

# Identification of a Deep Learning Technique for Facial Emotional Datasets for Song Recommendation System

Albert Mayan J<sup>1</sup>, A Mary Posonia<sup>2</sup>, Hariharan B P<sup>3</sup>, Jagadeesh N<sup>4</sup>, Mohnish Devaraj<sup>5</sup>

<sup>1,2</sup>Associate Professor, School of Computing, Sathyabama Institute of Science and Technology, Chennai.

<sup>3,4,5</sup> U.G Student, Department of CSE, Sathyabama Institute of Science and Technology, Chennai

<sup>1</sup>albertmayan@gmail.com, <sup>2</sup>maryposonia.cse@sathyabama.ac.in, <sup>3</sup>bphariharan1301@gmail.com, <sup>4</sup>jagadeeshshirs2456@gmail.com, <sup>5</sup>monish.devaraj312002@gmail.com

**Abstract**—An Android application called "SONG Recommender System Using Emotion Detection" allows users to play music in accordance with their mood (emotion). The user can choose their language and employ emotion recognition using this app. All of the prior research employed various algorithms, such as fisher faces, SVM, K-nearest algorithms, etc., however those methods were not always effective. Recent research recommended combining emotion recognition (ER) with deep learning (DL) applications to lessen the amount of manual work required from the user to search for and listen to songs. Convolutional neural networks (CNN) can be used to determine users' emotional recognition using the FER2013 dataset, according to this systematic review.

We provide a fresh method for automatically playing music while analyzing facial expression. The majority of the current methods include manually playing music, wearing wearable computers, or categorizing based on auditory attributes. We suggest changing the manual sorting and playing instead. For emotion identification, we have employed a convolutional neural network. The computational Duration needed for produce the findings and the overall price involved in the system are both likely to be reduced by our proposed approach, boosting the system's overall accuracy. The FER2013 dataset is used to test the system. An internal camera is used to record facial expressions. On the input face photos, feature extraction is done to identify emotions including happiness, rage, sadness, surprise, and neutrality. A music selection is created automatically by determining the user's current feeling.

**Keywords** - CNN, Convolutional Neural Network, Deep Learning, Song Recommender, Face Recognition, Feature extraction, Emotion detection, Music Player, Camera.

## I. INTRODUCTION

By identifying and capturing the user's emotions in real time, this notion suggests music to the user. We presented a method to categorize various types of music into distinct

moods, such as joyful, sad, furious, etc.

Existing techniques were employing collaborative techniques that would use user data from prior sessions to select music. However, these strategies take a lot of manual work.

Additionally, the branch of research known as Computer Vision (CV) enables computers to learn by employing various strategies and methodologies in order to capture what is present in an image or a video. Currently, computer vision has a diverse array of applications, such as recognizing faces, autonomous vehicles, medical analysis, and more. Today, we will delve into the exciting field of detecting emotions through facial expressions, a noteworthy use of CV. By seeing your facial expressions, CV can identify and describe your sentiment. It can tell if you're upset, joyful, glad, etc.

## II. METHODS

This study is solely focused on determining the best deep learning (DL) algorithm for emotional recognition that provides almost perfect accuracy. The necessary data was collected using a systematic review process, and quantitative methods were used to analyze and filter the data collected in accordance with the needs of the study.

In this work, the following critical methods have been applied: -

1. Facial Detection — The capability to detect a face in any given image or frame using facial detection. The results are the detected faces' bounding box coordinates.
2. Facial Recognition—To determine which faces belong to the same person, compare various faces together. Comparing facial embedding vectors allows for this.
3. Emotion Detection — identifying the various facial expressions as pleased, angry, sad, neutral, or surprised.

### CREATING RESEARCH QUESTIONS (RQ):

The following table1 shows the developed RQs according to the author's motivation for this research.

S.No	RQs	Motivations
1	Which is the most suitable DL technique for emotion recognition?	Researchers have used various DL techniques but, we are yet to know which has the near perfect accuracy.
2	What is the most suitable freely available face dataset to train and test the developed DL model?	Researchers have used various datasets to train their particular model but, for our study we have chosen The FER2013 datasets from Kaggle which has a wide variety of data's in it (50,000) and the test accuracy has Reached 63.2%.

### III. LITERATURE REVIEW

Examining the techniques is carried out to understand their weaknesses and how to improve them. The text of a scholarly paper, known as a literature review, encompasses the latest research findings and key theoretical and methodological advancements in a specific field. There is worldwide interest among students, researchers, engineers, and other professionals in the latent abilities of individuals to provide inputs to systems through various means. Facial expressions can give insight into a person's emotional state. In interpersonal communication, we often convey our emotions through hand gestures, facial expressions, and vocal tones. Many users have reported that creating and maintaining a large playlist takes a considerable amount of time. According to the report, the music player adapts to the user's mood and selects a song accordingly. To generate mood-based playlists, the application classifies audio files based on audio attributes. To evaluate the efficacy of the developed model, Support Vector Machine (SVM) was used to categorize features into five major universal emotions: anger, joy, surprise, sadness, and disgust.

