2023 6th International Conference on Information Systems and Computer Networks (ISCON)
GLA University, Mathura, India. Mar 3-4, 2023

Music Recommendation System through Hand Gestures and Facial Emotions

Dr. Meeta Chaudhry
Associate Professor
Department of CSE, ABESEC,
Ghaziabad
meeta.chaudhry@abes.ac.in

Sunil Kumar
Department of CSE, ABESEC,
Ghaziabad
ssunil2408@gmail.com

Suhail Qadir Ganie
Department of CSE
ABESEC,
Ghaziabad,
suhailqadir2019@gmail.com

2023 6th International Conference on Information Systems and Computer Networks (ISCON) | 979-8-3503-4696-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISCON57294.2023.10112159

***Abstract*— Music can be a powerful tool to describe the human mood. Hand Gestures and Facial emotions are forms of fast non-linguistic communication. The current research on Music recommendation either using a hand gesture music controller (that only controls the operations for playing music) or an emotion based music player but not both. In this work, a new and hybrid approach for playing music both using hand gestures and facial emotions is proposed that can help the user to recommend and play music. In this research facial expression recognizer(FER) algorithm is used that extract the features from the image for emotion detection and the MediaPipe framework and Tensorflow library are used for hand detection and gesture recognition respectively. The music will play based on the most recent gesture and emotion by using a pygame. First, priority is given to hand gestures and then to facial emotions. The accuracy of the proposed work is also compared with existing approaches to music recommendation.**

Keywords*— *Convolutional Neural Network, Facial Expression Recognizer, TensorFlow, Mediapipe, Hand Gesture Recognition, Pygame, Feature extraction, Emotion detection, Tkinter, Music Player, Webcam

I. INTRODUCTION

Numerous studies done over the last few years have shown that music has an impact on people's moods, actions, and thought processes. One of the most crucial purposes of music, according to researchers studying the reasons people listen to music, is its relationship with pleasure and mood. A wide range of applications, including virtual reality, sign language interpretation, video games, image database analysis, police departments, video indexing, civilian applications, privacy, and human cognitive integrations manifest in many ways, currently regard gesture and emotion recognition as the most useful techniques. A wide variety of feasible applications, such as musical entertainment and human-computer interaction systems, are significantly impacted by this technology.

It is not enough to describe the human mood, therefore, gestures also play an important role to describe the mood of humans. So this work has extended the music player so that it can recommend and play the music based on emotions as well as hand gestures such as 'rock', 'peace', 'stop', 'smile', 'okay', 'call me', 'fist', 'devotee hands', etc. and emotions such as 'angry', 'sad', 'happy', 'surprise', 'neutral', 'fear', and 'disgust'. Each hand gesture represents a different expression of the user's mood like a rock hand gesture represents that user wants to listen to something like rock dance music, peace hand gesture for peaceful (like meditation/yoga) music, live long gesture for devotee/divine music, etc.

Human-computer interaction technology is actively researching gesture recognition. It may be used for a variety of things, including music production, robot control, sign language translation, and virtual environment control. With this machine learning project on hand gesture recognition, we will use MediaPipe and Tensorflow to create a real-time hand gesture recognizer.

A person's facial expressions are frequently a nonverbal means of expressing emotion, and they can be used as direct evidence to determine whether they are telling the truth because they are correlated with emotions, facial expressions serve as essential markers for human emotions.

The objective of this research is to design a music recommendation system based on gestures and emotions, which applies machine learning algorithms system that will help the users to manage music playlists with less effort based on the user's gestures and emotions captured by a web camera in real-time basis

II. LITERATURE SURVEY

The research is carried out to obtain information about the methodologies, their limitations, and proposed solutions. A research paper's content called a "literature survey" contains the most up-to-date information on a particular subject, as well as important discoveries, conceptual and methodological improvements, and research findings. The implicit skills of humans who can provide inputs to any system in a variety of ways have attracted the attention of numerous students, researchers, engineers, and other people from all around the world. Facial expressions might provide information about a person's current state of mind. Nonverbal cues including hand gestures, facial expressions, and voice tones are widely used when talking with people to express our feelings.

