International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 9
Article Received: *25 July 2020* ***Revised:*** *12 September 2020* ***Accepted:*** *30 September 2020*

Facial Expression-Based Emotion Detection and Music Player using Convolutional Neural Network

Rahul Mannade[1], Ganesh Dongre[2], Sonu Ballal[3], Renuka Shinde[4], Pramod Dighole[5]

[1,2,3,4,5] Department of Information Technology, Government College of Engineering, Station Road, Aurangabad, Maharashtra 431005, India

Corresponding Author: mannade.rahul@gmail.com

Abstract

Music is a divine gift to humanity. It exists in many forms, and every individual has unique musical preferences. The choice of music often depends on a person's mood. For example, when a person is happy, they tend to listen to cheerful and pleasant music, whereas during moments of sadness they may prefer melancholic songs. However, selecting music based on mood is usually done manually. To address this issue, we propose a system that automatically recommends music according to the user's emotional state. The system captures an image of the user through a webcam and analyzes it to detect the user's mood, such as Angry, Sad, Happy, Neutral, or Surprised. Based on the detected emotion, the system provides a predefined playlist of songs that correspond to that mood. The proposed system was tested and analyzed on a desktop or laptop running the Windows operating system. The experimental results show that the system correctly detects emotions with an accuracy of 68.75%. This paper is organized as follows: Section 1 presents the introduction, Section 2 reviews related work, Section 3 describes the design of the proposed system, Section 4 discusses the test results and performance analysis, and Sections 5 and 6 present the conclusion and future scope, respectively.

Keywords- Music; CNN; Haar, Face Detection, Emotion Detection, OpenCV, Keras, TensorFlow.

1. Introduction

Music is a soothing form of expression that has the ability to improve our mood. A large portion of the population—particularly youngsters, middle-aged individuals, and music enthusiasts—listen to music to gain mental comfort, relieve anxiety, and reduce everyday stress.

People usually choose songs according to their current emotional state. As emotions can be clearly interpreted through facial expressions, this work considers the important role of emotions in music selection. Therefore, we developed a Facial Expression–Based Emotion Detection and Music Player System using CNN, which identifies a person's emotional state and generates a playlist of songs that match the detected emotion [1]. This system helps users avoid spending time browsing through large playlists and assists them in selecting appropriate songs based on their feelings. As a result, it enhances the overall music-listening experience in a more user-friendly manner.

The goal of this work is to design a system that captures a person's image and detects their emotion to provide customized, emotion-based music playlists. The proposed music player also considers the user's preferred song language to create a more personalized and user-friendly environment. The main objective of this approach is to move from manual song selection to an automated method, making the process of listening to music simpler and more enjoyable.

2. Literature Review

Various methodologies have been used by researchers to implement an emotion-based music player. The following papers address this problem statement and provide implementations along with a clear understanding of this approach.

Table 2. Research papers reviewed

Title of Paper	Authors	Published on	Methodology Used	Drawbacks
EMO-MUSIC (Emotion based	Sarvesh Pal, Ankit Mishra, Hridaypratap Mourya and	26th January 2020	In this approach Haar cascade classifier is used to extract features of face image in real	• It can be used only on windows operating system.

