An Improved Music Recommendation System for Facial Recognition and Mood Detection

Shakthi Sri S[1,*], *Sathya* S[1], and *Pandiarajan* T[1]
[1]Rajalakshmi Institute of Technology, Chennai, India

Abstract. A user can choose the song using a number of strategies. This system finds the user's mood and recommends music in accordance with that mood. The main objective is to accurately assess the user's mood in a friendly way. Listeners are introduced to a machine learning application that plays song based on the user's mood. Finding the user's expression using Artificial Neural Network model, which expresses their emotions, can help with this. Compared to speaking, facial expressions reveal a lot more about a person. Users are finding it challenging to select the song that they want to listen to due to the daily increase in the number of songs being created. In order to address, our recommendation system gives users a selection of songs that are automatically added to their customized playlist based on the user's mood.

1 Introduction

People tend to judge another person's mood based on their facial expressions. If this human ability is acquired by electronic devices using machine learning, applications with enormous utility may be created. Music has long been a favorite among people.It encourages greater expression and helps us better understand our moods and emotions. The power of music is to improve our feelings. For instance, we can listen to cheerful tunes if we want to feel joyful. A sorrowful song may be comforting to listen to after an awful incident. Science claims that listening to depressing music can actually improve one's mood. This research suggests a related application, namely an emotion-based song recommendation system. The facial expression is an essential component that defines a person's mood. To capture a face, a camera is required, and this image serves as the input. Prepare a song list based on person's mood to avoid human effort in order to construct a playing list and segregate tracks into distinct lists. This technique assists in creatinga playlist that is suitable for a person's mood. The goal of an emotion-based song recommendation system is to analyze the user's feelings and build a playlist using the results. As a result, the main focus of our suggested system is on characterizinghuman emotions in order to develop an emotion-based song player. It takes a long time to run a typical music player since the user has to go through their song collection to locate music that fits their mood. This is a stressful task, and choosingthe right song based on the mood is usually a challenging issue. Some recommendation systems

* Corresponding author: shakthisri.s.2019.cse@ritchennai.edu.in

with mood detection have a high level of complexity, which influences how effectively the application performs in real time.Our application uses a user's facial expressions to determine their stress and emotion in order to allowing them to listen to any number of songs they choose in a custom playlist.

2 Methodologies

2.1 CNN

A face detector model was made using the Convolution Neural Network (CNN), which takes faces and lines from images, including such facial landmarks. A dataset consisting of about 5 different human emotions is obtained, which is preprocessed such as image reshaping, resizing and conversion to an array form. The training images are then put into a convolution neural network, which employs the derived pixel information to predict the mood of the person in the image as shown in Fig.1. This is where the action begins as shown in Fig.2. Convolution layer, Pooling layer, Dropout layer, Flatten layer, Dense layer, Activation Function, and Optimizers are some of the layers that must be processed.

After the model is processed and trained successfully, the software can identify the Human feelings Classification image contained in the dataset and then comparing the test and trained images to predict the emotion.

