iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.47996

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Emotion Based Music Recommendation System

Aryan Shirwadkar[1], Pratham Shinde[2], Sahil Desai[3], Samuel Jacob[4]
[1, 2, 3, 4] *Information Technology, Vidyalankar Institute of Technology, Mumbai*

Abstract: ***We propose a new approach for playing music automatically using facial emotion. Most of the existing approaches involve playing music manually, using wearable computing devices, or classifying based on audio features. Instead, we propose to change the manual sorting and playing. We have used a Convolutional Neural Network for emotion detection. For music recommendations, Pygame & Tkinter are used. Our proposed system tends to reduce the computational time involved in obtaining the results and the overall cost of the designed system, thereby increasing the system's overall accuracy. Testing of the system is done on the FER2013 dataset. Facial expressions are captured using an inbuilt camera. Feature extraction is performed on input face images to detect emotions such as happy, angry, sad, surprise, and neutral. Automatically music playlist is generated by identifying the current emotion of the user. It yields better performance in terms of computational time, as compared to the algorithm in the existing literature***
Keywords: Emotion Detection, Face Recognition, Music

I. INTRODUCTION

Many of the studies in recent years admit that humans reply and react to music and this music has a high impression on the activity of the human brain. In one examination of the explanations why people hear music, researchers discovered that music played a crucial role in relating arousal and mood. Two of the most important functions of music are it is ability is participants rated to help them achieve a good mood and become more self-aware. Musical preferences have been demonstrated to be highly related to personality traits and moods.

The meter, timbre, rhythm, and pitch of music are managed in areas of the brain that affects emotions and mood. Interaction between individuals may be a major aspect of lifestyle. It reveals perfect details and much of data among humans, whether they are in the form of body language, speech, facial expression, or emotions. Nowadays, emotion detection is considered the most important technique used in many applications such as smart card applications, surveillance, image database investigation, criminal, video indexing, civilian applications, security, and adaptive human-computer interface with multimedia environments. With the increase in technology for digital signal processing and other effective feature extraction algorithms, automated emotion detection in multimedia attributes like music or movies is growing rapidly and this system can play an important role in many potential applications like human-computer interaction systems and music entertainment.

We use facial expressions to propose a recommender system for emotion recognition that can detect user emotions and suggest a list of appropriate songs.

The proposed system detects the emotions of a person, if the person has a negative emotion, then a certain playlist will be shown that includes the most related types of music that will enhance his mood. And if the emotion is positive, a specific playlist will be presented which contains different types of music that will inflate the positive emotions. The dataset we used for emotion detection is from Kaggle Facial Expression Recognition. Dataset for the music player has been created from Bollywood Hindi songs. Implementation of facial emotion detection is performed using Convolutional Neural Network which gives approximately 95.14% of accuracy

II. PROBLEM STATEMENT

In old-style music players, a user had to manually browse through the playlist and select songs that would soothe his mood. In today's world, with ever increasing advancements in the field of multimedia and technology, various music players have been developed with features like fast forward, reverse, variable playback speed, local playback, streaming playback with multicast streams and including volume modulation, genre classification etc. These features may satisfy the user's basic requirements, but the user has to face the task of manually browsing through the playlist of songs and select songs based on the current mood and behaviour.

That is the requirement of an individual, a user sporadically suffered through the need and desire of browsing through his playlist, according to his mood and emotions

©IJRASET: All Rights are Reserved

III. LITERATURE SURVEY

It was observed in a cross-database experiment [1] that raw features worked best with Logistic Regression for testing RaFD (Radboud Faces Database) database and Mobile images dataset. The accuracy achieved was 66% and 36% respectively for both using CK+ dataset as a training set. The additional features (distance and area) reduced the accuracy of the experiment for SVM (Support Vector Machine) from 89%. The algorithm that had been implemented generalized the results from the training set to the testing set better than SVM and several other algorithms. An average accuracy of 86% was seen for RaFD database and 87% for CK+ database for cross-validation=5.The main focus was feature extraction and analysis of the machine algorithm on the dataset. But accurate face-detection algorithms become very important if there are multiple people in the image. One of the works [10] was tested by deriving expression from the live feed via the system's camera or any pre-existing image available in the memory. It has been implemented using Python 2.7, OpenCV and NumPy. The objective was to develop a system that can analyse the image and predict the expression of the person. The study proved that this procedure is workable and produces valid results.

