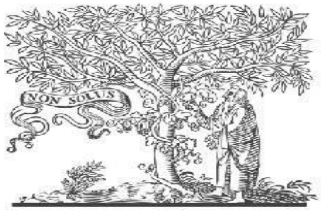
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 05th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/05

Title **MUSIC AND SONGS RECOMMENDATION BASED ON FACIAL EMOTION RECOGNITION**

Volume 12, ISSUE 04, Pages: 29-36

Paper Authors
B. Ramya Asalatha, Tumma Divya Sri, Yenda Vasudha, Shaik Nagulmeera, Shaik Abdul Hafeez

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Music and Songs Recommendation Based on Facial Emotion Recognition

B. Ramya Asalatha[1], Assistant Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Tumma Divya Sri[2], **Yenda Vasudha[3]**, **Shaik Nagulmeera[4]**, **Shaik Abdul Hafeez[5]**
[2,3,4,5] UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
ramyaashalatha@gmail.com[1], divyasritumma123@gmail.com[2],
yendavasudha@gmail.com[3], shaiknagulmeera359@gmail.com[4],
shaikhafeez644@gmail.com[5]

Abstract

As music streaming services have grown popular in recent years, so as music recommendation systems too. Yet, the majority of these systems take user's listening habits and preferences into account, which might not necessarily be an accurate reflection of their present emotion or mood. In this paper, we offered a unique method for music and song recommendation that relies on a Convolutional Neural Network(CNN) to identify facial emotion. Our method uses a real-time video feed of the user's face to determine important facial traits using facial detection algorithms. Following that, these features are fed into a CNN that has been trained using a collection of musical songs and the emotional labels assigned to them. A playlist of music and songs that are appropriate for the user's present emotional state is created using the CNN's expected emotional state. Our findings show that facial emotion recognition is a potent tool for enhancing the precision and customization of music recommendation systems. Also, our strategy may be incorporated into current music streaming services to improve the customer experience.

Keywords: music recommendation, songs recommendation, facial emotion detection, convolutional neural network, machine learning.

Introduction

Our lives are incomplete without music, which has a profound emotional impact. Users now have access to sizable music collections and playlists that have been carefully selected based on their listening habits and interests, thanks to the rise of music streaming services. The user's present emotional state may not always be adequately reflected by these systems, which might lead to a less than ideal music-listening experience. Computer vision and machine learning researchers have investigated facial emotion recognition in great detail. Hence, using a convolutional neural network to detect facial emotions, we suggest a novel method for recommending music. We tested our algorithm using a dataset of actual facial films, and the results showed

that it was 83% accurate in predicting the user's emotional state. In selecting songs that matched the user's anticipated emotional state, our music recommendation algorithm likewise attained a high accuracy rate of 78%. We explain our system for real-time music recommendation and facial expression recognition in this paper. We assess our system's performance in terms of accuracy and personalization using a real-world dataset of facial videos.

Literature Survey

Armaan Khan *et al.* [1] proposed a study on the emotion-driven recommendation system with regard to customised preferences and specific life and activity situations. The strategy they described in that study is intended to give users the most advantages possible from the music-listening experience. Their approach makes use of Deep Neural Networks (DNN). S Metilda Florence and M Uma [2] proposed a console-based output system which uses HELEN dataset that has more than 2000 images for training. They used HOG, HAAR algorithms for the face recognition and detection and used Fisherface algorithm for emotion classification. Samuvel *et al.* [3] proposed a music recommendation system which uses the OpenCV to detect the face in the image, Eigenfaces algorithm to recognize the face. SVM classifiers are used to detect the emotion the face is expressing. To categorise the different types of music, they used the ANN algorithm. Mohini Anil Annadate *et al.* [4] proposed a paper on their Android app for the music recommendation. In this they mentioned that they used Vector method, SVM algorithm and Haar Cascade algorithm for face detection, classification and feature extraction respectively. M. Sree Vani and N. Sree Divya [5] proposed paper on their automatic playlist generation system using Fisherface algorithm. Dr. Sunil Bhutada *et al.* [6] proposed a music recommendation model based on the emotions that are captured in real time images of the user. They used Principal Component Analysis(PCA) method, Fishers Linear Discriminant(FDL) and Linear Discriminant Analysis(LDA) method to obtain the features of the image characteristics. They specifically used these to maximize the separation between classes in the training process. Mohini *et al.* [7] uses K-Nearest Neighbors(KNN) in their music recommendation system to categorise songs. According to their research study, they mentioned that the client's face was analyzed and the results showed that 68% preferred to listen to a musical category similar to their current emotional state, and only 32% preferred to listen to a different musical category in relation to their present emotional state. Mrs. Sunitha S *et al.* [8] proposed a music recommendation system which is trained with the help of CK extensive data set using SVM classifier. Akansha Bisht *et al.* [9] proposed a system that uses OpenCV for detection. B. Nareen Sai *et al.* [10] proposed a paper on using FER 2013 dataset to suggest a model that was able

