Indian Institute of Information Technology Pune

Department of Computer Science Engineering

"Musical Playlist Generation using Facial Expression Recognition"

Presented by

U. Vinod Kumar

INDEX

INTRODUCTION

- Music plays an important role in our life. It's not just a source of entertainment in our life. It gives us relief and reduces our stress thus music also imparts a therapeutic approach. It helps to improve our mental health.
- Computer vision is a field of study which encompasses on how computer see and understand digital images and videos. Computer vision involves seeing or sensing a visual stimulus, make sense of what it has seen and also extract complex information that could be used for other machine learning activities.
- We will implement our use case using the Haar Cascade classifier. Haar Cascade classifier is an effective object detection approach which was proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. T

INTRODUCTION

- This project recognizes the facial expressions of user and play songs according to emotion. Facial expressions are best way of expressing mood of a person. The facial expressions are captured using a webcam and face detection is done by using Haar cascade classifier
- The captured image is input to CNN which learn features and these features are analyzed to determine the current emotion of user then the music will be played according to the emotion. In this project, five emotions are considered for classification which includes happy, sad, anger, surprise, frustration, fear, neutral.
- This project consists of 4 modules-face detection, feature extraction, emotion detection, songs classification. Face detection is done by Haar cascade classifier, feature extraction and emotion detection are done by CNN. Finally, the songs are played according to the emotion recognized.

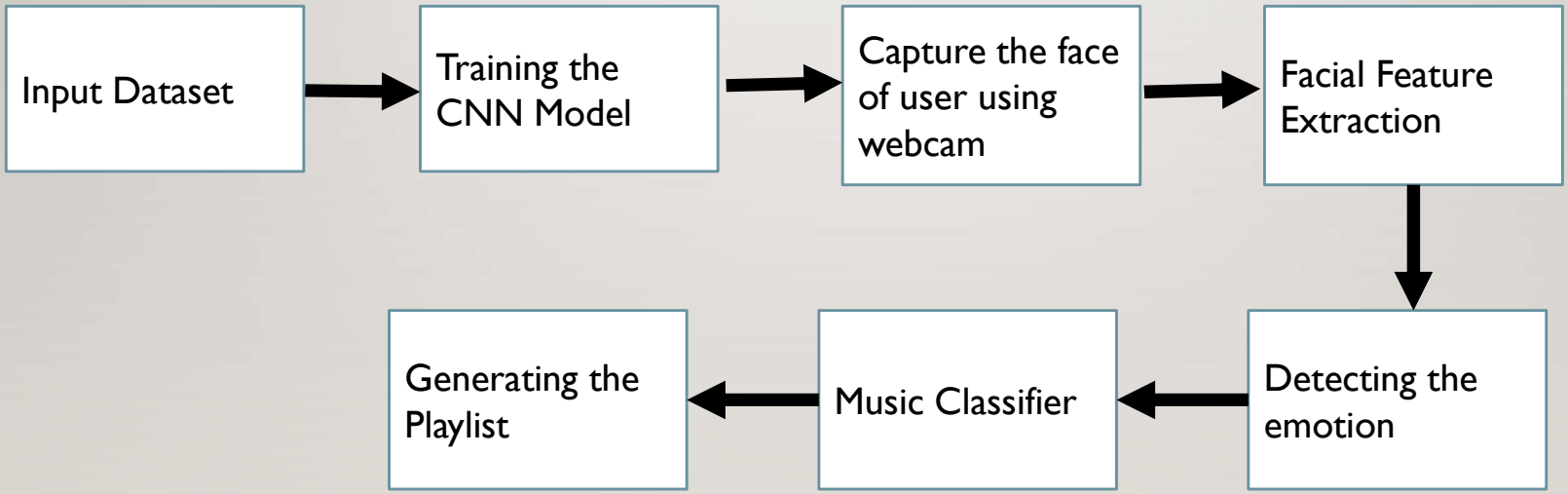
PROBLEM STATEMENT

"Develop a Musical system which will generate a musical playlist for the user based on his/her mood, by capturing his/her facial expressions."

