Image Verification and Emotion Detection using Effective Modelling Techniques

Sumana Maradithaya, Vaishnavi S
Department of Information Science and Engineering
M S Ramaiah Institute of Technology, Bangalore, India

***Abstract*—The feelings expressed on the face reflect the manner of thinking and provide useful insights of happenings inside the brain. Face Detection enables us to identify a face. Recognizing the facial expressions for different emotions is to familiarize the machine with human like capacity to perceive and identify human feelings, which involves classifying the given input images of the face into one of the seven classes which is achieved by building a multi class classifier. The proposed methodology is based on convolutional neural organizations and works on 48x48 pixel-based grayscale images. The proposed model is tested on various images and gives the best accuracy when compared with existing functionalities. It detects faces in images, recognizes them and identifies emotions and shows improved performance because of data augmentation. The model is experimented with varying depths and pooling layers. The best results are obtained sequential model of six layers of Convolutional Neural Network and softmax activation function applied to last layer. The approach works for real time data taken from videos or photos.**

Keywords*—Face detection; face recognition; emotion detection; data augmentation

I. Introduction

Facial emotion recognition is a part of nonverbal communication. Recognizing emotions or detecting face from input image is a challenging task. Facial emotion detection is very useful in the research field of robotics, in detecting mental disorders and in marketing division as well. Human face detection has numerous applications in computer vision domain. Facial expression detection and analysis helps in sentiment analysis of humans. The proposed methodology presents a model, which can detect, recognize and understand the emotion of any given input image by creating different pipelines. Output of face detection pipeline is the bounding box faces. Then the output will be fed to face recognition pipeline. Face recognition verifies the given two images and outputs the Boolean value. Using convolutional neural network, model is built to detect seven classes of emotions that is sad, fear, disgust, neutral, angry, surprise and happy. Model is trained using open-source dataset that is FER (Face Emotion Recognition), downloaded from Kaggle. Using image augmentation model performance is improved.

The capability to recognize and differentiate among different faces is a boon for humans. Now with the evolution of technology in machine learning and computer vision, even computers can recognize different faces and are able to distinguish between them. Using face detection and recognition access to security applications can be improved. Example iPhone does the payments and the processes doesn't require any actual card. The model can also be used in crime detection and other healthcare related applications. Existing functionalities or models focus on analysis of complete facial features, and hence with multiple features it will be very confusing to train the model using convolutional neural network. The proposed model mainly focuses on detecting seven classes of emotions that is happy, sad, anger, disgust, fear, surprise and neutral. The proposed methodology uses face recognition library for detecting the facial locations.

The study [28] indicates the importance of identifying emotions in people, so that any physiological expressions can be detected and handled at an early stage. Facial Detection is nothing but the ability to identify the facial location of any given input image or in any video frame. Bounding box will be the output of the identified face. Face Recognition is nothing but comparing multiple faces to verify whether the face belong to same person. Face recognition is done using by differentiating among the embedding vectors of the face. Detected faces emotion is recognized using the model built based on convolutional neural network.

Face detection can be regarded as object class detection where facial locations can be identified and specify the size of the objects being detected in a specific image or images. With the advent of technology, it is now possible to identify the faces, in image or frame without any constraints like head posture, lightning, illumination, skin color. Face recognition determines whether the given images have face that is positive or negative. To achieve this motive, algorithm applied for the purpose will be trained on large datasets. Once the algorithm is trained, algorithm will be capable of analyzing the images to detect the object that is face in any images. When it comes to face detection, it is very essential for any algorithm used to know the parts of the given image to generate the face prints that are compared with the face prints stored previously to check the match.

The research [29] discusses the need of facial detection and emotion detection during the COVID -19 pandemic times. It further elaborates on the effectiveness of face recognition, face detection and emotion detection in the nonverbal communication.

Section II elaborates on the earlier works done, available methodologies and model building approaches. Section III focuses on the proposed work. Section IV discusses the obtained results. Section V` summarizes the proposed work done.