Ajinkya Khobragade[2] suggested a song recommendation system utilising Python and Tableau, but he desired for this system to operate in a real-time setting and to capture the alignment of the facial structure to be developed for the comfort of users.

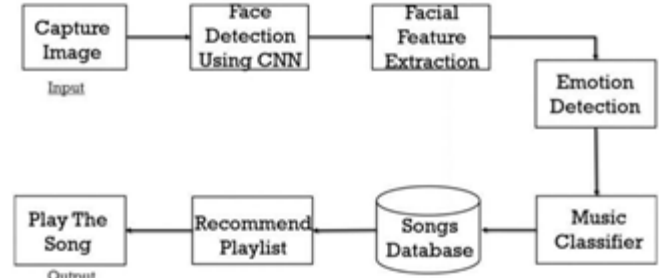


Fig 3.1: Yusuf's work flow

Yusuf Yaslan<sup>[12]</sup> and his colleagues proposed an emotion-based music recommendation system that utilizes wearable technology with physiological sensors for galvanic skin response (GSR) and photo plethysmo graphy (PPG) to detect the user's emotions. Emotions are a key element of human nature and play a vital role in our lives. This study uses multi-channel physiological signals to predict arousal and valence in order to address the difficulty of emotion detection.

According to Ayush Guidel et al., facial expressions make it simple to determine a person's emotional state of mind at any one time.

### IV. PROPOSED SYSTEM OVERVIEW

The user-music player interaction in the suggested system is advantageous to us. This system's objective is to use the camera to acquire a facial image. The CNN then utilized to forecast the emotion using the captured visuals. A customized list of songs is created based on the detected response in the captured image. The objective of the system is to automatically generate a music list that reflects the user's responsiveness.

Four modules make up the response-based music system:

- On-spot Capture: The system must accurately capture the person response in this module.
- Face Recognition: Here, the person response will be used like data. The CNN is set up to analyze the pictorial characteristic of a person.
- Emotion Detection: In this stage, the user image's features are extracted in order to identify their emotions, and the algorithm then generates captions based on those emotions.
- Music Recommendation: By matching the user's emotions to the song's mood type, the recommendation module suggests songs to them

### V. METHODOLOGY

We used the Kaggle dataset to create the CNN model. The dataset, FER2013, is Split into combination of two different dataset, said to be training and testing dataset. There are 24,176 photos in the training dataset and 6043 in the testing dataset.

The collection contains 48x48 pixel grayscale portraits of people. In FER-2013, each picture is assigned to a particular emotion among the set of five pre-loaded emotions such as

neutral, happy, sad, angry, and surprised. Faces present in the

image are automatically recognized in order to make sure it is approximately centered and occupies the same amount of area in each photograph. Google image searches for every emotion were used to compile the dataset.



Fig 5.1 An Angry Face

When trained on datasets with unequal distribution of emotions, facial expression recognition systems may excel in recognizing dominant emotions such as happy, sad, angry, neutral, and surprised, but fall short in accurately identifying less frequently represented emotions.

A more detailed view on how the algorithm works: -

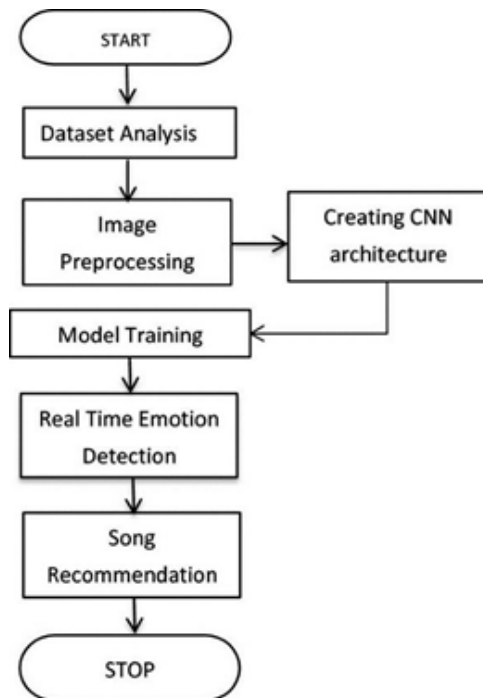


Fig: 5.2 Algorithm Flow Chart

This issue is typically addressed using the weighted-SoftMax loss strategy attempts to address this issue by assigning higher weights to the underrepresented classes. This way, the model is encouraged to pay more attention to the underrepresented classes and improve its accuracy in recognizing them. The weighted-SoftMax loss is calculated as the cross-entropy loss between the predicted probability distribution and the true

distribution, where the weights are applied to the true distribution. We have applied the categorical cross-entropy loss function to handle missing and outlier values. A chosen loss function is used for each iteration to estimate the error value. Therefore, we have utilized a loss function called categorical cross-entropy to treat missing and outlier values.