Shantha Shalini. Ka, Jaichandran. Ra, Leelavathy. S A, Raviraghul [3] published a paper titled "Facial Emotion Based Music Recommendation System using computer vision and machine learning technique" In the proposed approach, the facial expression is connected to music suggestion using computer vision and machine learning algorithms. We coded with PyCharm to get experimental results. Real human faces are entered using a camera in this case, and the input collected image is then processed using image-processing techniques. An approach called point detection is used to extract the characteristics from the input photos. The input photos are trained for facial expression recognition using the classification system OpenCV. Music would automatically play from the coded folder in accordance with the emotions identified.

Devikarani Patil, Varalakshmi B.D [7] published a paper titled “Hand Gesture Recognition for MP3 Player using Image Processing Technique and PIC16F8779 ”.In the proposed approach Gesture control is used to operate the MP3 (music) player[1][8]. The input image is taken using a camera and processed in a mat lab after that. The math lab compares the input and dataset images, and the outputs are subsequently transferred via RS232 to the PIC16F8779 controller.

The MP3 player is linked up with the PIC16F8779 controller and relay drivers; the relays act as an electric switching button. The relay will turn on instantly and the MP3 player will begin playing when the PIC16F8779 controller recognizes the first image input. Only a few audio signals were played by the MP3 player, and the LCD displayed its information.

T.S. Gnaneswar, N. Goutham Siddarth, M.Hariprasad, S.M.Mehzabeen[8] published a paper titled “hand gesture controlled digital music player using enhanced image processing techniques” The main goal of the project is to use hand gestures and improved image processing to control a digital music player, bridging the impact of technology in today's hectic environment. The music player will carry out the designated action when a gesture is made in front of the camera. Every action of the music player is controlled by a specific gesture, thus if no gesture is provided or if the gesture is displayed incorrectly, the system will respond with an error message stating that the gesture input is invalid. Thus, it can be inferred that the music player may be effectively used with great ease simply by making the proper gestures.

Sarhan K S, Shivakumar[9] published a paper titled “Gesture Based Music Player Controller ” In the proposed approach, a few hand gestures may be used to control the music playing on your computer[10]. You can play, pause, stop, or next to your music by using a few basic hand motions. The development of the Microsoft Kinect for PCs and the Xbox 360 has been a major factor in the recent rise in gesture-based technology[16]. Future technologies include computer-integrated programs, which make it easier to do things like change the music while reading something important.

New interfaces and interaction strategies are frequently displayed via controlling music playback with the buttons play, stop, pause, and next. Gesture recognition algorithms have been tested and shown to work by controlling music playback with a set of functions.

G. Deepa, G. Bhaskar Phani Ram[11] published a paper titled “Emotion-based Music Retrieval for Gesture Recognition” The suggested strategy This study presents a method for real-time recognition of static hand gestures. in HCI. This real-time method is composed of two phases: the hand gesture recognition phase and the hand gesture tracking phase. OpenCV, the Haar classifier, and the Linux operating system are all used in this system.

In this specific instance, the live input is a real-time image taken by a camera. Therefore, starting at this moment, each conceivable gesture or movement made in front of the camera will be evaluated for recognition. The left, right, top, and bottom positions of the hand gestures in the image affect how they are recognized in the proposed system. Anger, relaxation, happiness, and sadness were assigned musical emotions for certain hand positions.

Phaneendra, Madhusmitha Muduli, and Siri Lakshmi Reddy[14] published a paper titled “emuse – an emotion-based music recommendation system” In the proposed method, computer vision and machine learning methods are combined with deep learning techniques to identify facial expressions of emotion and make music recommendations based on such expressions. To learn the best feature abstraction, Deep Neural Networks are used as the method. Deep Neural Networks have proven to be effective in many different tasks, including visual object detection, facial verification, estimating human position, and many others. In fields like image identification and classification, Convolutional Neural Networks have proven to be quite successful. A convolutional neural network[13]model is used in the proposed system to detect the user's facial expressions. The user would be shown the song that best matched their emotion once the emotion had been assessed. In this project, the streamlit framework is used to construct the main web page, where a user's image is taken. The model is then given the taken image to forecast the user's sentiment[4][5]. When an emotion is identified, the python script Spotify calls the Spotify API to request music tracks, which are subsequently shown in the user interface.

