International Journal for Modern Trends in Science and Technology
Volume 12, Issue 03, pages 68-73.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol12issue03.html
DOI: https://doi.org/10.5281/zenodo.18880655

Mood-Based Music Recommendation System using Machine Learning

K.V. Lalitha, M. Sai Nadha Swamy, D. Steni, D. Arjun, K. Venkateswarlu

Department of ECE, Godavari Global University, Rajamahendravaram, INDIA.
Department of ECE, Godavari Institute of Engineering & Technology (A), Rajamahendravaram, INDIA.

To Cite this Article
K.V. Lalitha, M. Sai Nadha Swamy, D. Steni, D. Arjun & K. Venkateswarlu (2026). Mood - Based Music Recommendation System using Machine Learning, 12(03), 68-73. https://doi.org/10.5281/zenodo.18880655

Article Info
Received: 28 January 2026; Revised: 26 February 2026; Accepted: 02 March 2026.

KEYWORDS

Mood-Based Music Recommendation, Emotion Recognition, Deep Learning, Vision Transformer (ViT), Machine Learning, K-Nearest Neighbors(KNN), Music Personalization, Artificial Intelligence (AI).

ABSTRACT

In today's busy and stressful world, many people are facing mental and emotional problems due to heavy workloads and personal pressure. Most people don't get enough time to care for their mental health. Research shows that music play's a powerful role in improving a person's emotions and can help them feel better. So, our project idea is to detect a person's mood using a camera and suggest songs that match their current emotional state. We use different methods to understand the mood — like reading facial expressions using a camera, allowing users to upload photos or choose emojis, and even checking the current weather. Based on the mood, the system gives music suggestions that can help shift the person into a more positive mood. The goal of the project is to support emotional wellbeing by using music as a healing tool with the help of artificial intelligence and machine learning. This project uses smart technology to understand a person's emotions and recommend music. We use a machine learning model called Vision Transformer (ViT) to detect facial expressions from the user's image or live camera. For suggesting music, we use the K-Nearest Neighbors(KNN) algorithm, which chooses songs based on features like how fast the song is (BPM), how happy or energetic it feels (Valence, Energy, Danceability), and more. Users can also select emojis or let the system check the weather to detect mood. The project is built using Python programming and tools like TensorFlow, OpenCV, NumPy, and Streamlit for the user interface. A basic computer with 8GB RAM and a camera is needed to run the system. The system also allows language selection for Indian users. Compared to old methods, this new system gives better, real-time suggestions and helps users enjoy music that truly fits their feelings.

INTRODUCTION

Music plays an important role in human life because it has the power to influence emotions and mood. People often listen to music to relax, feel motivated, reduce stress, or express their feelings. Traditional music recommendation systems mainly depend on listening history, genre preferences, or collaborative filtering techniques. However, these systems do not understand the user's real time emotional state, which limits the personalization of recommendations.

To overcome this limitation, this project introduces a mood-based music recommendation system that combines deep learning and machine learning techniques. The main goal of this system is to detect the user's current emotion and recommend songs that match their mood. The system provides four different input methods: emoji selection, climate-based mood prediction, image upload for emotion analysis, and live camera feed for real time facial emotion detection. This multi modal approach makes the system more flexible and user friendly. Additionally, a language based filtering option is included so that users can receive recommendations in their preferred language.

In this upgraded version of the project, the traditional Convolutional Neural Network (CNN) is replaced with a Vision Transformer (ViT) for facial emotion recognition. Vision Transformer is a modern deep learning architecture that uses attention mechanisms to understand image features more effectively. It divides the input image into smaller patches and analyzes relationships between different parts of the image, which helps in capturing facial expressions more accurately. The model classifies emotions into categories such as happy, sad, neutral, and angry. For uploaded images, DeepFace is also used for additional emotion analysis support.

