STM JOURNALS

STM Journal
ISSN

www.stmjournals.com

Facial Emotion Based Music Recommendation System Using CNN.

[1]Ankita Gholap, [2]Bachkar Tejashri, [3]Shubham Patil, [4]Dr.S.U.Ghumbre
Department of Computer Engineering,
Government College of Engineering and Research, Avasari Khurd, Pune, India

Abstract:

In the today's era the field of technologies and intelligence provide solutions on various problems which faced by individuals every day. This paper is focusing on the human emotion based music recommendation. Music or songs are the powerful tools to describe human emotions and it has the power to change the mood, attitudes, and even behavior of users. Music has the ability to trigger various feelings, spanning from happiness and enthusiasm to sorrow and gloominess. Music become an integral part of individuals and emotion-based music recommendation system have potential to enhance the music listening experiences by providing more personalized and meaningful recommendation to users. The research has focused on identifying and changing emotions of the user through the music which help to increase user's engagement and satisfaction. Music is used for refreshing mood and therapy for changing the psychological and physiological states of human mind that responsible for enjoyment and entertainment to made one's life comfortable. Face is important aspect in capturing the human emotions. Deep learning used for mood modulation, genre classification, volume modulation. The objective of this research paper is to detect the current emotion of user through web camera and suggest songs according to emotion detected.

Keywords: Emotion prediction, Face detection, Music recommendation, Convolutional neural network(CNN), Image processing, OpenCV.

E-mail:
tejashri.bachkar122020@gcoeara.ac.in
ankita.gholap122020@gcoeara.ac.in
shubham.patil122019@gcoeara.ac.in
shashi.ghumbre@gmail.com

1. INTRODUCTION

Music has the power to evoke emotions, memories, and moods. It helps to relieve feelings of anxiety and depression. It is often used as a tool to regulate the emotions for example; happy, sad, energetic, neutral, angry, fear. However, huge amount of content available, users often face the challenge of finding music that suits their current emotional state or mood.To address this challenge, researchers and developers have been working on emotion based music recommendation system. This system uses machine learning and data mining techniques to analyze the user's current emotional state, to provide personalized music recommendations. By leveraging the emotional content of music and user data, the system can provide more relevant and meaningful music recommendations to users. The purpose of the research paper is to investigate the state-of-the-art in emotion-based music recommendation systems. The paper discusses about the theoretical foundations, design principles, and evaluation metrics used in these system.Dataset used for emotions detection has taken from Kaggle i.e. Facial Expression Recognition FER2013. FER contains 7 emotions data, out of them happy, sad, neutral and angry chosen for system implementation which contains 21,005 images of training dataset and 5,212 images for testing dataset. Dataset for the music player has been created from Spotify that categorized into happy, sad, neutral and angry emotions based playlist. Implementation of facial emotion detection is performed using Convolutional Neural Network, which gives 68 % of training and 70% of testing accuracy.

2. PREVIOUS RESEARCHES:

It was observed in previous researches that includes, The study presented by Ahlam Alrihaili ,Alaa Alsaedi ,Kholood Albalawi and Liyakathunisa Syed [1] introduces a recommender system for recognizing user emotions and suggesting appropriate music tracks to improve their mood. The proposed system is capable of detecting negative emotions and offering a playlist of suitable music tracks to improve the user's mood. Alternatively, if the system detects positive emotions, it will suggest a playlist of diverse music tracks to enhance positive emotions. The proposed system was implemented using the Viola-Jonze algorithm and Principal Component Analysis (PCA) techniques, and successfully developed in MATLAB(R2018a). Arushi Raghuvanshi and a Vivek Choksi [2], the convolutional neural network (CNN) in deep learning is used for performing the facial expression recognition. The seven layer CNN trained using FER-2013 dataset obtained an accuracy of 48%. The study presented by Kunjal Gajjar and Siddhi Shah [3] in the article is centered around the development of an automated system for generating Hindi popular music playlists based on user mood, with minimal user intervention. The system comprises two key modules. The initial module identifies the user's mood, taking cues from social media platforms and messaging applications, such as WhatsApp. The second module involves the tagging of songs from an existing database, which is done based on a variety of factors, including Genre, Artists, Tempo, and Lyrics. By adding these modules, the system can create mood-based playlists tailored to the user's preferences. Vijay Chakole, Aniket Choudhary, Kalyani Trivedi, Kshitija Bhoyar, Ruchita Bodele, Sayali Karmore [4] in Research Article "Emotion Based Music Player" they used Viola and Jones algorithm for face detection and facial features extraction. The conversion audio files (i.e. from Stereo signalsto 16 bit PCM mono signal) is done using Audacity technique. For classification of emotions and song into different categories they used Support Vector Machine (SVM). Surekha Samsani and Vineel Abhinav Gottala [5] also conducted experiments using the convolutional neural network. The model was trained using the FER-2013 dataset. The model was also compared with the other pre -trained models. The proposed CNN model achieved an accuracy of 61.4%.

