ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 01, January 2023

Human Emotion Based Live Music Using Deep Learning

[1] Dr D.V. Nagarjana Devi, [2] Kadava Yamuna, [3] Pallem Kartheek, [4] Gosu Likhitha

[1] Assistant Professor, CSE Dept, IIIT, RGUKT, Nuzvid, India
[2][3][4] UG student, IIIT, RGUKT, Nuzvid, India
Email: [1] devi.duvvuri@rguktn.ac.in, [2] yamunak139@gmail.com, [3] N170247@rguktn.ac.in, [4] gosulikhitha@gmail.com

***Abstract*— It is often confusing for a person to decide which music he/she has to listen to from a massive collection of existing options. There are several music players like Wynk music, JioSaavan, Spotify, Gaana, etc. But they show us only weekly playlists or some singers' playlists. They all (Music Players) also use AI to recommend us songs that we used to like and play that zoner of songs. But sometimes we may feel irritated by listening to them as they do not match up with our emotions. The main objective is to play songs based on user facial expressions i.e, emotion which is captured from our PC webcam while we are working on it. It analyses the current emotion or mental state of the user. Sometimes our words may lie, but our facial expressions never lie (i.e, will never hide our feelings).**

Most of the users believe that at a certain point in time it is very difficult to choose and play a song from their vast music playlist. Developing this Emotion Aware Music player,helps the user to choose the song easily and helps to reduce their level of stress. In this racing world, time is very precious, and with our project user need not search for a song that suits his current mood,it will play automatically the best music according to the user's mood.

***Keywords*— Human, Emotion, Live Music, Deep Learning**

I. INTRODUCTION

Generally, people express their emotions through their facial expressions. And people like to listen to music to enlighten their feelings. So, capturing the user's emotions and playing appropriate songs according to his emotion can increase the calmness in his mind.

This project aims to capture user emotion expressed through his facial expression which is taken by the inbuilt camera available on our computer. It will segregate the emotion from this captured picture into a few categories like Angry, Happy, Sad, Fear, Neutral, Disgust, and Surprise and plays music based on the categorized emotion.

2.1 Purpose

The project aims to lighten the mood of the user, by playing songs that cope with the user's emotions. There is a saying that music can heal wounds that medicine cannot touch. Music pleases us and releases our suffering (mental and sometimes physical suffering also). It helps to overcome depression and sadness. Many health risks can be avoided through music and there are chances to bring the mood of the user to a better stage.

2.2 Overview

This is a Python-based application that helps the user to play a song based on the emotions that he currently carries. Nowadays listening to music while working became a habit for all. So, to listen to music, he/she needs to search for them, but our project helps users to listen to music without searching. It will play automatically based on user emotion.

2.3 Scope

The scope of this application is wide as it can be used by anyone, anywhere, and anytime. This application helps the user to change his/her mood to a joyful one as music helps people to change their mood. This application also acts as a stress booster which boosts emotions to a great extent.

II. ANALYSIS OF EXISTING AND PROPOSED SYSTEMS

2.1 Problem Statement

Music is one of the most effective media as it can install deep feelings and swamp listeners with subliminal messages. It deftly plays with our emotions which in turn play at mood. It can aid us when we are feeling low and empower us. To enhance the mood of the user by designing a system that detects the emotion and plays the music accordingly. The main aim of this project is to play a song based on the emotion of the user who uses the application.

2.2 Existed System

The existing system will play a song based on the user's request that he/she requests the player play this song or that song. He manually requests the song. Its aves the user data that he/she likes which they prefer more and suggest that type of song all the time. And it plays songs that relate to the singer which he used to listen to most of the time. Or it plays a song based on recent hits, weekly, and monthly hits.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 01, January 2023

2.2.1 Disadvantages of Existed System

- Manual

Manually searching the desired song from plenty of playlists.

- Cached & History Based Cached-based songs and typing history will lead to suggestions for some sort of songs, but it's not based on the current emotion of the user.
- Random Pickups

Random Weekly/Monthly hits suggestions from music application

Ex: Spotify, Wynk, etc.