II. Literature Survey

The author in [7] has presented a model, which automatically recognizes face and facial emotions and classifies the emotions into eight different classes. This model uses Support vector machine algorithm for classification and the accuracy is almost 94%. This machine learning based model is used to detect real time images along with static images, and it cannot detect multiple faces in a given frame. Initially for face detection, HAAR cascade algorithm is used. Then using the face landmarks, trained dataset along with support vector machine algorithm face detection can be performed.

In [20], the author proposed a smart vision model, which can detect human face and emotions respectively. This model uses deep convolutional network for training the model. Even after applying some of the filters, accuracy achieved in the range of 48% to 80%. The model is used for classifying seven different emotions along with face recognition. Model uses swish activation function which is more efficient in terms of performance and maximum accuracy can be reached. Filters are nothing but parameters and layers in a fully connected neural network. Basically, shallow network (single layer) and modern deep network (8 layers) results are compared and dropout mechanisms are used to avoid overfitting.

The author in [10] proposes a deep learning model for emotion detection. Two different datasets are considered and performance evaluation of the proposed methodology is carried out. Validation accuracy and loss, computationally complex and time per step, learning rate, and detection rate are the factors considered for performance evaluation. Model is trained using convolutional neural network. The accuracy achieved is 79%. In [23], Viola Jones face detection algorithm is used for face detection in the image. The proposed methodology is used for human voice detection based on Mel frequency components. KNN classifier algorithm is used for facial emotion detection. Accuracy achieved is more than 90%. Features or a component of the face is calculated and these features are stored as database. Then face detection and emotion recognition can be done using these features stored in the database.

The authors in [5] proposed face detection model based on auto assistance neural network. Earlier approaches had both patch-based auto assistance neural networks integrated with global local-based auto assistance neural network. Accuracy achieved is 90% for detecting six classes of emotion. Model can be used to detect four classes of emotions in videos. The model can be used for detecting dynamic emotion detection from human videos.

The author in [25], proposes a model which makes use of two algorithms, one is based on colour intensity estimation and the other is based on Viola-Jones algorithm. Model is used to detect emotions based on spotting the micro expressions on the face. Colour intensity-based algorithm is very useful for smaller images and faster as well on coloured images. Viola-Jones algorithm can be used for greyscale images and high-resolution images. Both algorithms considered provide efficient outcomes in recognizing the face and eyes areas in the image. Model cannot predict entire face with higher accuracy.

The article [12], presents a model, which can be used for facial classification of emotion recognition. When considered face, there will be numerous factors that need to considered, like skin colour, textures, eye position and mouth and so on. The first step in this model is to recognize the skin using elliptical boundary. For facial features detection, an algorithm for extracting geometric and other anthropometric features from face image. Training and testing are done based on the classifiers and the accuracy achieved is 57% for detecting six different classes of emotions.

The author in [24] proposed the facial expression detection model based on Fisher face model. Facial features can be extracted using fisher face and facial expression detection is done using neural networks. The model is very sensitive in terms of illumination, noise and different backgrounds images are not suitable for this model. Accuracy is almost 86.85%.

The emotion detection of facial expression requires an image that is face of the person that needs to be recognized. Speech recognition doesn't require any image or person details. This model can be used in health department, education and customer service. Intelligence service makes use of this method to find the truthfulness of the investigation. The model uses MFCC for various frequency sound processing and it also uses Universal background models which eliminate the overfitting problems along with Gaussian mixture models for normal distribution as discussed in [1].

The authors in [8] propose a model which will capture image and detect the facial emotion based on Viola-Jones algorithm and prepares a list of song based on the emotion recognized. Viola-Jones algorithm is used for face detection and Fisher face features classifier is used for emotion detection. Model is used for detecting the face and based on the face detected, emotion recognition is performed and based on the emotion of the user, and music player is built. Accuracy is less.

One of the greatest challenges faced during face recognition is occlusion detection. This model is used for detecting the missing information in the occluded face. Viola-Jones algorithm is used for face detection and principal component analysis is used for face reconstruction and neural networks are used for face recognition. When considered other algorithms or methods, neural network yields the best result as seen in [19].

