IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

Emotion Detection with CNN Model and Song Recommendations using Machine Learning Techniques

Giridhar Sunil[1] and Abraham Kuriakose[2]
Student, Department of Computer Science[1,2]
Vellore Institute of Technology, Chennai, Tamil Nadu, India
giridhar.s2020@vitstudent.ac.in and abraham.kuriakose2020@vitstudent.ac.in

Abstract: *Music is an exemplary tool to judge a person's emotional state. It is the language of the soul. What cannot be articulated through words are easily conveyed through a melody. Music not only speaks to a person's emotional and mental state, but it is also known to have a therapeutic effect on the listener. The traditional method of music recommendation uses collaborative or content-based filtering to recommend songs but a person's song choices does not depend only on the song they usually listen to but depends mostly on their emotional state. With the fast-paced innovations pertaining to the music application industry, there is still scope of further improvement in the user experience and creating an encompassing application that not only allows the app users to enjoy listening to their favorite songs but also caters to their recommendation based on their emotional state. Thus, an emotion detection system using Convolution Neural Networks has been proposed. The user feeds in a custom playlist containing a mixture of musical genres that are classified into different emotions using K-Means Clustering. The CNN model detects the emotional state of the user and recommends a series of songs from the classified playlist. This interactive interface is a revolutionary innovation for users who need song recommendations that suit their current mind-state.*

Keywords: Emotion, Recommendation, Convolution Neural Network, K-Means Clustering, Content Based Filtering, Detection

I. INTRODUCTION

The software industry faces tremendous pressure to innovate and develop new applications to keep users entranced and ensure their loyalty. With the rapid increase in variety of apps created each day, users are no longer bound to limited choices and can now traverse through apps each day to find one that fits their liking the best. This creates great competition for app developers to satisfy the demands of their potential customers by displaying what the customers would like to see [9]. This has led to the extensive use of recommender systems in apps. Researchers have developed methods to automatically analyze and understand the diverse variety of music content available. Using computer science, signal processing, mathematics and statistics information such as genre, loudness, valence, etc. have been extracted and making music recommendation easier [2]. The traditional recommender system uses content based filtering, or collaborative filtering. These recommender systems use the past search data for recommendation. A person's song choices not only depend on the songs they have listened to earlier but also depend on their emotions. Emotions can be classified using the valence and arousal [1]. The songs have been classified based on these 2 features. Facial expression plays an important role in detection of emotions. A person's emotion can be analyzed from the facial features. These facial features can be used to detect emotions. Face recognition is a commonly used security method. Using the facial features, we can both detect the person who is using the application as well as the emotion the person is experiencing.

DOI: 10.48175/IJARSCT-11409

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

1.1 Objective

The objective of this project is to use Machine Learning and Deep Learning Algorithms to build a Recommender System for songs. The main reason to build this project is so that we can give users recommendation according to their needs. The recommendation based on previous songs gives users suggestions of songs they may like and the emotion-based music recommendation makes a custom playlist.

II. LITERATURE REVIEW

Table 1: Describes in detail the various literature survey papers.

Sn. No.	Journal Name	Author	Summary
1.	Induced Emotion-Based Music Recommendation through Reinforcement Learning	Roberto De Prisco, Alfonso Guarino, Delfina Malandrino and Rocco Zaccagnino	The paper "Induced Emotion-Based Music Recommendation through Reinforcement Learning" presents a method for recommending music to users based on their emotional state. The method uses reinforcement learning to model the relationship between users' emotional states and the songs they listen to. The proposed method outperforms existing music recommendation algorithms in terms of recommendation accuracy and user satisfaction. The emotion detection model achieved an accuracy of 94.6% in the training data and 67.4% in the testing data.
2.	Emotion Based Music Recommendation System Using LSTM - CNN Architecture	Saurav Joshi; Tanuj Jain; Nidhi Nair	The paper "Emotion based music recommendation system using LSTM-CNN architecture" presents a music recommendation system based on emotions. The system uses Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models to analyze audio features and lyrics of songs. The models are used to predict the emotions elicited by the songs. The system recommends songs to users based on their current emotional state, as determined by physiological signals such as heart rate and skin conductance. The authors evaluate the performance of the system using a dataset of songs and physiological signals.
3.	Emotion based music recommendation system	James, H. I., Arnold, J. J. A., Ruban, J. M. M., Tamilarasan, M., & Saranya, R	The use of an emotion recognition system to detect the emotions elicited by music. The use of a recommendation algorithm that takes into account the emotions elicited by music and the user's current emotional state. The use of physiological signals, such as heart rate and skin conductance, to determine the user's current emotional state. An evaluation of the performance of the recommendation system using a dataset of songs and physiological signals.
4.	Facial Emotion Based Music Recommendation System using computer vision and	Shalini, S. K., Jaichandran, R., Leelavathy, S., Raviraghul, R., Ranjitha, J., &	The use of a facial expression recognition system to detect the emotions of the user. The use of machine learning techniques, such as decision trees, to make music recommendations based on the user's emotional state.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