There has also been research done on the Music Recommendation System. According to one such research [11], a preliminary approach to Hindi music mood classification has been described, that exploits simple features extracted from the audio. MIREX (Music Information Retrieval Evaluation eXchange) mood taxonomy gave an average accuracy of 51.56% using the 10-fold cross validation. In addition to this, there is an article [10] that states that the current music recommendation research results from the perspective of music resources description. It is suggested that there is a lack of systematic research on user behaviour and needs, low level of feature extraction, and a single evaluation index in current research. Situation was identified to be an important factor in the music personalized recommendation system. Finally, it was concluded that when the weights given to all the contextual factors were the same, greatly reduced the accuracy of the recommendation results.

In a particular system [8], Anaconda and Python 3.5 softwares were used to test the functionality and Viola-Jones and haar cascade algorithms were used for face detection. Similarly, KDEF (Karolinska Directed Emotional Faces) dataset and VGG (Visual Geometry Group) 16 were used with CNN (Convolution Neural Network) model which was designed with an accuracy of 88%, for face recognition and classification that validated the performance measures. However, the results proved that the network architecture designed had better advancements than existing algorithms. Another system [9] used Python 2.7, OpenSource Computer Vision Library (OpenCV) & CK (CohnKanade) and CK+ (Extended Cohn-Kanade) database which gave approximately 83% accuracy. Certain researchers have described the Extended Cohn-Kanade (CK+) database for those wanting to prototype and benchmark systems for automatic facial expression detection. The popularity and ease of access for the original Cohn-Kanade dataset this is seen as a very valuable addition to the already existing corpora. It was also stated that for a fully automatic system to be robust for all expressions in a myriad of realistic scenarios, more data is required. For this to occur very large reliably coded datasets across a wide array of visual variabilities are required (at least 5 to 10k examples for each action) which would require a collaborative research effort from various institutions.

IV. METHODOLOGY

We built the Convolutional Neural Network model using the Kaggle dataset. The database is FER2013 which is split into two parts training and testing dataset. The training dataset consists of 24176 and the testing dataset contains 6043 images. There are 48x48 pixel grayscale images of faces in the dataset. Each image in FER-2013 is labeled as one of five emotions: happy, sad, angry, surprise, and neutral.

The faces are automatically registered so that they are more or less centered in each image and take up about the same amount of space. The images in FER-2013 contain both posed and unposed headshots, which are in grayscale and 48x48 pixels. The FER-2013 dataset was created by gathering the results of a Google image search of every emotion and synonyms of the emotions. FER systems being trained on an imbalanced dataset may perform well on dominant emotions such as happy, sad, angry, neutral, and surprised but they perform poorly on the under -represented ones like disgust and fear.

Usually, the weighted-SoftMax loss approach is used to handle this problem by weighting the loss term for each emotion class supported by its relative proportion within the training set. However, this weighted-loss approach is predicated on the SoftMax loss function, which is reported to easily force features of various classes to stay apart without listening to intra-class compactness. One effective strategy to deal with the matter of SoftMax loss is to use an auxiliary loss to coach the neural network.

To treating missing and Outlier values we have used a loss function named categorical crossentropy. For each iteration, a selected loss function is employed to gauge the error value. So, to treating missing and Outlier values, we have used a loss function named categorical crossentropy.

A. Face Detection

Face detection is one of the applications which is considered under computer vision technology. This is the process in which algorithms are developed and trained to properly locate faces or objects in object detection or related system inimages. This detection can be real-time from a video frame or images. Face detection uses such classifiers, which are algorithms that detect what's either a face (1) or not a face (0) in an image. Classifiers are trained to detect faces using numbers of images to get more accuracy. OpenCV uses two sorts of classifiers, LBP (Local Binary Pattern) and Haar Cascades. A Haar classifier is used for face 11 detection where the classifier is trained with pre -defined varying face data which enables it to detect different faces accurately. The main aim of face detection is to spot the face within the frame by reducing external noises and other factors. It is a machine learning-based approach where the cascade function is trained with a group of input files. It is supported the Haar Wavelet technique to research pixels inside the image into squares by function [9]. This uses machine learning techniques to urge a high degree of accuracy from what's called "training data"

B. Feature Extraction

While performing feature extraction, we treat the pre-trained network that is a sequential model as an arbitrary feature extractor. Allowing the input image to pass on it forward, stopping at the pre-specified layer, and taking the outputs of that layer as our features. Starting layers of a convolutional network extract high-level features from the taken image, so use only a few filters. As we make further deeper layers, we increase the number of the filters to twice or thrice the dimension of the filter of the previous layer. Filters of the deeper layers gain more features but are computationally very intensive. Doing this we utilized the robust, discriminative features learned by the Convolution neural network [10]. The outputs of the model are going to be feature maps, which are an intermediate representation for all layers after the very first layer. Load the input image for which we want to view the Feature map to know which features were prominent to classify the image. Feature maps are obtained by applying Filters or Feature detectors to the input image or the feature map output of the prior layers. Feature map visualization will provide insight into the interior representations for specific input for each of the Convolutional layers within the model.

