MUSIC RECOMMENDATION THROUGH FACE RECOGNITION AND EMOTION DETECTION

Manoj sabnis[1], Bhavesh Bhatia[2], Laveena Punjabi[3], Navin Rohra[4]

[1] Professor, Dept. of Information Technology, VESIT, MUMBAI
[234] Student, Dept. of Information Technology, VESIT, MUMBAI

---***---

Abstract - *Human facial expressions convey a lot of information visually rather than articulately. Facial expression recognition plays a very crucial role in the area of human-machine interaction.Face recognition technology has many applications, but they are generally limited to the understanding of human behavior, the detection of mental disorders, and synthesizing human expressions. This project aims to apply various deep learning methods to identify the seven key emotions: happiness, sadness, disgust, anger, fear, surprise, and neutrality, as well as suggest some mood booster songs.*

Keywords- ***Automated, Time-Saving, Face detection, Feature Extraction, Face Recognition, Music-recommendation***

I. INTRODUCTION

An interesting combination of psychology and technology is emotion analytics. Rather reductively, many tools for detecting facial expressions group human emotions into seven basic categories: anger, disgust, fear, happiness, sadness, surprise, and neutral.A playlist is also suggested after analyzing the person's mood. Music plays an important role in coping with stressful situations and triggering emotional responses. Hence, it is important to recommend music that suits the current emotional needs of the user. There are already numerous audio and video recommendation systems like Spotify, Netflix, Gaana, YouTube etc which are based on a search query rather than the emotional needs of the user. If, for instance, a person's emotion is detected as sad, the app will suggest happy songs and other mood songs to cheer him up.

Typically, digital music is sorted and categorized based on attributes such as the artist, genre, album, language, popularity, etc. Online music streaming services often recommend music based on users' preferences and previous listening histories, and they employ collaborative filtering methods and content-based recommendations. However, these recommendations may not suit the mood of the user at that moment. As a result, classifying songs manually by learning the user's preference for emotions is time-consuming.In other words, recommendations can also be achieved based on the physiological and emotional status of the user, which is mainly captured by facial expressions.

Studies have been conducted on detecting emotions by using facial landmarks.

Based on real-time analysis of facial emotional expressions, this paper proposes a CNN-based approach to recommend music.

Presented in section II of the paper is a brief survey of related work in the field of emotion detection. The proposed system architecture is discussed in section III and the hardware and software requirements are discussed in section IV. The proposed system architecture is discussed in section III and the hardware and software requirements are discussed in section IV.

II. RELATED WORK

Facial expressions can be used to analyze emotions in different ways. There has been much research conducted on detecting and classifying the emotional status expressed on the face by users using different approaches. For feature extraction and classification of the facial image, it is preprocessed and subjected to different algorithms.

In paper [4], the author proposes collaborative filtering, which consists of filtering the data similarly to what we feel. The paper examines the use of collaborative filtering techniques for music recommendation systems. Collaborative filtering is a technology that makes predictions based on the relationships between users and items. By using CF, they recommend the music by different parameters to enhance the effectiveness of the recommender system based on aggregating results of computing similarities between users and between items.

In paper [5], the author proposes EmoPlayer, an Android application that suggests music based on the user's mood. The system uses a camera to detect the user's face and to capture the user's image.As the songs are played, the program creates a list of songs that can improve his mood. EmoPlayer detects faces using the Viola Jones algorithm, and classifies emotions according to Fisherfaces.

There are three stages to the proposed method described in paper [6]: (a) face detection, (b) feature extraction, and (c) facial expression recognition. An YCbCr color model is used to detect skin color, lighting compensation for getting equilibrium on the face, and phonological operations to retain the required parts of the face. AAM (Active Appearance Model) is used to extract the facial features like the eyes, nose, and mouth from the first phase output. In the third stage, automated facial expression recognition, a simple Euclidean distance method is used to compare the feature points of the training and query images.