	machine learning techniques	Saravanakumar, N	The use of a dataset of songs and facial expressions to train and evaluate the performance of the recommendation system. An evaluation of the performance of the recommendation system using accuracy measures and a comparison with other existing methods.
5.	A survey of music recommendation systems and future perspectives	Song, Y., Dixon, S., & Pearce, M.	The paper surveys various music recommendation systems and categorizes them into content-based, collaborative filtering, and hybrid approaches. The authors discuss the challenges faced by these systems, such as scalability, sparsity, data quality, and user diversity. The paper highlights the importance of incorporating context information and social network data into recommendation systems. The authors suggest that the future of music recommendation systems lies in personalized and interactive systems that incorporate multiple sources of data and multiple objectives.

III. METHODOLOGY

The following section outlines the flow diagram of the implemented project, the datasets that have been used and the overall implementation of the project.

3.1 Process Flow

The flow diagram of the project has 3 parts: -

Creating the custom playlist and classifying songs based on the emotions. The songs are classified into 4 emotions "happy", "sad", "angry" and "surprised".

Emotion based song recommendation system that takes the webcam feed as input and uses the classified playlist made in the first part to make custom playlist.

Song based recommendation takes the Spotify link of the song you want recommendation for and the number of songs you want recommended, it searches for similar songs and accordingly gives recommendations.

The Process Flow Diagram explains the way in which we have made our model interactiveusing python.

We first make the custom song dataset after which the user is asked to select between emotion based and song-based recommendation and accordingly a playlist of songs is recommended.

3.2 Dataset Description

In this project, we used two datasets. One is a publicly available dataset and another is a custom dataset to develop the proposed system. The first database is "fer2013.csv", that is available on Kaggle under the google FER2013 challenge. The dataset has 3 columns and 35,557 rows with the pixel value of images, the usage and the emotion expressed in the images. These images are classified into 7 classes namely "angry, disgust, fear, happy, sad, surprised and neutral".

The second dataset is a custom dataset created using the Spotify API. The dataset consists of songs from our own playlists that have been added to the file using the "Spotipy" library and pandas. The dataset has 416 rows and 24 columns. The various features of the song like "danceability, energy, valence, "speechiness", "etc. have been described in the dataset. The songs are classified into 4 classes of emotions namely sad, angry, happy and surprised. The first dataset is used for emotion detection and second dataset is used for song recommendation.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