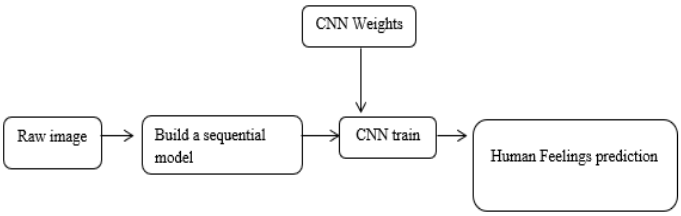

Fig. 1. Flow of Preprocessing and Training the CNN Model

2.1.1 Convolution Layer

A major element of CNN, the convolutional layer performs the majority of the calculations. Dot product is a mathematical operation that this layer performs between the filters and the input image. which are represented as matrices. Thesecond matrix is recognized as the filter or kernel, while the initial matrix is animage matrix that includes the pixel values of the mage (typically 0-255). detection of activities. The kernel and picture matrix both support three channels because the majority of images have three channels (RGB). At its core, 2D convolution is a relatively straightforward process: you begin with a kernel, which is just a small matrix of weights. Using the portion of the input it is currently on, this kernel "slides" over the 2D input data, executing an elementwise multiplication, and combining the results into a single output pixel. Every region the kernel slides over is subjected to this process again, resulting in the creation of a second 2D feature matrix. The input features are effectively positioned roughly in the same location as the output pixel on the input layer, and the output features are the weighted sums of those features, where the weights are the values of the kernel itself. Whether an input feature is in the same region of the kernel as the output or not determines whether it falls within this "roughly same location". This means that the number of input features that are integrated to create a new output feature depends directly on the size of the kernel. All of this stands in stark contrast to a layer that is entirely integrated. 5 x 5 = 25 input features and 3 x 3 = 9 output features are present in the example above. Every output feature would be the weighted sum of every input feature if this were a typical fully connected layer, which has a weight matrix of 259 = 225 parameters. Instead of

"looking" at every input feature, convolutions let us to perform this transformation with only 9 parameters, with each output feature only getting to "look" at the input features that come from roughly the same position. Do keep this in mind because it will be important for our interaction later.

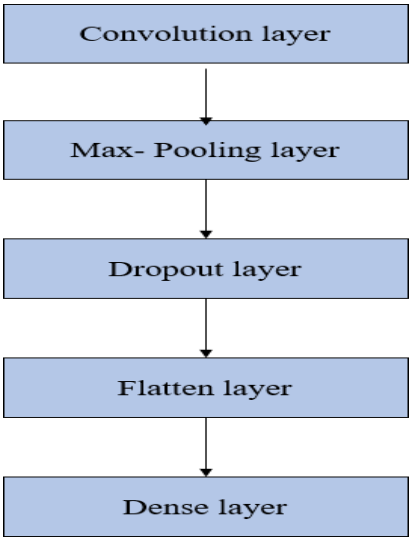

Fig. 2: Flow of Mood Prediction Model

2.1.2 Max-Pooling layer

The pooling layer's important usage is for feature reduction of the previous output layer. A minimum feature matrix accelerates computation and makes features quite resistant to noise and outliers. There are different methods for pooling: Average and maximum pooling. The most usual, but even so, Maximum Pooling. It moves through the tiny windows of the matrix and calculates the peak power from a small matrix. The input is down sampled along its spatial dimensions (height and width) by obtaining the maximum value for each input channel over an input window with a size determined by pool_size. Each dimension of the window is moved one step at a time.

When the "valid" padding option is used, the output has the following spatial form (number of rows or columns):

output_shape = math.floor((input_shape - pool_size) / strides) + 1 (when input_shape >= pool_size) (1)

When utilising the "same" padding option, the output shape is as follows:

output_shape = math.floor((input_shape - 1) / strides) + 1 (2)

2.1.3 Dropout layer

The Dropout layer is a filtering layer that decreases the contributions of a chosen few synapses to zero whereas allowing the contributions of the majorityunchanged. A Dropout layer can be given to the feature matrix, in which case it removes some of its features, or it may be implemented to a hidden layer, inwhich case it decreases the input of a small number

of neurons to zero. Dropout is a great method for decreasing overfitting and increasing accuracy.

2.1.4 Flatten layer

The feature matrix is reshaped by a flatten layer to have a shape that corresponds to the number of pixels it includes. It is used to flatten the dimensions of the image obtained after convolving it. Making a component a one- dimensional array is equal to doing this. For example, if flatten is applied to layer having input shape as (batch_size, 2,2), then the output shape of the layer will be (batch_size, 4)

2.1.5 Dense layer

An artificial neural network layer connected after CNN is termed the dense layer. The most widely and frequently used layer is this kind. This layer, whichis included in the artificial neural network, is added just after convolution layerto derive the result prediction.

