International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-11 Issue-3, January 2022

Music Player using Emotion Recognition

Abhay Chopde, Anuradha Lohar, Mitul, Swaraj Mane, Shaurya Mhaske

Abstract*: Expression of humans has a very crucial role in determining the present state and the mood of a person. It helps in the extraction of the emotion by understanding the different features of the face like cheeks, forehead, eyes, or maybe the curvature of the smile. Music plays a crucial role in the daily lifestyle. It is essentially a kind of thing that soothes and calms the body, brain and soul of a human. During this fast-paced lifestyle, everyone goes through a roller coaster of emotions and changes in moods within seconds. People tend to listen to music according to their emotions. Hence, we have chosen emotion and music and combined them together. Our program will first detect facial expressions and using that it will recognize the emotion of a human being. This is done through counting of repeated integers in a trained model and playing music favorable with the mood detected. This will alleviate the mood or just calm the individual. It may also be able to fetch quicker songs consistent with the mood, saving time looking up for different songs. Parallelly developing a software which will be used anywhere with the assistance of providing the functionality of playing music consistent with the emotion detected. People tend to concentrate on songs consistent with their moods, and it feels very frustrating if a tragic song starts playing when you are happy. It seems like a burden to constantly visit the music application to vary the song. Therefore, in this work, we've come up with a solution in which we will use computer vision to detect the emotion of an individual and therefore the system shall play a song consistent with the mood of the client.

***Keywords*: Music, Tensor flow, Emotion, OpenCV.**

I. INTRODUCTION

People always want to listen to songs according to their mood and the solution to understand the mortal geste and how it behaves and interacts with the surroundings. Latest, computing provides technologies that analyze mortal and computing commerce. Expressed feelings on the face convey the intention of a person.

Manuscript received on January 30, 2022.
Revised Manuscript received on February 05, 2022.
Manuscript published on February 28, 2022.

* Correspondence Author
Abhay Chopde, Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune (Maharashtra), India. Email: abhay.chopde@vit.edu
Anuradha Lohar*, Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune (Maharashtra), India. Email: anuradha.lohar18@vit.edu
Mitul, Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune (Maharashtra), India. Email: mitul.kumar18@vit.edu
Swaraj Mane, Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune (Maharashtra), India. Email: swaraj.mane18@vit.edu
Shaurya Mhaske, Department of Electronics & Telecommunication Engineering, Vishwakarma Institute of Technology, Pune (Maharashtra), India. Email: shaurya.mhaske18@vit.edu

The emotion lets others know the state of the person similar to sadness, joyfulness and wrathfulness. The ways in which people communicate has 35% of verbal and 65% of expressional communication. Also, expressions of a person's face are a very prominent means of social communication. Thus, these expressions on a human face are recognized means for discovering the feelings. Communicating technique which is not verbal among humans is usually done by facial expression. The reason behind this commerce is that humans can identify the emotion in an effective and prompt manner. Therefore, there exists a demand to develop a machine that can scan mortal emotion and take actions according to them. The identification of facial expressions plays a crucial part in relating patterns and image processing, and relating facial expressions through three main stages: recognition of the face, extracting its features, and classification of the emotion.

Music plays a truly primary part in elevating people's lives as it's a really prominent mode of entertainment for people who are music lovers and addicted to listening to it. In the present world, with the modernization of the entertainment sector and technologies, the development of a lot of different music-playing applications is also taking place. These applications have a number of features like the pause, play, fast forwarding, reducing the playtime, variable playback speed, genre category, playing songs with different music player applications and having volume control, etc. These features might satisfy the person's introductory demands, but the listener certainly faces the hardship of manually going through the playlist containing a variety of songs and choosing the one that supports their mood and geste at that specific moment. An application that plays songs by detecting the emotions of a person is a really fresh approach towards this problem that helps the person to play different songs just according to his mood, sentiments, and feelings.

