IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 7.047 ⸎ Vol. 10, Issue 4, April 2022

DOI: 10.17148/IJIREEICE.2022.10477

Portable Media Player

Uma Thakur[1], Jayesh Chauhan[2], Sneha Nagdeve[3], Rajani Meshram[4], Saloni Pillewan[5]

Professor of Computer Science & Engineering at Nagpur's Priyadarshini College of Engineering,
India [1]

Research Scholar, Computer Science & Engineering Department, Priyadarshini College of Engineering,

Nagpur, India [2, 3, 4, 5.]

Abstract: Despite the fact that each person's face is unique, their expressions tell the same story and play a significant influence in determining an individual's emotions and behaviour. Music is the most pure form of art and a vehicle of expression, with a stronger emotional connection. It has a unique ability to improve one's mood. This project system uses facial recognition algorithms to generate an effective music player that is based on the user's sentiment. The retrieved face features will be used to create a system, reducing the time and effort required to do it manually. Facial data is recorded using a camera. The emotion module use deep learning techniques to detect the precise mood with a specific term The system's mood identification module has an accuracy of over 80% for real-time movies, while static images have an accuracy of 95 to 100%. As a result, time and performance precision are improved.

Keywords: Computer Vision, Deep Learning Techniques, Face Recognition, Emotion and Mood Detection, Mood Extraction Module, Computer Vision

I. INTRODUCTION

In communication systems, human emotions must be expressed and identified. Emotions can be expressed and recognised by humans. To recognise human emotions, computers use picture analysis or sensors. In our daily lives and professional lives, we interact with a large number of people, either directly or indirectly through phone calls, and it is sometimes critical for people to be aware of the current emotions of the people with whom they are conversing. Surprise, fear, wrath, happiness, sadness, disgust, and neutral are examples of human emotions. Facial movement and verbal tone are critical when transmitting emotions. The physicality and tone of the face, which can be altered to express various emotions, TELL the energy in the emission of words. Humans can quickly recognise these signal variances, as well as the information seen by any other sensory ORGAN. This study investigates how emotions can be captured via photos, sensors, and speech. Music is a vital source of happiness for music lovers and listeners, and it can even be therapeutic at times. Where words fail, music speaks, and as a result, it can simultaneously and gradually turn a PERSON'S unpleasant emotion into a happy one. Voice, gestures, and facial expressions Emotions can be represented in a variety of ways, including body language. We employ facial expressions to assist the algorithm in determining the user's mood. We can capture the user's face expression using the mobile device's camera. Emotion detection systems that employ gathered photographs to analyse the emotion are numerous. We employ neural networks to recognise emotions in this application.

II. REVIEW OF LITERATURE

There are a number of programmes with features and services for building music playlists or playing a certain song, and this method necessitates the use of human labour. Several methodologies and approaches have been proposed and developed to classify human emotional states of behaviour. The proposed approaches have only focused on a few of the main emotions, using sophisticated methodology like Viola and Jones.

The papers listed above provide an overview of the methodologies used and demonstrate approaches and algorithms such as machine learning with support vector systems for recording facial expression and generating a playlist for the user.Several research papers giving a brief about the idea are:

[1] The authors claim that recording, playing, processing, and maintaining digital audio is straightforward. Because of its ubiquitous use, handling equipment is relatively inexpensive, allowing more people to record and perform music and speech. Furthermore, the internet has made collecting recorded sounds much easier. As a result, the number of people owning recorded music has increased dramatically. Audio files are compressed and stored in internal memory by most modern audio players. The amount of music that will be saved has exploded as storage costs have constantly lowered. If each song is 5 Mbytes in size and saved in compressed format, a player with 16 Gbytes of memory may hold roughly

3,200 songs. It's difficult to organise such a large volume of music effectively. People tend to listen to a small number of favourite songs repeatedly, while others are unfairly ignored. We developed Affection as a tool for organising music collections. Affection groups together pieces of music that express similar emotions and gives each category a symbol. Listeners can easily select music that matches their mood with these icons.