The following procedure is implemented by Dense:

$$\text{output} = \text{activation}(\text{dot}(\text{input}, \text{kernel}) + \text{bias}) \quad (3)$$

where activation is the element-wise activation function passed as the activation argument, kernel is a weights matrix created by the layer, and bias is a bias vector created by the layer (only applicable if use_bias is True). These are all the attributes of Dense.

2.2 Image Data Generator

It is that rescales the image, applies shear in some range, zooms the image and does horizontal flipping with the image. This includes all possible orientation of the image.

2.3 Training Process

The function used to prepare data from the train_dataset directory is train_datagen.flow_from_directory. The target size of the image is specified by target_size. To prepare test data for the model, use Test_datagen.flow_from_directory. Everything is done in a same manner as above. In addition to steps_per_epochs, fit_generator is used to fit the data into the model created above. Fit_generator determines how many times the model will run on the training data.

2.4 Epochs

It tells us the number of times that the model will be trained in forward and backward pass.

2.5 Validation Process

Validation_data is used to feed the validation/test data into the model. Validation_steps denotes the number of validation/test samples.

3 Existing System

A system that trains input photographs to recognise facial expressions using the classification algorithm OpenCV and the point detection algorithm to extract features from the input images. Pre-processing is applied to the recovered image when it is extracted from the camera. Canny edge detection is used to apply edge detection technology. The image with edge detection is segmented. Following this, face detection and feature extraction occur. Then, this is put into use by online services. Facial expression recognition using VGG-16

CNN is also created. When an emotion was identified, the user's customised playlist would play the song that best matched their feelings.

4 Proposed System

This system classifies the face expression and planned to design a deep learning technique. It proposed a system for predicting face expression. Samples of more images are collected that compared with different classes. The primary attributes ofthe image are relied upon the shape and texture-oriented features. Initially preprocessing our dataset and implementing more than two CNN architecture. Eacharchitecture gives a different kind of accuracy, so compare each one architecture. Finally, build a hierarchical model for saving our trained model. Once the model issaved then we can deploy it in any web browser by using Django Framework.

5 Human Emotion Prediction Model

5.1 Dataset Information

The Kaggle datasets that we are employing in this model's creation include the dataof 1300 trained and 180 test image records of features extracted. These data sets include classes like angry, disgust, happy, sad and surprise.

5.2 Data Preprocessing Stage

In this process, we will load the data, extract features from it, then split the dataset into training and testing sets. Then, we'll initialize an Artificial Neural Network andtrain the model. Finally, we'll calculate the accuracy of our model.

5.3 Algorithm Implementations

The approach for recognizing human emotions is based on a two-channel architecture that can do effectively. The CNN's inception layer receives the human emotions after being trimmed and removed. Utilizing a convolution neural network, the training process entails feature extraction and classification as shown in Fig.3.

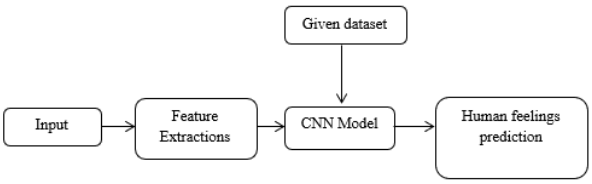

Fig. 3: Flowchart for the Emotion Prediction

5.3.1 Types of CNN Architecture

5.3.1.1 Res NET

Convolutional neural networks, notably those used in the application of deep learning to image processing, have significantly influenced the field of machine learning. One such network is known as Res Net. The first convolutional network to utilise a GPU to improve performance was Res Net. Five convolutional layers, three max-pooling levels, two normalisation layers, two fully connected layers, and one softmax layer make up the Res Net

architecture as represented in Fig.4. Convolutional filters and a nonlinear activation function called ReLU make up each convolutional layer. Max pooling is carried out using the pooling layers.

