Emotion based Movie Recommendation System using Deep Learning

Manjusha Sanke
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Shane Furtado
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Saloni Naik
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Shravani Nevagi
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

Vishal Bidikar
Dept. of Information Technology
Shree Rayeshwar Institute of Engineering and Information Technology
Goa, India

ABSTRACT
Emotions play a significant role in human perception and decision making. Recent studies on emotion detection and recognition are the subject for attention by researchers from different fields. This paper presents a web-based application which will be able to identify the emotion of user from the image uploaded on the server and stream the movies online. This paper focuses on seven emotional states: happiness, surprise, sadness, disgust, angry, fear and a neutral state which can be always found in daily life. The proposed system includes uploading the image to the server, detecting emotion from the image and finally presenting movie recommendation list based on the emotion of the user.

The system uses FER2013 (Facial Expression Recognition 2013) dataset and IMDB (Internet Movie Database) dataset. A deep learning model is trained using CNN technique for detection and classification of emotions. User's emotion is identified and a list of top-rated movies associated with it is presented.

General Terms
Image Processing, Neural Networks, Deep Learning

Keywords
Emotion recognition, FER13, Movie recommendation, Deep learning.

1. INTRODUCTION
A recommender system [9] refers to a system that is capable of predicting the future preference of a set of items for a user and recommends the top items. The customer usually provides the recommender system with characteristic data. The recommender system applies one or several recommendation techniques on these data and then recommends products to the customers. However, for subjective and complexes products such as movies, music, perfume, the task of rating or describing the desired product characteristics is quite difficult for customers.

While language is crucial to human communication, in most interactions it is supplemented by various forms of expressive information, such as facial expressions, vocal nuances, gesture and posture. Facial expressions are studied since ancient times, as it is one of the most important channel of non-verbal communication. Image Recognition is one of the most significant machine learning and artificial Intelligence examples. Face detection is a form of image recognition which automatically finds the face region from the input images.

The general approach to automatic facial expression analysis consists of three steps:

- face detection and tracking
- feature extraction involves extracting meaningful or discriminative information caused by facial expressions
- expression classification / recognition.

Although humans are filled with various emotions, modern psychology defines six basic facial expressions that are Happiness, Sadness, Surprise, Fear, Disgust, and Anger as universal emotions [1]. Facial muscles movements help to identify human emotions. Basic facial features are eyebrow, mouth, nose and eyes, cheeks. These features are actively used in training the emotion recognition model.

2. RELATED WORKS
Many researchers have used various methods in emotion recognition.

Alramzana Nujum Navaz et al. [1] proposed a solution that applies and compares four deep learning models for image pre-processing with namely Alexnet, GoogLeNet, and CNN. The maximum overall accuracy improvement from experiments was identified in CNN, from 23.5 to 59.4 % (≈ 36%increase).

P. Kaviya et al. [2] The proposed model achieves a final accuracy of 65% for Facial Expression Recognition (FER)-2013 and 60% for custom datasets. The proposed model achieves a final accuracy of 65% for Facial Expression Recognition (FER)-2013 and 60% for custom datasets.

Akriti Jaiswal et al. [3] This paper proposed a Convolutional Neural Networks (CNN) using two datasets Facial emotion recognition challenge (FERC-2013) and Japanese female facial emotion (JAFFE). The accuracies achieved with proposed model are 70.14 and 98.6.

Zeynab Rzayeva et al. [4] proposed a CNN model that is trained

on Cohn-Kanade and RAVDESS datasets to find 5 major facial emotions.

Renuka S. Deshmukh et al. [5] The system focused on live images taken from the webcam and developed automatic facial emotion recognition system for stressed individuals thus assigning them music therapy so as to relief stress.

Noel Jaymon et al. [6] experimented with three models as follows Simple, Inception, Xception model and showed test accuracy of 54%, 61.42% and 65.2% respectively.

Rabie Helaly et al. [7] used convolutional neural network model with dataset FER 2013. When the developed system is trained and tested on the GPU, the accuracy of the results is 94%.

Shlok Gilda et al. [10] developed an Emotion music player, which recommends music based on the real-time mood of the user. The Emotion Module takes an image of the user's face as an input and makes use of deep learning algorithms to identify their mood with an accuracy of 90.23%.

