International Journal of Multidisciplinary Research and Growth Evaluation.

EMUS: An Intelligent Music Recommendation System

Shaik Sohel [1*], Vanukuri Manideepa [2], Alla Sai Pavan [3], Danaboina Vamsi Krishna [4], KRMC Sekhar [5]

[1-4] Student, Department of Information Technology, Kallam Haranadhareddy Institute of Technology (Autonomous), Chowdavaram, Guntur, Andhra Pradesh, India

[5] Assistant Professor, Department of Information Technology, Kallam Haranadhareddy Institute of Technology (Autonomous), Chowdavaram, Guntur, Andhra Pradesh, India

* Corresponding Author: **Shaik Sohel**

Article Info

ISSN (online): 2582-7138
Volume: 06
Issue: 02
March-April 2025
Received: 05-02-2025
Accepted: 01-03-2025
Page No: 751-755

Abstract

Music plays a prominent role in various aspects of human life, culture, and society by influencing emotions, strengthening social bonds, preserving traditions, and shaping personal and collective identities. As AI emerges as a powerful tool to automate various tasks, music recommendation systems have become an integral part of this transformation. These systems automatically generate personalized music playlists for users based on their mood and listening behavior. By analyzing factors like facial expressions, voice tone, text input, and listening history, AI-driven music recommendation systems identify the user's emotional state and suggest songs that match or enhance their mood. Emotion-based music recommendation systems significantly enhance the way people experience music by improving emotional well-being, boosting user engagement, and broadening musical preferences. In this work, we propose an application called EMUS, an intelligent music recommendation system designed to suggest music based on the user's emotional state.

DOI: https://doi.org/10.54660/.IJMRGE.2025.6.2.751-755

Keywords: Deep Learning, Machine Learning, Music Recommendation System, CNN, Face Detection

1. Introduction

A music suggestion system is a smart application that aims to propose songs to users according to their listening habits, likes, and situational elements. As digital music libraries and streaming services expand rapidly, many users find themselves inundated with the vast amount of content available. Recommendation systems assist users in managing this extensive selection by examining their listening history, user activity, and metadata like genre, artist, and tempo [1]. These systems utilize a range of methods, such as collaborative filtering, content-based filtering, and hybrid models, to provide customized music recommendations [2]. By analyzing trends in user preferences and music features, recommendation systems enhance the overall listening experience and enable users to find new music that resonates with their tastes.

Recent progress in machine learning and artificial intelligence has considerably enhanced the performance of music recommendation systems. Deep learning techniques, such as CNNs and RNNs, are increasingly utilized to uncover intricate patterns in audio data and user behavior [3]. The integration of emotion recognition and sentiment analysis allows for music recommendations that align with the listener's emotional state. Additionally, real-time data processing and adaptive algorithms enable the system to modify recommendations instantly based on user feedback and evolving preferences. These developments have not only raised the precision of recommendations but have also increased user engagement and retention on music streaming services. In this work, we propose an intelligent music recommendation system that utilizes a Convolutional Neural Network (CNN) to recommend music from the library based on the user's emotions.

2. Related Work

Gaikwad Uday Vijaysinh *et al.* [4] proposed a system that captures the user's real-time emotions, potentially through facial expressions or conversation, and then recommends songs categorized by mood, such as happy, sad, or neutral. It mainly focusses on content-based features (acoustic parameters) for recommendation, potentially missing opportunities for personalized recommendations based on user history and feedback.

H. Immanuel James *et al.* [5] introduced a music recommendation system that uses a camera to capture facial expressions of an individual and then analyzes these expressions to determine their mood. This detected emotion is then used to automatically generate a playlist of songs appropriate to that emotional state, eliminating the need for manual song categorization. In this work, they applied Histogram of Oriented Gradients (HOG) and emotion classification using an SVM, aiming to provide a more personalized and efficient music listening experience.

Tina Babu *et al.* [6] proposed AI model that detects users' real-time emotions and generates personalized song recommendations that match their emotional state2. This approach aims to offer music that resonates with their current mood, elicits desired emotions, and creates a more immersive and meaningful listening experience. They used CNN model and Alchemy Blockchain API for secure storage and retrieval of system data, including user preferences, song metadata, and emotional tags. This model may lead to overfitting and processing inefficiency due to network delays.

