OPEN ACCESS Freely available online

Autism-Open Access

Research Article

Emotion Based Music Player Using Deep Learning

K Praveen Kumar*, **Nikhila Telakuntla**

Department of Information Technology, Kakatiya Institute of Technology and Science, Warangal, India

ABSTRACT

Every single person has a different face and the expressions on those faces tell the same story and are a major indicator of how someone is feeling and acting. Music is considered to be the purest form of expression and creativity and it is known to have a stronger emotional impact on listeners. It has a special ability to elevate one's mood. Emotion detection is the process of determining a person's feelings based on various facial cues and visual information. This field has flourished as a result of how popular deep learning has become. Emotion recognition has also opened the door to a wide range of previously unthinkable applications. One of the things with a strong emotional connection is music. A listener may look for music that conveys a specific emotion when they are feeling that emotion. Using our emotion detection model, we associate these emotions with a music player that plays music that improves user experiences. The main goal of the proposed system is to create a powerful music player that makes use of facial recognition technology to access user emotion. When compared to doing it manually, the system produced by the extracted facial features will save time and effort. However, the results of research into categorising music based on emotions have not been the best. In this project, we offer an affective cross-platform music player that recommends songs based on the user's current mood. EMP provides intelligent mood based music suggestions by integrating the capabilities of emotion context reasoning within our adaptive music recommendation engine.

Keywords: CNN Layers; Face recognition; Emotion detection

INTRODUCTION

Music has a vast history; it predates languages. Both the medical industry and the study of human emotions rely heavily on music. According to the study, music is a good approach to connect with people who are resistant to talk therapy. In the medical field, a method known as music therapy has proven effective in treating patients with depression and anxiety. The usage of music emotion analysis can be advantageous for the music suggestion feature. In order to provide relevant music suggestions and improve the user experience, major online music services have developed the "music recommendation" function by studying the listening habits of a range of users [1].

Users still need to search their playlists for music that convey their emotions even when this tool effectively meets their needs. A user of a traditional music player must autonomously browse through his playlist and select songs that will improve his mood and emotional experience. This method of song selection is challenging and time-consuming and the user could struggle to find the ideal music. The CNN method is used by the emotion module to accurately determine the user's mood from an image of their face. While classifying songs into 4 different mood classes, the music classification module achieves an impressive result by utilising audio features. By matching the user's emotions to the song's mood type while taking into account their preferences, the recommendation module proposes music to the user.

The researchers suggest an emotion-based music classification and recommendation framework for accurately categorising songs by observing how people interact with one another with emotionally charged music. Procedures that quickly classify the characters based on user emotions must be established specifically for the introduction of new songs to an IoT software application. It gives a description of the many methods researchers have investigated to identify moods from human

Correspondence to: K Praveen Kumar, Department of Information Technology, Kakatiya Institute of Technology and Science, Warangal, India; E-mail: kpk.it@kitsw.ac.in

Received: 11-Mar-2024, Manuscript No. AUO-24-30099; **Editor assigned:** 13-Mar-2024, PreQC No. AUO-24-30099 (PQ); **Reviewed:** 27-Mar-2024, QC No. AUO-24-30099; **Revised:** 09-Apr-2025, Manuscript No. AUO-24-30099 (R); **Published:** 16-Apr-2025, DOI: 10.35248/2165-7890.25.15.418

Citation: Kumar KP, Telakuntla N (2025) Emotion Based Music Player Using Deep Learning. Autism-Open Access. 15:418.

emotion expressed through facial expressions. The location of a person's eyes, mouth and nose on their face can be determined, as well as the detection of facial movements, using the calculation of extracted facial structure. The results imply that by employing a multimodal method from biological signals to determine emotional states, emotion detection can be enhanced and made more accurate. We provide a deep learning strategy that can focus on significant face features and outperforms existing models on a variety of datasets, including FER-2013, CK +, FERG and JAFFE. The attentional convolutional network is the foundation of this approach. The use of a Deep Convolutional Neural Network (DCNN) in conjunction with Google TensorFlow machine learning technologies to recognise facial expressions in videos is covered in ten emotions from the Amsterdam Dynamic Facial Expression Set-Bath Intensity Variations (ADFES-BIV) dataset were analysed using the approach, which was evaluated on two datasets [2].

The SVM algorithm has been used to work on the two features of emotion recognition using neural networks: Emotion tagging and music segment selection. The dataset used in this work, AMG1608, is a publicly available dataset that makes use of an AV dimensional model and contains emotional expressions as points in the AV emotional space. Based on State Vector Machine (SVM), MER was studied. The AV model has undergone some modifications, including AV4Q, AV11C and AV4Q-UHM9, to better classify emotions. In this study, the sound has been classified into seven levels based on its pitch, intensity and the Twelve Mean Law. 50 frames are taken from the pieces to create a 72-dimensional feature vector based on this dataset. The KNN algorithm is the foundation for MER in this paper and a KNN-based classifier has also been investigated. This study attained an accuracy of 88%, with pleasant emotions being recognised at a rate of 90% and angry emotions at a rate of 75%.