III. METHODOLOGY -

Images are composed of pixels with a 2-dimensional array. Pixels in an image can be classified based on their features. For example, Scikit-learn algorithms like SVM, decision-tree, Random-Forest, etc, which excel at solving classification problems, do not extract any appropriate features. This is where the Convolutional Neural Network comes into the picture. CNN, which stands for Convolutional Neural Network, is a combination of multiple layers of neural networks applied to image processing. consist of following layers - input layer, convolutional Layer, pooling Layer, dense Layer. An example of a convolutional neural network designed for mobile and embedded vision applications, MobileNet is used to combine deep neural networks with streamlined architectures that have low latency, allowing them to be used for low-latency applications. MobileNet can be used to build classification, detection, embedding, and segmentation models on top of small, low-latency, low-power models that meet the resource constraints of a variety of use cases.

IV. PROPOSED SYSTEM

Knowing the user's emotions is the key to selecting preferred music. Figure 1 shows the system architecture for generating playlists based on the emotion of the user. The proposed framework is a single input CNN model that detects emotion via facial landmarks.

By using CNN algorithms to classify the acquired facial expression, the proposed model derives the exact emotional state of a user from their facial expressions and recommends the appropriate music from the predefined directories.

A machine learning algorithm is used to detect emotions using supervised learning. With the use of CNN, unsupervised learning models are associated with learning algorithms that use data used for clustering and regression analysis, thus identifying an optimal boundary between the possible outputs.

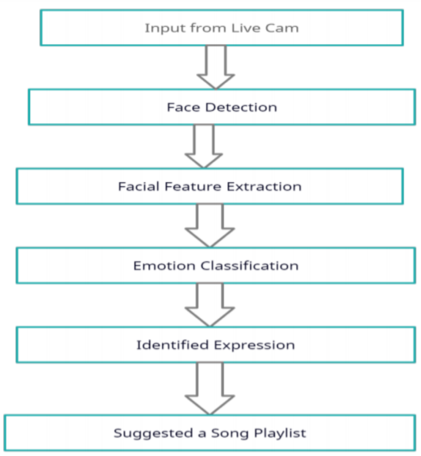


Fig 1: System Architecture

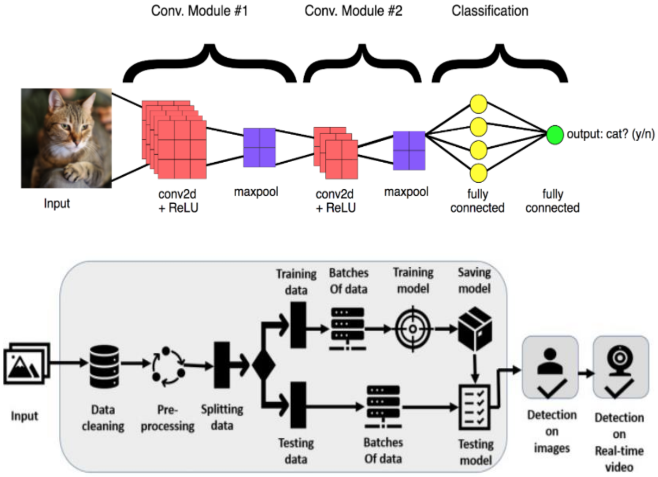


Fig 2: CNN Model Implementation

Emotion detection using facial expression

A wide-scale application of deep learning algorithms, such as Convolutional Neural Networks, for emotion detection is its high accuracy in classifying and recognizing facial expressions.

For emotion detection, the FER2013 dataset is used. This dataset includes 38,887 grayscale images of face sizes 48 x 48 with 7 emotion categories:

Table 1. Emotion labels and images in dataset

Label	Emotion	No. of images
0	Angry	4593
1	Disgust	547
2	Fear	5121
3	Happy	8989
4	Sad	6077
5	Surprise	4002
6	Neutral	6198

From 38,887 images, 28,709 were used to train the model, and the rest were used for testing.

Faces are captured using a webcam. Images are preprocessed and rescaled to 48x48 gray scale values and tested with the testing data. Using the OpenCV framework, we then preprocess the image to detect the face using the Haar Based Cascade Classifier.

Further, The input image is then subjected to a series of convolution layers where high level features are extracted. By using ReLu as an activation function, this layer examines the spatial and temporal dependencies of the image in order to identify both the low-level and high-level features.

