

# Music Recommendation based on User Mood

Vicky Kumar

Department of ECE  
GB Pant DSEU Okhla-I Campus  
ask4vicky@gmail.com

Sumit Kumar

Department of ECE  
GB Pant DSEU Okhla-I Campus  
sumitsingh1698@gmail.com

Venkateshwari P

Department of ECE  
GB Pant DSEU Okhla-I Campus  
venkateshwarigbpec@gmail.com

**Abstract**—This work detects the mood of the user and it plays the songs accordingly. The proposed work is implemented as a system that can run on an android and its main goal is to determine user's mood accurately. Facial images are used to recognise the emotion of the user. In this process, the front view of the facial images is taken to detect the mood from the images. Haar Cascade Algorithm is used for detecting the face and the CNN Algorithm is used for detecting the emotion of the user from the facial expression. The feedback system is also used to get much accuracy in Emotion Detection.

**Keywords:** Facial expression, CNN algorithm, mood detection, music classification, HCI.

## I. INTRODUCTION

In the past many times, it was proven that the Face expression is a better way in knowing anyone's Emotions. Emotion carries the thought or feeling which a person wants to listen to. So, Emotions play a vital role in listening to music. "If God wanted woods to be quiet, He would not be given birds songs to sing." If everyone started the day singing, just think how happy they'd be. "song", mode of expression, forever it is the best choice to know and understand human emotions.

Current research in the music field has demonstrated that music increases a clear emotional response in its listeners. The Musical choices appeared to be highly connected with the personality of a person and their moods. Characteristics of music are understood and managed by the motor regions brain that deal with mood [1]. The user response to the music fragments is based on various pivotal factors like age, gender, emotions, tradition, personal taste, etc [2],[3],[4] (e.g. Situation at that time). Apart from these external facts, people used to categorize the music based on their mood like joyful, sorrowful, neutral, shocked, etc. Recent researches on feelings-based music recommendation systems pivot on two main things, music audio and lyrical features [5][6][7][8][9]. This system is taking care of language-related issues. Facial expression is the immemorial way of naturally expressing the emotions and feelings of a person. In this project, the facial expressions are categorized into 7 basic types like happy, sad, angry, neutral, surprise, disgust, and fear.

The main goal of this work is to cast a least cost and accurate song recommender system that creates and opens an emotion-aware playlist for the user. The MRS system is designed to use fewer amounts of system resources. The song recommender system merges the outcomes of the emotions, state of mind and songs list and provides the mood-based song list. This system provides better precision, efficiency, and its performance finer than the existing applications.

## II. LITERATURE SURVEY

In [10], Random forest method is used for accurate classification of the emotions and music recommendation based on that detected emotions. LSTM method is used for labelling based training data sets on emotions. The reinforcement learning is further implanted to perform continuous training on the given dataset. The author analysed various factors such as feeling parameter

In [11], author uses a two step method, one for facial expression and other is emotion detection. Facial position and location of eyes and mouth is analysed extract the features from the facial image. Hausdorff distance and Bezier curve is used for training and taking measurement of the facial image and image available in datasets.

In [12], Mood play is an interactive platform discovers and recommends music based on the real time mood of the persons. It focuses on optimization and accurate prediction of music and ranking those song list in order.

In [13], the author analysed a personality of a person and emotional states of a person. The author analysed the various psychological factors of the user and enhanced the accuracy of the system.

## III. METHODOLOGY

In this work, the facial expression datasets are used which are taken out from the Kaggle source, it contains about 40,000 grayscale facial pictures, and all of them whereof size 48x48 pixels. The pictures are virtually centered and pre-processed in a way so that the magnitude of the images will be maintained correctly [14][15]. Images are then categorized into seven classes of facial emotions. The types of facial emotions are as 0= Repulsiveness, 1= Exasperated, 2= Blissful, 3= Fear, 4= Surprise, 5= Woebegone, and 6=Neutral. Fig 1 shows one example from some face expression categories.



Fig. 1. Facial Expressions

The images are divided into three sets for training, validating, and testing purpose. There are about five thousand validating and testing images each and 35,000 training images are available for serving the purpose. The image raw pixel data is taken and subtracted from the mean of all datasets. For the information augmentation purpose, the mirrored images are engendered by flipping images within the training 6 set horizontally. To relegate the expressions, convolution layers are utilized for engendering the features from the raw pixel data

#### A. Emotion Detection

- 1) CNN is used to train the face database.
- 2) Seven facial probabilities are extracted from each of the facial frame.
- 3) The single facial frame probability is fine tuned to a lengthy image descriptor. This process continuous for all the images in database.
- 4) Then the images are classified by utilizing a SVM machine [16-17].