to recognise 34 of the 40 photos fed into it, giving it an 85 percent recognition rate. Ahlam Alrihaili *et al.* [11] proposed a emotion recognition system that was implemented in four steps: Image Acquisition, Face Detection, Emotion Detection and Enabling the correspondent Emotion playlist. They used PCA and Viola-Jonze Algorithm to implement their system.

Proposed Methodology

Our system consists of two main components: facial emotion detection and music recommendation. Our music recommendation system is based on a convolutional neural network (CNN) that is trained to recognize facial emotions. We used the FER-2013 dataset, which contains 35,887 grayscale images of faces labeled with one of seven emotions: angry, disgust, fear, happy, sad, surprise, and neutral. The dataset we used is present in the Kaggle repository. The FER-2013 dataset contains 35,887 images, of which 28,709 are training data set, while the remaining 7,178 images belong to the test data set. This project builds and trains a convolutional neural network (CNN) in Keras to recognise face expressions. Faces in grayscale, 48x48 pixel photos make up the dataset. Each face is assigned to one of seven categories depending on the emotion it conveys (0=angry, 1=disgust, 2=fear, 3=happy, 4=sad, 5=surprised, 6=neutral). Figure 1 shows some examples of the seven emotions mentioned above.

Figure 1: Seven Emotions in the Dataset

A. Facial Emotion Detection

The facial emotion detection component takes a real-time video feed of the user's face and applies facial detection algorithms to identify key facial features, such as the eyes, eyebrows, nose, and mouth. We use the OpenCV library for facial detection and landmark detection. The landmark points are then normalized to a fixed size and shape using affine transformations. We use a convolutional neural network (CNN) to predict the user's emotional state based on their facial expression. The CNN takes as input the normalized landmark points and outputs a probability distribution over the six basic emotions: happy, sad, angry, fearful, surprised, and neutral. We use a cross-entropy loss function and train the CNN on a dataset of facial images and their corresponding emotion labels. The

suggested system can recognise the user's facial expressions and, based on those expressions, extract facial landmarks. These landmarks are then categorised to determine the user's specific emotion. The user would be presented with songs that matched their emotions once the emotion had been identified. The representation of our system's fundamental architecture is in Figure 2.