M. Sree Vani” and “N.Sree Divya[17] published a paper titled “EMOTION BASED MUSIC RECOMMENDATION SYSTEM” in 2022 in which the proposed automatic playlist generating system combines several existing schemes. Various forms of emotions are taken into account from the user's expressions in this work, and it explores how this information could be used to enhance the user experience with music players. The suggested solution is built on the notion of automating a lot of user interaction with the music player[2][6]. It features a "smart" music player that picks songs based on how its user is feeling after learning about them. The smart music player can use its internal algorithms to make an educated choice of the song that would best suit its user's feelings after an initial period of training.

III. PROPOSED METHODOLOGY

To distinguish gestures, the Hand Gesture Recognition dataset from Kaggle is used, which consists of two folders, one for training and the other for testing purposes. Each folder has 20 directories containing various gestures with names ranging from 0 to 19, and there are 24000 total photos in this collection representing distinct movements. There are 900 photos in each directory for training purposes and just 300 images in each directory for testing purposes. This dataset is mostly used for tasks involving hand gesture recognition.

FER2013, a dataset divided into training and testing halves, is used for emotion identification. 6043 photos make up the testing dataset, whereas 24176 are in the training set. In the dataset, there are faces represented as grayscale 48x48 pixel pictures. Each image in FER-2013 is assigned one of seven emotions: surprise, fear, disgust, anger, happiness, sadness, and anger. The faces are automatically registered such that they are essentially in the middle of each image and occupy roughly the same amount of space. FER-2013's pictures, which are grayscale and 48x48 pixels in size, include both posed and candid headshots. Google image search results for each emotion and their synonyms were

compiled to form the FER-2013 dataset. When trained on an unbalanced dataset, FER systems may perform well on dominant emotions including happiness, sadness, anger, surprise, neutrality, disgust, and fear. The proposed model will perform the following flowchart as shown in figure 1.

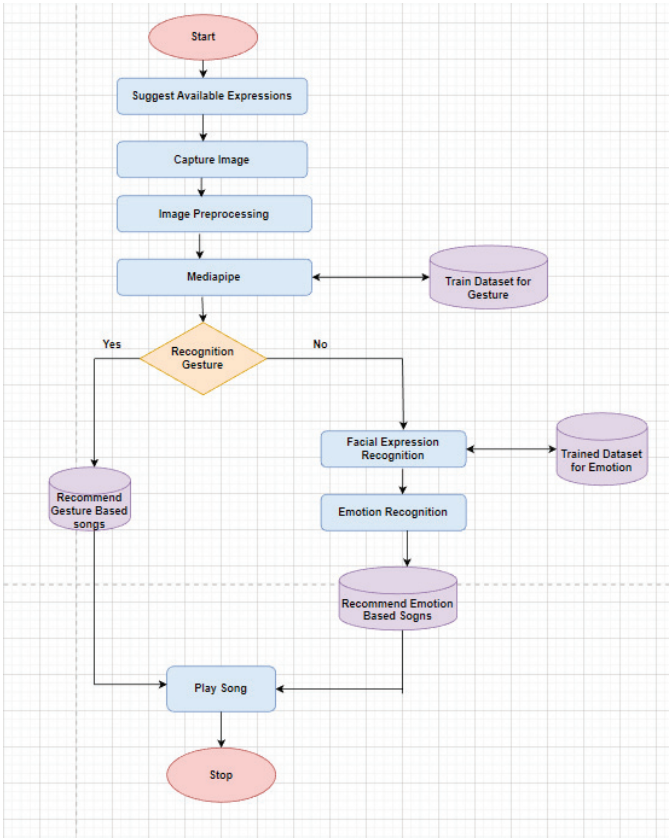

Fig. 1. Flowchart of the model.

A. Emotion based expressions

To get familiar with the proposed system the user can show the available expressions as shown in figure 2. The system will work only on these seven emotion based expressions.

Fig. 2. Emotions used for detection

B. Gesture based expressions

The various expressions used by the proposed system are illustrated in figure 3. The system will only perform with these ten gesture-based expressions and each gesture indicates its own meaning.