Once the emotion is detected, it is sent to the Music Recommendation Module. This module uses the K-Nearest Neighbors (KNN) algorithm to recommend suitable songs. The recommendation is based on important audio features such as BPM (tempo), energy, valence, danceability, acousticness, and loudness. These features help the system match the emotional state of the user with appropriate songs in a data-driven way.

By integrating Vision Transformer for emotion detection and machine learning for song recommendation, this system provides a smarter and more context-aware music experience. Unlike traditional systems, it considers real time emotional input along with user preferences, making it more dynamic and interactive.

However, some challenges still exist. Facial emotion detection can be affected by lighting conditions, facial occlusions, and individual differences in expressions. Also, mood-based recommendations may not always perfectly match personal music taste. For example, a user who appears sad may still prefer energetic songs. Future improvements can include hybrid recommendation models that combine mood detection with listening history for better personalization.

Overall, this project demonstrates how advanced deep learning techniques like Vision Transformer can enhance emotion aware music recommendation systems. By enabling real-time mood analysis and multi-modal inputs, the system offers a more personalized and engaging music listening experience.

LITERATURE SURVEY

The development of music recommendation systems has gradually shifted from traditional preference-based filtering methods to more intelligent, AI-driven mood-based systems. Earlier recommendation techniques mainly relied on collaborative filtering and content-based filtering, which used user listening history and genre preferences to suggest songs. Although these systems provided reasonable recommendations, they were unable to recognize the user's emotional condition in real time. This limitation encouraged researchers to explore emotion-aware music recommendation approaches.

The study titled *"Melodic Mood: Emotion-based Music Recommendation System Using Deep Learning Techniques"* (2024), published in *Multimedia Tools and Applications*, proposed a system that detects user emotions from images and recommends songs accordingly. The authors implemented deep learning techniques for facial emotion recognition and selected the best-performing model based on accuracy evaluation. Their work demonstrated that integrating visual emotion detection significantly improves the personalization of music recommendations.

Similarly, the research work *"Music Recommendation System Based on Real-Time Emotion Analysis"* (2022) focused on detecting facial expressions

in real time and mapping them to suitable songs. The system analyzed facial features to recognize user emotions and dynamically generated recommendations based on the detected mood. This study highlighted the importance of real-time emotion analysis in improving user engagement and recommendation accuracy.

Another relevant work, *"Mood-Based Music Recommendation System"* (2021), published in Springer Proceedings, implemented machine learning techniques for mood classification and song selection. Instead of relying only on historical user data, this system categorized songs based on emotional characteristics and matched them with detected moods. The research emphasized the role of audio feature analysis, including parameters such as energy, valence, and tempo, in building an effective recommendation framework.

With advancements in deep learning, transformer-based architectures have gained significant attention in computer vision tasks. Vision Transformer (ViT) has emerged as a powerful alternative to traditional Convolutional Neural Networks for image classification and facial analysis. Unlike CNNs, which focus primarily on extracting local features through convolution operations, ViT divides an image into smaller patches and processes them as sequential inputs using self-attention mechanisms. This allows the model to capture global relationships between different facial regions, leading to improved understanding of complex emotional expressions.

Recent studies suggest that transformer-based models can achieve competitive or superior performance compared to CNN-based systems when sufficient training data is available. The ability of Vision Transformer to model long-range dependencies makes it particularly suitable for facial emotion recognition, where overall facial structure and contextual relationships play a crucial role.

In addition to emotion detection, feature-based music recommendation has become an important research direction. Many studies classify songs based on audio features such as BPM (beats per minute), energy, valence, danceability, acousticness, and loudness. These features help determine the emotional tone of a song and improve the alignment between detected mood and recommended tracks. Machine learning algorithms such as K-Nearest Neighbors (KNN) are commonly used to match user emotions with songs having similar audio characteristics.

Despite significant progress, certain research gaps still exist. Many existing systems lack multilingual filtering, making them less effective in culturally diverse regions. Furthermore, environmental factors such as weather conditions, which can influence mood and music preference, are rarely integrated into recommendation models. Hybrid systems that combine real-time emotion detection with contextual and preference-based filtering are still limited.