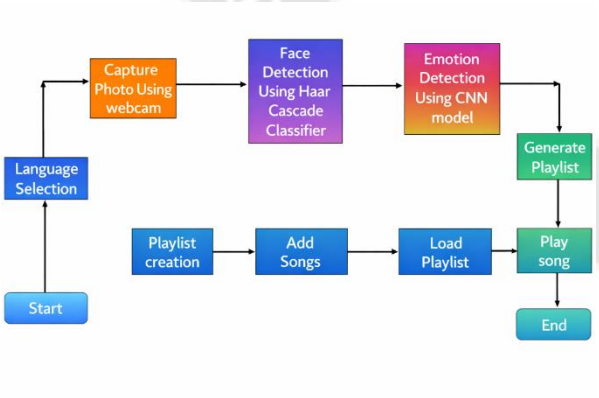
Title of Paper	Authors	Published on	Methodology Used	Drawbacks
Music player) [2]	SupriyaDicholkar		time. After feature extraction Cohn Kanade dataset is used to classify emotion and detect emotional state of user using SVM classification algorithm of machine learning. After emotion detection song is played from predefined directories.	• This model is not able to detect emotions in poor lighting conditions and is not able to process blur images.
Emotion Based Music Player [3]	Hafeez Kabani , Sharik Khan , Omar Khan , Shabana Tadvi	January – February 2015	In this approach face detection is performed using viola and jones algorithm and then SVM is used to classify emotions. Audio feature extraction is done to classify songs based on emotions and play according to mood of the user.	• This model needs high esolution pictures up to 4000X3000 typical image resolution in pixels. • It doesn't perform well in extreme bad light conditions.
Emotional Detection and Music Recommendation System based on User Facial Expression [4]	S Metilda Florence and M Uma	2020	In this approach Haar and HOG algorithm used for face detection. CK extensive dataset is used to train model. Facial features are extracted using facial landmarks and then emotion classification is done using the Fisherface algorithm.	• The quality of image is required higher than 320pixels in order to detect exact emotion. • System is not able to detect all emotions correctly due to less image availability in dataset.
Emotion Based Music Playing Device [5]	Dhruvisha Bansal, Pinkal Bhatt, Megha Dusane, Avneet Saluja, Kushal Patel	6th June 2020	In this approach a hardware device raspberry pi is developed which works as an emotion based music playing device. Python libraries like OpenCV is used for face detection. Keras is used to develop and estimate deep learning model. Tensorflow is used to define and train neural network model.	• Hardware equipment are required which increases the implementation cost.
A Novel Method To Design Emotion Based Music Player [6]	Mr T.M. Hayath, Ms.Sulthana, Mr.Kampli Shabbir	2020	In this approach a web application is designed using java script. CNN model is trained to detect emotions of user and Fisherface algorithm used for face detection purpose.	• Image size (128X128) is predefined for processing. • Only four emotions (Sad, Happy, Angry, Neutral) are

Title of Paper	Authors	Published on	Methodology Used	Drawbacks
				considered.
Emotion Based Music Player – Xbeats [7]	Aditya Gupte, Arjun Naganarayanan, Manish Krishnan	9th September 2016	In this approach they have developed an android application called XBeats. They have used anroid programming for application development and for face detection and emotion detection OpenCV and haar classifier is used.	• Needs working mobile device with android SDK and working front camera. • It can be used only by android users.
Face Player: Facial Emotion Based Music Player [8]	Pushkar Mahajan et al.	March 2020	Proposed a system that detects facial expressions and plays music accordingly. The system focuses on real-time emotion recognition and personalized playlists	• Faces challenges in varying environmental conditions and detection accuracy

3. Methodology

This section proposes a The Facial Expression (Emotion)-Based Music Player (FEBMP). It is a system designed and developed for desktop devices. The system aims to enhance the music listening experience by automatically playing songs based on the user's detected emotion. This section presents an overview of the proposed system and describes the methodology used for its implementation.

Figure 1. FEBMP Model Block Diagram (schematic)



In the initial step, the user selects the preferred language of songs such as Hindi, English, or Marathi. After selecting the language, the user captures a snapshot using the camera provided in the system interface. The user interface allows the user to take the snapshot and provides an option to retake the image if necessary.

Once the user is satisfied with the captured image, it can be submitted to the system for further processing.

The system then analyzes the user's facial expression from the captured image to detect the user's emotion. Based on the recognized emotion, the system automatically selects and plays a predefined playlist corresponding to that particular emotional state.

The proposed system comprises the following modules:

- Image capture with the camera
- Face Detection
- Emotion Detection
- Music Player
- Personalized Playlist

3.1 Image capture with Camera

With this module user can capture images using the system's webcam. It includes all the necessary functions required for image capture, such as switching the camera, taking snapshots, and sending the captured image to the model. This module is implemented using OpenCV.

3.2 Face Detection

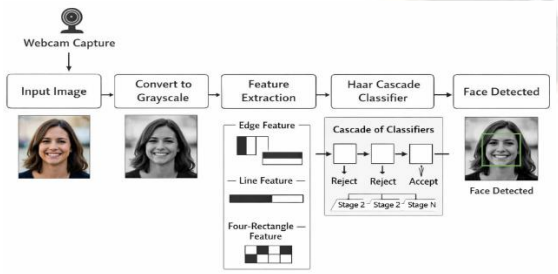
In this module, face detection is performed by processing images captured through a webcam. The system utilizes the Haar Cascade Classifier for feature extraction, an efficient method proposed by Paul Viola and Michael Jones [2], [3]. This classifier primarily

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 9
Article Received: *25 July 2020* ***Revised:*** *12 September 2020* ***Accepted:*** *30 September 2020*

focuses on detecting facial regions while ignoring other irrelevant parts of the image.

