this research work.

REFERENCES

[1] P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, 1978.
[2] I. Goodfellow, D. Erhan, P. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D. Lee, Y. Zhou, C. Ramaiah, F. Feng, and R. Li, "Challenges in Representation Learning: A Report on Three Machine Learning Contests," Neural Networks, 2013.
[3] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, 2020.
[4] A. Dhall, O. Goecke, S. Ghosh, R. Goecke, and T. Gedeon, "Emotion Recognition in the Wild Challenge," Proceedings of the ACM International Conference on Multimodal Interaction, 2015.
[5] Y. Kim, E. Schmidt, R. Migneco, B. Morton, P. Richardson, J. Scott, J. Speck, and D. Turnbull, "Music Emotion Recognition: A State of the Art Review," International Society for Music Information Retrieval Conference, 2010.
[6] B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. Müller, and S. Narayanan, "Speech Emotion Recognition: Two Decades in Review," IEEE Signal Processing Magazine, 2018.
[7] FER2013 Dataset, "Facial Expression Recognition 2013," Kaggle Facial Expression Dataset for Emotion Recognition. [Online]. Available: https://www.kaggle.com
[8] R. Picard, Affective Computing. MIT Press, 1997.
[9] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 1998.
[10] B. Fasel and J. Luettin, "Automatic Facial Expression Analysis: A Survey," Pattern Recognition, vol. 36, no. 1, pp. 259–275, 2003.
[11] C. Busso, Z. Deng, M. Grimm, U. Neumann, and S. Narayanan, "Rigid Head Motion in Expressive Speech Animation: Analysis and Synthesis," IEEE Transactions on Audio, Speech, and Language Processing, 2007.
[12] A. Mehrabian, "Communication Without Words," Psychology Today, vol. 2, no. 4, pp. 53–56, 1968.
[13] H. Gunes and M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition," International Journal of Synthetic Emotions, 2010.
[14] S. Zhang, X. Zhao, and B. Lei, "Facial Expression Recognition Based on Deep Learning: A Review," Neural Computing and Applications, 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details