Table 1. Survey table

Ref. No.	METHODS	MERITS	DEMERITS	DATASET USED
[1]	Alexnet, GoogLeNet, CNN	1.Manual labeling of images (Alexnet), 2. Removal of confusion emotion thereby narrowing the classes (GoogLeNet), 3. Image enhancement using CNN	Disgust and neural emotions were never classified correctly.	Facial image data of Indian film stars.
[2]	CNN	Haar filter to detect and extract face features	Several neural images are predicted as sad neighbor	FER-2013
[3]	CNN	Computing time was reduced in the Japanese model as compared to FER-2013	In FER dataset we train on 32,298 samples which is validate on 3589 samples and in JAFFE dataset we train 833 samples.	FER-2013 & Japanese female facial expression (JAFFE)
[4]	CNN	Predicted surprise and happy emotions better in comparison to others	-	Cohn-Kanade & RAVDESS
[5]	SVM (Support vector machine)	The pre-processing step include converting the facial expression into binary image.	SVM have a good generalization performance, but their results are in general less sparse	CK+, FERG, NVIE
[6]	Simple, Inception, Xception model	-	Lack of robustness	FER- 2013
[7]	Xception CNN	This exception model is trained on more than millions of images. The gained results confirmed that this system can be effectively implemented and utilized using low powered embedded system.	Vibrant environment	FER-2013
[8]	CNN	iCVMEFED was selected as it offered completely raw images.	Data Augmentation was needed. Faced problems while detecting feared and surprised emotion/face.	FER-2013, AffectNet & iCVMEFED

3. DATASETS

Below mentioned are some of the datasets used by our model.

3.1 FER2013 (Facial Emotion Recognition 2013)

The FER13 dataset [14] is provided by Kaggle. It was introduced in 2013. The dataset contains images along with categories describing the emotion of the person in it. The dataset contains 48×48 pixel grayscale images with 7 different emotions such as Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The dataset contains 28709 examples in the training set, 3589 examples in the public testing set, and 3589 examples in the private test set.

3.2 IMDB

The IMDB **(**Internet Movie Database**)** [15] dataset consists of 50K movie reviews for natural language processing or Text analytics. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. It provides a set of 25,000 highly polar movie reviews for training and 25,000 for testing.

4. MODEL USED

Convolution Neural Network (CNN) [4] is a type of deep learning model for processing data that has a grid pattern, such as images, which is inspired by the organization of animal visual cortex and designed to automatically and adaptively learn spatial hierarchies of features, from low- to high-level

patterns.

CNN is a mathematical construct that is typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification.

A convolution layer plays a key role in CNN, which is composed of a stack of mathematical operations, such as convolution, a specialized type of linear operation. In digital images, pixel values are stored in a two-dimensional (2D) grid and a small grid of parameters called kernel, an optimizable feature extractor, is applied at each image position, which makes CNNs highly efficient for image processing, since a feature may occur anywhere in the image.

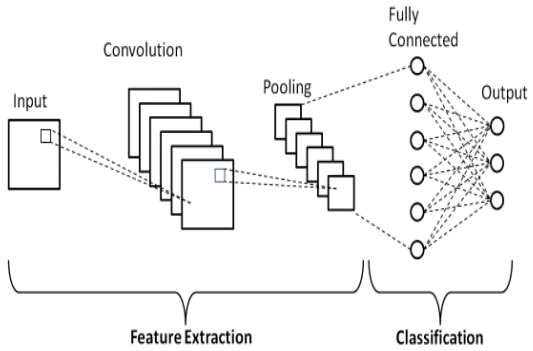

Fig 1: CNN Model

5. METHODOLOGY

The system architecture is as follows: -

i. Dataset - FER 13 dataset contains approximately 30,000 images which will give a great accuracy in detection of the image. The images are of 48*48 pixels and already converted into grayscale image. It provides categories of image which are angry, fear, happy, sad, surprise and neutral.

ii. Pre-processing - As known that FER 13 dataset is beforehand pre-processed, but to avoid any errors a pre-processing step in added in the model. In Pre-processing the input image is cropped into 48*48 pixels resolution and image is converted to gray scale image.