During the preprocessing stage, the captured image is first converted into a grayscale format. Converting the image to grayscale reduces the amount of data that needs to be processed, thereby simplifying the computation and improving the efficiency of the face detection process.

Figure 2. Face detection with Haar Cascade Classifier

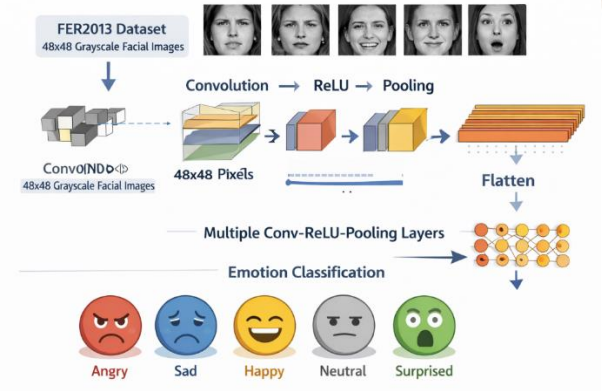


3.3 Emotion Detection

This module is one of the most crucial components of the Emotion-Based Music Player system. In this module, a Convolutional Neural Network (CNN), which is a deep learning algorithm, is used to train the model for facial emotion recognition [1], [6].

To train the CNN model, the FER2013 dataset is utilized. This dataset contains 48×48 pixel grayscale images representing different human facial expressions. The primary objective of the model is to learn the patterns of facial expressions and classify them into five emotion categories: Angry, Sad, Happy, Neutral, and Surprised. Based on this classification, the system detects the user's current emotional state. A Sequential model architecture is employed to construct the CNN, where layers are added one after another in a structured manner.

Figure 3. Emotion detection with CNN



3.3.1 Convolution layer:

It is primarily used as a feature detection and image filtering layer for two-dimensional inputs. This layer plays an important role in summarizing and extracting meaningful features from the input image. In the proposed approach, features are extracted from images contained in the FER-2013 dataset, which includes multiple emotion classes such as Happy, Sad, Angry, Surprised, and Neutral [4], [5].

For each emotion category, several facial attributes—such as the nose, lips, and eyes—are analyzed to capture distinctive patterns that help in identifying the expressed emotion. Based on these facial characteristics, different parameters are evaluated and the following feature points are calculated [4].

- Eyebrows raised
- Upper eyelid to eyebrow distance
- Inter-eyebrow distance
- Upper eyelid
- Mouth width
- Mouth open

3.3.2 Pooling layer

It is used to reduce dimensions of feature map by reducing parameters. It summarizes the feature map instead of features generated by convolution layer. This makes the model more robust to variations

3.3.3 Activation Layer

It is used to introduce non-linearity to a system that has computed linear operations in Convolution layer

3.3.4 Dropout Layer: In this layer neurons are randomly dropped while training.

3.3.5 Flatten Layer: This layer flattens all the result of previous layer

3.3.6 Dense layer

Neurons contained in dense layer receive input from all of its previous layers and then they are combined to generate a specific output.

After training the model we saved it into .h5 file.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 9
Article Received: *25 July 2020* ***Revised:*** *12 September 2020* ***Accepted:*** *30 September 2020*

Algorithm: Emotion Detection using Convolutional Neural Network (CNN)

