JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Musical Therapy Using Facial Emotion Detection

Priyadarshini M [1], Sneha V B [2], Thimmanna [3], Vishal Prajapathi P [4], Veena Kumari S [5]

[1,2]*Assistant Professor Department of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*

[2,3,4,5]*Students Department of Computer Science and Engineering, Cambridge Institute of Technology, Bangalore, India*

Abstract - *The proposed project aims to efficiently generate a music playlist based on the current mood of users by capturing their facial expressions through a webcam. By using a learning algorithm to recognize the most probable emotion from the captured facial image, the system can suggest a playlist that matches the user's mood, saving them time and effort. The detected emotion is then used by the Spotify API to generate a playlist that aligns with the user's emotion.*

***Key Words*: CNN, Detection ,Emotion ,Face Recognition, Music ,Mood, Mental Health**

1.INTRODUCTION

Music plays an important role in our daily lives, but manually browsing for music can be time-consuming. Computer vision, a field of study focused on how computers perceive and understand images and videos, can automate this task. The Haar Cascade classifier, proposed by Paul Viola and Michael Jones in 2001, is an effective approach for object detection. This project utilizes Haar Cascade classifier for face detection to recognize facial expressions of users. A webcam is used to capture facial images, which are then input to a Convolutional Neural Network (CNN) for feature extraction and emotion detection. CNN is a type of deep learning model widely used for image classification. It is inspired by the organization of animal visual cortex and designed to automatically learn spatial hierarchies of features from images. CNN consists of convolution, pooling, and fully connected layers, which perform different functions in the feature extraction process. Convolution layers use a kernel to extract features from the 2D grid of pixel values in an image, making CNNs efficient for image processing as features can occur anywhere in the image. Extracted features are hierarchically combined in subsequent layers to capture more complex patterns.

The CNN analyzes the features to determine the current emotion of the user based on the facial expression according to the recognized emotion. The project consists of four modules: face detection, feature extraction, emotion detection, and song classification. Haar cascade classifier is used for face detection, while CNN is used for feature extraction and emotion detection. The CNN learns features from the input images to identify the user's emotion accurately. Finally, the system classifies and plays songs based on the recognized emotion, providing a personalized music experience. This project showcases the potential of computer vision and CNNs in automating tasks related to music and emotion recognition. The integration of Haarcascade classifier and CNN allows for efficient and accurate detection of facial expressions and emotion recognition for personalized music recommendation.

1.1 OBJECTIVE

The primary objective of using musical therapy with facial emotion detection is to improve the emotional well-being of patients. Music has long been known to have a powerful effect on human emotions, and using technology to analyze a patient's facial expressions while listening to music can help therapists tailor their treatment to the patient's emotional needs. By identifying the patient's emotional state and using music to promote positive emotions, therapists can help patients achieve greater emotional stability and resilience, reducing the risk of developing mental health disorders and improving overall quality of life.

1.2 MOTIVATION

The motivation for using musical therapy with facial emotion detection is to provide a more personalized and effective form of therapy for individuals struggling with mental health issues. Traditional forms of therapy, such as talk therapy, may not be effective for everyone and can take a long time to produce results. By incorporating technology, therapists can better understand a patient's emotional state and tailor the therapy to their individual needs, improving the chances of success. Additionally, the use of music can make therapy more engaging and enjoyable, which can motivate patients

to continue with treatment and improve their mental health. Overall, the motivation for using this technology is to provide a more efficient, effective, and enjoyable form of therapy that can improve the lives of individuals struggling with mental health issues.

1.2 PROBLEM STATEMENT

Despite the potential benefits of using musical therapy with facial emotion detection, there are several challenges that must be addressed before this technology can be widely adopted. One problem is the lack of standardization in facial emotion detection algorithms, which can lead to variability in the interpretation of facial expressions and potentially inaccurate assessments of emotional state. Another challenge is the need for large datasets of emotional responses to music in order to train these algorithms effectively. Finally, there is a need for more research to establish the efficacy of this approach and determine which types of patients are most likely to benefit from this technology. Addressing these challenges will be critical for realizing the potential benefits of musical therapy with facial emotion detection and improving mental health outcomes for individuals struggling with mental health issues.

