# EMOTION PERSONALIZED MUSIC AND MOVIE RECOMMENDATION USING CONVOLUTIONAL NUERAL NETWORK

Aishwarya N[1], Aishwarya RB[2], Anuka Sushma[3], Aryan Kaul[4], Chaitra P C[5]
[1] *Student, Computer Science and Engineering, DSATM, Karnataka, India*
[2] *Student, Computer Science and Engineering, DSATM, Karnataka, India*
[3] *Student, Computer Science and Engineering, DSATM, Karnataka, India*
[4] *Student, Computer Science and Engineering, DSATM, J&K, India*
[5] *Professor, Computer Science and Engineering, DSATM, J&K, India*

## ABSTRACT

*In the day-to-day stressful environment of the IT Industry, proper relaxation time is essential for all working professionals. To keep people stress-free, various technical and non-technical stress-relieving methods are currently being used.[1] Administrators, programmers, and other computer users can be classified as administrators, programmers, and other computer users, each of whom requires a unique method of relaxation. A person's emotions can express workplace stress and vexation of any kind. The key to determining a person's current psychology is to observe their facial expressions. In this paper, we discuss a user-friendly smart music player. This player will capture a person's computer-related facial expressions and detect their current mood.*

*Thanks to this music player, working professionals will be able to stay comfortable despite their heavy workloads. In recent years, several Internet companies have experimented with sentiment analysis to recommend content based on human emotions conveyed through informal language posted on social media. Sentiment analysis measures, on the other hand, simply classify a sentence's intensity as positive, neutral, or negative, and do not detect sentiment fluctuations based on the user's profile. User's attitudes are derived from social media sentences, and the music recommendation engine is run on mobile devices using a simple framework that suggests songs based on the current user's sentiment intensity. Furthermore, the framework was designed with usability ergonomics in mind. The dataset is pre-loaded with both music and movies that can be changed based on the needs of the user, as many artists release songs or movies on a daily basis.*

**Keyword:** - *emotion, face, interaction, mood, extraction, recommendation, happy, sad, angry, CNN, facial landmark, histogram.*

## 1. Introduction

Humans have an unintentional tendency to express their emotions through their faces. We can provide user-music system interaction using the proposed approach. This project focuses primarily on the user's favorite music, which is recommended based on emotional awareness. At the outset, we presented two options for the proposed system, each with its own set of features. As an example, we've provided a list of music, movies, and emotions based on spatial identification. When the application is launched, it captures photos using the webcam or any other physiological device. Our primary goal with this system is to develop a sophisticated music player capable of improving the user's mood, and music is one of the best ways to do so.

1. Emotion extraction module

2. Music Classification module

3. Music Recommendation module

After capturing the photo from the webcam, the facial features are extracted by which the emotion is classified into its categories. Human emotion is mainly classified into 7 categories mainly anger, disgust, fear, joy, neutral, sadness and surprise. After classifying the emotion to its category, respective movie and songs is recommended by the system to the user

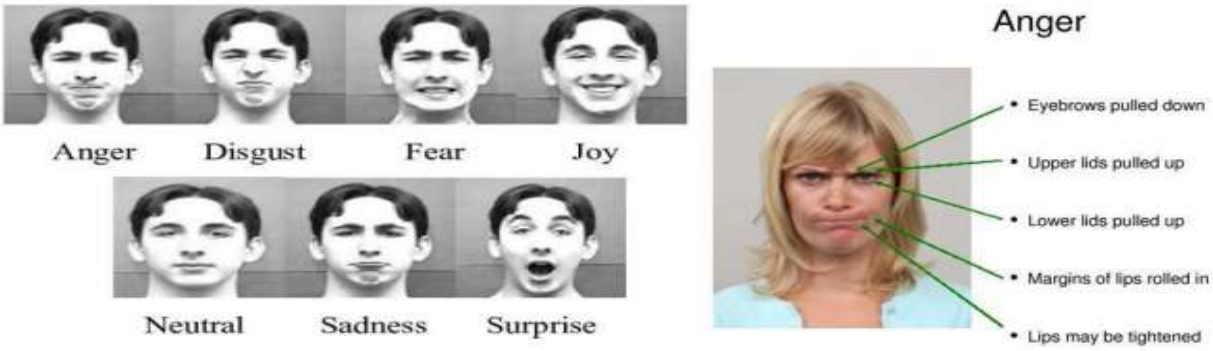


**Fig.1.** different types of emotion in a human-being

### 1.1 Face Detection

The primary goal of the face detection technique is to identify the face in the frame by reducing external noises and other factors.

