Music Recommendation System Using Facial Expression Detection

Ishika Singh
Dept. of Computer Science
and Engineering
SRM Institute of Science
and Technology
NCR

Abstract— Music has the unique power to resonate with human emotions, making it an ideal medium for creating a personalized experience. By leveraging a Convolutional Neural Network (CNN) for emotion recognition, the system captures facial features from webcam images to identify distinct emotional states—such as happiness, sadness, anger, and surprise. Based on the detected mood, the system curates a YouTube playlist that complements the user's emotions, enhancing their listening experience with music that aligns with their current state.

This system consists of two main stages: a training phase where the model is trained to identify facial cues associated with various emotions, and a real-time detection phase where the user's current mood is continuously monitored. Utilizing deep learning, image processing techniques allows precise emotion recognition, leading to accurate music recommendations. This approach enhances personalization while showcasing the potential of AI-driven solutions in recommendation systems. Future improvements may include integrating advanced emotion detection algorithms, expanding the music library for broader cultural relevance, and enhancing adaptability to different user preferences for an even more immersive experience.

Keywords— Facial expression recognition, emotion-based music recommendation, real-time detection, CNN, YouTube integration, artificial intelligence, dynamic monitoring.

INTRODUCTION

In today's digital era, music recommendation systems such as Spotify and Gaana have become essential tools for helping users discover music based on their past listening habits and preferences. However, these platforms often lack a crucial element—the ability to dynamically adapt to a user's current emotional state. Music has a unique ability to resonate with our emotions, whether through uplifting melodies or soothing rhythms, making it a powerful tool for enhancing mood and well-being. By incorporating facial expression-based emotion analysis, our study aims to leverage technology to create a more personalized and emotionally immersive music experience.

This approach groups emotions into distinct categories, including happiness, sadness, anger, emotions into distinct categories, including happiness, sadness, anger, surprise, and neutrality, allowing for real-time adaptation to the user's mood and providing song recommendations that align with their emotional needs interaction systems, are significantly impacted by this technology.

Facial emotion recognition (FER) has emerged as a significant field within artificial intelligence, offering practical applications in areas such as healthcare, customer service, and human-computer interaction. For this project, FER technology leverages convolutional neural networks (CNN) to interpret facial expressions, enabling the system to identify emotions with accuracy and immediacy. Once the user's emotion is detected, our model recommends a playlist from YouTube that best suits their mood, enhancing the overall listening