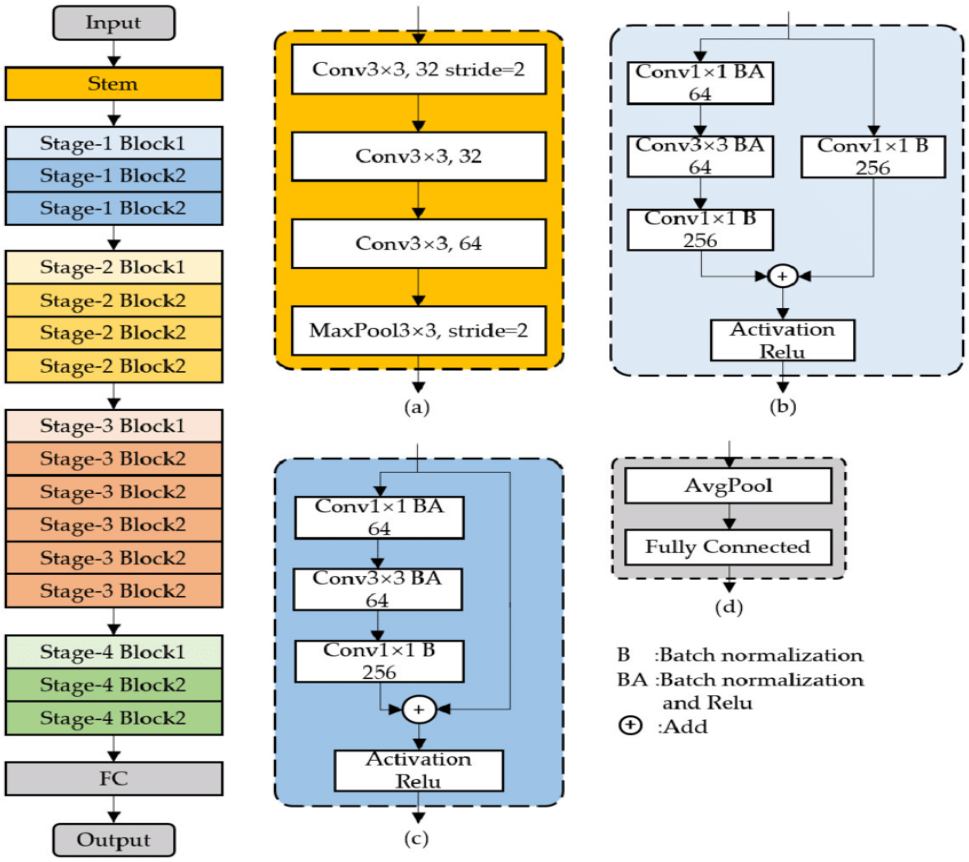

Fig. 4: Architecture of Res Net

5.3.1.2 Le NET

Convolutional layer, pooling layer, and full link layer are the three fundamental deep learning modules that make up Le Net, a compact network shown in Fig.5. Other deep learning models are built on this foundation. Here, we look closely at LeNet5. Develop a deeper grasp of the convolutional layer and pooling layer at the same time through example analysis. Let's examine Lenet-5's architecture. The network is known as Lenet-5 since it contains five layers with learnable parameters. It combines average pooling with three sets of convolutional layers. We have two fully linked layers following the convolution and average pooling layers. Finally, a Softmax classifier places the photos in the appropriate class.

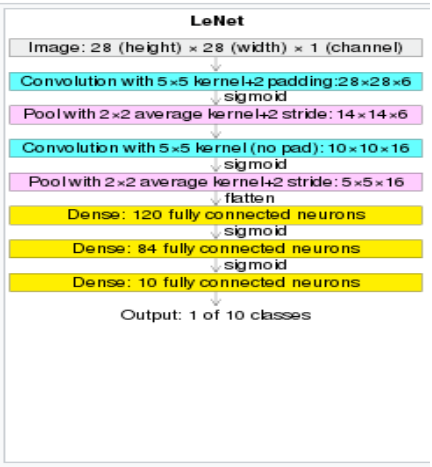

Fig. 5: Architecture of Le Net

5.3.2 Manual Net

Utilising the keras preprocessing image data generator tool, we must input our data set and construct size, rescale, range, zoom range, and horizontal flip. Then, using the data generator tool, we import our image dataset from the folder. Here, we specify the parameters for the train, test, and validation phases as well as the target size, batch size, and class-mode. After executing this function, we must train the network we built by adding CNN layers.

