IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.047 ⋇ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10733

Emotion based smart music player using Deep Learning

Bharath K.V[1], Chandan M.N[2]

PG Scholar, Dept. Of MCA,P.E.S Colege of Engineering,Mandya[1]

Assistant Professor, Dept. Of MCA,P.E.S Colege of Engineering,Mandya[2]

Abstract: Songs have long been a common choice as a means of expression to describe and comprehend human feelings. We can greatly benefit from trustworthy emotion-based classification systems when it comes to interpreting their significance. The outcomes of study into the classification of music based on emotions, however, have not been the best. We provide an affective cross-platform music player in this project that makes music suggestions based on the user's current mood. By merging the capabilities of emotion context reasoning within our adaptive music recommendation engine, EMP offers intelligent mood-based music recommendations.

Keywords: Deep Learning, CNN Algorithm, Face Expression, Music Classification.

I. INTRODUCTION

One's life is significantly impacted by music. It serves as a significant form of entertainment and is frequently used therapeutically. The development of technology and ongoing breakthroughs in multimedia have led to the creation of sophisticated music players that are rich in features like volume modulation, genre classification, and more. Despite the fact that this function effectively addresses users' needs, users still need to browse their playlists for songs that express their feelings.In a conventional music player, a user is required to independently go through his playlist and choose songs that would enhance his mood and emotional experience. This way of selecting songs is difficult and time-consuming, and the user may struggle to discover the right music.Our music player contains three modules: Emotion Module, Random music player Module and queue based Module. The Emotion Module takes an image of the user’s face as an input and makes use of CNN algorithm to identify their mood in a accurate way. The Music Classification Module makes use of audio features to achieve a remarkable result and while classifying songs into 4 different mood classes. The Recommendation Module suggests songs to the user by mapping their emotions to the mood type of the song, taking into consideration the preferences of the user.

II. LITERATURE SURVEY

Here in [1], the researchers propose an emotion-based music recommendation and classification framework for precisely categorising songs by watching how individuals engage with one another and with emotional songs. When introducing new songs to an IoT software application, procedures must be developed specifically that rapidly categorise the characters based on user emotions. [2] It provides a summary of the numerous approaches researchers have looked into to determine moods based on facial expressions of human emotion. The calculation of extracted facial structure enables the location of a person's eyes, mouth, and nose to be identified on their face, as well as the detection of facial movements.The SVM algorithm is utilised for emotion recognition, while the Haar cascade is employed for noise reduction, feature calculation, and picture contrast. We look through and describe numerous multimodal deep learning approaches for human emotion detection in [3]. The findings suggest that emotion detection can be improved and made more accurate by using a multimodal approach from biological signals to identify emotional states. In [4], we provide a deep learning approach that outperforms existing models on a number of datasets, including FER-2013, CK+, FERG, and JAFFE and can concentrate on important facial traits. This method is built on an attentional convolutional network. In [5], we go over how to utilise a deep convolutional neural network (DCNN) with the Google TensorFlow machine learning tools to identify facial expressions in videos. The method was tested on two datasets and used to analyse ten emotions from the Amsterdam Dynamic Facial Expression Set-Bath Intensity Variations (ADFES-BIV) dataset.

III. EXISTING SYSTEM

The classification of the user's behaviour and emotional state has been recommended using a variety of approaches.

- Existing music player plays the songs randomly which are present in the music folder.
- There is no emotion tracing or identification of the human facial emotions and play the song.

DOI: 10.17148/IJIREEICE.2022.10733

- Music is recommended based on what type of songs user plays every time.
- If the user listens sad songs then the same kind of pitch songs are always recommended to the user.
- Music pitch is always monitored based on the user login id and always recommends same kind of music irrespective of human facial emotions.
- Existing system has recommendation system but not classification system based on real time input.

DISADVANTAGES

- Always same kind of songs is played.
- User need to manually search songs based on his mood.
- Very old methodology.

IV. PROPOSED SYSTEM

- We gain knowledge about the application of convolutional neural networks (CNN) to visual perception.
- CNN is used to create an effective differentiating calculation model Feelings were 4 emotions, namely, happiness, sadness, anger and neutrality, with better accuracy.
- Introduces a new platform-based music player called EMP that makes song recommendations in real time based on user experience.
- Through the integration of the strength of emotional context thinking with our adaptable music complementary system, the EMP offers a music-based music suggestion.
- The user's choices are taken into account when the recommendation module transfers the user's thoughts about the genre of song.
- User emotions are determined by the emotional module. The music classification module reveals the most crucial and vital audio information in a song.
- The proposed solution is more compact and performs better over time..
- The system can be trained and tested to get real-time input or real-time images.