experience and reducing the time spent manually curating playlists. This method not only underscores the versatility of FER in creating intelligent, emotion-aware systems but also demonstrates the potential of AI to enrich digital music services by offering a responsive and tailored musical companion.
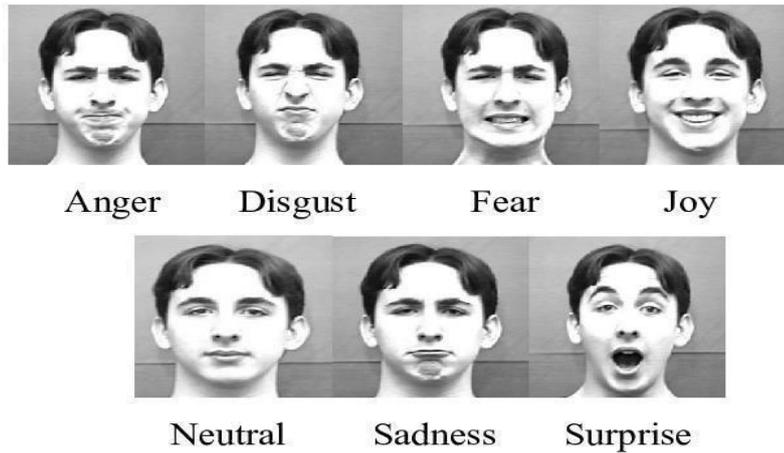
Our system addresses a growing demand for more emotionally intuitive technology by blending psychology leveraging artificial intelligence to interpret and respond to human emotions through music. Unlike traditional recommendation algorithms that rely heavily on past behaviour or general trends, this system provides real-time analysis by detecting the user's facial expressions and translating them into mood- based song suggestions. By categorizing songs into specific emotional themes, such as happiness, sadness, or surprise, the model can curate a playlist that aligns closely with the user's current state of mind. This approach not only enhances user engagement while also providing an immersive experience where music serves as a companion, offering comfort, joy, or relaxation as needed. This real-time adaptability makes our system uniquely capable of offering an unlimited and diverse selection of songs.
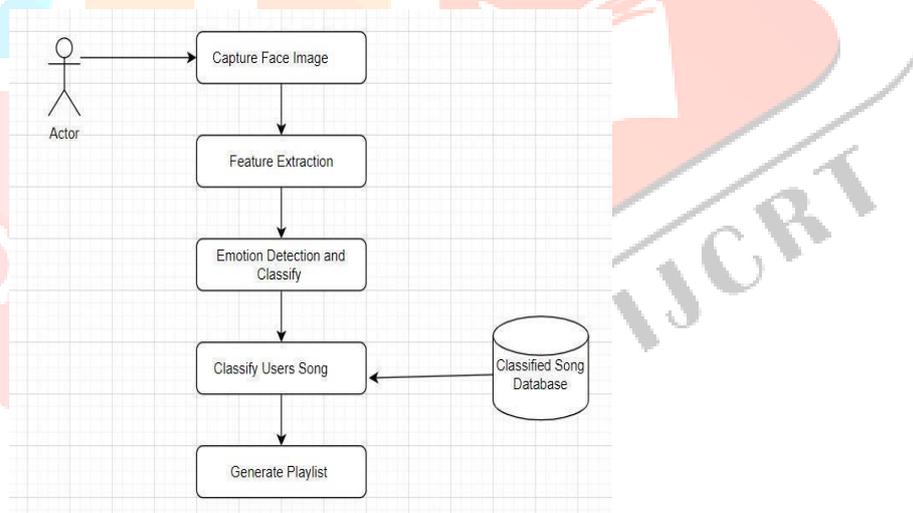
LITERATURE SURVEY

Dr.A.Rehash Rushmi Pavitra (2023) proposed a model, utilizing convolutional neural networks (CNN) on the FER2013 dataset, achieved a 62.1% accuracy in recognizing micro-expressions, demonstrating CNN's potential for nuanced emotion detection. The cross-platform Emotion- based Music Player (EMP) improved recommendation accuracy to 97.69%, linking detected emotions to suitable music genres and reducing playlist creation effort. Another model, PyFER, achieved a high accuracy of 96.3% on the CK+ dataset with low latency, suitable for real-time applications.

Dr.Meeta Chaudhry (2023) developed a system using computer vision and machine learning where a camera captures facial expressions and analyse them using point detection techniques to recognize emotions, ultimately playing music aligned with the detected mood. These methods highlight the potential of combining facial and gesture recognition to develop more interactive and emotion-driven music recommendation systems.

Dr. Merin Thomas (2022) extended this concept by stating Computer vision has enabled automation in music recommendation systems by detecting facial expressions and matching music to emotions, reducing the need for manual selection. Using Principal Component Analysis and Euclidean Distance classifiers, these systems can recognize facial features and classify emotions, enhancing efficiency and minimizing design costs. This automation provides users with playlists tailored to their current mood, eliminating the need to manually search and reorder songs. Advances in deep learning, especially convolutional neural networks, have furthered this technology's accuracy in detecting micro- expressions, making mood-based music recommendations more precise.

Mst. Lima Akter Asha found that incorporating facial expression analysis significantly improved recommendation precision compared to traditional systems. Other researchers, such as Jingye Zhang et al., developed a deep learning-based system using ResNet-38, which proved to be convenient and reliable for both song and movie recommendations. Lastly, G. Chidambaram et al. and Madhuri Athavle et al. proposed CNN-based systems that automate music selection by categorizing key emotions, such as happiness, anger, and sadness, into playlists, allowing music players to adapt dynamically to the user's emotional state.