5.3.3 Human Feelings

5.3.3.1 Angry

```
Trainned data for angry:

 ====== Images in:  Dataset/train/angry
images_count:   3995
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

5.3.3.2 Disguist

```
Trainned data for disgust:

 ====== Images in:  Dataset/train/disgust
images_count:   436
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

5.3.3.3 Happy

```
Trainned data for happy:

 ====== Images in:  Dataset/train/happy
images_count:   7215
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

5.3.3.4 Sad

```
Trainned data for sad:

 ====== Images in:  Dataset/train/sad
images_count:   492
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

5.3.3.5 Surprise

```
Trainned data for surprise:

 ====== Images in:  Dataset/train/surprise
images_count:   519
min_width:      48
max_width:      48
min_height:     48
max_height:     48
```

5.3.4 Training the module by given dataset

To train our dataset using classifier and fit generator function also we make training steps per epoch's then total number of epochs, validation data and validation steps using this data we can train our dataset and predict the emotion as shown in Fig.6.

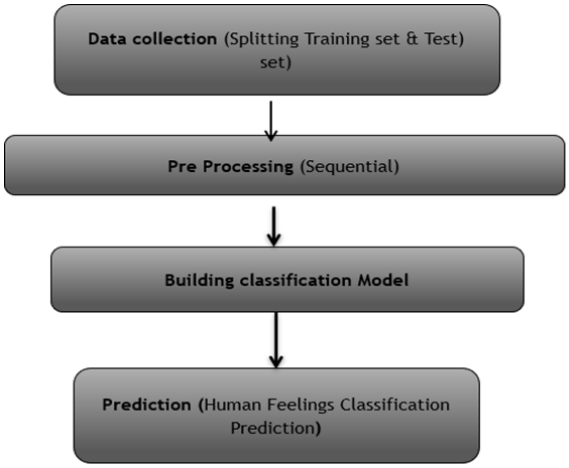

Fig. 6. Architecture of Emotion Prediction Model

5.3.5 Deployment Stage

In this module the trained deep learning model is converted into hierarchicaldata format file (.h5 file) which is then deployed in our Django framework for providing better user interface and predicting the output whether the givenimage has human feelings and recommends song based on mood.

