Journal of Informatics
Electrical and Electronics Engineering, 2021,
Vol. 02, Iss. 02, S. No. 018, pp. 1-11
ISSN (Online): 2582-7006

Music Recommendation Based on Face Emotion Recognition

Madhuri Athavle[1], Deepali Mudale[2], Upasana Shrivastav[3], Megha Gupta[4]

[1,2,3,4]Department of Computer Engineering, NHITM, Thane, India
[1]madhuriaathavle@gmail.com, [2]deepalimudale35@gmail.com, [3]upasanashrivastav172@nhitm.ac.in, [4]meghagupta@nhitm.ac.in.

How to cite this paper: M. Athavle, D. Mudale, U. Shrivastav and M. Gupta (2021) Music Recommendation Based on Face Emotion Recognition. *Journal of Informatics Electrical and Electronics Engineering, Vol. 02, Iss. 02, S. No. 018, pp. 1-11, 2021*.

https://doi.org/10.54060/JIEEE/002.02.018

Received: 07/04/2021
Accepted: 27/05/2021
Published: 09/06/2021

Abstract

We propose a new approach for playing music automatically using facial emotion. Most of the existing approaches involve playing music manually, using wearable computing devices, or classifying based on audio features. Instead, we propose to change the manual sorting and playing. We have used a Convolutional Neural Network for emotion detection. For music recommendations, Pygame & Tkinter are used. Our proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the system's overall accuracy. Testing of the system is done on the FER2013 dataset. Facial expressions are captured using an inbuilt camera. Feature extraction is performed on input face images to detect emotions such as happy, angry, sad, surprise, and neutral. Automatically music playlist is generated by identifying the current emotion of the user. It yields better performance in terms of computational time, as compared to the algorithm in the existing literature.

Keywords

Face Recognition, Feature extraction, Emotion detection, Convolutional Neural Network, Pygame, Tkinter, Music, Player, Camera.

1. Introduction

Many of the studies in recent years admit that humans reply and react to music and this music has a high impression on the activity of the human brain. In one examination of the explanations why people hear music, researchers discovered that music played a crucial role in relating arousal and mood. Two of the most important functions of music are it is ability is participants rated to help them achieve a good mood and become more self-aware. Musical preferences have been demonstrated to be highly related to personality traits and moods [1].

The meter, timbre, rhythm, and pitch of music are managed in areas of the brain that affects emotions and mood [2]. Interaction between individuals may be a major aspect of lifestyle. It reveals perfect details and much of data among humans, whether they are in the form of body language, speech, facial expression, or emotions [3]. Nowadays, emotion detection is considered the most important technique used in many applications such as smart card applications, surveillance, image database investigation, criminal, video indexing, civilian applications, security, and adaptive human-computer interface with multimedia environments.

With the increase in technology for digital signal processing and other effective feature extraction algorithms, automated emotion detection in multimedia attributes like music or movies is growing rapidly and this system can play an important role in many potential applications like human-computer interaction systems and music entertainment. We use facial expressions to propose a recommender system for emotion recognition that can detect user emotions and suggest a list of appropriate songs [13-24]. The proposed system detects the emotions of a person, if the person has a negative emotion, then a certain playlist will be shown that includes the most related types of music that will enhance his mood. And if the emotion is positive, a specific playlist will be presented which contains different types of music that will inflate the positive emotions [4].

The dataset we used for emotion detection is from Kaggle Facial Expression Recognition [5]. Dataset for the music player has been created from Bollywood Hindi songs. Implementation of facial emotion detection is performed using Convolutional Neural Network which gives approximately 95.14% of accuracy [2].

2. Literature Review

The review is done to get insights into the methods, their shortcoming which we can overcome. A literature review, a literature survey is a text of a scholarly paper, which includes the current understanding along with great findings, as well as theoretical and methodological contributions to a particular topic. The latent qualities of humans that can provide inputs to any system in various ways have brought the attention of several learners, scientists, engineers, etc. from all over the world.

The current mental state of the person is provided by facial expressions. Most of the time we use nonverbal clues like hand gestures, facial expressions, and tone of voice to express feelings in interpersonal communication. Preema et al [6] stated that it is very time-consuming and difficult to create and manage a large playlist. The paper states that the `music player itself selects a song according to the current mood of the user. The application scans and classifies the audio files according to audio features to produce mood-based playlists. The application makes use of the Viola-Jonas algorithm that is used for face detection and facial expression extraction. Support Vector Machine (SVM) was used in the classification extracted features into 5 major universal emotions like anger, joy, surprise, sad, and disgust.

Yusuf Yaslan et al. proposed an emotion-based music recommendation system that learns the user's emotion from signals obtained through wearable computing devices that are integrated with galvanic skin response (GSR) and photoplethysmography (PPG) physiological sensors in their paper [3]. Emotions are a basic part of human nature. They play a vital role throughout life. In this paper, the emotion recognition problem is taken into account as arousal and valence prediction from multi-channel physiological signals. In [7] Ayush Guidel et al stated that human being’s state of mind and current emotional mood can be easily observed through their facial expressions. This system was developed by taking basic emotions (happy, sad, anger, excitement, surprise, disgust, fear, and neutral) into consideration. Face detection in this project was implemented by using a convolutional neural network. Music is usually told as a "language of emotions" throughout the planet.

The paper proposed by Ramya Ramanathan et al [1] conveyed the intelligent music player using emotion recognition. Emotions are a very basic part of human nature. They play the most important role throughout life. Human emotions are meant for sharing feelings and mutual understanding. The user's local music selection is initially grouped based on the emotion conveyed by the album. this is often calculated taking into consideration the song's lyrics. The paper specifically makes a

specialty of the methodologies available for detecting human emotions for developing emotion-based music players, the approach a music player follows to detect human emotions, and the way it is ideal to apply the proposed system for emotion detection. It additionally offers a brief idea about our systems working, playlist generation, and emotion classification.CH Radhika et al [8] advised manual segregation of a playlist and annotation of songs, following the current emotional state of a user, as a labor-intensive and time-consuming task. Numerous algorithms had been proposed to automate this manner. However, the prevailing algorithms are slow, increase the overall cost of the system by using additional hardware (e.g., EEG structures and sensors), and feature much less accuracy. The paper presents an algorithm that automatically does the process of generating a playlist of audio, based on the facial expressions of a person, for rendering salvage of time as well as labor, invested in performing this process manually. The algorithm given in the paper directs at reducing the overall computational time and the cost of the designed system. It additionally aims at growing the accuracy of the system design. The system's facial expression recognition module is validated by comparing it to a dataset that is both user-dependent and user-impartial.

3. Problem Definition

Develop a system that presents a cross-platform music player, which recommends music based on the real-time mood of the user through a web camera using Machine Learning Algorithms.

4. Proposed System Overview

The proposed system benefits us to present interaction between the user and the music player. The purpose of the system is to capture the face properly with the camera. Captured images are fed into the Convolutional Neural Network which predicts the emotion. Then emotion derived from the captured image is used to get a playlist of songs. The main aim of our proposed system is to provide a music playlist automatically to change the user's moods, which can be happy, sad, natural, or surprised. The proposed system detects the emotions, if the topic features a negative emotion, then a selected playlist is going to be presented that contains the foremost suitable sorts of music that will enhance the mood of the person positively. Music recommendation based on facial emotion recognition contains four modules.

- Real-Time Capture: In this module, the system is to capture the face of the user correctly
- Face Recognition: Here it will take the user's face as input. The convolutional neural network is programmed to evaluate the features of the user image.
- Emotion Detection: In this section extraction of the features of the user image is done to detect the emotion and depending on the user's emotions, the system will generate captions.
- Music Recommendation: Song is suggested by the recommendation module to the user by mapping their emotions to the mood type of the song.

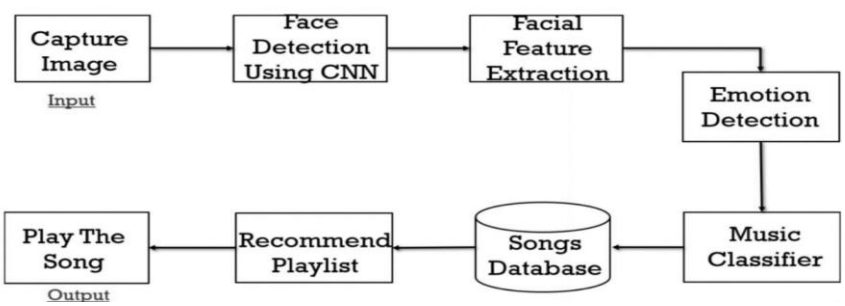

Figure 1. Block diagram of the proposed system.

5. Methodology

5.1 Database Description

We built the Convolutional Neural Network model using the Kaggle dataset. The database is FER2013 which is split into two parts training and testing dataset. The training dataset consists of 24176 and the testing dataset contains 6043 images. There are 48x48 pixel grayscale images of faces in the dataset. Each image in FER-2013 is labeled as one of five emotions: happy, sad, angry, surprise, and neutral. The faces are automatically registered so that they are more or less centered in each image and take up about the same amount of space. The images in FER-2013 contain both posed and unposed headshots, which are in grayscale and 48x48 pixels.

The FER-2013 dataset was created by gathering the results of a Google image search of every emotion and synonyms of the emotions. FER systems being trained on an imbalanced dataset may perform well on dominant emotions such as happy, sad, angry, neutral, and surprised but they perform poorly on the under-represented ones like disgust and fear. Usually, the weighted-SoftMax loss approach is used to handle this problem by weighting the loss term for each emotion class supported by its relative proportion within the training set. However, this weighted-loss approach is predicated on the SoftMax loss function, which is reported to easily force features of various classes to stay apart without listening to intra-class compactness. One effective strategy to deal with the matter of SoftMax loss is to use an auxiliary loss to coach the neural network. To treating missing and Outlier values we have used a loss function named categorical crossentropy. For each iteration, a selected loss function is employed to gauge the error value. So, to treating missing and Outlier values, we have used a loss function named categorical crossentropy.

Figure 2. Samples from FER2013 dataset.

5.2 Emotion Detection Module

5.2.1 Face Detection

Face detection is one of the applications which is considered under computer vision technology. This is the process in which algorithms are developed and trained to properly locate faces or objects in object detection or related system in

images. This detection can be real-time from a video frame or images. Face detection uses such classifiers, which are algorithms that detect what's either a face (1) or not a face (0) in an image. Classifiers are trained to detect faces using numbers of images to get more accuracy. OpenCV uses two sorts of classifiers, LBP (Local Binary Pattern) and Haar Cascades. A Haar classifier is used for face detection where the classifier is trained with pre-defined varying face data which enables it to detect different faces accurately. The main aim of face detection is to spot the face within the frame by reducing external noises and other factors. It is a machine learning-based approach where the cascade function is trained with a group of input files. It is supported the Haar Wavelet technique to research pixels inside the image into squares by function [9]. This uses machine learning techniques to urge a high degree of accuracy from what's called "training data".

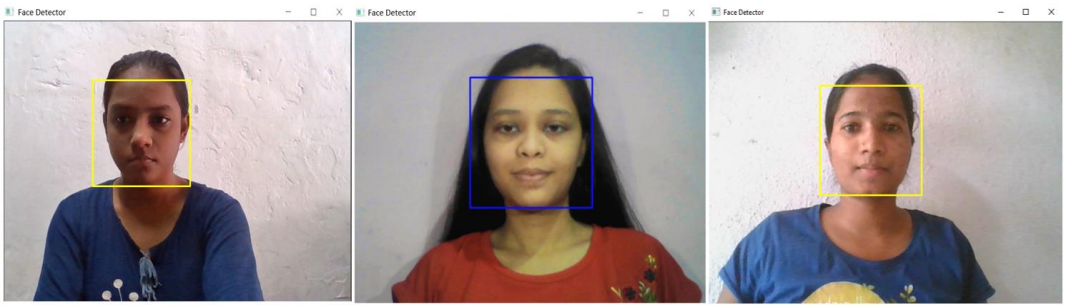

Figure 3. Face detection.

5.2.2 Feature Extraction

While performing feature extraction, we treat the pre-trained network that is a sequential model as an arbitrary feature extractor. Allowing the input image to pass on it forward, stopping at the pre-specified layer, and taking the outputs of that layer as our features. Starting layers of a convolutional network extract high-level features from the taken image, so use only a few filters. As we make further deeper layers, we increase the number of the filters to twice or thrice the dimension of the filter of the previous layer. Filters of the deeper layers gain more features but are computationally very intensive.

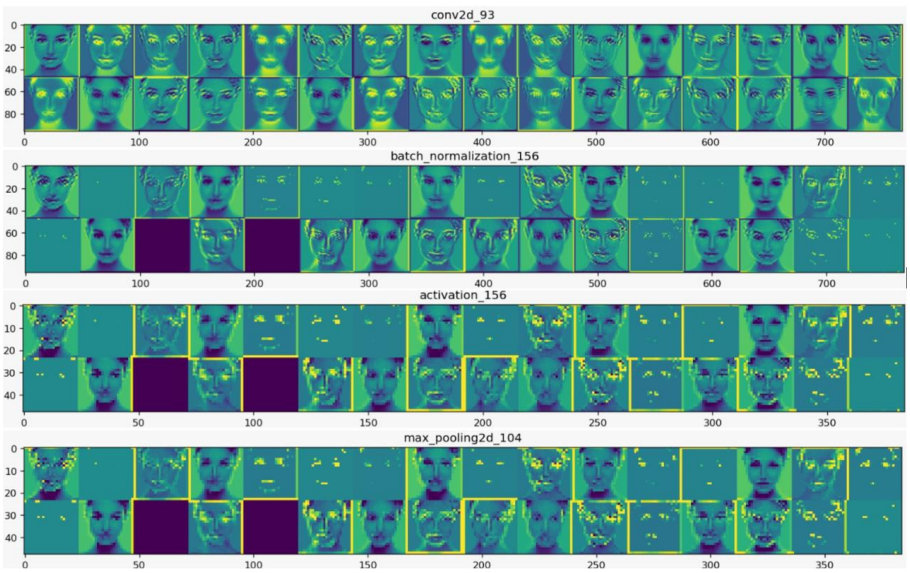

Figure 4. Visualization of The Feature Map.

Doing this we utilized the robust, discriminative features learned by the Convolution neural network [10]. The outputs of the model are going to be feature maps, which are an intermediate representation for all layers after the very first layer. Load the input image for which we want to view the Feature map to know which features were prominent to classify the image. Feature maps are obtained by applying Filters or Feature detectors to the input image or the feature map output of the prior layers. Feature map visualization will provide insight into the interior representations for specific input for each of the Convolutional layers within the model.

5.2.3 Emotion Detection

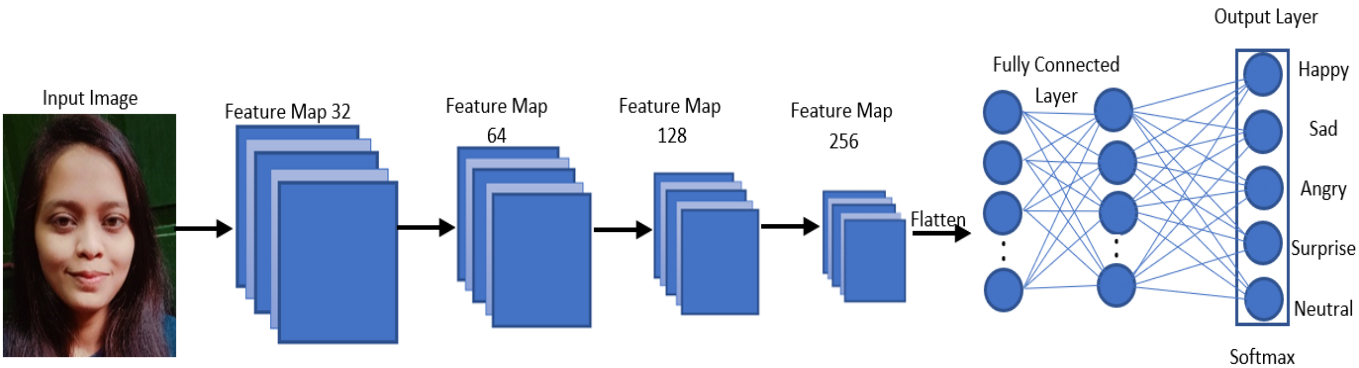

Figure 5. Convolution neural network Architecture.

Convolution neural network architecture applies filters or feature detectors to the input image to get the feature maps or activation maps using the Relu activation function [11]. Feature detectors or filters help in identifying various features present in the image such as edges, vertical lines, horizontal lines, bends, etc. After that pooling is applied over the feature maps for invariance to translation. Pooling is predicted on the concept that once we change the input by a touch amount, the pooled outputs don't change. We can use any of the pooling from min, average, or max. But max-pooling provides better performance than min or average pooling. Flatten all the input and giving these flattened inputs to a deep neural network which are outputs to the class of the object.

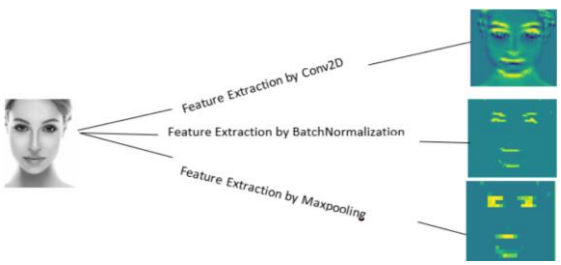

Figure 6. Feature Extraction by each layer in Convolutional Neural Network.

The class of the image will be binary, or it will be a multi-class classification for identifying digits or separating various apparel items. Neural networks are as a black box, and learned features in a Neural Network are not interpretable. So basically, we give an input image then the CNN model returns the results [10]. Emotion detection is performed by loading the model which is trained by weights using CNN. When we take the real-time image by a user then that image was sent to the pre-trained CNN model, then predict the emotion and adds the label to the image.

Figure 7. Results of Emotion Detection.

5.3 Music Recommendation Module

5.3.1 Songs Database

We created a database for Bollywood Hindi songs. It consists of 100 to 150 songs per emotion. As we all know music is undoubtedly involved in enhancing our mood. So, suppose a user is sad then the system will recommend such a music playlist which motivates him or her and by this automatic mood will be delighted.

5.3.2 Music Playlist Recommendation

By using the emotion module real-time emotion of the user is detected. This will give the labels like Happy, Sad, Angry, Surprise, and Neutral. Using the os.listdir() method in python we connected these labels with the folders of the songs database which we have created. **Table 1** shows the list of songs. This method of os.listdir() is used to get the list of any file in the specified directories.

```
if label== 'Happy':
            os.chdir("C:/Users/deepali/Downloads/Happy")
           self.mood.set("You are looking happy, I am playing song for You")
# Fetching Songs
            songtracks = os.listdir()
```

Table 1. Database of songs.

Emotion	Songs
Happy	Track 1"Dil Dhadakne Do"
	Track 2"Aaj Mai Upar"
	Track 3 "Ilahi"
Sad	Track 1 "Apna Time Aayega"

	Track 2 "Ruk Jana Nahi"
	Track 3 "All is Well"
Angry	Track I "Dushman Na Kare Dost Ne"
	Track 2 "Thukra Ke Mera Pyaar"
	Track 3 "Khalbali'
Surprise	Track I "Zindagi Kaisi Hai Paheli"
	Track 2 "Aao Milon Chalen"
	Track 3 "Jaane Kyun"
Neutral	Track I "Buddhu Sa Mann"
	Track 2 "Matargashti"
	Track 3 "Dildara"

```
# Inserting Songs into Playlist
        for track in songtracks:
            self.playlist.insert (END, track)
```

This will result in the recommended playlist for the user in the GUI of the music player by showing captions according to detected emotions. We have used a library called Pygame for playing the audio as this library supports playing various multimedia formats like audio, video, etc. Functions of this library such as playsong, pauseong, resumesong, and stopsong are used to working with the music player. Variables like playlist, songstatus, and root are used for storing the name of all songs, storing the status of currently active songs, and for the main GUI window respectively. For developing the GUI, we have used Tkinter.

Here are the results:

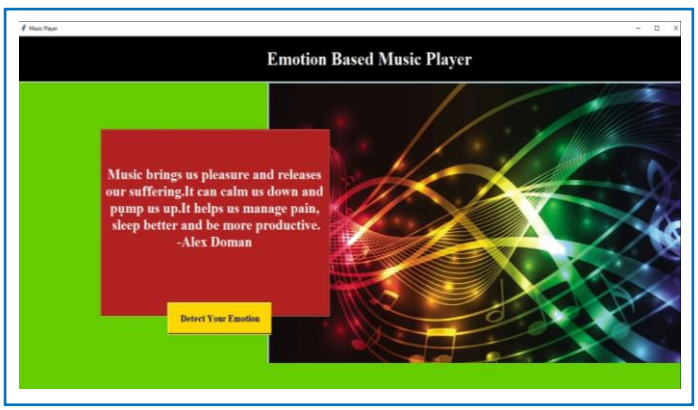

Figure 8. GUI of the front page.

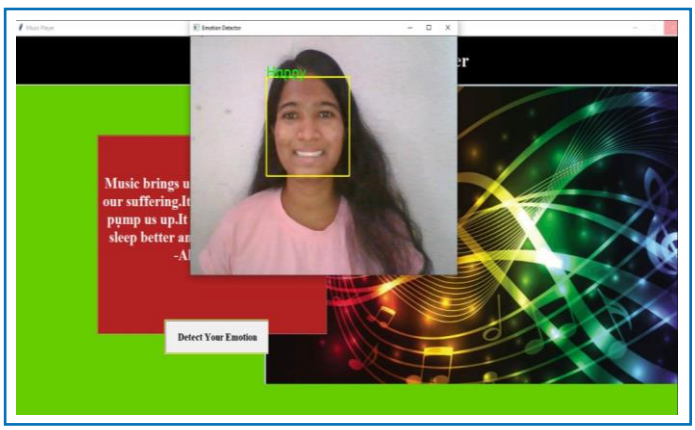

Figure 9. Detection of emotion.

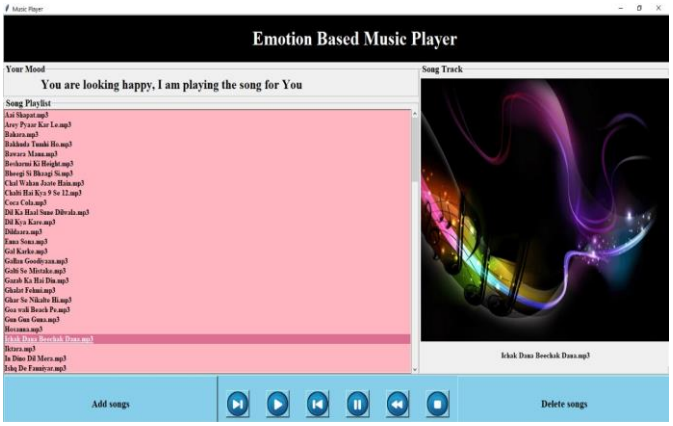

Figure 10. Recommendation of music playlist.

6. Result & Analysis

We evaluated a number of the studies which use support vector machine (SVM), extreme learning machine (ELM), and convolutional neural network [12]. **Table 2** shows the comparison of related algorithms. Corresponding algorithms and accuracy values are given for each study. The usage of a Convolutional Neural Network improves the efficiency of the emotion detection accuracy.

Table 2. Validation and Testing accuracy for the three algorithms on the Fer2013 Dataset.

Algorithm	SVM	ELM	CNN
Validation Accuracy	0.66	0.62	0.95
Testing Accuracy	0.66	0.63	0.71

Table 3 shows hyperparameters for the trained CNN network. The learning rate regulates the update of the weight at the end of each batch. Several epochs of the iterations of the entire training dataset to the network during training. Batch size the number of patterns shown in the network before the weights are updated. Activation functions allow the model to learn nonlinear prediction boundaries. Adam may be a replacement optimization algorithm for stochastic gradient descent for training deep learning models. The loss function categorical- crossentropy is employed to quantify deep learning model errors, typically in single-label, multi-class classification problems.

Table 3. Hyperparameter for trained CNN network.

Hyperparameters	Values
Batch size	128
No. of classes	5
Optimizer	Adam
Learning rate	0.001
Epoch	48
No. of Layers	28
Activation function	Relu, SoftMax
Loss function	Categorical-crossentropy

7. Conclusion

A thorough review of the literature tells that there are many approaches to implement Music Recommender System. A study of methods proposed by previous scientists and developers was done. Based on the findings, the objectives of our system were fixed. As the power and advantages of AI-powered applications are trending, our project will be a state-of-the-art trending technology utilization. In this system, we provide an overview of how music can affect the user's mood and how to choose the right music tracks to improve the user's moods. The implemented system can detect the user's emotions. The emotions that the system can detect were happy, sad, angry, neutral, or surprised. After determining the user's emotion, the proposed system provided the user with a playlist that contains music matches that detected the mood. Processing a huge dataset is memory as well as CPU intensive. This will make development more challenging and attractive. The motive is to create this application in the cheapest possible way and also to create it under a standardized device. Our music recommendation system based on facial emotion recognition will reduce the efforts of users in creating and managing playlists.

8. Future Scope

This system, although completely functioning, does have scope for improvement in the future. There are various aspects of the application that can be modified to produce better results and a smoother overall experience for the user. Some of these that an alternative method, based on additional emotions which are excluded in our system as disgust and fear. This emotion included supporting the playing of music automatically. The future scope within the system would style a mechanism that might be helpful in music therapy treatment and help the music therapist to treat the patients suffering from mental stress, anxiety, acute depression, and trauma. The current system does not perform well in extremely bad light conditions and poor camera resolution thereby provides an opportunity to add some functionality as a solution in the future.

Acknowledgement

The building of a B.E project needs the co-operation and guidance of several people. We, therefore consider it our prime duty to thank all those who helped us during this venture.

We would like to thank our Principal, Dr. P. D. Deshmukh, for inspiring us and providing us with the requisite resources during our time working on this project. We would also like to express our gratitude to Dr. Sanjay Sharma, Head of the Department of Computer Engineering, for his kind co-operation.

It is with great pleasure that we express our appreciation to Ms. Megha V. Gupta, our Project Guide, for providing us with constructive and encouraging feedback during the planning of this project, as well as for constantly directing us and providing us with useful insights.

Last but not the least, we are thankful to our friends, the teaching and the non-teaching staff whose encouragement and suggestions helped us to enhance our B.E Project. We also are thankful to our parents for their constant support and best wishes.

References

[1]. R. Ramanathan, R. Kumaran, R. R. Rohan, et al., "An intelligent music player based on emotion recognition," in 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS), pp.1-5,2017.

[2]. S. Gilda, H. Zafar, C. Soni, et al., "Smart music player integrating facial emotion recognition and music mood recommendation," in 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), pp.154-158, 2017.

[3]. D. Ayata, Y. Yaslan, and M. E. Kamasak, "Emotion based music recommendation system using wearable physiological sensors," IEEE trans. consum. electron., vol. 64, no. 2, pp. 196–203, 2018.

[4]. A. Alrihaili, A. Alsaedi, K. Albalawi, et al., “Music recommender system for users based on emotion detection through facial features,” in 12th International Conference on Developments in eSystems Engineering (DeSE), pp.1014-1019,2019.

[5]. I. J. Goodfellow, D. Erhan, Y. Bengiou, et al., “Challenges in representation learning: A report on three machine learning contests,” in Neural Information Processing, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 117–124 2013.

[6]. J.S.Preema Rajashree, M.Sahana, et al.,” Review on facial expression-based music player,” International Journal of Engineering Research & Technology (IJERT), vol. 6, no.15, 2018.

[7]. A. Guidel, B .Sapkota, K. Sapkota, “Music recommendation by facial analysis,” 2020.

[8]. C.H. Sadhvika, G. Abigna, P. S. Reddy, “Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology,” Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR), vol.7, no. 4, April 2020.

[9]. V. Tabora,” Face detection using OpenCV with Haar Cascade Classifiers,” Becominghuman.ai,2019.

[10]. Z. Qin, F. Yu, C. Liu,et al, “How convolutional neural network see the world - A survey of convolutional neural network visualization methods,” 2018.

[11]. K. Chankuptarat, R. Sriwatanaworachai and S. Chotipant, "Emotion-Based Music Player," *5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1-4, 2019.

[12]. F. Norden and F.V.R. Marlevi, “A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition, “TRITA-EECS-EX:143 pp.9,2019.

[13]. P. Singhal, P. Singh, and A. Vidyarthi, “Interpretation and localization of Thorax diseases using DCNN in Chest X-Ray, “Journal of Informatics Electrical and Electronics Engineering, vol.1, no.1, pp.1-7,2020.

[14]. M. Vinny, P. Singh, “Review on the Artificial Brain Technology: Blue Brain, “Journal of Informatics Electrical and Electronics Engineering, vol.1, no.1, pp.1-11,2020.

[15]. A. Singh and P. Singh, “Object Detection. Journal of Management and Service Science, “vol.1, no.2, pp. 1-20,2021.

[16]. A. Singh, P. Singh, “Image Classification: A Survey, “Journal of Informatics Electrical and Electronics Engineering, vol.1, no.2, pp. 1-9,2020.

[17]. A. Singh and P. Singh, “License Plate Recognition, “Journal of Management and Service Science, vol.1, no.2, pp. 1-14,2021.