The author [22] proposed a model that is an electric wheelchair with emotion detection of the user with impairment. For capturing images, an RPI board with camera is used to record the frames captured in a video stream, along with Raspberry P12 and B model incorporated. This model can recognize the face using the images captured from the streams through sample python scripting, for face detection and sample emotion detection. Only three emotions classes were detected that is anger, fear or danger.

Study [17] elaborates on a model that is built based on hidden markov model, which can detect real time user's emotion. The model can detect six different classes of

emotions. Using hidden markov model, facial expressions can be separated and identified automatically. Model is not suitable for different positions and texture or orientation of the face with respect to images. And the accuracy is very low.

The research [3] proposes a methodology where local histograms are normalized along with their features. These features have orientations that is gradient and give better results for human detection along with densely overlapped grids. A new pedestrian database is introduced by the authors. The impact of different descriptor boundaries for fine-scale angles, fine direction binning, coarse spatial binning, nearby differentiation standardization in covering descriptor blocks is immensely significant for great execution.

The focal point of this investigation is on PC computerized impression of human feeling as discussed in [6]. The investigation was based on utilization of Fisher faces for the acknowledgment of human feeling in facial pictures. Large number of Fisher face models are trained and the overall model is considered in contrast to an autonomous test set. The results obtained are used to construct and test a compound progressive framework that endeavors to decipher human feeling continuously utilizing face discovery and calculations in combination with the look examination technique. Outcomes demonstrate that Fisher faces can be valuable in anticipating feeling dependent on content recovered from facial pictures.

Research [16] manages two particular utilizations of PCA in picture handling. The principal application comprises of picture shading that decrease, while the three shading segments are diminished into one containing a significant part of data. The second utilization of PCA exploits eigenvectors properties for assurance to choose object direction. Different strategies can be utilized for past object identification. Nature of picture division suggests to consequences of the accompanying cycle of item direction assessment dependent on PCA as well. Introduced paper momentarily presents the PCA hypothesis from the start and proceeds with its applications referenced previously. Results are archived for the chosen genuine pictures.

The study [2] features the improvement of an Android stage-based application named XBeats which goes about as a Music Player dealing with Picture Preparing essentials to catch, investigate and present music according to the feeling or temperament of the client utilizing this application. Android application was created utilizing the Android SDK programming and OpenCV programming was utilized to execute facial acknowledgment calculations and falls. The one-of-a-kind part of this task is that it centers around facial acknowledgment on the Android stage dissimilar to that on PC frameworks which utilize normally accessible programming projects for something similar. This article additionally gives correlation between utilization of different arrangement calculations utilized for facial location.

Recent FER frameworks spotlight on two significant issues: overfitting brought about by an absence of adequate preparing information and articulation inconsequential varieties, like light, head posture and personality predisposition. For the best in class in deep FER, the article [14], presents existing novel profound neural organizations and related preparing methodologies that are intended for FER dependent on both static pictures what's more, powerful picture groupings and talk about their benefits and restrictions. Several demonstrations and exploratory correlations on generally utilized benchmarks are likewise summed up. Later stretch out the overview to extra related issues and application situations. At the end, audit the leftover difficulties and comparing open doors in this field just as future headings for the plan of vigorous profound FER frameworks.

In [21], author proposes an emotion recognition framework which can be created by using the advantages of profound learning and various applications like criticism investigation, face opening and so forth can be carried out with great exactness. The primary focal point of this work is to make a Deep Convolutional Neural Network (DCNN) model that groups five unique human facial feelings. The model is prepared, tried and approved utilizing the physically gathered picture dataset.