To address these gaps, the proposed study develops a Mood-Based Music Recommendation System that integrates Vision Transformer (ViT) for facial emotion detection and K-Nearest Neighbors (KNN) for song recommendation. By incorporating multi-modal inputs, audio feature analysis, and language-based filtering, the system aims to provide a more intelligent, adaptive, and user-centred music recommendation experience.

METHODOLOGY

The proposed Mood-Based Music Recommendation System integrates deep learning-based emotion recognition with machine learning-based song classification to overcome the limitations of traditional music recommendation models. Unlike conventional systems that rely only on listening history or genre preferences, this approach focuses on detecting the user's real-time emotional state and recommending songs accordingly.

The system collects user input through multiple modalities, including live camera feed, image upload, emoji selection, and climate-based mood inference. This multi-modal design ensures flexibility and improves the reliability of emotion detection. By allowing users to interact with the system in different ways, the model becomes more adaptable to real-world usage scenarios.

For facial emotion recognition, the system replaces traditional Convolutional Neural Networks (CNN) with a Vision Transformer (ViT) model. Vision Transformer is a modern deep learning architecture that processes images by dividing them into smaller patches and analyzing them using a self-attention mechanism. Unlike CNNs, which focus mainly on local feature extraction, ViT captures global relationships between different regions of the face. This enables more accurate detection

of facial expressions, especially in cases where emotions are subtle or complex. The model classifies user emotions into categories such as happy, sad, angry, neutral, and relaxed.

When users upload an image instead of using the live camera, the same Vision Transformer-based emotion detection process is applied to ensure consistency and accuracy. In addition to facial analysis, emoji selection is handled through predefined mood mappings, where each emoji corresponds to a specific emotional category. Similarly, climate-based mood inference maps weather conditions such as sunny, rainy, or cloudy to associated emotional states based on established correlations. The combination of these input methods enhances the robustness of mood identification.

Once the user's emotion is identified, it is passed to the Music Recommendation Module. This module uses the K-Nearest Neighbors (KNN) algorithm to retrieve songs that best match the detected mood. Each song in the database is characterized using key audio features such as BPM (beats per minute), energy, valence, danceability, acousticness, and loudness. These features help represent the emotional intensity and musical characteristics of each track.

The KNN algorithm calculates similarity between the detected mood profile and the feature vectors of songs in the database. By identifying the K-nearest matching tracks, the system ensures that recommended songs closely align with the user's current emotional state. Additionally, a language preference filter is applied to refine the recommendations, allowing users to receive songs in their preferred Indian language, such as Telugu, Hindi, or Tamil.

The complete system is implemented as a web-based application using Streamlit, providing a simple and interactive user interface. The backend integrates TensorFlow/Keras for implementing the Vision Transformer model, Scikit-Learn for the KNN algorithm, OpenCV for image processing, and MongoDB for storing song metadata and feature information. This architecture ensures scalability, real-time processing capability, and efficient performance.

Overall, the proposed methodology combines transformer-based emotion detection, multi-modal input processing, and feature-driven music recommendation to create a smart, adaptive, and user-centered music recommendation system. By incorporating Vision Transformer instead of CNN, the system achieves improved contextual understanding of facial expressions and delivers more accurate mood-based song suggestions.

ARCHITECTURE DIAGRAM

The four particular components of the Mood-Based Music Recommendation System are User Interface Module, Emotion Detection Module, Music Recommendation Module, and Database Module. The system is thus developed in the modular manner. End users have multiple ways in which they can contribute to the User Interface Module powered by Streamlit, such as using a live camera feed, uploading pictures, selecting emoji, or letting the module infer based on the climatic condition. However, to address the input, the Emotion Detection Module relies on various means such as Convolutional Neural Networks (CNN) for facial recognition, DeepFace for image-based analysis, and the preset mappings from weather and emoji to infer mood meaning. Once an emotion is identified, the Music Recommendation Module determines the K-Nearest neighbors (KNN) to identify music for the identified mood.

