e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

MUSIC THERAPY USING FACIAL RECOGNITION

Satyam Shivam[*1], Brij Raj Singh[*2], Rimjhim Bhagat[*3], Pranita Shah[*4]

[*1,2,3,4]Department Of Information Technology, Sinhgad College Of Engineering Pune, India.

DOI : https://www.doi.org/10.56726/IRJMETS34347

ABSTRACT

Music plays a very important role in human daily life. Everyone wants to listen to music to their individual taste, mostly based on their mood. Users always face the task of manually browsing music and creating a playlist based on their current mood. The proposed project is very efficient and generates a music playlist based on the current mood of users. Facial expressions are the best way of expressing the ongoing mood of the person. The objective of this project is to suggest songs for users based on their moods by capturing facial expressions. Facial expressions are captured through webcam and such expressions are fed into a learning algorithm that gives the most probable emotion. Once the emotion is recognized, the system suggests a playlist for that emotion, thus saving a lot of time for a user.

Facial recognition is a system built to identify a person's face from an image or video. This technology has been around for decades, but its usage has become more noticeable, and accessible, in the past few years as it now powers innovative solutions, such as personal photo applications and secondary authentication for mobile devices.

The emotion Recognition System analyzes an individual's facial expression. For example, if the corners of a person's mouth are raised, the machine might rule that the person is in a good mood, whereas a wrinkled nose suggests anger or disgust.

Once the emotion is detected by CNN then the emotion is used by Spotify API and then the Spotify API generates a playlist according to the emotion of the user.

Keywords: Face Detection, Emotion Recognition, Webcam, CNN Classification, Spotify API, Music Playlist.

I. INTRODUCTION

Music plays an important role in our daily life. Users have to face the task of manually browsing the music. Computer vision is a field of study that encompasses how computers see and understand digital images and videos. Computer vision involves seeing or sensing a visual stimulus, making sense of what it has seen, and also extract complex information that could be used for other machine-learning activities. We will implement our use case using the Haar Cascade classifier. Haar Cascade classifier is an effective object detection approach that was proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. This project recognizes the facial expressions of users and plays songs according to emotion. Facial expressions are the best way of expressing the mood of a person. The facial expressions are captured using a webcam and face detection is done by using the Haar cascade classifier. The captured image is input to CNN which learns features and these features are analyzed to determine the current emotion of the user then the music will be played according to the emotion. In this project, five emotions are considered for classification which includes happy, sad, anger, surprise, and neutral. This project consists of 4 modules-face detection, feature extraction, emotion detection, and song classification. Face detection is done by the Haar cascade classifier, and feature extraction and emotion detection are done by CNN. Finally, the songs are played according to the emotion recognized. Convolutional Neural Networks (CNN) is a specific type of Artificial Neural Network which are widely used for image classification. CNN is a type of deep learning model for processing data that has a grid pattern, such as images, which is inspired by the organization of animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level patterns. CNN is a mathematical construct that is typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into the final output, such as classification.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

II. LITERATURE REVIEW

1. A Survey of AI-Based Facial Emotion Recognition: Features, ML & DL Techniques, Age-Wise Datasets, and Future Directions Author: CHIRAG DALVI1, MANISH RATHOD1, SHRUTI PATIL 2, SHILPA GITE 2, AND KETAN KOTECHA

Publish Date: 2021

Facial expressions are mirrors of human thoughts and feelings. It provides a wealth of social cues to the viewer, including the focus of attention, intention, motivation, and emotion. It is regarded as a potent tool of silent communication. Analysis of these expressions gives a significantly more profound insight into human behavior. AI-based Facial Expression Recognition (FER) has become one of the crucial research topics in recent years, with applications in dynamic analysis, pattern recognition, interpersonal interaction, mental health monitoring, and many more. However, with the global push towards online platforms due to the Covid-19 pandemic, there has been a pressing need to innovate and offer a new FER analysis framework with the increasing visual data generated by videos and photographs. Furthermore, the emotion-wise facial expressions of kids, adults and senior citizens vary, which must also be considered in the FER research. Lots of research work has been done in this area. However, it lacks a comprehensive overview of the literature that showcases the past work done and provides aligned future directions. In this paper, the authors have provided a comprehensive evaluation of AI-based FER methodologies, including datasets, feature extraction techniques, algorithms, and recent break throughs with their applications in facial expression identification. To the best of the author's knowledge, this is the only review paper stating all aspects of FER for various age brackets and would significantly impact the research community in the coming years.