DATASET DETAILS

- Fer-2013 dataset was prepared by Pierre-Luc Carrier and Aaron Courville, as part of an ongoing research project. They have graciously provided the workshop organizers with a preliminary version of their dataset to use for this contest.
- The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression in to one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).
- The train.csv contains two columns, "emotion" and "pixels". The "emotion" column contains a numeric code ranging from 0 to 6, inclusive, for the emotion that is present in the image. The "pixels" column contains a string surrounded in quotes for each image. The contents of this string a space-separated pixel values in row major order. test.csv contains only the "pixels" column and your task is to predict the emotion column.
- This dataset consists of 35,887 grayscale images. The training set consists of 28,709 examples. The public test set consists of 3,589 examples.

METHODOLOGY



METHODOLOGY

Step I : Proposed System

Convolution neural network algorithm is a multilayer perceptron that is the special design for the identification of two-dimensional image information. It has four layers: an input layer, a convolution layer, a sample layer, and an output layer.

Step II : Face Detection

The Viola-Jones Algorithm, developed in 2001 by Paul Viola and Michael Jones, the Viola-Jones algorithm is an object-recognition framework that allows the detection of image features in real-time.

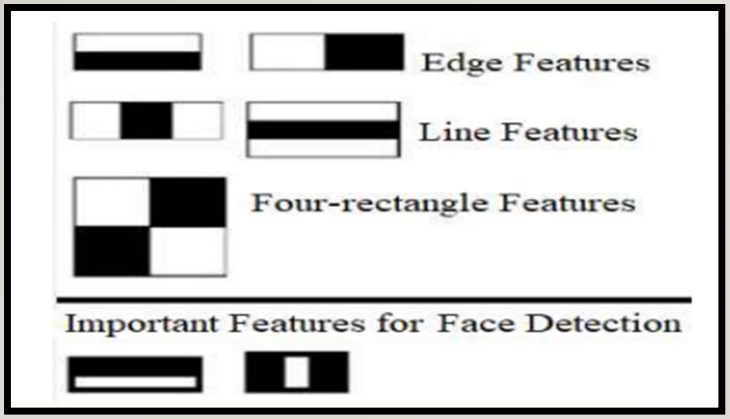
The Viola-Jones Object Detection Framework combines the concepts of Haar-like Features, Integral Images, the AdaBoost Algorithm, and the Cascade Classifier to create a system for object detection that is fast and accurate.

1.Haar-like Features :

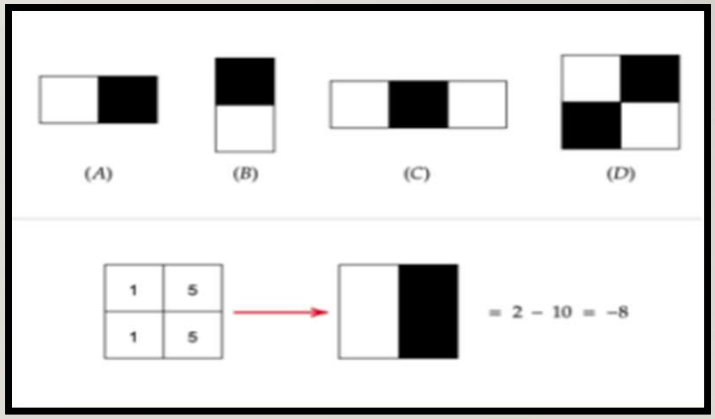
Haar-like features are named after Alfred Haar, a Hungarian mathematician in the 19th century who developed the concept of Haar wavelets (kind of like the ancestor of haar-like features). The features below show a box with a light side and a dark side, which is how the machine determines what the feature is. Sometimes one side will be lighter than the other, as in an edge of an eyebrow. Sometimes the middle portion may be shinier than the surrounding boxes, which can be interpreted as a nose.

There are 3 types of Haar-like features that Viola and Jones identified in their research:

- Edge features
- Line-features
- Four-sided features



Dig. Haar-like Features



Dig. Feature Value Calculation

2.Integral Image :

We calculated the value of a feature. In reality, these calculations can be very intensive since the number of pixels would be much greater within a large feature. The integral image plays its part in allowing us to perform these intensive calculations quickly so we can understand whether a feature of a number of features fit the criteria. To calculate the value of a single box in the integral image, we take the sum of all the boxes to its left.