Ashu Abdul *et al.* [7] proposed Emotion-Aware Personalized Music Recommendation System (EPMRS) designed to improve user listening experience by considering user data, including their emotions. This model employed a hybrid approach, combining a Deep Convolutional Neural Networks (DCNN) method and a Weighted Feature Extraction (WFE) method to correlate user data with music data. The performance analysis of this work is not demonstrated, limiting the validation of its effectiveness and real-world applicability.

3. Proposed Methodology

In this work, we propose an intelligent music recommendation system called EMUS that leverages facial emotion recognition to provide personalized music recommendations. The system aims to enhance the user's listening experience by analyzing their emotional state and suggesting songs that align with their mood.

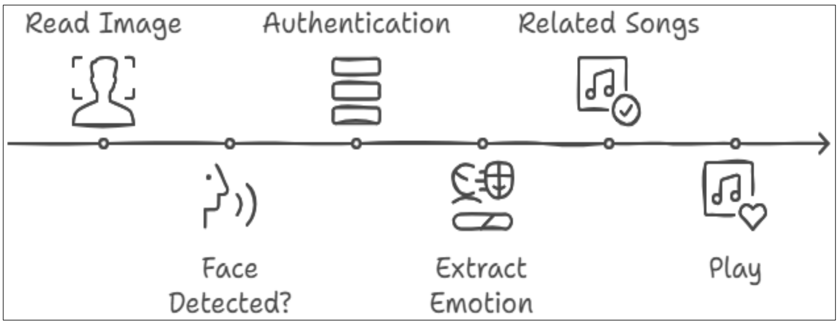

Fig 1: Proposed Model Architecture

Generalized Algorithm:

Step 1: Read Image

- The system begins by capturing an image of the user's face using a camera or other input device.

Step 2: Face Detection

- The captured image is processed to determine if a face is present.
- If no face is detected, the system loops back to the "Read Image" step.
- If a face is detected, the system proceeds to the next step.

Step 3: Authentication

- The detected face is compared to a database of registered users' facial features.
- If a match is found, the user is authenticated, and the system proceeds to the "Extract Emotion" step.
- If no match is found, the system can either:
- Prompt the user to register their face.
- Return to the "Read Image" step, or terminate the process.

Step 4: Extract Emotion

- Once the user is authenticated, the system extracts emotional features from the detected face using a CNN or other machine learning techniques.

Step 5: generate recommended songs list

- Based on the extracted emotion, the system retrieves a list of recommended songs from the song database, potentially personalized based on the authenticated user's preferences.

Step 6: Play recommended songs

- The system plays the recommended song(s) to the authenticated user.

In our proposed system, a Convolutional Neural Network (CNN) is used for feature extraction from facial images. The CNN consists of convolutional layers for detecting patterns, pooling layers for dimensionality reduction, and fully connected layers for final feature representation and classification. Further, the model is trained using the Adam Optimizer, an adaptive optimization algorithm that dynamically adjusts learning rates for efficient and faster convergence.

The model's performance is optimized using Categorical Crossentropy, a loss function designed for multi-class classification problems, which measures the difference between predicted and actual class probabilities to improve model accuracy**.** Early Stopping is also applied to halt training when the validation loss stops improving, thereby preventing overfitting and enhancing the model's generalization ability.

The proposed system uses a Haar Cascade Classifier for face detection from a live webcam feed or image, efficiently identifying facial features based on trained patterns. To further improve model performance, Dropout Regularization is employed to reduce overfitting by randomly disabling neurons during training. Image Data Augmentation (such as rescaling) is used to normalize pixel values, improving the model's ability to generalize across different inputs. Finally, the Softmax Activation function is applied to the final layer, converting outputs into probability values for accurate multi-class classification.

Algorithm for the proposed music recommendation system (EMUS)

Step 1: Initialize System
1.1 Load the trained CNN model.
1.2 Load the Haar Cascade Classifier for face detection.

Step 2: Face Detection (Using Haar Cascade)
2.1 Capture an image from a live webcam feed or input image.
2.2 Apply Haar Cascade Classifier to detect the face region.
2.3 If a face is detected, crop and extract the face region.
2.4 If no face is detected, return to Step 2.1.