According to the authors of this study, music is pervasive throughout our lives. People hear music in a variety of contexts, either actively or passively, and feel it as a kind of emotion expression, consciously or unconsciously. We describe a new location and emotion aware web-based interactive music system in this study. Its purpose is to give the user with their favourite music while simultaneously tracking their whereabouts and emotions. The technique begins with professional expertise-based guidance. If the user doesn't like the recommendation, he or she can disregard it and choose their own music. The user's interactions with the system, present location, and emotion are all logged during the process of learning music preferences. As a result, the system may be able to adapt to the user's current musical preferences. Additionally, the more the user interacts with the system, the more personalised music is generated for him or her.

III. METHODOLOGY

Module 3.1: Face Recognition

The technique of recognising a face from an image or video input is known as face detection. Face recognition software is available in a number of flavours. The Viola Jones algorithm is used to recognise faces. The following are the main steps of Viola Jones' algorithm:

A. HAAR characteristic

HAAR characteristics are used to depict several aspects of the face. Har features are related to convolution kernels, which are used to detect the existence of a feature in an image. The total number of pixels in the black rectangle is subtracted from the total number of pixels in the white rectangle to give each feature a single value. the number of pixels in the white rectangle as a whole In this feature, the black areas are replaced with plus ones, while the white regions are replaced with minus ones.

B. Image integral

All pixels in the black and white zones must be summed up every time the window moves in HAAR feature computation. With an integrated image solution, it's a time-consuming process. Instead of summing all pixels beneath a rectangle with all four corner values, it saves time by summing all pixels beneath a rectangle with just four integral image corner values. Simply add the values of the pixels on the top and left to get the value of any pixel.

C. Adaboost

The Viola Jones approach uses a 24*24 window as the fundamental window to analyse characteristics in each photo. If we took into account all accessible feature properties such location size and kind, we'd have to calculate 160,000+ features in this timeframe, which is almost impossible. The fundamental idea is to eliminate many features that are redundant or inefficient, keeping just those that are really valuable. In Adaboost's version, 160 thousand features are removed, decreasing the number of features we need to evaluate to a few thousand. Adaboost's recovered features are poor classifiers. Adaboost combines weak classifiers in a linear fashion.

3.1. Module for Extracting Facial Features

CNN is used for feature extraction. To train the system for the emotion recognition module, we must use datasets that include photographs of happy, angry, sad, and neutral emotions. CNN has the unique ability to use automatic learning to uncover properties from dataset images for model construction. To put it another way, CNN has the ability to learn on its own. A two-dimensional image can be represented internally by CNN. This is depicted as a three-dimensional matrix on which training and testing operations are performed. All nodes in one layer are connected to nodes in the following layer in other neural networks, such as fully connected networks. Each connection has a corresponding weight. The computational complexity will rise as a result. In CNN, nodes in one layer are still only connected to valid nodes in the next layer. As a result, the computational complexity will be reduced. There are several levels for training and assessing input photos in this. The final layer is completely integrated and contains a classification task that categorises images based on their emotional content. The emotions seen should fall into one of four categories: angry,

