The Pharma Innovation Journal 2019; 8(2): 850-855

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(2): 850-855

www.thepharmajournal.com
Received: 01-12-2018
Accepted: 06-01-2019

Varun Tiwari
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Correspondence Author;
Varun Tiwari
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Music recommendation based on face emotion recognition

Varun Tiwari

DOI: https://doi.org/10.22271/tpi.2019.v8.i2n.25417

Abstract
This paper proposed a new way to play music on the go. Most current methods involve manual music processing, using wearables, audio classification. Instead, it is recommended switching to manual connection and working traditionally. There convolutional neural network is used for emotion recognition. Using Pygame and Tkinter for better visualization. Our suggested approach will increase the system's overall efficiency by reducing the amount of time needed to get results and the overall cost of installation. Face expressions are captured by the built-in camera during system evaluation experiments on the FER2013 dataset. Concept face images are subjected to feature extraction in order to identify various emotions, including happy, angry, sad, surprised, and moderate. A user's current preferences are analyzed to update the music playlist. When compared to the algorithm in the current literature, it performs better in terms of computation time.

Keywords: Face recognition, music, player, camera, Tkinter, Pygame, convolutional neural network, feature extraction, and emotion detection

1. Introduction
Previous research has demonstrated that people react to music, which has a profound effect on how human brains function. Researchers have discovered that music has a significant influence on emotions and sentiments, which helps to explain why people listen to it. Among the most significant functions of music are its ability to enhance participants' mood and self-awareness. Research has demonstrated the correlation between musical tastes and one's attitudes and emotions [1].

The area of the brain that affects the mind and emotions is where beats, timbre, rhythms, and music is controlled. A vital part of life can be interacting with other people. It provides users with a wealth of information while paying close attention to facial expressions, body language, speech, and emotions. Today, it is an important technique used in many applications such as emotion recognition, smart-cards, invigilation, image detection, crime, video indexing, public use, security, and adapting human-machine interfaces to various environments [2].

The rapid advancement of signal processing technology and other efficient extraction algorithms has facilitated the automatic recognition of diverse materials, including music and videos. This system holds great potential for numerous applications, including human-computer interaction. A facial recognition system is used for emotional recognition, which allows it to identify the user's emotions and recommend appropriate playlists [3].

The system readies itself to detect an individual's emotional state; in the event that the person is depressed, a customized playlist featuring the most upbeat music will be displayed. It has been demonstrated that a certain list contains a variety of musical genres that elicit happy feelings when the atmosphere is upbeat [4].

The data utilized for facial expression recognition is from Kaggle Facial Recognition. The facial recognition application is done using a convolutional neural network that provides up to 95.14%.

2. Literature Review
The fact that humans commonly convey our emotions in interpersonal relationships through non-verbal cues like gestures, facial expressions, and voice pitch demonstrates how challenging and time-consuming it is to create and manage huge playlists [5]. According to the newspaper, the music chooses songs based on the user's current mood.

The application scans as well as categorizes audio files based on audio characteristics to create mood playlists. The app uses Viola-Jonas algorithm for face detection and face removal.
Five themes—angry, joyful, surprised, depressed, and embarrassed—are expanded into features of a Support Vector Machine (SVM), which is used for classification [6].
In article, galvanic skin response and photoplethysmography physiological sensors have proposed an emotion-based music recommendation from the signal obtained from a wearable computing device with working user perception. Emotions are an important aspect of the human body [7].
Cognitive issues in the article are related to perceiving and estimating different motor systems. A person's emotional state and present mood can be readily read on their face, as was mentioned in. The main emotions—happy, sad, angry, surprised, disgusted, fearful, and neutral—were taken into consideration when designing the system. In this project, convolutional neural networks are used for face detection. Music is frequently viewed as a "emotional language" worldwide [8]."
Emotions are important part of the human body. They play the most crucial role in life. The human desire is unity and understanding. One user of the local music selection started the band from the perspective conveyed by the album. This is usually included in the lyrics of the account. In particular, the article presents a technique that can be used to analyses human emotions to create emotion-based music, a music-based technique to capture emotions, and the excellent way to use this technique for researching theory [9]. Along with a brief explanation of playlist creation and distribution theory, it also describes how our system operates. The desire to divide playlists and song annotations based on the user's current mood was expressed by CH Radhika *et al*. This is a tedious and time-consuming task. Various algorithms have been suggested to mechanize the procedure [10].
However, widely suggested algorithms are less sensitive, slow, and require extra hardware, like sensors and EEG models, which raises the system's overall cost. Provide an algorithm that saves time and labor when generating a playlist based on a person's facial expressions instead of requiring human intervention. The algorithms presented in this paper aim to reduce the overall computation time and system creation cost. It also aims to improve the efficiency of the design. The system's facial emotion recognition system is verified by comparison with both user and non-independent data [11].

3. Problem Definition
The paper focuses on creating a cross-platform music player that uses machine learning algorithms to provide music recommendations based on the user's current mood via any camera.

4. Proposed System Overview
The way the user and musician interact is presented more simply in the suggested application. The camera's primary job on the system is to take pictures of faces. A neural network that predicts emotions is fed captured photos. To obtain a playlist, requirements from the captured images are utilized [11]. Our system's second major goal is to give users access to a music playlist that can alter their mood from happy to sad, natural to unexpected. It is advised to act to observe the subject's emotions. If the subject exhibits negative emotions, a list accompanied by the most uplifting music is shown to help the person become more compassionate. The following modules comprise facial recognition-based music recognition:

- Real-Time Face Capturing: This module correctly captures the user's face.
- Face Recognition Module: It will take in facial data from the user. The end-user image's features are evaluated by the convolutional neural network according to its programming [12].
- Emotion Detection Module: In this section, the end-user image's features are extracted in order to identify the user's emotions. The system then generates captions in accordance with those emotions.
- Music Recommendation Module: The user's expressions are mapped to the mood type of songs, allowing the recommendation function to suggest a list of songs [13].

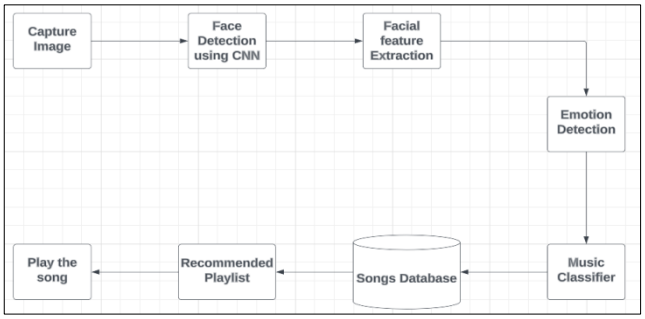

Fig 1: Workflow of the system

5. Methodology
5.1 Database Description
The dataset FER2013 consists of thousands of images with different human emotions along with a caption that describes the emotion detected. The following figure shows some of the sample images.

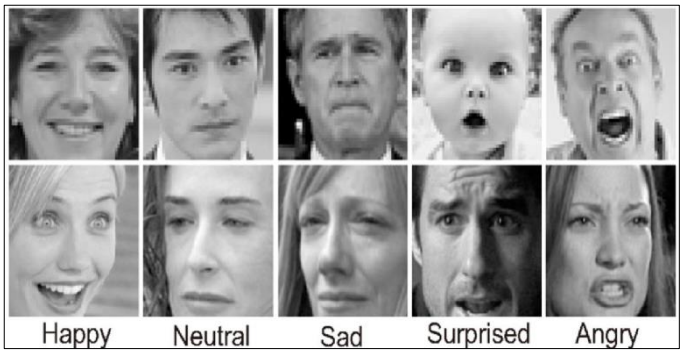

Fig 2: Samples images from FER2013 dataset.

5.2 Emotion Detection Function
using the Kaggle dataset to create a convolutional neural network (CNN) model. The FER2013 database was divided into training and test data sections. There are 24170 images in the training file and 6040 images in the test file. Grayscale images of human faces, 48 by 48 pixels, are included in the file. Five moods are labeled on each picture in FER-2013: neutral, surprise, anger, sadness, and happiness [14].
The faces are recorded so that they are more or less in the same place in each image. Images in FER-2013 have exposed and unexposed heads, grayscale and 48x48 pixels.
FER-2013 was crafted by collecting all the thoughts and feelings of the mind from Google Image Searches. While the FER system trained with random data will perform well on

emotions such as happiness, sadness, anger, neutrality, and surprise, it will not perform well on Negative emotions such as hatred and fear. This problem is typically solved using the weighted SoftMax method, which involves weighing the time loss for each class hypothesis that is supported by its relative equivalent in the training program.

Nonetheless, the SoftMax function—which is registered to forcefully separate the properties of various classes without taking into account the compactness class—is the foundation of the weight loss strategy. Training neural networks with the loss program is a smart way to address the SoftMax loss issue. To deal with missing values, categorical cross-entropy is used. For every iteration, the failure option is used to evaluate the error rate. Therefore, a loss function called categorical cross-entropy to check for values and missing values.

5.2.1 Face Detection Function

One of the main uses for computer vision systems is face detection. It is the process of creating and refining algorithms to track faces or objects in photos accurately enough to detect objects or related processes. This can be discovered in real time via snapshots or images. Face detection uses segmentation, a process that detects faces (1) or no faces (0) in images. Classifiers are trained to identify facial expressions using multiple images for greater efficiency. Haar cascades and LBP (Local Binary Patterns) are the two classification methods used by OpenCV. When using Haar classifiers for face detection, a predefined set of distinct faces is used to train the classifier, enabling it to recognize a variety of faces. The primary goal of face detection is to identify the face in a frame by minimizing outside noise and other distractions. This method, which is based on machine learning, uses input data to train the cascade function. Encourage the use of Haar Wavelet technology to enable function-based learning of image pixels in frames.

It extracts a high degree of efficiency from the "learning data" by utilizing a variety of machine learning techniques.

```
class VideoCamera(object):

    def get_frame(self):
        global cap1
        global df1
        cap1 = WebcamVideoStream(src=0).start()
        image = cap1.read()
        image=cv2.resize(image,(600,500))
        gray=cv2.cvtColor(image,cv2.COLOR_BGR2GRAY)
        face_rects=face_cascade.detectMultiScale(gray,1.3,5)
        df1 = pd.read_csv(music_dist[show_text[0]])
        df1 = df1[['Name','Album','Artist']]
        df1 = df1.head(15)
        for (x,y,w,h) in face_rects:
            cv2.rectangle(image,(x,y-50),(x+w,y+h+10),(0,255,0),2)
            roi_gray_frame = gray[y:y + h, x:x + w]
            cropped_img = np.expand_dims(np.expand_dims(cv2.resize(roi_gray_frame, (48, 48)), -1), 0)
            prediction = emotion_model.predict(cropped_img)

            maxindex = int(np.argmax(prediction))
            show_text[0] = maxindex
            #print("===================================",music_dist[show_text[0]],"===================================")
            #print(df1)
            cv2.putText(image, emotion_dict[maxindex], (x+20, y-60), cv2.FONT_HERSHEY_SIMPLEX, 1, (255, 255, 255), 2, cv2.LINE_AA)
            df1 = music_rec()
```

Fig 3: Filtering images based on emotions

In figure 4, various facial emotions are passed to the model to check its accuracy and performance. A few shortcomings were noticed that can be overcomed in the upcoming versions.

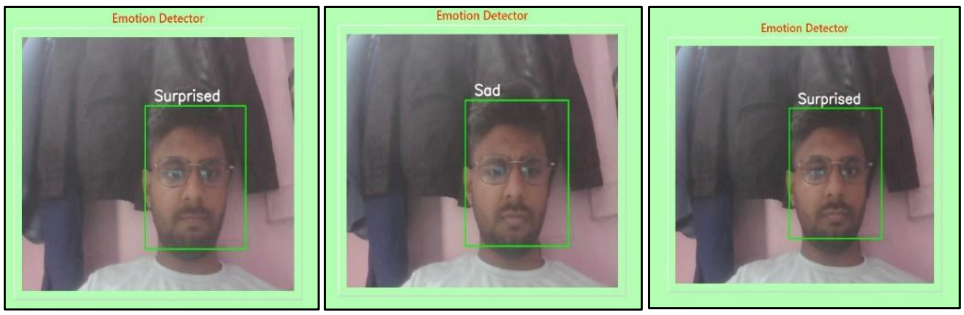

Fig 4: Testing of model on random images

5.2.2 Feature Extraction

The pre-trained network is used as a random feature extractor or sequence model when feature extraction is successful. We can proceed based on our properties, pause in the previous process, and utilize the process's output by using the input view. Only a few filters are used because the convolutional mesh layer extracts high-level features from the captured image. Increase the filter size to double or triple the filter size of the preceding layer as deeper layers are generated. Deep filters are more expensive but require more features.

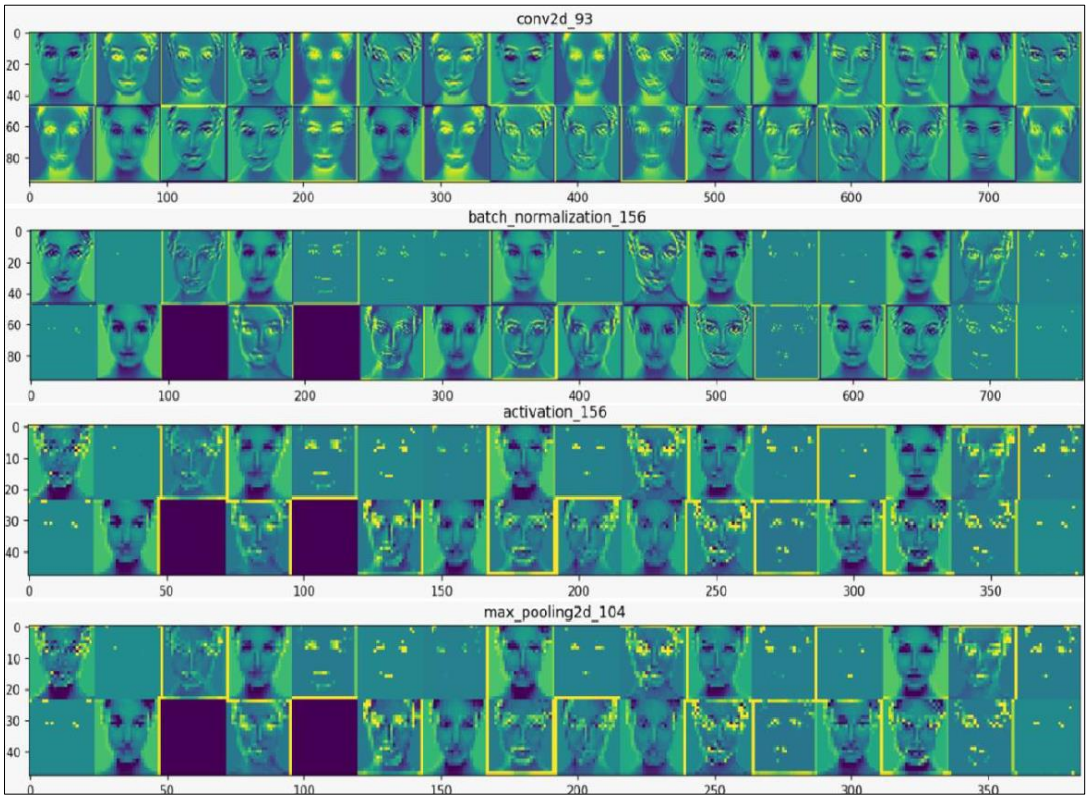

Fig 5: Classifying images based on facial heat maps

Neural networks' power and discrimination are applied for this purpose. A feature map, or the average representation of each layer beyond the first, is the model's output. Load the input images, and use the feature map that is produced to identify the primary features for image classification. Applying filters or feature detectors to the input image or feature map output of the preceding layer produces feature maps. You can see how each convolutional layer in the model represents its unique features by utilizing the feature map view.

5.2.3 Emotion Detection

Filters or detectors are used in convolutional neural network (CNN) architecture to access the image and obtain a map, which is then used to apply the Relu activation function. It is possible to identify edges, curves, vertical and horizontal lines, etc. using feature detectors or filters, found in images. It helps to identify various features such as Then integration is applied to the specific text to achieve seamless translation. Po oling makes predictions based on the assumption that the pool ed results will not change when the input is replaced with a touch number.

But maximum pooling yields superior results than minimum and average pooling. Flattening all inputs and feeding the flattened inputs into the deep neural network, which is the out put of the product category.

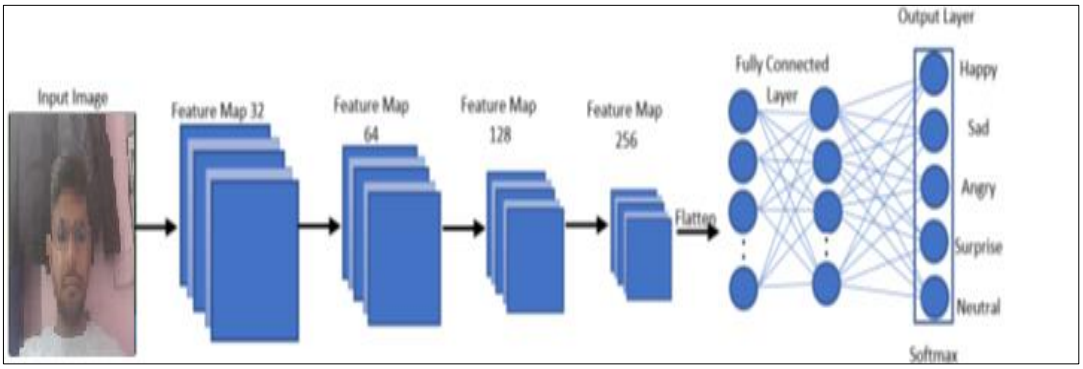

Fig 6: CNN model view

Figure 7 explains the different steps that are opted while extracting various facial features from any image of 48x48pixels. The major focus areas include eyebrows, nose, and lips postures. Multiple layers are overlapped to get to the final prediction value.

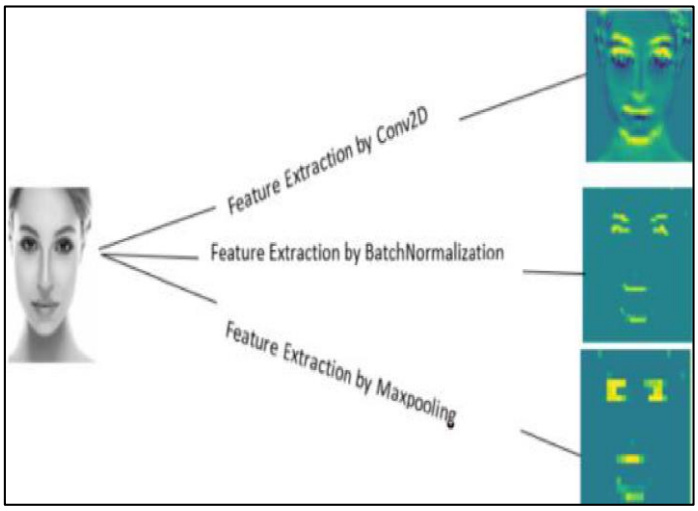

Fig 7: Feature description

Display group can be double or multiple groups to identify nu mbers or separate different clothes. The neural network is like a blackbox and the learned properties in the neural networks are not defined. Basically, system took the input image, and th e CNN model gave results. The search hypothesis is impleme nted by loading the model trained with CNN weights. When p rovided with a live image of an enduser, that image is first conveyed to a CNN model that predicts emotions and add s tags to the image.

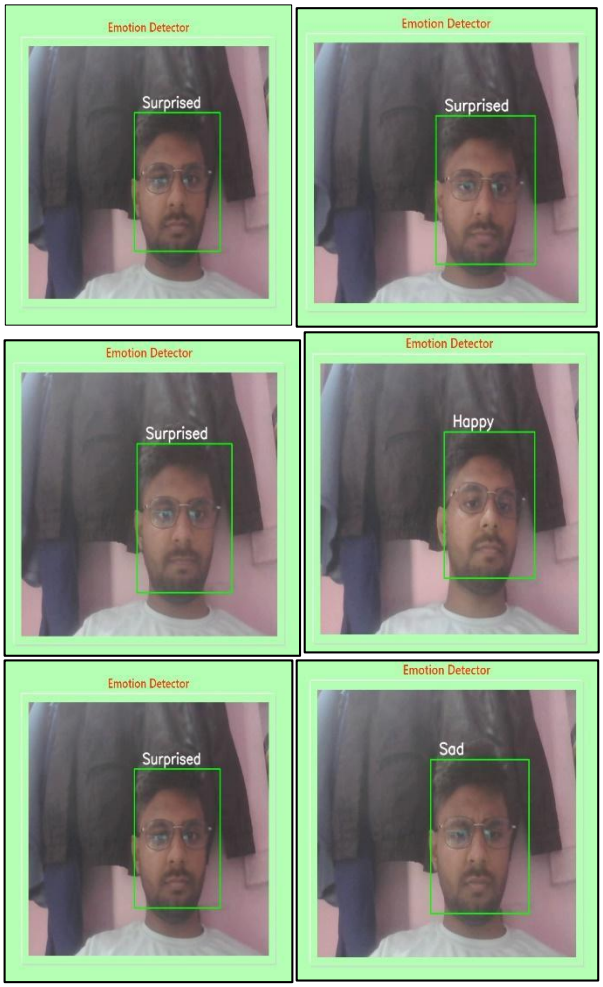

Fig 8: Testing various samples to improve accuracy

5.3 Music Recommendation Module

5.3.1 Songs Database

Thesystem creates a database for Bollywood/Hollywood sons. Each csv has 100 to 150 songs. It is evident that music improvises our mood. Therefore, if the user is feeling sad, the model will suggest playlists to support him and cheer him up from the automatic mood.

5.3.2 Music Playlist Recommendation

Real-time analysis of the user's mood using the mood module. This will provide text like Happy, Sad, Angry, Surprised and Neutral. The system associated directory with the folders of the song database that is generated using the os.listdir() method in Python. This way listdir() is utilised to get a list of files of a directory.

```
auth_manager = SpotifyClientCredentials('537f836f3f1547c0b2f468ab7590ac09','57db1d304f53432ab0efd141f15632cb')
sp = spotipy.Spotify(auth_manager=auth_manager)

def getTrackIDs(user, playlist_id):
    track_ids = []
    playlist = sp.user_playlist(user, playlist_id)
    for item in playlist['tracks']['items']:
        track = item['track']
        track_ids.append(track['id'])
    return track_ids

def getTrackFeatures(id):
    track_info = sp.track(id)

    name = track_info['name']
    album = track_info['album']['name']
    artist = track_info['album']['artists'][0]['name']
    # release_date = track_info['album']['release_date']
    # length = track_info['duration_ms']
    # popularity = track_info['popularity']

    track_data = [name, album, artist] #, release_date, length, popularity
    return track_data
```

Fig 9: Classification of songs on the basis of emotions

6. Result & Analysis

Following a review of several studies utilizing convolutional neural networks, extreme learning machines, and support vector machines (SVM).

A comparison of related algorithms is presented in Table 2. For each study, corresponding algorithms and efficiency values are provided. Convolutional neural networks, or CNNs, are used to increase the accuracy and efficiency of emotion detection.

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

Algorithm	SVM	ELM	CNN
Validation Accuracy	0.660	0.620	0.950
Testing Accuracy	0.669	0.634	0.715

7. Conclusion

A comprehensive overview of the literature shows that there are a number of ways to use visual music. The process was provided by previous researchers and the developers were trained. According to the findings, our system's target was adjusted. As the strength and quality of AI applications continue to increase, our operations will be cutting-edge using modern technology. This system provided information on how music affects an end-user's mood and how to select the right music to lift that mood.

Facial emotions can be recorded by a useful system. The model is capable of capturing the following emotions: surprised, happy, sad, angry, or neutral. The model tracks the end-user's mood and then gets ready to show the user a list of songs that go with the desired appearance. Processing large files requires a lot of CPU and memory. This will make the development process more difficult.

The goal is to implement this application on standard hardware using the simplest and least expensive method possible. Our facial recognition-based music recommendations will reduce users' workload of designing and managing playlists.

8. Future Scope

While the model is running, there is still room for future impr ovements. Many aspects of the app can be customized to prov ide users with better results and a better experience. Some of t hese are alternatives based on other emotions that do not invol ve our body, such as hatred and fear. This consideration inclu des support for music playback. The future opportunity for thi s technique will be to develop techniques that will assist in m usic therapies and help music therapists treat patients with stre ss, anxiety, depression pressure and injury.

There is a chance to quickly add some extra functionality as a solution because the current system struggles in dim lighting and unfavorable camera conditions.

References

1. AlDeeb AH. Emotion-Based Music Player Emotion Detection from Live Camera. ResearchGate. June 2019. Available from: [link]
2. Ayata D, Yaslan Y, Kamasak ME. Emotion-based music recommendation system using wearable physiological sensors. IEEE Trans Consumer Electron. 2018 May;14(8):[page range]. DOI: 10.1109/TCE.2018.2844736.
3. Norden F, von Reis Marlevi F. A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition. TRITA-EECS-EX-2019:143.
4. Kaushik P, Yadav R. Reliability design protocol and block chain locating technique for mobile agent. J Adv Sci Technol. 2017;14(1):136-141. DOI: 10.29070/JAST.
5. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(6):590-595. DOI: 10.29070/JASRAE.
6. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(6):606-611. DOI: 10.29070/JASRAE.
7. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(6):590-595. DOI: 10.29070/JASRAE.
8. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing. J Adv Scholarly Res Allied Educ [JASRAE]. 2018;15(1):1439-1442. DOI: 10.29070/JASRAE.
9. Preema JS, Rajashree, Sahana M, Savitri H. Review on facial expression-based music player. Int J Eng Res Technol (IJERT). 2018;6(15). ISSN-2278-0181.
10. Ramanathan R, Kumaran R, Rohan RR, Gupta R, Prabhu V. An intelligent music player based on emotion recognition. 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions; 2017. DOI: 10.1109/CSITSS.2017.8447743.
11. Research Prediction Competition. Challenges in representation learning: facial expression recognition challenges. Learn facial expression from an image. Kaggle.
12. Gilda S, Zafar H, Soni C, Waghurdekar K. Smart music player integrating facial emotion recognition and music mood recommendation. Pune Institute of Computer Technology, Pune, India. IEEE. 2017. DOI: 10.1109/WiSPNET.2017.8299738.
13. Tabora V. Face detection using OpenCV with Haar Cascade Classifiers. Becominghuman.ai. 2019.
14. Qin Z, Yu F, Liu C, Chen X. How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. Math Found Comput. 2018 May.