Step 3: Preprocessing
3.1 Convert the extracted face to grayscale to reduce complexity.
3.2 Resize the image to match the CNN input size.
3.3 Normalize pixel values through rescaling for consistent input.

Step 4: Feature Extraction (Using CNN)
4.1 Convolutional Layers – Apply convolution to detect patterns and extract low-level features.
4.2 ReLU Activation Function – Introduce non-linearity and prevent vanishing gradient issues.
4.3 Max Pooling – Reduce spatial dimensions while retaining key features.
4.4 Repeat steps 4.1 to 4.3 for deeper feature extraction.

Step 5: Classification (Using Fully Connected Layers)
5.1 Apply Flattening – Convert the 2D feature maps into a 1D vector.
5.2 Pass the flattened vector to Dense (Fully Connected) Layers to combine extracted features.
5.3 Use Dropout Regularization – Randomly disable neurons to prevent overfitting.
5.4 Apply Softmax Activation on the final layer to output probabilities for each emotion category.

Step 6: Training
6.1 Use Batch Processing – Divide the dataset into mini-batches for efficient training.
6.2 Set Steps per Epoch – Define how many batches to process in each epoch.
6.3 Set Validation Steps – Define how many validation batches to process per epoch.
6.4 Train the model using Adam Optimizer
6.5 Monitor Categorical Crossentropy
6.6 Apply Early Stopping – Stop training if validation loss stops improving to prevent overfitting.

Step 7: Emotion-Based Recommendation
7.1 Predict the emotion using the trained CNN model.
7.2 Match the detected emotion with the music library.
7.3 Recommend and play songs based on the predicted emotion.

Step 8: Termination
8.1 Continue monitoring for new face inputs or terminate based on user input.

4. Results and Discussion

Our proposed system is implemented and validated by using the following tools and libraries to ensure user-friendly and efficient development.

a) OpenCV:
- Used for face detection through the Haar Cascade Classifier.
- Facilitates real-time image processing, making it ideal for capturing facial data from live webcam feeds or static images.

b) TensorFlow:
- Employed for developing and training the CNN-based emotion recognition model.
- It provides a powerful framework for deep learning with efficient model building, training, and evaluation tools.

c) Pandas & SQL:
- Pandas is used for handling structured data such as music track details.
- SQL used to manage the storage and retrieval of recommended songs, ensuring fast query processing and data management.

d) Flask/Django:
- Used for implementing web-based recommendation system interface.
- It supports integration between the backend logic (emotion recognition model) and the user interface, ensuring smooth user interactions.

This combination of tools and libraries enables our system to efficiently detect emotions, classify them using CNN, and provide personalized music recommendations in real-time.

Train and Test

We considered the dataset FER2013 from kaggle. Data is processed in batches of 64 to improve memory efficiency and training speed. The model is trained for 75 epochs with backpropagation and gradient descent optimization. The CNN architecture employs various convolutional layers with ReLU activation for feature extraction. Dropout (set to 0.25) is used to prevent overfitting. The final dense layer uses softmax activation to classify music genres or preferences. A validation set is used to monitor model performance during training. The model's performance is tracked using accuracy and loss curves. The model is tested on unseen data to evaluate its generalization ability.

We implemented a user-friendly interface that allow users to use a webcam to detect face and emotions. The system analyzes the emotional state and provides personalized music recommendations in real-time. The interface also includes options for user feedback and customization of music preferences. The sample output screens are depicted in figure 2 and figure 3. The figure 2 and 3 represents the facial emotions and recommended music playlist. Our proposed system is designed to learn from user feedback, enabling it to continuously adapt and improve recommendation accuracy, thereby enhancing its practical applicability. The sample feedback is depicted n figure 4.

