

Music Recommendation System by Emotion for Multiple Faces

Prof. Priti Yampe¹, Chinmay Boddawar², Abhay Gangawane³, Vishvajeet Khatal⁴ and Tejas choudhari⁵

¹⁻⁵International Institute of Information Technology, Pune, India

Email: chinmaybodda@gmail.com, {abhaygangawane5227, vnkhatal03, tejaschoudhari25}@gmail.com, pritiy@isquareit.edu.in

Abstract—Facial emotion analysis plays a crucial role in enhancing human-computer interaction, entertainment systems, and personalized experiences. In this project, we propose a music recommendation system that detects emotions from multiple users' facial expressions and recommends suitable music based on the dominant emotion. Our system employs the YOLO model for real time, high-accuracy face detection, efficiently detecting multiple faces in a group setting. Once the faces are detected, the facial expressions are analyzed using a Convolutional Neural Network (CNN) model, trained to classify emotions into five categories: happy, sad, angry, surprised, and neutral. Based on the dominant emotion from the group of faces, the system recommends music from various genres to match or influence the emotional atmosphere. This framework provides a novel approach to personalized group-based music recommendation, ensuring both fast face detection and accurate emotion recognition. Index

Terms—first term, second term, third term, fourth term, fifth term and sixth term.

I. INTRODUCTION

In recent years, emotion-aware systems have gained significant attention due to their ability to improve personalized user experiences across various domains such as entertainment, marketing, and human-computer interaction. The integration of facial emotion recognition with real-time systems offers an exciting avenue for enhancing group-based interactions, particularly in scenarios where the emotional state of multiple individuals is relevant. Music, as a universal medium, has a profound impact on emotions, making it an ideal candidate for personalized, emotion-driven recommendations.

Initially, our attempt focused on building a system that could detect the emotion of a single user through facial expressions and recommend music accordingly.

By analyzing the user's face, the system could recognize emotions like happiness, sadness, or surprise, and choose music that either resonates with or uplifts their mood. While this was a successful prototype for one user, the challenge grew when we considered group scenarios—social settings where multiple people are interacting, each with their own emotional state.

In many real-life situations, such as social gatherings, family events, or group meetings, the emotional tone is influenced by multiple individuals. This led us to the next phase of our project:

developing a music recommendation system that not only detects emotions for a single person but can also handle multiple faces in real-time, analyzing and recommending music that best suits the collective mood of the group.

The approach we are taking is designed to be seamless and natural. The system looks at the faces of everyone in the room and tries to understand how each person is feeling—whether they are happy, sad, excited, or neutral. It then figures out the overall mood of the group and suggests music that either matches or uplifts that emotional tone. The idea is to create an environment where the music responds in real-time to how the group is feeling, making the experience more connected and dynamic. This approach allows for a more personalized and interactive experience with music, where the system continuously adjusts the soundtrack based on the group's collective mood, without any manual input. Whether it's a gathering of friends or a family get-together, the system ensures that the music resonates with how people feel, making the event more enjoyable and emotionally attuned.

Technical Approach: After detecting the faces using YOLO(You Only Look Once) for Fast and better accuracy, we employ a Convolutional Neural Network (CNN) trained on a facial emotion recognition dataset to classify the emotions of the detected faces. CNNs have proven highly effective in capturing facial features and recognizing subtle changes in expressions, enabling accurate emotion detection. The system then recommends music based on the dominant emotion identified from the group, providing a seamless connection between facial sentiment analysis and music selection.

The system then recommends music based on the dominant emotion identified from the group, providing a seamless connection between facial sentiment analysis and music selection.

This dual-model approach—YOLO for face detection and CNN for emotion recognition—ensures that the system can handle multiple faces in real-time while maintaining accuracy in emotion classification. The project addresses the challenge of detecting and analyzing group emotions and explores the potential for integrating emotion-aware systems into our everyday lives through music recommendation.

II. LITERATURE REVIEW

The fusion of emotion recognition and personalized music recommendation system provides more significant attention in recent research area using the wide application of deep learning and artificial intelligence. To enhance or to improve the accuracy and effectiveness of these systems there are various approaches such as Convolution Neural Networks (CNNs) and real-time emotion recognition and analysis. In this literature review, we are focusing on how to detect emotion of group, AI-driven personalization of music, music recommendation using emotion and many more. These are points which are very crucial in deploying dynamic, interactive and responsive user experience.