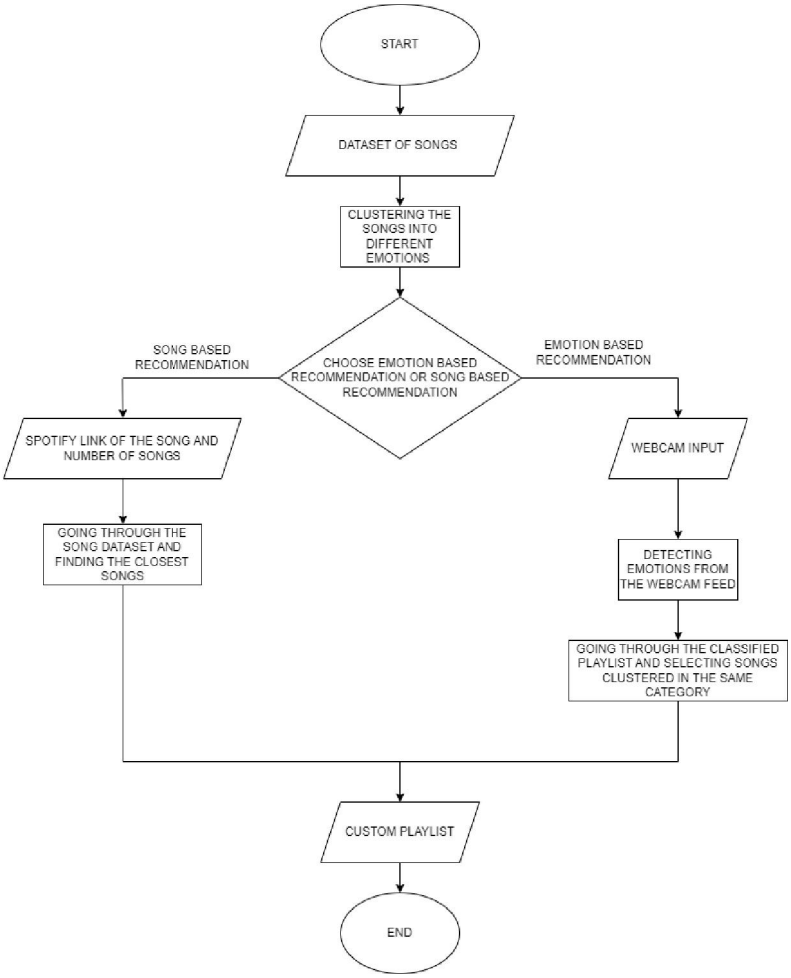

Figure 1 – The Process Flow Diagram

Table 2: This is a sample of the custom song dataset that was made. The actual dataset has over 400 rows and 26 attributes

	track_id	artist_name	track_name	danceability	energy	loudness	valence	mood
0	16nHjvEhYAxpdpeHWfBZZK	Flo Rida	Right Round	0.717	0.772	-4.264	0.734	surprised
1	3eekarcy7kvN4yt5ZFzltW	Travis Scott	HIGHEST IN THE ROOM	0.598	0.427	-8.764	0.0605	angry
2	6xcJyGpfZbuuiequtnlKt4	Travis Scott	BUTTERFLY EFFECT	0.764	0.628	-5.851	0.196	sad
3	2xLMifQCjDGFmkHkpNLD9h	Travis Scott	SICKO MODE	0.834	0.73	-3.714	0.446	sad
4	6gBFPUFcJLzWGx4lenP6h2	Travis Scott	goosebumps	0.841	0.728	-3.37	0.43	surprised

3.3 Implementation

In order to create an efficient model to perform clustering and classification tasks, the methodology has been implemented using 2 algorithms:

- K-Means Clustering
- Convolutional Neural Network 2D model
- Content Based Filtering

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

Pseudocode

```
Start
get credentials from spotify API
playlist_link = "<playlist link>"
for each song in playlist:
        extract features from song
        append features to list
create dataset using all the features extracted
save the csv file
drop the unnecessary columns – use only valence and energy
classes = fit the dataframe for KMeans(n_clusters=4, init = 'k-means++')
find the average of the cluster centers found and subtract the average from the centres
moodind = dict()
for i in centre:
        if i[0]>0 and i[1]>0:
    moodind[centre.index(i)]='surprised'
  elif i[0]<0 and i[1]<0:
    moodind[centre.index(i)]='sad'
  elif i[0]>0 and i[1]<0:
    moodind[centre.index(i)]='angry'
  elif i[0]<0 and i[1]>0:
    moodind[centre.index(i)]='happy'
for i in range(len(classes))
        classes[i] = moodind[classes[i]]
add the column to the dataframe
save the csv file with the clustered songs
end
```

The first algorithm is used to cluster songs from the dataset into emotions. The Convolutional neural network model is used for emotion detection and the Content based filtering algorithm is used to give recommendation of songs based on the user's song choice.

The user can add a wide variety of songs into the dataset from different genres. The K-means clustering algorithm will group the songs into its respective moods. Once the dataset is set up, the user can now run the program where they get a choice of song-based recommendation or emotion-based recommendation. If the user chooses emotion-based recommendation the camera opens where the Convolutional Neural Network model detects the user's emotions and on closing the camera the required song playlist is automatically created. This mode has a face detection mechanism as well for security purpose. Once set up, only if the user's face is detected the playlist is created, else it doesn't. The song-based recommendation asks the user to input the Spotify link of the song based on which the user wants recommendations and the number of songs he wants recommended. The Content based Filtering algorithm works and provides the songs that have most similar characteristics to the inputted songs.