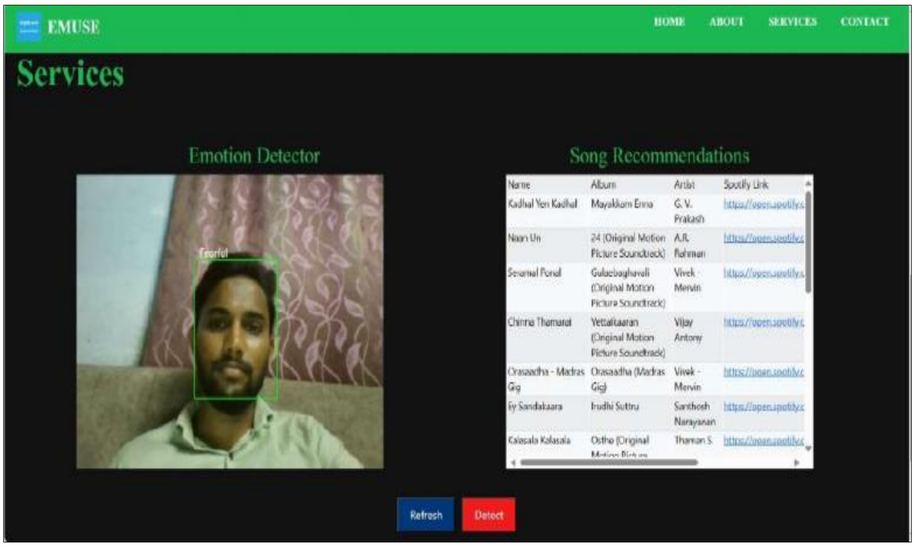

Fig 2: Sample Output Screen 1

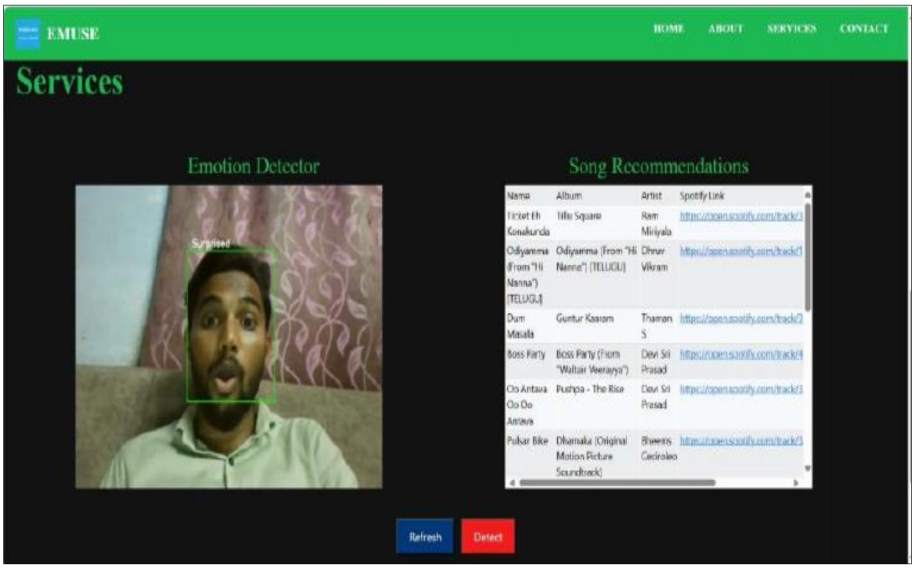

Fig 3: Sample Output Screen 2

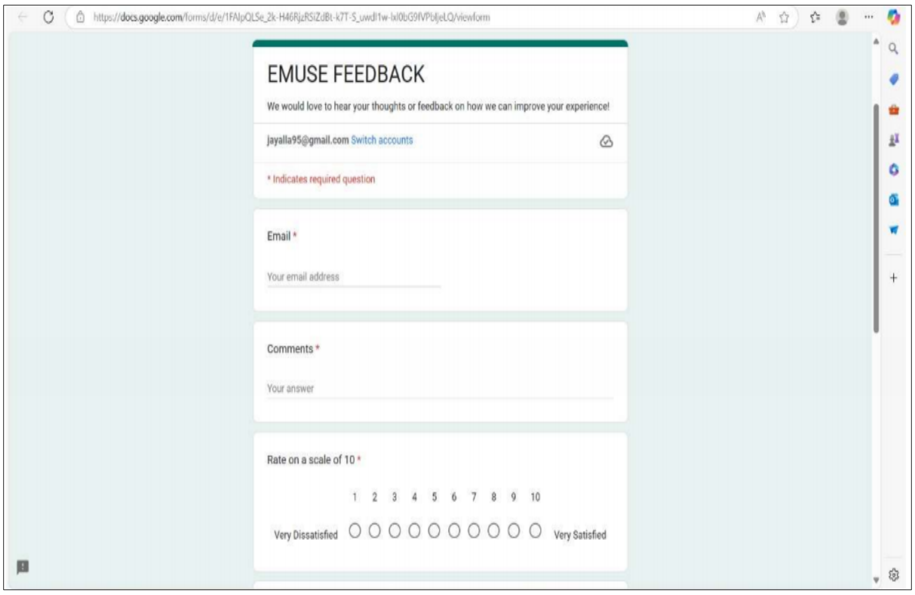

Fig 4: Sample Output Screen 3

The proposed model shows good learning behavior during training as both accuracy increases and loss decreases. It is shown in figure 5.