It does facial sentiment detection of a person and randomly picks and plays a song from the data provided based on the emotion of that person. The face is a prominent organ of the human body and it essentially has a vital role in the descent of a person's actions and his emotional and sentimental state. The computer's camera first detects the face in a grid and then scans the image of the person. It also excerpts the specific features of the person's face from the obtained image. Facial expression distributed into 2 classes, smiling and non-smiling. The foremost conception of this program will be focused on playing the song on its own grounded on the feelings of the listener. It focuses on giving self- suggested music with respect to the feelings detected. In the being system, the human has to handpick the tune, aimlessly playing it might not match the thoughts of the listener; the person must categorize various genres into multiple feelings, and also for playing any suggested music, the person must choose a specific emotional directory.

Then the selected song would be streamed from the directories (already defined) according to the emotion detected previously.

II. LITERATURE REVIEW

Dr. R.K. Kulkarni and A.S. Dhavalikar [1] worked on a system which automatically detects facial features and uses different kinds of algorithms to detect the emotions. In this system, there were three levels. First was detection of face, also feature extraction and eventually expression recognition based on different algorithms and methods.

A music player based on emotion and expressions which uses IP was proposed by Dr. Jayshree Jha et al., [2] and it proved how different techniques, algorithms and trained models that have already been proposed earlier by a lot of other authors in their research could be used for the purpose of combining the human with the computer by putting the emotion and music player together for working. It has therefore corrected and reduced the workload of users in making and controlling playlists and provided an amazing benefit to the listeners of the song by bringing them the most favorable song according to their expression at that time.

Anukritine et al., [3] deduced an algorithm which works by giving a list of songs from a specific playlist related to the listener's emotions which are shown on his face. It basically works by reducing the time of computation and cost included in using a lot of different hardware. The main idea included the division of numerous emotions into categories like happiness, anger, grief, surprise, fright. It gave an appropriate audio information retrieval approach that took less time to extract the related information from an audio signal.

A similar android application was developed by Aditya et al., [4] which acted as a music streamer for a person. It used image processing to analyze the emotion and play songs according to a person's emotion. The app is made using OpenCV to perform emotion based methodologies. The research also made progress in comparing various algorithms that are used for facial detection. The photos of the person were captured with the use of the front camera of the mobile device. It majorly aimed at providing satisfaction to music addicts by extraction and detection of their emotions.

Algorithm which recognizes facial emotion that includes three levels: pre-processing of the image, extracting and classification was developed by A. Habibzad et al., [5]. Firstly it describes numerous levels in image processing. It includes pre-processing and a feature extraction filtering technique. The other part worked by optimizing the oval characteristics of lips/eyes. Finally at the third level, the optimal parameters of eyes and lips helped in the categorization of their sentiments. The conclusion claimed that velocity of recognizing facial expressions came to be a lot improved than fellow existing methods.

Charvi Jain et al., [6] worked on creating a system which used Support Vector Machine (SVM) and grid search to refine images. Similarly, Shivam Gupta et al., [7] were getting an accuracy of around 93.7% by using the HAAR filter from openCV and SVM.

S. Metilda Florence and M Uma [8] proposed a system that can detect the facial expressions of the user and based on it, the extraction of the facial landmarks was done.It would then be classified to get a particular emotion of the user and then the songs matching the user's emotions would be shown to the user. A similar work was proposed by Charu Agrawal et al., [9].

Chavi Ralhan et al., [10] proposed a system of emotion detection using fisher face model and compared it with existing models. They used a webcam to capture the face image, resized it and converted it into grayscale for acquiring better results.