DOI: 10.17148/IJIREEICE.2022.10477

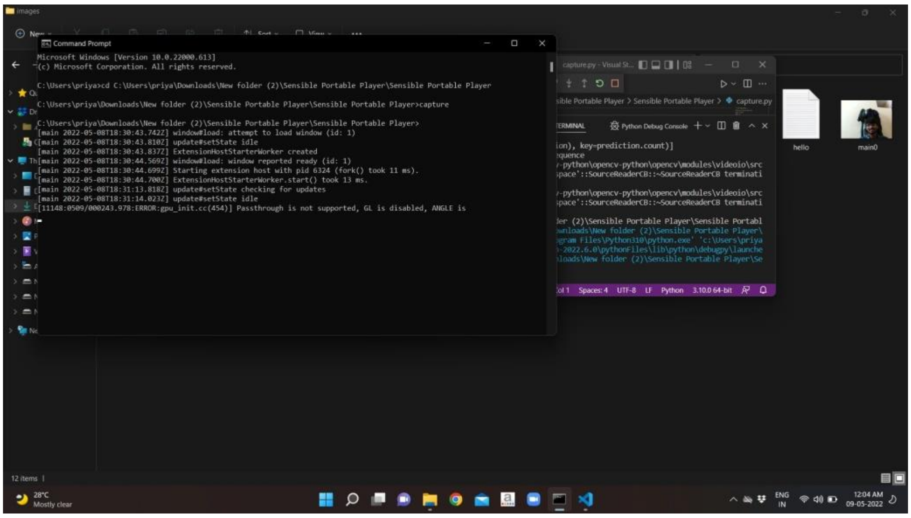
happy, sad, or neutral. Before entering the CNN, the full dataset will be divided in half. The great majority of it will be utilised for training, with the remaining 20% used for testing. After that, the model will be put to the test. At the convolution layer, a process known as filtering takes place. The arithmetic that goes into matching is known as filtering. The feature and the image patch must first be aligned. After that, multiply each image pixel by the corresponding feature pixel. Add them all together and divide by the total amount of pixels in the feature. In convolution, one picture becomes a stack of images after filtering. The ReLU, tanh, and sigmoid functions can be used to implement nonlinear processes. All negative numbers in ReLU are converted to zero, while positive values are kept alone. The leaking stops when the gadget is switched off. ReLU can be used to create a subtle gradient. As a result, no information will be overlooked. The issue of the dying ReLU is also resolved. The other two non-linear functions are outperformed by ReLU. After the convolution layer, the pooling layer will be applied. This is done to make the picture stack of the convolution layer smaller. The overall number of parameters will be lowered as a consequence [20]. You may select a window size here (usually 2 or 3). Then choose a stride (usually 2). The number of pixels that shift across the input matrix is called a stride. We change the filters one pixel at a time when the stride is one; when the stride is two, we change the filters two pixels at a time, and so on. After then, the window is shifted over the filtered photographs. Pooling can take three different forms. Maximum pooling, average pooling, and sum pooling are the three forms of pooling. Max pooling identifies the largest element from a rectified feature map. The procedure of taking the average of the components in the window is known as average pooling. The total of all elements in a feature map is referred to as sum pooling. Additional convolution and pooling layers can be added until the accuracy required is met. The matrix is flattened to a vector and transferred to the fully connected layer after the pooling layer. The purpose is to create a feature vector from a two-dimensional feature matrix that may be used to train a neural network or classifier. The vector elements mix to form models after the layers are fully joined. Finally, activation functions such as the SoftMax and Sigmoid functions are used to classify data.

3.2. Emotion Detection and Song Classification Module

The neural network classifier assigns one of four emotion labels based on the detected emotion: joyous, angry, sad, or neutral.

IV. IMPLEMENTATION

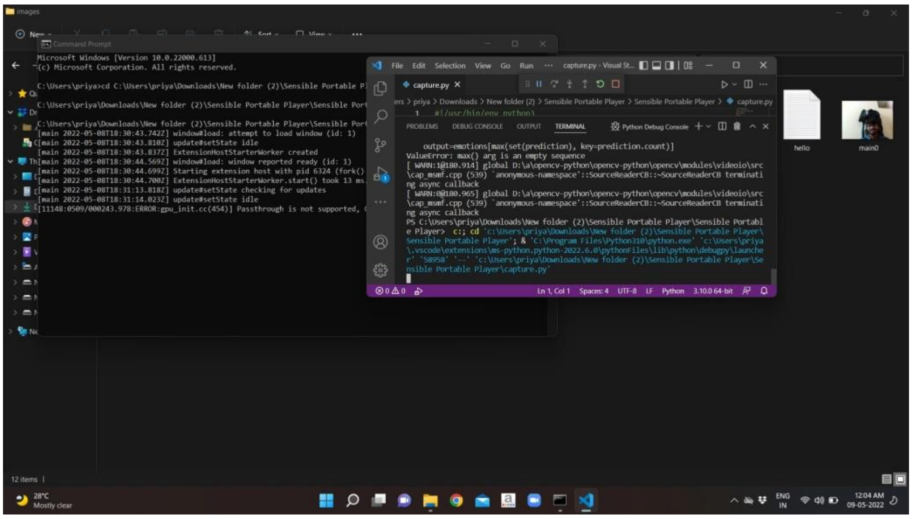
The approach presented is based on an artificial music suggestion system that plays music depending on a person's mood or current state. The person's photo is captured when the app is accessed, and the current emotion is recorded and recognised. The music is played in line with the feeling, according to the information offered by the image. Your phone's music has already been divided into four categories: happy, sad, furious, and neutral. Newly submitted songs are automatically categorised into moods. Facial expression recognition, song emotion recognition, and system integration are the three components of the system. Recognizing facial expressions and recognising auditory emotions are incompatible modules. As a consequence, the system integration module links two modules in the system in order to find the most appropriate match for the indicated emotion