MATERIALS AND METHODS

CNN algorithm: One of the deep learning approaches, CNN is a feed forward neural network with a deep structure that uses the convolution kernel to complete the convolution operation. Although it is widely used across many different industries, image recognition makes the most use of it. In most CNNs, convolutional, pooling and full connection layers are present. The convolution layer primarily extracts the local elements of the image. The main goal of the pooling layer is to create new, smaller image features while retaining the primary input images by compressing the relevant picture qualities that were extracted from the convolution layer's previous layer. The full connection layer is an effective approach to get global features since it cascades and fuses all of the local data retrieved from the convolution layer and the full connection layer. Softmax is used to produce the classification effect in the end. The CNN network frame diagram is shown in Figure 1 [3].

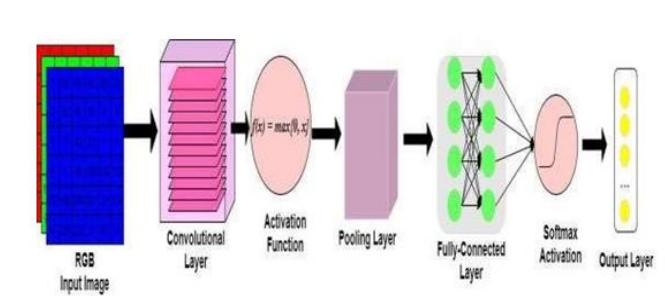

Figure 1: CNN network.

When the convolutional neural network is layered too deeply, it is easy to make the gradient disappear, which will lower performance. The CNN model utilised in this article is composed of three parts that are each connected by a full connection layer and a Softmax layer after being partitioned into a convolutional level and a max-pooling layer. First, the first convolution layer is convolutioned with ten convolution tests after 48 48 grayscale images of facial expressions are input. The output of the first pooling layer is then convolved 20 times. Second, the output of the second pooling layer is convolved with 40 convolution checks. Extend it to create a framework that is entirely integrated [4].

There are 100 neurons, a 5 5 convolution kernel size, a pooling layer that employs 2 2 maximum pooling and a softmax layer that categorises expressions. In general, the convolution layer's calculation formula is as follows:

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

The input matrix Mi, the convolution kernel value Li,j, the offset term pj, the output matrix Cj and the activation function f(x) are all used. The activation function used in this work is called RELU and it is defined as follows.

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

Here, we create a CNN network that is completely trainable. To extract global information, a convolutional kernel with a size of 5 5 is used. In contrast to CNNs with many parameters, we build a network structure that is simple to extract global information, while the training time is rapid and more suited for the experiment data.

Feature extraction using CNN: In this model, a webcam is used to capture live video using the OpenCV library. The different face emotions are captured by the web cam and gets trained. This input is given to different layers of CNN.

Convolutional layer: The image will be reduced to pixels and stored as an array in this layer. Extraction of the image's characteristics and dimensionality reduction are both beneficial. Any 2 × 2 filter will be multiplied by the array of pixels (a filter with low values will facilitate calculation). The result matrix is thus obtained following calculation and is displayed in Figure 2 [5].

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 2: Calculation for matrix.

Relu layer: This layer is also referred to as the activation layer since it examines the matrix to see if f(y)=0 when y0 and f(y)=y when y>0 respectively. It performs as a half rectifier to reduce the size of the image and conserve computational resources. Enough time to make computations. For this layer, the output matrix is shown in Figure 3.

3	22	0	10
23	25	87	139

Figure 3: Output 2D Matrix for RELU Layer.

Pooling layer: This layer helps to reduce overfitting and to extract features from the image. Overfitting happens when a model works perfectly on test data but has problems with real-time data. This layer uses stride and the stride value ought to be reduced to avoid loss of data [6].

Flattening of data: The data array is straightened at this stage and supplied as input to the following layer.

Fully connected layer: Data from all other layers are integrated in this layer to provide the final product. The entire dimension will be shrunk and the output will then be supplied as input to every layer. This process will result in the output depicted in Figures 4 and 5.

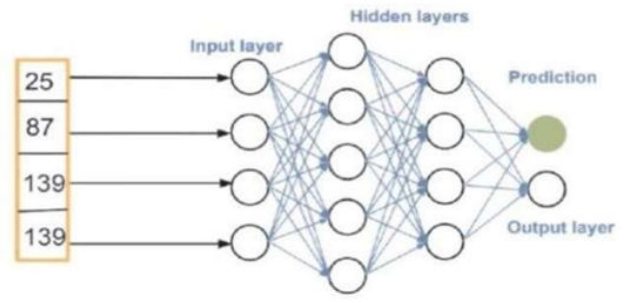

Figure 4: Fully connected layer.