The steps in the face detection process are as follows:

1. Image pyramid

2. Histogram of Oriented Gradients

3. Linear Classifier

The obtained data is decomposed into the sampling image using an image pyramid into multiple scales. This technique is simply used to extract features while reducing noise and other factors. [1,3] The low pass image pyramid technique (also known as Gaussian pyramid) consists of smoothing the frame and sub sampling it by decreasing its resolution; the process must be repeated several times to achieve a perfect result; at the end of the process, we obtain a frame similar to the original but with a lower resolution and a higher smoothing level. HOG is a feature descriptor, a technique that counts occurrences of gradient orientation in a localized portion of an image that is commonly used to detect objects in images in the field of image processing. The main goal of using this technique is to describe the face within the image using a set of intensity gradient distributions.

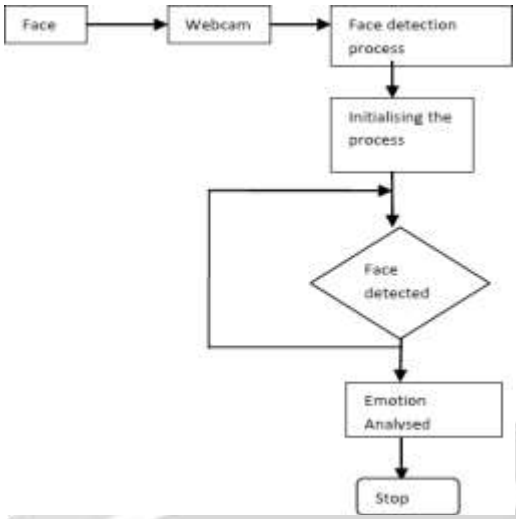


**Fig.2.** Flow diagram of Face Detection

### 1.2 Emotion Classification

When the face is successfully detected, a bounding box is applied to the image as an overlay to extract the ROI (face) for further analysis. The extracted ROI will then be processed with the "Predictor" function, also known as a script, to extract the 68 facial landmark points and save them in an array. Following that, the data stored in the features array will be fed into a PCA reduction code, which will reduce the size of the data and remove any correlated coordinates, leaving only the necessary points as principal components. The data is a 68x2 array with 68 points, each with coordinates on the x and y axes. The array will be converted into a vector with 136 rows and 1 column.

The "Predictor" facial landmark extraction code is trained using a set of images and landmark maps for each image. Using regression trees trained with the gradient boosting algorithm, the code learns how to extract the [4] facial landmark map of a given face image based on the pixel intensity values indexed of each point. The data obtained after the [3] CNN reduction operation will be used for classification. A multiclass CNN with a linear kernel is used to compare inputted data to stored data to determine which class (emotion) it belongs to. If one of the three emotions, anger, fear, or surprise, is detected, a speed decreasing command will be executed to reduce the wheelchair's speed in order to keep the user safe.

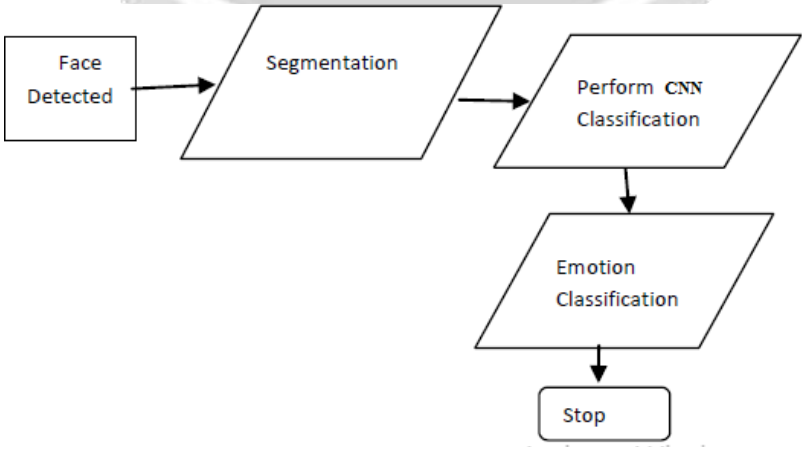


**Fig.3.** flow diagram of emotion classification

### 1.3 Music Recommendation

Because the input is acquired in real-time, the camera is used to capture the video before framing. The framed images are processed using the hidden Markov model classification. For the purpose of emotion classification, the obtained frames are considered in all frames and pixel formats. The value of each landmark in the face is computed and saved for later use. The classifier's efficiency is approximately 90-95 percent, which means that even if the face changes due to environmental conditions, the system can still identify the face and the emotion being expressed.