Research [26] proposes a methodology to examine expressions of the face, dependent on both permanent and transient facial features, in an almost front facing view face picture grouping. The AFA framework perceives fine-grained changes in expressions of the face into action units (AU) of the Facial Action Coding System (FACS), rather than a couple of prototypic articulations. Multistate face and facial segment models are proposed for following and demonstrating the different facial highlights, including lips, eyes, foreheads, cheeks, and wrinkles. During following, point by point parametric depictions of the facial highlights are separated. With these boundaries as the sources of information, a gathering of activity units are perceived whether they happen alone or in blends. The framework has accomplished normal acknowledgment paces of 96.4% for upper face AU and 96.7% for lower face AU.

The authors of [4] have proven that emotional facial expressions are universal. Many other studies signify that different other culture, literate people show up the same expressions for few similar emotions. So by this fact, one can draw the conclusion that facial expressions are universal. And different cultured people have different expressions, but people in all variety of cultures are exposed to one single emotion and almost majority had the same expressions, irrespective of their facial shape, the expression was similar. Dataset was generated by considering three sets of faces using New Guinea. From the hypothesis curve, the results prove that facial expressions for certain emotions are universal.

In speech, emotion recognition is a theme on which little examination has been done to date. In [9], the authors talk about why emotions to be recognized in speech is an intriguing and relevant exploration point and present a framework for feeling acknowledgment utilizing one-class-in-one neural organizations. By utilizing an enormous data set of phonemes adjusted words, our framework is speaker and setting autonomous. Upon testing eight emotions approximately 50% of recognition rate was achieved.

Facial recognition framework is utilized in numerous applications going from HCI, reconnaissance to feelings. The

extension is wide and a great deal should be possible in this field. Highest accuracy is achieved through Neural Network (NN) as a classifier and the blend of Gabor Wavelet (GW) and Local Binary Pattern (LBP) is likewise great with a precision of 91.2 and 90%. The other techniques like Principal Component Analysis and Support Vector Machine's exhibition were low as discussed in [11].

The study [15] primarily considers face recognition with the segments by face parsing (FP). Considering the burden that various pieces of face contain distinctive measure of data for look and the weighted capacity are not the equivalent for various appearances, a thought is proposed to perceive look utilizing segments which are dynamic in demeanor exposure. The face parsing finders are prepared through deep conviction network and tuned by logistic regression. The identifiers initially recognize face, and afterward identify nose, eyes and mouth progressively. A profound engineering pretrained with stacked autoencoder is applied to face recognition with the concentrated highlights of distinguished parts. The parsing segments eliminate the excess data in articulation acknowledgment, and pictures should not be adjusted or some other fake treatment.

The authors in [13] proposed a deep convolutional neural network to arrange the 1.2 million high-goal pictures in the ImageNet challenge that is images are classified into 1000 variety of classes. Based on the testing data, the model accomplished top-1 along with top-5 blunder paces of 37.5% which is more than 17.0%, much better compared to the existing best in class. The neural organization, had 60 million boundaries and 650,000 neurons, which consists of five different convolutional layers, and some of the layers had max-pooling layers, along with three completely associated layers and last layer had 1000-way softmax activation function applied. To make processing quicker, model had non-immersing neurons along with effective GPU execution of the convolution activity. In order to reduce overfitting in the fully associated layers the model had created regularization technique called "dropout" which end up being exceptionally successful.

In the extended cohn-kanade dataset (ck+), [18] the quantity of arrangements is expanded by 22% and the quantity of subjects by 27%. The objective articulation for each succession is completely FACS coded and feeling names have been reconsidered and approved. What's more, non-presented arrangements for a few kinds of grins and their related metadata have been added. The article presents standard outcomes utilizing Active Appearance Models (AAMs) and a straight help vector machine (SVM) classifier utilizing a leave-one-out subject cross-approval for both AU and feeling location for the presented information. The feeling and AU marks, alongside the all-encompassing picture information and followed milestones will be made accessible July 2010.