3.3.1 K-Means Clustering for songs

Emotions are classified into 2 types – Continuous and Discrete. In the continuous class, the emotions are described in a 2D Arousal-Valence plane. Clustering is done based on the Energy and Valence of the song into 4 clusters – "Surprised", "Happy", "Sad" and "Angry".

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

3.3.2 CNN 2D Model for Emotion Detection

The fer2013 dataset was used for training the model, containing 35,887 images from 7 different emotion-classes. The pixel values from the dataset are extracted and are converted into NumPy arrays. Image Data Generator is used to augment the image dataset for greater training samples.

The CNN model is a 13-layer model that uses Convolution 2D layers. The images that were extracted created 48 x 48-pixel grayscale images of faces. 64 in the first set of layers, 128, 256,512 in the subsequent sets were used for this. A kernel size is the size of the filter matrix for the convolution. We use a kernel size of 3 means we will have a 3 x 3 filter matrix. The mathematical representation for Convolution layers is given in Equation 1. A max pooling layer has been added after every Conv2D layer. It helps in down sampling the input represented to reduce its dimensionality and allow networks to be invariant to small translations. The mathematical explanation for maxpooling is given in Equation 2.

Let x be the input data, w be the filter and y be the output feature map, then the convolution is given by:

$$y[i,j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h[m,n] \cdot x[i-m, j-n]$$

Equation 1 - CNN

If x is the input featuring map, and y is the output feature map and s is the stride. The max pooling operation is given by:

$$y[i,j] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h[m,n] \cdot x[i \times s + k, j \times s + 1]$$

Equation 2 – Max pooling

The activation function for the hidden layers is ReLU or Rectified Linear Activation which is a piece wise activation function that is used to get the output at each neuron. It is a non-linear function that helps improve the accuracy of the classification. The function is is given in Equation 3.

$$\text{ReLU activation function: } f(x) = \max(0, x)$$

Equation 3 - ReLU

The input layer takes an input of 48 x 48 which is the size of the image. We have used batch normalization after every Convolution layer to normalize the output of the batches. There is a 'Flatten' layer between the Convolution layer and the fully connected Dense layers. Flatten between the convolution and dense layers acts as a connection. The output layer gives output of 7 classes and uses the activation function 'SoftMax'.

Figure 2. explains the custom model that was created. This model was used for emotion detection.

For compiling the model, we take 3 parameters: optimizer, loss and metrics. The optimizer uses 'Adam' which controls the learning rate. Adam optimizer helps adjust the learning rate throughout the training. The learning rate determines how fast the weights that give a good accuracy is reached. The smaller the learning rate the better chance of getting a better accuracy. The loss function we will be using is 'categorical-crossentropy' which is one of the most commonly used loss function for classification. To understand the performance of the model better a performance metric is used.

The model was trained for 100 epochs on the training data with a batch size of 64 and verbose 1.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

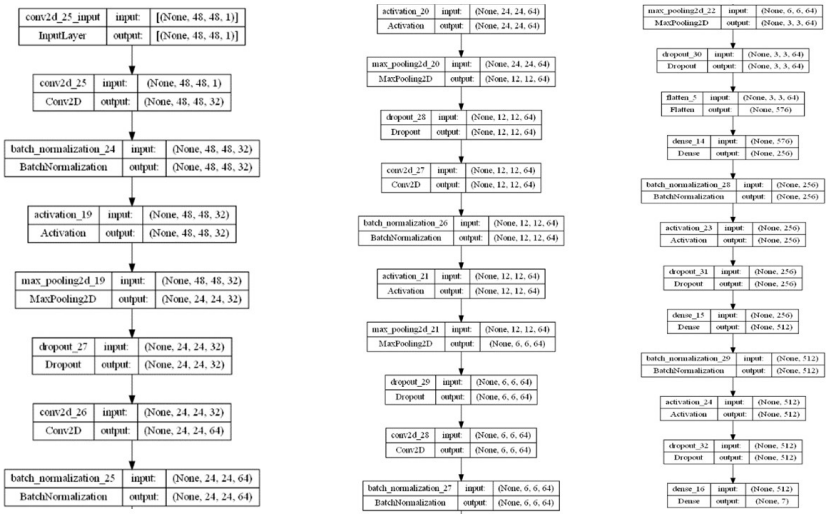

Figure 2 - Model Layers

Pseudocode

Start
Open the FER2013.csv dataset
Append the pixel values into its respective lists - X_train, Y_train, X_test, Y_test
Change all the lists to numpy arrays
Y_train= to_categorical(Y_train, num_classes=7)
Y_test = to_categorical(Y_test, num_classes=7)
X_train = X_train.reshape(X_train.shape[0], 48, 48, 1)
X_test = X_test.reshape(X_test.shape[0], 48, 48, 1)
Augment the images to increase the number of image data
Create a custom model
Fit it for 100 epochs on training data
Plot the accuracy and loss
Save the model
end