1) ***Okay***- represents that everything is going according to plan and play all time good songs.

2) ***Thumbs Up***- In response to points we have in common. We frequently play the music that gives us more confidence and correlates the motion with saying "Yes" or "agree."

3) ***Thumbs down***- A gesture in which the user holds his hand out with his thumb pointed down to say no, to show *a feeling of displeasure,* disapproval, disagreement, etc., and play the song accordingly**.**

4) ***Stop-*** If the user wants someone to stop, pause, avoid something, or just be still, flash them your palm and play songs accordingly.

5) ***Peace***- a symbol for peace that is created by holding the palm outward and creating a V with the index and middle fingers representing the feeling of peace and resting.

6) ***Rock***- A gesture where the *middle two fingers are held down by the thumb is commonly used in rock music.*

7) **Call me-** The traditional meaning of this gesture is to request a phone call because the "call me" gesture matches the "hang ten" sign, it can also be used to express that something is nice or has a good vibe.

8) ***Live long***- A gesture that represents the feeling of devotee/divine music.

9) ***Fist***- A gesture with the finger closed tightly towards the palm to hit someone, to make an angry gesture.

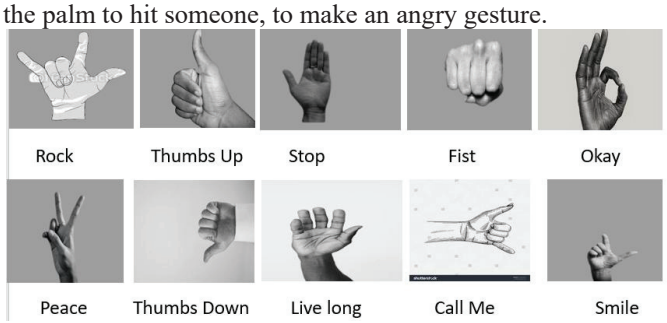

Fig. 3. Gestures used for detection.

C. Gesture detection

Hand gestures are identified using the Mediapipe framework. MediaPipe includes several machine learning pre-trained solutions, including face detection, pose estimation, hand recognition, object detection, etc. Mediapipe is used for creating processing pipelines to begin exploring from several sources, including audio and video. This method uses machine learning to identify 21 3D key - points of a hand from a single picture, increasing overall hand and finger tracking.

Each of the hand gestures mentioned above in figure 3 has a distinct meaning, allowing the user to represent a variety of moods in addition to emotions. Gestures can be predicted using the following flowchart as shown in figure 4.

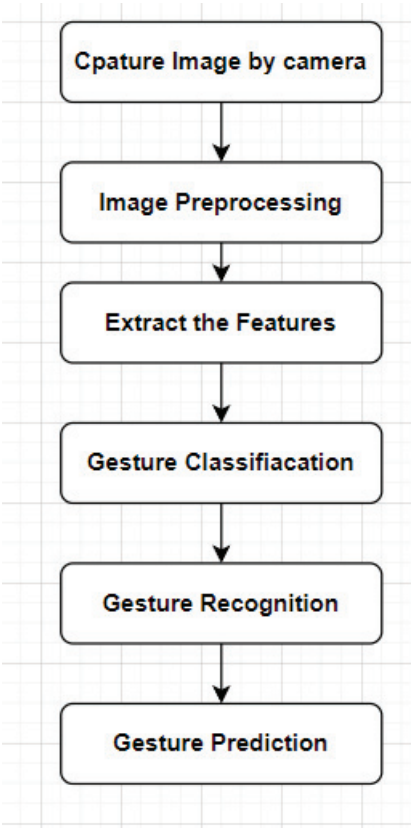

Fig. 4. flowchart for gesture detection

1) Initialize mediapipe framework

This module executes the hand recognizer algorithm defined by mediapipe. In comparison to MediaPipe, which can identify multiple hands in a single frame, the recommended system can only recognize one hand at a time. Mp.solutions.drawing utils will perform it for the user instead of forcing them to manually draw the discovered key areas. Making a VideoCapture() object and passing the value 0 will allow the user to read frames. It is the camera's system-assigned ID.

