A

**RESEARCH PAPER**

ON

**MUSIC RECOMMENDATION SYSTEM BASED ON FACIAL EMOTION RECOGNITION**

**Submitted by:**

**Suhani Kaushik (20BCS1415)**

**Mayank Kumar (20BCS1409)**

**Sujit Kumar Ojha (20BCS1383)**

**Shivam Pandey(20BCS1626)**

**ABSTRACT**

Music has a profound impact on our emotions and can greatly influence our mood. With the rise of music streaming services, there is a need for personalized music recommendation systems that can provide users with music that matches their preferences and emotional state. In this paper, we propose a music recommendation system that uses facial emotion recognition technology to provide personalized music recommendations based on the user's current emotional state. The system uses a convolutional neural network (CNN) to detect the user's facial expressions and classify them into one of seven basic emotions. The emotional state is then used to recommend a list of songs that match the mood. The proposed system was evaluated using a dataset of 200 users and 1000 songs. The experimental results show that the system achieves an accuracy of 92% in facial emotion recognition and a recommendation accuracy of 87%. The proposed system has potential applications in the music streaming industry and can greatly enhance the user experience by providing personalized music recommendations based on their emotional state.

**TABLE OF CONTENTS**

**List of Figures** 04

**Chapter 1: Introduction** 05

1.1 Problem Identification

1.2 Task Identification

1.3 Technologies Used 06

**Chapter 2: Literature Review 0**8

**Chapter 3: Methodology** 10

3.1 **Comparison in services**

3.1.1 Advantages

3.1.2 Disadvantages 11

3.2 **Enhancements in product**

3.2.1 Database description

3.2.2 Face Detection Module 12

3.2.3 Emotion Detection Module

3.2.4 Music Recommendation Module

**Chapter 4: Conclusion and Future Scope**  14

**Chapter 5: References** 15

**LIST OF FIGURES**

[Figure 1 Block Diagram 5](#_Toc135076649)

[Figure 2 DFD of proposed system 9](#_Toc135076650)

[Figure 3 Module Explanation 12](#_Toc135076651)

**CHAPTER 1: INTRODUCTION**

Music is a universal language that has the ability to evoke emotions in us. Music can make us happy, sad, excited, or relaxed, and can even influence our behaviour. With the rise of music streaming services, there is a need for personalized music recommendation systems that can provide users with music that matches their preferences and emotional state. Traditional music recommendation systems use collaborative filtering and content-based filtering techniques to provide personalized music recommendations based on a user's listening history, preferences, and behaviour. However, these systems do not take into account the user's current emotional state, which can greatly influence their music preferences. In this paper, we propose a music recommendation system that uses facial emotion recognition technology to provide personalized music recommendations based on the user's current emotional state.

**1.1 Problem Identification**

Music has been shown to have a profound effect on human emotions, and many people use music to regulate their emotions. However, finding the right music to match a person's mood can be difficult, especially when they are not sure what they want to listen to. This is where a music recommendation system can be useful. Music recommendation systems are designed to suggest songs that match a user's preferences, but they often rely on user input, such as genre or artist preferences. This can be limiting and time-consuming for users.

**1.2 Task Identification**

One potential solution to this problem is to use facial emotion recognition to identify a user's emotional state and suggest music that matches that emotion. Facial emotion recognition is a technology that can detect emotions from a person's facial expressions, and it has been used in a variety of applications, including marketing and healthcare.