3.3.3 Content Based Filtering

Content based filtering method is the general recommendation algorithm used in most music apps. For all the songs that they listen it checks for songs that have similar energy and valence and accordingly recommends songs that the user might like.

IV. RESULTS AND DISCUSSION

The project uses 2 machine learning models and 1 deep learning model. It uses K means clustering and content-based filtering for recommendation of songs and the CNN model for emotion detection. The K means clustering algorithm is an efficient method for grouping songs as we have used the continuous classification of emotions based on energy and valence. If the energy and valence are both high then the song is suitable for surprised, if the valence is high but the energy is low then the song is suitable for happy, if the valence and energy both are low the song is suitable for sad and if energy is high and valence is low then the song is suitable for angry mood. The final project has 2 modes the first mode being the emotion detection mode where the camera switches on and detects the emotion of the user. On closing the camera, a playlist is created by filtering the mixed playlist that is custom built for the mood. If no emotion is detected then it shows neutral and gives a mixed set of songs from all the emotions. In the content-based filtering model

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

it takes the link of the song and extracts the information of the song using the Spotify API. It then takes the valence and energy of the song and uses the algorithm to get the closest songs and gives the output.

4.1 Result Analysis

The emotion detection model provides an accuracy of 94.6% and a validation accuracy of 67.4%. The model was saved and tested and it successfully worked.

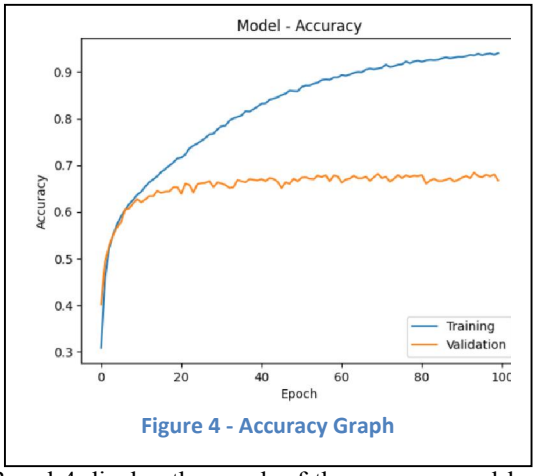

Figure 4 - Accuracy Graph

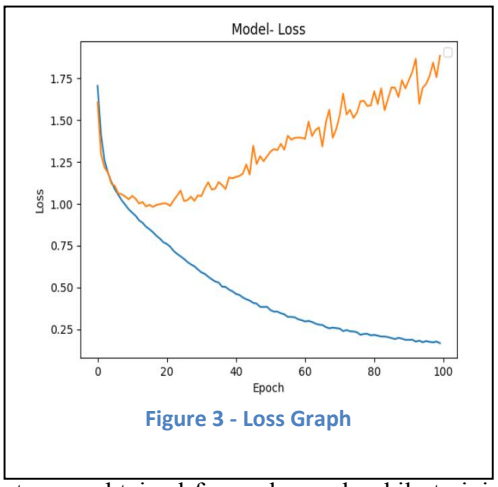

Figure 3 - Loss Graph

Figure 3 and 4 display the graph of the accuracy and loss that were obtained for each epoch while training the CNN model.