3. RESEARCH METHODOLOGY

Objectives

1. To detect face using OpenCV.
2. To extract and classify emotion using CNN.
3. To recommend emotions based music playlist.
4. To play the song which are based on the emotions of the users.
5. To reduce time and efforts for making manual playlist.

3.1 Database Description:

The FER-2013 dataset is a popular benchmark dataset in the field of computer vision and facial expression recognition. It consists of 35,887 grayscale images of size 48x48 pixels, with each image belonging to one of seven different emotion categories: anger, disgust, fear, happiness, sadness, surprise, and neutral. The FER-2013 dataset has been widely used to develop and evaluate machine learning models for facial expression recognition. It is often used as a benchmark dataset for academic research and has been used to evaluate deep learning models, convolutional neural networks (CNNs), and other machine learning algorithms. But in this project we focused on just four emotion anger, happy, sad and neutral. So after removing other remaining emotion related images dataset consists of 21,005 grayscale images belongs to four categories. Then dataset will be further divided into to two parts i.e. training and testing dataset.

Fig 1. Type of Emotions

3.2 Emotion Detection Module:

3.2.1 Face Detection:

OpenCV (Open Source Computer Vision Library) is a popular open-source computer vision and machine learning library that provides pre-built functions and algorithms for various image and video processing tasks, including face detection. OpenCV includes several pre-trained classifiers for face detection, such as Haar Cascades, Local Binary Patterns (LBP), and Deep Neural Networks (DNN). The Haar Cascade classifier is the most commonly used algorithm for face detection in OpenCV. The haarcascade_frontalface_default.xml is commonly used and pre-trained Haar Cascade classifier for face detection provided by OpenCV.

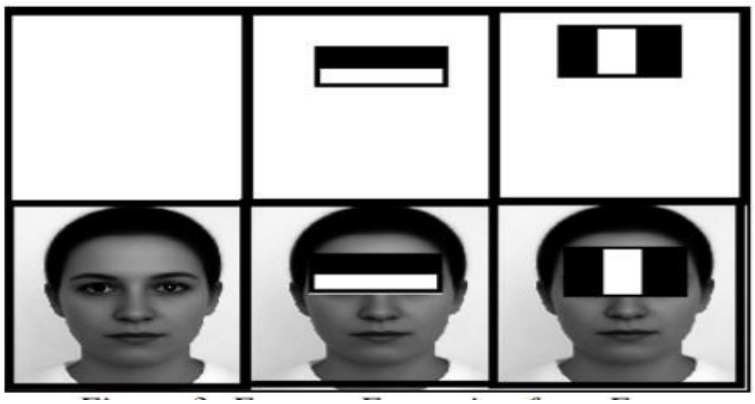

Fig 2. face detection.

The haarcascade_frontalface_default.xml classifier is trained on a large dataset of positive and negative images and has been optimized to achieve high accuracy in detecting frontal faces in various lighting conditions, angles, and poses. It is widely used in various face detection applications and has become a standard benchmark for evaluating the performance of face detection algorithms. The haarcascade_frontalface_default.xml classifier is trained to detect frontal faces in an image. It uses Haar-like features, which are simple rectangular features that describe the contrast between neighboring regions of an image. The classifier works by scanning an image with a sliding window of various sizes and positions and checking whether the features of a face are present in that window. If the features match, the window is marked as a potential face. OpenCV also provides several functions for drawing rectangles around the detected faces and displaying the image or video stream with the detected faces. Additionally, OpenCV can be used in combination with other libraries and frameworks, such as NumPy and Tensor Flow, to build more complex applications for face detection and recognition.