III. PROPOSED SYSTEM

Considering the issues in the existing system our proposed system benefits us to present interaction between the user and the music player. The purpose of the system is to capture the face properly with the camera. Captured images are fed into the Convolutional Neural Network which predicts the emotion. Then the emotion derived from the captured image is used to get the playlist of songs using an audio tool. The main aim of our proposed system is to provide a music playlist automatically according to the user's moods, which can be happy, sad, angry, fearful, disgusted, or surprised. Wechose the Jupyter platform which solves all the issues of the previous system. So, it can avoid manual systems.

3.1 Advantages of the Proposed System

- Automatic Recommendation of the music
- Accuracy around 90%
- Easy to use Accessible from any where

IV. MODEL

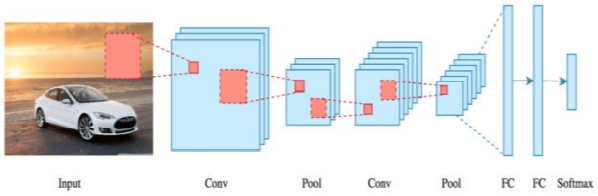
In our model, we are using deep learning CNN Algorithm and sequential model. As we know CNN algorithm works better for image-related training and testing. We are using the CNN algorithm rather than other algorithms.

4.1 Deep Learning

Deep Learning is a kind of machine learning which includes many neurons for dealing with a big amount of data with good accuracy. Deep learning is an element of data science, which includes statistics and predictive modeling. It helps the data scientist who has to work with a large amount of data. It makes their task easier and faster. Machine learning algorithms are linear whereas deep learning algorithms are stacked in a hierarchy of increasing complexity and abstraction. Therefore, deep learning works better than traditional machine learning when working with complex feature inputs and a large amount of data.

4.2 CNN Algorithm

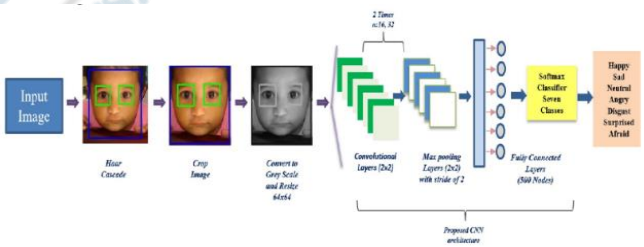
In deep learning, a convolutional neural network (CNN/ConvNet) is a class of deep neural networks, mostly applied to analyze visual images. It works best when we are working with images rather than other algorithms. CNN uses a special technique called convolution instead of matrix multiplications like the other neural networks.



Convolutional neural networks consist of one input layer, one output layer, and many hidden layers. These layers are composed of many artificial neurons. Artificial neurons are mathematical functions that calculate the weighted sum of multiple inputs and output an activation value. When we give input to ConvNet, each layer generates several activation functions that are passed onto the next layer to activate the neurons in that layer.

The first layer usually extracts basic features from the input image such as horizontal or diagonal edges. The output of the previous layer moves on to the next layer which determines more complex features of that image such as corners or combination edges. As it moves deeper into the network it can identify even more complex features from it such as the faces of people.

Based on the activation map of the final convolution layer, the classification layer outputs a set of confidence scores (probability values between 0 and 1) that specify the image likely to belong to a “class”.



4.3 Sequential Algorithm

A sequential algorithm executes sequentially from starting to ending one layer at a time. In the CNN algorithm, a sequential model means, it is an arrangement of a linear stack of layers of ConvNet.

4.4 Layers and Activation functions

Our model consists of several layers and to activate neurons in those layers we have used a few activation functions. The layers are the Convolution layer, Max Pooling, Dropout, Flatten, and Dense.

Activation functions used are Relu and Softmax.

While compiling we have used adam optimizer, categorical cross-entropy loss.