Input*: FER-2013 dataset containing 48×48 grayscale facial images*

Output*: Predicted emotion class (Angry, Sad, Happy, Neutral, Surprised)*

1: Load the FER-2013 facial image dataset

2: Preprocess images

a) Resize images to 48×48 pixels

b) Normalize pixel values

3: Split dataset into training and testing sets

4: Construct CNN architecture

a) Apply Convolution layer

b) Apply ReLU activation

c) Apply Max-Pooling

d) Repeat for multiple Conv–ReLU–Pooling layers

5: Flatten the extracted feature maps

6: Add fully connected (dense) layers

7: Apply activation function for emotion classification

8: Train the CNN model using the training dataset

9: Input a new facial image for testing

10: Preprocess the input image

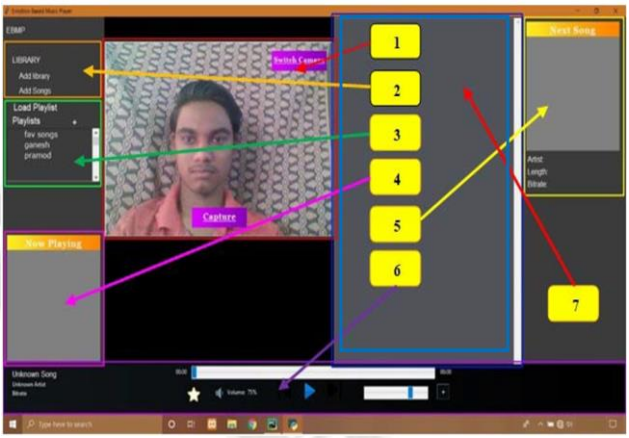
11: Predict emotion using the trained CNN model

12: Output the predicted emotion label

3.4 Music Player

This module primarily focuses on providing a user interface to perform the various functions of the music player system. In this work, the Tkinter Python library is used to design the graphical user interface (GUI), which offers different types of controls for user interaction. The Pygame library is utilized to manage audio-related operations such as playing a song, pausing playback, playing the next track, and returning to the previous track. Additionally, the Mutagen Python library is employed to extract metadata from audio files, including details such as the artist's name, song duration, and album artwork. The following figure illustrates the graphical user interface of the proposed system.

Figure 4. GUI of EBMP



1. Image Capture Section:

The section highlighted with red block is the camera frame provided to capture image through webcam. Here two button controls are given that is capture button to click image and switch camera button to switch the camera in case more than one camera are available in the user's system.

2. Add Library Section

The section highlighted with orange block represents the section from where user can add library of their own and as well as add songs of their choice in their customized playlist library.

3. Playlist Selection Section

The section highlighted with green block represents the section from where user can choose their created playlist/library from given list. The playlist list shows all the already created library/playlist by the user. So user can choose any playlist and click load playlist to load all the songs contained in it.

4. Now Playing Section

The section highlighted with pink block represents the song which is being played currently and it shows information about song like name of the song, artist of the song and the bit rate of the song etc.

5. Next Song Section

The section highlighted with yellow block represents the next song which will be played after current playing

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 9
Article Received: *25 July 2020* ***Revised:*** *12 September 2020* ***Accepted:*** *30 September 2020*

song. It also shows information like artist of the song, length of the song and its bit rate.

6. Player Control Section

The section highlighted with purple block represents the various music player controls like progress bar of song, volume controller, play and pause button, next song control, previous song control etc.

7. Current Playlist Section

The section highlighted with blue block represents the list area where songs playlist will be shown when the emotion of user is detected and when user will load their own created playlist.

3.5 Personalized Playlist

This component of the system enables the creation of personalized playlists. A *.csv* file is generated using the playlist name specified by the user, and songs can be added to the playlist according to the user's preferences. Whenever the user wishes to listen to music from their personalized playlist, they can load the stored songs and play them as desired.

4. Results and Discussion

The system was tested extensively to evaluate its performance and reliability. The testing environment consisted of a computer with the Windows 10 operating system, 8 GB RAM, an Intel Core i3 processor, and an integrated 2 MP webcam. The experimental results indicate that the proposed system functions effectively and produces accurate outcomes. The obtained accuracy is comparable to existing emotion-based systems reported in literature [3], [4]. Snapshots illustrating the detection of various emotions along with the corresponding playlist selections generated by the system are presented below.

