Emotion Based Music Recommendation System

Article · May 2022

DOI: 10.15680/IJIRSET.2022.1105032

4 authors, including:

Harsh Raj Singh
Indian Institute of Technology BHU
11 PUBLICATIONS **32** CITATIONS

SEE PROFILE

IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 5, May 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

Emotion Based Music Recommendation System

Harsh Raj Singh[1], Atul Upadhyay[2], Mayuresh Sutar[3], Dr. Najmuddin Aamer[4]

UG Students, Dept. of Computer Engineering, TCOE, Mumbai University, Maharashtra, India[1,2,3]

Professor, Dept. of Computer Engineering, TCOE, Mumbai University, Maharashtra, India[4]

ABSTRACT: Listening to music affects the human brain activities. Emotion based music player with automated playlist can help users to maintain a particular emotional state. This research proposes an emotion based music player that creates a playlists based on captured photos of the user. Manual sorting of a playlist and annotation of songs, in accordance with the current emotion, is more time consuming and quite tedious. Numerous algorithms have been implemented to automate this process. However, existing algorithms are slow, increase cost of the system by using additional hardware and have quite very less accuracy. This paper presents an algorithm that not only automates the process of generating an audio playlist, but also to classify those songs which are newly added and the main task is to capture current mood of person and to play song accordingly. This enhances the system's efficiency, faster and automatic. The main goal is to reduce the overall computational time and the cost of the designed system. It also aims at increasing the accuracy of the system. The most important goal is to make change the mood of person if it isa negative one such as sad, depressed. This model is validated by testing the system against user dependent and user independent dataset.

KEYWORDS: Convolution neural network, Long Short-term memory, Emotion detection, audio classification, hidden layers, Max-pooling.

I. INTRODUCTION

Expressing and recognizing emotions of human are very much important in communication system. Human beings have the ability to express and recognize emotions. Computer seeks to identify the human emotions either by image analysis or through sensors. In our day to day life and in our professional life we interact with many people face to face or indirectly by phone calls, sometimes it is necessary for people to be aware of their present emotions of the person with whom they are interacting. Human emotions are classified as: surprise, fear, anger, happy, sad, disgust and neutral.

Facial movement and the tone of speech play a major role in expressing emotions. The physique and tone of the face tells the energy in the utterance of speech, which can be firstly modified to communicate different feelings. Humans can easily recognize these changes in signals along with the information felt by any other sensory organs. This project analyses the use of image or sensors or speech to capture the emotions.

Music plays a vital role in enhancing an individual's life as it is an important medium of entertainment for music lovers and listeners and sometimes even imparts a therapeutic approach. ―Where words fail music speaks‖, and hence it can change person's negative emotion simultaneously and slowly into a positive mood.

Emotions can be expressed through gestures, speech, facial expressions, body language etc. For the system to understand the user's mood, we use facial expression. Using the mobile device's camera, we can capture the user's facial expression. There are many emotion recognition systems which take captured image as input and determine the emotion. For this application, we are using neural networks for recognition of emotion.

II. LITERARTURE REVIEW

S Metilda Florence and M Uma (2020) proposed a paper "Emotional Detection and Music Recommendation System based on User Facial Expression" where the proposed system can detect the facial expressions of the user and based on his/her facial expressions extract the facial landmarks, which would then be classified to get a particular emotion of the user. Once the emotion has been classified the songs matching the user's emotions would be shown to the user. It could assist a user to make a decision regarding which music one should listen to helping the user to reduce his/her stress

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

levels. The user would not have to waste any time in searching or to look up for songs. The proposed architecture contained three modules, namely, Emotion extraction module, Audio extraction module and Emotion-Audio extraction module. Although it had some limitations like the proposed system was not able to record all the emotions correctly due to the less availability of the images in the image dataset being used. The image that is fed into the classifier should be taken in a well-lit atmosphere for the classifier to give accurate results. The quality of the image should be at least higher than 320p for the classifier to predict the emotion of the user accurately. Handcrafted features often lack enough generalizability in the wild settings.

H. Immanuel James, J. James Anto Arnold, J. Maria Masilla Ruban, M. Tamilarasan (2019) proposed "Emotion Based Music Recommendation" which aims at scanning and interpreting the facial emotions and creating a playlist accordingly. The tedious task of manually Segregating or grouping songs into different lists is reduced by generating an appropriate playlist based on an individual's emotional features. The proposed system focuses on detecting human emotions for developing emotion-based music players. Linear classifier is used for face detection. A facial landmark map of a given face image is created based on the pixel's intensity values indexed of each point using regression trees trained with a gradient boosting algorithm.