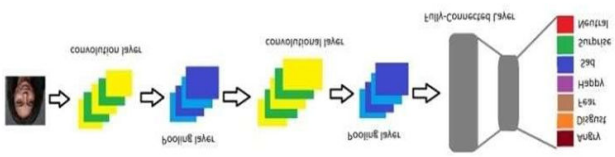

Figure 5: CNN architecture for music emotion recognition.

Music player: Java script is used to create the backend code for the music player. It has three extra options for playing music in addition to the mood-based feature. It enables the adding of songs to the songs queue, as well as the option of selecting a queue at random. We know that HTML and CSS give e-mail utilising Java script a beautiful appearance and help us communicate with users, which improves their friendliness and user accessibility. It functions outside of the console and gives users access to manually control it [7].

Once the user's expression is predicted, appropriate music is checked and played in the music player, which the user can view from a player interface. When a user wants to use the space as a regular player, there are options for random play mode and queue play mode that can be used. In above figures shows the user's feeling and the music player interface. The ultimate outcome is returned.

Implementation

Step 1: Importing the libraries. In the initial step, we will import the libraries that are necessary for the project emotion based music player (Figure 6).

```
import mediapipe as mp
import numpy as np
import cv2
from tensorflow.keras.utils import to_categorical
from keras.layers import Input, Dense
from keras.models import Model
import streamlit as st
from streamlit_webrtc import webrtc_streamer
import av
import webbrowser
```

Figure 6: Importing libraries.

Step 2: Collecting the data from the web cam (Figure 7).

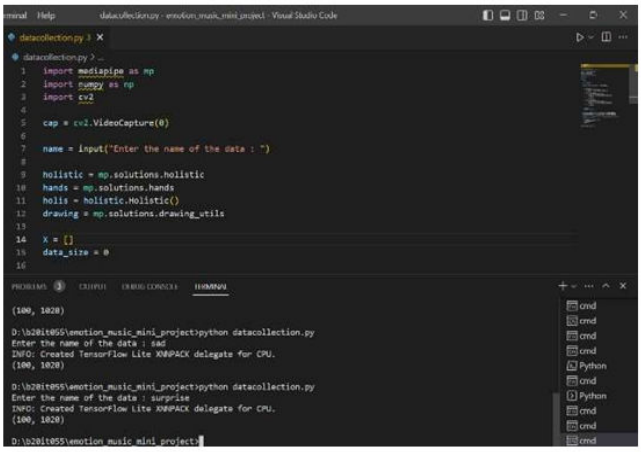

Figure 7: Collecting the data.

Step 3: Training the model. A new model is created in this step (Figure 8).

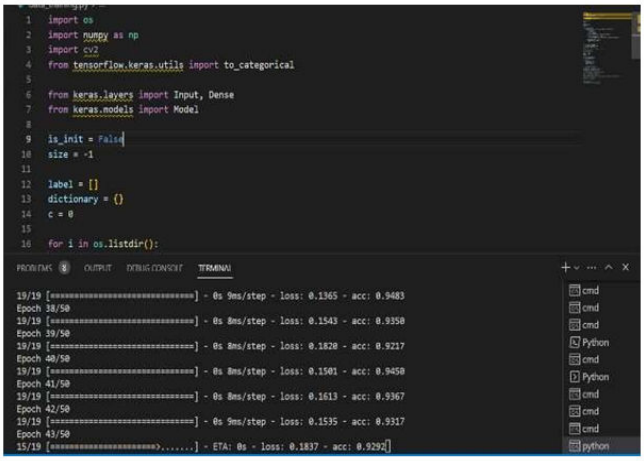

Figure 8: Training the model.

Step 4: Execute the program. When we execute the program, the web app gets opened in the browser. We need to enter the language and singer name if we want to listen any specified songs and should click the "recommend me songs" button. Then automatically web cam collects the data and this data is send to the trained model in which it detects the face emotions. According to the emotion, this project displays the recommended songs of that particular emotion in youtube. The output is shown in Figure 9 [8].

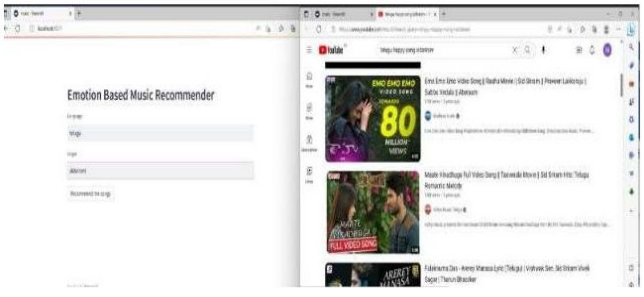

Figure 9: Output.