2) Detect hand landmarks

To detect hand landmarks each frame from the webcam is read using the cap. read(). We add the coordinate to a landmarks list after iterating over each detection. every landmark in the frame is drawn as shown in figure 5.

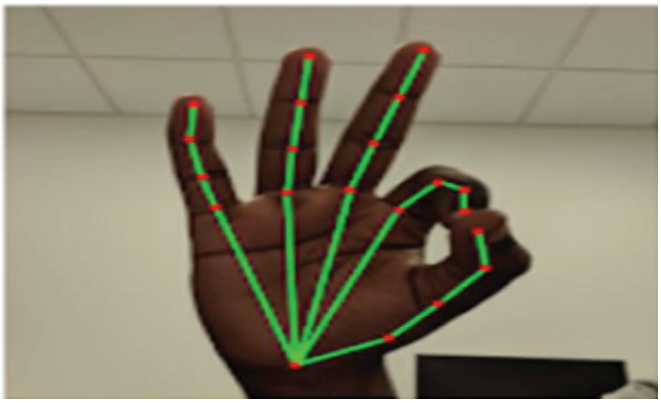

Fig. 5. Detect Hand Landmarks

3) Recognize gestures

The index of the list's highest value is returned by the function. Once we have the index, we can simply select the class name from the gesture list. Once the gesture has been identified, we display it in the frame using the cv2.putText() function as shown in figure 6.

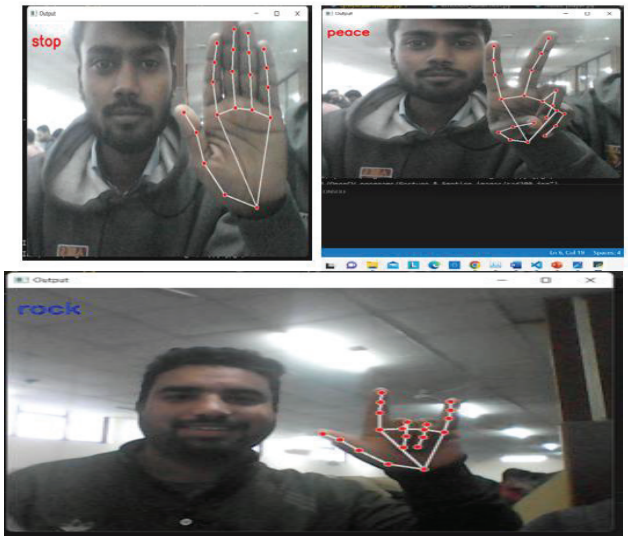

Fig. 6. Detect Gesture Names

D. Emotion Detection

When execution commences following gesture detection, it starts to access the final image that was collected during gesture detection for additional processing and emotion detection. Emotion detection is the next stage of the project. The feature of the image attributes is obtained using the facial expression recognition method. To identify emotions users can recognize emotions in photos by using the open-source Facial Expression Recognition (FER) library that follows the steps shown in figure 7. As a result, each of the seven fundamental emotions was given a value in a range of values between 0 and 1. The emotion that matches the expression of the user will be shown once the emotion has been assessed.

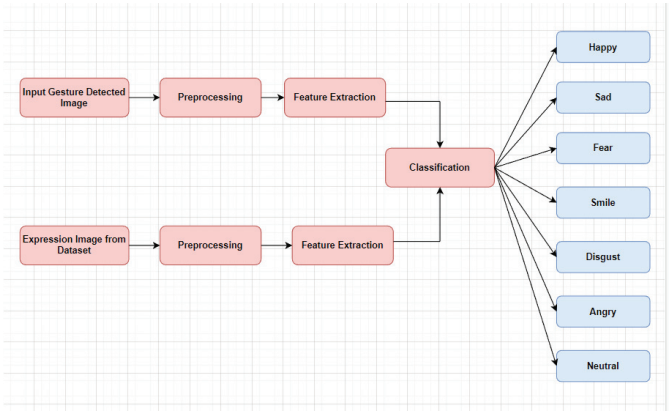

Fig. 7. Emotion classification

1) Music player

The Tkinter module is used for frontend GUI design and pygame for playing music. The most well-known and user-friendly library is Tkinter, which includes a variety of components that make it easy to build products in good GUI applications. The Tk GUI toolkit that comes with Python has a normal Python interface. The easiest and simplest method for developing GUI projects is using Python with Tkinter. Using Tkinter to build a GUI is simple. The widgets'