During the first part of a Convolutional Layer, the kernel/filter is responsible for wrapping up the convolution operation. In the Convolution Operation, the input image is used to extract the high-level features like edges. A Convolutional Network need not be limited to just one Convolutional Layer. Conventionally, the primary ConvLayer is charged with capturing the low-level features like edges, color, gradient orientation, etc. As added layers are added, the architecture continues to adapt to High-Level features. An activation function called ReLU (Rectified Linear Unit) is applied after the convolution operation. It converts negative values contained within the feature map to '0' in order to bring non-linearity to the model.

The Pooling layer is answerable for reducing the spatial size of the Convolved Feature. Through dimensionality reduction, this is often done to reduce the computational power needed to process the information. Furthermore, it can be used to extract dominant features that are rotational and positional invariant, allowing the model to be effectively trained. The two types of pooling are Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the kernel. Average Pooling returns the average of all the values in the area covered by the kernel

The neurons during this layer have full connectivity with all neurons during the previous and subsequent layers, as seen in a regular FCN. Full Connected Layer is also referred to as Dense Layer, since it provides learning features from every combination of the features in the previous layers. A feed-forward neural network is fed the flattened output and back propagation is applied to each coaching iteration, helping to map the representation between the input and output. After a series of epochs, the model can differentiate between dominating and certain low level features and classify them using SoftMax Classification

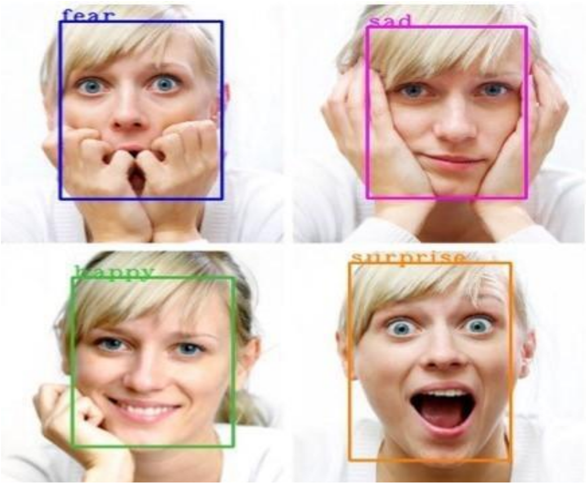


Fig 3: Sample images of few emotions

B. Music Recommendation

After detecting an emotion by using the proposed CNN model, the next step is to recommend the music that matches the detected emotion based on predefined directories.

V. HARDWARE AND SOFTWARE REQUIREMENTS

Server Side:

Hardware Requirements:

- Android OS version 8 & above
- Minimum 4GB RAM
- Minimum 16MP Resolution front camera (for testing on android device)
- 30 MB Memory space for the app
- Having access to the Internet
-
- **Software Requirements:**
-
- At least Python 3.6
- OpenCV 3.1
- Android Studio version 4.1
- Webcam (for testing on laptop/desktop)

Client-Side :