RESULTS AND DISCUSSION

The CNN model has been tested and trained. RELU is utilised as the activation function in the CNN model and the model was trained over 120 iterations, with iterations consisting of 448 steps each. We have presented the training and validation accuracy in Figure 10.

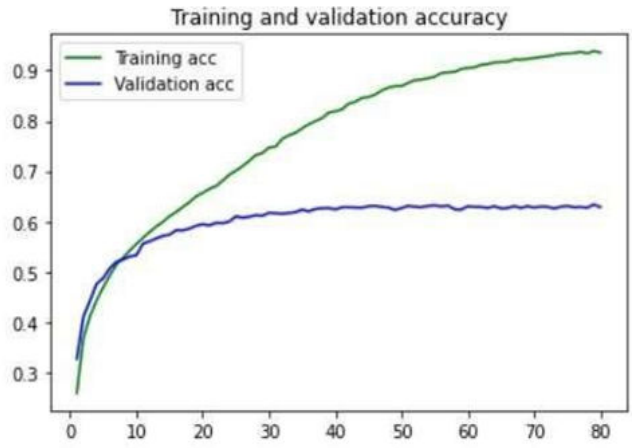

Figure 10: Training and validation accuracy.

We have achieved training accuracy of 93.58% and validation accuracy of 63.40% using the CNN model. Figure 11 shows the loss during training and validation. Since the model was trained over a lengthy period of time, it is clear that the fit is good up to 10 epochs; hence, the validation loss is bigger than the training loss [9].

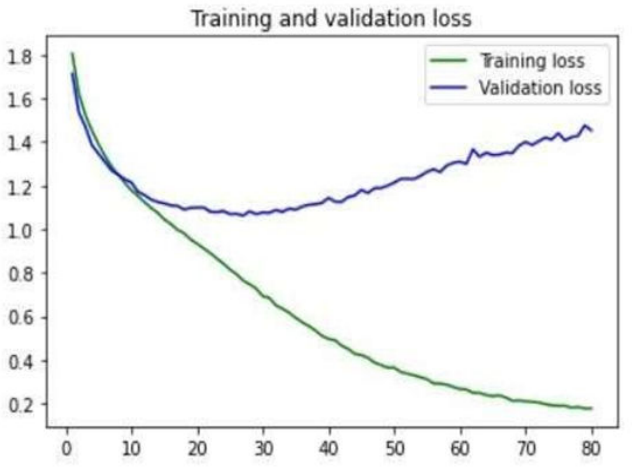

Figure 11: Training and validation loss.

CONCLUSION

The suggested method analyses photos of facial expressions, spots behaviours indicative of based emotions and then selects music indicative of these feelings. We work to investigate a novel method for classifying music based on the emotions and expressions of individuals. Thus, it was proposed to classify the four primary universal emotions expressed by music-happiness, sadness, anger and neutrality-using neural networks and visual

processing. First, a face detection process is applied to the input image. The feature points are subsequently extracted with a feature point extraction approach based on image processing. In order to determine the emotion included in a set of values produced by processing the gathered feature points, a neural network is then given instructions.

The key benefit of this technique is that it doesn't require you to choose the music by hand at all. Therefore, the created project will offer users the songs that are most appropriate for them based on their current emotions, lessening the workload of users when creating and managing playlists, making listening to music more enjoyable and allowing songs to be organised systematically in addition to helping users.

REFERENCES

1. Quasim MT, Alkhammash EH, Khan MA, Hadjouni M. Retracted article: Emotion-based music recommendation and classification using machine learning with IoT Framework. Soft Comput. 2021;25(18): 12249-12260.
2. Abdulsalam WH, Alhamdani RS, Abdullah MN. Facial emotion recognition from videos using deep convolutional neural networks. Int J Mach Learn Comput. 2019;9(1):14-19.
3. Shalini SK, Jaichandran R, Leelavathy S, Raviraghul R, Ranjitha J, Saravanakumar N. Facial emotion based music recommendation system using computer vision and machine learning techiniques. Turkish J Comput Math Educ. 2021;12(2): 912-917.
4. Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors. 2021;21(9):3046.
5. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014;1409:1556.
6. Mageshwaran L, Keerthivasan SE, Hariharan K, Rajasekaran T. Analysing meticulous behavior of learners from non-verbal cues. J Pat Recogn. 2020;7(2):7-13.
7. Abdullah SM, Ameen SY, Sadeeq MA, Zeebaree S. Multimodal emotion recognition using deep learning. J Appl Sci Technol Trend 2021;2(2):52-58.
8. Chowdary MK, Nguyen TN, Hemanth DJ. Deep learning-based facial emotion recognition for human computer interaction applications. Neural Comput Appl. 2021;35:1-18.
9. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.