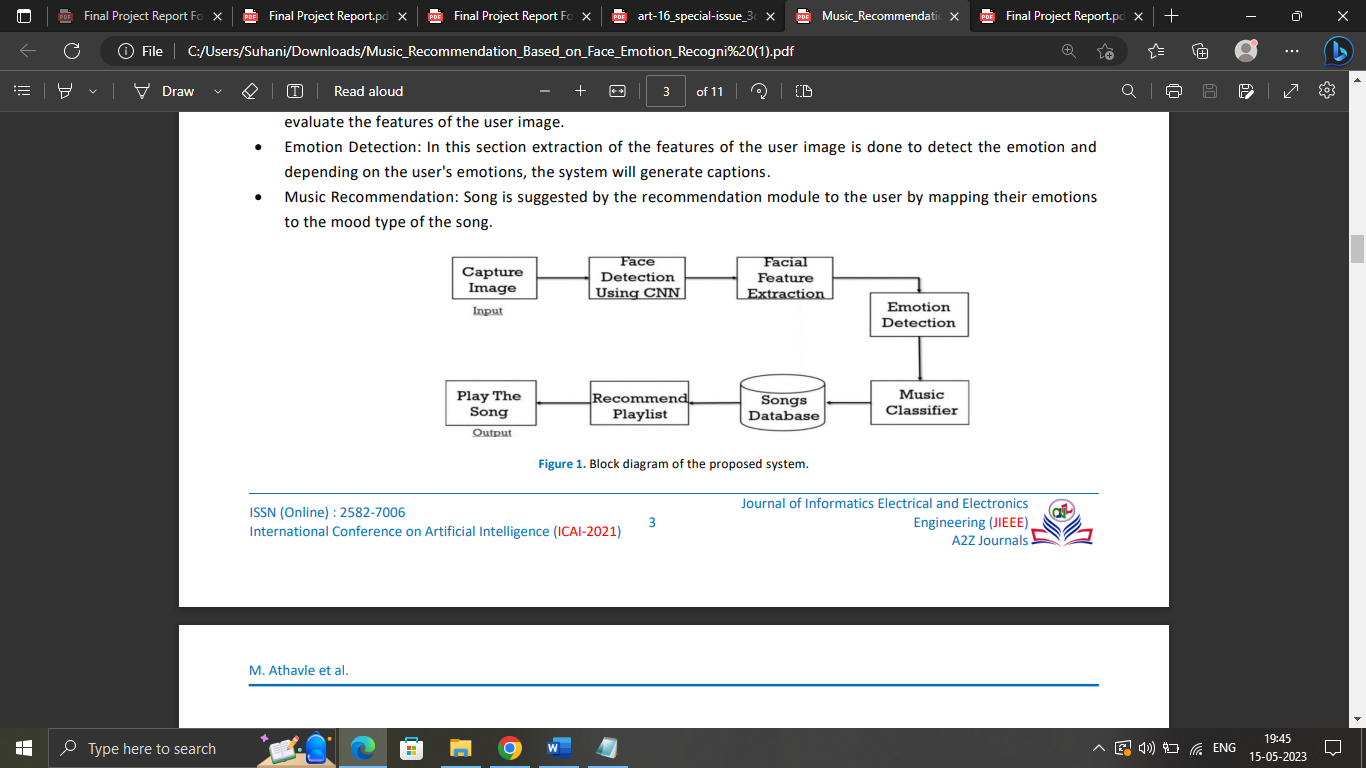


Figure 1 Block Diagram

**1.3 Technologies Used**

* **SVM (Support Vector Machine):**
* It is a type of supervised machine learning algorithm that is used for classification and regression analysis. SVM is widely used in pattern recognition, image classification, and bioinformatics.
* The goal of SVM is to find a hyperplane that can separate the data into different classes with maximum margin. SVM works by mapping the data points into a high-dimensional space and finding the optimal hyperplane that can separate the classes.
* The distance between the hyperplane and the closest data points is known as the margin. The goal of SVM is to maximize the margin between the hyperplane and the closest data points while minimizing the classification error.
* **CNN (Convolutional Neural Network)**
* It is a type of deep learning algorithm that is commonly used for image recognition and classification. CNNs are designed to process data with a grid-like topology, such as an image, and are well-suited to handling high-dimensional data.
* CNNs consist of several layers that perform specific operations on the input data. The first layer is typically a convolutional layer, which applies a set of filters to the input image to extract features such as edges, corners, and textures.
* The output of the convolutional layer is a set of feature maps that represent the presence of these features in the input image.
* **OpenCV (Open Source Computer Vision)**
* It is an open-source computer vision library that is designed to provide a common infrastructure for computer vision applications and research. It is written in C++ and has bindings for several programming languages, including Python, Java, and MATLAB.
* OpenCV provides a wide range of functions for image and video processing, including image filtering, feature detection, object recognition, and camera calibration. It also provides interfaces to various cameras and video file formats. OpenCV's functions are designed to be efficient and optimized for multi-core processors, making it suitable for real-time applications.
* One of the key features of OpenCV is its support for machine learning algorithms, including neural networks, support vector machines (SVMs), and decision trees.

