IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Emotion Based Music Recommendation

Vrajkumar Parekh

Department of computer engineering

G. H. Patel college of Engineering and Technology (CVM University)

Abstract: **The feelings contained in the human face have a great influence on decisions and arguments about various issues. Listening to music can reduce stress. Music plays a significant role based on mood. Automatic extraction of these emotions from images of human faces can help in human-computer interaction as well as many other applications. Machine learning algorithms, especially deep neural networks, can learn complex features and classify complex patterns. In this paper, a deep learning-based framework for human emotion detection is presented. The proposed framework uses deep convolutional neural networks and transfer learning methods. The experimental results show that the proposed features increase the speed and accuracy of training the neural network.**

***Keywords*—Facial emotion recognition, Convolution neural network, Transfer learning**

1. INTRODUCTION

Emotions have an important role in our everyday lives and directly affect decisions, reasoning, attention, prosperity, and the quality of life of human beings. Music, a tool for arousing emotions and feelings, is far more powerful than language. Music is something that taps deeply into our emotional core as human beings. Thus, listening to good music can help us elevate our mood from a negative sense to a positive one. For example, listening to upbeat songs when the person is feeling sad can help him come out of his sadness and start feeling better. Nowadays, with the influence of computers on human lives and the mechanization of the lives of individuals, the establishment of human-computer interaction (HCI) has played a crucial and very important role [1].

Most of the existing music recommendation systems use collaborative or content-based recommendation engines. However, the music choice of a user is not only dependent on their historical preferences or music

contents. But it is also dependent on the mood of the user. This paper proposes an emotion-based music recommendation framework that learns the emotion of a user from the signals obtained via images. In this project, we are using images with faces from the FER-2013 dataset for training, to detect the user's mood, and to recommend music.