III. METHODOLOGY

1. We have trained the model using retrain.py. It creates a model file retrain. pb.

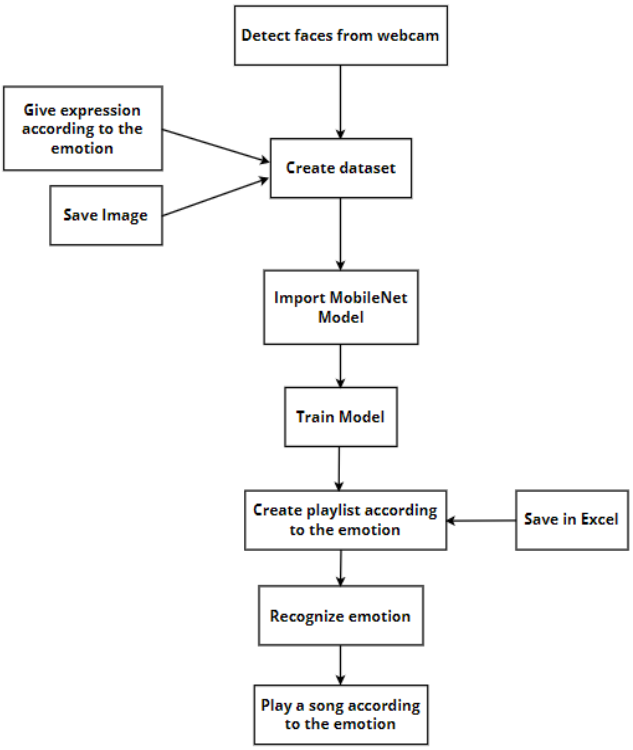

Fig. 1 Block diagram of system

Fig. 1 shows the block diagram of our proposed system.

2. We then run the label.py file which detects the face using cv2 (OpenCV) and it makes use of the haar cascade classifier which is an object and obstacle detecting algorithm. It also helps in identifying faces on a screen, a captured picture/motion video in real time. The given algorithm works by detecting features like edges and lines to identify the obstacles clearly.
3. After doing that, we then provided the face images from the database to label_image.py. Now it returns the detected emotion (in integer 1/2/3/4) by using a pre-trained model file. The emotions are stored up to 10 times. Then we have checked the emotion with the highest count and mapped the integer to the correct emotion (as defined in the dictionary).
4. Now the emotion is given to the play_music_pygame.py under '/emotions file' (in the dataset). This dataset contains different images related to the different emotions: happy, angry, neutral or sad.

Retrieval Number: 100.1/ijitee.C97610211322
DOI: 10.35940/ijitee.C9761.0111322
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

5. Finally it reads any random song name from the dataset and plays any favorable song from the '/songs' folder. Also, the necessary logs are shown on the terminal and the music starts playing in the program itself.

Libraries used:

- Cv2
- Sys
- Pandas
- Random
- Datetime
- Numpy
- Tensorflow
- Pygame

1. **Why OpenCV?**
   OpenCV provides many functions for facial detection and facial recognition. It comes with a trainer and a detector. If you want to train your own classified objects such as mobile phones, pens, smart devices, computers etc., you can use OpenCV to create it.
2. **Tensorflow:** TensorFlow is a library in the python programming language which is used for machine learning, deep learning structures, artificial intelligence, etc. This open-source and free library plays an important role in training models of deep neural networks. F.Ertam and G.Aydin [11] proposed a study in which they used Tensorflow to classify MNIST dataset. This dataset is repeatedly used in data analysis studies.
3. **Haarcascade:** A Haar classifier or Haar cascade classifier is a program which involves different machine learning object detection algorithms to detect obstacles and things in a media.
4. **Normalization:** Normalization technique is usually applied for preparation of a dataset in a machine learning program. Normalization majorly focuses on changing the numbers and values of a numeric column to use a common scale without disturbing the range difference of numerous values and to prevent data loss.
5. **MobilNet:** The MobileNet is a model which is designed to be used in various mobile applications and it is TensorFlow's first mobile computer vision model.
6. **MobileNet Architecture:** Fig.2 shows different MobileNet Layers that are used in our system.

Type / Stride	Filter Shape	Input Size
Conv / s2	3 × 3 × 3 × 32	224 × 224 × 3
Conv dw / s1	3 × 3 × 32 dw	112 × 112 × 32
Conv / s1	1 × 1 × 32 × 64	112 × 112 × 32
Conv dw / s2	3 × 3 × 64 dw	112 × 112 × 64
Conv / s1	1 × 1 × 64 × 128	56 × 56 × 64
Conv dw / s1	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 128	56 × 56 × 128
Conv dw / s2	3 × 3 × 128 dw	56 × 56 × 128
Conv / s1	1 × 1 × 128 × 256	28 × 28 × 128
Conv dw / s1	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 256	28 × 28 × 256
Conv dw / s2	3 × 3 × 256 dw	28 × 28 × 256
Conv / s1	1 × 1 × 256 × 512	14 × 14 × 256
5× Conv dw / s1	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 512	14 × 14 × 512
Conv dw / s2	3 × 3 × 512 dw	14 × 14 × 512
Conv / s1	1 × 1 × 512 × 1024	7 × 7 × 512
Conv dw / s2	3 × 3 × 1024 dw	7 × 7 × 1024
Conv / s1	1 × 1 × 1024 × 1024	7 × 7 × 1024
Avg Pool / s1	Pool 7 × 7	7 × 7 × 1024
FC / s1	1024 × 1000	1 × 1 × 1024
Softmax / s1	Classifier	1 × 1 × 1000