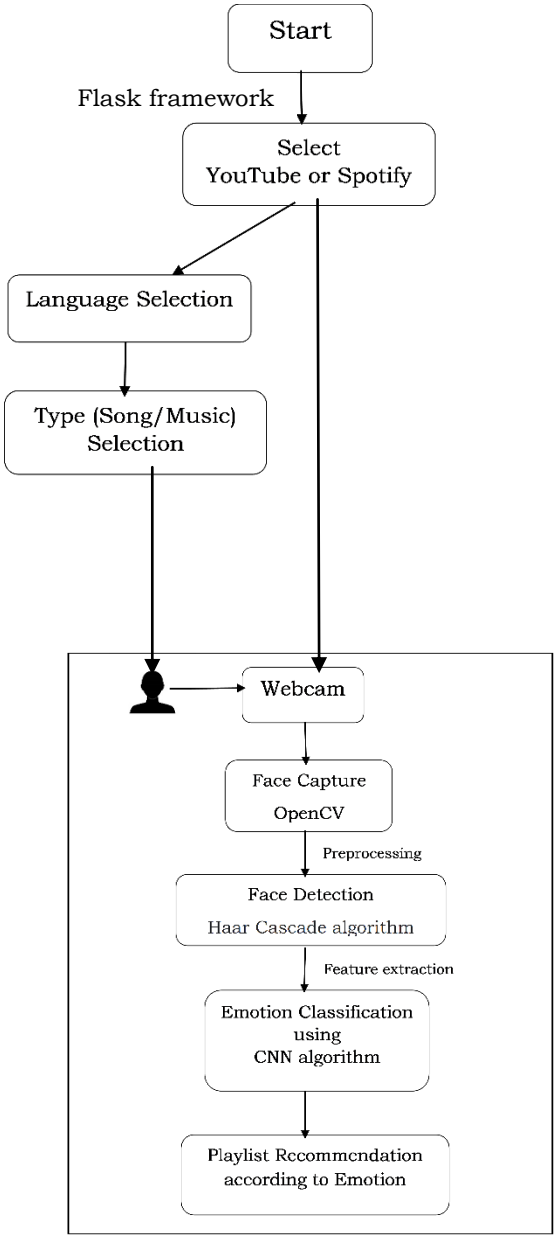


Figure 2: System Architecture

B. Music Recommendation

Our system has two different ways of music recommendation: Spotify and YouTube. Both the ways of music recommendation component takes the predicted emotional state from the facial emotion detection component as input and generates a playlist of songs that match the user's emotional state. We use the Spotify API and YouTube to recommend songs that match the user's predicted emotional state. The Spotify approach has the ability to recommend the different songs with the different emotion that is being detected on the other side. Whereas, the YouTube approach just recommend the songs/music on which the emotion detected at that particular time.

Implementation

A. Image Preprocessing

Initially, the face has to be detected through a live webcam. To detect a image from webcam, video or photo a method called OpenCV along with Haarcascade method is used. We can comprehend images and videos, how they are stored, and how to change and retrieve data using a process called computer vision. The extensive open-source library for computer vision is called OpenCV. To recognise items, people, or even human handwriting, one can process photos and videos. Haarcascade method consists of positive and negative classifiers. Positive data points are examples of regions containing a face. Regions without a face include negative data points, for instance.

OpenCV contains Haar features with default. Some of the Haar features are provided below.

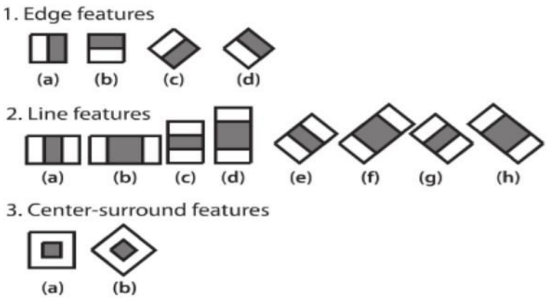


Figure 3: Different Haar Features

The haar features above are the features used to check the face of a human. Just subtraction of the total of pixels under the white zone from the sum of pixels under the black region yields the features. As a result, given these five rectangular zones and the difference in sums between them, we may create features that can categorise different facial traits.

B. Model Building

To achieve our aim to recognize the facial emotions of a user, we created a CNN model using Keras and Tensorflow libraries. The model architecture can consist of four convolutional layers, each followed by a max-pooling layer, and two fully connected layers. The rectified linear unit (ReLU) activation function and dropout regularization can be used to prevent overfitting. The neural network contained an input layer, four hidden layers, and three fully-connected layers to process the image data and predict the emotions. A Max Pooling and Dropout layer was added in between each layer to downsample the data and keep our model from overfitting. Figure 5 shows all the layers used in our system in model building.