geometric configuration, which can arrange the tools in the primary windows, is also accessible through the Tkinter. There are primarily three types of geometry management classes. Some of the popular methods that are available in tkinter are pack(), grid(), place() and tools button, canvas, CheckButton, Entry, Frame, Label, ListBox, Menu, Message, RadioButton, ScrollBar, Text, etc.

The pygame. mixer module is used to play (), pause (), stop (), and many methods to perform operations on music and audio files. There are classes in a module that can be used to load

Sound objects and manage playing. To do this, one must follow four stages.

IV. RESULTS AND DISCUSSION

This music recommendation system through gesture and emotion can recognize all real-time gestures and emotions. The current system has been trained and tested on the Hand Gesture Recognition dataset that has 20 different folders having over 900 images on each gesture and for emotion detection, the model is prepared on the FER2013 dataset having 7 different folders named happy, surprise, sad, neutral, angry, disgust and fear and each folder consist over five hundred images. For playing music we have created our dataset that contains 17 unique folders having about 100-150 mp3 songs files based on 10 gestures and 7 emotions. The proposed model works in three modules gesture detection, emotion detection, and music playing.

TABLE I. ACCURACY TABLE FOR GESTURES

Gesture	Number of Images	Number of right read images	Accuracy (%)
Okay	30	24	80
Peace	30	26	86.66
Thumbs Up	30	29	96.66
Thumbs Down	30	28	93.33
Call me	30	27	90
Stop	30	28	93.33
Rock	30	29	96.66
Live Long	30	26	86.66
Fist	30	28	93.33
Smile	30	26	86.66

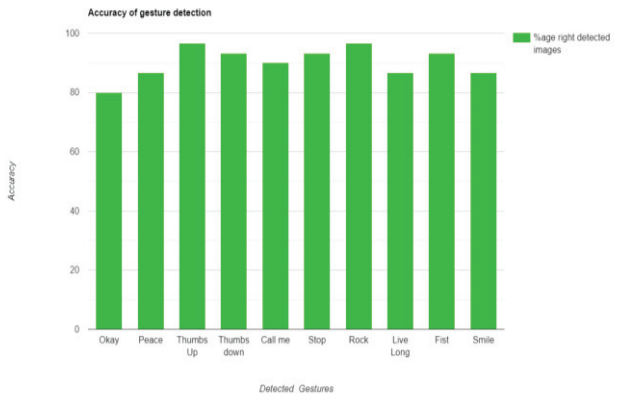

Fig. 8. Accuracy chart for gesture detection.

TABLE II. ACCURACY TABLE FOR EMOTIONS

Emotion	Number of Images	Number of right read images	Accuracy (%)
Happy	30	28	93.33
Sad	30	27	90
Anger	30	25	83.33
Fear	30	26	86.66
Disgust	30	24	76.66
Surprise	30	27	90
Neutral	30	28	93.33

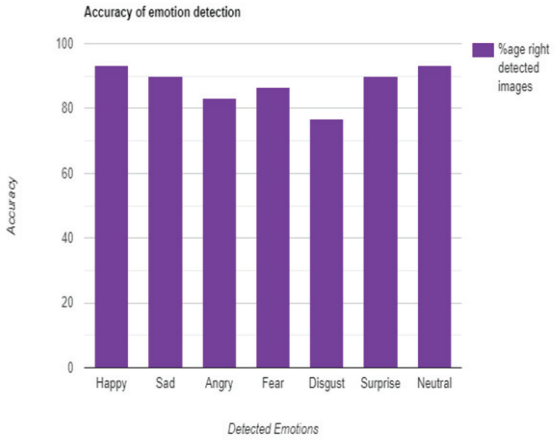

Fig. 9. Accuracy chart for emotions detection

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, as seen in figure 10.