ADVANTAGES

- User will have no worries of searching the songs based on his emotions manually.
- Every time the song is played the system verifies the human facial emotions.
- The system is less complex.
- Better performance.
- Good accuracy.

V. METHODOLOGY

1. Dataset

The data is fetched from Kaggle Facial Expression Recognition Challenge, FER2013, was used to train the model. Grayscale pictures of faces measuring 48 x 48 pixels make up the data. Each of the seven emotion categories—angry, disgusted, fear, pleased, sad, surprise, and neutral is depicted by one of the faces. Here, we are making use of four emotions namely anger, happiness, sadness, and neutral. There are 26,217 photos in all that go with these feelings. The photos were divided into four categories: happy (7213 samples), sad (4830 samples), neutral (4965 samples), and angry (3995 samples). The dataset is publicly accessible and open source.

2. CNN Algorithm

CNN is a feedforward neural network with a deep structure that uses the convolution kernel to complete the convolution operation, making it one of the deep learning techniques. It is widely utilised in many different industries, but image recognition uses it the most.
Convolutional, pooling, and full connection layers are typically present in CNNs. The local features of the image are mostly extracted by the convolution layer. The pooling layer's major purpose is to compress the pertinent picture characteristics that were taken from the convolution layer's previous layer in order to create new, smaller image features while still preserving the main input images. By cascading and fusing all of the local features obtained from the convolution layer and the full connection layer, the full connection layer is an efficient way to retrieve global features. Finally, the categorization effect is achieved using softmax. 4.1 displays the CNN network frame diagram..

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified ⁘ Impact Factor 7.047 ⁘ Vol. 10, Issue 7, July 2022
DOI: 10.17148/IJIREEICE.2022.10733

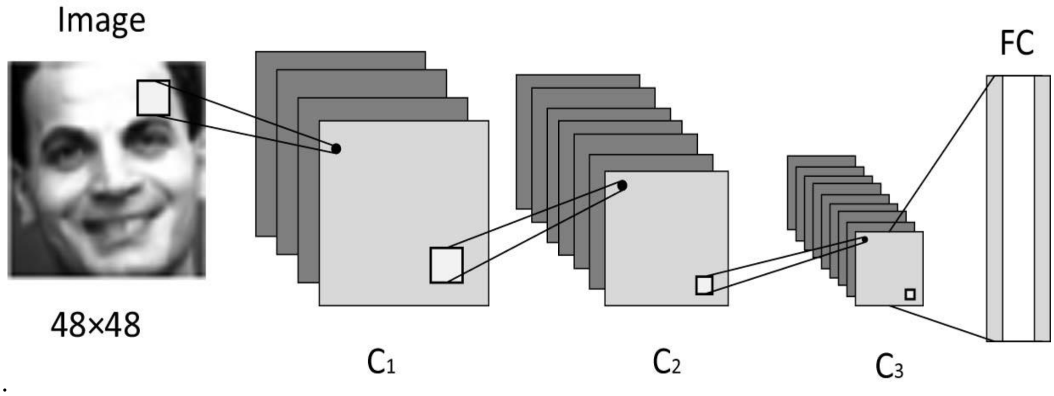

Figure 2.1 CNN network.

It is simple to make the gradient vanish when the convolutional neural network is layered too deeply, which will lead to a reduction in performance. The CNN model used in this article consists of three sections, each of which is coupled by a complete connection layer and a Softmax layer after being divided into a convolutional level and a max - pooling. First, 48 48 grayscale images of face expressions are input, and the first convolution layer is convolutioned with ten convolution tests. Next, 20 convolution checks are convolved on the first pooling layer's output. Second, 40 convolution checks are convolved onto the output of the second pooling layer. Expand it to become a fully integrated framework. They include the size of the convolution kernel is 5 5, the pooling layer uses 2 2 maximum pooling, there are 100 neurons, and the softmax layer classifies expressions.
Generally, the calculation formula of the convolution layer is as follows :

$$Cj=f\sum i=1NMi*Li,j+pj.$$

where Mi denotes the input matrix, Li,j the convolution kernel value, pj the offset term, Cj the output matrix, and f() the activation function. In this study, RELU is utilized as the activation function, and its definition is as follows.

$$f(x)=max(0,x).$$

Here we develop an end-to-end trainable CNN network. A convolutional kernel with a size of 5 5 is utilized in order to extract global information. We create a network structure that is simple to extract global information compared to CNNs with numerous parameters, while the training time is quick and better suited for the experiment data.