#### Steps:

**Step 1:** The user gives the input, in the form of the captured image using the camera.

**Step 2:** The image is then analyzed by the model and gets classified into a particular category of emotion.

**Step 3:** The data gets extracted and detected with the training datasets, which are refer-13 datasets.

**Step 4:** A particular customized list of songs(playlist) is chosen according to the user's emotion.

**Step 5:** The music is played depending on the emotion to boost the mood of the user.

#### Face Detection:

One of the uses of Computer Vision is facial detection. It is the procedure by which a specific algorithm is created and taught to find faces. A video or image can be used to perform this detection in real time. This approach employs classifiers, algorithms designed to determine the presence (1) or absence (0) of a face in an image. To boost accuracy, the classifiers undergo training with a substantial number of photos to recognize faces. The technology OpenCV utilizes two distinct types of classifiers, namely Local Binary Pattern (LBP) and Haar Cascades. In order to effectively detect faces, a Haar Cascade classifier is utilized. It is trained with pre-defined variable face data. Facial detection primarily focuses on locating the face within the frame by disregarding external disturbances and distractions. The technique is grounded in machine learning and utilizes a set of input files for training the cascade function. The Haar Wavelet approach is used to divide the image's pixels into squares based on their functions [9]. The extraction of training data is achieved with a high level of accuracy through the utilization of machine learning techniques.

#### Emotion Detection

The CNN architecture creates feature maps or activation maps using the Relu activation function by applying filters or feature detectors to the input image. These filters or feature detectors can identify edges, lines, curves, and other characteristics present in the picture. The feature maps are then subjected to pooling to maintain translation invariance, meaning that even small changes in the input should result in unchanged outputs. While pooling can be performed using minimum, maximum, or average, max-pooling is considered superior in performance compared to the other methods. Finally, the inputs are flattened before being fed into a deep neural network, which outputs the class of the object.

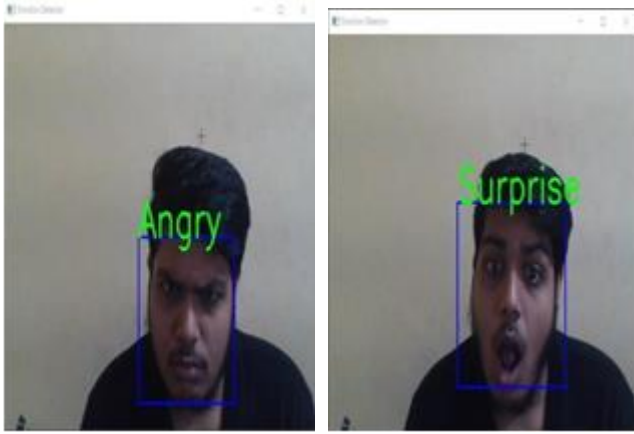


Fig: 5.3 Emotion Screen Shots



Fig: 5.4 Dataset Image Samples

The basic process involves feeding an input image into a CNN model, which then produces the results [10]. A CNN-trained model, with learned weights, is utilized for emotion detection. When a real-time photo is submitted by a user, it is processed by the pre-trained CNN model, which assigns an emotion label to the image based on its prediction.



Fig: 5.5 Features Extracted for identification

## VI. RESULT & ANALYSIS

By default, this solution recognizes all faces in the webcam image feed that exhibit emotions. The test accuracy for a straightforward 4-layer CNN was 63.2% after 50 iterations.

We also encountered overfitting during training due to the high complexity of transfer learning models and the comparatively modest size of our datasets. Our ResNet transfer learning model soon over fit to the training data, with train accuracy starting to flatten after only 25 epochs, even though we added 30% dropout for the last three layers.

The model comprises of a softmax output layer, a 1024-by-1024-pixel FC layer, and three stages of convolutional and max-pooling layers.

Filters of sizes 5x5, 4x4, and 5x5 are used in 32, 32, and 64 convolutional layers, respectively.

The kernels used by the max-pooling layers are 3x3 in size. As the activation function, ReLU was used.