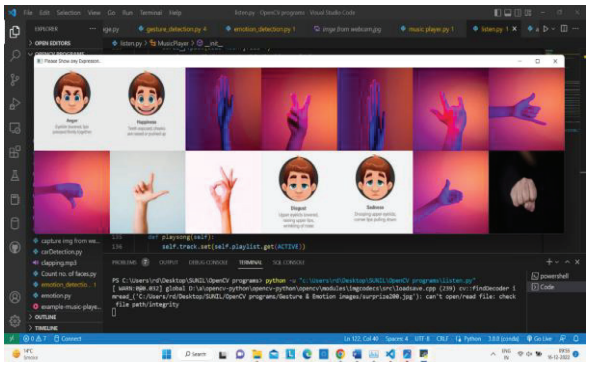

Fig. 10. GUI result for available expressions

The system will suggest a music playlist depending on the user's gesture when it has identified a gesture, as seen in figure 11.

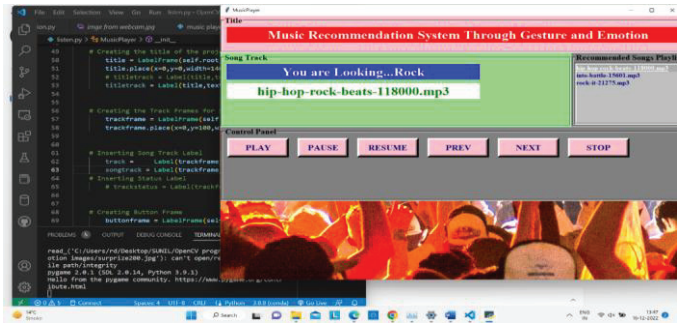

Fig. 11. Detecting rock gestures and playing songs accordingly.

The model will recommend a music playlist based on the detected emotion if a gesture has not been detected, as illustrated in Figures 12 and 13 since gesture is given priority over emotion in that manner.

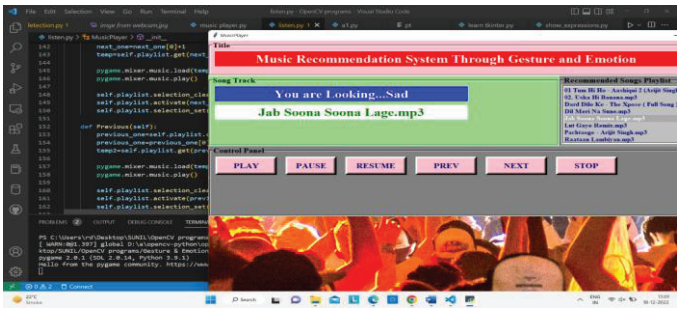

Fig. 12. Detecting sad emotions and playing songs accordingly

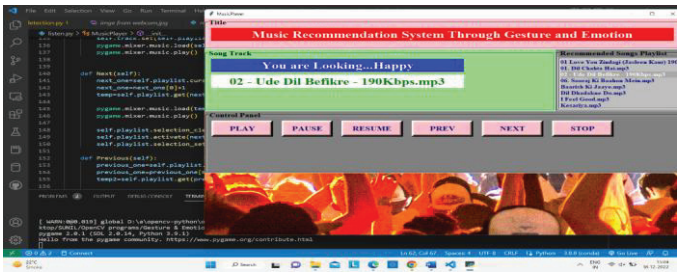

Fig. 13. Detecting happy emotions and playing songs accordingly

The proposed hybrid model for music recommendation has been compared with the existing work as discussed in the literature review. Table 3 shows the comparative analysis chart. From the table, it can be seen that the proposed hybrid model is achieving an accuracy of 89%, which is better than many of the state of art techniques and allows the user to use both emotions and hand gestures.