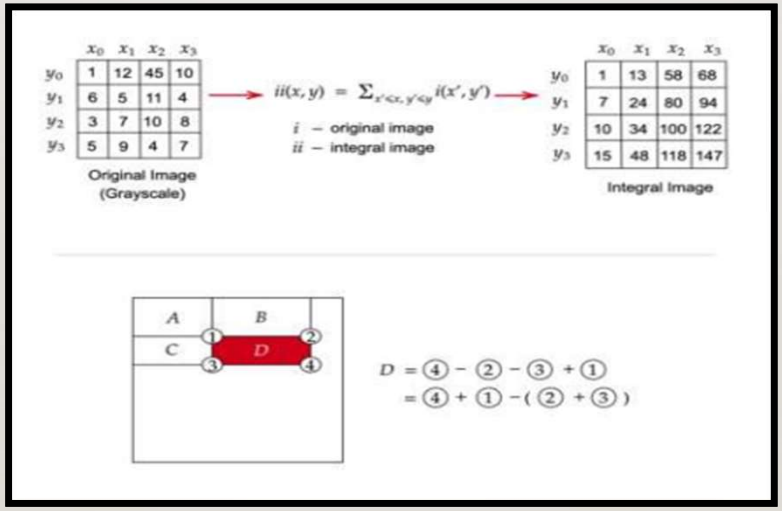
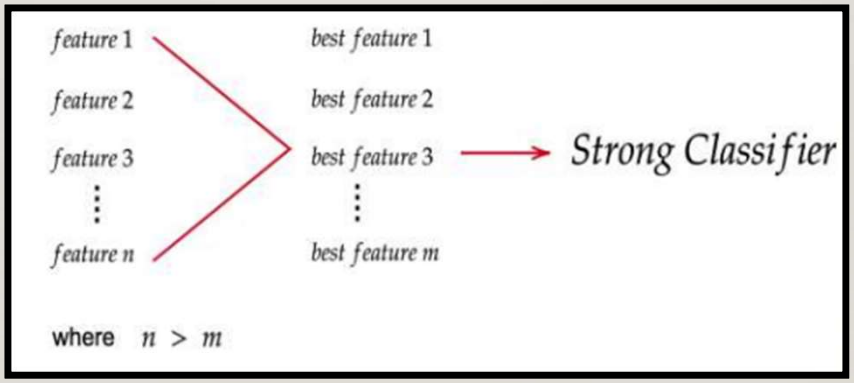


Fig. Integral Image

Haar-like features are actually rectangular, and the integral image process allows us to find a feature within an image very easily as we already know the sum value of a particular square and to find the difference between two rectangles in the regular image, we just need to subtract two squares in the integral image. So even if you had 1000 x 1000 pixels in your grid, the integral image method makes the calculations much less intensive and can save a lot of time for any facial detection model.

3.Adaptive Boosting :

- The AdaBoost (Adaptive Boosting) Algorithm is a machine learning algorithm for selecting the best subset of features among all available features. The output of the algorithm is a classifier (Prediction Function, Hypothesis Function) called a "Strong Classifier". A Strong Classifier is made up of a linear combination of "Weak Classifiers" (best features).
- The algorithm learns from the images we supply it and is able to determine the false positives and true negatives in the data, allowing it to be more accurate. We would get a highly accurate model once we have looked at all possible positions and combinations of those features. Training can be super extensive because of all the different possibilities and combinations you would have to check for every single frame or image.
- Let's say we have an equation for our features that determines the success rate (as seen in the image), with f1, f2 and f3 as the features and a1, a2, a3 as the respective weights of the features. Each of the features is known as a weak classifier. The left side of the equation F(x) is called a strong classifier. Since one weak classifier may not be as good, we get a strong classifier when we have a combination of two or three weak classifiers. As you keep adding, it gets stronger and stronger. This is called an ensemble.