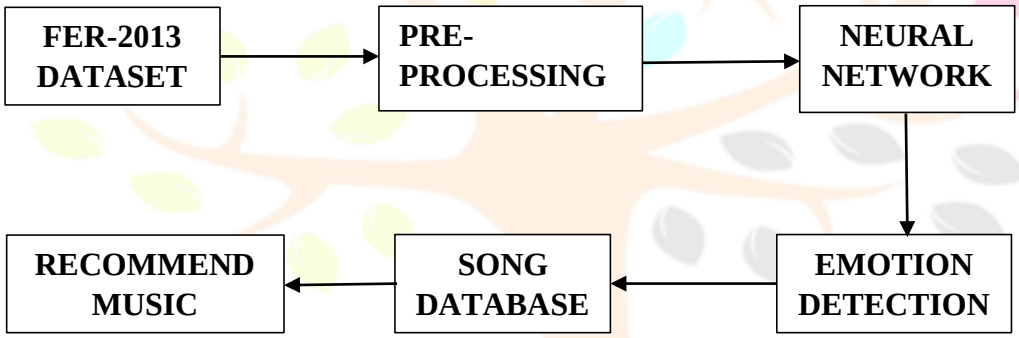
2. LITERATURE SURVEY

- A Genetic Algorithm (GA) for Music Data Recommendation System. This recommendation system designed by Hyun-Tae Kim, Eungyeong Kim, Jong-Hyun Lee and Chang WookAhn is a joint filtering (CF) and genetic algorithm hybrid solution [2].
- M. T. Quazi [3] applies the heart rate to classify human emotions. The research proposes 4 different types of human emotions; namely neutral, happy, sad and angry. The neutral emotion has the most stable heart rate (60 – 80 bpm), while the happy emotion has the highest variation rate (70 – 140 bpm) which depends on the type of happiness. The sad emotion has the second highest variation rate (80 – 100 bpm). The heart rate of angry emotion is in the same range as the happy emotion, but it will not lower than 100bpm [3].
- Dolly Reney and Dr.Neeta Tripaathi in their paper "An Efficient Method to Face and Emotion Detection" have detected face from the input image using Viola-Jones face detection algorithm and evaluated the face and emotion detection using KNN classifier [4].
- Santamaria-Granados et. al applies the deep learning approach using a deep convolutional neural network on a dataset of physiological signals (electrocardiogram and galvanic skin response), in this case, the AMIGOS dataset. The detection of emotions is done by correlating these physiological signals with the data of arousal and valence of this dataset, to classify the affective state of a person. In addition, an application for emotion recognition based on classic machine learning algorithms is proposed to extract the features of physiological signals in the domain of time, frequency, and non-linear. This application uses a convolutional neural network for the automatic feature extraction of the physiological signals, and through fully connected network layers, the emotion prediction is made. The experimental results on the AMIGOS dataset show that the method proposed in this paper achieves a better precision of the classification of the emotional states, in comparison with the originally obtained by the authors of this dataset [5].
- Li et al. presented a deep fusion convolutional neural network (DF-CNN) method using 2D+3D images. In this method, three-dimensional scan images of the faces are used. These images make up a total of 32 dimensions. This paper explains that prediction is done using two methods. 1- Classification

by using the SVM from the 32- dimensional features. 2- The normal prediction of the Softmax function using the six-state probability vector. Experimental results shown DF-CNN achieved good results [6].

3. SYSTEM ANALYSIS AND DESIGN

This work proposes an emotion-based music recommendation framework that learns the emotion of a user from the signals obtained via images. In this project, we are using images with faces to detect user moods and recommend music. Faces are the best option to predict the mood of the users. In this application, uploading an image and then using Python OPENCV and pre-processing the image to extract features is done. This feature is then applied to the Deep Learning Neural Network Training Model to predict the moods of the user, and based on the user's mood, all songs will be detected and shown in tabular form. The results of comprehensive experiments on real data confirm the accuracy of the proposed emotion classification system, which can be integrated into any recommendation engine.



[Flow diagram]

3.1 DATA DESCRIPTION

We built the Convolutional Neural Network model using the Kaggle dataset. The database is FER-2013, which is split into two parts: the training and testing datasets. 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 seven emotions: happy, sad, angry, surprise, fear, disgust, 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, fear, disgust, and surprised. 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 treat missing and outlier values, we have used a loss function named categorical cross-entropy. For each iteration, a selected loss function is employed to gauge the error value. So, to treat missing and outlier values, we have used a loss function named categorical cross-entropy.

[Types of emotions]

3.2 CNN ARCHITECTURE

The structure of the convolutional neural network is shown below. First, the image is received and after passing through different layers and the learning process returns a vector with seven modes as output. In fact, these seven modes are: Angry, Disgust, Fear, Happy, Neutral, Sad and Surprise.

Rectified Linear Unit (Relu) function: $f(x) = x^{+} = max(0, x)$

SoftMax function: $\sum_{i} = Xi = 1 \qquad Xi = [0, 1]$

→Convolution model

Layer		Feature maps	Size	Kernel size	Activation	Param
Input	Operation					
1	Convolution	64	(48, 48)	(3, 3)	Relu	640
2	Max pool	64	(24, 24)	(2, 2)	Relu	0
3	Convolution	128	(24, 24)	(5, 5)	Relu	204928
4	Max pool	128	(12, 12)	(2, 2)	Relu	0
5	Convolution	512	(12, 12)	(3, 3)	Relu	590336
6	Max pool	512	(6, 6)	(2, 2)	Relu	0
7	Convolution	512	(6, 6)	(3, 3)	Relu	2359808
8	Max pool	512	(3, 3)	(2, 2)	Relu	0