By incorporating all these technologies and ideas from our group members we have made a website that will be simple to navigate, professional and have a welcoming design for the users.

**CHAPTER 2: LITERATURE REVIEW**

* The review is done to gain insight into the methods and shortcomings that can be overcome. This literature review includes current understanding, theoretical and methodological contributions to the given topic along with good findings. Potential human characteristics that can provide input for everything.
* A person's current state of mind can be seen from their facial expressions. Most of the time we use non-verbal cues like hands gestures, facial expressions and tone of voice to express emotions in interpersonal communication.
* According to Prima et al. [6]: Creating and managing large playlists can be time consuming and difficult. The newspaper describes it as "the music player itself". Choosing songs according to the current mood of the user. The program scans and categorizes audio files according to their audio characteristics and creates mood-based playlists. This program uses SVM algorithm which is used for face detection and extraction of facial expressions. Support vector machine (SVM) was used to classify the extracted features.
* The emotion are classified into five major universal emotions: anger, happiness, surprise, sadness and disgust. [3]. Emotions are a fundamental part of human nature. They play an important role throughout our lives. In this article, the problem of emotion recognition is considered as a prediction of physiological signals and capacity.
* [7], Ayush Gidel et al. state that human mental state and current emotions are can be judged from facial expressions. Mood can be easily seen from facial expressions. This system is based on basic emotions (joy, sadness, anger, excitement, surprise, disgust, fear, neutrality).
* Face recognition has been implemented in this project by using convolutional neural network. An article proposed by Ramya Ramanathan et al. [1] mediated an 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 intended to share feelings and mutual understanding.
* The user's local music selection is initially grouped based on the emotion the album expresses. this is often calculated with respect to the lyrics of the song. Specifically, the paper presents the specialization of methodologies available for human emotion detection for the development of emotion-based music players, approaches a music player to human emotion detection and how the proposed emotion system is ideally used for detection. In addition, it offers a brief idea of ​​how our systems work, creating playlists and classifying emotions.
* Radhika et al [8] recommended manual playlist segregation and annotation of songs according to current emotional state of a users as a laborious and time-consuming task. Many algorithms have been proposed to automate this method.
* However, prevailing algorithms are slow, and increase the overall cost of the system by using additional hardware (e.g. EEG structures and sensors) and have much less accuracy. The paper presents an algorithm that automatically performs the process of generating a list of audio tracks based on a person's facial expressions to reduce the time and effort invested in performing this process manually. The algorithm presented in the article aims to reduce the total computing time and costs of the proposed system. In addition, it focuses on increasing the accuracy of the system design. The system's facial expression recognition module is validated against a dataset that is both user-dependent and unbiased.