4.Cascade Classifier :

A Cascade Classifier is a multi-stage classifier that can perform detection quickly and accurately. Each stage consists of a strong classifier produced by the AdaBoost Algorithm.

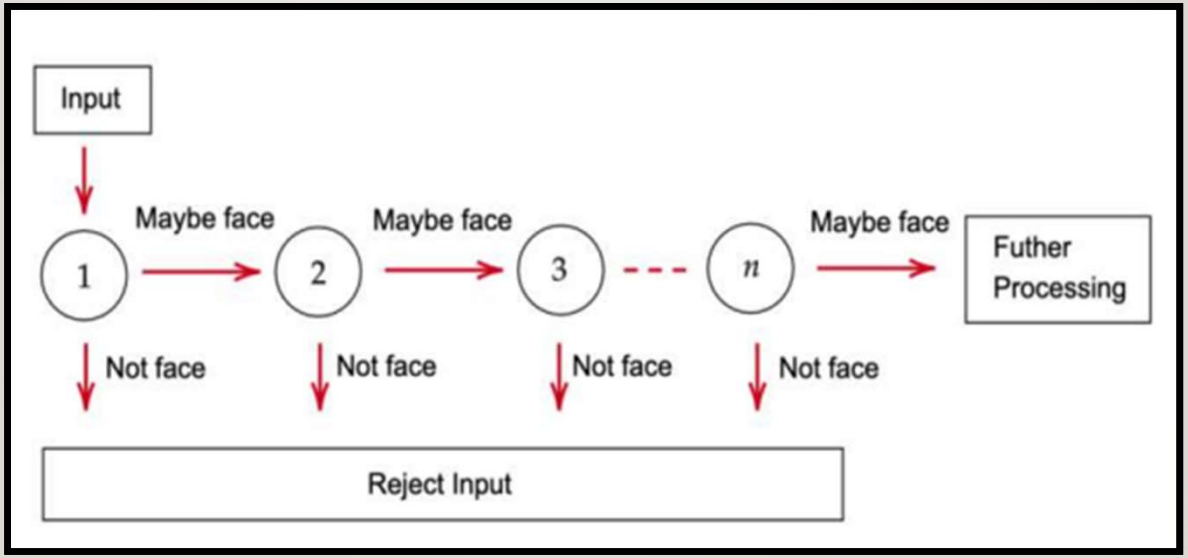


Fig. : Cascade Classifier

METHODOLOGY

Step III : Facial Feature Extraction

Convolution neural network (CNN) is an efficient recognition algorithm which is widely used in pattern recognition and image processing. It has many features such as simple structure, less training parameters and adaptability.