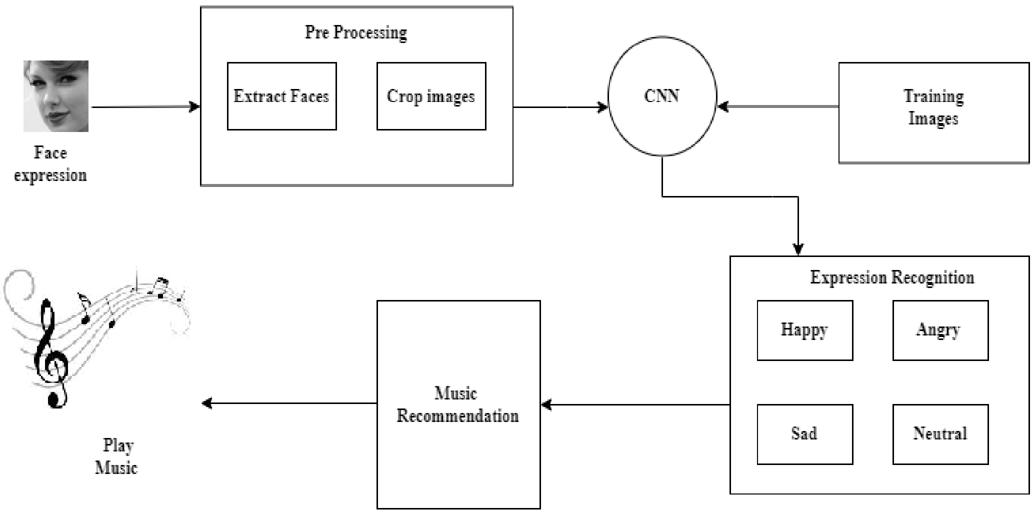

Figure 2.2 System Architecture

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ∺ Impact Factor 7.047 ∺ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10733

3. Feature Extraction using CNN

In this model, using the OpenCV library, the live video is captured with the help of web camera the Haar cascade method is adopted. The images in FER2013 have been trained and then the input is passed to several layers as shown in Figure 5.3.

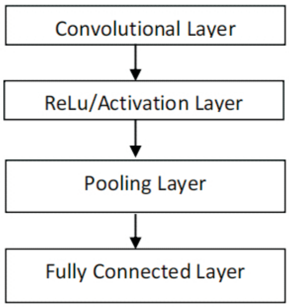

Figure 3.1 CNN Layers

3.1 Convolutional Layer

In this layer, the image will be converted into pixels and stored as array. It helps to extract the image features and reduce the dimensionality. The array of pixels will be multiplied with the filter having any 2 x 2 values (filter with low values will make calculation easier). Thus after calculation the result matrix is obtained as shown in Figure 3.1.1.

4	21	54	92
2	22	54	36
3	42	37	86

★

0	-1
1	1

3	22	-2	10
23	25	87	139

Figure 3.1.1 Calculation for Matrix

3.2 ReLu Layer

If f(y)=0 when y<0 and f(y)=y when y > 0; this layer is also known as the activation layer since it checks the matrix in this layer.To minimize the size of the image and save on computational resources, it functions as a half rectifier. sufficient time for calculations. Figure 3.2.1 shows the output matrix for this layer.

3	22	0	10
23	25	87	139

Figure 3.2.1. Output 2D Matrix for ReLu Layer.

3.3 Pooling Layer

This layer is useful for extracting features from the image and for reducing overfitting. When a model performs flawlessly on test data but has issues with real-time data, overfitting occurs. Stride is employed in this layer, and the stride value should be lower to prevent data loss.

3.4 Flattening of Data

In this step array of data is straightened and it is given as input to next layer.

ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 7.047 ⋇ Vol. 10, Issue 7, July 2022
DOI: 10.17148/IJIREEICE.2022.10733

3.5 Fully Connected Layer

In this layer data from all the layers are combined to give the overall output. The entire dimension will be reduced and the output is passed as input to the entire layer and this procedure is followed to give the output as shown in Figure 3.5.1.After passing through several layers, the face emotion is identified.