3.2.2 Feature Extraction

Convolutional Neural Networks (CNNs) are a type of deep learning algorithm that are commonly used for image recognition and analysis tasks, including face recognition and emotion detection. In the context of emotion detection from faces, a CNN can be used to extract features from facial images that are then used to classify the emotion expressed in the image. The process of feature extraction involves passing the input image through a series of convolutional layers, which apply a set of filters to the image and produce a set of feature maps. Each feature map represents a specific aspect of the image, such as the presence of edges or textures, and is the result of applying a convolutional operation to the input image.

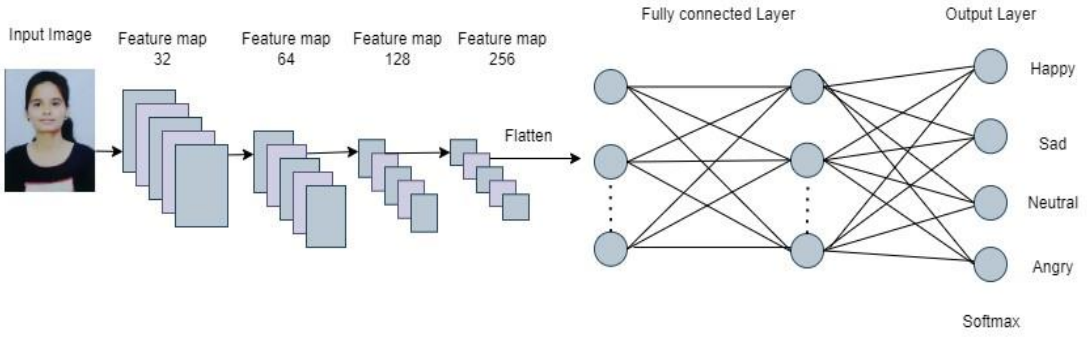

Fig 3. Feature Extraction using CNN.

3.2.2 Emotion Detection:

After the convolutional layers, the feature maps are flattened into a vector and passed through one or more fully connected layers, which perform a non-linear transformation of the features to produce an output vector. The output vector represents the probability of each possible emotion label for the input image. During the training phase, the CNN is trained on a dataset of labeled facial images, where each image is associated with a corresponding emotion label. The CNN learns to adjust the weights of its filters and fully connected layers in order to minimize the difference between the predicted emotion label and the true emotion label of each image in the training set. Once the CNN is trained, it can be used to predict the emotion label of new facial images. The process of emotion detection involves passing the input image through the CNN and obtaining the output vector of emotion probabilities. The predicted emotion label is then the label with the highest probability in the output vector.