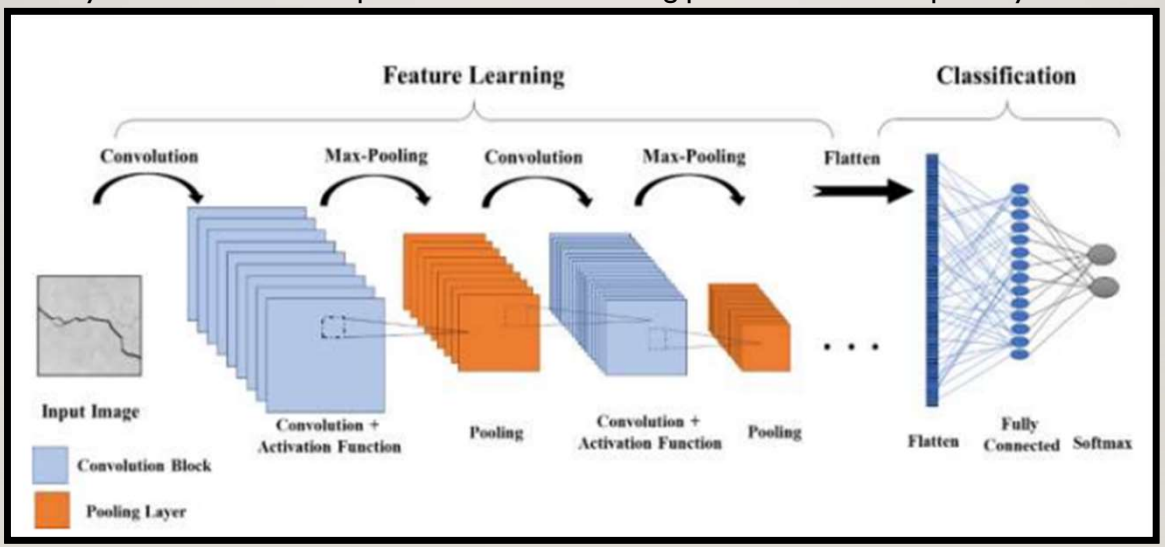


Fig. : CNN Architecture

A CNN typically has three layers: a convolutional layer, a pooling layer, and a fully connected layer.

1.Convolutional Layer :

- The element involved in carrying out the convolution operation in the first part of a Convolutional Layer is called the Kernel/Filter. The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. Conventionally, the first Convolution Layer is responsible for capturing the Low-Level features such as edges, color, gradient orientation, etc. With added layers, the architecture adapts to the High-Level features as well.
- Convolution is a mathematical operation to merge two sets of information. In our case the convolution is applied on the input data using a convolution filter to produce a feature map.
- We perform the convolution operation by sliding this filter over the input. At every location, we do element-wise matrix multiplication and sum the result. This sum goes into the feature map. The green area where the convolution operation takes place is called the receptive field. Due to the size of the filter the receptive field is also 3x3.

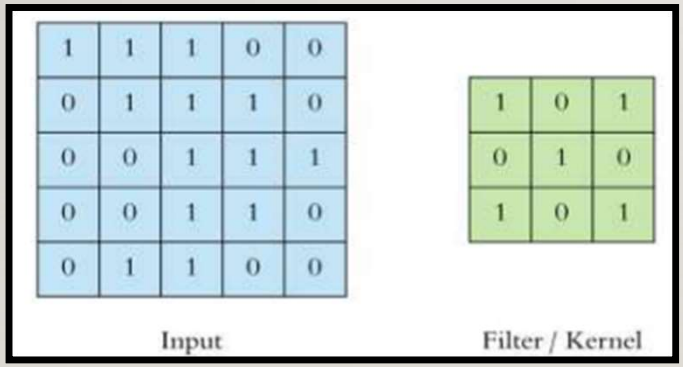


Fig. : Input and Filter

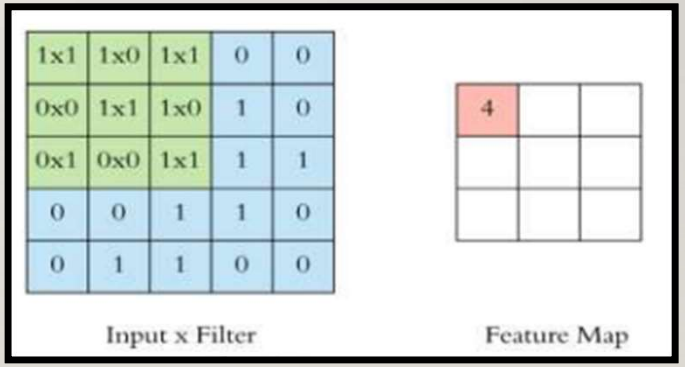


Fig. : Feature Map

2.Pooling Layer :

- Similar to the Convolutional Layer, the Pooling layer is responsible for reducing the spatial size of the Convolved Feature. This is to decrease the computational power required to process the data through dimensionality reduction.
- Furthermore, it is useful for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training of the model.
- There are two types of Pooling: Max Pooling and Average Pooling. Max Pooling returns the maximum value from the portion of the image covered by the Kernel. On the other hand, Average Pooling returns the average of all the values from the portion of the image covered by the Kernel.