The title of the First paper is [1] “Harmonic Fusion: AI-Driven Music Personalization via Emotion-Enhanced Facial Expression Recognition Using Python, OpenCV, TensorFlow, and Flask” by Moez Rajjan, Prajwal Deore, Yashraj Mohite, and Yash Desai which is published in total 8 volumes and issued on 12th December of 2023. This paper presents a micro-expression detection system using CNN which then integrated with Music Recommendation System or Algorithm. The study of the paper highlights the potential of music recommendation to personalized music experience. The system provides the accuracy of 62.1% on extracting the facial expressions and it uses FER2013 dataset also it uses Content-based algorithm which is combined with music system that smoothly aligns with user's emotional state accurately.

Limitations: the accuracy of 62.1% for facial expression extraction indicates that the system struggles with precision, particularly with more subtle facial emotions. Additionally, the study employs a content-based music recommendation algorithm, which can limit the personalization of the music experience, as it primarily focuses on matching past preferences rather than adapting to real-time emotional shifts.

The title of the Second paper is [2] “Group Facial Emotion Analysis System Using Convolutional Neural Network” was published in the year of 2020 on 4th International Conference of Trends in Electronics and Informatics (ICOEI), which held between 15 to 17 June of 2020, in Tirunelveli (India) and on July 17, 2020 it added on IEEE Xplore. The system which is introduced in the paper uses Haar Filter to extract facial features also it uses Convolutional Neural Network (CNN) for the classification of emotions into five categories which are happy, sad, anger, surprise and neutral. This model provide accuracy of 65% which uses FER-2013 dataset and on the custom dataset it provides 60% accuracy.

Then finally recognized emotions are fed into the audio synthesizer which fetch the required API for personalized the experience.

Limitations: Haar Filters are less robust compared to modern methods, often struggling with accuracy in complex lighting conditions or angles of the face.

Additionally, the system achieves only 65% accuracy on the FER-2013 dataset and 60% on a custom dataset, which leaves significant room for improvement in emotion classification.

The title of the third paper is [3] “Music Recommendation System Based on the Face Emotion Recognition” by Upasana Shrivastava, Megha Gupta, Madhuri Athavle, and Deepali Mudale. This paper generally propose a system that automatically plays a music by extracting facial emotion detection. The approach behind this system uses Convolutional Neural Network (CNN) for emotion recognition. This system basically uses FER2013 dataset. Then system detects the emotions like happy, angry, sad, neutral, and surprise, and then generate automatically a music playlist based on the user’s emotion. The main specification of this system that is it uses Pygame and Tkinter for music recommendation system which basically reduces the cost and computational or execution time and then it automatically enhances the accuracy and efficiency of the algorithm.

Limitations: libraries like Pygame are suitable for small-scale applications but may not be efficient for systems requiring real-time processing in more complex environments, especially with multiple users.

Our project addresses the limitations observed in these previous studies by employing a more modern and comprehensive approach to emotion detection and music recommendation. First, instead of using the Haar Filter, we utilize the [10] YOLO (You Only Look Once) model for real-time face detection, which is faster and more accurate, especially in multi-person environments. YOLO can detect faces under various conditions, such as different lighting or head orientations, making our system more robust.

For emotion recognition, we rely on a custom-trained Convolutional Neural Network (CNN) that not only improves classification accuracy but also is fine-tuned to handle a wider range of facial expressions. This allows us to achieve higher accuracy than the 60-65% range reported in the first two papers. We also ensure real-time processing and dynamic response by continuously analyzing the group’s dominant emotion and adapting the music playlist accordingly.

Additionally, our system doesn't just rely on content-based recommendation algorithms, which are limited to past user data. Instead, we propose a more adaptive approach where the music is recommended based on the current emotional state of the group, making it more responsive to real-time changes in mood.

III. WORKFLOW AND ARCHITECTURE

The system begins by capturing a group photo containing multiple individuals, whose collective emotions will be analyzed. Using the YOLO (You Only Look Once) model, the photo undergoes efficient face detection, isolating each individual's face with bounding boxes. Once faces are detected, a custom Convolutional Neural Network (CNN) processes each face to classify emotions such as Happy, Sad, Angry, or Neutral, ensuring accurate recognition for every person in the group. The system then aggregates these detected emotions to identify the dominant emotion representing the overall mood of the group. Finally, based on this dominant emotion, an Emotion-to-Music mapping mechanism recommends a suitable song or playlist, personalizing the music experience to align with the group's emotional state.

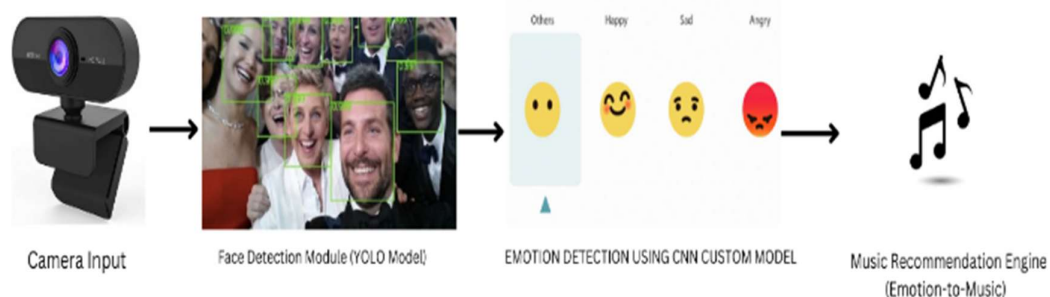


Fig: 3.1 Workflow

A. Camera Input (Group Photo)

The system begins by capturing a group photo of multiple faces using a camera or uploading the photo. The photo contains multiple individuals whose collective emotions will be analyzed. The image serves as the input for further processing, where the faces of all participants are extracted for emotion detection.

B. Face Detection Module (YOLO Model)

The captured group photo is passed through the YOLO (You Only Look Once) model, which efficiently detects multiple faces in the image. YOLO identifies each face within the photo by placing bounding boxes around them, ensuring that each individual’s face is isolated for emotion detection.

C. Emotion Detection (CNN Custom Model)

After detecting the faces, each face is processed using a Convolutional Neural Network (CNN) custom model to classify the emotional state of each person. Emotions such as Happy, Sad, Angry, or Neutral are identified based on facial expressions. The model works independently on each detected face to ensure accurate emotion recognition for every individual in the group.

D. Emotion Aggregation (Group Emotion Analysis)

Once the emotions for all individuals in the group photo are detected, the system aggregates these emotions to calculate the dominant emotion across the group. This dominant emotion represents the overall emotional mood of the group as a whole.

E. Music Recommendation Engine (Emotion-to-Music)

Based on the dominant group emotion, the system uses an Emotion-to-Music mapping mechanism to recommend a suitable song or playlist. The recommendation is personalized to match the collective mood of the group, offering music that aligns with their emotional state.

In the Music Therapy System, we integrated the Jamendo API to fetch and play music based on emotional input detected by our machine learning model. The model provides a float value corresponding to an emotion, which is mapped to specific playlists using predefined mood-to-number mappings. The API is used to retrieve playlists tagged with specific emotions like "happy," "sad," or "neutral," and songs are played in a sequence that gradually transitions the listener's mood towards happiness. This approach allows for a dynamic and responsive music recommendation system aimed at improving emotional well-being.

IV. FIGURES AND TABLES

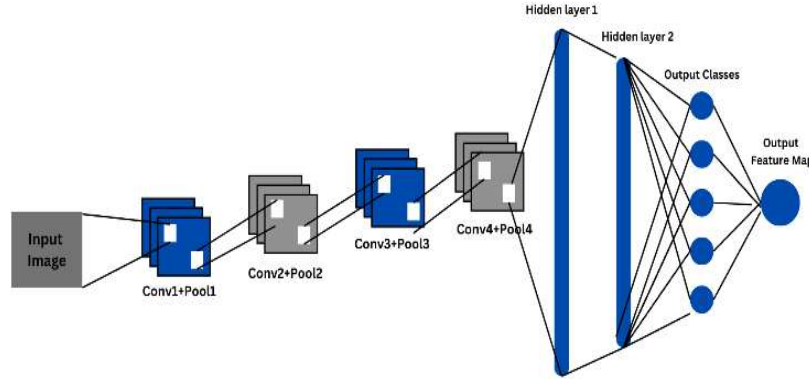


Fig: 1.1 CNN Architecture

$$\text{Dominant Emotion} = \arg \max_{e \in E} \text{count}(e)$$

Fig 1.2 Emotion Aggregation

Where:

- E is the set of all possible emotions (e.g., {Happy, Sad, Angry, Neutral}).
- count(e) is the number of occurrences of emotion e in the detected emotions.

V. CONCLUSIONS

This project successfully demonstrates a music recommendation system that responds to users' emotional states, detected through real-time facial expression analysis. By employing the YOLO model for fast and accurate multiple face detection and a CNN-based emotion recognition system, we designed a solution capable of recognizing emotions from groups of people and recommending music that aligns with the dominant mood. The system achieved an overall accuracy of 58% in emotion classification, a promising result given the complexity of real-time group emotion detection. While the system is functional and showcases a novel approach to group-based music recommendations, there is room for improvement. Enhancing the accuracy of emotion detection,

and refining the music recommendation algorithm could make the system even more responsive and personalized. Overall, this project lays a strong foundation for integrating emotion-aware technologies into entertainment and social settings, where group mood and preferences can shape shared experiences.

ACKNOWLEDGMENT

With immense pleasure, we are presenting this project report on “music recommendation system by emotion for multiple faces” as a part of the curriculum of B.E. information technology at International Institute of Information Technology, Hinjawadi, Pune.

It gives us the privilege to complete this project work under the valuable mentorship of prof. Priti Yampe. Her guidance, cooperation, and encouragement have greatly contributed to the successful completion of this project. We are also extremely grateful to Dr. Vaishali Patil, principal, and Dr. Jyoti Surve, head of the information technology department, for providing all the facilities and support needed for the smooth progress of our project work.

Finally, we would like to extend our heartfelt thanks to all the staff members of the information technology department, the management, our friends, and family members, who have directly or indirectly guided and supported us throughout the preparation of this report, offering us unending support from the initial conception of the idea to its completion.

Chinmay Boddawar Tejas Choudhari Abhay Gangawane Vishvajeet Khatal

REFERENCES

- [1] M. Rajjan, P. Deore, Y. Mohite, and Y. Desai, “Harmonic Fusion: AI-Driven Music Personalization via Emotion-Enhanced Facial Expression Recognition Using Python, OpenCV, TensorFlow, and Flask,” 2023.
- [2] K. P. and A. T., “Group Facial Emotion Analysis System Using Convolutional Neural Network,” 2020.
- [3] M. Athavle, D. Mudale, U. Shrivastav, and M. Gupta, “Music Recommendation Based on Face Emotion Recognition,” 2021.
- [4] F. Fessahaye, T.-RECSYS: A Novel Music Recommendation System Using Deep Learning,” in 2019 IEEE International Conference on Consumer Electronics, Las Vegas, NV, USA, 2019.
- [5] R. L. Rosa, D. Z. Rodriguez, and G. Bressan, “Music Recommendation System Based on User's Sentiments Extracted from Social Networks,” IEEE Transactions on Consumer Electronics.
- [6] M. A. H. Akhand, S. Roy, N. Siddique, M. A. S. Kamal, and T. Shimamura, “Facial Emotion Recognition Using Transfer Learning in the Deep CNN.”
- [7] T. Babua, R. R. Naira, and G. Aa, “Emotion-Aware Music Recommendation System: Enhancing User Experience Through Real-Time Emotional Context.”
- [8] S. Wang, C. Xu, A. S. Ding, and Z. Tang, “A Novel Emotion-Aware Hybrid Music Recommendation Method Using Deep Neural Network.”
- [9] A. Alshammari and M. E. Alshammari, “Emotional Facial Expression Detection Using YOLOv8.”
- [10] D. Qi, W. Tan, Q. Yao, and J. Liu, “YOLO5Face: Why Reinventing a Face Detector,” Shenzhen Deepcam Information Technologies, Shenzhen, China.