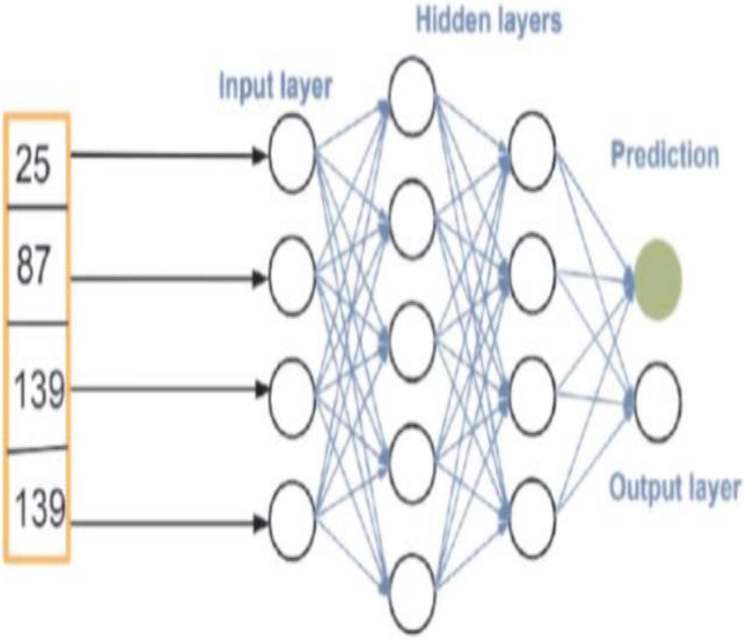

Figure 3.5.1. Fully Connected Layer

4. MUSIC PLAYER

The music player's backend code is written in the Java Script programming language. In addition to the mood-based feature, it includes three more modes for playing tunes. It allows for the addition of songs to the Songs and the queue can alternatively be chosen at random. As we understand that HTML and CSS give a beautiful look to e-mail using JavaScript, and assist us in communicate with users, which enhances their friendliness and user accessibility. It operates not just on the console, it grants the user access so they can manipulate it manually

Once the expression is being predicted music relevant to the emotion of the user is checked and played in the music player which can be seen with a player interface back to the user. There are options of random play mode and queue play mode which can be made use by the user whenever he/she wants to use the place as a normal player.The user's emotion and music player interface is depicted in the below Figure 4.1 and Figure 4.2.The is taken as the final result given back to the user.

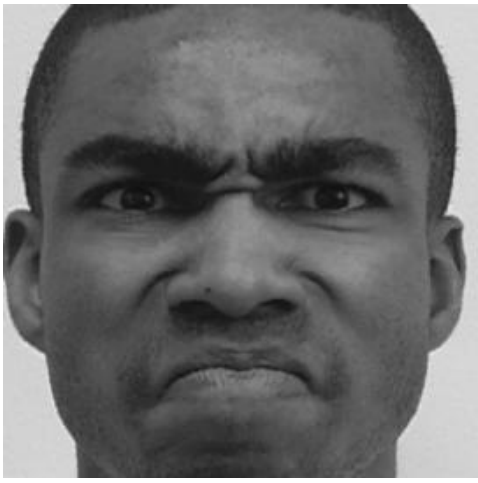

Figure 4.1 Face emotion

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁑ Impact Factor 7.047 ⁑ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10733

