

Personalized Music Recommendation System Using Deep Learning

Miss. Eshwari Wankhade
Information Technology Department
Prof. Ram Meghe Institute of
Technology & Research, Badnera
Amravati, India

Miss. Vaishnavi Kulat
Information Technology Department
Prof. Ram Meghe Institute of
Technology & Research, Badnera
Amravati, India

Mr. Sanket Dok
Information Technology
Department
Prof. Ram Meghe Institute of
Technology & Research, Badnera
Amravati, India

Mr. Aditya Lahe
Information Technology Department
Prof. Ram Meghe Institute of
Technology & Research, Badnera
Amravati, India

Prof. (Dr.) A. W. Burange
Assistant Professor,
Information Technology Department
Prof. Ram Meghe Institute of
Technology & Research, Badnera
Amravati, India

Abstract— In this paper, Personalized Music Recommendation System using deep learning is proposed which integrates facial emotion recognition and machine learning and deep algorithms to provide personalized music suggestions based on a user's current emotion state. The system allows users to register and log in, following which their facial expression is captured in real-time using the OpenCV library. A Convolutional Neural Network (CNN) model is employed to detect the user's current emotion from their facial expression, classifying it into the seven categories: angry, disgust, fear, happy, sad, surprise, or neutral. Based on the detected emotion of user, the system recommends emotion-specific music tracks. The user's preferences are tracked over time, and a decision tree algorithm is applied to refine the recommendations based on both the user's emotional state and historical preferences. The system also features an admin panel where administrators can manage the music dataset, upload expression-labeled images for training, and classify new songs based on their lyrics into appropriate emotional categories. By combining real-time emotion detection and personalized music recommendations, this system enhances user experience and provides a dynamic music selection process tailored to individual emotional states. The experimental results demonstrate the effectiveness of CNN for emotion detection and the suitability of the decision tree for generating personalized recommendations.

I. INTRODUCTION

Music is an integral part of human life, often used as a medium to express, evoke, and influence emotions. With the rapid growth of digital music platforms, the demand for personalized music recommendation systems has grown exponentially. Conventional recommendation systems typically rely on collaborative filtering, content-based filtering, or a combination of both to suggest songs to

users. While these methods are effective in suggesting popular songs or songs similar to those already liked by the user, they do not account for the user's real-time emotional state, which can greatly affect the type of music one desires to listen to at any given moment. This limitation presents an opportunity to create a more dynamic and emotionally aware music recommendation system that can enhance user experience by tailoring music recommendations based on their emotions.

Proposed work is an Emotion-Based Music Recommendation System that integrates real-time emotion recognition using Convolutional Neural Networks (CNN) with a lyrics-based Random Forest classifier to suggest songs that align with the user's current emotional state. Furthermore, the system tracks user preferences over time, analyzing user interactions with the recommended songs to refine the music suggestions using a decision tree algorithm. The system identifies seven emotions mainly angry, disgust, fear, happy, sad, surprise, and neutral which correspond to different musical moods. This blend of emotion detection and personalized recommendation bridges the gap between the user's emotional context and their music preferences, providing a more immersive and emotionally responsive music selection process.

The system consists of two main components: Emotion Detection and Music Recommendation. The emotion detection component leverages CNN to analyze facial expressions captured by a webcam in real-time. This module assigns an emotion label to the user based on their current expression. Concurrently, the music recommendation module uses a Random Forest classifier to analyze song lyrics and classify them into one of the seven emotional categories. The final recommendations are not solely based on the detected emotion but also take into account the user's historical preferences and previous song choices, which are tracked using click-through data.