Fig. 2 MobileNet Layers

7. Mobile networks on the principle of separating convolutions. It is done on the ground of depth. It has been proven that it reduces the count of parameters when its comparison is done with the network having regular complications of same depths in the net. As a result of this, a lightweight deep neural network is created. A divisible complication based on depth is made from the two operations:
   Depth wise convolution
   Point-wise convolution
   Fig. 3 shows classification of images with MobileNet.

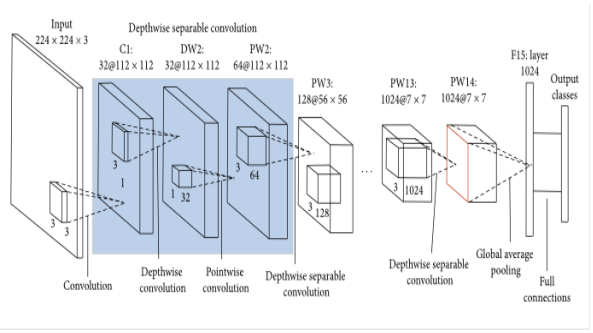

Fig. 3 Image classification with MobileNet

8. **Difference between standard convolution and depth wise separable convolution:** The major difference would be that the MobileNet architecture splits the convolution into a 3x3 depth wise convolution and a 1x1 point-wise convolution. This has been displayed in the image below. Whereas, the traditional CNN has a single 3x3 conv. layer, batch norm and ReLU.
   Fig.4 shows the difference between the Standard convolutional layer and Depth-wise separable convolution layer.

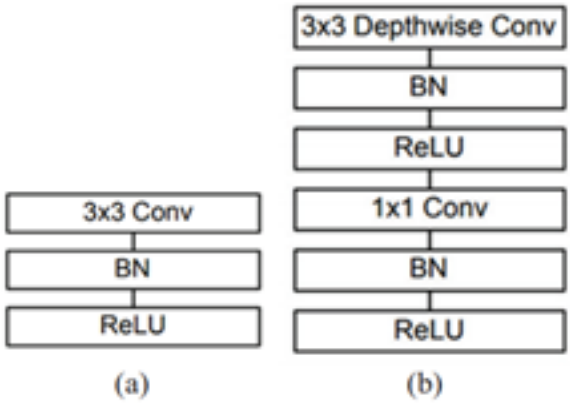

Fig. 4 Difference between layers

Fig. (a) Standard convolutional layer with batch normalization and ReLU.
Fig. (b) Depth-wise separable convolution with depth-wise and pointwise layers followed by batch normalization and ReLU.

The open-sourced model MobileNet (a category of CNN) has proven to be excellent in giving a beginning for the existing classifiers to be trained which are extremely small and rapid.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Table. 1 Shows Different Models and Accuracy

Model	Accuracy
1.00 Mobilenet-224	83.3%
0.75 Mobilenet-224	81.9%
1.00 Mobilenet-192	81.9%
0.75 Mobilenet-192	80.5%

Table. 1

9. **Softmax:** A softmax function that transforms the output into a probability distribution value. The softmax function is additionally a sort of sigmoid function but is handy once we attempt to handle classification problems. It's a nonlinear function. The softmax function is utilized within the output layer of the classifier where we are actually trying to analyze the chances to define the class of each input.

IV. EXPERIMENTATION

A. Dataset:

a. Images: Our dataset contains images with emotions like angry, happy, sad or neutral for Proper Training of Data.