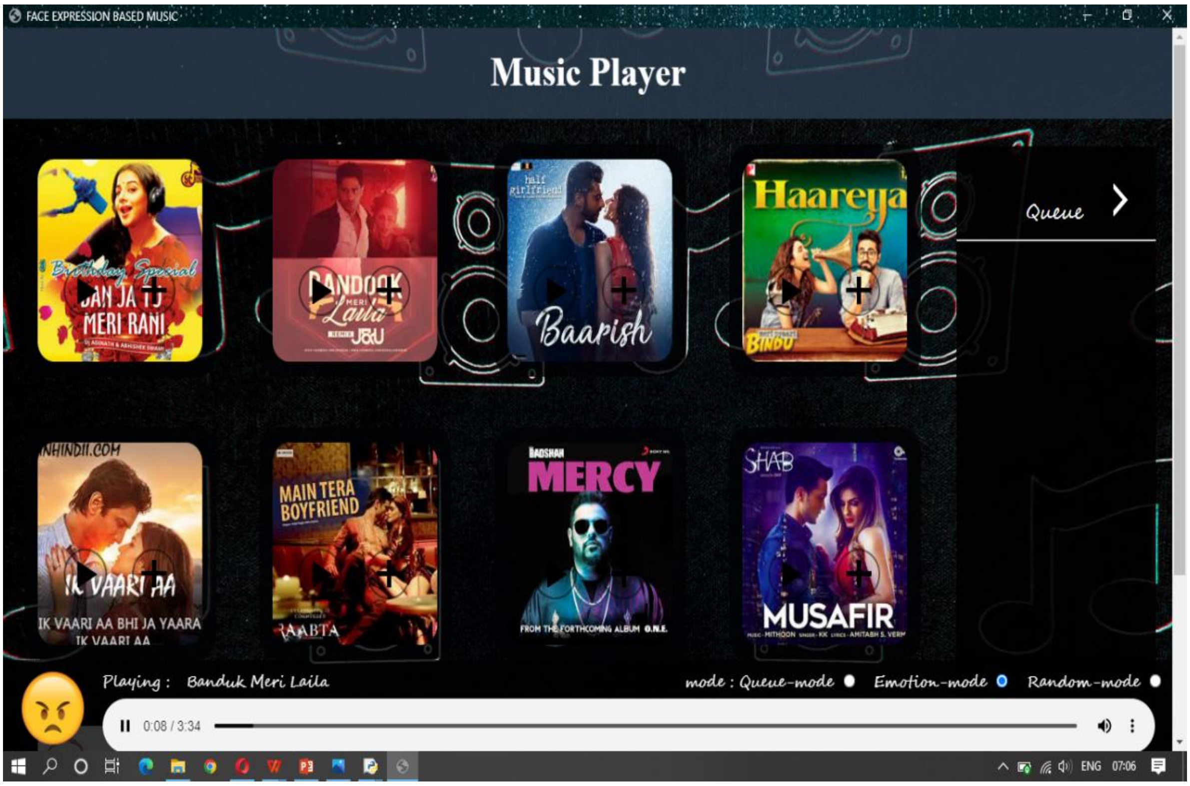

Figure 4.2 Music player interface

VI. CONCLUSION

In this project, we strive to investigate a fresh approach to categorizing music based on the feelings and expressions of people. So it was suggested to use neural networks and visual processing to categorize the four main universal emotions expressed by music: happiness, sadness, anger, and neutral. On the input image, a face detection procedure is first carried out. The feature points are then extracted using an image processing-based feature point extraction approach. Finally, a neural network is instructed to identify the emotion included in a set of values generated from processing the collected feature points. The project is still in progress, and it is anticipated to succeed in the field of emotion recognition and play the music from the supplied dataset.

Though our project has covered all the parameters to be implemented but still as per the future thinking our project can be enhanced with more features like increasing the number of expressions, the code can be still optimized, can also implement voice based volume control or gesture from the hands to manage the music that is being played.

Including the different sensors and gesture to control the volume or change to the next song, as iot is the trending technology that is making human lives easier day by day. Using the resources available to make the music player faster and smarter.

REFERENCES

[1]Quasim, Mohammad Tabrez, et al. "Emotion-based music recommendation and classification using machine learning with IoT Framework." Soft Computing 25.18 (2021): 12249-12260.

[2]Shalini, Shantha K., et al. "Facial Emotion Based Music Recommendation System using computer vision and machine learning techiniques." Turkish Journal of Computer and Mathematics Education 12.2 (2021): 912-917.

[3]Abdullah, Sharmeen M. Saleem Abdullah, et al. "Multimodal emotion recognition using deep learning." Journal of Applied Science and Technology Trends 2.02 (2021): 52-58.

[4]Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. "Deep-emotion: Facial expression recognition using attentional convolutional network." Sensors 21.9 (2021): 3046.

[5] Abdulsalam, Wisal Hashim, Rafah Shihab Alhamdani, and Mohammed Najm Abdullah. "Facial emotion recognition from videos using deep convolutional neural networks." International Journal of Machine Learning and Computing 9.1 (2019): 14-19.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.047 ⋇ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10733

[6] Mahajan, P., Khurad, P., & Chaudhari, P. (2020). Face Player: Facial Emotion Based Music Player. International Journal of Research and Analytical Reviews (IJRAR).

[7]Mageshwaran, L., et al. "Analysing Meticulous Behavior of Learners from Non-Verbal Cues." i-manager's Journal on Pattern Recognition 7.2 (2020).

[8] Talegaonkar, Isha, et al. "Real time facial expression recognition using deep learning." Proceedings of International Conference on Communication and Information Processing (ICCIP). 2019.

[9] Chowdary, M. Kalpana, Tu N. Nguyen, and D. Jude Hemanth. "Deep learning-based facial emotion recognition for human–computer interaction applications." Neural Computing and Applications (2021): 1-18.

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