iii. FER 13 is further divided into 2 sets; training and testing set. Training set contains 28709 images and testing contains 7187 images which will be used for training and testing the model.

iv. Face detection - Extracting the properties which will lead to the particular emotion, it will use facial expression and parts of an image to determine the emotions.

v. Feature extraction - process of extracting the face features from a face image.

vi. Classification - In classification the emotion output from the face detection and feature extraction will be distinguished to its particular categories, weather it falls under the category of angry, sad, happy or neutral.

vii. Trained model - Training the model with FER 13 dataset which will give use an admissible accuracy. It will consist of sample output data and corresponding set of input data. The input data will run through some algorithm to get the output against the sample output so it will conclude in which emotion categories the input data falls.

viii. User-interface - To take input image from user's a user interface is created. First the user will login, after login the user will upload the input image which will go through the pre-processing, face detection feature extraction and to the trained model which will predict the emotion.

ix. Template matching - With the help of template matching the emotion with movies which corresponds to the predicted emotion.

x. IMDB database - Contains movies that are correlated with the particular category of emotions.

xi. Playlist - On the basis of the input data and the emotion predicted by the trained model the user is been provided with the playlist of movies.

xii. Play - The User can then pick a movie of his liking from the provided movie list.

5.1 Pre-Processing

Image pre-processing [13] involves the steps taken to format images before they are used for model training and inference. This includes, but is not limited to, resizing, orienting, and colour corrections.

Image pre-processing may also decrease model training time and increase model inference speed. If input images are particularly large, reducing the size of these images will dramatically improve model training time without significantly reducing model performance.

5.2 Face Detection

Face detection applications use algorithms and ML to find human faces within larger images, which often incorporate other non-face objects such as landscapes, buildings and other human body parts like feet or hands. Face detection algorithms typically start by searching for human eyes. The algorithm might then attempt to detect eyebrows, the mouth, nose, nostrils and the iris.

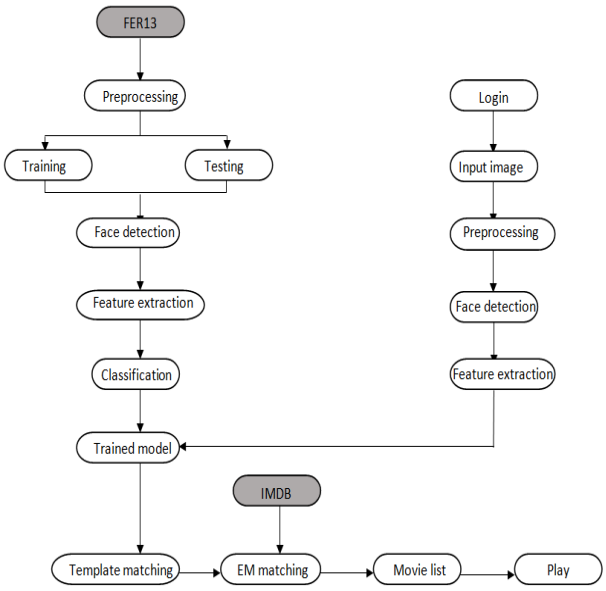

Fig 2: System Block Diagram

6. RESULTS

Below shown are some of the emotions tested with the model.