Unlike traditional music recommendation systems, this particular project does not store songs' metadata in a database. Songs are processed and classified directly by the system, using given parameters of danceability, valence, energy, and BPM. The KNN algorithm provides successful recommendations based on a labored collection of songs with identified features, without the need for any separate means of storage. Another criterion for specifying recommendations is the language preference of the user. The overall architecture is built to achieve real-time mood-based song choice with a smooth and enjoyable user experience in the scalable, lightweight, and optimized design.

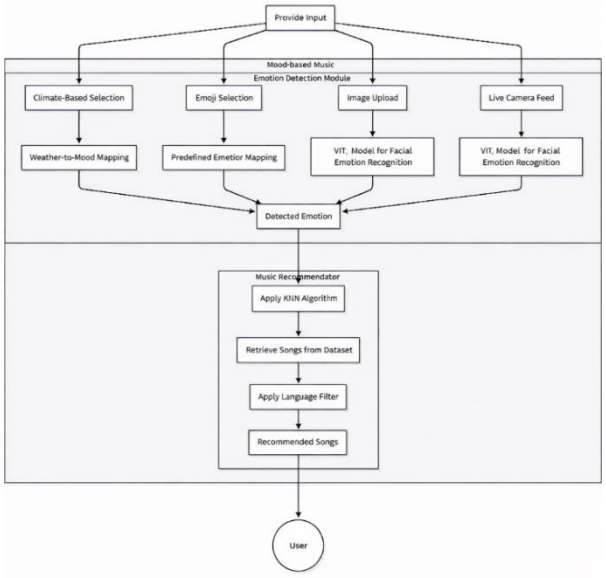

Figure 1: Architecture diagram

RESULTS AND DISCUSSION

Using important criteria like accuracy and loss, the mood-based music recommendation system's performance was assessed. Training and testing stages' model loss is depicted in the first graph. The training loss steadily dropped as the number of epochs increased, suggesting that the model was learning efficiently, even if the testing and training loss values were initially large. Nevertheless, the testing loss stabilized after a specific number of epochs, indicating that the model had achieved its point of optimal generalization. A little amount of overfitting is indicated by the difference between the training and testing losses, which can be improved by regularization methods

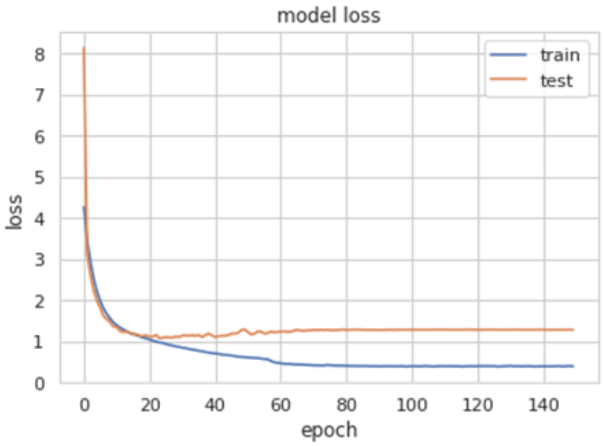

Figure 2: Model Loss Diagram

The model's accuracy throughout several epochs is depicted in the second graph. As evidence of the model's capacity to pick up significant patterns, the training accuracy steadily rose, surpassing 85%. The test accuracy, meanwhile, reached a plateau at about 65%, suggesting that although the
model does well on the training dataset, it can still be improved for use with unseen data. The discrepancy in accuracy between testing and training indicates that more fine-tuning, such adding more data or changing hyperparameters, is required to improve the model's resilience.

CONCLUSION

The Mood-Based Music Recommendation System successfully integrates transformer-based emotion recognition with machine learning-based song recommendation. By replacing CNN with Vision Transformer, the system achieves improved contextual understanding of facial expressions and more accurate emotion classification.