Figure 5. '*Sad*' Emotion Detection

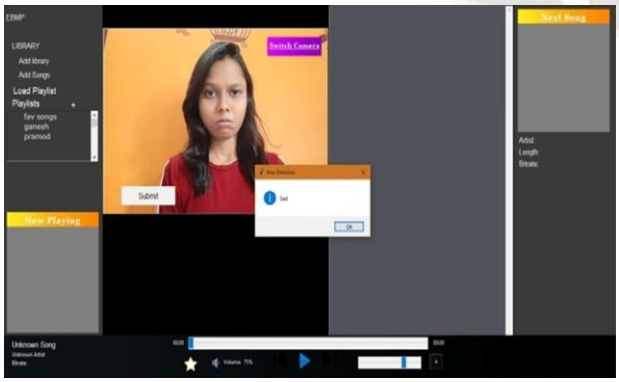


Figure 6. '*Neutral*' Emotion Detection

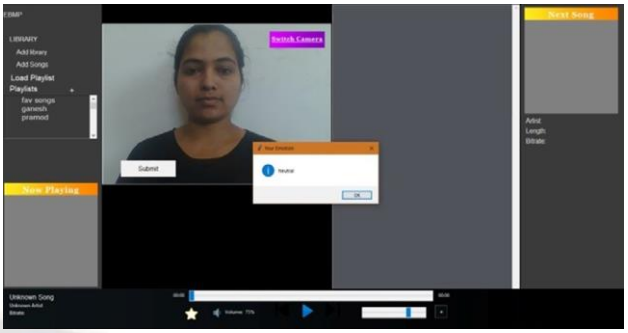


Figure 7. '*Angry*' Emotion Detection

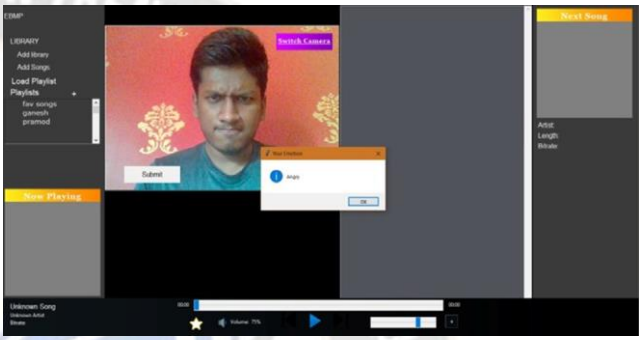


Figure 8. '*Surprized*' Emotion Detection

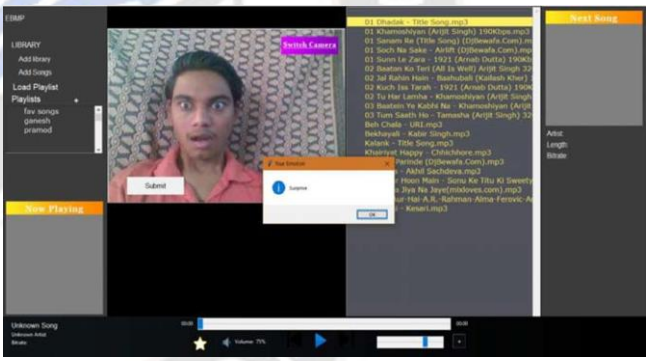
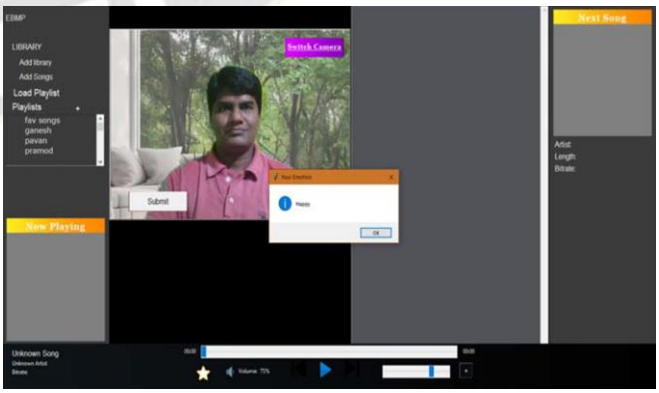
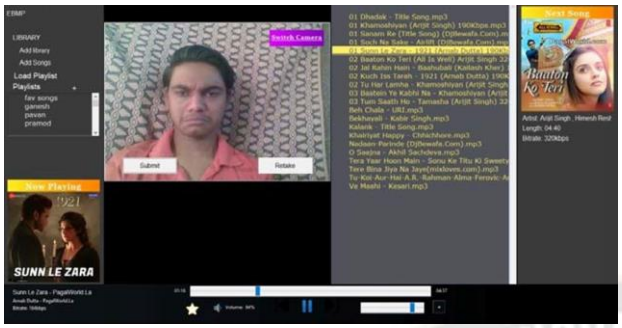


Figure 9. '*Happy*' Emotion Detection



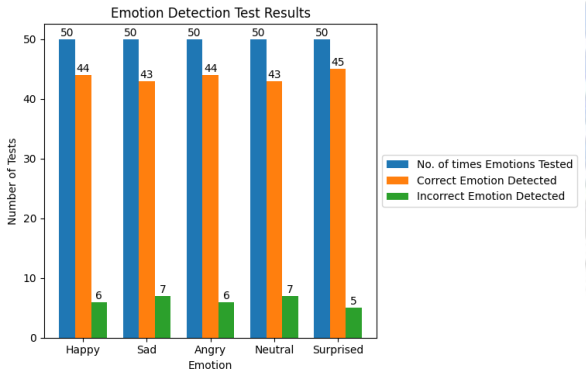
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 8 Issue: 9
Article Received: *25 July 2020* ***Revised:*** *12 September 2020* ***Accepted:*** *30 September 2020*