TABLE III. COMPARISON WITH EXISTING WORK

S.No	TITLE OF RESEARCH ARTICLE	APPLIED MODEL/TECHNOLOGIES	ACCURACY
1.	Emotion Integrated Music Recommendation System[7]	Generative Adversarial Networks	61.35%
2.	Emotional Detection and Music Recommendation System based on User Facial Expression[4]	Deep Neural Networks, VGG16	69.14%
3.	Emotion-Based Music Player Emotion Detection from Live Camera[5]	LeNET CNN	85.0%
4.	Facial emotion recognition in real-time and static images[6]	SVM	92.1%
5.	Hand Gesture Controlled digital Music player Using Enhanced Image Processing Techniques[8]	Image Processing,	89.6%
6.	Emotion based Music Retrieval for Gesture Recognition[11]	ARM9 controller, RS232, OpenCV library, Phonon multimedia, Qt Application	73.6%
7.	Music Controller using Gesture Recognition to Control Music Playback[10]	HSV model, Binary Picture	79.6%
8.	Proposed hybrid "Music Recommendation System Through Hand Gestures and Facial Emotions"	FER, MediaPipe, Tensorflow, Tkinter, PyGame	89%

V. CONCLUSION

The goal is to develop a hybrid gesture and emotion-based music recommendation system that will let users construct and manage playlists with less effort. This research work provides a method for recommending music that is based on the actions and feelings of the user. The input is a human gesture, and the music is played automatically based on the gesture expressions that are recognized. On the Hand Gesture Recognition dataset, the work presented here successfully extracted gestures using the gesture recognizer algorithm provided by mediapipe, and the human face is used as an input and based on the detected facial expressions, music is automatically played. With impressive results, the Facial Expression Recognition algorithm is used for feature extraction on the FER2013 dataset Therefore; the proposed technique offers a high level of accuracy of 89% on live images. In comparison with existing work, the proposed hybrid model is found to be better in performance and accuracy.

REFERENCES

[1] Deepali N. Kakade Prof. Dr. J.S. Chitode, "Dynamic Hand Gesture Recognition: A Literature Review", International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 9, November- 2012.

[2] S Metilda Florence1 and M Uma, "Emotional Detection and Music Recommendation System based on User Facial Expression", 3rd International Conference on Advances in Mechanical Engineering (ICAME 2020) IOP Conf. Series: Materials Science and Engineering 912 (2020) 062007 IOP Publishing doi:10.1088/1757-899X/912/6/062007.

[3] 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.

[4] 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.

[5] Ahmed Hamdy, "Emotion-Based Music Player", The British University in Egypt Thesis ·June 2019.

[6] Shivam Gupta, "Facial emotion recognition in real-time and static images". India Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018).

[7] Devikarani Patil, Varalakshmi B.D "Hand Gesture Recognition for MP3 Player using Image Processing Technique and PIC16F8779". International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-4 Issue-4, September 2014.

[8] T.S.Gnaneswar, 2N.Goutham Siddarth, 3M.Hariprasad, 4 S.M.Mehzabeen . "Hand Gesture Controlled Digital Music Player Using Enhanced Image Processing Technique" SSN: 2455- 2631 © September 2016.

[9] 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.

[10] Aditya R. Parab1 Janhavi R. Ghosalkar2 V. Akshay B. Medge3 Prof.mamta Meena. "Music Controller using Gesture Recognition to Control Music Playback". Vikrant Institute of Technology & Management Conference Paper · January 2016.

[11] G.Deepa M.Tech, Student, "Emotion based Music Retrieval for Gesture Recognition" International Journal of Computer Applications (0975 – 8887) Volume 78 – No.14, September 2013.

[12] Rafiqul Zaman Khan and Noor Adnan Ibraheem,"Hand Gesture Recognition: A Literature Review" International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012.

[13] Authors: Madhuri Athavle1, Deepali Mudale, Upasana Shrivastav, and Megha Gupta. "Music Recommendation Based on Face Emotion

Recognition" Received: 7 April 2021 Accepted: May 27, 2021 Published: 09 June 2002.

[14] A. Phaneendra*1, Madhusmitha Muduli*2, Siri Lakshmi Reddy, "EMUSE – An Emotion Based Music Recommendation System", e-ISSN: 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.

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

[16] Ms. Kamal Preet Kour, Dr. (Mrs) Lini Mathew, "Literature Survey On Hand Gesture Techniques For Sign Language Recognition", International Journal of Technical Research & Science., ISSN No.: 2454- 2024, Volume 2 Issue VII, August 2017.

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