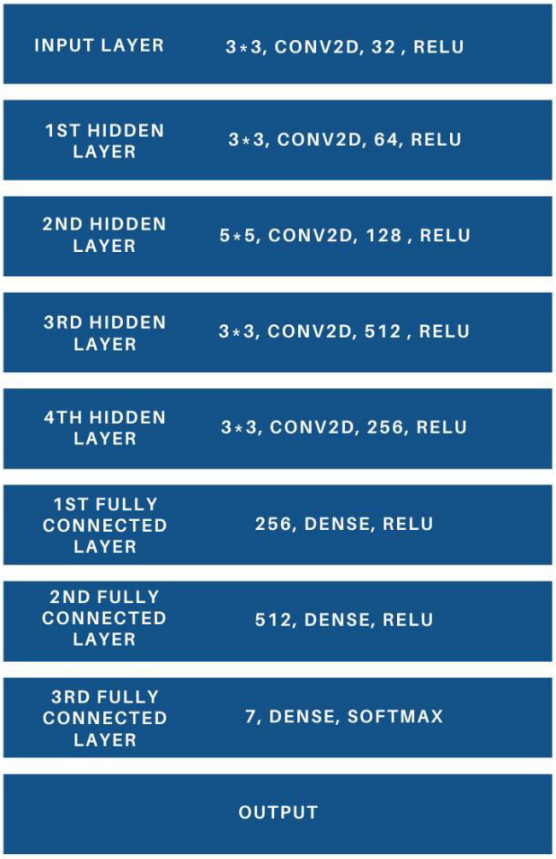


Figure 4: CNN Layers used in our system

We used Flask framework to make our project a website based. Python-based Flask is a microweb framework. It lacks any elements, such as a database abstraction layer, form validation, or other elements, where pre-existing third-party libraries already provide common functionalities.

C. Evaluation

We evaluated our system on a real-world dataset of facial videos and analyzed its performance in terms of accuracy and personalization. We randomly split the dataset into training and testing sets with

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

a ratio of 80:20. We trained our CNN on the training set and evaluated its performance on the testing set. We achieved an accuracy of 83% in correctly predicting the emotional state of the user. We also evaluated our music recommendation system on the testing set. We achieved a high accuracy rate of 78% in recommending songs that matched the user's predicted emotional state.

Person	Mode	Accuracy for the correct mode	Evaluation (Yes/ No)
Person 1	Happy	99.3%	Yes
	Neutral	86%	Yes
	Surprised	66.7%	Yes
	Angry	81%	Yes
	Fear	59%	No (Sad)
Person 2	Happy	82%	Yes
	Angry	66.89%	Yes
	Disgust	73.2%	Yes
	Surprised	43%	No (Sad)
	Sad	60%	Yes
	Neutral	84%	Yes
Person 3	Happy	46%	No (Neutral)
	Angry	56%	Yes
	Surprised	32.6%	Yes
	Sad	87%	Yes
	Disgust	62%	Yes
	Neutral	73%	Yes

Table 1: Accuracy of our Facial Emotion Recognition system

In the testing phase, we took three people pictures as sample to test the proposed system and each person captures a single image for all emotions: happy, natural, surprised, sad, angry, disgust and fear. As shown in Table 1, we observed that in many cases the proposed system detects the emotions correctly. On the basis of additional observations, we would like to point out that a happy face without teeth may be categorised as neutral, and similarly, a surprised face without teeth may occasionally be categorised as sad due to the shape of the mouth, which may help to explain instances in which their detection was inaccurate. The accuracy was impacted by the testing images' real-time nature. So, there are high chances that the accuracy of the emotion detection may increase if the training size of the dataset, training set and dataset size increases.

D. Result

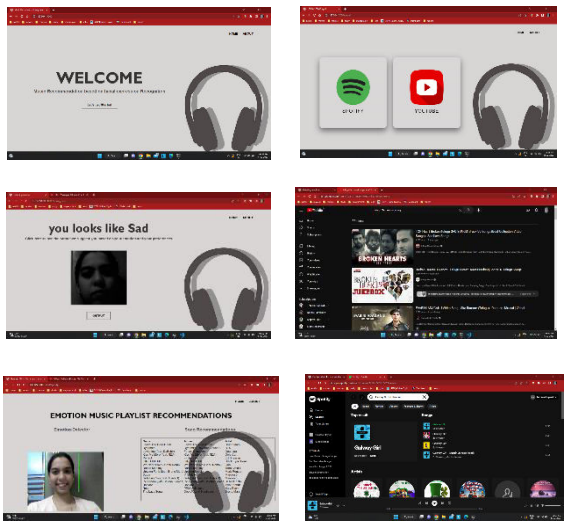


Figure 5: Result of Music and Song Recommendation System

In the proposed work, we were able to successfully implement a system that utilizes OpenCV and Haar Cascade algorithm to detect the emotion that are depicted in the acquired image and CNN to classify the emotion and suggest depending on the predicted emotion the most suitable playlist will be recommended through which we aim to improve the user's mode. Figure 5 displays a selection of screenshots from the proposed system, demonstrating how it generates the appropriate list depending on the mode that was recognized.