2. RELATED WORK

There has been significant research on predicting user emotions. Some of which are as follows: 1. Facial expression-based automatic emotion recognition is an intriguing study area that has been presented and used in a number of fields, including safety, health, and human-machine interactions. Researchers in this discipline are interested in creating methods for human machine interfaces, safety, and health. Researchers in this discipline are interested in creating methods to decipher, encode, and extract these characteristics from facial expressions in order to improve computer prediction. Due to deep learning's exceptional success, its various architectures are being utilised to produce greater results.

2. Facial expression recognition (FER) has gained considerable attention due to its potential applications in various fields such as safety, health, and human-machine interactions. architectures to produce better results in FER. Emotion recognition has several potential applications, including software engineering, website personalisation, education, and gaming. This study presents a brief overview of affect recognition techniques that use various inputs such as biometrics, video channels, and behavioural data. The scenarios discussed in this review illustrate the complexity and challenges of deploying emotional computing in different sectors. The analysis of these scenarios leads to some conclusions and highlights the need for further research to address the difficulties with automatic recognition.

3. IMPLEMENTATION AND WORKING

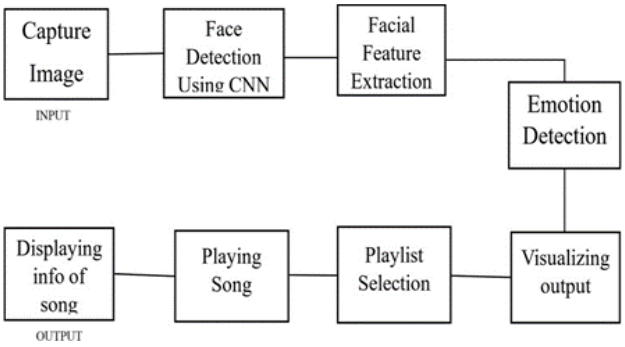


Fig 1: System Architecture

The proposed system aims to provide an interactive music player that detects the user's emotions through facial recognition technology. Images captured by the camera are analyzed by a Convolutional Neural Network to predict the user's emotional state, and a playlist of songs is suggested based on that emotion, such as happy, sad, natural, or surprised. If the emotion detected is negative, the system will present a playlist of songs that can positively enhance the user's mood. Overall, the system seeks to improve the user's music experience by providing personalized song recommendations based on their emotional state Emotion Detection: This module extracts the features of the users image and analyzes them to detect the users emotional state. Based on the users emotions, the system will generate captions. Music Recommendation: The recommendation module suggests songs to the user by mapping their emotions to the mood type of the song. The system will provide the user with a playlist that matches their current emotional state, such as happy, sad, natural, or surprised

4. METHODOLGY

A Convolutional Neural Network (CNN) is a specific type of artificial neural network that excels at recognizing patterns in images, making it a valuable tool for image processing and recognition. However, to achieve high accuracy in its predictions, a CNN requires a vast amount of labeled data for training, and powerful processors such as GPUs or NPUs to produce results quickly. While CNNs are primarily used for visual imagery, they can also be applied to other areas, such as natural language processing, drug discovery, and health risk assessments. Additionally, CNNs have become increasingly important in depth estimation for self-driving cars. Compared to other neural networks, CNNs have shown superior performance in analyzing image, speech, and audio signals. They have three main types of layers, which are:

4.1 Convolutional layer

The convolutional layer is a crucial component of a Convolutional Neural Network (CNN) and is responsible for most of the computation. It requires three main elements: input data, a filter, and a feature map. In the case of a color image, the input will be a 3D matrix of pixels representing the height, width, and

RGB color channels. A feature detector, also called a filter or kernel, will move across the receptive fields of the image, searching for specific features using a convolution operation. Convolutional neural networks (CNNs) apply a Rectified Linear Unit (ReLU) transformation to the feature map after each convolution operation, which introduces nonlinearity to the model. Additionally, multiple convolutional layers can be stacked in a hierarchical manner, where later layers have access to pixels within the receptive fields of prior layers.