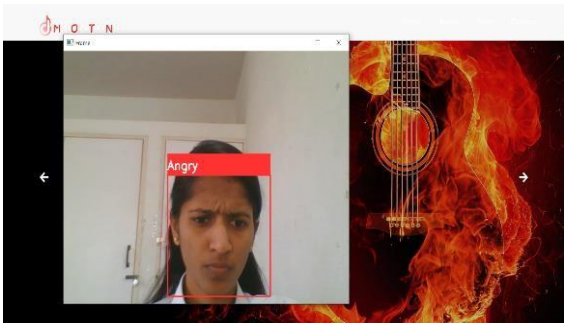

Fig 4. Angry Emotion

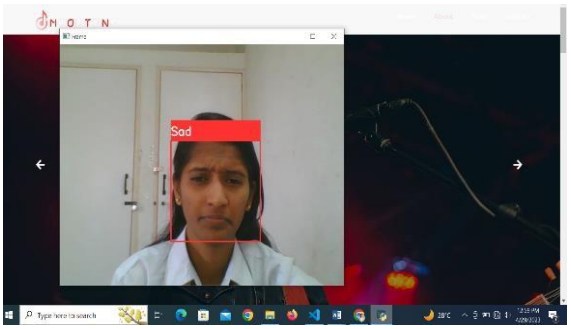

Fig 5. Sad Emotion

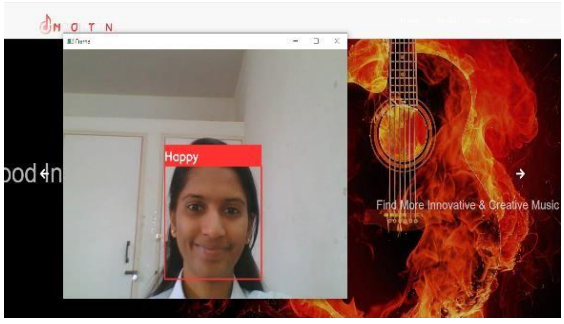

Fig 6. Happy Emotion

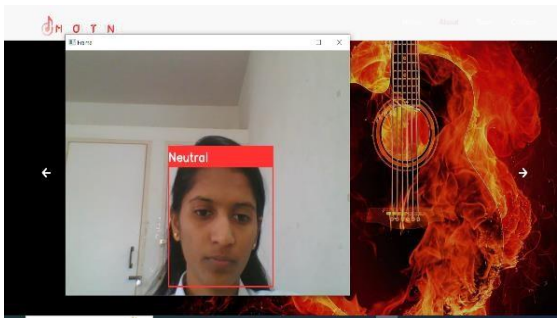

Fig 7. Neutral Emotion

3.3 MUSIC RECOMMENDATION MODULE:

There are many techniques used for music recommendation that includes content-based and collaborative techniques. Content-based filtering selects information based on semantic content and collaborative based filtering gathers opinion of users and make predictions. For music recommendation, playlist categorized from Spotify which includes Hind, Marathi songs. System uses Spotify API to fetch song details such as song name, image, artist name, etc. if

emotion detected is happy then happy song playlist displayed on GUI. Similarly, for angry, sad and neutral playlist shown. In this system the real time face detected by uses of Haar Cascade classifier which provided by OpenCV library. Result of recommended playlist displayed on GUI and based on emotion detected system plays song. GUI contains song list, detected emotion name, functions like, play, pause, next, previous to perform action on buttons in music player. Music plays important role in day to day life. So suppose user is sad then system recommends playlist of sad and motivational songs which automatically change the state of user mood. Similarly, songs recommended for happy, neutral, angry as mentioned below.

Emotion Detected	**Song Recommended**
1. Angry	Energetic + Calm
2. Happy	Happy
3. Sad	Sad + Motivational
4. Neutral	All songs

RESULT & ANALYSIS:

In Convolutional neural network model, System uses the Adam optimizer for reducing losses. 50 epochs use to train the model. The output layer is uses softmax function and other layers are uses RELU as activation function. Adam optimizer gives the best results as compared to other optimizers. We achieved accuracy of 68% for training and 70% for the testing.

Hyperparameters	**Values**
Batch Size	64
No. of Classes	4
Optimizer	Adam
Learning Rate	0.001
Epoch	50
Activation function	Relu, SoftMax
Loss function	Categorical-cross entropy

In Fig (a), the graph displays the accuracy of model, where the x-axis specifies the number of epochs and the y-axis specifies the accuracy. As it can be seen in the figure, model has achieved approximately 70% accuracy. (b) The graph displays the training and validation loss of our model, where the x-axis specifies the number of epochs and the y-axis specifies the loss.

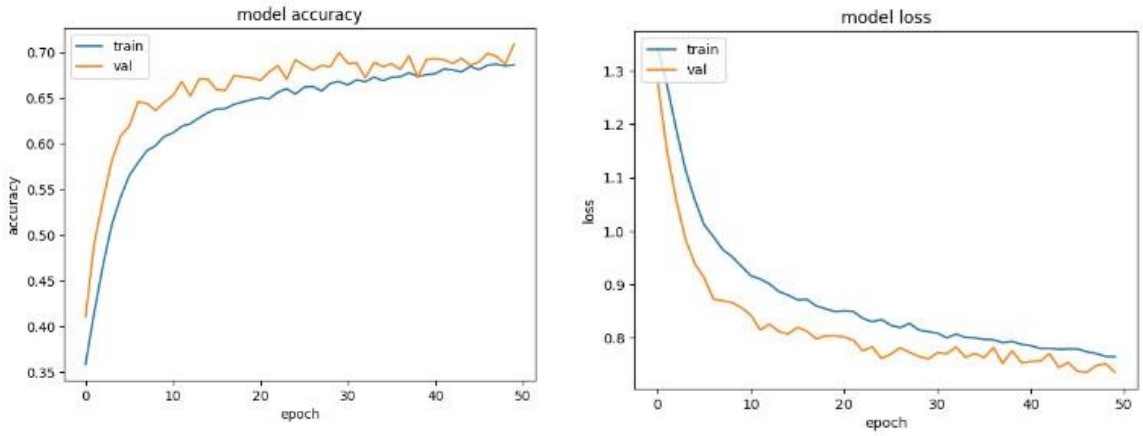

Fig 8. (a) Accuracy (b) Loss and graph for Genre Classification using CNN Model.

