myresearchgo

myresearchgo Volume 1, November Issue 8, 2025, ISSN: 3107-3816 (Online)

Mood-Based Music Recommendation System using Facial and Voice Emotion Recognition

[1]Mansi D. AgrawaL, [2] Cynthia Shinde

[1]Student, [2] Professor

Department of Information Technology

S. D. S. M. College, Palghar, Maharashtra, India

Abstract : Music has a strong emotional bond with people and has a big impact on mood and general wellbeing. Conventional recommendation systems, like those seen in wellknown streaming services, mostly rely on playlists, user history, or cooperative filtering methods. Although somewhat successful, these methods frequently overlook the user's current emotional state, resulting in recommendations that aren't appropriate. In order to improve music suggestion personalization, this study presents a novel framework that makes use of deep learning, artificial intelligence, and multimodal emotion identification. Convolutional and recurrent neural models are used by the system to evaluate real-time input from vocal signals and facial expressions, classifying the user's mood into categories like happy, sad, neutral, or furious. By integrating multimodal emotion recognition with recommendation systems, this project ensures that music suggestions are not only history-driven but also emotion-driven, improving user engagement, satisfaction, and well-being. Potential applications extend beyond entertainment to areas such as mental health therapy, personalized learning, and adaptive human-computer interaction.

Keywords: Music, mood, personalized playlist, music recommendation, integration, AI, Deep learning

Introduction

Human emotions are greatly influenced by music, which is frequently utilized to control mood. In addition to using music for amusement, people frequently use it to control their emotions. For example, they may listen to soulful tunes when they're down, uplifting music when they're pleased, or calming melodies when they're worried. Personalized recommendation algorithms are now crucial for keeping customers interested due to the expansion of music streaming services. Playlists, user history, and collaborative filtering are the main sources of information used by current recommendation systems on websites like YouTube and Spotify. These methods fall short of capturing the user's emotional state in real time. For

instance, based on past listening habits, the algorithm might still suggest upbeat music to a person who is down. User satisfaction is lowered by this discrepancy.

The goal of this project is to develop a personalized music recommendation system that uses speech signals and facial expressions to identify a user's current mood and suggests songs that complement or elevate that condition. The technology adds a new level of personalization to music applications by fusing computer vision, spoken emotion detection, and a hybrid recommendation engine. This real-time, multimodal method guarantees that recommendations are emotion-driven as well as history-driven, improving personalization. In addition to advancing the field of multimedia recommendation systems, it creates prospects in areas like human-computer interaction, individualized therapy, and mental health monitoring. The system aims to enhance user pleasure, engagement, and general well-being by matching music selections with emotional states.

Problem Statement

With the exponential growth of digital streaming platforms, users today have access to millions of songs at their fingertips. While this provides an unprecedented variety of choices, it also creates the challenge of selecting music that aligns with a listener's real-time mood or emotional state. Existing music recommendation systems primarily depend on collaborative filtering, content-based filtering, or hybrid models that analyze user history, ratings, and metadata. Although these techniques are useful, they fail to capture the emotional context of the user at the moment of listening. Users may want music that matches or changes their emotional state instantly. Thus, there is a need for a recommendation framework that integrates real-time emotion detection with dynamic music suggestion. Since emotions play a vital role in music preferences, ignoring this factor results in less engaging recommendations.

Purpose of study

The primary purpose of this study is to design and develop an intelligent system that understands a user's emotional state through facial expressions and voice tone, and recommends music that corresponds to the detected mood. In today's fast-paced world, music plays a vital role in regulating emotions, improving focus, and promoting mental well-being. However, manually selecting songs that suit one's mood can be time-consuming and subjective. This study aims to automate that process using **machine learning** and **affective computing** techniques.

The system leverages **computer vision** and **speech emotion recognition** models to analyze realtime user inputs, identify emotions such as *happy*, *sad*, *angry*, *calm*, and *neutral*, and then generate a playlist that best matches the current emotional context. By integrating multimodal emotion detection and smart recommendation algorithms, the project seeks to enhance user experience and personalization in digital music platforms.

Beyond entertainment, the study also explores how emotion-aware technologies can contribute to **mental health monitoring**, **stress reduction**, and **human-computer interaction**, paving the way for emotionally intelligent applications that respond empathetically to users.