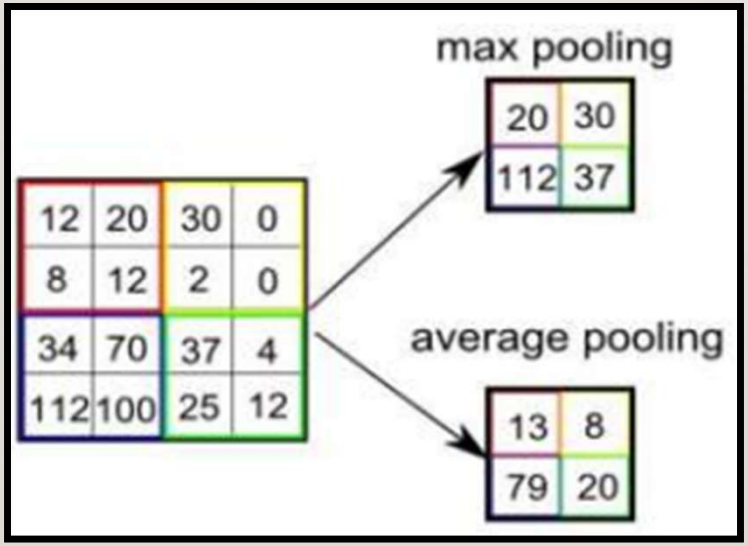


Fig. : Pooling

3.Fully Connected Layer :

- Neurons in this layer have full connectivity with all neurons in the preceding and succeeding layer as seen in regular FCNN.
- Fully Connected Layer is also called as Dense Layer. It provides learning features from all the combinations of the features of the previous layer. The FC layer helps to map the representation between the input and the output.
- The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training.
- Over a series of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the SoftMax Classification technique.