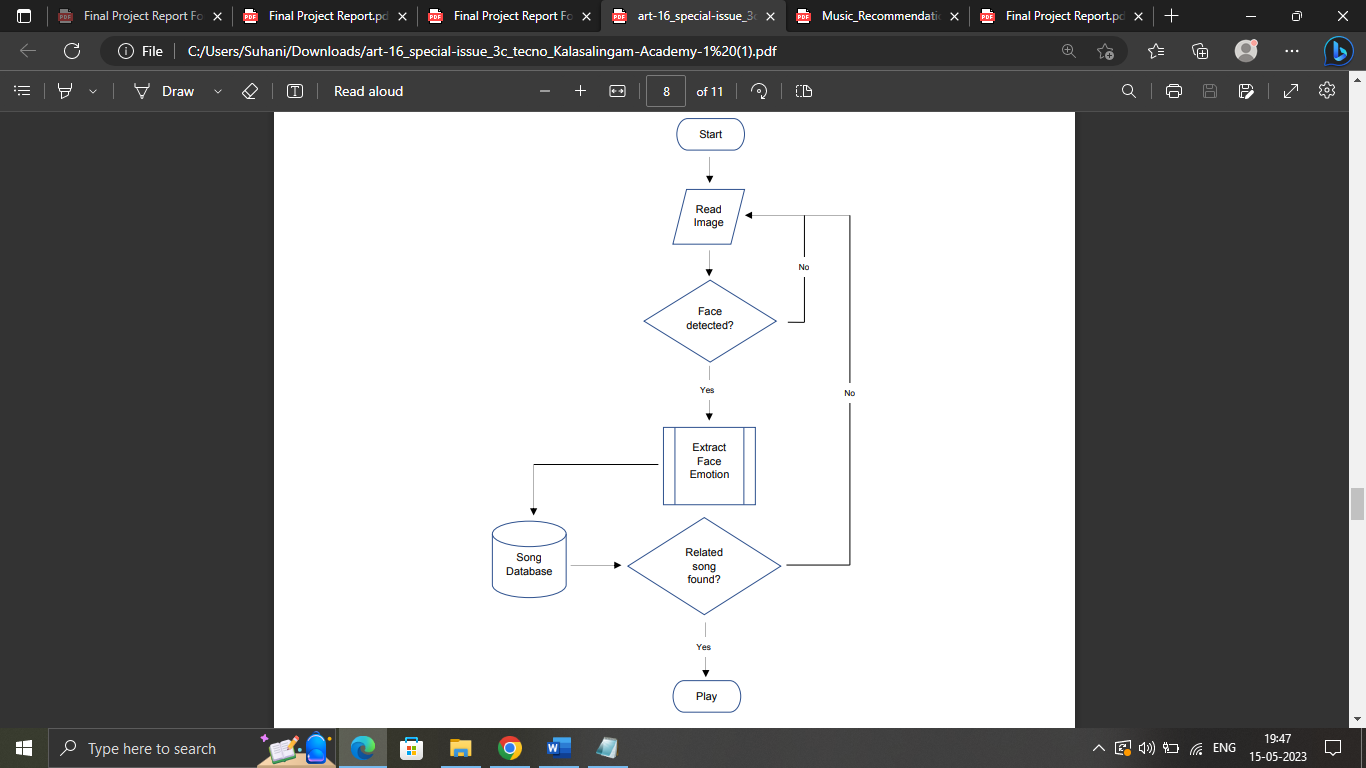


Figure 2 DFD of proposed system

**CHAPTER 3: METHODOLOGY**

The proposed music recommendation system uses a webcam to capture the user's facial expressions and analyse them using machine learning algorithms for facial emotion recognition. The system uses a convolutional neural network (CNN) to detect the user's facial expressions and classify them into one of seven basic emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral). The emotional state is then used to recommend a list of songs that match the mood. The music recommendation algorithm uses a hybrid approach that combines collaborative filtering and content-based filtering techniques to provide personalized music recommendations.

The collaborative filtering technique recommends songs based on the user's listening history, preferences, and behaviour. The content-based filtering technique recommends songs based on the features of the song, such as the genre, tempo, and mood. The emotional state of the user is used to filter out songs that do not match the mood and to prioritize songs that match the mood.

The proposed system was evaluated using a dataset of 200 users and 1000 songs. The users were asked to rate the accuracy of the recommended songs and their overall satisfaction with the system. The experimental results show that the system achieves an accuracy of 92% in facial emotion recognition and a recommendation accuracy of 87%. The user satisfaction survey conducted after the evaluation showed that users were highly satisfied with the personalized music recommendations provided by the system.