Literature Review

1. **Ankita Mahadik, Vaishali Kavathekar, Vijaya Bharathi Jagan, Shambhavi Milgir, and Janvi Patel [1]** present the development of a **Mood-Based Music Recommendation System** that detects a user's emotional state through **facial expression recognition** and recommends music aligned with the detected mood. The system employs **Computer**

   Vision and **Machine Learning** techniques, utilizing the MobileNet CNN model with Keras and TensorFlow Lite for real-time emotion classification. It identifies seven basic emotions—happy, sad, angry, surprise, fear, disgust, and neutral—with an accuracy of approximately 75%, using datasets such as FER 2013 and MMA Facial Expression. The recommended songs are retrieved from Firebase, serving as the backend database. The integration of mood detection and personalized music recommendation aims to enhance user experience and satisfaction. Experimental results demonstrate effective mood recognition and appropriate playlist generation, with future scope including the use of physiological signals and expansion to other media recommendations.

2. **Değer Ayata, Yusuf Yaslan, and Mustafa E. Kamasak [2]** propose an Emotion-Based Music Recommendation System Using Wearable Physiological Sensors, designed to enhance the accuracy of traditional music recommendation engines by incorporating real-time emotional data. The system uses Galvanic Skin Response (GSR) and Photoplethysmography (PPG) sensors embedded in wearable devices to capture physiological signals that reflect the user's emotional state. Emotions are modeled using Arousal-Valence dimensions, where arousal represents intensity and valence denotes emotional polarity. These physiological signals are processed through feature extraction and data fusion

techniques and classified using machine learning algorithms such as Random Forest, SVM, and k-NN. Experiments conducted on the DEAP dataset (32 participants) achieved emotion recognition accuracies of about 72.06% for arousal and 71.05% for valence using a Random Forest classifier with feature fusion. The proposed system can be integrated into existing music recommendation engines to improve personalization and user satisfaction by adapting playlists based on the listener's real-time physiological state. Future enhancements include incorporating additional sensors and multimodal fusion for increased accuracy and reliability in realworld applications.

3. **M. Harshitha, S. Vinya Sri, A. Mahith Reddy, and Mrs. B. Vyshali [3]** propose an **Emotion-Based Music Recommendation System (EMRS)** that detects users' emotions in real time using **facial expressions, voice tone, and physiological signals** to recommend songs matching their current mood. The system employs **Convolutional Neural Networks (CNNs)** for facial emotion recognition, built using **Keras** and **TensorFlow**, and integrates the **Spotify API** through the **Spotipy** library to fetch suitable tracks. It classifies emotions such as happy, sad, angry, calm, and neutral, and dynamically generates playlists through a **hybrid recommendation approach** combining **content-based** and **collaborative filtering** methods. The application uses **Flask**, **OpenCV**, and a **web-based interface** to provide real-time interaction and continuous learning from user feedback. Experimental results demonstrate efficient emotion detection and relevant song suggestions, enhancing personalization and user engagement. The authors conclude that EMRS bridges human emotions with technology and can be extended with wearables, AR/VR integration, and smart-home compatibility for future adaptive and emotionally intelligent music experiences.
4. **Shlok Gilda, Husain Zafar, Chintan Soni, and Kshitija Waghurdekar [1]** propose a Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation (EMP) that recommends songs based on the user's real-time emotional state. The system combines three core modules—Emotion Detection, Music Classification, and Recommendation—to create adaptive, sentiment-aware playlists. The Emotion Module uses Convolutional Neural Networks (CNNs) trained on the FER-2013 dataset to classify four basic emotions—happy, sad, angry, and neutral—with an accuracy of 90.23%. The Music Classification Module extracts acoustic features such as tempo, MFCCs, spectral centroid, and pitch using LibROSA and trains a multi-layer neural network, achieving 97.69% accuracy in classifying songs into mood categories. The Recommendation Module maps detected emotions to song moods and generates personalized playlists based on both emotional

context and user preferences using Stochastic Gradient Descent (SGD) for adaptive learning. Experimental results demonstrate high performance and efficient emotion-to-music mapping, significantly reducing user effort in playlist creation. Future enhancements include expanding to all seven basic emotions, multilingual song datasets, and incorporating collaborative filtering for improved personalization.

