Emotion based Music Recommendation using FaceRecognition

Anjali Salvi
Student, Dept. of MCA
Sardar Patel Institute of Technology,
Mumbai, India
anjali.salvi22@spit.ac.in

Prajakta Shevkar
Masters of Computer Application
Sardar Patel Institute of Technology
Mumbai, India
prajakta.shevkar22@spit.ac.in

Prof. Harshil Kanakia
Assistant Professor, Dept. of MCA
Sardar Patel Institute of Technology,
Mumbai, India
harshil_kanakia@spit.ac.in

***Abstract*—In an age propelled by cutting-edge technology and data analytics, emotion-based music recommendation systems have risen to paramount significance. This project introduces an innovative approach to elevate the music recommendation process by incorporating advanced technologies. By combining the functionalities of Mediapipe, Keras, OpenCV, and Streamlit, our system utilizes facial recognition, machine learning, and real-time analysis to identify and interpret the user's emotional state. By analyzing facial expressions and emotional cues through a Streamlit web application with integrated Streamlit-webrtc webcam capture, the system tailors music suggestions to harmonize with the user's mood. This amalgamation of technologies creates a personalized and emotionally resonant music selection, ultimately enriching the user's auditory experience. This research contributes to the burgeoning domain of emotion-based recommendation systems, paving the way for more perceptive and emotionally-aware technology applications.**

I. INTRODUCTION

In an era driven by technology and data analysis, the concept of music recommendation systems based on emotions has gained significant importance. This project proposes an innovative approach to enhance the music recommendation process by integrating facial recognition technology. Utilizing facial expressions and emotional cues, the system gauges the user's prevailing emotional state, offering personalized music recommendations that resonate with their mood. [1]

By harnessing the remarkable capabilities of Mediapipe, Keras, OpenCV, and Streamlit, our system transcends conventional music recommendation approaches. It leverages facial expressions and emotional cues captured through a webcam, which allows us to decode the user's prevailing emotional state in real-time. [2]

The magic behind this innovation lies in the fusion of machine learning techniques and computer vision. What sets the project apart is its remarkable integration of leading-edge technologies: Mediapipe, Keras, OpenCV, and Streamlit.

These robust components form the foundation upon which our system is built. Mediapipe and Keras infuse machine learning capabilities, while OpenCV and Streamlit provide the invaluable tools of computer vision and web application creation, respectively.The outcome is a personalized and emotionally resonant selection of music, elevating the user's auditory journey. [3]

Why Computer Vision and Emotion-Based Systems?

Computer vision stands at the intersection of artificial intelligence and computer science, dedicated to empowering machines to comprehend and interpret the visual realm. This field encompasses the creation of algorithms and systems designed to analyze and derive meaning from visual data, ranging from images to videos. Computer vision strives to emulate the intricate processes of human visual perception, enabling systems to undertake tasks such as image recognition, object detection, facial recognition, motion tracking, and beyond. [4]

Emotion detection through facial expressions is a technology for automatically identifying human emotions, mirroring the cognitive process of the human brain. [5]

Emotion-based systems, in the context of technology and artificial intelligence, refer to systems and applications that can detect, interpret, and respond to human emotions. These systems use various data sources, including facial expressions, voice tone, physiological signals, and text analysis, to identify the emotional state of individuals.

THE TECHNOLOGIES: Mediapipe, Keras, Opencv, Streamlit

Mediapipe: Mediapipe, an open-source framework created by Google, offers tools for the development of machine learning and computer vision applications.

Keras: Keras is an open-source deep learning framework that acts as an interface for the TensorFlow library. Keras is known for its user-friendliness and is often used for tasks like image recognition, natural language processing, and more.

OpenCV (Open Source Computer Vision Library): OpenCV is a widely used computer vision library that provides a variety of tools and functions for image and video processing.

Streamlit: Streamlit, a Python library, simplifies the creation of web applications with minimal effort. It has gained popularity, particularly for rapidly constructing data-driven and interactive web apps. Utilizing Streamlit allows you to develop web-based user interfaces for your projects without requiring extensive knowledge of web development. [6]

II. LITERATURE REVIEW

In recent years, the intersection of technology, music, and emotions has given rise to a rapidly growing field of research focused on emotion-based music recommendation systems...