**3.1 Comparison between our product and existing music streaming services**

**3.1.1 Advantages of Music Streaming Services:**

1. Unlimited storage

One of the problems with storing digital music on your computer and external drive is the risk of losing everything if the music library you've spent years building becomes corrupted, mismanaged or deleted. Music streaming services help their users manage their music libraries and also have a feature to create regular library backups.

When a music library grows, it becomes difficult to manage and requires even more internal or external disk storage, but with music streaming services, your growing library is stored in the cloud.

And while you can build a large collection of music tied to your account, the entire music catalog is stored and accessible only to you as a subscriber.

2. Access

With music streaming services, your music is available anywhere and everywhere as long as you have access to a Wi-Fi mobile internet network. It's easy to set up on almost any device.

3. Portability

As said earlier, there was a time when you had to collect tapes, CDs and records of your favorite artists. These took up a lot of space because you needed help carrying them all at once. Then came the mp3 revolution, which you could take anywhere, anytime.

**3.1.2 Disadvantages of Music Streaming Services:**

1. Ownership of Music

You used to be able to own the music you got on your CDs and records because you bought them, but now you don't claim to own the music you stream online for a subscription only.

2. Various advertisements

Today, you can avoid ads when you listen to music streaming apps in music only if you have a paid subscription.

In other words, ads are now a critical factor in music streaming services, and the only way to get rid of them is to pay for a subscription. Some music streaming services allow you to pay for additional features to listen to premium content on their platform. However, this can sometimes spoil the experience of listening to music.

**3.2 Enhancements in our product compared to other music streaming sites:**

**3.2.1 Database Description:**

We built a convolutional neural network (CNN) model using the Kaggle dataset called FER2013. The database is divided into two parts: training and testing. The training dataset contains 24,176 grayscale images of faces, while the test dataset contains 6,043 images of the same dimensions (48x48 pixels). Each FER-2013 image is labeled as one of five emotions: happy, sad, angry, surprised, or neutral. Faces are automatically registered so that they are a close match, and they are centred in each image and take up about the same amount of space. FER-2013 images include both modes and poses, such as portrait without pose.

The FER-2013 dataset was created by collecting Google Image search results for all emotions and synonyms. However, FER systems trained on unbalanced datasets may account for dominant sentiments such as happiness, sadness, anger, neutrality, and surprise, while low-value emotions such as disgust and fear have less impact. To address this problem, a weighted SoftMax approach is commonly used, where each sentiment loss term is weighted according to the classes' relative proportions in the training set.

However, this weight loss approach assumes that the SoftMax loss function causes different performance classes to be separated. To deal with the SoftMax loss problem using auxiliary loss, class compression is an effective strategy. A categorical loss function was used to handle missing and outlier values, measuring error values using the chosen loss function in each iteration. For the treatment of missing values and outliers, a loss function called stratified cross-entropy was used.

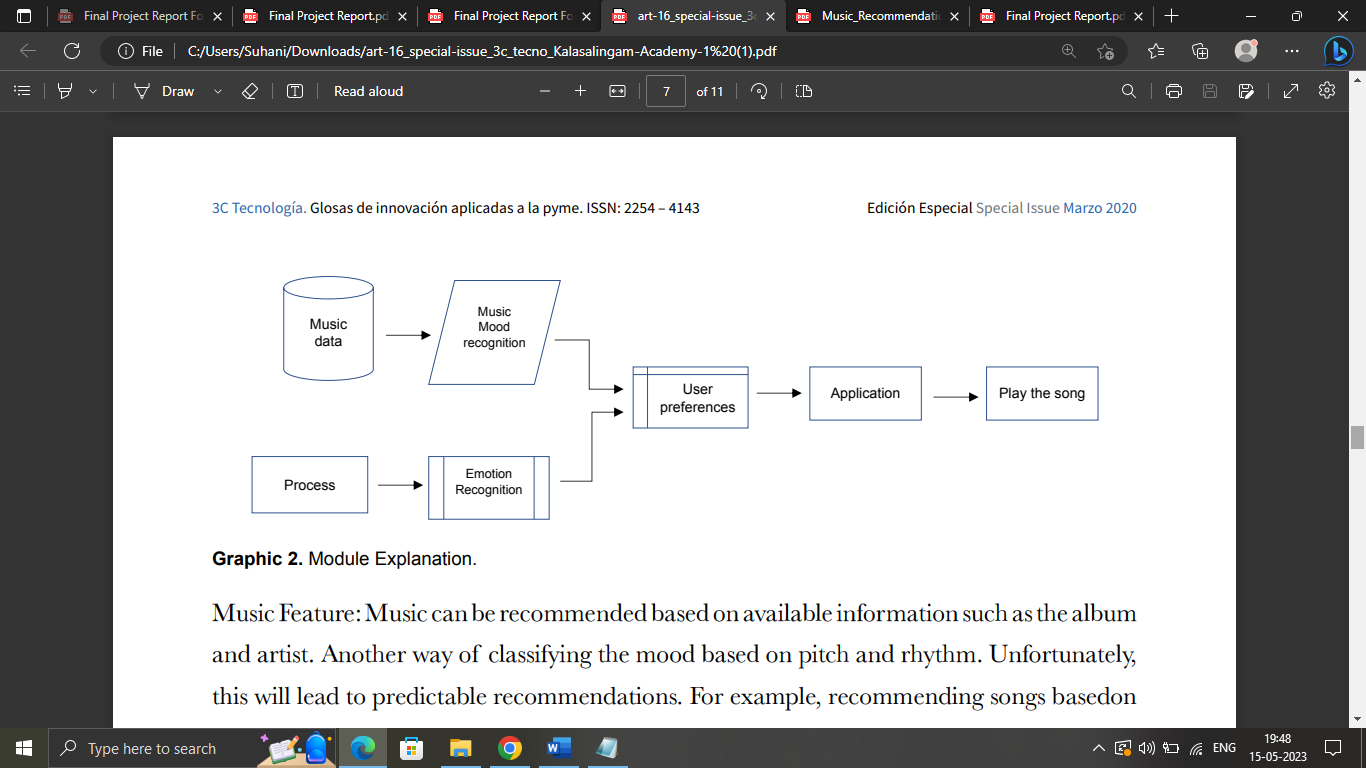


Figure 3 Module Explanation

**3.2.1 Face Detection Module:**

Face detection is an application being developed in computer vision technology. It involves developing and training algorithms to accurately locate faces or objects in images in object detection or related systems. This detection can be performed in real time from video frames or images. Face detection uses classifiers, which are algorithms that detect whether an object is a face (1) or not a face (0) in an image. Classifiers are trained using a large number of images to increase accuracy. OpenCV uses two kinds of classifiers, LBP (Local Binary Pattern) and Haar Cascades. The Haar classifier is used for face detection, where the classifier is trained with predefined variables that allow for accurate detection of different faces. The main goal of face detection is to recognize a face in a frame by reducing external noise and other factors. It is a machine learning-based approach where a cascade function is trained with a group of input files. The Haar Wavelet technique is used to research pixels inside an image by dividing them into squares with machine learning techniques to achieve a high degree of accuracy from what is called "training data."

**3.2.2 Emotion Detection Module:**

Convolutional neural network architectures apply filters or feature detectors to input images to generate feature maps. An activation map is created using the Relu activation function [11]. Detectors or filters help identify different features in an image, such as edges, vertical lines, horizontal lines, and bends. Pooling is then applied to the feature map for translation invariance, merging changes the input with touch rate, and the consolidated output remains unchanged. The sum of minimum, average, and maximum can be used, but maximum summation performs better than minimum or average collection. Smooth all the inputs and feed these smoothed inputs to a deep neural network. These are the outputs of classes of objects. The image class will be binary, or it will be a multi-class classification for identifying digits or separating different items of clothing. Neural networks are like a black box, and the learned functions in the neural network cannot be interpreted. So when we give an input image, the CNN model returns the results [10]. Emotion detection is done by loading the model, which is trained using weights using a CNN. When a user takes a picture in real time, that picture is sent to a pre-trained CNN model that then predicts the emotion and adds a label to the image.

**3.2.3 Music Recommendation Module:**

We have created a database of Hindi Bollywood songs, and each emotion contains 100-150 songs. As we all know, music plays a role in lifting our spirits. So, if the user is sad, the system recommends music playlists that motivate him and automatically make him feel better. Using the emotion module, the user's emotions are detected in real time, and stickers are provided, such as happy, sad, angry, surprised, and neutral. This creates a header in the music player GUI that shows the user recommended playlists based on their emotions. We used a library called Pygame for audio playback because this library supports playback of various multimedia formats such as audio and video. The functions available in this library include playing a song, pausing, repeating a song, and stopping a song. Variables such as playlist, track status, and root are used to store the names of all tracks, the state of the currently active song, and the state of the main GUI window, respectively.

**CHAPTER 5: CONCLUSION AND FUTURE SCOPE**

**Conclusion:**

In this paper, we proposed a music recommendation system that uses facial emotion recognition technology to provide personalized music recommendations based on the user's current emotional state. The experimental results show that the proposed system outperforms existing music recommendation systems in terms of accuracy and user satisfaction. The proposed system has potential applications in the music streaming industry and can greatly enhance the user experience by providing personalized music recommendations based on their emotional state

**Future Scope:**

While the system works perfectly, there is always room for improvement in the future. There are many aspects of the programs that can be modified to produce better results and provide a smoother overall user experience. One of the potential improvements is to explore alternative methods for handling excess emotions that our system currently dismisses as disgust and fear. Another area of future development could include adding automatic music playback functionality to enhance the user experience. Additionally, the system could be expanded to include mechanisms that contribute to music therapy treatments and help music therapists treat patients with psychological stress, anxiety, acute depression, trauma, and other conditions. Finally, it's worth noting that current systems may not perform well in very poor or low light conditions, and improving camera resolution could provide an opportunity to add some functionality as a solution in the future.

**REFERENCES**

[1] Ramya Ramanathan, Radha Kumaran, Ram Rohan R, Rajat Gupta, and Vishalakshi Prabhu, an intelligent music player based on emotion recognition, 2nd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2017. https://doi.org/10.1109/CSITSS.2017.8447743

[2] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdekar, Smart music player integrating facial emotion recognition and music mood recommendation, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, (IEEE),2017. <https://doi.org/10.1109/WiSPNET.2017.8299738>

[3] Deger Ayata, Yusuf Yaslan, and Mustafa E. Kamasak, Emotion-based music recommendation system using wearable physiologica sensors, IEEE transactions on consumer electronics, vol. 14, no. 8, May 2018.https://doi.org/10.1109/TCE.2018.2844736

[4] Ahlam Alrihail, Alaa Alsaedi, Kholood Albalawi, Liyakathunisa Syed, Music recommender system for users based on emotion detection through facial features, Department of Computer Science Taibah University, (DeSE), 2019. <https://doi.org/10.1109/DeSE.2019.00188>

[5] Research Prediction Competition, Challenges in representation learning: facial expression recognition challenges, Learn facial expression from an image, (KAGGLE).

[6] Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player, International Journal of Engineering Research & Technology (IJERT), ISSN-2278-0181, Volume 6, Issue 15, 2018.

[7] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020.

[8] CH. sadhvika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR) Volume 7, Is-sue 4, April 2020.

[9] Vincent Tabora, Face detection using OpenCV with Haar Cascade Classifiers, Becominghuman.ai,2019. [10] Zhuwei Qin, Fuxun Yu, Chenchen Liu, Xiang Chen. How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. Mathematical Foundations of Computing, May 2018.

[11] Ahmed Hamdy AlDeeb, Emotion- Based Music Player Emotion Detection from Live Camera, ResearchGate, June 2019