Essential backpropagation, which is a straightforward strategy currently being broadly utilized in regions like example acknowledgment and deficiency finding, is inspected. The fundamental conditions for backpropagation through time, and applications to regions like example acknowledgment including dynamic frameworks, frameworks recognizable proof, and control are examined. Further expansions of this strategy, to manage frameworks other than neural organizations, frameworks including concurrent conditions, or genuine intermittent organizations, and other pragmatic issues emerging with the technique are portrayed. Pseudocode is given to explain the calculations. The chain rule for requested subordinates the hypothesis which underlies backpropagation is momentarily talked about. The emphasis is on planning an easier form of backpropagation which can be converted into PC code and applied straight by impartial organization clients as seen in [27].

III. Proposed Work

- To train the model training samples or images are collected using FER Dataset. FER is nothing but Facial Emotion or Expression Recognition. Dataset contains 48x48 pixel-based grayscale facial images. FER Data is used to identify the emotions of the unseen data into one of the seven classes of emotions. The dataset is created by browsing each emotion or expression and their synonyms.
- To detect the face and recognize from the given input images, face-recognition library is used. Given any input, image is converted into grayscale image and data is stored in the form of array using numpy array. Face-recognition library loads the given input image and then encodes the loaded image into feature vector and returns the Boolean value. To detect the emotion, faces being detected are taken as input and return the emotion type.
- Using data augmentation performance of the model is improved. The performance of the model is calculated by plotting the confusion matrix that is by considering the true positive and negative values. The accuracy is nothing but the precision, recall and f1-score.

A. Issues, Controversies, Problem

Improving the automatic facial emotion recognition will help in upgrading the social knowledge in the machines. In the field of computer vision, multiple applications make use of facial emotion recognition system and it is very challenging to identify multiple facial emotion in a given frame or image.

B. Proposed Methodology

Face detection is the key mechanism for emotion detection. Because of mental thinking and physical circumstances, humans tend to change their expression from dusk till dawn. The proposed methodology consists of input image, preprocessing, face detection, face recognition, feature extraction, emotion detection/classification. During face detection, the model automatically finds the face locations from the given input image. These face locations are the output of the face detection that is bounding box face and this achieved using face detection library. Once face is detected using face detection library face recognition is done. Once face is recognized, meaningful features are extracted and face emotions or expressions will be analyzed. Basic facial features are eyebrows, eyes, nose, cheeks and mouth. Facial

expressions are classified under seven classes, happy, sad, fear, angry, disgust, neutral and surprise.

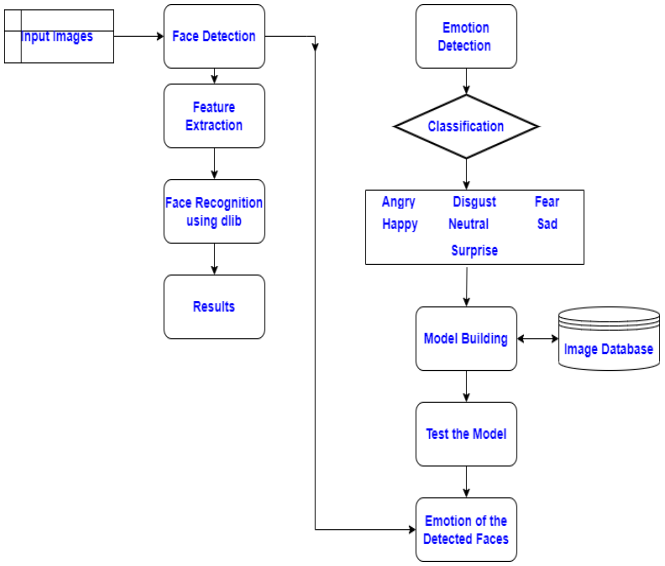

Fig. 1. Proposed Methodology.

As shown in the Fig. 1, the proposed methodology contains three pipeline stages.

1) *Face Detection*
2) *Face Recognition*
3) *Emotion Detection*

C. Face Detection

Detecting the faces in a given image is the first stage of the pipeline. Face recognition is the python library that is used for facial detection. This python library scans any given input image and outputs the bounding box coordinates of the faces being detected in the given input image. Face detection is special case of object detection technology in which human faces are detected from given input image digitally.