images
Command Prompt
capture.py - Visual St...
sible Portable Player > Sensible Portable Player > capture.py
TERMINAL
Python Debug Console
Col 1 Spaces: 4 UTF-8 LF Python 3.10.0 64-bit
hello
main0
12 items
28°C
Mostly clear
ENG
IN
12:04 AM
09-05-2022

```
Microsoft Windows [Version 10.0.22000.613]
(c) Microsoft Corporation. All rights reserved.

C:\Users\priya>cd C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Portable Player

C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Portable Player>capture

C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Portable Player>
[main 2022-05-08T18:30:43.742Z] window#load: attempt to load window (id: 1)
[main 2022-05-08T18:30:43.810Z] update#setState idle
[main 2022-05-08T18:30:43.837Z] ExtensionHostStarterWorker created
[main 2022-05-08T18:30:44.569Z] window#load: window reported ready (id: 1)
[main 2022-05-08T18:30:44.699Z] Starting extension host with pid 6324 (fork() took 11 ms).
[main 2022-05-08T18:30:44.700Z] ExtensionHostStarterWorker.start() took 13 ms.
[main 2022-05-08T18:31:13.818Z] update#setState checking for updates
[main 2022-05-08T18:31:14.023Z] update#setState idle
[11148:0509/000243.978:ERROR:gpu_init.cc(454)] Passthrough is not supported, GL is disabled, ANGLE is
```

```
ion), key=prediction.count)]
equence
-python\opencv-python\opencv\modules\videoio\src
space'::SourceReaderCB::~SourceReaderCB terminati

-python\opencv-python\opencv\modules\videoio\src
space'::SourceReaderCB::~SourceReaderCB terminati

der (2)\Sensible Portable Player\Sensible Portabl
wnloads\New folder (2)\Sensible Portable Player\
gram Files\Python310\python.exe' 'c:\Users\priya
-2022.6.0\pythonFiles\lib\python\debugpy\launche
loads\New folder (2)\Sensible Portable Player\Se
```



```
Microsoft Windows [Version 10.0.22000.613]
(c) Microsoft Corporation. All rights reserved.

C:\Users\priya>cd C:\Users\priya\Downloads\New folder (2)\Sensible Portable Pl

C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Port

C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Port
[main 2022-05-08T18:30:43.742Z] window#load: attempt to load window (id: 1)
[main 2022-05-08T18:30:43.810Z] update#setState idle
[main 2022-05-08T18:30:43.837Z] ExtensionHostStarterWorker created
[main 2022-05-08T18:30:44.569Z] window#load: window reported ready (id: 1)
[main 2022-05-08T18:30:44.699Z] Starting extension host with pid 6324 (fork()
[main 2022-05-08T18:30:44.700Z] ExtensionHostStarterWorker.start() took 13 ms.
[main 2022-05-08T18:31:13.818Z] update#setState checking for updates
[main 2022-05-08T18:31:14.023Z] update#setState idle
[11148:0509/000243.978:ERROR:gpu_init.cc(454)] Passthrough is not supported,
```

```
    output=emotions[max(set(prediction), key=prediction.count)]
ValueError: max() arg is an empty sequence
[ WARN:1@180.914] global D:\a\opencv-python\opencv-python\opencv\modules\videoio\src
\cap_msmf.cpp (539) `anonymous-namespace'::SourceReaderCB::~SourceReaderCB terminati
ng async callback
[ WARN:0@180.965] global D:\a\opencv-python\opencv-python\opencv\modules\videoio\src
\cap_msmf.cpp (539) `anonymous-namespace'::SourceReaderCB::~SourceReaderCB terminati
ng async callback
PS C:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Sensible Portabl
e Player>  c:; cd 'c:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\
Sensible Portable Player'; & 'C:\Program Files\Python310\python.exe' 'c:\Users\priya
\.vscode\extensions\ms-python.python-2022.6.0\pythonFiles\lib\python\debugpy\launche
r' '58958' '--' 'c:\Users\priya\Downloads\New folder (2)\Sensible Portable Player\Se
nsible Portable Player\capture.py'
```