The emotions are then identified using the values that are obtained and set, and the value of the pixel that is received is compared to the values that are present as threshold in the code. The values are sent to the web service. The song is played based on the emotion detected. Each song has its own set of emotions. When the emotion is transferred, the appropriate song is played. There are seven emotions that can be used: happy, angry, sad, and surprised. When the happy emotion is recognized, the songs assigned to that emotion are played, and the same is true for the other emotions; that is, the songs are played for the emotions detected respectively.

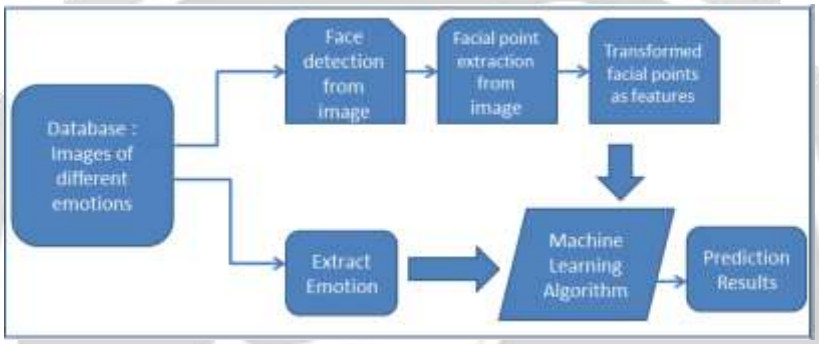


**Fig.4.** Flow diagram of Music Recommendation

## 2. Phases in Facial Expression Recognition

The supervised learning approach is used to train the facial expression recognition system, which takes images of various facial expressions. The training and testing phases of the system are followed by image acquisition, face detection, image preprocessing, feature extraction, and classification. [2,3] Face detection and feature extraction are performed on face images, and the results are classified into six classes corresponding to seven4 basic expressions, as detailed below:

### 2.1 Image Acquisition

Images used for facial expression recognition are static images or image sequences. Images of face can be captured using camera.

### 2.2 Face Detection

Face Detection is useful for identifying facial images. Face detection is performed in the training dataset using a Haar classifier called the Voila-Jones face detector, which is implemented in OpenCV. [1] Haar-like features encode the difference in average intensity in different parts of the image and are made up of black and white connected rectangles, with the value of the feature being the difference in the sum of pixel values in the black and white regions.

### 2.3 Image Preprocessing

Image pre-processing includes the removal of noise and normalization against the variation of pixel position or brightness.

a) Color Normalization
b) Histogram Normalization

### 2.4 Feature Extraction

The most important aspect of a pattern classification problem is the selection of the feature vector. After pre-processing, the image of the face is used to extract the important features. Scale, pose translation, and variations in illumination level are all inherent problems in image classification.

## 3. CONCLUSIONS

In this project, we presented a music recommendation system based on emotion detected. The system uses a two-layer convolution network model for facial emotion recognition. The model classifies 7 different facial emotions from the image dataset. The model has comparable training accuracy and validation accuracy which convey that the model is having the best fit and is generalized to the data. Since we are using CNN along with facial landmark the accuracy of the system is almost 87%. The rest of the accuracy can be improved by regularly monitoring the user like around every 3 minutes. Facial landmark recognizes the face because of which the process can be fastened. This recommendation also includes movie. We also recognize the room for improvement. It would be interesting to analyze how the system performs when additional emotions are taken into consideration. User preferences can be collected to improve the overall system using collaborative filtering.

## 4. REFERENCES

[1]. Arnaja Sen, Dhaval Popat, Priyanka Kuwor and Hardik Shah. Music Playlist Generation using Facial Expression Analysis and Task Extraction DOI: 10.17951/AI.2016.16.2.1 .2019.

[2]. G. Yang, J. S Saumell, J Sannie, "Emotion Recognition using Deep Neural Network with Vectorized Facial Feature" 2018 IEEE International Conference on Electro/Information Technology (EIT).

[3]. Jie Wang and Zihao Li. Research on Face Recognition Based on CNN. 2nd International Symposium on Resource Exploration and Environmental Science,2018.

[4]. Adil Sarsenov and Konstantin Latuta. Face Recognition Based on Facial Landmarks, Published in: 2017 IEEE 11th International Conference on Application of Information and Communication Technologies (AICT). Date of Conference: 20-22 Sept. 2017.