D. Face Recognition

Facial recognition is done using face-recognition library which manipulates the faces from input image. Face-recognition library is built using dlib along with deep learning. Through command line options, facial recognition can be done and that is achieved through dlib's face-recognition library. Facial detection is done using face-recognition library which return bounding box coordinates as output. Then find the required features from given input image and manipulate the features of the face from the given image. That is face recognition. Face landmarks will return person's eyes, chin, nose and mouth. And this is how any makeup related options in the camera works.

Using encoding and based on the images trained or the dataset used, who appears in the image can be recognized using face-recognition library. Using face-recognition library, one can verify any given image and recognize who appears in the image digitally. Create two folders, one with training the model with all the images required. And the other folder which contains all the images present in the training dataset but in a different angle. Then on unseen data if the input image resembles any of the trained images, then the output will be, the model verifies the image and recognizes the face. Using face-recognition library the proposed model outputs Boolean value if the face matches, then returns true else false.

E. Emotion Detection

Human emotions can be recognized in multiple ways that is through voice, action, facial expressions, and body language and so on. Facial expressions for each emotion are innate and they are not any other part of cultural learning. But making the computer to do so is a challenging task. Facial expressions can be categorized as a form of nonverbal communication. For any given input image face detection and recognition is part of object detection and recognition from any given input images. Humans have tendency to express their emotions through facial expressions at most of the times.

The emotions can be classified into seven different classes apparently. That is happy, sad, anger, disgust, fear, neutral and surprise. The model is built using six layers of convolutional neural network as shown in the below Table I.

TABLE I. SUMMARIZING LAYERS OF CONVOLUTIONAL NEURAL NETWORK

Layer (type)	Output Shape	Parameters
conv2d (Conv2D)	(None, 46, 46, 32)	320
conv2d_1 (Conv2D)	(None, 44, 44, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 22, 22, 64)	0
conv2d_2 (Conv2D)	(None, 20, 20, 128)	73856
max_pooling2d_1 (MaxPooling2	(None, 10, 10, 128)	0
conv2d_3 (Conv2D)	(None, 8, 8, 128)	147584
max_pooling2d_2 (MaxPooling2	(None, 4, 4, 128)	0
conv2d_4 (Conv2D)	(None, 4, 4, 7)	903
conv2d_5 (Conv2D)	(None, 1, 1, 7)	791
flatten (Flatten)	(None, 7)	0
activation (Activation)	(None, 7)	0

Total params: 241,950
Trainable params: 241,950
Non-trainable params: 0

F. Convolutional Neural Network

Fully connected convolution neural network consists of four inputs that is input layer, and one hidden layer with five neurons and an output neuron in output layer. In fully connected neural network each neuron of the input layer is connected to all the neurons of hidden layer. Then all the computations are performed at hidden layer and the output is presented at the output layer. When any input image is fed into the convolutional layer, apply filters or choose proper parameters along with strides or padding required. Perform the operation that is convolutional on the input. For non-linearity apply Relu Activation function to the convolution operation. Pooling is performed to reduce the dimensionality. Until the error rate is reduced convolution layers are added. Then this flattened output is fed to fully connected convolution layer.

Then model will output based on the activation function used and then object detection and their classification are done.

Artificial Intelligence helps to view the world through computer vision as human does. It is nothing but bridging the gap between humans and computers. It helps in analyzing the world as humans do and perceiving the inputs in the similar fashion how human brain works. Convolutional neural network is a deep learning algorithm which can take input, assign the weights based on the objects in the image and identify these objects or different objects can be identified. Image pre-processing required is very lower when compared to other algorithms available for classification or recognition of objects. Using convolutional neural network, the model can identify any filters applied in the image without any training and they have the ability to learn these filters or characteristics of the image.

To obtain maximum values of the image portion covered by the kernel, max pooling is performed. Max pooling reduces the noise in the input along with the dimensionality reduction. Convolutional neural network will have these layers increased if the input dataset contains more complex images or to capture more low-level detailed information of the image. The output of the final layer can be fed as input for other neural networks, flatten the image which is transposing the row to column vector. Then backpropagation algorithm can be applied to get the network for each training iteration. Then after several series of epochs, model will learn through experience and would provide very accurate results.