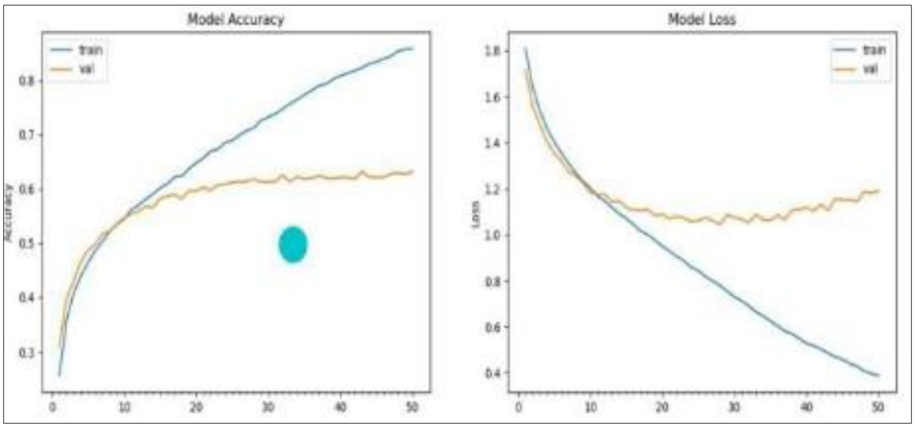

Fig 5: Proposed Model Accuracy

5. Conclusion

The proposed intelligent music recommendation system EMUS demonstrate a significant improvement over existing CNN-based models for music recommendation. EMUS achieves higher accuracy, faster training, and better user personalization. The system enhances the music experience by incorporating user feedback, allowing it to adapt to individual preferences over time and provide more accurate and diverse recommendations. Experimental results confirm that EMUS performs better than the traditional CNN-based models in terms of accuracy, precision, recall, and recommendation relevance. In future work will focus on expanding the dataset, refining the model's ability to handle multi-genre recommendations, and improving real-time performance to enhance user satisfaction.

6. References

1. Kumar R, Rakesh. Music Recommendation System Using Machine Learning. 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). Greater Noida, India; 2022:572–6. doi:10.1109/ICAC3N56670.2022.10074362.
2. Zeng Z, Wang Y, Zhao Y, Shi W. A Survey of Music Recommendation Systems. Proceedings of the 5th International Conference on Computer Information and Big Data Applications (CIBDA '24). Association for Computing Machinery; 2024:507–19. doi:10.1145/3671151.3671243.
3. Kodama Y, *et al*. A music recommendation system. Digest of Technical Papers. International Conference on Consumer Electronics (ICCE). Las Vegas, NV; 2005:219–20. doi:10.1109/ICCE.2005.1429796.
4. Gaikwad UV, Ghodake SS, Mokalkar RA, Jagtap HS. Emotion-based Music Recommendation System. International Journal of Innovative Science and Research Technology (IJISRT). 2022;7(12):542–5. doi:10.5281/zenodo.7490716.
5. James HI, *et al*. Emotion-Based Music Recommendation System. International Journal of Engineering and Technology (IRJET). 2019;6(3).
6. Babua T, Naira RR, Aa G. Emotion-Aware Music Recommendation System: Enhancing User Experience Through Real-Time Emotional Context. arXiv preprint arXiv:2311.10796.
7. Abdul A, Chen J, Liao HY, Chang SH. An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach. Applied Sciences. 2018;8(7):1103. doi:10.3390/app8071103.
8. Mahadik A, *et al*. Mood-Based Music Recommendation System. International Journal of Engineering Research & Technology. 2021;10(6).
9. Kaggle. FER2013 Dataset. Available from: https://www.kaggle.com/datasets/msambare/fer2013.