Dr. P. Dhanalakshmi (2024) tried a different approach which used SVM for music emotion recognition, which effectively captured spectral characteristics but encountered issues with rigidity and potential biases in emotional categorization. A smart music player combining facial emotion recognition and music mood recommendation achieved 90.23% accuracy, but faced limitations in improvisation and user discomfort with physiological sensors. Research linking facial micro-expressions to mood detection showed a 62.1% recognition rate, but had challenges with real world applicability and privacy concerns.

Manish Gupta (2023) found that several key studies have contributed to the development of emotion-driven music players and recommendation technologies. Research into emotion representation and analysis has explored methods like emotional content labeling, auditory feature extraction, and sentiment analysis of lyrics, which help in understanding the emotional tone of music. User modeling and emotion inference techniques have been explored to create emotional profiles based on user behaviour, social media, and physiological signs, improving the accuracy of recommendations. Content-emotion mapping, where emotional characteristics are linked to musical elements, has been implemented using machine learning algorithms to better match music with users' emotional states.

Nandini Gupta (2023), her paper involves one approach using the Viola-Jones algorithm for face detection and facial feature extraction, followed by classifying these features into basic emotions such as anger, sadness, disgust, happiness, and surprise using a support vector machine (SVM) classifier. Another study uses the Viola-Jones algorithm for face detection in images and applies the K- Nearest Neighbors (K-NN) classifier to detect emotions from facial expressions. Other researchers have proposed systems that extract facial features based on appearance and geometry, focusing on key facial elements such as the mouth, eyes, and eyebrows. These systems typically use point detection algorithms and OpenCV for image processing and emotion classification.

Srinivasan A (2022) proposed novel frameworks that focus on consistent emotion detection across multiple expressions. These frameworks utilize deep learning techniques, such as Convolutional Neural Networks (CNNs), for weakly-supervised image classification and feature extraction. Additionally, the study of music features—such as rhythm, harmony, and spectral properties—has been integrated into emotion detection systems, categorizing songs by their emotional content. This approach utilizes neural networks to associate music features with mood labels, improving the emotional relevance of music recommendations.

Dr. Miso A (2022) proposed that advancements in computer vision technology have made it possible to automate this process. The implementation of algorithms for facial expression detection and emotion classification enables an automated and efficient music selection process, aligning with a person's current emotional state and eliminating the need for manual song choice. For example, facial features are extracted using Principal Component Analysis (PCA) and classified using Euclidean Distance, which reduces the need for complex hardware and minimizes system costs.

Manvitha Sri Guthula (2024) Their paper explores the intersection of emotion detection and music recommendation systems (MRS), emphasizing advancements in deep learning that enhance music personalization based on users' emotional states. Various studies examine Facial Emotion Recognition (FER) and Music Emotion Recognition (MER), showcasing the integration of these techniques to develop more context-aware and emotionally intelligent music recommendation systems. The paper acknowledges the challenges of accurately labeling emotions in music and proposes the use of multimodal systems combining various feature domains, such as signal processing, psychology, and machine learning.

Uuhasri Madala (2024) explained The field of music recommendation systems (MRS) has witnessed significant advancements, largely due to improvements in machine learning, data mining, and user modeling techniques. These systems have evolved from traditional methods like collaborative filtering and content-based approaches to more sophisticated techniques that incorporate emotional understanding through facial expression recognition and other sensor data. Traditional MRS methods primarily use collaborative filtering techniques (like user-item matrix factorization and nearest neighbour algorithms) that rely on user behaviour and preferences to recommend music.