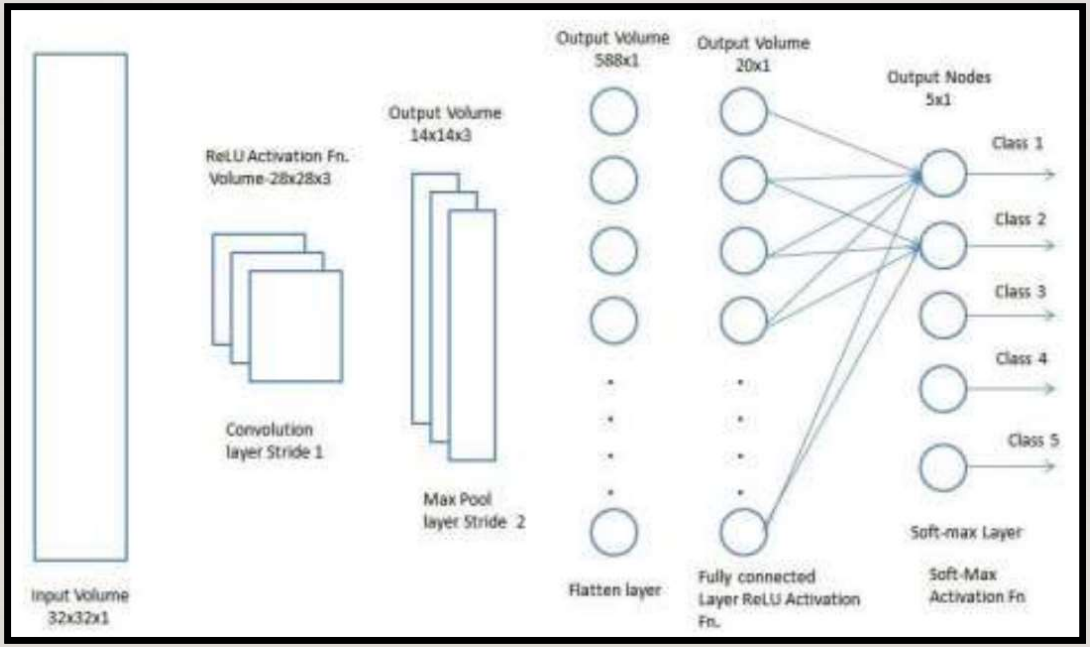


Fig. : Fully Connected Layer

RESULT

Fig. Neutral Emotion detected and Accordingly playlist is generated

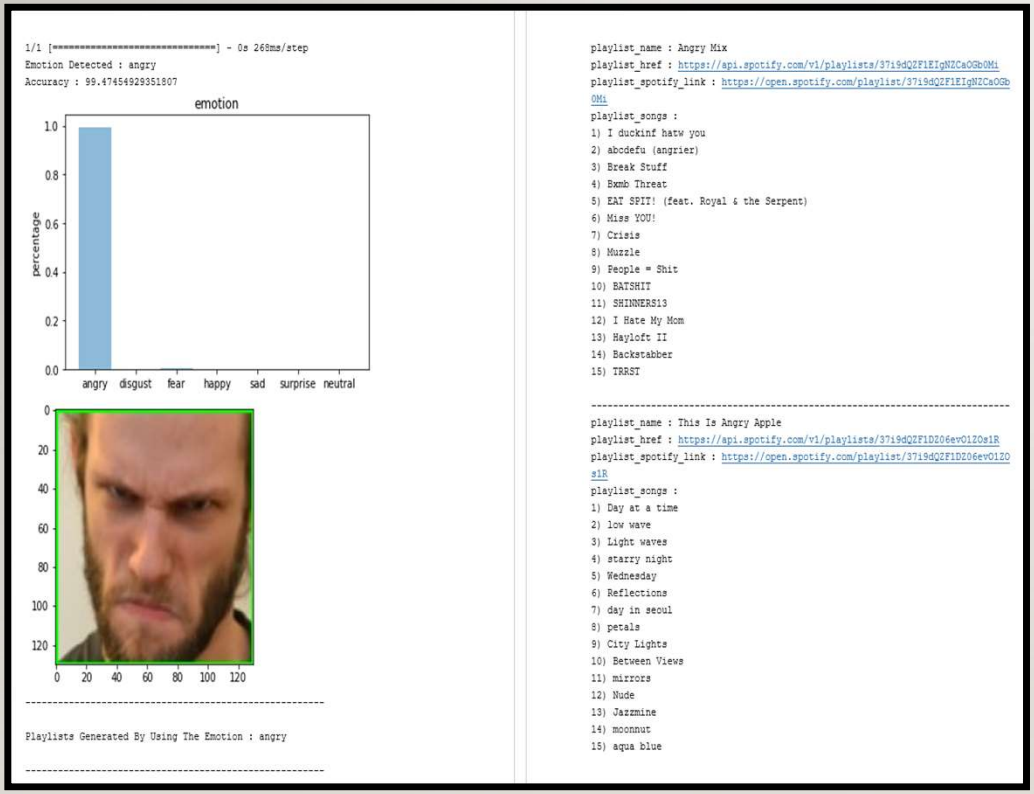


Fig. Angry Emotion detected and Accordingly playlist is generated

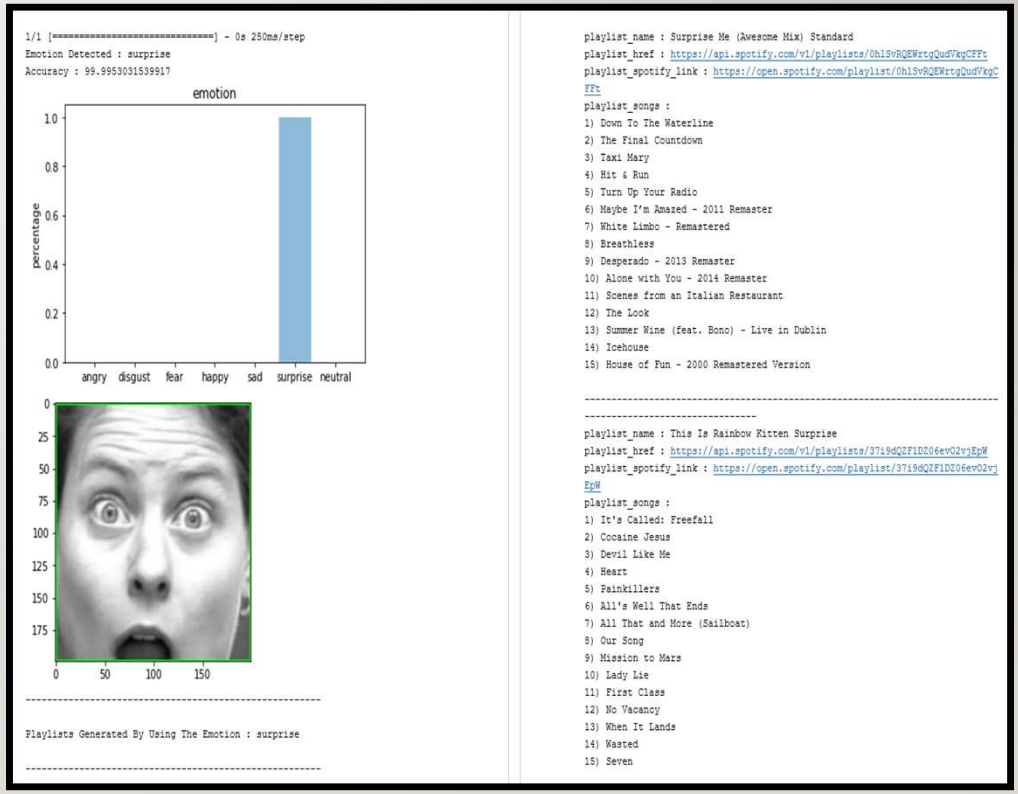


Fig. Surprise Emotion detected and Accordingly playlist is generated

CONCLUSION

1. A thorough review of the literature tells that there are many approaches to implement Music recommendation system using facial expressions.
2. Implementation of this project will help the user to automatically generate musical playlist using his facial expressions, saving his time and manual labour.
3. In this project, we are generating the playlist according to the emotion of the user, we developed a program for predicting the emotion of the user using Convolution neural networks and for generating the playlist we have used Spotify API.
4. We have applied it on various images and achieved a very high accuracy of more than 99% for happy, anger, surprise and fear emotions.

FUTURE WORK

1. To improve accuracy for sad and disgust emotions.
2. To create a web or an android application.
3. To tackle constraints like web camera resolution, lighting conditions, and other such issues that hinder the emotion recognition.
4. To observe more human emotions and incorporate them into our project.

REFERENCES

- R. S. Deshmukh, V. Jagtap and S. Paygude, "Facial emotion recognition system through machine learning approach," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 272-277.