Figures 5,6 and 7 shows a dataset of images with different emotions.

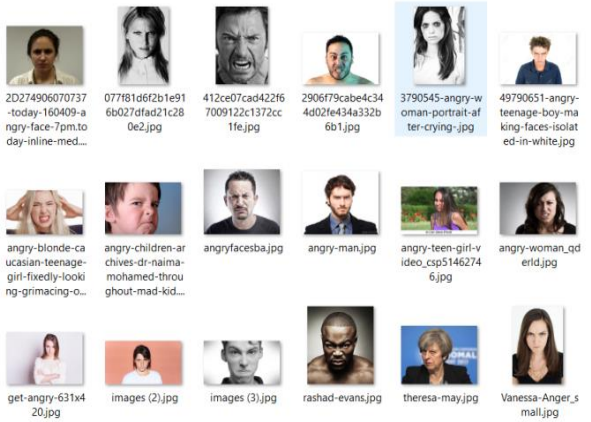

Fig. 5 Angry Images Dataset

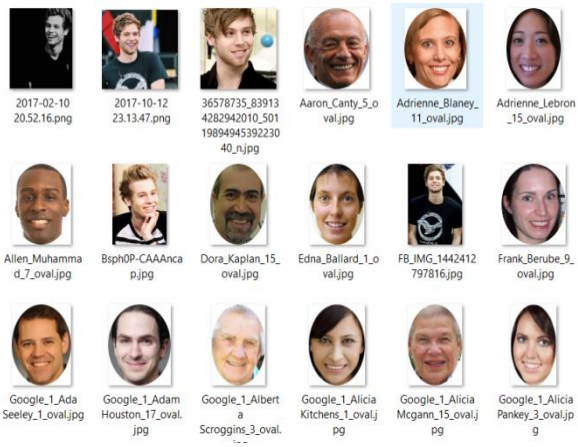

Fig. 6 Happy Images Dataset

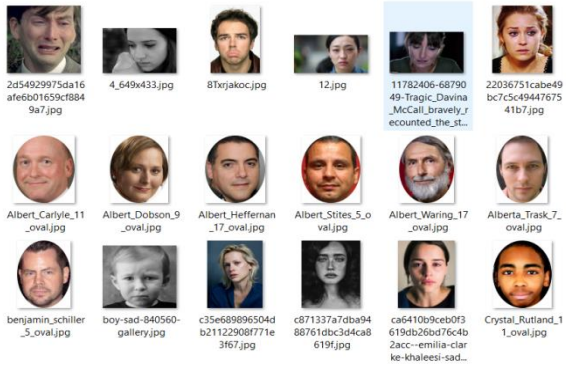

Fig. 7 Neutral or sad Images Dataset

b. Emotions file: We have added .csv files which have a dataset of different songs having respective emotions.
Fig. 8 shows the emotions files dataset in the form of .csv files.

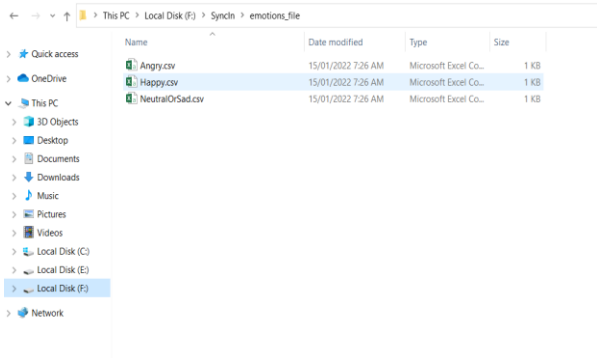

Fig. 8 Emotions files Dataset

c. Songs dataset: We have added a list of songs having different moods which will play after detecting the particular facial emotion.
Fig. 9 shows a dataset of songs which has songs with different moods.

Fig. 9 Songs Dataset

B. Process:

We have divided our project into three processes:

1. The first process loads the xml file and detects the faces available using the web camera.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

2. The second process is getting the dataset ready and dividing them into three classes. i.e. angry, happy, neutral or sad and training the model.
3. Last process is getting the emotion and playing the song accordingly.