The frontal view of the facial images is given as input then facial features are extracted from the input image. The expressions are analyzed and emotions are detected for the given data from the extracted facial features.

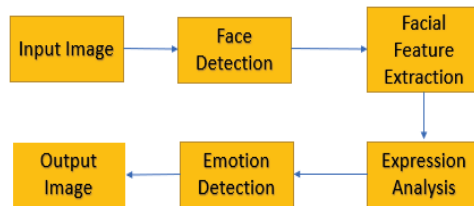


Fig. 2. Block diagram shows the emotion detection part.

#### B. Emotion Database

This is utilized in both realistic and virtual world to acquire as much data as possible at the time of data collection. Different types of emotional photographs of friends and family members, relatives, and some recognized unknown people's facial expressions can be found in the real world. Those data are filtered and saved for later analysis. The data set comes from kaggle.com and is gathered from internet media. The data is transferred into grayscale images at a resolution of 48 x 48 pixels. The datasets are divided into two parts: pixels and feelings. The sensation section contains a numeric code with values ranging from 0 to 6. Furthermore, each picture's pixel part comprises a string that is included in statements. Furthermore, the image should only be a portrait of a person's face. As a result, the gathered images are shrunk and cropped to create a portrait of a face. In addition, there is a distinct picture.

#### C. Training phase using deep learning method

CNN is a sublime technique to utilize deep learning method for identifying the images (CNN). It is developed by using a keras package in python. Image edges are nothing but a group of pixels to form a pattern. That pattern is

utilized by CNN to identify the images. The network used here multiplies a filter with a pixel matrix and then it integrates the results. This process continuous for the next pixel and it runs throughout all the pixels in the picture. Sequential method is a simplest approach in keras for building any model. This method is basically used for constructing the layer-by-layer model. Conv2D layers is used, it first constructs the initial two layers. Then these layers are used to handle input images with this 2D matrices. Conv2D, first layer contains 64 nodes and second layer contains 32 nodes. Based on the dataset size the value may be changed to high or low. In order to perform convolution, the filter matrix size and kernel size should be matched. A 3x3 filter matrix is taken and a kernel is of size 3. The ReLU is a type of Linear Activation function which is applied to the input image. This activation function acts as a stimulus to the neural network used. The input images are in this form 28,28,1, where 1 is an image which are greyscale. A 'Flatten' layer is a layer which is formed in between the Conv2D layers and the dense layer. This layer is utilized for connecting the convolutional and dense layers. This model provides the future values by predicting and cutting down the unnecessary data from the image. The model is reassembled to utilize three important parameters like optimizer, loss, and metrics. Optimizer is used to regulate the cognition rate. It makes utilization of the optimizer 'adam'. The Adam optimizer transmutes the cognition rate throughout the training. If a learning rate is increased gradually it results in using precise weight points, but this in turn increases the computation time. Categorical cross-entropy is employed for finding the loss function parameter and this is the widely used relegation method. The model is performing better for the lower score. To make things even more facile to understand, when training the model, the 'precision' measure will be habituated to precise the score of the validation dataset. The network is trained by using fit() function to validate the following terms like train data (X) and the targeted data (y), validating data, and epochs. Validation data uses the test dataset and it divides that into x and Y test set. The more no. of epochs will improvise the model. Once max limit is achieved, the model will be get saturated and never changes after that. The number of epochs in our model will be set to three. On that validation set, it has achieved 93 percent precision after three epochs.

#### D. Detection

In K- betokens clustering method two sets of clusters are made. Maximum and minimum values are found in each and every row and averages are taken out from all the calculated values. Utilizing these average values as a linear line, pixels which are more proximate to the max average values are grouped together as one cluster, and pixels which are more proximate to the min average values are grouped as another cluster. The clustering used for predicting the number of components used in image computation. The bounding box function is utilized to segment the person's ocular perceivers first. The ocular perceivers are segmented first because it is the first component encountered in the column-sagacious. Then the remaining face feature elements are segmented by utilizing a distance-predicated technique employed in the ocular perceiver matrix. The image obtained after doing k-denotes clustering for sundry expressions is exhibited. The important fact about this algorithm is that it takes a long time to train yet only takes a short time to detect. The Haar basis

feature is used in this technique. There are many different sorts of characteristics, including:

- 1) Edge characteristics
- 2) Line Characteristics
- 3) Four Rectangle Characteristics

Edge features are detected in the face like eyebrows, nose and teeth. The image moves on finding the next features using Haar method. Emotion recognition is based on the ratio of these identified features.