4.2 Pooling layer

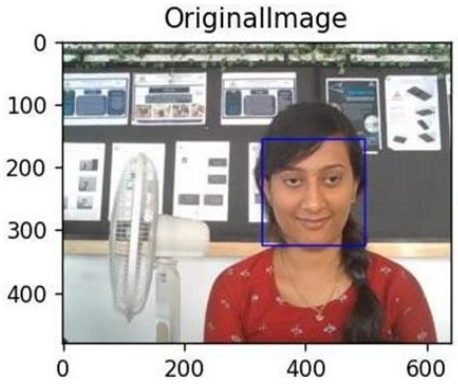
Pooling layers are a type of layer in a neural network that help reduce the dimensionality of the input by reducing the number of parameters. This is done through a process called down sampling. In contrast to the convolutional layer, the pooling operation does not have any weights associated with the filter. Instead, the filter applies an aggregation function to the values within its receptive field and populates the output array. Max pooling and average pooling are common techniques used in convolutional neural networks (CNNs) to reduce the size and complexity of the input data. During max pooling, the filter selects the pixel with the highest value within the receptive field to transmit to the output array. On the other hand, during average pooling, the filter calculates the mean value within the receptive field to send to the output array.

4.3 Fully-connected

The fully-connected layer is so named because every node in the output layer is directly connected to a node in the previous layer, allowing for classification based on features extracted through previous layers and their filters. Unlike partially connected layers, the pixel values of input images are not filtered directly. ReLu functions are commonly used in convolutional and pooling layers, while softmax activation functions are typically used in FC layers to classify inputs by producing a probability between 0 and 1.

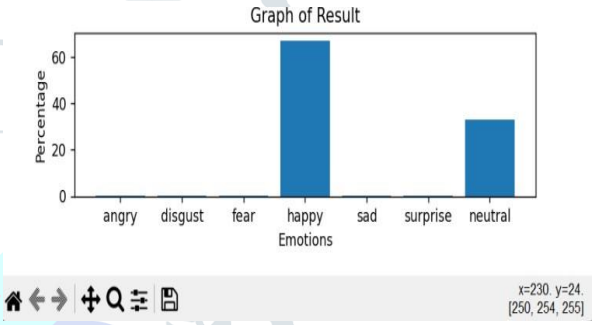
5. EXPERIMENTAL RESULTS

5.1Reading and Saving Image



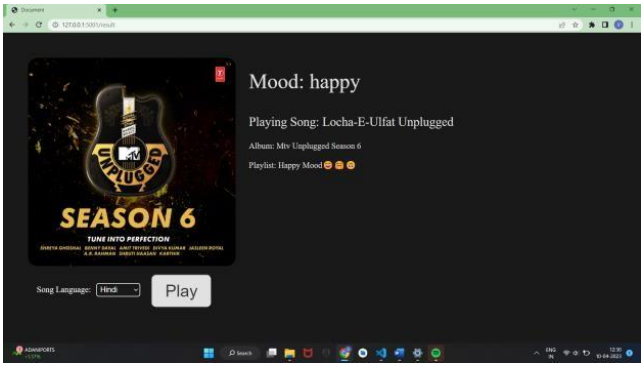
First the image will be captured by the web cam. The web cam will be opened and will be waiting for an image to be detected using opency. After this we will stop the process and the image would be cropped. Essential features will be extracted that is nose, eyes and lips. Then cropped image would be stored in order to send it for emotion detection.