In the convolution layer if our input image is of size N × N and the kernel ω is of size m × m then our output of the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 01, January 2023

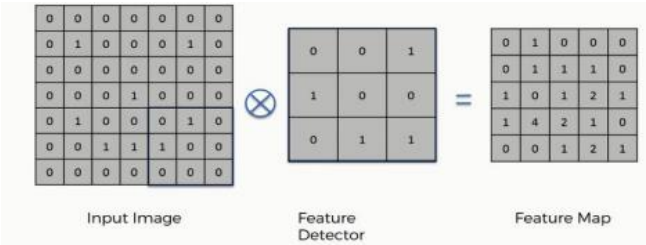
convolution layer is (N-m+1) × (N-m+1). The mathematical function for this layer is

$x\ell ij = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab}\, y_{(i+a)(j+b)}^{\ell-1}$ [8]

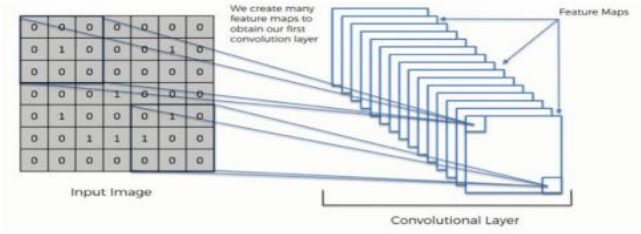
conv2(x, ω,'valid')

Then, the convolutional layer applies its nonlinearity:

$y_{ij}^{\ell} = \sigma(X_{ij}^{\ell})$ [8]



Feature Map generation through convolution operation [9]



Creation of Convolution Layer [9]

In jupyter notebook,

classifier.add(Conv2D(32, (3, 3), input_shape = (128, 128, 3), activation = 'relu'))

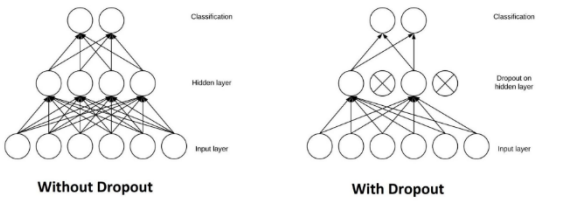
(3,3) is the kernel size and the input shape of the image is given as (128,128,3), and to activate the neurons we are using the relu activation function.

Max pooling is a pooling operation in which it selects the maximum elements from the region of the feature map covered by a filter.

Maximum Pooled Feature Map [9]

Dropout determines which neuron should be considered for further layer.It randomly deactivates some neurons. So, it reduces the complexity.

q=1-p



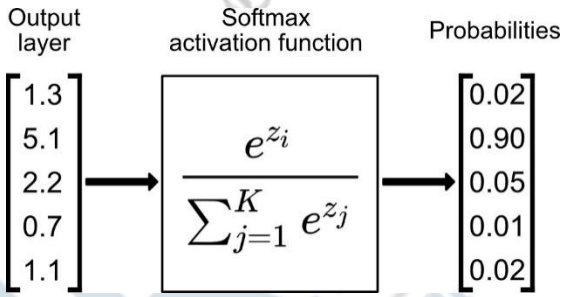
Flatten layer is used to convert the 2D into 1D for inputting into the next layer.

The dense layer is used in the final stage of the neural network to classify the image.

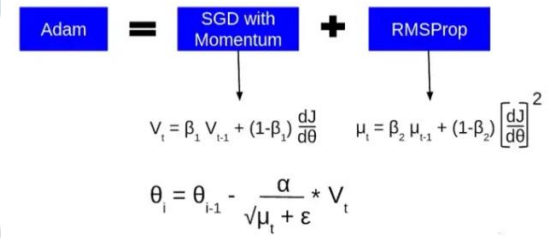
Relu is simple to compute it outputs the input directly if it is positive and zero if it is negative.

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

Softmax is used to convert the raw outputs into probability values.



Adam optimizer is used for optimizing gradient descent. Adam optimizer is a combination of RMSProp and SGD with Momentum optimizers.



Adam optimizer function [10]

V. MAIN FUNCTIONALITY

Emotion Aware Music is a Python-based application, which helps the user (player) to play a song based on his/ her current emotion and helps the user to get more focussed on work. Emotion Aware Music acts as a medium to make users more focused on work by playing some music based on their emotions. The application does not tell the user that it was capturing the image. This makes the application capture the original natural facial expression so that better accuracy can be achieved.

5.1 Image Capturing

To utilize this application, the user has to give access to the webcam. Then the image is captured from the webcam without the user's intimation to record the natural emotion. Based on the image captured, the emotion is detected.