6 Sample Output

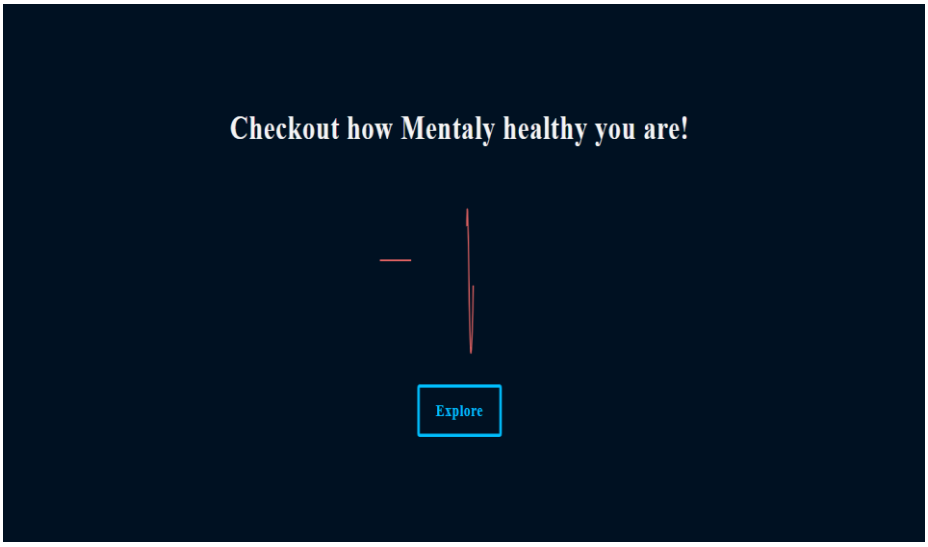

Fig .7. Opening page

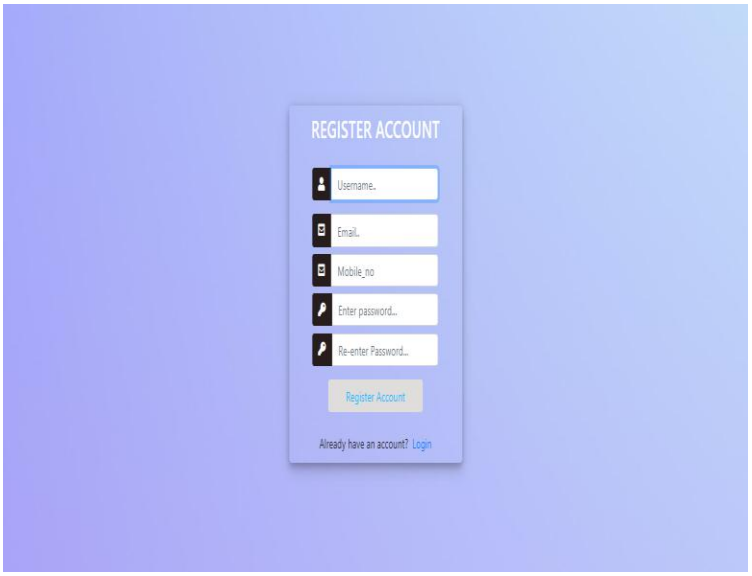

Fig .8. Registration Page

Fig .9. Login Page

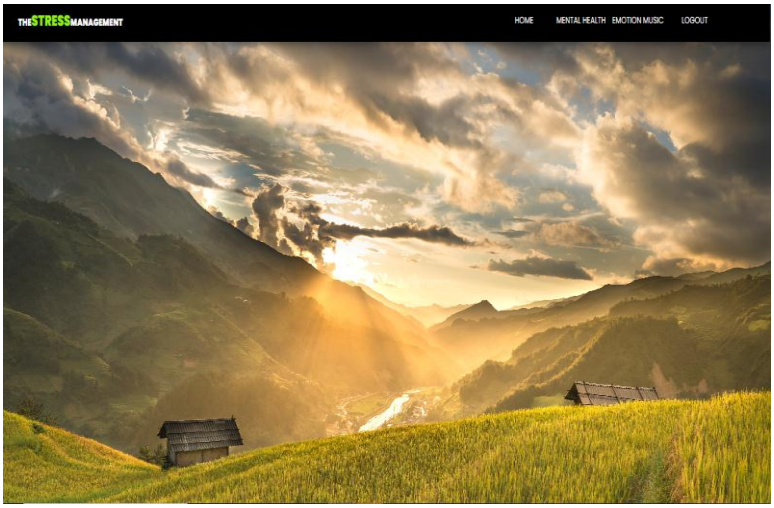

Fig.10. Home Page

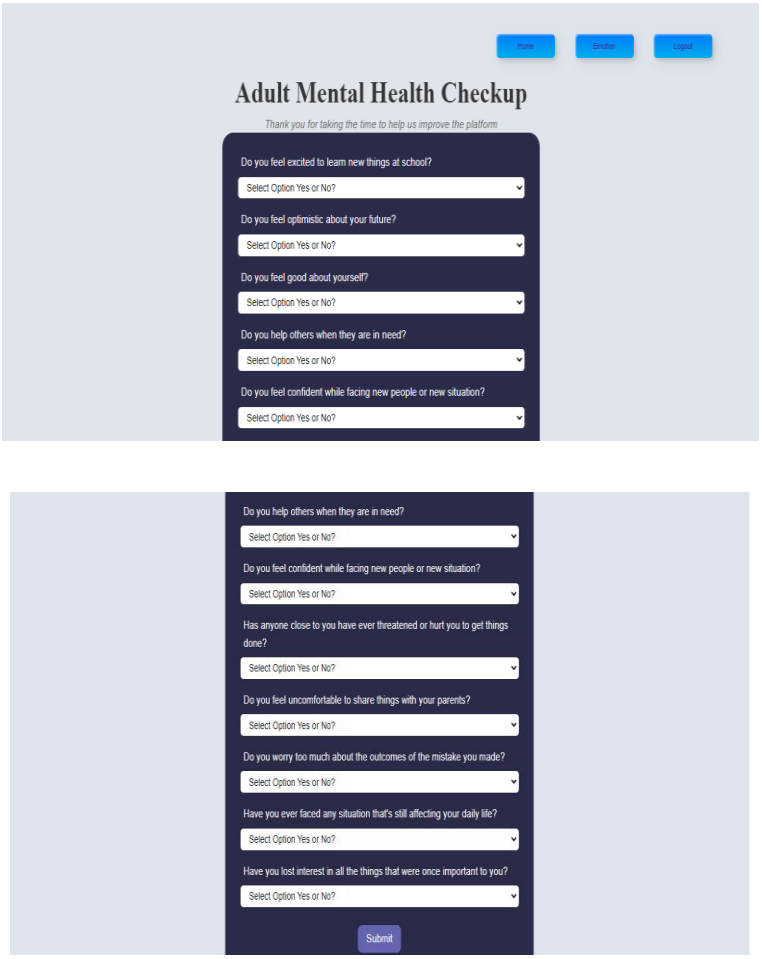

Fig .11. Mental Health Checkup

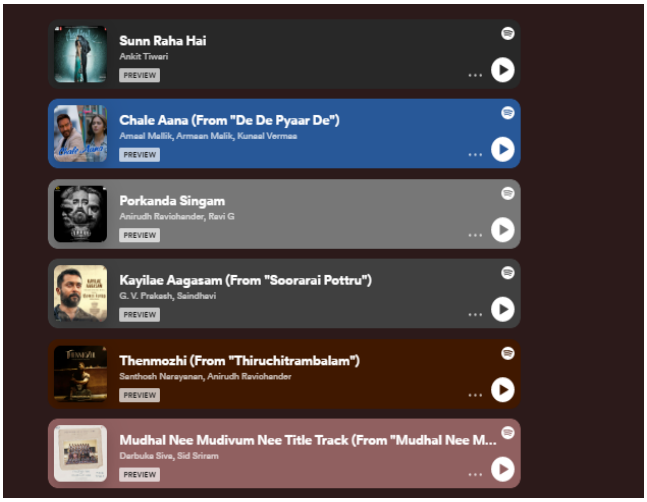

Fig. 12. Sad Songs Recommendation from Spotify

7 Results and Discussion

This offers two distinct descriptions of a song recommendation system, utilising image. The image-based model, which uses CNN with the Kaggle dataset achieves a face recognition accuracy score of 98.9%.

Table 1. Represent the emotion accuracy.

Emotion	Expected Result	Calculated Result	Accuracy
Happy	93	90	96.77%
Sad	91	89	95.65%
Disgust	87	86	98.85%
Angry	96	92	95.83%
Surprise	90	89	98.88%

8 Future Works

As one of the most popular streaming music services, Spotify also offers a number of developer tools that make integrating with other systems simple. Future work will concentrate on creating a music recommender that can recognize various emotions not just from an image, but also from movies or real-time camera feeds. So that it can be used in electronic and automotive systems, including those in automobiles, laptops, and mobile phones, etc.

9 Conclusion

With the purpose of supporting face expression recognition, a wide variety of image processing systems have been created. Our contributions give a method for designing and implementing emotion-based devices in addition to the theoretical basis. The suggested system can analyze facial photos to identify fundamental emotions and create playlists in response.