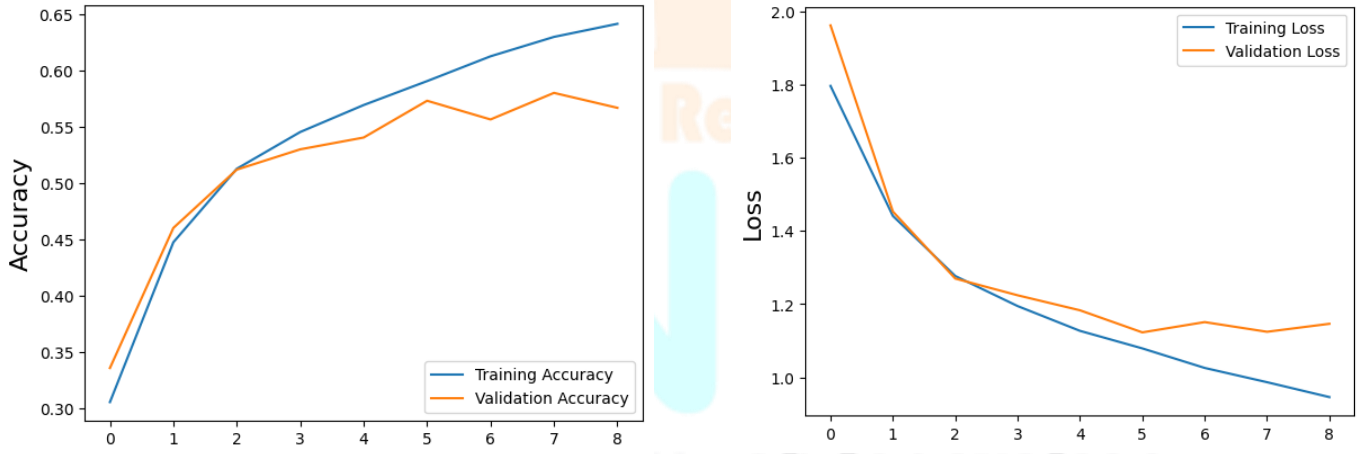
9	Flatten	-	4608	-	Relu	0
10	Dense	-	256	-	Relu	1179904
11	Dense	-	512	-	Relu	131584
Output	Dense	-	7	-	SoftMax	3591