The first file is facecrop.py, which is the manual preprocessing for the images. After running this file it finds the number of folders in the images directory.
Fig. 10 shows different images directories with emotions that are saved in the folder.

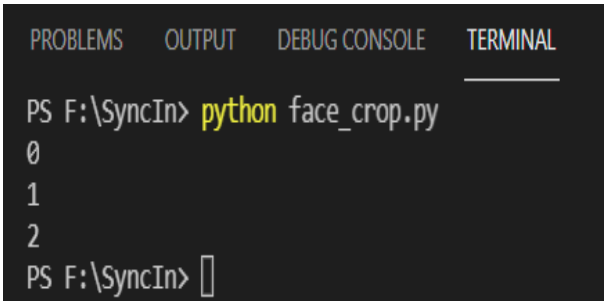

Fig. 10 Images Directory

The label.py is importing the label_image.py and the cv2 library. label_image.py is used to get the label for the emotion. Then subtracted, resized the files then used normalization. After running a command using command prompt and training the model, run label.py as it is the entry point to the project. It detects the emotion in real time and captures the emotion and streams the musical number grounded on the person's mood.
Fig.11 shows that the face is detected with a particular emotion ID.

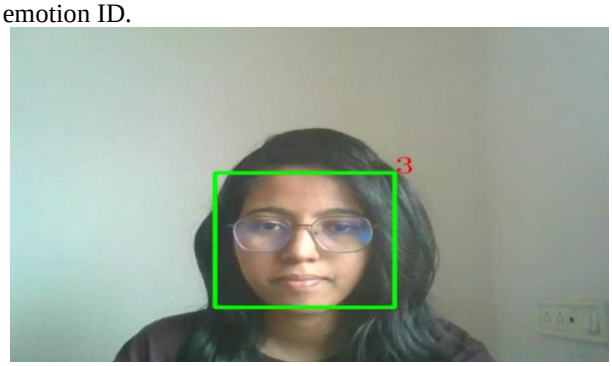

Fig. 11 Face Detection

Fig. 12 shows that the emotion is recognized with predictions and particular frequency.

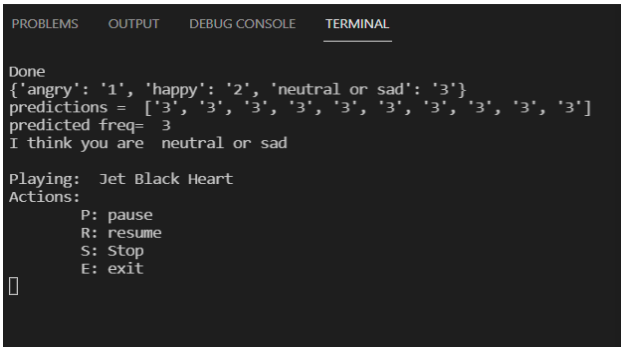

Fig. 12 Emotion Recognition

V. RESULTS

Our dataset was mainly divided into two parts: The training (80%) and the testing (20%) part. Dataset contains images which have different emotions. The model which has quite decent accuracy is nothing but MobileNet. So it is being compared with existing Models based on the three parameters Training, Validation, Accuracy.

1. Detecting emotion: After running the label.py file system uses video capture feature to capture face in real time and the model works towards getting and recognizing the user's emotion.
2. Playing songs according to the emotion: After capturing the face, the model tries to get the specific emotion of the person and fetches the music accordingly. It also gives pause, stop, resume, exit features.
3. Fig. 13 shows that the song is playing according to the emotion of the person which is recognized by the webcam.

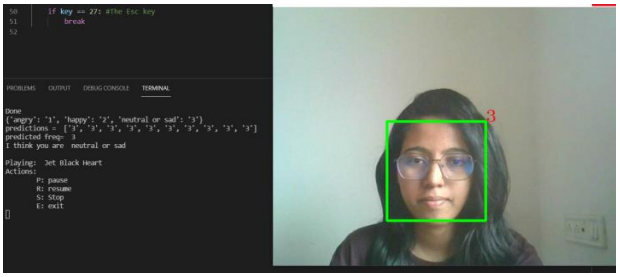

Fig. 13 Playing Song According to Emotion

VI. CONCLUSION

This system can be used to bring automation in the entertainment sector and provide a much better song listening experience to the world. The accuracy of the trained model came to be 83.3%. Other pre-trained models can also be used for variation accuracy. The program meets the original requirements of music listeners without bothering them like other music players do. It increases the commerce of the system with the listener in numerous ways. It facilitates a person's work by getting the image with a camera, determining his feelings through facial expressions and proposing a favorable song through a more advanced and interactive system. But this currently only works inside the program and python ide. When the program is run, it detects the emotion and fetches the song from the dataset and plays it inside the program only. No other media player is embedded currently for playing the song which could be done in the near future to make the program more prominent and usable.

FUTURE SCOPE

1. This music player can be used in different music apps like gaana and spotify where users will be able to access songs according to their emotion.
2. It can be used in different musical events where songs can be played using audience mood.
3. This system can be used in mental therapies for psychological treatment where patients will be cured by musical therapy.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

REFERENCES

1. Anagha S. Dhavalikar and Dr. R. K. Kulkarni (2014) Face Detection and Facial Expression Recognition System in International Conference on Electronics and Communication System (ICECS - 2014)
2. Prof. Jayshree Jha, Akshay Mangaonkar, Deep Mistry, Nipun Jambaulikar and Prathamesh Kolhatkar (2015) Facial Expression Based Music Player in International Journal of Advanced Research in Computer and Communication Engineering
3. Dureha A (2014) An accurate algorithm for generating a music playlist based on facial expressions International Journal of Computer Applications 100 33-9
4. Gupte A, Naganarayanan A and Krishnan M Emotion Based Music Player-XBeats International Journal of Advanced Engineering Research and Science 3 236854
5. A. habibzad, ninavin and Mir kamalMirnia A new algorithm to classify face emotions through eye and lip feature by using particle swarm optimization 2012 4th International Conference on Computer Modeling and Simulation
6. Charvi Jain, Kshitij Sawant, Mohammed Rehman and Rajesh Kumar (2018) Emotion Detection and Characterization using Facial Features 2018 3rd International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE)
7. Shivam Gupta (2018) Facial emotion recognition in real-time and static images in 2018 2nd International Conference on Inventive Systems and Control (ICISC)
8. S Metilda Florence and M Uma (2020) Emotional Detection and Music Recommendation System based on User Facial Expression in 3rd International Conference on Advances in Mechanical Engineering (ICAME 2020)
9. Charu Agrawal, Meghna Varma, Anish Varshaney, Khushboo Singh, Chirag Advani and Dr. Diwakar Yagyasen (2021) Emotions Based Music Player International Research Journal of Engineering and Technology (IRJET) Volume: 08 Issue: 07 | July 2021
10. Chavi Ralhan, Kodamanchili Mohan, Vinay Raj Kalleda, Pendli Anirudh Reddy and Pannamaneni Saiprasad (May 2021) Emotion Based Smart Music Player in International Journal of Scientific Research in Computer Science Engineering and Information Technology
11. Fatih Ertam and Galip Aydın (2017) Data classification with deep learning using Tensorflow in 2017 International Conference on Computer Science and Engineering (UBMK)000000

AUTHORS PROFILE

Abhay Chopde, Professor at Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, Pune. E-mail: abhay.chopde@vit.edu

Anuradha Lohar, Pursuing BTech in Electronics and Telecommunication Engineering from Vishwakarma Institute of Technology, Pune. E-mail: anuradha.lohar18@vit.edu

Mitul, Pursuing BTech in Electronics and Telecommunication Engineering from Vishwakarma Institute of Technology, Pune. E-mail: mitul.kumar18@vit.edu

Swaraj Mane, Pursuing BTech in Electronics and Telecommunication Engineering from Vishwakarma Institute of Technology, Pune. E-mail: swaraj.mane18@vit.edu

Shaurya Mhaske, Pursuing BTech in Electronics and Telecommunication Engineering from Vishwakarma Institute of Technology, Pune. E-mail: shaurya.mhaske18@vit.edu