5.2Detecting the Emotion



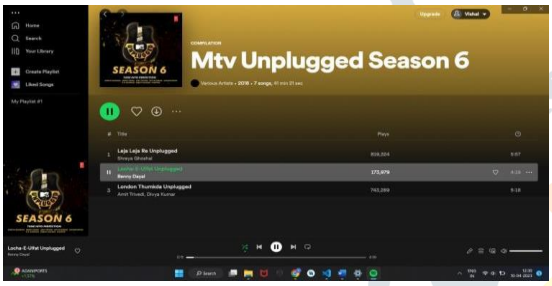
Emotion detection is performed and the image is classsified among the 7 emotions. They are Angry , Happy, Sad, Surprise, Disgust,Fear,Neutral. The output is plotted and represented in the Bar Graph. The X-axis represents the emotions and the Y-axis represents the percentage of confidence for the emotions. In the graph for each and every emotion the percentage of confidence that the image would belong to either one of 7 emotions. Finally which emotion has the highest Percentage that image would be classified as the the final output. As with repect to our example image the percentage for Happy emotion is highest so the final output is Happy

5.3 Choosing Playlist and Playing Song



Based on emotion detected and appropriate playlist is selected For each emotion there is one playlist and the playlist inturn has the appropriate songs. There are random number of songs under each playlist. The user can select his preferrable choice of language. There are 3 languages – Hindi, English, Kannada when the user clicks on the play button the language would be selected and webcam would be open for capturing the image For the above process we have used the spotify api. It is an API which was available from spotify co-operation. From which we can access the already available database according to our requirement.

4)Displaying the Information on the webpage



The right emotion is detected according to highest percentage of confidence and appropriate playlist is selected. The information displayed on the webpage are emotion detected, Song name, Album, Playlist and an image which contains basic details of the song like singer name, lyrics writer etc. The snapshot of the spotify app where the selected song is being played is added for reference.

6 . CONCLUSION

The proposed work presents a holistic review of facial expression recognition using machine learning approaches. The PCA method is used for feature extraction and classification based on Euclidean distance, with high accuracy and consistency even with a small number of training images. The system is designed for musical therapy based on facial emotion recognition, using a webcam and a CNN to detect emotions and play suitable music. The system is user-friendly and intuitive, with an option for further improving the trained model by choosing to retrain it. The online music player is built using the Spotify API, making it robust and sturdy.The proposed system has the potential to improve the work of therapists who use music therapy and provides personalized music therapy based on the user's emotional state.However, the proposed system does not consider age variation factor, which could be a topic for further research. Overall, the proposed system shows promise in providing personalized music therapy and has the potential to be further developed and improved.

5. FUTURE SCOPE

The Emotion Based Music System offers personalized music recommendations based on the user's emotions, saving time and increasing efficiency. As technology continues to advance, the system has the potential to become even more effective at understanding and meeting user needs. Visionify's custom computer vision solutions can help tailor such systems to meet specific requirements. The system is a promising step towards leveraging technology to improve our daily lives, and further development could enhance user experience and satisfaction.

6.REFERENCES

[1] Olufisayo S, Ekundayo, Serestina Viriri “Facial Expression Recognition: A Review of Trends and Techniques” 2021.

[2] W. Mellouk and W. Handouzi, ``Facial emotion recognition using deep learning: Review and insights,'' Proc. Comput. Sci., vol. 175, pp. 689694, Jan. 2020.

[3] A. Koakowska, A. Landowska, M. Szwoch,W. Szwoch, and M. Wróbel, ``Emotion recognition and its applications,'' in Proc. Adv. Intell. Syst. Comput., vol. 300, 2014.

[4] André Teixeira Lopes, Edilson de Aguiar, Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order, 2017.

[5] Richard Jiang, Anthony T.S. Ho, Emotion recognition from scrambled facial images via many graph embedding, 2017.