Total Params: 4,478,727
Trainable params: 4,474,759
Non-trainable params: 3,968

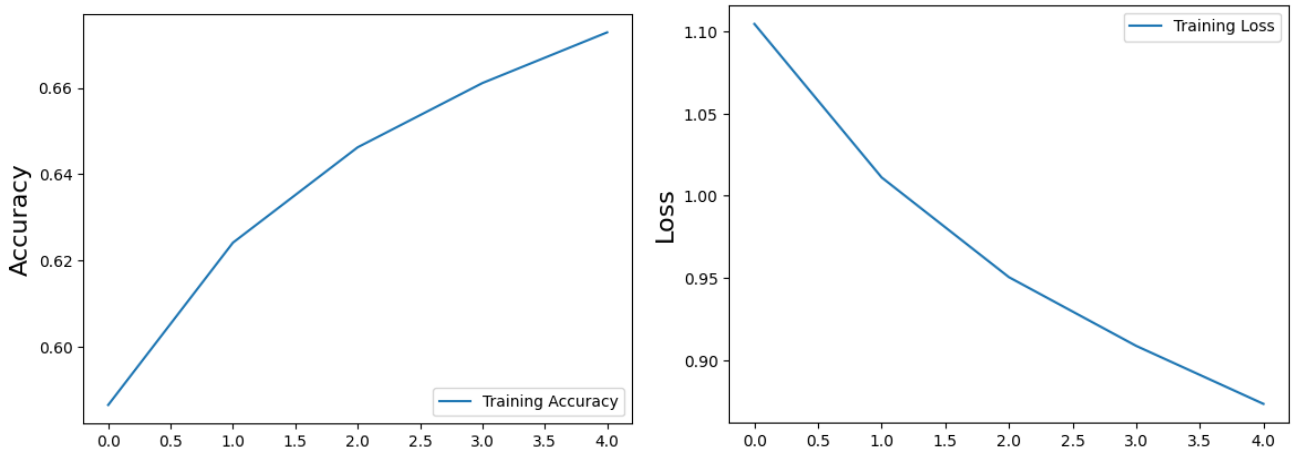
→Results

Model	**CNN**	**MobileNet**
Accuracy	64%	67%
Loss	94%	87%

- CNN



- MobileNet



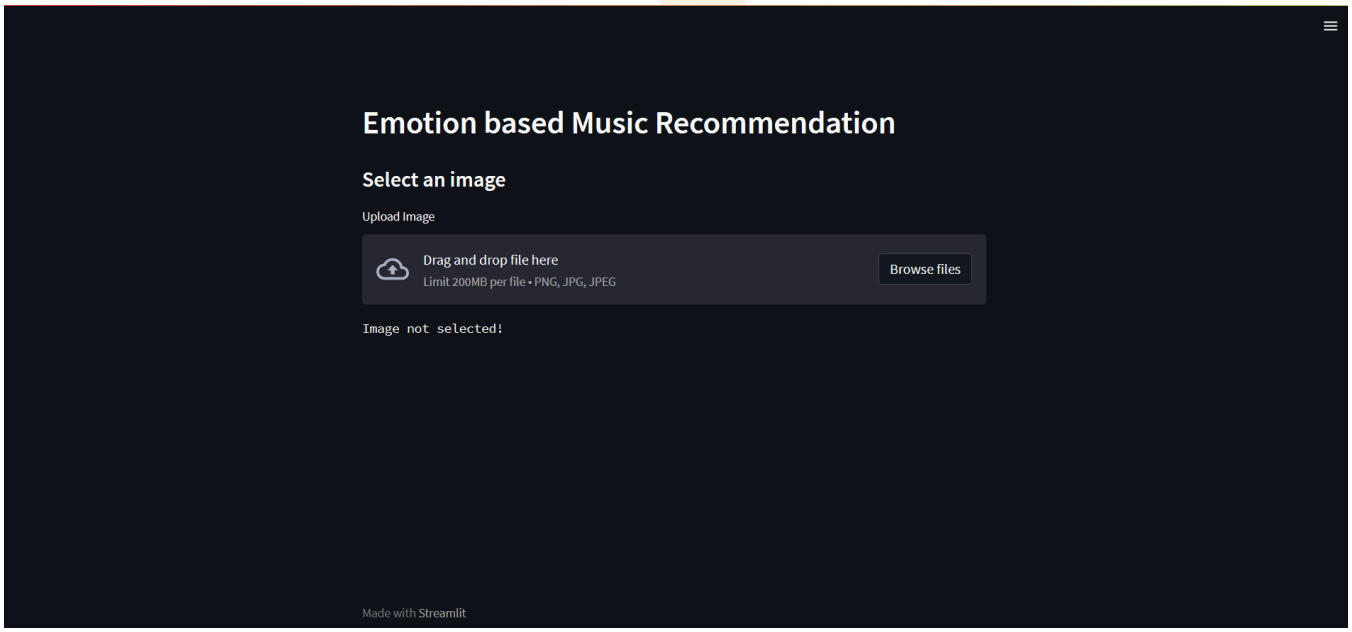
→Modules Description

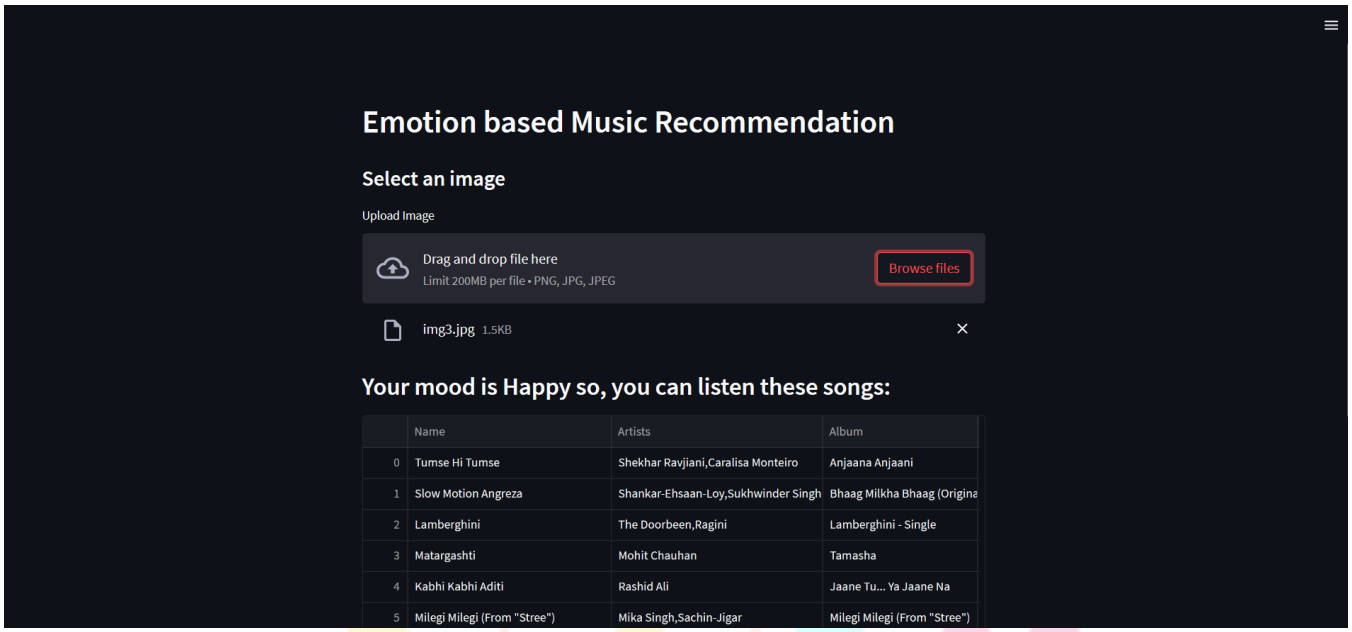
1. *Upload Image with Face*: Upload Image with Face module is used to upload image. And selecting the one image.

2. *Pre-process & Detect Face in Image*: Pre-process & Detect Face in Image module is used to perform pre-processing and to extract face from images. We can see in uploaded image one face is detected.

3. *Detect Emotion:* Detect Emotion Module to detect emotion and we can see emotion Angry is detected.

4. *Predicted Song*: Now you can see the list of Bollywood songs which you can hear which helps in healing of health.





4. CONCLUSION

This paper proposes an emotion-based music player using a convolutional neural network. The application aims to suggest songs based on the users' emotions. To classify the emotion, the user's facial image is analyzed. If the user has positive mode, the application will recommend positive songs. In contrast, it will recommend songs with a negative mood because the users want to unleash their anger, sadness, or stress. After transfer learning was applied, the system learned faster, and