The foundation of emotion-based music recommendation systems lies in the ability to accurately recognize and interpret human emotions. In the work of Lartillot and Toiviainen (2007), techniques for emotion recognition in music were explored, setting the stage for integrating emotional cues in music recommendation. Subsequent studies, such as Saari and Lartillot (2012), emphasized the importance of aligning music with user emotions to enhance the listening experience.

Facial recognition technology has rapidly evolved. Ekman's groundbreaking research on facial expressions (Ekman and Friesen, 1971) provided the foundational framework for understanding emotions through facial cues. More recent works, such as the deep learning-based models by Mollahosseini et al. (2019), showcase the robustness of facial recognition technology in detecting emotions.

The integration of machine learning techniques in music recommendation has seen a surge of interest. Research by Law and von Ahn (2016) introduced personalized music recommendation through collaborative filtering, opening avenues for enhancing music discovery. Further, the utilization of machine learning models for music preference analysis, as demonstrated by Li and Ogihara (2006), laid the groundwork for algorithms capable of aligning music with emotional states.

The ability to recognize real-time emotional states is an evolving area of research. Works like Patel et al. (2016) explored real-time emotion detection from video streams, which aligns with the goals of emotion-based music recommendation systems that seek to adapt to users' ever-changing moods.

Personalization in music recommendation has been a central theme in recent research. The study by McFee et al. (2012) demonstrated how machine learning can be employed to create personalized music recommendation systems, providing a foundation for the integration of emotions as a key factor in user profiling. [7]

III. METHODOLOGY

To build Emotion based Music Recommender, we adhered to a systematic and structured process involving a series of crucial steps with an innovative algorithm designed to discern human emotional states via facial expressions.

The approach encompasses three key stages:

1. User Preferences Input: Gathering user input for preferred song language and favorite artist.

2. Feature Extraction and Emotion Classification: Gathering user input for preferred song language and favorite artist. Analyzing and extracting features from user preferences and classifying emotions based on these inputs.

3. Song Recommendation: Recommending songs tailored to the user's preferences and emotional state.

Figure 1 illustrates the general flow of our System.

Our system explores an idea that facilitates automated communication between users and a music player. It involves training the music player with user preferences, emotions, and activities. The result is the display of a customized music playlist. The system analyzes users' facial expressions to gauge their mood and predict the genre of the track based on favorite artist input. [8]

The authors introduce a mood-based system that identifies a user's emotional state and subsequently generates a playlist tailored to that mood. For instance, if the user is feeling low, the system compiles a playlist with songs to uplift their spirits, while a positive mood prompts a playlist with a mix of music to enhance the user's good feelings.

The system focuses on automatically generating music that aligns with the emotions detected by a webcam. The system rapidly evaluates a person's emotional state with a time frame of approximately 0.95 to 1.05 seconds. Periodically, the algorithm reevaluates the user's emotional state and, based on comparisons with previous assessments, makes decisions such as changing the currently playing song or suggesting a break. This process involves capturing images from a webcam using the OpenCV library and analyzing them with the OpenFace library.

Also the Convolutional Neural Network (CNN) algorithm assists in comprehending the underlying rules applied in classification or regression problems. It employs a technique known as the kernel trick to transform your data based on these differences, ultimately determining the optimal boundary between potential outcomes. [9] We applied a similar approach utilizing the CNN algorithm to recognize and categorize emotions based on the user's detected facial expressions. The CNN algorithm necessitates image input to predict facial expression in the form of distinct labels of six moods, such as happy, neutral, anger, sad, and rock. [10]