V. WORKING SYSTEM PROPOSED

The proposed system can recognise the user's facial expressions and extract facial landmarks from them, which can then be classified to identify the user's mood. When the emotion is recognised, the user will be presented music that fits their feelings.

The design and operation phases of the application are depicted in this section. The Emotion-Based Music Player, which is loaded on a mobile device, allows the user to access their personalised play-lists and play music based on their feelings.

VI. WORKING MODULES:

1. Frame Extraction / Live Camera:

The user may upload/capture photographs using the app's live camera, and the software subsequently pulls frames from the video. These frames are kept locally on the computer. The most common frame size is 640x480.

2. Face Recognition:

Images can be uploaded or captured. Use the Haar cascade Classifier to recognise faces in pictures.

3. Image post-processing:

Images can be uploaded or captured. Once we have the faces, we may perform preprocessing to the photos, such as noise reduction and normalisation.

a. Image Conversion from RGB to Grayscale: Convert the picture to grayscale by averaging the RGB values of each pixel.

b. Normalization of images: Normalization is a technique that alters the range of pixel intensity levels in order to prevent mental weariness or distraction from the visuals.

c. Noise Abatement: Errors in the picture acquisition process that cause pixel values to differ from the true intensities of the real scene.

4. Extraction of Features

An input and output layer make up an SVM. On the basis of the training dataset, SVM will categorise the features. Face features such as the nose, mouth, and eyes are extracted from the picture as points, as seen below.

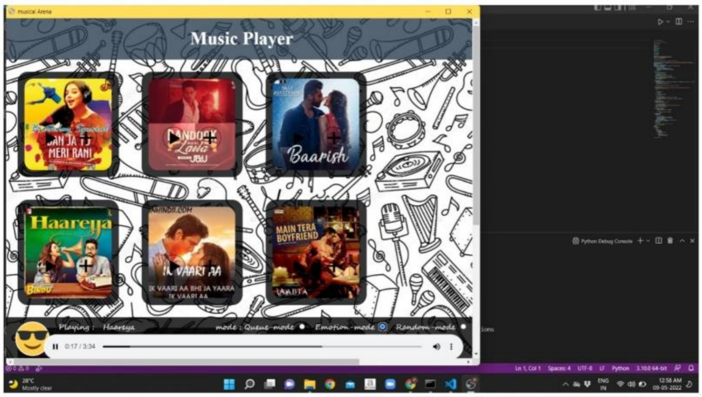
i. Raise your brows
ii. Distance from upper eyelid to brow
iii. Inter-eyebrow distance
iv. Upper lash line
v. Mouth width
vi. Mouth Open

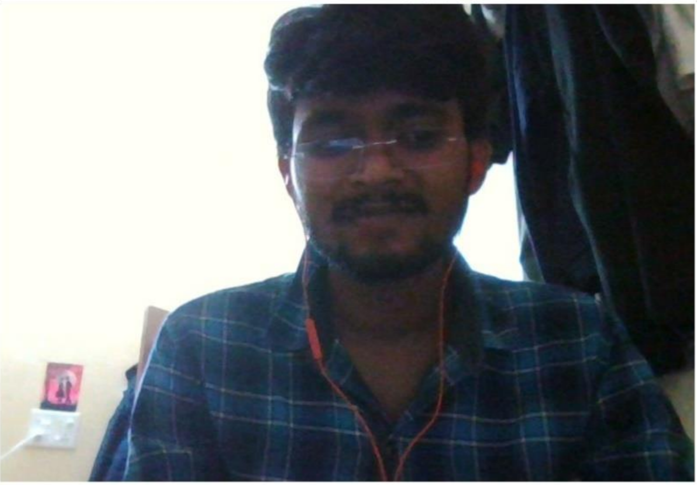
5. Calculation of Features:

During this step, all extracted characteristics are computed and the placement of the eyes, mouth, and nose on the person's face is determined. Face motion is detected using this calculation.