the accuracy improved. The learning speed of the convolutional neural network has increased profoundly. The convolutional neural network receives a number of sub-features and takes one step further in extracting the emotions from the faces. To improve the performance of the application, we plan to explore more techniques to eliminate the uncontrolled environment. Moreover, we plan to expand the song database in order to support more users' interests.

REFERENCES

[1] S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," IEEE Transactions on Affective Computing, vol. 6, no. 1, pp. 1-12, 2015.

[2] Sturm Bl. Classification Accuracy Is Not Enough: On the Evaluation of Music Genre Recognition Systems. J Intell Inform Syst. (2013)

[3] M. T. Quazi, “Human Emotion Recognition Using,” MASSEY UNIVERSITY, PALMERSTON NORTH, New Zealand, 2012.

[4] Dolly Reney and Dr.Neeta Tripaathi, “An Efficient Method to Face and Emotion Detection”, Fifth International Conference on Communication Systems and Network Technologies, 2015.

[5] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," in IEEE Access, vol. 7, pp. 57-67, 2019, doi:10.1109/ACCESS.2018.2883213.

[6] H. Li, J. Sun, Z. Xu, and L. Chen, "Multimodal 2D+3D Facial Expression Recognition With Deep Fusion Convolutional Neural Network," IEEE Transactions on Multimedia, vol. 19, no. 12, pp. 2816-2831, 2017.