C. Emotion Detection Figure

Convolution neural network Architecture. Convolution neural network architecture applies filters or feature detectors to the input image to get the feature maps or activation maps using the Relu activation function [11]. Feature detectors or filters help in identifying various features present in the image such as edges, vertical lines, horizontal lines, bends, etc. After that pooling is applied over the feature maps for invariance to translation. Pooling is predicted on the concept that once we change the input by a touch amount, the pooled outputs don't change. We can use any of the pooling from min, average, or max. But max-pooling provides better performance than min or average pooling. Flatten all the input and giving these flattened inputs to a deep neural network which are outputs to the class of the object. 12 The class of the image will be binary, or it will be a multi-class classification for identifying digits or separating various apparel items. Neural networks are as a black box, and learned features in a Neural Network are not interpretable. So basically, we give an input image then the CNN model returns the results [10]. Emotion detection is performed by loading the model which is trained by weights using CNN. When we take the real-time image by a user then that image was sent to the pre-trained CNN model, then predict the emotion and adds the label to the image.

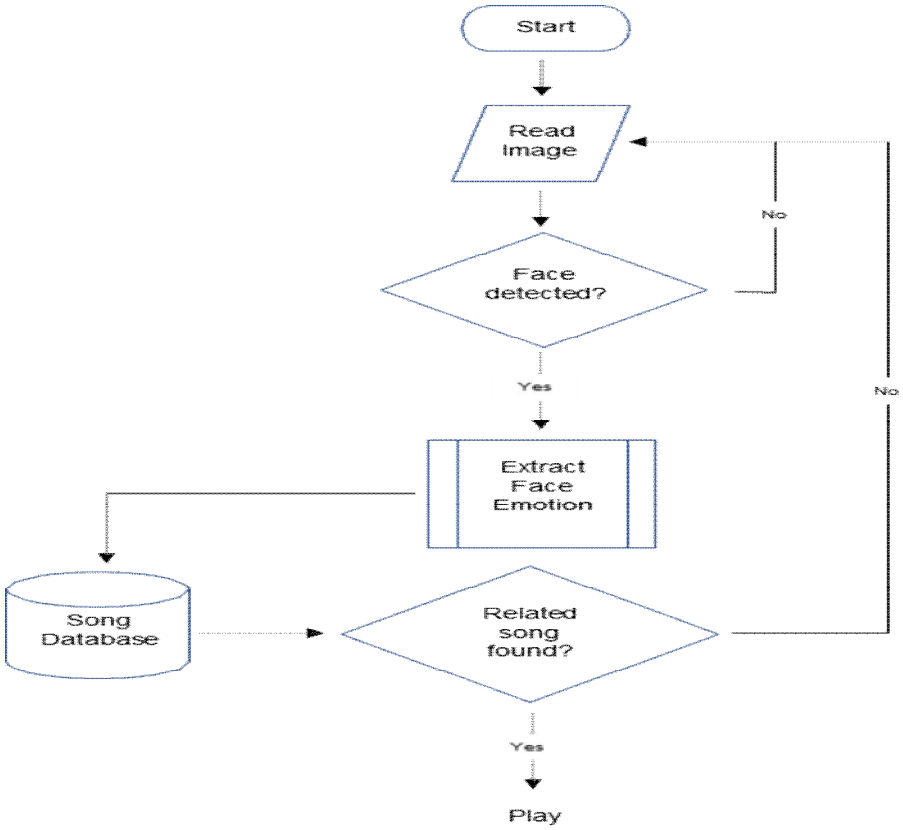
D. Music Recommendation Module

Songs Database We created a database for Bollywood Hindi songs. It consists of 100 to 150 songs per emotion. As we all know music is undoubtedly involved in enhancing our mood. So, suppose a user is sad then the system will recommend such a music playlist which motivates him or her and by this automatic mood will be delighted. Music Playlist Recommendation By using the emotion module real-time emotion of the user is detected. This will give the labels like Happy, Sad, Angry, Surprise, and Neutral. Using the os.listdir() method in python we connected these labels with the folders of the songs database which we have created. Table 1 shows the list of songs. This method of os.listdir() is used to get the list of any file in the specified directories. if label== 'Happy': os.chdir("C:/Users/deepali/Downloads/Happy") self.mood.set("You are looking happy, I am playing song for You") # Fetching Songs songtracks = os.listdir()] # Inserting Songs into Playlist for track in songtracks: self.playlist.insert (END, track) This will result in the recommended playlist for the user in the GUI of the music player by showing captions according to detected emotions. We have used a library called Pygame for playing the audio as this library supports playing various multimedia formats like audio, video, etc.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