Using the Fourier equation, values are calculated.

$$\Delta = \text{dark} - \text{white} = \frac{1}{n} \sum_{\text{dark}}^n I(x) - \frac{1}{n} \sum_{\text{white}}^n I(x) \quad (1)$$

#### E. APP Development.

API is designed for creating a mobile application which connects the emotion detection with the mobile application. API is created using Django. The images are sent as a request to the app. The app responds by asking some basic questions to the user. Then based on the response received from the user, app suggests the music according to the mood of the user. The Fig 3 shows the final implementation of the proposed system.

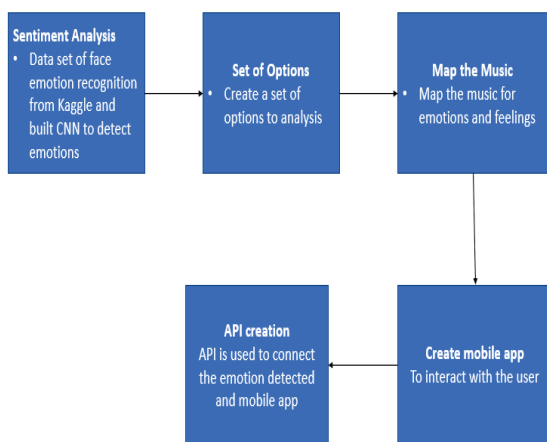


Fig. 3. Proposed system diagram

#### IV. RESULTS AND DISCUSSION

The black region value is 1 and the white area value is 0 for optimal Haar features. For optimal Haar characteristics, the difference between dark and white is 1-0=1. In an actual image, black region value is 0.74 and white region value is 0.18. The difference between these two values is 0.56.

Layers are the processing stages used to form neural networks. These layers are composed of a group of interconnected nodes. The input layer forms a pattern and provides to the other layers of a network, then in turn it communicates with one or more obnubilated layers, which perform the further processing via a weighted connections system. Face and facial components detection utilizing the viola-Jones algorithm, facial Feature extraction, and feature relegation utilizing CNN. These are the three stages of the facial emotion apperception system.

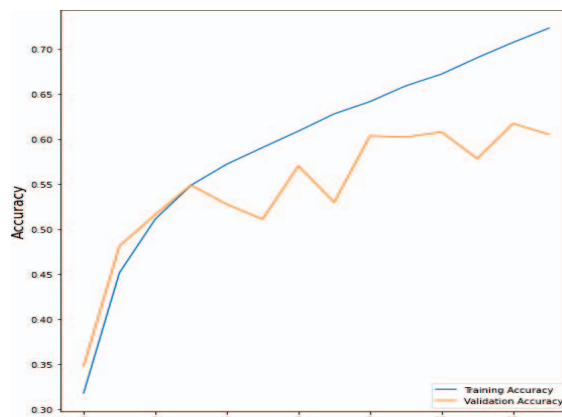


Fig. 4. Comparison b/w Training accuracy and Validation accuracy.

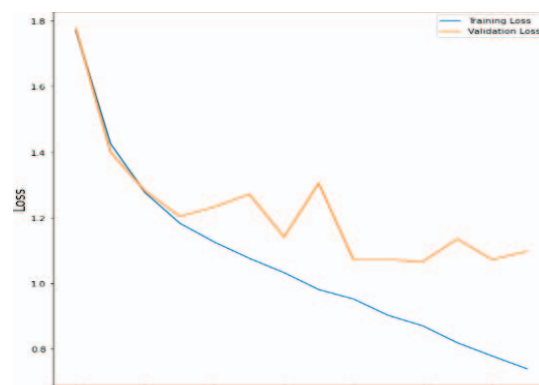






Fig. 5. Comparison b/w Training loss and Validation loss function.