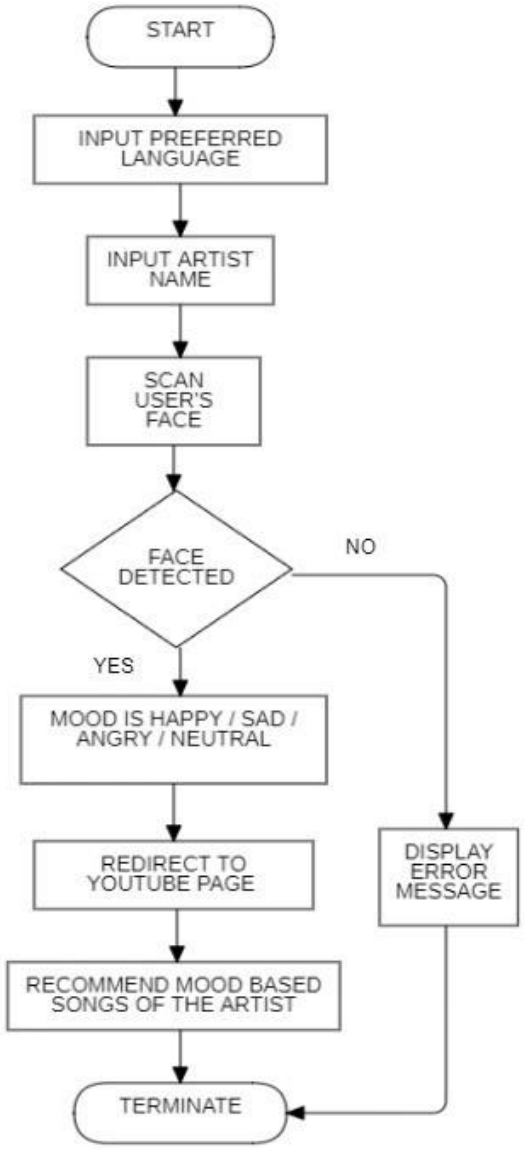

Fig. 1 General Flow of System

IV. RESULTS AND FINDINGS

We have categorized the results in accordance with the primary objectives of our proposed system, which are as follows:

1. Identifying unobstructed facial features.
2. Accurately extracting emotions and categorizing them into the six predefined emotional states.
3. Taking the detected emotion and user input for language and artist into account, we recommend suitable songs to enhance the customer's mood.

Results Discussion:

Figure 2 illustrates the Mood Accuracy Graph, displaying the results of mood analysis conducted on a sample of 100 individuals. The test group consisted of 50 males and 50 females, spanning ages from 15 to 80 years, in order to evaluate the effectiveness of our proposed system across various age groups. The emotion analysis results were derived using a Convolutional Neural Network (CNN), which generates feature maps to capture image regions. [11]

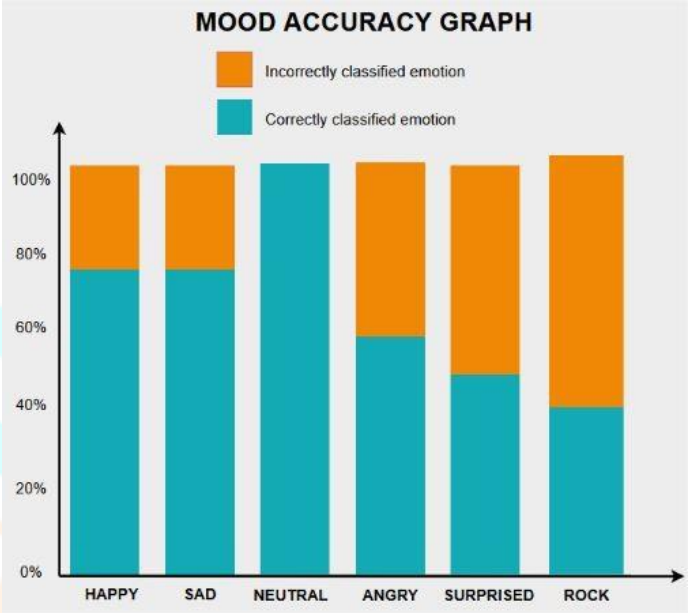

Fig. 2 Mood Accuracy Chart

V. CONCLUSION

The domain of emotion identification through facial expressions has gained significant attention. Emotion recognition using image processing algorithms has become increasingly challenging. Researchers continually explore diverse features and image processing techniques to address this challenge.

This paper introduces a system that leverages two predicates, namely "language" and "singer," to comprehend user preferences. Once these predicates are input, the webcam initiates image capture of the user's face, and the captured emotion is subsequently analyzed by a machine learning model. The inputs, including "Language," "Singer," and "Emotion," are then incorporated into the URL query. The user is then directed to a relevant YouTube page based on their requirements. [12]

This research introduces an approach to identify core universal emotions through frontal facial expressions. Following the deployment of a machine learning model for facial recognition, we take the additional step of transforming it into a web application using Streamlit. Emotion detection is achieved through deep learning, specifically utilizing the well-established Deep Learning model. The library, keras is employed, and the Convolution Neural Network (CNN) algorithm plays a pivotal role. [13]

CNN is an artificial neural network capable of various tasks, including object detection, facial recognition, and image processing, among others. CNNs use a series of convolutional

layers to extract features from input images. Convolution is a mathematical operation that involves applying a small filter (also called a kernel) to the input image, which scans the image in a grid-like fashion. This process helps identify patterns and features at different levels of abstraction. [14]

VI. FUTURE SCOPE

While this system is fully functional, there exists considerable room for enhancement in the future. Various aspects of the application can be refined to yield superior results and offer a more seamless user experience. Potential improvements include the exploration of alternative methods for recognizing additional emotions like disgust and fear. The incorporation of these emotions could support the automatic playback of music.