Pseudocode

```
Start
Load the spotify API with credentials
Path = “images”
Append all the images into a list images
Append the file names of images into a list classnames
Define function findencoding(images):
        encodelist = []
   for img in images:
      img = cv2.cvtColor(img,cv2.COLOR_BGR2RGB)
      encode = face_recognition.face_encodings(img)[0]
      encodelist.append(encode)
   return encodelist
encode the images
load the emotion detection model
use Cascade classifier to detect frontal face
emotion_detection = ('angry', 'disgust', 'fear', 'happy', 'sad', 'surprise', 'neutral')
def recommend_base_mood(mood,name):
   df = pd.read_csv('songdataset_withmood.csv')
   if mood=='happy':
      df_mood = df[df.mood==mood]
   elif mood=='surprised':
      df_mood = df[df.mood==mood]
   elif (mood=='sad')or(mood=='fear'):
      df_mood = df[df.mood=='sad']
   elif (mood=='angry')or(mood=='disgust'):
```

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

```
        df_mood = df[df.mood=='angry']
    else:
        df_mood = df.sample(n=100)
    df_mood.drop(['mood'],axis = 1)

    df_mood.to_csv(f'{name}_{mood}.csv')
open video capture
pass the frames through face recognition and emotion detection model
place bounding boxes
once window closes give recommendation based on last detected emotion
end
```

V. CONCLUSION

This paper summarizes the creation of a recommender system for music based on emotions detected from facial expression. This technology identifies basic human emotions and comprehends the suitable action/recommendation required. This system efficiently captures the mood of the listener and customizes a playlist based on this data. To assess the robustness of the produced system, it must be tested under various lighting scenarios. Additionally, the system was able to obtain the user's updated photos and properly update its classifier and training dataset. The system was developed utilizing a facial landmarks method and it was put to the test in variety of circumstances to see how it would perform.

REFERENCES

[1]. Divya Garg, Gyanendra K. Verma, Emotion Recognition in Valence-Arousal Space from Multi-channel EEG data and Wavelet based Deep Learning Framework, Procedia Computer Science, Volume 171, 2020, Pages 857-867, issn 1877-0509, https://doi.org/10.1016/j.proc.2020.04.093.

[2]. James, H.I., Arnold, J.J.A., Ruban, J.M.M., Tamilarasan, M. and Saranya, R., 2019. Emotion based music recommendation system. Emotion, 6(3).

[3]. Author Hupont, I., Baldassarri, S. & Cerezo, E. Facial emotional classification: from a discrete perspective to a continuous emotional space. Pattern Anal Applic 16, 41–54 (2013). https://doi.org/10.1007/s10044-012-0286-6

[4]. Song, Y., Dixon, S., & Pearce, M. (2012, June). A survey of music recommendation systems and future perspectives. In 9th international symposium on computer music modeling and retrieval (Vol. 4, pp. 395-410).

[5]. Jaiswal, A., Raju, A. K., & Deb, S. (2020, June). Facial emotion detection using deep learning. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-5). IEEE.

[6]. Dagar, D., Hudait, A., Tripathy, H. K., & Das, M. N. (2016, May). Automatic emotion detection model from facial expression. In 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT) (pp. 77-85). IEEE.

[7]. Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017, September). Emotion detection: a technology review. In Proceedings of the XVIII international conference on human computer interaction (pp. 1- 8).

[8]. Sun, Y., Sebe, N., Lew, M. S., & Gevers, T. (2004, May). Authentic emotion detection in real-time video. In International Workshop on Computer Vision in Human-Computer Interaction (pp. 94-104). Springer, Berlin, Heidelberg.

[9]. Chen, HC., Chen, A.L.P. A Music Recommendation System Based on Music and User Grouping. J Intell Inf Syst 24, 113–132 (2005). https://doi.org/10.1007/s10844-005-0319-3

[10]. De Prisco, R., Guarino, A., Malandrino, D., & Zaccagnino, R. (2022). Induced Emotion-Based Music Recommendation through Reinforcement Learning. Applied Sciences, 12(21), 11209

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 3, June 2023

[11]. Joshi, S., Jain, T., & Nair, N. (2021, July). Emotion based music recommendation system using LSTM-CNN architecture. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 01-06). IEEE

[12]. Shalini, S. K., Jaichandran, R., Leelavathy, S., Raviraghul, R., Ranjitha, J., & Saravanakumar, N. (2021). Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques. Turkish journal of computer and mathematics education, 12(2), 912-917

[13]. James, H. I., Arnold, J. J. A., Ruban, J. M. M., Tamilarasan, M., & Saranya, R. (2019). Emotion based music recommendation system. Emotion, 6(03)

[14]. Song, Y., Dixon, S., & Pearce, M. (2012, June). A survey of music recommendation systems and future perspectives. In 9th international symposium on computer music modeling and retrieval (Vol. 4, pp. 395-410)