To facilitate the overall system, we plan to implement an admin panel where the administrator can upload new music datasets and images labeled with emotional expressions to further train and improve the CNN model. Additionally, the admin panel allows the classification of new songs based on their lyrics into appropriate emotional categories, ensuring the music recommendation system stays updated with fresh content. This system aims to enhance user satisfaction by dynamically adjusting to the user's emotional state, providing real-time, personalized music recommendations that reflect the user's mood. The combination of deep learning for emotion detection, machine learning for song classification, and user preference tracking presents a powerful, holistic approach to music recommendation. The experimental results demonstrate that such a system can significantly improve the emotional engagement of users with their music libraries, thereby offering a more enjoyable and personalized listening experience.

II. LITERATURE REVIEW

1. Emotion-Based Music Recommendation Systems

Emotion-based music recommendation systems grasp user emotions to provide personalized music experiences. Various studies have explored this intersection of affective computing and music retrieval.

For instance, **Li et al [1]**, proposed a music recommendation system that utilizes emotional information from the user to personalize music suggestions. Their approach integrated sentiment analysis of user-generated content, which helped refine recommendations based on detected emotional states. Similarly, **Hernandez et al [2]**, developed a hybrid music recommendation system that combines collaborative filtering with emotion detection, improve the relevance of recommendations by considering the emotional context of songs.

Another notable work by **Jiang et al [3]** utilized a multimodal approach, combining textual and acoustic features to classify songs based on emotions. Their findings highlighted that integrating multiple sources of information improved the accuracy of emotion classification, thereby enhancing recommendation quality.

2. Facial Emotion Recognition Techniques Facial emotion recognition (FER) is a critical component of emotion-based systems, enabling the capture of user emotions in real-time. CNNs have become the predominant method for FER due to their ability to learn hierarchical feature representations.

Zhang et al [4], presented a comprehensive review of CNN architectures for FER, outlining various models and their performance across standard datasets. Their work underscores the importance of feature extraction in achieving high accuracy in emotion recognition tasks. In another study, **Mollahosseini et al [5]**, developed a large dataset specifically for facial emotion recognition, known as AffectNet. They demonstrated that training CNNs on this dataset could significantly improve classification performance for various emotions. The effectiveness of their approach illustrates the value of high-quality, labeled datasets in training robust emotion recognition models.

3. Decision Tree Algorithms in Music Recommendation Decision tree algorithms are popular in machine learning for classification tasks due to their interpretability and ease of use. In the context of music recommendation, decision trees can efficiently classify songs based on user preferences and emotional attributes.

Yam et al [6], explored the application of decision trees in music recommendation systems, demonstrating how the model could effectively capture user preferences based on previous listening habits and contextual information. Their results indicated that decision trees could provide relevant recommendations while being transparent about the decision-making process.

Additionally, **Bishop et al [7]**, examined the hybrid use of decision trees and other machine learning techniques, such as clustering, to improve music recommendation outcomes. Their findings suggested that combining different algorithms could enhance the diversity and relevance of recommendations.

4. User Preference Tracking in Music Recommendation Systems

User preference tracking is essential for refining recommendation algorithms over time. Research has shown that incorporating user interactions can significantly improve the personalization of music recommendations. In a more recent study, **Karydis et al [8]**, investigated the impact of real-time user feedback on music recommendation systems. They found that incorporating immediate user interactions led to more relevant recommendations, suggesting that real-time tracking of user preferences is crucial for personalization.

To incorporate IoT concepts into personalized music recommendation systems, research work [10] highlights the foundational role of IoT in data connectivity and management across platforms. They discuss how IoT enables seamless data exchange between devices without direct human input, a feature valuable for real-time, adaptive recommendation engines. IoT's capacity to connect diverse devices and support autonomous data-driven decisions makes it essential for systems that need constant updates based on user preferences, thus aligning with the needs of personalized music recommendation.

The literature indicates a growing interest in emotion-based music recommendation systems, particularly those integrating facial emotion recognition and machine learning techniques. The combination of these fields offers promising avenues for enhancing user experience and engagement in music consumption. Future research could focus on improving the accuracy of emotion detection, refining recommendation algorithms, and exploring the impact of user feedback in real time.

III. PROPOSED METHODOLOGY

The proposed methodology for the Emotion-Based Music Recommendation System involves three key modules: emotion recognition using CNN, song classification using Random Forest, and personalized music recommendation based on emotion and user preferences. These modules interact with each other to create a seamless experience where the system dynamically recognizes the user's current emotional state and suggests emotionally relevant music. Below is a detailed explanation of each component.

1. Emotion Recognition Using CNN

Emotion recognition is a critical component of the system, as the user's real-time emotional state serves as the basis for music recommendation. Convolutional Neural Networks (CNN) are highly effective for tasks involving image classification and recognition. The CNN model in this system processes facial images captured via a webcam and classifies the user's current emotion into one of seven categories: angry, disgust, fear, happy, sad, surprise, and neutral.

CNN Architecture:

- **Input Layer:** The input to the CNN is a facial image captured from the user's webcam. The image is pre-processed (e.g., resized, normalized, and converted to grayscale) to prepare it for feature extraction.
- **Convolutional Layers:** The CNN applies several convolutional filters that extract important features from the facial image, such as eye movement, mouth shape, and eyebrow positioning. These layers detect low-level features (e.g., edges) in the initial stages and more complex patterns (e.g., facial expressions) in deeper layers.
- **Pooling Layers:** The pooling layers reduce the spatial dimensions of the image, making the model computationally efficient while retaining essential features.
- **Fully Connected Layers:** These layers take the feature maps generated by the convolutional and pooling layers and combine them to produce a final classification decision. The fully connected layers aggregate all the information learned by the previous layers and map it to an output emotion.
- **Output Layer:** The output is a probability distribution over the seven emotions, with the highest-probability emotion selected as the detected emotion.

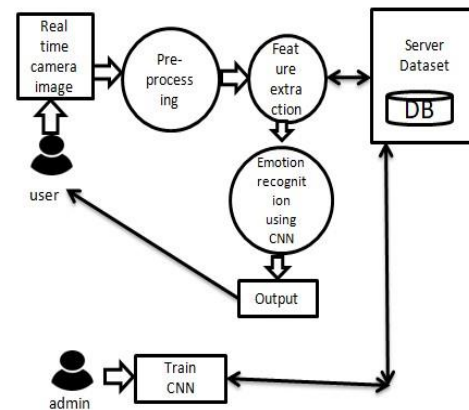


Fig. 1 Emotion Recognition using CNN

The Facial Expression Recognition (FER-2013) dataset is commonly used for training CNN models in facial emotion detection. This dataset contains over 35,000 labeled facial images representing seven different emotions. The training process involves feeding the CNN with labeled facial images and adjusting the model's weights using backpropagation to minimize classification errors. Data augmentation techniques, such as image rotation, flipping, and zooming, are applied to increase the robustness of the model against varying facial orientations and lighting conditions.

Once the model is trained, it is deployed to detect emotions in real time. The system captures the user's facial expression at regular intervals, classifies it, and uses this information to drive the music recommendation process.



Fig. 2 Emotion Categories

2. Song Classification Based on Lyrics Using Random Forest

Music recommendations in the system are driven not only by user emotions but also by the classification of songs based on their lyrics. Since song lyrics often reflect emotional themes, they can be analyzed to classify songs into emotional categories corresponding to the seven detected emotions. The Random Forest algorithm is employed for this classification task due to its robustness and ability to handle high-dimensional data, such as text. Random Forest for Lyrics Classification:

- **Feature Extraction from Lyrics:** Before applying the Random Forest classifier, song lyrics are processed using Natural Language Processing (NLP) techniques. Key NLP features include:

- Bag of Words (BoW): This technique represents the frequency of words in the lyrics. Commonly occurring words across emotionally similar songs form a basis for classification.
- TF-IDF (Term Frequency-Inverse Document Frequency): TF-IDF captures the importance of words in the lyrics by penalizing common words that appear in many songs and emphasizing those that are more unique to certain emotions.
- Word Embeddings: Advanced word embeddings such as Word2Vec or GloVe can be used to capture the semantic meaning of words, allowing the model to understand contextual relationships in lyrics.

Random Forest Model: Once the lyrics are converted into feature vectors using the above techniques, the Random Forest classifier trains multiple decision trees using subsets of the data. Each tree in the forest learns different rules for classifying the lyrics into one of the seven emotional categories. The final classification decision is made by aggregating the

- predictions of all decision trees, improving overall accuracy and reducing overfitting.

songs based on their emotional tone, in this work we use “..”. After training, the Random Forest classifier can automatically label new songs with one of the seven emotions, enabling real-time classification of new music added to the system.

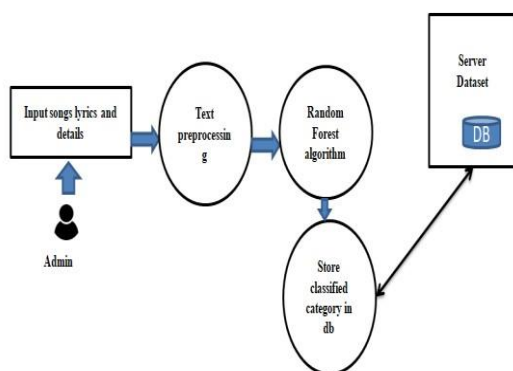


Fig. 3 Song Classification using Random Forest

Dataset for Lyrics Classification:

A labeled dataset containing song lyrics categorized by emotion is required for training the Random Forest classifier.

This dataset can be curated from publicly available music datasets or generated manually by labeling

3. Personalized Music Recommendation Based on Emotion and Preferences

The final step in the system is to generate personalized music recommendations based on both the detected emotion and the user's historical preferences. The Decision Tree algorithm is used to model user preferences by analyzing their interaction with the recommended songs.

Preference Tracking and Decision Tree:

- **Click-Through Data:** The system tracks which songs a user clicks on or listens to, capturing the emotional context of each interaction (i.e., the user's emotion at the time of the click and the emotion of the recommended song). This information is stored in a user profile, which grows richer over time.
- **Decision Tree-Based Recommendation:** The Decision Tree algorithm learns patterns from the click-through data, establishing decision rules based on the user's preferences. For example, if a user consistently prefers listening to "happy" songs when they are feeling neutral, the decision tree will prioritize recommending happy songs during neutral emotional states.

Hybrid Recommendation Model:

The system employs a hybrid approach by considering both the detected emotion and the user's long-term preferences. While the emotion recognition module provides a real-time snapshot of the user's mood, the preference model ensures that the recommended songs align with the user's listening habits and genre preferences, leading to a more personalized and satisfying music recommendation experience. The proposed work integrates CNN-based emotion detection, Random Forest for lyrics classification, and a Decision Tree for user preference tracking, resulting in a comprehensive music recommendation system that is both emotionally responsive and personalized. The system dynamically adapts to the user's real-time emotions while continuously learning from their preferences, enhancing the overall user experience.

IV. EXPECTED OUTPUT

In this personalized music recommendation system, the system will work as per the following flow-diagram shown below.

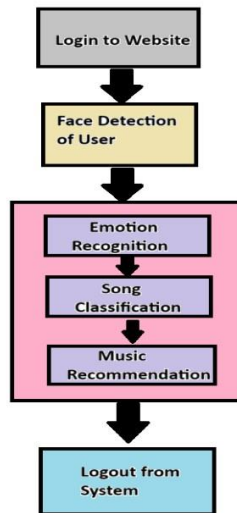


Fig. 4 Expected Output Flow-Diagram

Login to Website: The process starts with the user logging

This flowchart represents a system designed to provide music recommendations based on user emotions. Here's a breakdown of the different stages in the diagram: recommendations by aligning lyrical content with the detected emotions, into a website or application.

Face Detection of User: Once logged in, the system detects the user's face using face recognition technology. This likely initiates the next step of emotional analysis.

Emotion Recognition: The system then analyzes the user's face to recognize their current emotional state, such as happiness, sadness, anger, etc.

Song Classification: Based on the detected emotion, the system classifies songs that match the user's mood. This classification could be based on predefined categories of emotions and music genres or moods.

Music Recommendation: The system recommends music that corresponds to the classified emotion. This could involve selecting songs from a database based on mood, genre, personal preference

Logout from System: After receiving music recommendations, the user can log out of the system.

CONCLUSION:

The Emotion-Based Music Recommendation System integrates advanced machine learning techniques, including Convolutional Neural Networks (CNNs) for facial emotion recognition and Random Forest algorithms for lyrics-based classification, to deliver personalized, emotionaligned music recommendations. By detecting the user's emotional state through facial expressions and classifying them into seven

core emotions—angry, disgust, fear, happy, sad, surprise, and neutral—the system provides music tailored to the user's mood. The use of Random Forests for analysing song lyrics enhances the relevance of pp. 357-360, doi: 10.1145/3383583.3398623.

Additionally, the system tracks user interactions over time,

refining its future suggestions based on evolving preferences, which addresses the limitations of static models. This hybrid approach ensures that recommendations are not only emotionally relevant but also personalized, leading to higher user satisfaction and engagement. Literature supports the effectiveness of combining facial emotion recognition and adaptive learning in enhancing music recommendations. Future improvements could involve incorporating contextual data, such as time or location, and analysing more nuanced emotional states, further improving the system's accuracy and responsiveness in real-time music streaming services.

REFERENCES:

- [1] **H. Tran, T. Le, A. Do, T. Vu, S. Bogaerts, and B. Howard**, "Emotion-aware music recommendation," in Proc. 37th AAAI Convention on Artificial Intelligence (AAAI-23), 2023, pp. 16087-16095.
- [2] **K. Kushwaha and S. Sharma**, "A Hybrid Collaborative Filtering-Based Music Recommendation System," International Journal of Recent Development in Engineering and Technology, vol. 11, no. 10, pp. 15-23, Oct. 2022.
- [3] **R. Liu and X. Hu**, "A Multimodal Music Recommendation System with Listeners' Personality and Physiological Signals," in Proc. ACM/IEEE Joint Conf.
- A. Dixit and T. Kasbe**, "A Survey on Facial Expression Recognition using Machine Learning Techniques," in Proc. 2020 IEEE 2nd Int. Conf. Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 1-6, doi: 10.1109/IDEA49133.2020.9170706.
- [4] **A. Mollahosseini, D. Chan, and M. H. Mahoor**, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," IEEE Transactions on Affective Computing, vol. 10, no. 1, pp. 18-31, Jan.-Mar. 2019, doi: 10.1109/TAFFC.2017.2740923.
- [6] **A. Utku, H. Karacan, O. Yıldız, and M. Akçayol**, "Implementation of a New Recommendation System Based on Decision Tree Using Implicit Relevance Feedback," Journal of Software, vol. 10, no. 12, pp. 1367-1374, 2015, doi: 10.17706/jsw.10.12.1367-1374.

- [7] **M. M. Rahman, I. A. Shama, M. S. Rahman, and M. R. Nabil**, "Hybrid Recommendation System to Solve Cold Start Problem: A Survey," Journal of Theoretical and Applied Information Technology, vol. 100, no. 11, pp. 35623575, June 2022.
- [9] **Y. Xie and L. Ding**, "A Survey of Music Personalized Recommendation System," School of Computer, Guangdong University of Technology, Guangdong, China, 2020.
- [10] **Burange, Anup & Misalkar, Harshal.**, Review of Internet of Things in development of smart cities with data management & privacy. 189-195. 10.1109/ICACEA.2015.7164693., 2015.