G. Data Augmentation

Accuracy prediction for any supervised model largely depends on diversity of data and amount of data available during training. Success of any deep learning model depends on the relationship between deep learning models and the higher amount of data required. Data augmentation is a way of dealing with the problem of limited dataset. Image augmentation is nothing but synthesizing new data from the data available by applying various transformations.

Some of the simple geometric transformations applied on the images are flipping, cropping, scaling, rotating and color casting or pixel transformation that is color space transformations. The proposed model uses geometric transformation because of the positional biases present in the dataset. Color transformation helps in dealing with the problems related to lighting or illumination of the given input image. The proposed model uses keras image data generator augmentation.

As shown in the above Fig. 2, the workflow of the proposed methodology for any given input image, face detection and face recognition are done using face recognition library which python offers with upgraded version. For emotion detection, there are seven classes of emotions and the faces being detected from the image belongs to which class that needs to be detected. Dataset is downloaded from kaggle that is facial emotion recognition dataset. Once dataset is loaded, divide the number of classes of emotions. The data augmentation is done to improve the performance of the model.

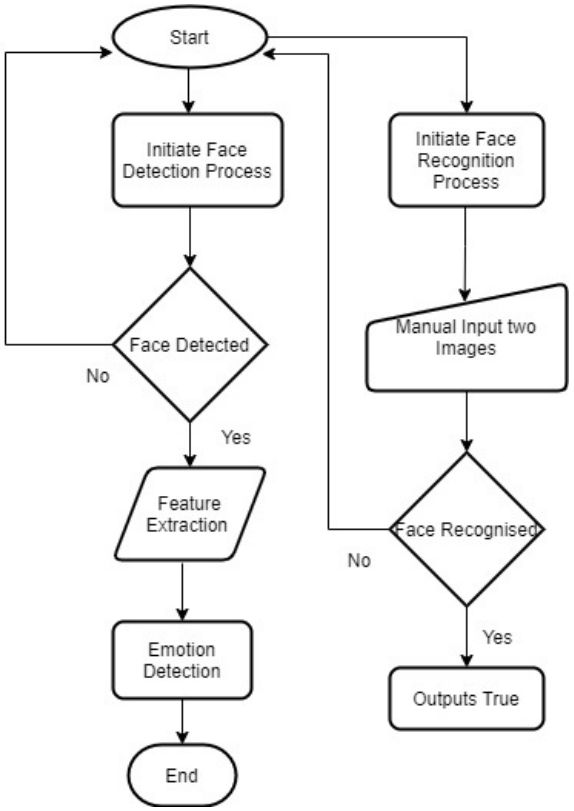

Fig. 2. Flow Chart of the Proposed Methodology.

IV. RESULTS

Face detection is done using face-recognition library. As shown in the Fig. 3 the output of face detection using dlib's face-recognition library. Given input image has three human faces, and the output is the bounding box coordinates of all the three images present in the input image.

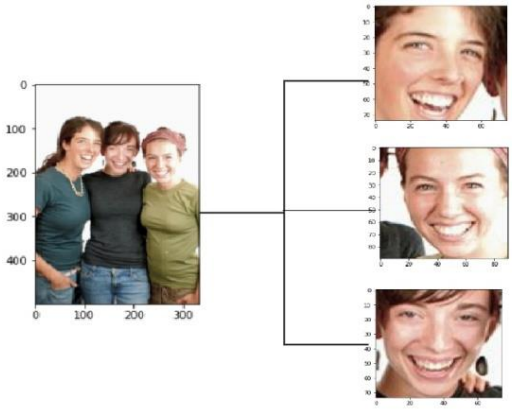

Fig. 3. Face Detection.

Face recognition is done is using face-recognition library. The output of the face recognition is shown in the below Fig. 4. As shown in Fig. 4, face recognition is nothing but verifying or comparing any given two input images and if two matches then output the Boolean value true or else false. In the below example, took two input images of the same actress and the output was true.