2. The Role of Music Therapy in the Emotional Regulation and Psychological Stress Relief of Employees in the Workplace Author: Nan

Publish Date: 29 Jan 2022

With the increasing global attention to the problem of staff stress, scholars in the fields of sociology, psychology, and medicine are seeking effective solutions. Music therapy has entered the field of vision of scholars with its unique advantages and is used to maintain the mental health of workers in various industries and improve employee productivity. Regarding the definition of music therapy , we must first understand the two basic elements of music therapy: treatment target and treatment goals; countries around the world have different definitions of music therapy, but they are inseparable from the two basic elements of treatment target and treatment goals. Mayer defined music therapy in a recent literature as follows: "Music therapy is the use of various ways of music experience to help the person being treated achieve the goal of mental or physical health." (e target of music therapy is mainly graduate students who are about to graduate, industry workers, people with autism, and other groups with psychological and physical needs for treatment. Music therapy is not a random or simple process of playing music. (e treatment cycle is divided into shortterm and long-term types . (e treatment process contains a variety of theories and methods. After determining the treatment target , a detailed treatment plan needs to be formulated. Music psychotherapy, as an auxiliary means in psychological counseling, counseling, and activities, is a choice that meets the public's psychological needs and appreciation habits.

3. Review of Facial Expression-Based Music Player

Author name: Hafeez Kabini, Sharik Khan, Omar Khan, Shabana Tadv

Publish Date: 05 Jan 2019

Human often use nonverbal cues such as hand gestures, facial expressions, and tone of the voice to express feelings in interpersonal communications. The face of the human is an important organ of an individual's body and it plays an important role in the extraction of an individual's behavior and emotional state. Facial expression provides the current mind state of a person. It is very time-consuming and difficult to create and manage large playlists and select songs from these playlists. Thus, it would be very helpful if the music player itself selects a song according to the current mood of the user. Manually segregating the list of songs associated, and generating an acceptable playlist that supported an individual's emotions could be a terribly tedious, time overwhelming, intensive, and upheld task Thus, an application can be developed to minimize these efforts of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

managing playlists. However, the proposed existing algorithms in use are computationally slow and less accurate. This proposed system based on facial expression extracted will generate a playlist automatically thereby reducing the effort and time involved in rendering the process manually. Facial expressions are given using an inbuilt camera. The image is captured using a camera and that image is passed through different stages to detect the mood or emotion of the user. We will study how to automatically detect the mood of the user and present him with a playlist of songs that is suitable for his current mood. The proposed paper has used the Viola-Jones algorithm and multiclass SVM (Support Vector Machine) for face detection and emotion detection respectively.

III. METHODOLOGY

A. Objective :

Therapists, psychologists, and related practitioners and theorists of mental health tends to hold one of two broad views about how to help clients. On the one hand, some maintain that, or at least act as though, the basic point of therapy is to help clients become clear about what they want deep down and to enable them to achieve it by overcoming mental blockages. On the other hand, there are those who contend that the aim of therapy should instead be to psychologically enable clients to live objectively good lives, say, ones that involve developing their inherent talents or exhibiting an authentic/integrated/resilient self.

The proper aim of therapy, according to this chapter, is to enable clients to live meaningfully.

B. System Overview

3.1 EXTERNAL INTERFACE REQUIREMENT

3.1.1 User Interface

Music Therapy Using Facial Expression using a deep learning approach

3.1.2 Hardware Interfaces:

RAM: 8GB

As we are using Machine Learning Algorithm and Various High Level

Libraries Laptop

The RAM minimum required is 8 GB.

Hard Disk: 40 GB

Data Set of Scan images is to be used hence minimum 40 GB Hard Disk memory is required.

Processor: Intel i5 Processor

Jupyter Notebook IDE that Integrated Development Environment is to be used and data loading should be fast hence Fast Processor is required IDE: Jupyter

Notebook

Best Integrated Development Environment as it gives possible suggestions at the time of typing code snippets that make typing feasible and fast.

Coding Language: Python Version 3.5

Highly specified Programming Language for Machine Learning because of availability of High-Performance Libraries.

Operating System: Windows 10

Latest Operating System that supports all types of installation and development Environment.

3.1.3 Software Interfaces

Operating System: Windows 10

IDE: Spyder

Programming Language: Python

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

IV. SYSTEM ARCHITECTURE

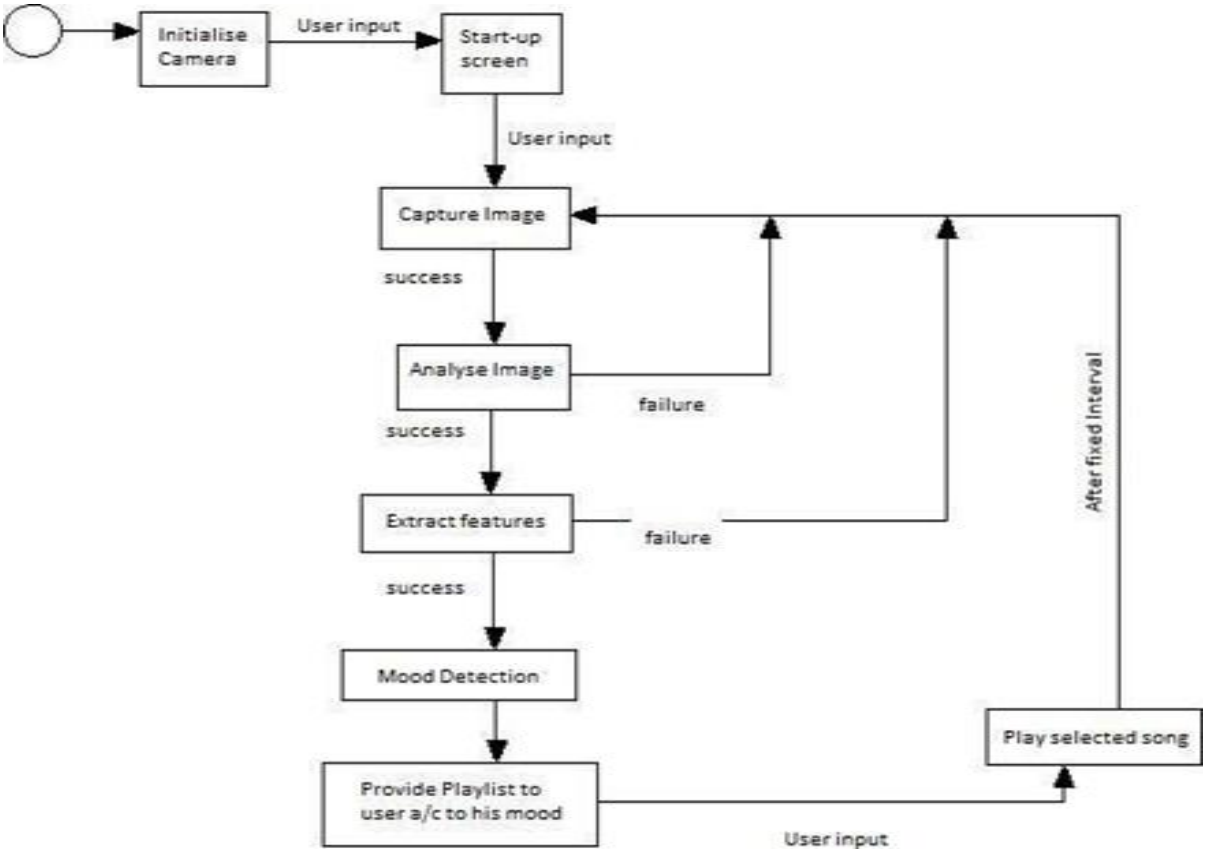

Fig. 4.1: System Architecture

Data Flow Diagram

In the input unit, a preconfigured camera is arranged such that it is interfaced with a display device like a TFT monitor and a switch or a key is attached. Whenever the subject (the car) is in the frame of the picture, the key is pressed in order to capture the picture.

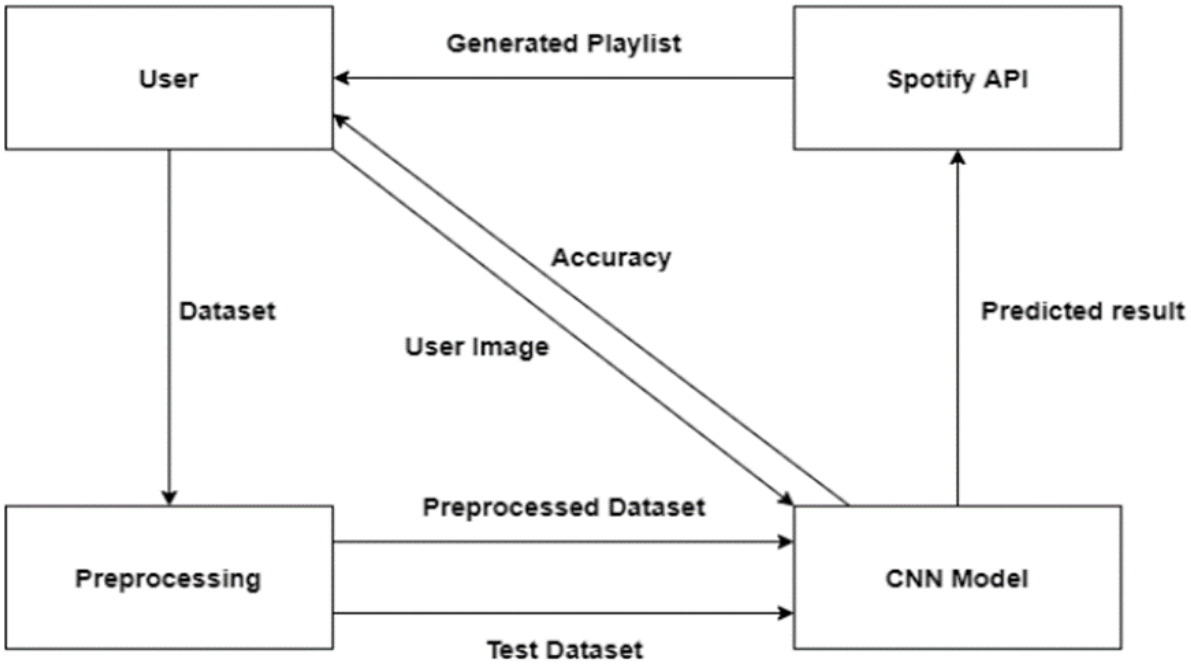

Figure 4.2: Data Flow(0) diagram.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:03/March-2023 Impact Factor- 7.868 www.irjmets.com

In the Output Unit, the input image is taken from the input unit and then processed in the processing unit. In the processing unit all sorts of enhancement and extraction is done and then the number on the license plate is extracted using Optical Character Recognition and then it can either be stored in a database or be displayed on a display device or both or can be used to excite an actuator.

UML DIAGRAMS

Unified Modeling Language is a standard language for writing software blueprints.The UML may be used to visualize,specify,construct and document the artifacts of a soft-wareintensive system.UML is process independent,although optimally it should be used in process that is use case driven,architecture-centric,iterative,and incremen-tal.The Number of UML Diagram is available.

Class Diagram.

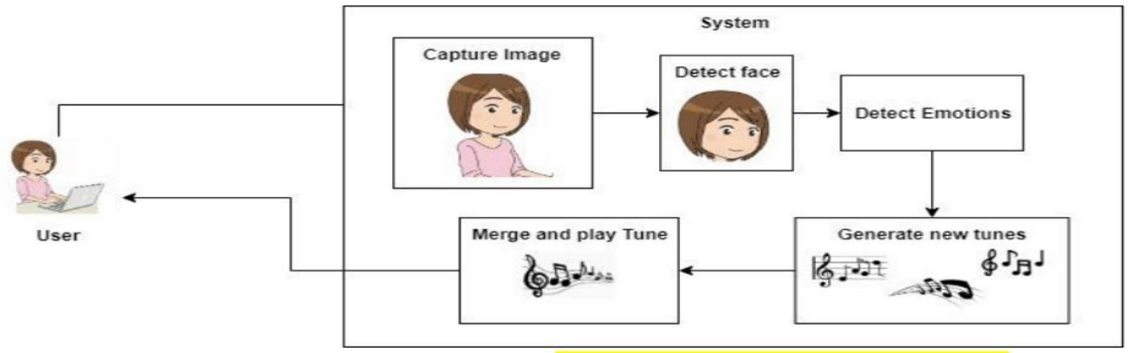

Fig 4.3: Class diagram

V. CONCLUSION

The algorithm proposed in this paper aims to generation music tunes based on facial expressions. Experimental results indicate that the proposed algorithm was successful in automating music tune generation on the basis of facial expressions and hence reduced the amount of labor and time, incurred in performing the task manually. The use of laptop or camera helped in eradicating the requirement of any additional hardware, such as EEG systems and sensors, and thus helped in curtailing the cost involved. Since, face emotion recognition is not performed in real time, the total time taken by the algorithm is equal to the amount of time taken by the algorithm to recognize facial expressions and the amount of time taken by the algorithm to query the meta data file. Hence, the proposed algorithm yields better performance, in terms of computational time, than the algorithms reported in the existing literature. Also, since the time taken by the algorithm to query the Meta data file is negligible (0.0008 sec), the total time taken by the algorithm is proportional to the time taken to recognize facial expressions. Viola Jones algorithm Detect face of the user accurately. Then this gives input to the convolutional neural network and we get emotions.

VI. REFERENCES

[1] M. Spezialetti, G. Placidi, and S. Rossi, "Emotion recognition for humanrobot interaction: Recent advances and future perspectives,"Frontiers Robot. AI, vol. 7, p. 145, Dec. 2020, doi: 10.3389/FROBT.2020.532279.

[2] S. Ramis, J. M. Buades, and F. J. Perales, ‘‘Using a social robot to evaluate facial expressions in the wild,’’ Sensors, vol. 20, no. 23, pp. 1–24, 2020, doi: 10.3390/s20236716.

[3] Y. K. Bhatti, A. Jamil, N. Nida, M. H. Yousaf, S. Viriri, and S. A. Velastin, ‘‘Facial expression recognition of instructor using deep features and extreme learning machine,’’ Comput. Intell. Neurosci., vol. 2021, pp. 1–17,Apr. 2021, doi: 10.1155/2021/5570870.

[4] S. Li and W. Deng, ‘‘Deep facial expression recognition: A survey,’’ IEEE Trans. Affect. Comput., early access, Mar. 17, 2020, doi: 10.1109/TAFFC.2020.2981446

[5] A. Fathima and K. Vaidehi, ‘‘Review on facial expression recognition system using machine learning techniques,’’ in Advances in Decision Sciences, Image Processing, Security and Computer Vision. Cham, Switzerland: Springer, 2020, pp. 608–618, doi: 10.1007/978-3-030- 24318-0_70