A multiclass SVM Classifier is used to classify emotions Emotion are classified as Happy, Angry, Sad or Surprise. The limitations are that the proposed system is still not able to record all the emotions correctly due to the less availability of the images in the image dataset being used. Diverse emotions are not found. Handcrafted features often lack enough generalizability in the wild settings.

Result of Literature Review

Emotions Classification from Facial Expressions part of the report is one the most major and important parts of the project. Because of this, publications of research papers, whitepapers, and previous results are near to my study. According to some scientists for this department. Since researchers tried to build a new app with a much more efficient and also computationally effective app, they started by classifying emotions into four major emotions of Happiness, Sadness, Anger, Neutral, Fear, Surprise, and Disgust. While other emotions are the fusion of emotions like grief is of Sadness and Anger. Thus, this proposed project can utilize Convolutional Neural Network (CNN) for the identification of facial expressions, as CNN can give good accuracy and precision in a reasonable amount of time.

Hardware and Tools Used

a) Wireshark Designing Tool: Figma
b) State Transition Designing Tool: yEd Live
c) Breakdown's Designing Tool: yEd Live
d) Proposed Image Capturing Tool: OpenCV
e) Proposed Algorithmic Tool (for Emotion Detection): Convolutional Neural Network
f) Preferred Programming Language: Flutter (Mobile Application Layer) and Python (Server Layer).
g) Preferred Platform: Android and iOS mobile phones.

III. IMPLEMENTATION

The proposed algorithm revolves automated music recommendation system that plays song according to the mood or current emotion of the person. The person's photo is captured whenever the application gets open; hence current emotion is captured and detected. According to the information given by the image, the song is played related to the emotion.

The songs whichever present in the phone are already classified into 7 different classes such as happy, sad, anger, surprise, fear, disgust and neutral. The newly added songs are also classified dynamically to appropriate mood. It is composed of three modules: Facial expression recognition module, Song emotion recognition module and System integration module. Facial expression recognition and audio emotion recognition modules are two mutually exclusive modules. Hence, system integration module maps two modules to find the correct match of detected emotion. ***(See Figure:-1)***

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

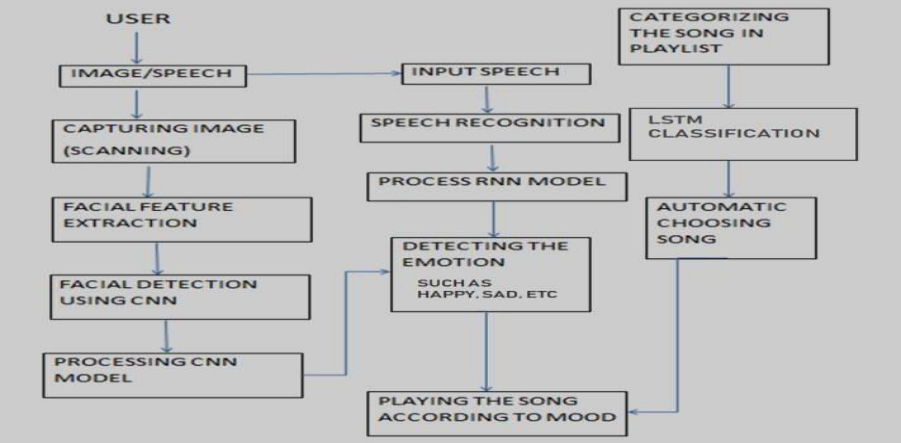

Figure 1. Block Diagram

A. Data Set

The Raw dataset is downloaded one by one from Google images for seven emotions. Extra dataset is taken from Kaggle datasets for facial expression detection.

B. Trained Dataset

Before processing the model, the training and testing phases is undergone. Trained dataset are those which are taught to the model or which learn

At the time of training, system takes dataset of faces (images) with their respective expression; eye should be in centre location mostly and learns a set of weights, which splits the facial expressions for classification. ***(See Figure:-2)***

Emotion Extraction Module -The image of the user is captured with the help of a camera/webcam. Once the picture captured, the frame of the captured image from webcam feed is converted to a grayscale image to improve the performance of the classifier, which is used to identify the face present in the picture. Once the conversion is complete, the image is sent to the classifier algorithm which, with the help of feature extraction techniques can extract the face from the frame of the web camera feed. From the extracted face, individual features are obtained and are sent to the trained network to detect the emotion expressed by the user. These images will be used to train the classifier so that when a completely new and unknown set of images is presented to the classifier, it is able to extract the position of facial landmarks from those images based on the knowledge that it had already acquired from the training set and return the coordinates of the new facial landmarks that it detected. The network is trained with the help of CK extensive data set. This is used to identify the emotion being voiced by the user.

Audio Extraction Module - After the emotion of the user is extracted the music/audio based on the emotion voiced by the user is displayed to the user, a list of songs based on the emotion is displayed, and the user can listen to any song he/she would like to. Based on the regularity that the user would listen to the songs are displayed in that order. This module is developed using web technologies like PHP, MySQL, HTML, CSS, JAVASCRIPT.

Emotion - Audio Integration Module - The emotions which are extracted for the songs are stored, and the songs based on the emotion are displayed on the web page built using PHP and MySQL. For example, if the emotion or the facial feature is categorized under happy, then songs from the happy database are displayed to the user.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

Sample Trained Dataset Is:-

Figure 2: Trained dataset

1.

For training, the sequence is:

1. Spatial normalization
2. Synthetic samples generation
3. Image cropping
4. Down-sampling
5. Intensity normalization.

Figure 3 : Test Dataset

C. Test Data

At the time of testing, classifier takes images of face with respective eye center locations, and it gives output as predicted expression by using the weights learned during training.

For recognizing an unknown image (testing), the sequence is:

1. Spatial normalization
2. Image cropping
3. Down-sampling
4. Intensity normalization

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

Convolution Neural Network

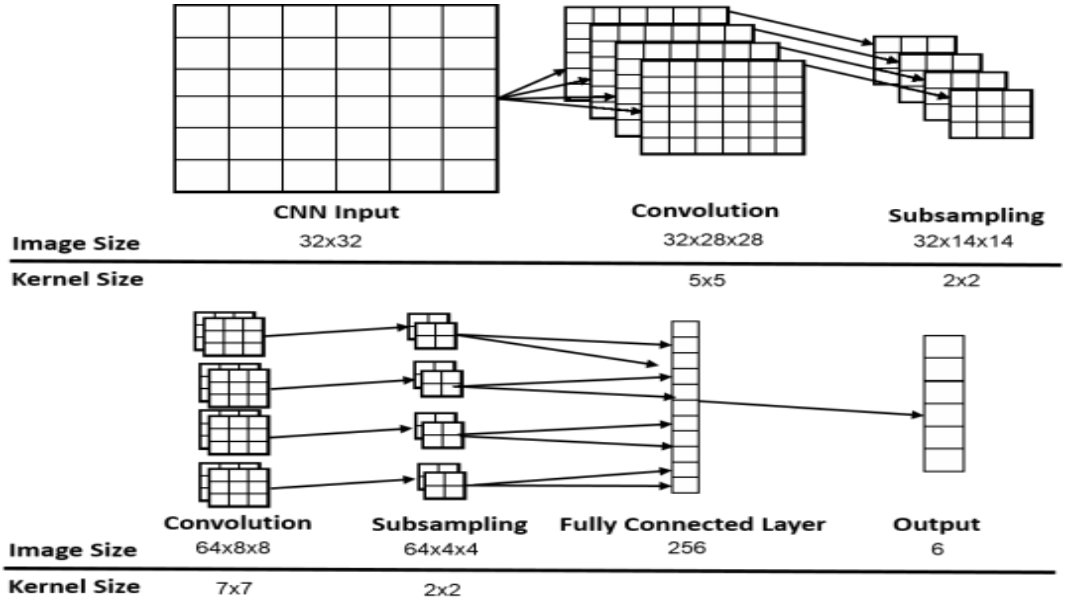

Figure 4: Architecture of proposed Convolution Neutral Network. It has five layers: first layer is convolution, second layer is sub-sampling, the third layer is convolution, fourth layer is sub-sampling, fifth layer is fully connected layer and final responsible for classifying facial image.

IV. FINAL RESULTANT MODEL

All the photos present in dataset are firstly converted to grayscale, for making preprocessing and detection more efficiently, faster and easier. Each input image is in form of pixels (e.g. 48x48). Now the pixel represented images are sent to the convolution layers (hidden layers). In Between each layer maximum pooling is done, the purpose of doing so, is to down-sample the input data or image, reducing the dimensions and allows assumption to be made about features contained in sub regions. This is done to avoid over-fitting. As well as it reduces computational cost by reducing number of parameters to learn. Example, if input image is of matrix 4x4 representation and let's say output we want is in 2x2, then pooling is performed in between all hidden layers. After that data is sent to dense layer, to prevent over-fitting. Dropout technique is used to reduce over-fitting in neural networks. The output layer conveys the detected class. Let's say if the detected expression is happy, then the next step is to select anyone training dataset for music model. Now, the dataset is trained according to the match for playing music. LSTM neural network is used for classifying the songs. One hot encoding is performed to represent categorical variables into binary vectors, so as to make the classification faster and better. Then the song is played according to the current mood of the person.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

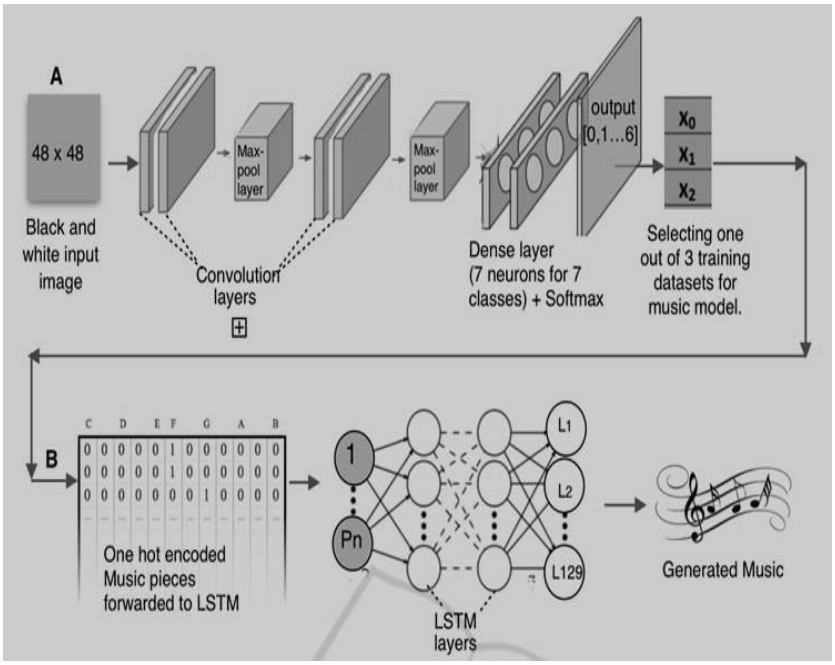

Figure 5 : Final Resultant Model

1. Result Analysis

Figure 6: Welcome Activity

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

Figure 7: Option Activity

Figure 8 : Playlist

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

Figure 9: Capturing Image And Emotions

Figure 10: Play Music According To Mood

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

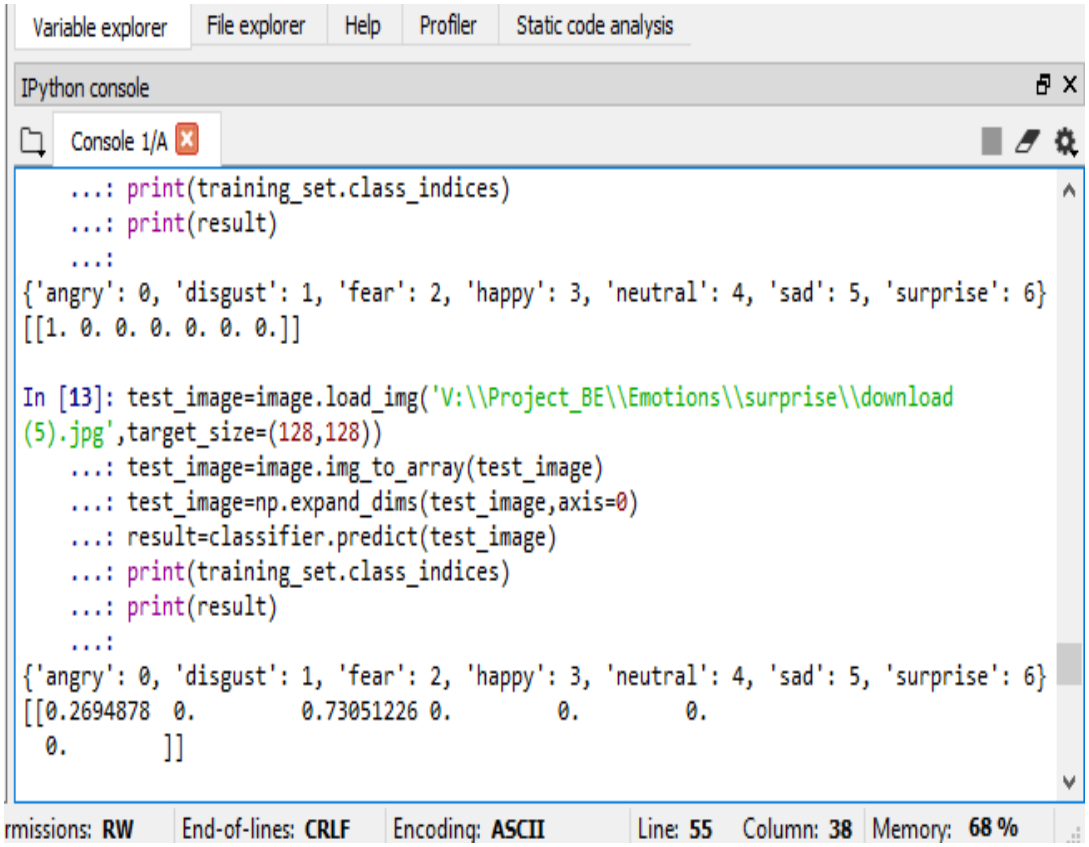

Figure 11: Backend (Emotion Detection)

V. CONCLUSIONS

The Emotion Based Music Player is used to automate and give a best music player experience for end user. Application solves all the basic needs of music listeners without troubling them as existing applications do. It uses technology to increase the interaction of the system with the user in numerous ways. It eases the work of user by capturing the image using phone's camera, detecting their emotion and suggesting a customized playlist with advanced features. The user's negative or bad thoughts are slowly converted to positive thoughts by changing the song from low tone to excited tone.

VI. FUTURE SCOPE

In future Music Player can be enhanced with Google play music, so songs which are not present in local storage can also be played and to access the whole application in speech based. The Emotion Based Music System will be of great advantage to users looking for music based on their mood and emotional behavior. It will help reduce the searching time for music thereby reducing unnecessary time and hence increasing the overall accuracy and efficiency of the system. The system will not only reduce physical stress but will also act as a boon for the music therapy systems and may also assist the music therapist to treat the patient. In future it can also be used to detect the sleepy mood of the driver, driving the car and many more uses. Also with its additional features mentioned above, it will be a complete system for music lovers and listeners.

REFERENCES

1. Kanade T, Cohn JF and Tian Y 2000 Comprehensive database for facial expression analysis InProceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition 46-53 IEEE.
2. Luoh L, Huang CC and Liu HY 2010 Image processing based emotion recognition In2010 International Conference on System Science and Engineering 491-494 IEEE.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 5, May 2022||

| DOI:10.15680/IJIRSET.2022.1105032 |

3. Vivek JD, Gokilavani A, Kavitha S, Lakshmanan S and Karthik S 2017 A novel emotion recognition based mind and soul-relaxing system In2017 International Conference on Innovations in Information, Embedded and Communication Systems 1-5 IEEE.
4. Jyoti Rani and Kanwal Garg 2014 Emotion Detection Using Facial Expressions A Review International Journal of Advanced Research in Computer Science and Software Engineering 4.
5. Joshi A and Kaur R 2013 A Study of speech emotion recognition methods Int. J. Comput. Sci. 75-82, 2004.
6. Shoaib M, Hussain I, Mirza HT and Tayyab M 2017 The role of information and innovative technology for rehabilitation of children with autism a systematic literature review In2017 17th International Conference on Computational Science and Its Applications 1-10 IEEE.
7. Dubey M and Singh L 2016Automatic emotion recognition using facial expression a review International Research Journal of Engineering and Technology (IRJET) 3 488-92.
8. Anwar S, Milanova M, Bigazzi A, Bocchi L and Guazzini A 2016 Real time intention recognition In IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society 1021-1024 IEEE.
9. Rázuri JG, Sundgren D, Rahmani R, Moran A, Bonet I and Larsson A 2015 Speech emotion recognition in emotional feedbackfor human-robot interaction International Journal of Advanced Research in Artificial Intelligence 4 20-7.
10. Dureha A 2014 An accurate algorithm for generating a music playlist based on facial expressions International Journal of Computer Applications 100 33-9.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com