OUTPUT IMAGES:

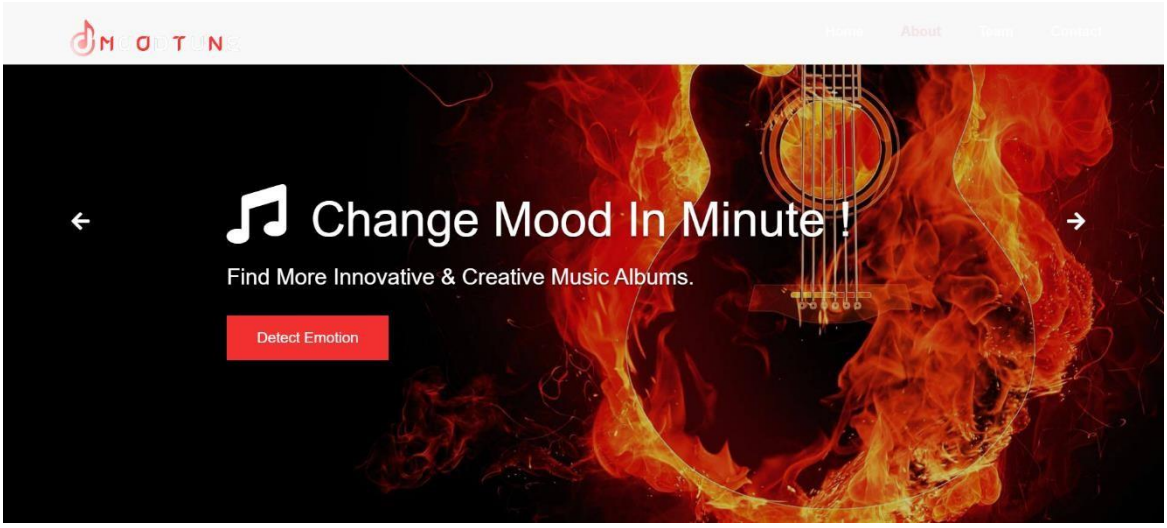

Fig 9. Home page

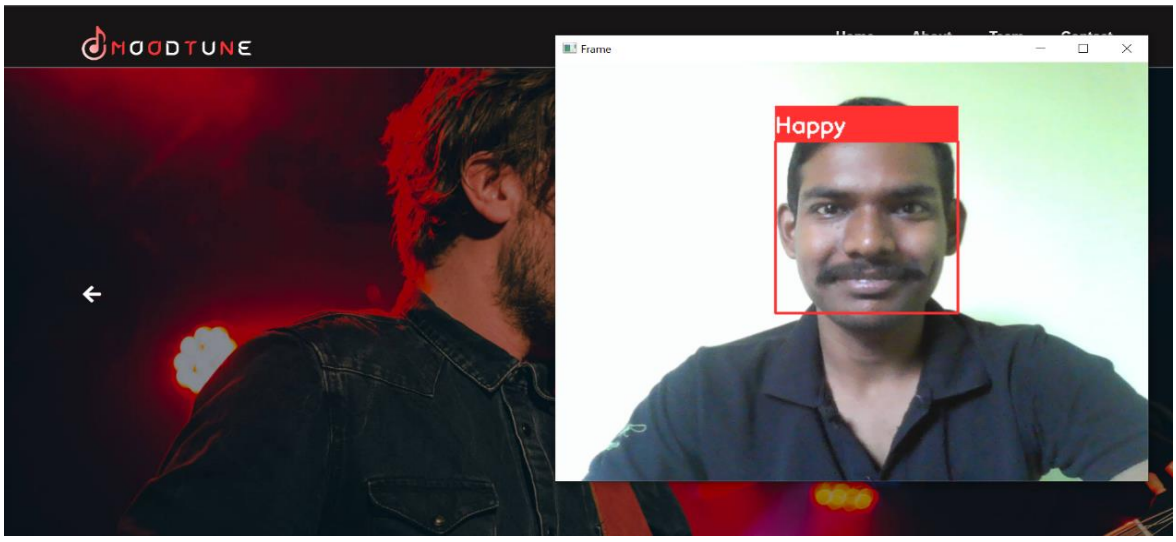

Fig.10 Emotion detection

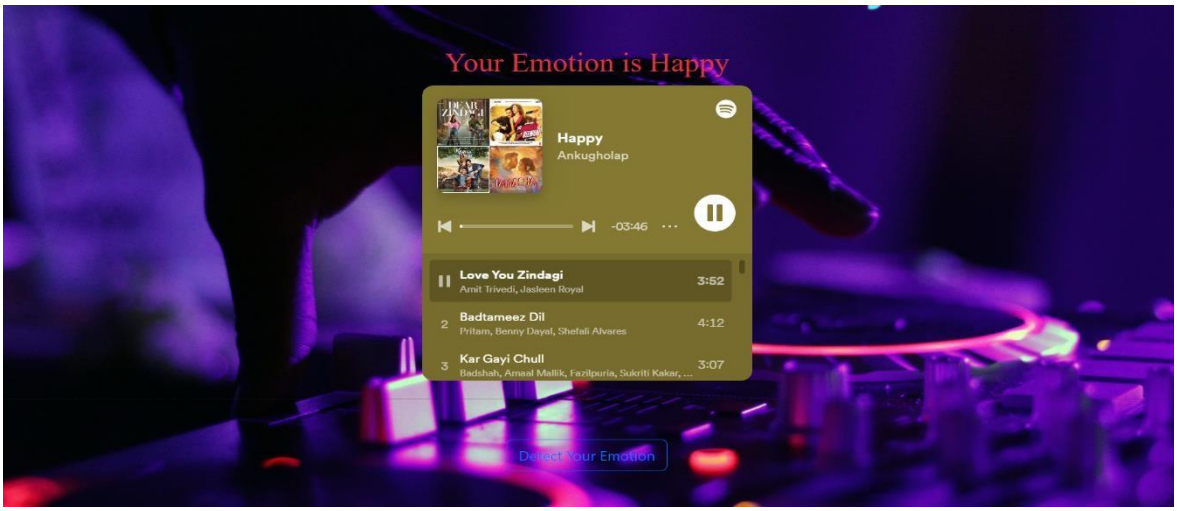

Fig 11. Music recommendation

Confusion Matrix:

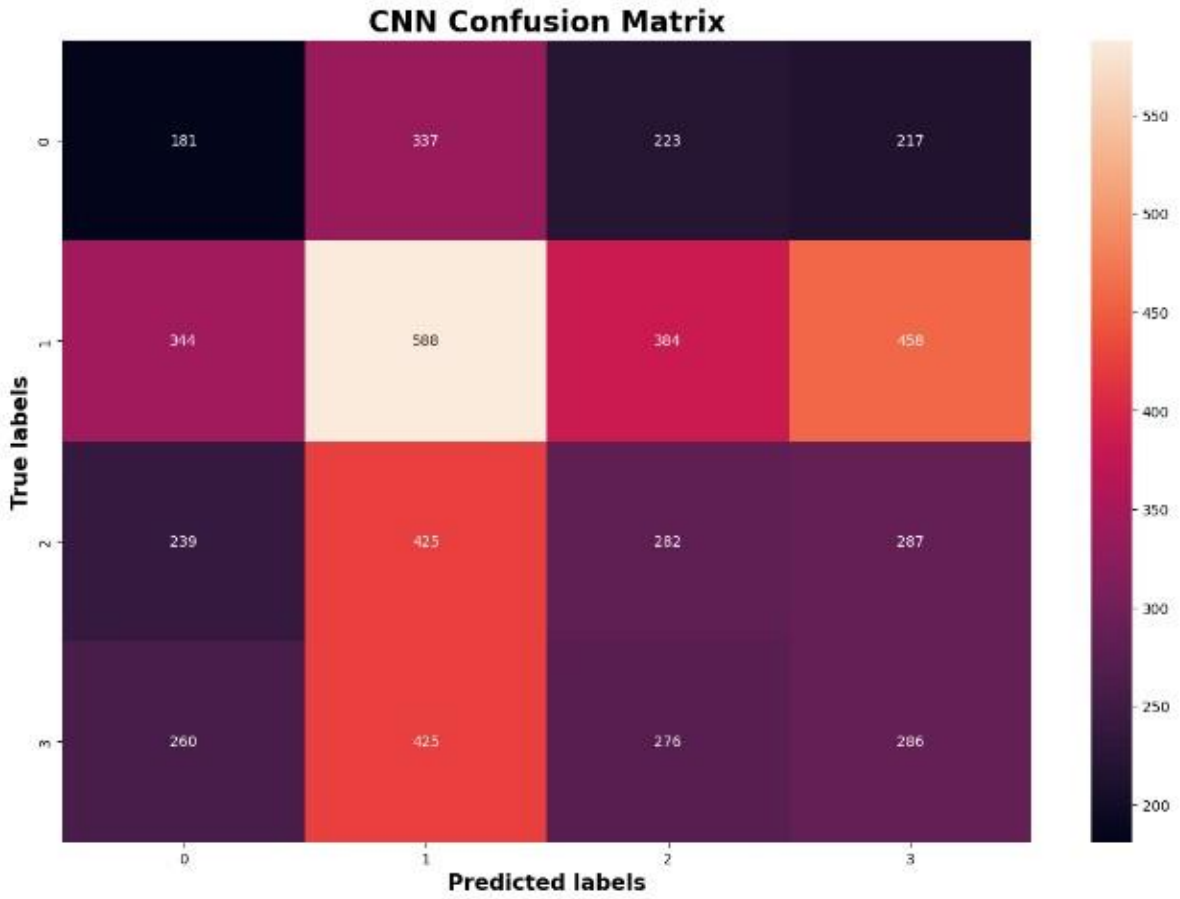

Fig. 12 Confusion Matrix

The confusion matrix represents the performance evaluation of the emotion detection component of the emotion-based music recommendation system. The confusion matrix is a table that summarizes the performance of a classification model by showing the predicted labels against the actual labels.

Label	**Emotion**
0	Angry
1	happy
2	sad
3	Neutral

CONCLUSION

In the research project, feature based recommendation system used which provides a way to classify songs into happy, sad, neutral, angry. Facial expressions are captured through web camera and extracting the features of emotions by using CNN. Thus, the present system helps for face(expressions) detection and Music plays according to current mood detected by web camera. The project is designed for the purpose of making better interaction between the music system and the users. The system is helpful in changing the mood and stress reliever of the individuals. This project fulfilled objectives that shows a wide prospective in the developing the emotion based on the music recommendation system. The proposed model reduces time and efforts of users for creating manual playlist. The system provides enjoyable and interactive platform for users with approximately 68% of training and 70% of testing accuracy.

ACKNOWLEDGMENTS

We would like to express our gratitude to all those who have contributed to the successful completion of our Bachelor of Engineering project. We would like to start by thanking our Head of Department (HOD) who is also our project guide Dr.S.U.Ghumbre for his guidance and constant support throughout the project. Their constant encouragement, insightful feedback and willingness to share their knowledge and expertise have been instrumental in the successful completion of this project.