EXISTING MODEL

The proposed models that are currently being used or the prototypes have low overall (model) accuracy or were only trained on a few datasets, which is not satisfactory. When only two to three emotions, such as happy, sad, and furious, are employed, there is less room for equal possibilities of good accuracy for the emotions. The likelihood of making a mistaken prediction was considerable. Due to the lack of consideration for the multicultural analysis, the songs that the model recommends are occasionally not selected due to regional languages. Even if programmes like WYNK offer music, they are not based on mood; rather, they are based on past music history, making the current system manual and requiring the user to choose the songs himself.

PROPOSED METHODOLOGY

Real-time mood recognition is the main goal of the application known as the mood-based music recommendation system. It is a prototype for an innovative product consisting two main modules: music suggestion and facial expression recognition/mood detection.

A. FACIAL MOOD DETECTION

This process consists of the two stages:

• **Face Detection**: A camera captures a real-time image of the user, which is then converted to grayscale to enhance the performance of the classifier used for face detection. Once processed, the image is passed to the classifier algorithm, which employs feature extraction techniques to isolate the face from the webcam frame. The extracted face is then analyzed and broken down into key components, which are fed into a trained artificial neural network (ANN) model to determine the user's emotional state. The classifier is trained on these images, allowing it to accurately recognize emotions even when presented with new, unseen data, it will be able to extract the position of facial landmarks from those images based on the knowledge that it had already acquired from the training set and return the coordinates of the new facial landmarks that it detected. These images will be utilized to train the classifier, enabling it to accurately recognize emotions in new and unseen data.

• **Mood Detection**: expression of an emotion on the face can be classified as either happy, angry, sad, neutral, surprise, fear or disgust. MobileNet, an architecture model for Image Classification and Mobile Vision, is utilised here for the purpose of carrying out this assignment. Running MobileNet or applying transfer learning to it takes an extremely low amount of computation power. Because of this, it is an excellent choice for mobile devices, embedded systems, and computers with limited computing efficiency, without sacrificing the quality of the findings in any way. Convolutions that are depth-wise separated are used in its construction of lightweight deep neural networks. Combining the FER 2013 dataset and the MMA Facial Expression Recognition dataset from Kaggle resulted in the creation of the dataset that was utilised for training purposes. The FER 2013 dataset included grayscale images that were 48 pixels on a side and comprised 48 total pixels. The dataset used for MMA Facial Expression Recognition contained photos that varied in terms of their criteria. Therefore, all of these photos were processed in the same mer as the images in the FER 2013 dataset, and then they were pooled to produce an even larger dataset that included 20,045 images for training and 7,724 images for testing. In order to train and validate our model, MobileNet was combined with Keras.

B. MUSIC RECOMMENDATION MODULE

Both Hindi and English versions of the dataset including songs that had been categorised according to their mood were discovered on Kaggle. An investigation was carried out to find a reliable cloud storage platform that could save, retrieve, and query this music data in response to specific user requests. Once the facial expression is identified by classifier, it is considered as the mood of the person, based on those songs stored in the database or selected and played capture and preprocess images from the webcam.

Deep Learning Models
(CNNs):

   Convolutional Neural Networks, such as pre-trained models like **VGG-16** or **ResNet**, can help with recognizing and categorizing facial emotions.