Functions of this library such as playsong, pauseong, resumesong, and stopsong are used to working with the music player. Variables like playlist, songstatus, and root are used for storing the name of all songs, storing the status of currently active songs, and for the main GUI window respectively. For developing the GUI, we have used Tkinter.

Flow diagram of proposed system



V. HARDWARE AND SOFTWARE REQUIREMENTS

1) *Hardware Requirements:* ➢ PC/ Desktop/ Laptop/ Tablet
2) *Software Requirements:* ➢ Windows Operating System ➢ Python 3 ➢ PyCharm IDE ➢ Jupyter notebook ➢ Streamlit ➢ Excel or MySQL for database

.

VI. CONCLUSION

In this project, music recommendation model it is based on the emotions that are captured in real time images of the user. This project is designed for the purpose of making better interaction between the music system and the user because Music is helpful in changing the mood of the user and for some people it is a stress reliever. Recent development it shows a wide prospective in the developing the emotion based music recommendation system. Thus the present system presents Face (expressions) based recognition system so that it could detect the emotions and music will be played accordingly

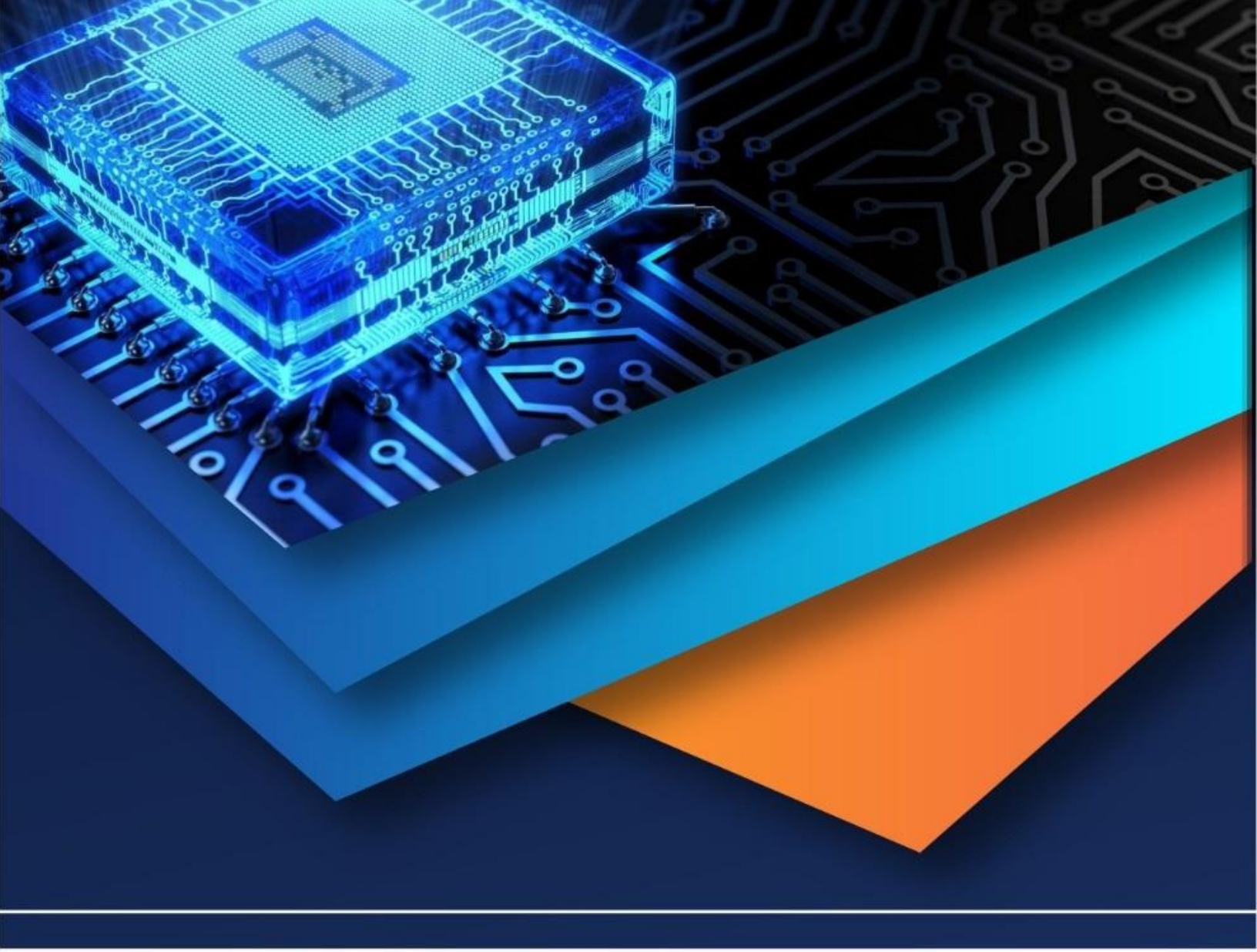
VII. ACKNOWLEDGEMENT

We would like to express our special thanks of gratitude to our guide Prof. Samuel Jacob who gave us the golden opportunity to do this wonderful project on the topic Emotion Based Music Recommendation which also helped us in doing a lot of Research and we came to know about so many new things we are thankful to him. Secondly, we would also like to thank our professors of review panel in finalizing and improving this project within the limited time frame. This project helped us in understanding the various parameters which are involved in the development of desktop application and the working and integration of front end along with the back end.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Raut, Nitisha, "Facial Emotion Recognition Using Machine Learning" (2018). Master's Projects. 632. https://doi.org/10.31979/etd.w5fs-s8wd
[2] Hemanth P,Adarsh ,Aswani C.B, Ajith P, Veena A Kumar, "EMO PLAYER: Emotion Based Music Player", International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 4, April 2018, pp. 4822-87.