References

1. Dr. Reena Sonkusare, Ghosh, Sanskar Laddha, Sudhanshu Kulkarni, *Music Recommendation System Based On Emotion Detection Using Image Processing And Deep Networks*, 2nd International Conference On Intelligent Technologies, CONIT Karnataka, India, June 24-26, (2022).
2. Freya Vora, Prof. Sanjay Vidhani, Arya Karambelkar, Pram Mamania, Jainam Chhadwa, *Mood Indicator: Music And Movie Recommendation System Using Facial Emotion*, 5th International Conference On Advances In Science And Technology, ICAST (2022).
3. Sumit Kumar, Vicky Kumar, P. Venkateshwari, *Music Recommendation based on User Mood*, 9th International Conference on Computing for Sustainable Global Development, (2022).
4. A. Panneerselvam, V. Sneharanthna, M. Gumasekar, K. Logeswaran, M. Suganneshan, *Improved Facial Emotion Recognition using Yolo and DeepFace for Music suggestion*, Proceedings of the Third International Conference on Electronics and Sustainable Communication Systems, ICESC, (2022).
5. Sanchit Gajam, Madhav Lahoti, Nataasha Raul, Aditya Kasat, *Music Recommendation system based on facial mood detection*, Third International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT, (2022).
6. Shreel Trivedi, Mridu Pant, Samiksha Aggarwal, Amita Dev, Ritu Rani, Poonam Bansal, *Driver's Companion Drowsiness Detection And Emotion-Based Music Recommendation System*, International Conference On Computing, Communication, and Intelligent Systems, ICCCIS, (2022).
7. Azeem Saleem Gaded, Vijay Prakash Sharma, Deevesh Chaudhary, Shikha Sharma, Sunil Kumar, *Emotion Based Music Recommendation System*, 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO, (2021).
8. Rajeev Kumar Gupta, Kevin Patel, *Song Playlist Generator System Based On FacialExpression and Song Mood*, International Conference on Artificial Intelligence and MachineVision, AIMV, (2021).
9. Shilpa Hari Prakash, Pranesh Ulleri, Kiran B Zenith, Jinesh, Gowri S Nair, Kannimoola, *Music Recommendation System based on Emotion*, 12th ICCCNT, (2021).
10. Tanuj Jain, Saurav Joshi, Nidhi Nair, *Emotion Based Music Recommendation System Using LSTM – CNN Architecture*,12th ICCCNT, (2021).
11. Rajdeep Mangrola, Shavak Chauhan, D.Viji, *Analysis of Intelligent movie recommender system from facial expression*, Proceedings of the Fifth International Conference on Computing Methodologies and Communication, ICCMC, (2021).
12. R. Jaichandran, K. ShanthaShalini, S. Leelavathy, R. Raviraghul, J. Ranjitha, N. Saravanakumar, *Facial Emotion Based Music Recommendation System using Computer Vision And Machine Learning Techniques*, Turkish Journal of Computer and Mathematics Education, **Vol.12 No.1**, pp.9012-917, 05 April, (2021).
13. A. Dhanush Ram, G. Chidambaram, G. Kiran, P. Shivesh Karthic, Abdul Kaiyum, *Music Recommendation System Using Emotion Recognition*, International Research Journal of Engineering and Technology, IRJET, **Vol. 08** , Issue: 07, July (2021).
14. Vincenzo, Moscato, Giancarlo Sperli, Antonio Picarriello, *An emotional recommender system for music*.", IEEE Intelligent Systems (2020).
15. Yu, Ziyang, *Research on Automatic Music Recommendation Algorithm Based on Facial Micro-expression Recognition*, 39th Chinese Control Conference, CCC, IEEE, (2020).
16. B. Perumal, Sanuvel, Muthukumaran Elangovan, Deny John, *Music recommendation system based on facial emotion recognition*, (2020).
17. B. Sriman, S.K. Shriram, R. Sathish Kumar, *Virtual Assistant for Automatic Emotion Monitoring using Perceived Stress Scale (PSS)*, (2022).