Additionally, we added batch norm at every layer and 30% dropout after the last FC layer to boost performance. Optimize the cross-entropy loss using a momentum of 0.3 to fine-tune the model. Initial learning rates of 0.2, 32, and 0.0001 are predetermined, as are batch size and weight decay. If the validation accuracy does not increase, the learning rate is cut in half.

Epoch occurs when the neural network only processes an entire dataset ONCE, both forward and backward.

## VII. CONCLUSION

The programme seeks to give operating system users a more user-friendly, versatile, and dependable emotion-based music system. People looking for music based on emotion and emotional behaviour would benefit from the Emotion-based Music Program. It could aid in reducing the amount of time spent searching for music, decreasing the amount of time wasted on computing, and ultimately improving the system's accuracy and efficiency. By employing a camera to capture the image, identifying the end-emotion, user's and providing a personalised playlist, it makes the end-job user's easier.

With the help of this initiative, machine learning technology will advance significantly. The purpose of an emotion-based music player is to organise the music according to the user's emotions, such as happiness or sadness.

## VIII. FUTURESOCPE

Even though this system is fully functional, there is still room for future development. The application can be altered in several ways to improve user experience overall and output better results. Some of these use an alternate approach based on extra emotions like disgust and fear that are not considered in our methodology.

The system's future plans include designing a mechanism that could aid in the treatment of individuals who are experiencing mental stress, anxiety, acute depression, and trauma through music therapy. Due to the current system's poor camera resolution and performance in extremely low light levels, there is a chance to add some functionality as a potential fix in the future.

In the future, this might be improved and added to iPods and other similar gadgets, making it much more practical.

## IX. LIMITATIONS

Due to the limited number of photos in the used image dataset, the system is currently unable to accurately record all emotions. For the classifier to produce accurate results, the image that it receives must be captured in a well-lit environment. The classifier can accurately estimate the user's sentiment if the image quality is at least higher than 320p

## **X. ADVANTAGES OF PROPOSED SYSTEM**

- Users do not need to select songs manually.
- No need of existing playlist.
- Users do not need to classify

[12] Yusuf Yaslan, Deger Ayata, Mustafa E. Kamasak, Emotion Based Music Recommendation System Using Wearable Physiological Sensors, IEEE journal, Volume 64, Issue 2, May 2018.

## **REFERENCES:**

[1] Preema J.S, Rajashree, Sahana M, Savitri H, Review on facial expression-based music player, International Journal of Engineering Research & Technology (IJERT), ISSN-2278-0181, Volume 6, Issue 15, 2018.

[2] AYUSH Guidel, Birat Sapkota, Krishna Sapkota, Music recommendation by facial analysis, February 17, 2020.

[3] CH. sadhika, Gutta.Abigna, P. Srinivas reddy, Emotion-based music recommendation system, Sreenidhi Institute of Science and Technology, Yamnampet, Hyderabad; International Journal of Emerging Technologies and Innovative Research (JETIR) Volume 7, Issue 4, 2020.

[4] Julian Menezes .R, Albert Mayan .J, M. Breezely George,” Development of a Functionality Testing Tool for Windows Phones”, Indian Journal of Science and Technology, Vol:8, Issue:22, pp: 1-7, 2015.

[5] Vincent Tabora, Face detection using OpenCV with Haar Cascade Classifiers, Becominghuman.ai, 2019.

[6] Zhuwei Qin, Fuxun Yu, Chenchen Liu, Xiang Chen. How convolutional neural networks see the world - A survey of convolutional neural network visualization methods. Mathematical Foundations of Computing, 2018.

[7] Frans Norden and Filip von Reis Marlevi, A Comparative Analysis of Machine Learning Algorithms in Binary Facial Expression Recognition, TRITA-EECS-EX-2019:143.

[8] Ahmed Hamdy AlDeeb, Emotion- Based Music Player Emotion Detection from Live Camera, ResearchGate, 2019.

[9] Mary Posonia A, S. Vigneshwari, Albert Mayan J, D. Jamunarani, "Service Direct : Platform that Incorporates Service Providers and Consumers Directly", International Journal of Engineering and Advanced Technology (IJEAT) , Vol.8 ,No.6, pp.3301-3304, 2019.

[10] John, A.M., Prince Mary, S., Mohana Prasad, K., Naveena, M. and Laveti, N. (2022), 'Authentication for cloud computing system through smartcard', Int. J. Cloud Computing, Vol. 11, Nos. 5/6, pp.518–528.

[11] Albert Mayan J, Karthikeyan S, Nikhil Chandak, Bharat Mundhra and Padmavathy J, "Facial attendance system technology using Microsoft Cognitive Services", International Journal of Engineering Systems Modelling and Simulation, Vol. 12, Nos. 2/3, pp. 180-187 ,2021