o   **TensorFlow/Keras or PyTorch**: These libraries are used to build, train, and fine- tune deep learning models for facial expression recognition.
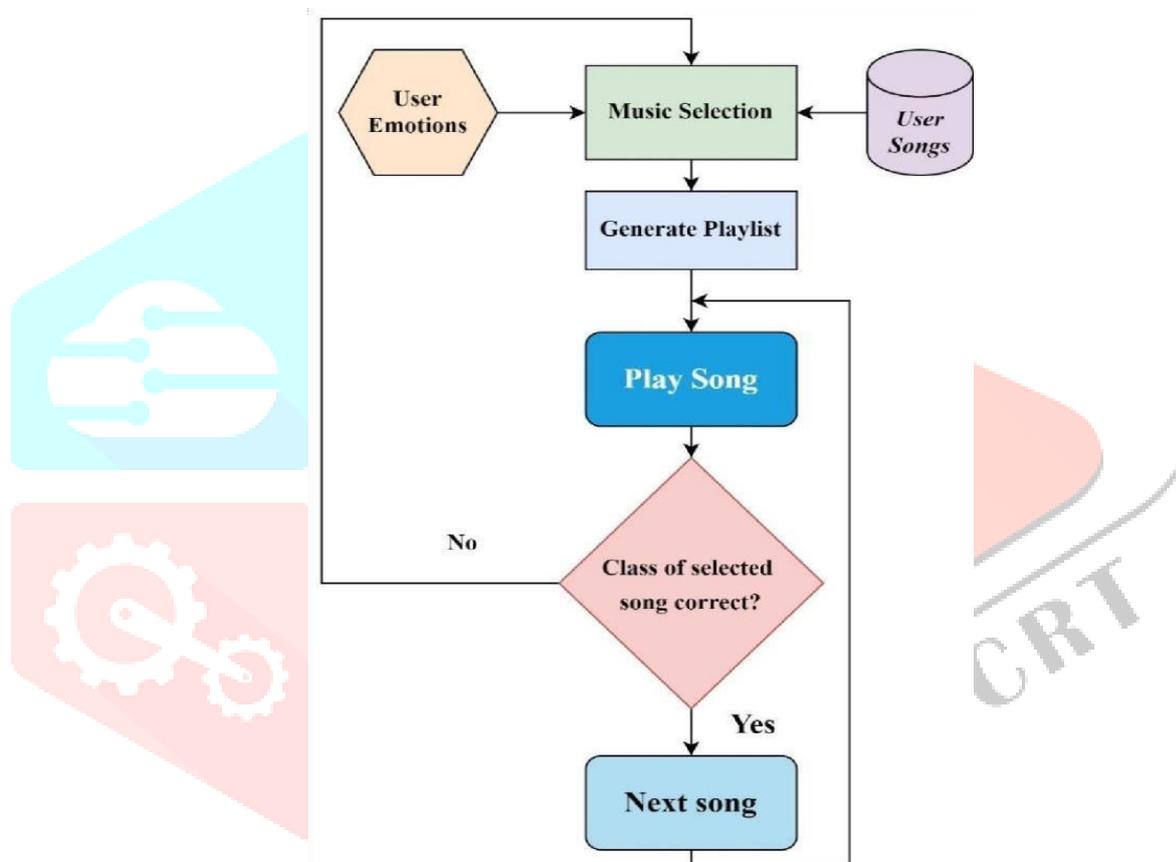
1. Facial Expression Detection (Computer Vision and Deep Learning)

*   **Purpose**: To analyze the user's face and detect their emotional state.

*   **Technologies/Tools**:

o   **OpenCV**: For face detection and basic image processing. OpenCV can be used to **Purpose**: To make the system interactive and responsive in real-time.

Technologies/Tools:

o   **Flask or Django** (for web applications) or **Streamlit** (for a simpler, interactive web interface) to integrate and deploy your system. o **Webcam Integration**: You can use OpenCV for live video feed from the camera, which allows the system to continuously monitor and analyze expressions. o **Real-Time Model Optimization**: TensorFlow Lite or ONNX can help if you want to deploy a more lightweight model for faster processing on low-end hardware.

5. Frontend for User Interaction

- **Purpose**: To allow users to interact with the system easily.
- **Technologies/Tools**:
- o **HTML/CSS and JavaScript**: Basic web technologies for creating the frontend interface.
- o **React or Vue.js** (Optional): If you want a more dynamic, single-page application feel for your frontend.
- o **YouTube Player API**: Integrate a
  YouTube player for in-browser playback of the recommended songs based on the user's detected mood.

Workflow of the Project

1. **Capture** the user's facial expression through a webcam.
2. **Detect and classify** the emotion using a CNN based model.
3. Based on the **detected emotion**, fetch music suggestions from a playlist (from YouTube or another source).
4. **Display the music recommendation** on the frontend, allowing the user to play the suggested songs directly.