The integration of multi-modal inputs such as live camera, image upload, emoji selection, and weather-based inference enhances system flexibility and robustness. The KNN algorithm effectively matches songs using audio features like BPM, energy, valence, and danceability, ensuring relevant mood-based recommendations.

Although minor overfitting and dataset imbalance exist, the system demonstrates strong potential for real-world applications. Future improvements may include expanding the training dataset, implementing hybrid recommendation models, applying reinforcement learning for adaptive personalization, and deploying the system as a mobile application.

Overall, the project highlights the practical application of Vision Transformer and machine learning techniques in building intelligent, emotion-aware music recommendation systems.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fss8wd

[2] Hemanth P,Adarsh ,Aswani C.B, Ajith P, Veena A Kumar, "EMOPLAYER: Emotion Based Music Player", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 4, April 2018, pp. 4822-87.

[3] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," Applied Sciences, vol. 8, no. 7, p. 1103, Jul. 2018.

[4] Emotion based music recommendation system using wearable physiological sensors. (2018, May 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8374807/

[5] Emotion-Based Music Recommendation System. (2021, September 3). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9596276/

[6] Emotion-Based Music Recommendation System- a deep learning approach. (2024, February 22). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10493773/

[7] Han, B., Rho, S., Jun, S., & Hwang, E. (2009). Music emotion classification and context-based music recommendation. Multimedia Tools and Applications, 47(3), 433–460. https://doi.org/10.1007/s11042-009-0332-6

[8] An emotional recommender system for music. (n.d.). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9204829/

[9] Chen, Y., Cheng, C., Chen, D., & Lai, C. (2015). A mood- and situation-based model for developing intuitive Pop music recommendation systems. Expert Systems, 33(1), 77–91. https://doi.org/10.1111/exsy.12132

[10] Emotion based mood enhancing music recommendation. (2017, May 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8256863/

[11] MOHAMED, M. H., KHAFAGY, M. H., IBRAHIM, M. H., KHALED ELMENSHAWY, & HAITHAM RIZK FADLALLAH. (2021). MUSIC RECOMMENDATION SYSTEM USED EMOTIONS TO TRACK AND CHANGE NEGATIVE USERS' MOOD [Journal-article]. Journal of Theoretical and Applied Information Technology, 99(17), 4358–4359. https://www.jatit.org

[12] Music recommendation system based on emotion. (2021, July 6). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9579689/

[13] Music recommendation system using emotion triggering low-level features. (2012b, May 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/6227467/

[14] An integrated music recommendation system. (2006, August 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/1706489/

[15] De Prisco, R., Guarino, A., Malandrino, D., & Zaccagnino, R. (2022). Induced Emotion-Based Music Recommendation through Reinforcement Learning. Applied Sciences, 12(21), 11209. https://doi.org/10.3390/app122111209

[16] A Novel Emotion based Music Recommendation System using CNN. (2023, May 17). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10142330/

[17] Rho, S., Han, B., & Hwang, E. (2009). SVR-based music mood classification and context-based music recommendation. Proceedings of the 30th ACM International Conference on Multimedia, 2007, 713–716. https://doi.org/10.1145/1631272.1631395

[18] Emotion based Music Recommendation System using LSTM - CNN Architecture. (2021, July 6). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9579813/

[19] Choi, H. J. M. J. H. H. K. (2010). Music recommendation system based on user emotion and music mood. https://koreascience.kr/article/CFKO201033254124211.page

[20] Personalized Mood-Based Song Recommendation System using a hybrid approach. (2023, November 9). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10355387/

[21] Smart music player integrating facial emotion recognition and music mood recommendation. (2017, March 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8299738/

[22] Florence, S. M., & Uma, M. (2020). Emotional Detection and Music Recommendation System based on User Facial Expression. IOP Conference Series Materials Science and Engineering, 912(6), 062007. https://doi.org/10.1088/1757-899x/912/6/062007 .