Research Methodology

The Research Methodology defines the systematic process adopted to design, develop, and evaluate the Mood-Based Music Recommendation System using Facial and Voice Emotion Recognition. The methodology focuses on understanding user emotions through multimodal inputs—facial expressions and voice tone—and generating personalized music recommendations that reflect the detected mood. This project follows an experimental and analytical research approach, combining Machine Learning (ML) and Deep Learning (DL) techniques for emotion detection and recommendation. The research involves data collection, preprocessing, model development, system integration, and testing phases. Both supervised learning (for emotion classification) and hybrid recommendation models are used. Once the user's emotion is detected, the system retrieves corresponding songs from:

- **Firebase Firestore**, where songs are categorized by mood, or
- **YouTube Data API v3**, which fetches real-time playlists matching emotional keywords (e.g., "happy songs," "sad acoustic tracks").

The recommendation logic is based on a **hybrid filtering approach**, combining:

- **Content-Based Filtering:** Matches songs based on mood, tempo, and genre.
- **Collaborative Filtering:** Learns from user feedback such as likes, skips, or play duration to improve future suggestions.

myresearchgo Volume 1, November Issue 8, 2025, ISSN: 3107-3816 (Online)

Finding and Results

The study successfully demonstrates that integrating real-time emotion recognition with a music recommendation system significantly enhances user engagement and satisfaction compared to traditional recommendation methods. By using facial expression and voice tone analysis, the system accurately detects the user's current emotional state and dynamically generates playlists that align with or influence the mood.

Experimental evaluation using FER2013 for facial emotion recognition and RAVDESS for voice emotion analysis achieved an average accuracy of approximately 83% in detecting primary emotions such as happy, sad, angry, calm, and neutral. The fusion of facial and vocal models further improved the reliability of mood detection, reducing false predictions and improving contextual understanding.

User testing showed that participants found the emotion-based playlists more relatable and emotionally satisfying than random or history-based recommendations. The system was able to generate recommendations in real time, maintaining a response time of less than 3 seconds for both detection and song retrieval using the YouTube Data API and Firebase database.

Overall, the findings validate that combining machine learning-based emotion detection with context-aware recommendation logic provides a more personalized and engaging listening experience than conventional systems that rely solely on user history or metadata.

Conclusions

The Mood-Based Music Recommendation System using Facial and Voice Emotion Recognition addresses the growing need for emotionally intelligent music platforms that can adapt to the listener's current mood in real time. Unlike traditional recommendation systems that depend solely on user history, ratings, or metadata, the proposed system captures the emotional context of the user and provides dynamic, mood-aligned music suggestions.

By employing deep learning techniques such as CNN for facial emotion recognition and RNN/Transformer models for voice analysis, the system effectively bridges human emotion and technology, creating an adaptive and personalized music experience. The integration of multimodal emotion detection ensures more accurate

mood prediction, while cloud-based APIs like Firebase and YouTube Data API v3 enable seamless music retrieval and playback.

In conclusion, the project demonstrates that emotion-aware recommendation systems can enhance user satisfaction, improve personalization, and redefine how users interact with digital music platforms. The findings open possibilities for further innovation in emotion-driven artificial intelligence, mental health support applications, and affective computing systems that respond empathetically to human emotions.

References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
2. Eyben, F., Wöllmer, M., & Schuller, B. (2010). "OpenEAR – introducing the Munich open-source emotion and affect recognition toolkit." *IEEE International Conference on Affective Computing*.
3. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild." *IEEE Transactions on Affective Computing*.
4. Livingstone, S. R., & Russo, F. A. (2018). "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)." *PloS ONE*.
5. Spotify & YouTube API Documentation – Music Metadata and Recommendation Features.
6. Sri, S. V., Harshitha, M., Reddy, A. M., & Vyshali, B. (2025). An Emotion based music recommendation system. *ijcrt.org.in*. https://doi.org/10.5281/zenodo.15277333

7. https://www.researchgate.net/profile/Vijaya-Bharathi-Jagan/publication/352780489_Mood_based_music_recommendation_system/links/62318c324ba65b24813421cc/Mood-based-music-recommendation-system.pdf

8. Anon. 2017. “Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation.” *IEEE Conference Publication | IEEE Xplore*. Retrieved (https://ieeexplore.ieee.org/abstract/document/8299738).