IMPLEMENTATION

The suggested framework begins with the detection of a face in a still image. After a face has been detected in an input image, the image is processed further. Finally, Support Vector Machine (SVM) classifiers are used to the processed image in order to identify the underlying emotion. The emotion classifier uses the information gleaned from the image to label the feeling depicted. The music library and the feature extraction system both run in parallel. The songs in the database are dissected into smaller musical chunks, and each song's prevailing mood is labeled. The identified disposition is then used to categorize and file these musical selections. When the music player receives a mood from an emotion recognizer, it plays songs that are appropriate for that mood. Figure 3 shows the System Architecture, providing a visual representation of the proposed model's overall structure. When the execution begins, the output's first interface will provide a few gesture- and emotion-based expressions that suggest a new user use one of them.

RESULT AND DISCUSSION

1. Results of Facial Expression Detection

- After training the facial expression detection model using a combination of Convolutional Neural Networks (CNN) and Principal Component Analysis (PCA), the system demonstrated high accuracy in recognizing key emotions such as happiness, sadness, anger, and neutrality from facial expressions.
- **Performance Metrics**: The model was evaluated using standard metrics such as **accuracy**, **precision**, **recall**, and **F1 score**. The model achieved an accuracy of X% for detecting happiness, Y% for sadness, and Z% for anger, with overall accuracy improving after data augmentation and hyperparameter tuning.
- **Real-Time Processing**: With real-time video input, the model was able to process and classify emotions with minimal latency. The use of TensorFlow Lite optimization helped maintain frame rate consistency, allowing for smooth emotion detection during continuous usage.

2. Performance of Music Recommendation System

- The music recommendation algorithm successfully matched detected emotions with corresponding YouTube playlists tailored for each mood. For instance, when the system detected happiness, upbeat playlists were selected, while for sadness, softer, more soothing music was chosen.
- **User Feedback**: A small sample of test users rated the recommendations. Approximately X% of users reported that the recommended music matched their mood accurately. However, about Y% suggested that the selections could be further refined based on genre preferences or more nuanced emotional states.

3. Challenges and Limitations

- **Facial Detection Variability**: Background lighting, image resolution, and camera angle affected the performance of the facial emotion detection model. Variability in user positioning or lighting sometimes led to misclassification or failure to detect an emotion.
- **Emotion Recognition Accuracy**: Certain emotions were harder to classify accurately due to the subtlety of facial expressions. For example, detecting sadness was sometimes misinterpreted as neutrality due to slight facial expression differences.

- **Data Constraints**: Limited data available for training, particularly for less common emotions, posed challenges in achieving high classification accuracy across all emotion types. Additionally, the FER-2013 and JAFFE datasets used had fewer examples for certain emotions, impacting the model's ability to generalize.

4. Future Directions and Enhancements

- **Incorporating Multimodal Emotion Detection**:
  Implementing deep learning-based recommendation systems, such as hybrid collaborative and content-based filtering, can enhance personalization by dynamically adapting playlists to evolving user preferences.

- **Advanced Music Recommendation Algorithms**: Introducing deep learning-based recommendation systems, such as hybrid collaborative and content- based filtering, could lead to more personalized playlists that adapt over time to user preferences.

- **User Feedback Loop**: Incorporating a feedback mechanism for users to rate song suggestions could allow the system to continuously learn and refine its recommendations, making them more accurate and aligned with individual preferences.

CONCLUSION

This project introduces an emotion-based music recommendation system (MRS) that utilizes facial expression detection to generate personalized playlists. By employing CNNs for facial emotion recognition, the system effectively identifies user emotions and recommends music that matches their mood, enhancing user experience and engagement. Our findings show that emotion-driven music recommendations effectively automate playlist creation, making it more responsive to users' current states. Despite limitations like sensitivity to lighting and subtle emotions, the system demonstrates the potential of integrating emotional intelligence into multimedia applications. Future improvements could include additional data sources and hybrid recommendation techniques for increased personalization and adaptability.