Hardware Requirements

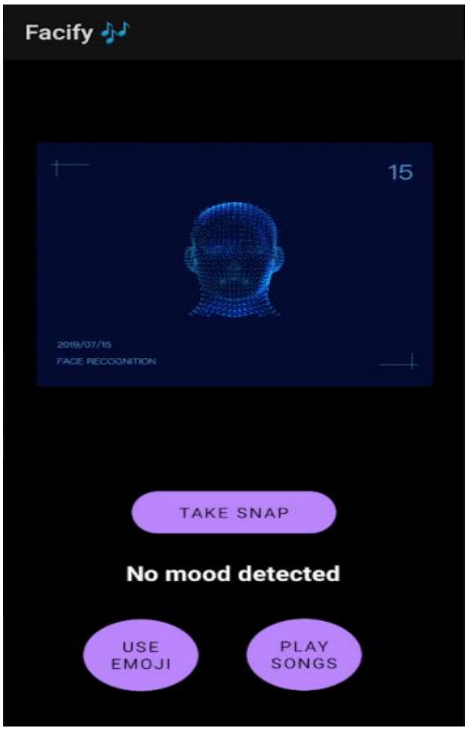
- Smart Mobile Phone

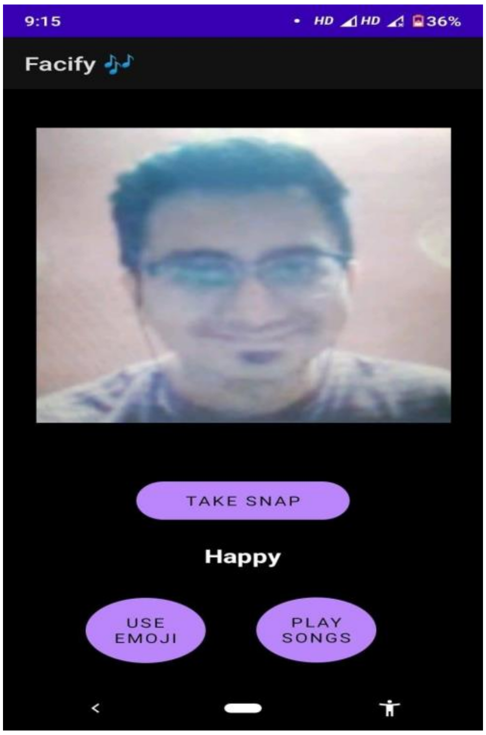
Software Requirements

- Android 6.0 or above

VI. RESULTS

This is the 1st page of our application





VII. SCOPE

In order to protect data, facial recognition can be integrated into devices such as smartphones and tablets.Furthermore, it can also be used to detect sleepy mood while driving. It can also be used to add extra features which result in Android Development. Social robot Emotion recognition system, Medical practices, and Feedback system for elearning are some of the other applications that can use this feature. Physical or mentally challenged people can use it to identify their emotional state, which can be used to treat them.

This feature can also be incorporated into automated counseling systems so that they provide the appropriate counseling to enhance the system.

VIII. CONCLUSION

Stress and emotions are often triggered by music, so it is necessary to recommend music according to the user's current emotional needs.There are many audio and video recommendation systems already in use such as Spotify,

Netflix, Gaana, YouTube, etc. that are based on search queries, not emotional needs. Therefore, the proposed CNN-based model detects the emotion and proposes a music playlist according to the mood of the user. It is supplemented with modules for detecting facially expressed emotions. The purpose of this project was to explore facial expression recognition for implementation of an emotion-based music player. The manual analysis of faces by people was completely replaced by reasonable computer programming.Apart from providing theoretical background, this study provides approaches to outline and execute emotion-based music players with a wide variety of image processing techniques.In the proposed system, facial images are processed and basic emotions are recognized, and then music is played based on the user's emotions, and also suggested music that enhances mood. We would like to improve our system's ability to recognize emotions in the future and also recognize more different emotions.

IX. REFERENCES

[1] In 2014 International Conference on Electronics & Communication Systems (ICECS -2014), Anaghia S. Dhavalikar and Dr. R. K. Kulkarni presented a paper titled "Face detection and facial expression recognition system".

[2] Krupa K S, Ambara G, Kartikey Rai, Sahil Choudhury, “Emotion aware Smart Music Recommender System using Two Level CNN”, 2020 Third International Conference on Smart Systems & Inventive Technology (ICSSIT).

[3] A study of face recognition based on LBPH & regression of Local Binary Features by Gao Xiang, Zhu Qiuyu, Wang Hui, Chen Yan, presented at the 2016 International Conference on Audio, Language and Image Processing (ICALIP).

[4] "Collaborative filtering for music recommender systems", 2017 IEEE Conference of Russian Young Researchers in Electrical & Electronic Engineering (EIConRus).

[5] In a paper entitled “Emotion based mood enhancing music recommendation”, presented at the IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017, Aurobind V. Iyer, Viral Pasad, Smita R. Sankhe, Karan Prajapati.

[6] IEEE International Conference on Robotics, Intelligent Systems & Signal Processing Proceedings. 2003. Shaoyan Zhang, Hong Qiao. "Face recognition with support vector machines.".