6. Music suggestion and emotion detection:

The emotions Happy, Neutral, and Sad are recognised using the SVM classifier on the retrieved characteristics. A song from the playlist is played based on the user's mood, such as sad, furious, rural, or cheerful.





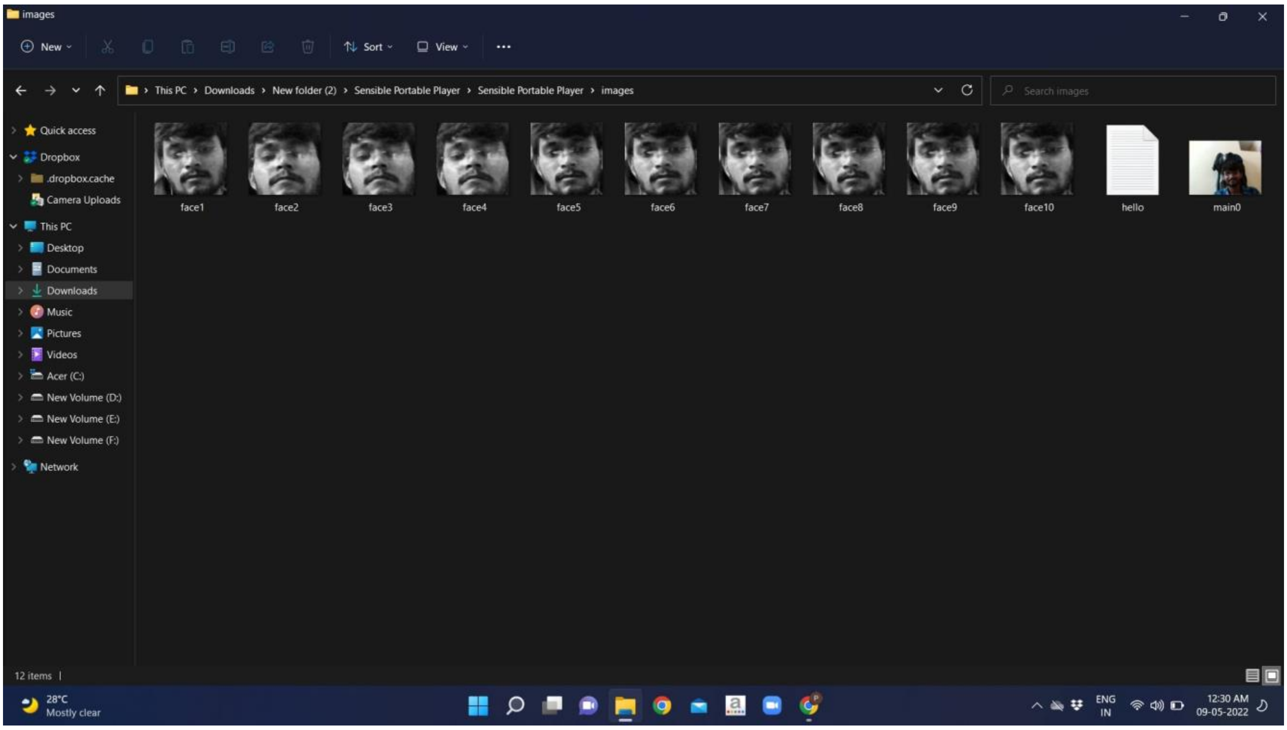
VII. EXPERIMENT RESULTS AND ANALYSIS

This study provides a music recommendation system that uses a camera attached to the computer platform to extract the user's picture. The classifier that is used to identify the face in a photograph after it has been taken is the name of the classifier that is used to transform th e collected frame of the image from the webcam feed to a grayscale image to improve performance.. The image is then supplied to the classifier algorithm, which classifies the image using feature extraction. extraction Procedures may be used to remove a person's face from the frame of a web camera broadcast. Once the face has been extracted, individual elements of the face are collected and passed to the trained network to determine the user's emotion.