We would also like to extend our thanks to all the teaching and non-teaching staff, who have provided us with the necessary resources, facilities, and support to complete this project successfully.

Finally, we would like to acknowledge our families and friends for their support and encouragement throughout this project. Their unwavering support and belief in us have been an important source of motivation and inspiration.

Thank you to all who have contributed to the development of this project.

REFERENCES

[1] Ahlam Alrihaili,Alaa Alsaedi ,Kholood Albalawi , Liyakathunisa Syed."Music Recommender System for Users Based on Emotion Detection throughFacial Features"IEEE 2019 12th International Conference on Developments in eSystems Engineering (DeSE) ,7-10 Oct. 2019.

[2] Aniket Jha,Sagar Gupta , Priyanshu Dubey , Aditi Chhabria ," Music Feature Extraction And Recommendation Using CNN Algorithm", ITM Web of Conferences 44, 03026 (2022).

[3] Arushi Raghuvanshi and Vivek Choksi, "Facial Expression Recognition with Convolutional Neural Networks"CS231n Course Projects Winter, 2016.

[4] Kunjal Gajjar,Siddhi Shah . " Mood based Playlist Generation for Hindi Popular Music:A Proposed Model " ,International Journal of Computer Applications (0975 – 8887)Volume 127 – No.14, October 2015.

[5] Nikhil Zaware,Tejas Rajgure,Amey Bhadang , Prof. D. D. Sapka, "Emotion Based Music Player", Vol 3 Issue 3, March 2014.

[6] Prof. Vijay Chakole, Aniket Choudhary, Kalyani Trivedi, Kshitija Bhoyar, Ruchita Bodele and Sayali Karmore ,"Emotion Based Music Player"Research Article IJESC Volume 8 Issue No.3 ,2018.

[7] Samsani, S., Gottala, V.A. "A real-time automatic human facial Expression recognitiom system using deep neuralnetworks", Information and Communication Technology for Sustainable Development.