[3] Music Recommendation System: "Sound Tree", Dcengo Unchained: Sıla KAYA, BSc.; Duygu KABAKCI, BSc.; Işınsu KATIRCIOĞLU, BSc. and Koray KOCAKAYA BSc. Assistant : Dilek Önal Supervisors: Prof. Dr. İsmail Hakkı Toroslu, Prof. Dr. Veysi İşler Sponsor Company: ARGEDOR
[4] Tim Spittle, lucyd, GitHub, , April 16, 2020. Accessed on: [Online], Available at: https://github.com/timspit/lucyd
[5] A. Abdul, J. Chen, H.-Y. Liao, and S.-H. Chang, "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," Applied Sciences, vol. 8, no. 7, p. 1103, Jul. 2018.
[6] Manas Sambare, FER2013 Dataset, Kaggle, July 19, 2020. Accessed on: September 9, 2020. [Online], Avalilable at: https://www.kaggle.com/msambare/fer2013
[7] MahmoudiMA, MMA Facial Expression Dataset, Kaggle, June 6, 2020. Accessed on: September 15, 2020. [Online], Available at: https://www.kaggle.com/mahmoudima/mma-facial-expression
[8] Dr. Shaik Asif Hussain and Ahlam Salim Abdallah Al Balushi, "A real time face emotion classification and recognition using deep learning model", 2020 Journal. of Phys.: Conf. Ser. 1432 012087
[9] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.
[10] Puri, Raghav & Gupta, Archit & Sikri, Manas & Tiwari, Mohit & Pathak, Nitish & Goel, Shivendra. (2020). Emotion Detection using Image Processing in Python.
[11] Patra, Braja & Das, Dipankar & Bandyopadhyay, Sivaji. (2013). Automatic Music Mood Classification of Hindi Songs.
[12] Lee, J., Yoon, K., Jang, D., Jang, S., Shin, S., & Kim, J. (2018). MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION.
[13] Kaufman Jaime C., University of North Florida, "A Hybrid Approach to Music Recommendation: Exploiting Collaborative Music Tags and Acoustic Features", UNF Digital Commons, 2014.
[14] D Priya, Face Detection, Recognition and Emotion Detection in 8 lines of code!, towards data science, April 3, 2019. Accessed on: July 12, 2020 [Online], Available at: https://towardsdatascience.com/facedetection-recognition-and-emotion-detection-in-8-lines-of-codeb2ce32d4d5de
[15] bluepi, "\Classifying Different Types of Recommender Systems, November 14, 2015. Accessed on: July 7, 2020. [Online], Availableon:https://www.bluepiit.com/blog/classifying-recommendersystems/#:~:text=There%20are%20majorly%20six%20ty pes,system %20and%20Hybrid%20recommender%20system



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)