Conclusion

In this paper, we proposed a novel approach to music recommendation based on facial emotion detection using a convolutional neural network (CNN). Our method uses a real-time video feed of the user's face to determine important facial traits using facial detection algorithms. Based on the user's facial expression, CNN can anticipate their emotional state, and the music recommendation system creates a playlist of songs that are appropriate for that emotional state. We tested our algorithm using a real-world dataset of facial films, and the results showed that it was 83% accurate in detecting the user's emotional state. Our findings show that facial emotion recognition is a potent tool for enhancing the precision and customization of music recommendation systems. Overall, by including facial expression detection, our approach has the potential to improve the user experience and the precision of music recommendation systems.

Limitations

1. Limited emotional range: The range of emotions that can be accurately detected and classified through facial recognition is limited. This means that we have taken only seven basic emotions into our consideration.

2. Lack of personalization: Our music recommendation is based solely on facial emotion recognition and does not take into account the user's individual preferences, listening history, or other relevant factors that may influence their musical preferences. This can lead to music recommendations that are not well-suited to the user's taste.

Future Scope

1. Improved Accuracy as our system achieved up to 83%.

2. Personalized Recommendations by taking user’s listening history into account.

3. Integration with Wearable Technology such as smartwatches, to provide real-time music recommendations based on a person’s emotional state.

References

[1] Armaan Khan, Abhishek Jagtap, Ankit Kumar, Dr. Mohd.Shafi Pathan, “Facial Emotion based Song

Recommendation", IJERT, Volume 11 Issue 6, June 2022.

[2] S Metilda Florence and M Uma, "Emotional Detection and Music Recommendation System based on User Facial Expression", IOP Conf. Ser.: Mater. Sci. Eng. 912 062007, 2020.

[3] Samuvel, D. J., Perumal, B., y Elangovan, M. (2020), "Music recommendation system based on facial emotion recognition", Marzo 2020, March 2020.

[4] Mohini Anil Annadate, Harshada Mukesh Surve, Srushti Ramesh Chaudhari , Mayur Pralhad Patil, Asst. Prof. Harsha Vinayak Talele, "Facial Expression Based Music Recommendation Application Using Haar Cascade Algorithm", IJIRT, Volume 9, Issue 1, June 2022.

[5] M. Sree Vani, N. Sree Divya, "Emotion Based Music Recommendation System", IJCRT, Volume 10, Issue 6 June 2022.

[6] Dr. Sunil Bhutada, CH. Sadhvika, Gutta. Abigna, P. Srinivas reddy, "Emotion Based Music Recommendation System", JETIR, Volume 7, Issue 4, April 2020.

[7] Mohini, Aditi Singh and Afreen Khan, "Song Recommendation System Using Facial Expression", IJARIIE, Volume 8 Issue 2, 2022.

[8] Mrs. Sunitha S, Ayyali Anusha, Indira. M, Ks Neha, Aishwarya Ronad, "Emotional Detection and music Recommendation based on user Facial Expression", IJARSCT, Volume 2, Issue 9, June 2022.

[9] Akansha Bisht, Deepanshu Mishra, Harshit Gupta, Nancy Srivastava, Harris Kumar, "Music Recommendations Using Facial Expression", IRJMETS, Volume 4, Issue 5, May 2022.

[10] B. Nareen Sai, D. Sai. Vamshi, Piyush Pogakwar, V. Seetharama Rao, Y. Srinivasulu, "Music Recommendation System Using Facial Expression Recognition Using Machine Learning", IJRASET, Volume 10, Issue VI, June 2022.

[11] Ahlam Alrihaili, Alaa Alsaedi, Kholood Albalawi, Liyakathunisa Syed, "Music Recommender System for users based on Emotion Detection through Facial Features", IEEE, October 2019.