5.2 Image Classification

After successfully capturing the image, the emotion is classified into 6 different types: Happy, sad, anger, surprise, fear, and disgust. The image is passed on to the neural

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 01, January 2023

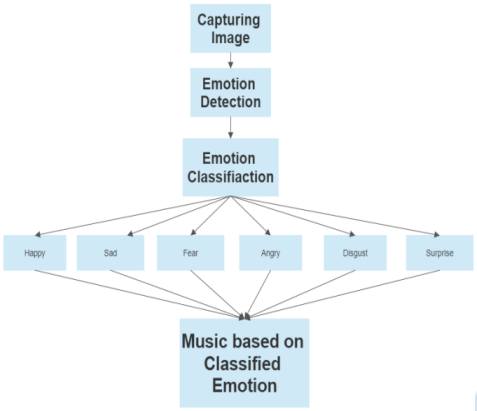
networks which classify it into one of these six emotions and the music is played based on the classified emotion.

5.3 Music Based on Classified Emotion

The classified emotion is recommended to the user and the song related to the emotion will be played from the music library.

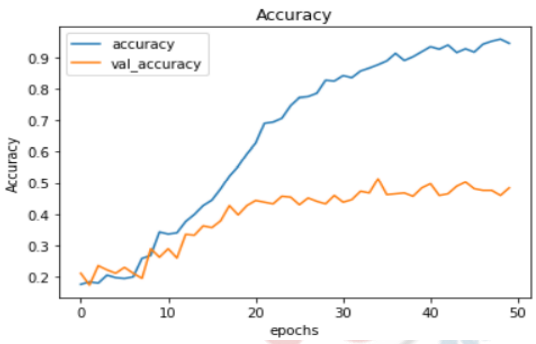
VI. ARCHITECTURE

Architecture diagram

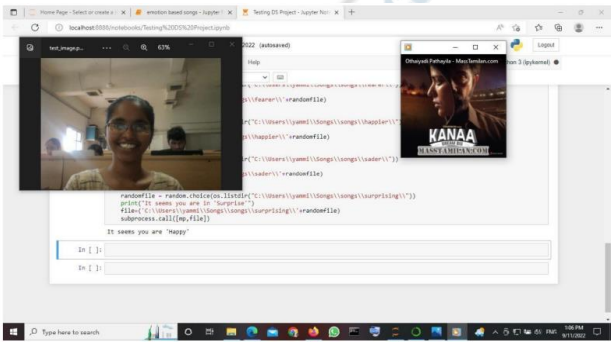


VII. ACCURACY

We have used Adam optimizer and categorical cross entropy as loss. We have got an overall accuracy of 94% on the training set and val_accuracy of 49% on the validation set.



VIII. OUTPUT



IX. CONCLUSION AND FUTURE SCOPE

Our application is a one-stop solution to all the issues and disadvantages of the existing system of various music players and music websites which don't use user emotion and use a search bar to play songs. This saves a lot of time and energy for the users. This enables the user to listen to the music without searching for it. This is just an application of the idea based on the disadvantages of the existing music player system. There are a lot of scopes to improve the application by training different users.

REFERENCES

[1] https://ikee.lib.auth.gr/record/329752/files/GRI-2021-30311.pdf
[2] https://ieeexplore.ieee.org/document/9189600
[3] Facial Emotion Detection Using CNN - Analytics Vidhya
[4] Frontiers | A Novel Music Emotion Recognition Model Using Neural Network Technology (frontiersin.org)
[5] Hands-On Computer Vision with TensorFlow 2: Leverage deep learning to create powerful image processing apps with TensorFlow 2.0 and Keras
[6] Hands-On Neural Networks: Learn how to build and train your first neural network model using Python
[7] Adil_Mohammed_Adnan_Meng_2021.pdf (uvic.ca)
[8] https://andrew.gibiansky.com/blog/machine-learning/convolutional-neural-networks/#:~:text=Convolutional%20Layers,-Suppose%20that%20we&text=In%20order%20to%20compute%20the,)(j%2Bb).&text=Then%2C%20the%20convolutional%20layer%20applies,(x%E2%84%93ij).
[9] https://www.irjmets.com/uploadedfiles/paper//issue_6_june_2022/27381/final/fin_irjmets1656670953.pdf
[10] https://medium.com/analytics-vidhya/optimizer-loss-functions-in-neural-network-2520c244cc22