TABLE I. IMAGES CLASSIFICATION

Images	Accuracy	Songs suggested
	Fear=0.93	Mantras enchancing
	Neutral=0.97	Melodies
	Angry=0.75	Melodies
	Happy=1.0	Bang

## V. CONCLUSION

The proposed system is implemented as an application that can run on android device and it determines the mood of the user. Facial images are utilized to detect the mood of the user. Haar Cascade Algorithm is used for detecting image features and the CNN Algorithm is applied to detect the emotion being expressed by the utilizer from the facial features. The feedback system is used to get more accuracy in Emotion Detection. In future this model will be extended to detect the emotions of the user by taking the side view of the image and also to serve the interest of the user in various fields like movie, education site, shopping apps etc based on the emotion detection and it will be applied on the various views of the image.

## VI. FUTURE SCOPE

The proposed system is implemented as an application that can run on android device and determines the mood of the user. The facial images are utilized to detect the mood of the user. Haar Cascade Algorithm is used for detecting image features and the CNN Algorithm is applied to detect the emotion being expressed by the utilizer from the facial features. The feedback system is used to get more accuracy in Emotion Detection. In future this model will be extended to serve the interest of the user in various fields like movie, education site, shopping apps etc based on the emotion detection.

## REFERENCES

- [1] Gilda, S., Zafar, H., Soni, C., & Waghurdekar, K. (2017, March). Smart music player integrating facial emotion recognition and music mood recommendation. In 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET) (pp. 154-158). IEEE
- [2] Divya, M. N. (2021). Smart Teaching Using Human Facial Emotion Recognition (Fer) Model. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(11), 6925-6932.
- [3] Swaminathan, S., & Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emotion review*, 7(2), 189-197.
- [4] Supriya, L. P., & Khilar, R. (2021, November). Affective Music Player for Multiple Emotion Recognition Using Facial Expressions with SVM. In 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 622-626). IEEE.
- [5] Laurier, C., Herrera, P., Mandel, M., & Ellis, D. (2007). Audio music mood classification using support vector machine. *MIREX task on Audio Mood Classification*, 2-4.
- [6] Shao, B., Wang, D., Li, T., & Ogihara, M. (2009). Music recommendation based on acoustic features and user access patterns. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(8), 1602-1611.
- [7] Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., & Liwicki, M. (2015). Dexpression: Deep convolutional neural network for expression recognition. *arXiv preprint arXiv:1509.05371*.
- [8] Anand, R., Singh, B., & Sindhwani, N. (2009). Speech Perception & Analysis of Fluent Digits' Strings using Level-By-Level Time Alignment. *International Journal of Information Technology and Knowledge Management*, 2(1), 65-68.
- [9] Bhat, A. S., Amith, V. S., Prasad, N. S., & Mohan, D. M. (2014, January). An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. In 2014 fifth international conference on signal and image processing (pp. 359-364). IEEE.
- [10] Rumiantcev, M., & Khriyenko, O. (2020). Emotion Based Music Recommendation System. In *Proceedings of Conference of Open Innovations Association FRUCT*. Fruct Oy.
- [11] Lee, Y. H., Han, W., & Kim, Y. (2013, September). Emotional recognition from facial expression analysis using bezier curve fitting. In 2013 16th International Conference on Network-Based Information Systems (pp. 250-254). IEEE.
- [12] Andjelkovic, I., Parra, D., & O'Donovan, J. (2016, July). Moodplay: Interactive mood-based music discovery and recommendation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (pp. 275-279).
- [13] Ferwerda, B., & Schedl, M. (2014, July). Enhancing Music Recommender Systems with Personality Information and Emotional States: A Proposal. In *Umap workshops*.
- [14] Kumar, R., Anand, R., & Kaushik, G. (2011). Image Compression Using Wavelet Method & SPIHT Algorithm. *Digital Image Processing*, 3(2), 75-79.
- [15] Vyas, G., Anand, R., & Holé, K. E. Implementation of Advanced Image Compression using Wavelet Transform and SPHT Algorithm. *International Journal of Electronic and Electrical Engineering*. ISSN, 0974-2174.
- [16] Jadhav, S. B., Udupi, V. R., & Patil, S. B. (2021). Identification of plant diseases using convolutional neural networks. *International Journal of Information Technology*, 13(6), 2461-2470.
- [17] Vijayalakshmi, M., and V. Joseph Peter. "CNN based approach for identifying banana species from fruits." *International Journal of Information Technology* 13, no. 1 (2021): 27-32.