Figure 10. '*Sad*' Emotion Detection with sad songs play list selection



To evaluate the performance of the proposed system, four users were asked to test each emotion category (Happy, Sad, Angry, Surprised, Neutral, etc.) 10 times using their system's webcam. The images were captured under adequate lighting conditions using webcams with a minimum resolution of 2 MP or higher. After conducting the experiments, the obtained results were recorded and are presented in the following graph.

Figure 11. Emotion detection test results



Based on the above data, the overall accuracy for correctly detecting emotions was calculated to be 87.60%. In addition, emotion-wise accuracy was also evaluated to analyze the model's performance for each emotional category. The results indicate that the "*Surprized*" emotion achieved the highest detection accuracy, while the "*Sad*" and "*Neutral*" emotions exhibited the lowest accuracy in correctly identifying the corresponding emotional state.

Figure 12. Emotion-wise and overall accuracy

Experimental Results

	Emotion	Tested	Correct Detection	Incorrect Detection	Accuracy (%)
0	Happy	50	44	6	88.0
1	Sad	50	43	7	86.0
2	Angry	50	44	6	88.0
3	Neutral	50	43	7	86.0
4	Surprised	50	45	5	90.0

Overall System Accuracy: 87.6 %

5. Conclusion and Future Scope

The developed Facial Expression–Based Emotion Detection and Music Player using CNN operates effectively on desktop and laptop systems, achieving an accuracy of 87.60% in detecting the user's correct emotion or mood. The system offers a user-friendly mechanism that shifts the traditional manual song selection process to an automated music recommendation system.

It enhances the overall music listening experience by automatically generating a playlist of songs based on the user's current emotional state. This approach simplifies the process of navigating through large music libraries by displaying only those songs that correspond to the detected emotion.

In addition to standard music player controls, the system also provides additional features, such as allowing users to select their preferred language of songs and enabling them to create and manage their own playlists, thereby offering a more personalized and convenient music experience.

The proposed system is currently designed to operate on desktop platforms only. In the future, it can be extended and developed as a mobile application to support greater mobility, enabling users to access the system anytime and anywhere.

At present, the playlist is generated manually based on the emotional categories of songs. In future work, the song classification process can be automated to generate playlists dynamically, which would significantly improve the efficiency and usability of the system.

Additionally, the image capturing mechanism can be further enhanced to perform more effectively under low-light conditions, thereby improving the overall accuracy and reliability of emotion detection.

References

[1] Ultimate Guide for Facial Emotion Recognition Using A CNN by PrudhviGnv on medium.com.

[2] EMO-MUSIC (Emotion based Music player) by Sarvesh Pal, Ankit Mishra, Hridaypratap Mourya and Supriya Dicholkar on EasyChair Preprint no. 2463.

[3] HafeezKabani, Sharik Khan, Omar Khan, Shabana Tadvi "Emotion Based Music Player", International Journal of Engineering Research and General Science Volume 3, Issue 1.

[4] S Metilda Florence and M Uma, "Emotional Detection and Music Recommendation System based on User Facial Expression", IOP Conference Series: Materials Science and Engineering.

[5] Dhruvisha Bansal, Pinkal Bhatt, Megha Dusane, Avneet Saluja, Kushal Patel, "Emotion Based Music Playing Device", International Research Journal of Engineering and Technology (IRJET) Volume: 07.

[6] Mr. T. M. Hayath, Ms. Sulthana, Mr. Kampli Shabbir, "A Novel Method To Design Emotion-Based-Music-Player", International Journal Of Engineering Development And Research.

[7] Aditya Gupte, Arjun Naganarayanan, Manish Krishnan , "Emotion Based Music Player – XBeats", Department of Computer Engineering, SIES GST, Mumbai University, Navi Mumbai, India, vol. 3.

[8] Pushkar Mahajan, Pratik Khurad, Prateek Chaudhari, "Face Player: Facial Emotion Based Music Player", International Journal of Research and Analytical Reviews (IJRAR), Volume 7.