Looking ahead, there's an opportunity to introduce a mechanism within the system that could be beneficial in music therapy treatment. This would assist music therapists in treating patients dealing with mental stress, anxiety, acute depression, and trauma. [13] It's worth noting that the current system's performance is compromised under extremely poor lighting conditions and with low camera resolution. This presents an avenue for adding functionality to address these challenges in the future.

Furthermore, a feedback system can be integrated to enhance the accuracy of emotion detection. Looking ahead, this model's scope extends to catering to the user's preferences in diverse domains such as movie recommendations, educational resources, shopping applications, and more, all guided by the system's capability to detect and interpret emotions. [15]

REFERENCES

[1] T. Gorasiya, A. Gore, D. Ingale and M. Trivedi, "Music Recommendation based on Facial Expression using Deep Learning," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1159-1165, doi: 10.1109/ICCES54183.2022.9835929.

[2] K. Patel and R. K. Gupta, "Song Playlist Generator System Based on Facial Expression and Song Mood," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, 2021, pp. 1-6, doi: 10.1109/AIMV53313.2021.9670976.

[3] M. Lahoti, S. Gajam, A. Kasat and N. Raul, "Music Recommendation System Based on Facial Mood Detection," 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), Kannur, India, 2022, pp. 284-289, doi: 10.1109/ICICICT54557.2022.9917956.

[4] J. Qiu, J. Liu and Y. Shen, "Computer Vision Technology Based on Deep Learning," 2021 IEEE 2nd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), Chongqing, China, 2021, pp. 1126-1130, doi: 10.1109/ICIBA52610.2021.9687873.

[5] M. Kumar and S. Srivastava, "Emotion Detection through Facial Expression using DeepLearning," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702451.

[6] Garima, S. Thapliyal, B. Bhatia, R. Tyagi, V. Jha and R. Kumar Singh, "Vibing: The Mood Based Music Recommendation System," 2022 4th International Conference on Artificial Intelligence and Speech Technology (AIST), Delhi, India, 2022, pp. 1-7, doi: 10.1109/AIST55798.2022.10065377.

[7] Z. Yu, M. Zhao, Y. Wu, P. Liu and H. Chen, "Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition," 2020 39th Chinese Control Conference (CCC), Shenyang, China, 2020, pp. 7257-7263, doi: 10.23919/CCC50068.2020.9189600.

[8] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[9] X. He and F. Ding, "An Efficient Face Recognition Method Based on CNN," 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 1788-1791, doi: 10.1109/ICPECA56706.2023.10076242.

[10] O. Ghosh, R. Sonkusare, S. Kulkarni and S. Laddha, "Music Recommendation System based on Emotion Detection using Image Processing and Deep Networks," 2022 2nd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2022, pp. 1-5, doi: 10.1109/CONIT55038.2022.9847888.

[11] A. Alrihaili, A. Alsaedi, K. Albalawi and L. Syed, "Music Recommender System for Users Based on Emotion Detection through Facial Features," 2019 12th International Conference on Developments in eSystems Engineering (DeSE), Kazan, Russia, 2019, pp. 1014-1019, doi: 10.1109/DeSE.2019.00188.

[12] R. Chauhan, K. K. Ghanshala and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 278-282, doi: 10.1109/ICSCCC.2018.8703316.

[13] K. Vayadande, P. Narkhede, S. Nikam, N. Punde, S. Hukare and R. Thakur, "Facial Emotion Based Song Recommendation System," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 240-248, doi: 10.1109/CISES58720.2023.10183606.

[14] V. Kumar, S. Kumar and V. P, "Music Recommendation based on User Mood," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022, pp. 575-578, doi: 10.23919/INDIACom54597.2022.9763120.

[15] A. Srivastava, D. K. Srivastava and M. Shandilya, "Facial Emotion-based Music Recommender System using CNN," 2023 International Conference on Artificial Intelligence and Smart Communication (AISC), Greater Noida, India, 2023, pp. 113-118, doi: 10.1109/AISC56616.2023.10085294.