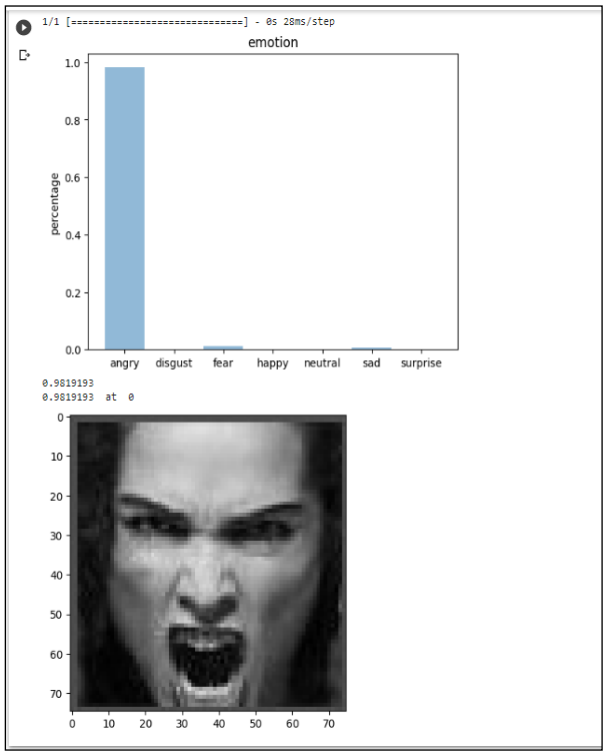

Fig 3: Anger emotion result

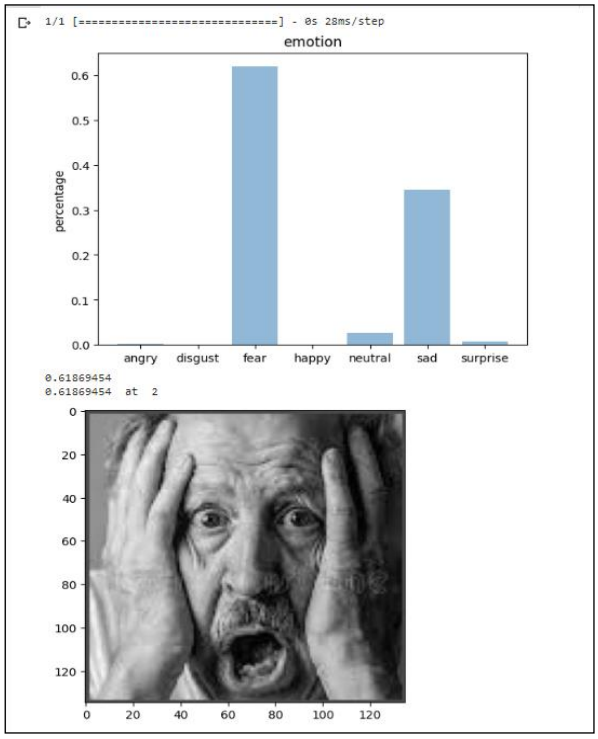

Fig 4: Fear emotion result

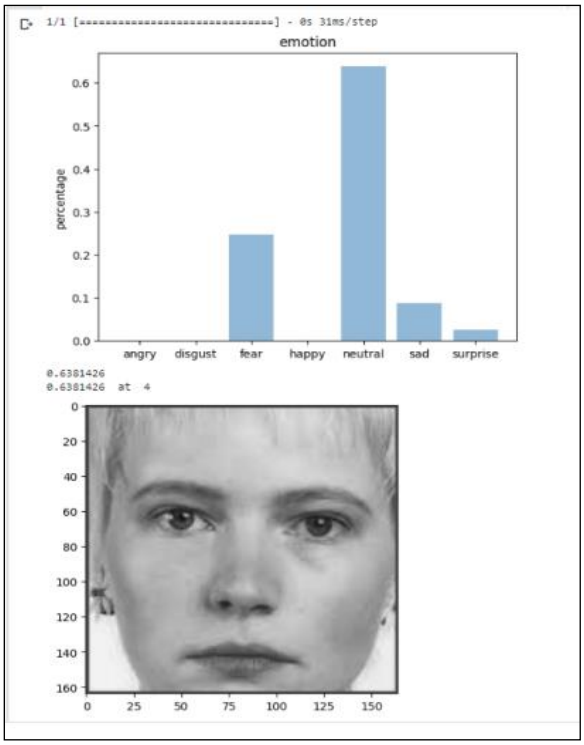

Fig 5: Neutral emotion result

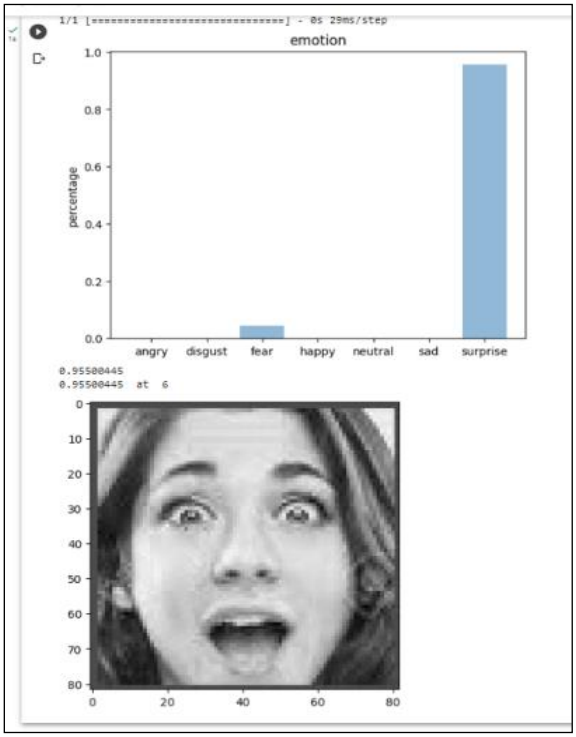

Fig 6: Surprise emotion result

7. CONCLUSION

Facial expression is an important channel for human communication and can be applied in many real-world applications. Many people like to watch movies, but they have to browse through movies according to their mood and then select one.

An emotion-based movie recommendation system proposed in this paper eases the job of the users. A web-based application captures the users' image via a webcam or uploads an image

stored on the user's system. The emotion is detected from the uploaded image and a movie playlist is displayed on the website according to the emotion of the user.

In this paper, a deep learning model is trained using CNN technique for detection and classification of emotions. Using image processing and machine learning, the user's emotion is successfully identified and presented a list of top-rated movies associated with it. The CNN model achieved an average accuracy of 87.72% when training with FER-13 dataset.

8. REFERENCES

[1] Alramzana Nujum Navaz; Serhani Mohamed Adel; Sujith Samuel Mathew "Facial Image Pre-processing and Emotion Classification: A Deep Learning Approach", 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications.

[2] P. Kaviya; T. Arumugaprakash "Group Facial Emotion Analysis System Using CNN", 2021 5th International Conference on Computing Methodologies and Communication.

[3] Akriti Jaiswal; A. Krishnama Raju; Suman Deb "Facial Emotion Detection using Deep Learning", 2020 International Conference for Emerging Technology.