The instructions for the user have been explained. The users were given instructions on how to anticipate the stated emotion in this situation, which resulted in the following findings. When the inner feeling is unhappy yet the outward appearance is joyful, it might lead to failure. The values are shown in Table 1, and the result is shown in;

Fig. The User's Instructions Explained

User	Emotion	Facial Expression	Accuracy
1	Happy	Happy	100
2	Sad	Happy	0
3	Happy	Happy	100
4	Sad	Sad	100



VIII. CONCLUSION

The recognition of human emotions using facial expressions has a wide variety of applications. It takes care of the time-consuming chore of choosing the right song for each occasion based on the person's mood. This paper gives an overview of the many techniques and approaches that have been suggested and developed to recognise human emotional states of behaviour while listening music, as well as an abstract image of the proposed system that we will design...

IX. ACKNOWLEDGMENT

We'd like to express our sincere gratitude to **Dr. Uma Thakur** for her invaluable guidance and continuous support. We also want to thank **Dr. S.A. Dr. Leena Patil**, the HOD of the Computer Science and Engineering Department, and **Dr. Dhale**, the principal of Priyadarshini College of Engineering **Nilesh Shelke**, the Project In-charge, for their kind cooperation, valuable guidance, and constant motivation, as well as for providing the necessary infrastructure and all the facilities for the project's development. We would also like to extend our gratitude to all of the faculty members in the department and institution. as well as all non-teaching personnel, for their assistance throughout the project. Finally, a big thank you to the researchers whose study pointed us in the right direction.

X. REFERENCES

V. Patchava, P. Jain, R. Lomte, P. Shakthi, H. B. Kandala, “Sentiment Based Music Play System”.

1. Viral Prasad and Aurobind V. Iyer, "Emotion Based Mood Enhancing Music Recommendation," in 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information, and Communication Technology (RTEICT), India, pp. 1573-1577, May 19-20, 2017.
2. Harshala Chaudhari, Amrapali Waghmare, Reshma Ganjewar International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 10, May 2015, Dr. Abhijit Banubakode A media player that is controlled by Human Emotions.
3. "Facial Expression Recognition based on Local Region-Specific Features and SVM" by Deepak Ghimire, Sung Wan Jeong, Joonwhoan Lee, and Sang Hyun Park in Multimedia Tools and Applications, Vol.76, Issue 6, pp. 7803–7821, March 2017
4. Mary Duenwald (2005). The Physiology of Facial Expressions. Retrieved on October 9 2012 from.
5. Frijda, N.H. (1986): The emotions. New York: Cambridge University Press.
6. Maria M. Ruxanda1, Bee Yong Chua, Alexandros Nanopoulos, Christian S. Jensen. (2007): In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Emotion-Based Music Retrieval on a

Well-Reduced Audio Feature Space was published (ICASSP): 181-184.

7. Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, Eva Cerezo1, Isabelle Hupont, Critina Manresa, Javier Varona, Sandra Baldassarri, Francisco J. Perales, Eva Cerezo1, Isabelle Hu. Perales, and Francisco J. Seron. (2007): Real-Time Facial Expression Recognition for Natural Interaction: In J. Martí et al. (Eds.): IbPRIA 2007, Part II, LNCS 4478, (pp. 40–47). Springer-Verlag Berlin Heidelberg.

8. Ekman, P. & Friesen, W. V. (1969): The repertoire of nonverbal behavior: Categories, origins, usage, and coding: Semiotica, Vol. 1: 49–98.

I. STOCK PHOTOGRAPHY [IMAGE]. RETRIEVED ON OCTOBER 9 2012 FROM HTTP://WWW.DREAMSTIME.COM.