- Athavle, M., Mudale, D., Shrivastav, U. and Gupta, M., 2021. Music Recommendation Based on Face Emotion Recognition. *Journal of Informatics Electrical and Electronics Engineering (JIEEE)*, *2*(2), pp.1-11.
- Rashma, T.V., Kannur, K. and Cheemeni, K., Smart Music Player Based on Emotion Recognition from Facial Expression.
- James, R., Sigafoos, J., Green, V.A. *et al.* Music Therapy for Individuals with Autism Spectrum Disorder: a Systematic Review. *Rev J Autism Dev Disord* **2**, 39–54 (2015).
- Florence, S.M. and Uma, M., 2020, August. Emotional detection and music recommendation system based on user facial expression. In *IOP Conference Series: Materials Science and Engineering* (Vol. 912, No. 6, p. 062007). IOP Publishing.
- Agrahari, P., Tanwar, A.S., Das, B. and Kunekar, P., 2020. Musical Therapy using Facial Expressions.
- EMOTION BASED MUSIC RECOMMENDATION SYSTEM M. Keerthana, M. Shruthi, S. Aravind Kumar
- Sarda, Pranav & Halasawade, Sushmita & Padmawar, Anuja & Aghav, Jagannath. (2019). Emousic: Emotion and Activity-Based Music Player Using Machine Learning. 10.1007/978-981-13-6861-5_16.
- Emotion based Music Player Shital Rewanwar , Siddhika Pawar , Swati Bade , Sujata Oak
- Li Y, Li X, Lou Z, Chen C. Long Short-Term Memory-Based Music Analysis System for Music Therapy. *Front Psychol*. 2022;13:928048. Published 2022 Jun 14. doi:10.3389/fpsyg.2022.928048
- Pandeya YR, Bhattarai B, Lee J. Deep-Learning-Based Multimodal Emotion Classification for Music Videos. *Sensors*. 2021; 21(14):4927. https://doi.org/10.3390/s21144927

THANK YOU