[4] Zeynab Rzayeva; Emin Alasgarov "Facial Emotion Recognition using CNN", 2022 IEEE INMIC. Technology.

[5] Renuka S. Deshmukh; Vandana Jagtap; Shilpa Paygude "Facial Emotion Recognition System through Machine Learning Approach", 2017 International Conference on Intelligent Computing and Control Systems.

[6] Noel Jaymon; Sushma Nagdeote; Aayush Yadav; Ryan Rodrigues "Real Time Emotion Detection using Deep Learning", 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies.

[7] Rabie Helaly; Mohamed Ali Hajjaji; Faouzi M'Sahli; Abdellatif Mtibaa "Face Recognition Model using Neural Network", 2020 IEEE INMIC. Technology

[8] Sabrina Begaj; Ali Osman Topal; Maaruf Ali "Emotion Recognition Based on Facial Expression using CNN", 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications.

[9] Leo Pauly; Deepa Sankar "Product Recommendation System from Emotion Detection", 2015 IEEE ICCICCT.

[10] Shlok Gilda; Husain Zafar; Chintan Soni; Kshitija Waghurdekar "Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation", 2017 International Conference on Wireless Communications, Signal Processing and Networking.

[11] Rabia Qayyum; Vishwesh Akre; Talha Hafeez "Android based Emotion Detection using Convolutions Neural Network", 2021 International Conference on Computational Intelligence and Knowledge Economy.

[12] J.Jayapradha Soumya Sharma, Yash Dugar "Detection and Recognition of Human Emotion Using Neural Network", International Journal of Applied Engineering Research.

[13] Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing Third Edition- Pearson Education. Inc, Prentice Hall,2008.

[14] FER2013 Dataset: https://www.kaggle.com/datasets/msambare/fer2013

[15] IMDB Dataset: https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews