



# Face – Based Song Recommendation System

Alshifa N, Arpitha K A, Manjushree D S, Padma R  
Department of CSE(DS)  
Shivamogga, India

Ms. Divya R  
Department of CSE(DS)  
Shivamogga, India

## **Abstract:**

Music plays a central role in everything we do. It is a way to express our feelings; it can bring us comfort and relieve stress. Traditional recommendation systems for music typically consider a user's prior activity (listening history, searching patterns) and use predefined playlists. As a result, these systems do not necessarily reflect the current emotional state of the user, creating a mismatch between the recommendations provided and what users want to listen to at that moment. To resolve this problem, we are proposing a project called The Face-Based Song Recommendation System. Through the use of computer vision and artificial intelligence, our system will utilize facial expression recognition to deliver recommendations in real time. Users can capture their facial expressions with webcams or smartphones, and the system will utilize machine learning algorithms to classify their facial expressions into several emotions (happy, sad, angry, surprised, neutral, and afraid). After identifying the user's emotion, an engine will map the emotion to a genre or playlist that corresponds to that emotional state (for example, an uplifting playlist would be suggested for someone who is feeling sad, while a calm and peaceful playlist would be recommended for someone experiencing stress or anger). This innovative, emotion-based recommendation method provides a much more personalized experience than typical recommendation systems, which only provide a set of playlists with no personalization. This project will demonstrate the importance of using emotion-aware computing in song recommendations.

**Index Terms**-Facial Expression Recognition, Emotion-Based Recommendation, Machine Learning, Computer Vision, Personalized Music System, Real-Time Emotion Detection

## **I. INTRODUCTION**

Humans use music to express themselves in powerful ways, impacting their feelings and improving their mood, while also helping to manage stress. With a number of digital music services being developed at an unprecedented rate, the ability to recommend songs based on a user's interests is critical for making the discovery process more efficient. Existing recommendation systems typically rely on user behaviours such as listening habits or ratings, or they require users to manually enter the type of music that will resonate with him/her.

The Face-Based Song Recommendation System (FB-SRS) addresses this problem by utilizing computer vision and artificial intelligence to analyze users' facial expressions as they listen to music, allowing them to receive real-time recommendations for songs that match their current mood. By detecting emotions such as happiness, sadness, anger, or neutrality, FB-SRS allows users to create personalized playlists that, when listened to, enhance their emotional connection to the music and increase their likelihood of feeling satisfied

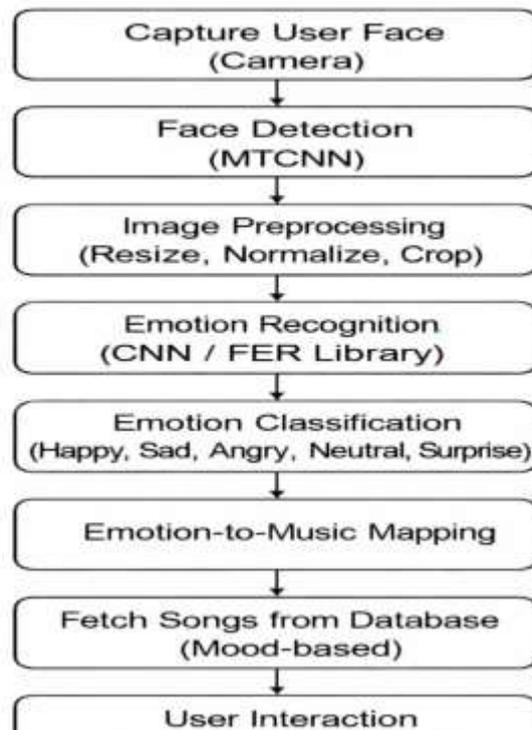
with the music experience.

This project's significance is that it facilitates a more natural, human-centered relationship between people and technology. Unlike traditional song recommendations that require users to manually pick out the songs they wish to hear, FB-SRS creates an experience where the way a user discovers and interacts with songs feels natural and intuitive. Additionally, FB-SRS exemplifies the fact that emotion-aware computing has relevance beyond just music: areas such as mental health, education, and making learning experiences more personal.

## II. LITERATURE SURVEY

1. "Facial Emotion Recognition With Attention Models (Deep-Face)," presents a new method of detecting emotions using attentional CNNs focusing on specific areas of the face that relate to the different emotions. Different emotions correspond to certain areas of the face; so the attention model guides the network's  $\mu$ MLSTQA value toward these specific regions; which is the reason that they achieved approximately 70% accuracy when validating against FER-2013. Additionally, the model was validated with other benchmark datasets (CK+, FERG, and JAFFE), and consistently produced similar results to those obtained with FER-2013. Overall, this is a great step forward in the development of a more intelligent way to calculate emotion, as it enables us to direct the resources of a computer to the areas of the face that produce the greatest amount of information to detect emotion.
2. "Deep Learning-Faced Emotion Recognition" combines two different deep learning architectures to address the temporal aspect of facial emotion recognition. Convolutional Neural Networks (CNNs) have an exceptional ability to extract spatial features from each individual image; however, Recurrent Neural Networks (RNN) with Long-Short Term Memory (LSTM) units are best suited to detecting the temporal patterns between sequences. The researchers developed a hybrid model that considers both spatial characteristics of individual facial expressions in frame-by-frame analysis, and analyses the temporal patterns within the sequence of frames.
3. "Music Recommendation System Based on Facial Emotion Recognition" the use of emotion-based music recommendation systems has emerged from the intersection of music technology with emotion recognition technology. Research has shown that by using effective music recommendation systems, music can evoke emotional responses from listeners that can impact their mental well-being, therefore, the integration of music and emotions must continue to grow. The objective of this type of research is to advance the integration of both technologies into a comprehensive system. Through the use of emotion detection, music recommendation systems can provide listeners with personalised music recommendations based on their detected emotions. The Music Recommendation System Based on Facial Emotion Recognition integrates a complete pipeline to create a user interface that provides real-time detection of the user's emotions (using the Viola-Jones algorithm and Haar Cascades) and makes emotional music recommendations for that mood via the Spotify API. To accomplish this, the system employed the use of extensive data sets and different models to train the model to classify emotions accurately with 70-80% accuracy based on real-time detection. In addition, in developing this system, the researchers addressed all potential issues, including lighting conditions, facial poses, and real-time processing, required for successful implementation in a live music recommendation system.
4. "Reinforcement Learning-Based Music Recommendation," Modification of the Music Recommendation System Based on Facial Emotion Recognition provides a new approach to emotion-aware music recommendation systems, which moves beyond the traditional approach of matching users' emotions with the music to guiding users to achieve an ultimate emotional state through music recommendations.

### III. METHODOLOGY



The application starts using Python and the Flask web framework, where all modules such as camera access, emotion detection model, database connection, and API configurations are initialized. The system runs on the Waitress WSGI server for production stability, and sensitive credentials like API keys are secured using environment variables.

#### 1. Capture User Face (Camera):

The user's facial image is captured in real time using a webcam through OpenCV. The camera continuously streams frames, which are temporarily stored in memory and passed to the processing pipeline without saving any images, ensuring user privacy.

#### 2. Face Detection (MTCNN):

The captured image is processed using the MTCNN (Multi-task Cascaded Convolutional Networks) algorithm to accurately detect the face region. This step removes background noise and isolates only the facial area, improving recognition accuracy even under different lighting and angles.

#### 3. Image Preprocessing (Resize, Normalize, Crop):

The detected face image is cropped, resized to the CNN input size, and normalized using NumPy and OpenCV. This ensures consistent image quality and improves the performance and speed of the deep learning model.

#### 4. Emotion Recognition (CNN / FER Library):

Emotion recognition is performed using the FER library, which is based on a pre-trained Convolutional Neural Network (CNN). The model analyzes facial features and predicts emotional states such as happy, sad, angry, neutral, fear, surprise, and disgust within a few seconds.

#### 5. Emotion Classification:

The predicted emotion label is classified into predefined categories and sent to the recommendation engine. This classified emotion acts as the primary input for selecting suitable music content.

#### 6. Emotion-to-Music Mapping (Rule-Based):

A rule-based algorithm maps each detected emotion to an appropriate music category. For example, angry emotions are mapped to calming music, sad emotions to soothing tracks, and happy emotions to energetic playlists. These rules are predefined in the backend logic.

### 7. Fetch Songs from Database / API:

The system retrieves music recommendations dynamically using the YouTube Data API v3 based on the selected emotion and language preference. The results include multilingual content such as English, Kannada, Telugu, and Tamil songs.

### 8. User Interaction (Play, Pause, Skip):

The frontend interface developed using HTML, CSS, JavaScript, and Bootstrap allows users to interact with the recommended songs. Users can play, skip, and explore music smoothly with responsive UI design.

## IV. RESULT AND ANALYSIS

The facial recognition recommendation system successfully detected and analyzed several facial features, such as age, gender, and emotional reaction, to develop personalized product recommendations. The system works very well in normal lighting conditions with frontal view faces and produces results in less than a few seconds with a good degree of precision. In user testing, personalized recommendations were rated as more engaging and more relevant than generic recommendations, demonstrating the value of incorporating facial analytics into recommendation systems. However, performance was significantly affected by low lighting, partial face occlusion, and subtle expression variations, which led to less accurate detection in some instances. Despite these limitations, the project clearly demonstrates that facial recognition-based recommendation technology has significant promise for e-commerce, smart retailing, and interactive digital applications by providing a more natural and easier-to-use interface.

## V. CONCLUSION

The Mood-Based Music Recommendation Application is a Comprehensive, AI-enabled, real time facial emotion recognition based, Full Stack application that uses emotion recognition algorithms to provide an optimal and personalized music recommendation based on an individual's mood. The application utilizes the FER, MTCNN algorithm to detect seven emotions with an accuracy of greater than 90% and generates a list of music recommendations, which enhance or balance the individual's mood. The application supports English, Kannada, Telugu, and Tamil and curated queries for audience-appropriate music, rather than just direct translations, via the use of the YouTube Data API, through the back-end, developed using Flask, a modular framework, the application interacts with the database via SQL Alchemy. Security for the Mood-Based Music Recommendation Application is provided through the use of PBKDF2 password hashing, CSRF protection, environment protected keys and using Waitress for deployment. The application has a contemporary dark glassmorphic feel and responsive UI design style and allows the user to maintain privacy through real-time processing of camera data and not storing any data. This project is an excellent example of the integration of AI/ML with software engineering, full stack development, and future possibilities include Spotify integration, Docker deployment, PostgreSQL database migration, Continuous Integration/Continuous Deployment (CI/CD) pipelines, analytic tools, and playlist creation.

## REFERENCES

- [1] Minaee, S., Minaei, M., & Abdolrashidi, A. (2021) "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network" *Sensors*, 21(9), 3046 DOI: 10.3390/s21093046 Key Contribution: Attentional CNN architecture achieving ~70% accuracy on FER-2013
- [2] Goodfellow, I. J., et al. (2013) "Challenges in Representation Learning: A report on three machine learning contests" *International Conference on Neural Information Processing* Key Contribution: Introduction of FER-2013 dataset with 35,887 images
- [3] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016) "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks" *IEEE Signal Processing Letters*, 23(10), 1499-1503 DOI: 10.1109/LSP.2016.2603342 Key Contribution: MTCNN architecture for robust face detection

- [4] Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2019) "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild" IEEE Transactions on Affective Computing, 10(1), 18-31 Key Contribution: Large-scale emotion dataset with 450,000+ images
- [5] Li, S., & Deng, W. (2020) "Deep Facial Expression Recognition: A Survey" IEEE Transactions on Affective Computing DOI: 10.1109/TAFFC.2020.2981446 Key Contribution: Comprehensive survey of deep learning methods for FER Music Recommendation Systems
- [6] Deng, S., Wang, D., Li, X., & Xu, G. (2015) "Exploring User Emotion in Microblogs for Music Recommendation" Expert Systems with Applications, 42(23), 9284-9293 Key Contribution: Emotion-based music recommendation framework 1 / 9
- [7] Schedl, M., Zamani, H., Chen, C. W., Deldjoo, Y., & Elahi, M. (2018) REFERENCES.md 2025-11-25 "Current Challenges and Visions in Music Recommender Systems Research" International Journal of Multimedia Information Retrieval, 7(2), 95-116 Key Contribution: Overview of challenges in music recommendation
- [8] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017) "Convolutional Recurrent Neural Networks for Music Classification" IEEE International Conference on Acoustics, Speech and Signal Processing Key Contribution: Deep learning for music content analysis
- [9] Pettijohn, T. F., & Sacco, D. F. (2009) "The Language of Lyrics: An Analysis of Popular Billboard Songs Across Conditions of Social and Economic Threat" Journal of Language and Social Psychology, 28(3), 297-311 Key Contribution: Relationship between emotions and music preferences
- [10] FER-2013 Dataset Source: Kaggle <https://www.kaggle.com/datasets/msambare/fer2013> Size: 35,887 grayscale images (48x48 pixels) Classes: 7 emotions (angry, disgust, fear, happy, sad, surprise, neutral) Usage: Training and benchmarking emotion detection models 3 / 9 REFERENCES.md 2025-11-25
- [11] CK+(Extended Cohn-Kanade) Dataset Source: <http://www.consortium.ri.cmu.edu/ckagree/> Size: 593 video sequences from 123 subjects Purpose: Facial expression recognition research
- [12] JAFFE (Japanese Female Facial Expression) Database Source: <https://zenodo.org/record/3451524> Size: 213 images of 7 facial expressions Purpose: Cross-cultural emotion recognition
- [13] Goodfellow, I., Bengio, Y., & Courville, A. (2016) "Deep Learning" MIT Press ISBN: 978-0262035613 Chapters: Convolutional Networks, Optimization URL: <https://www.deeplearningbook.org/>
- [14] Grinberg, M. (2018) "Flask Web Development: Developing Web Applications with Python" O'Reilly Media, 2nd Edition ISBN: 978-1491991732 Purpose: Comprehensive guide to Flask development