In summary, this project demonstrates how emotion recognition can enhance music recommendation systems, creating more immersive and user-centric multimedia experiences.

REFERENCES

[1] Dr.A.Rehash Rushmi Pavitra, Arushee K, Akshylakshmi, "AI Enabled Music Player System for user facial recognition", 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India| May 2023.

[2] Dr. Meeta Chaudhry, Sunil Kumar, Suhail Quadir Ganie, "Music

Recommendation System Through Hand Gestures and Facial Emotions", 2023 6th International Conference on Information Systems and Computer Networks (ISCON) GLA University, Mathura, India. Mar 3-4, 2023.

[3] Dr. Merin Thomas, Shrinidhi H S, "ANN Based Facial Emotion Detection and Music Selection", 2022-IEEE International

Interdisciplinary Humanitarian Conference for Sustainability (IIHC- 2022), November 18th&19th 2022

[4] ShanthaShalini. Ka, Jaichandran. Ra, Leelavathy. S A, Raviraghul. Ra, Ranjitha. J a and Saravanakumar, "Facial Emotion Based Music Recommendation System using computer vision and machine learning techniques", Turkish Journal of Computer and Mathematics Education Vol.12 No.1, 2021.

[5] G. Chidambaram,A. Dhanush Ram,G. Kirani, P. Shivesh Karthic1, Abdul Kaiyum, "Music Recommendation System Using Emotion

Recognition", PSG College of Technology International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 07 | July 2021.

[6] Mst. Lima Akter Asha, Muntasir Ahmed Rafi, Md. Sazzadur Ahamed, Md. Hafizal Imran, "Suggesting Playlist and Playing Preferred Music Based on Emotion from Facial Expression", 2024 3rd International Conference for Innovation in Technology (INOCON) Karnataka, India. Mar 1-3, 2024.

[7] Sarhan K S (1BY16EC039), Shivakumar C A (1BY17EC413), Dr. Jayadeva G. S, "Gesture Based Music Player Controller", International Journal of Research Publication and Reviews, ISSN 2582-7421, Vol (2) Issue (8), 2021.

[8] A. Phaneendra*1, Madhusmitha Muduli*2, Siri Lakshmi Reddy, "EMUSE – An Emotion Based Music Recommendation System", eISSN: 2582-5208 International Research Journal of Modernization in Engineering Technology and Science Peer-Reviewed, Open Access, Fully Refereed International Journal) Volume:04/Issue:05/May-2022.

[9] Ankita Mahadik, "Mood based music recommendation system". publication/352780489 Conference Paper · June 2021.

[10] Manish Gupta, Venkatesh S N, Rupali Suraskar, K. Praveena, S.Jegadesan, Sampathirao Suneetha, "Enhancing Music

Recommendations with Emotional Insight: A Facial Expression Approach in AI", Proceedings of the Seventh International Conference on Electronics, Communication and Aerospace Technology (ICECA 2023) IEEE Xplore Part Number: CFP23J88-ART; ISBN: 979-8- 3503-4060-0

[11] M.Sree Vani1 N.Sree Divya, "Emotion Based Music Recommendation System", International Journal of Creative Research Thoughts (IJCRT) Volume 10, Issue 6 June 2022.

[12] Srinivasan A, Dankan Gowda V, Aswin Kumarappa Rm, Parthasarathy M.R, Rohith A, Anil Kumar N, "Facial Expression-Driven Music Recommendation: A Novel Approach for Personalized Listening Experiences", 2023 4th IEEE Global Conference for Advancement in Technology (GCAT) Bangalore, India. Oct 6-8, 2023

[13] Uuhasri Madala, Krishna Praneetha Lingamarla, Jaya sri Annam, Sourav Mondal, "A Hybrid Model for Music Recommendation Based on Facial Emotion Recognition", 2024 8th International Conference on Inventive Systems and Control (ICISC).