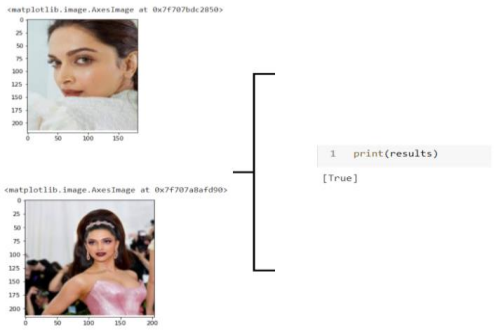

Fig. 4. Face Recognition.

Each time when validation accuracy is increased, model was saved. Calculated model loss for each epoch and plotted graph against the loss occurred vs. validation loss occurred in each epoch as shown in the Fig. 5.

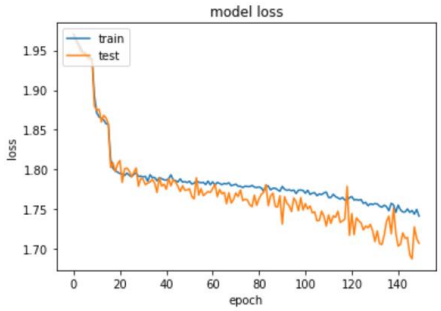

Fig. 5. Model Loss.

Model was trained using facial emotion recognition dataset. There were 28709 training samples which belong to seven different classes of emotions. And 3589 images were considered as validation samples and 3589 images were considered as testing samples. Facial emotion dataset was downloaded through kaggle. There are many open websites and from where image dataset was downloaded. Then image dataset was generated from .csv format into the specific folder. Then the dataset was used for training, validating the model and testing the model on unseen data.

As shown in the below figure, model performance was calculated by plotting confusion matrix for seven different classes of emotions considered for all validations samples considered through validation generator. The accuracy was achieved up to 85% when evaluated the model using confusion matrix as shown in the Fig. 6. Confusion matrix is actual a 2x2 matrix with actual data on one axis and predicted data on the other. Precision and recall are the important terms for performance evaluation, since positive class was considered the most and all other irrelevant or unwanted weeds are not considered while evaluating the performance of the model as shown in Fig. 7.

$$Precision = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (1)$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive+False Negative}} \quad (2)$$

F1-score is the mean of both precision and recall considered. And it takes both false positive and negatives and performs well on dataset which is imbalanced.

$$F1\ score = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}}$$

The output of the emotion detection is shown in the Fig. 8. The trained model is loaded and emotion is predicted for each face being detected.

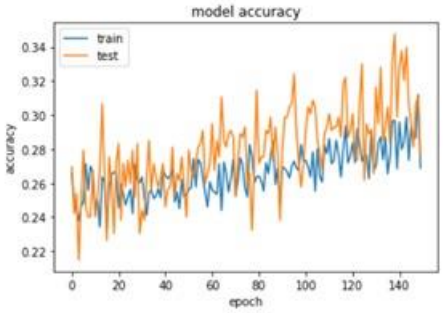

Fig. 6. Model Accuracy.

```
Confusion Matrix
[[331   1  39   8  56  52   4]
 [ 18  30   3   1   1   2   0]
 [ 88   2 243  12  59  94  30]
 [ 28   1  16 754  45  19  16]
 [114   8  79 148 113 105  59]
 [ 56   1  57  16 148 307   9]
 [ 12   0  67  19  18   6 294]]
Classification Report
              precision    recall  f1-score   support

       Angry       0.51      0.67      0.58       491
     Disgust       0.70      0.55      0.61        55
        Fear       0.48      0.46      0.47       528
       Happy       0.79      0.86      0.82       879
     Neutral       0.26      0.18      0.21       626
         Sad       0.52      0.52      0.52       594
    Surprise       0.71      0.71      0.71       416

avg / total       0.56      0.58      0.56      3589
```

Fig. 7. Performance Evaluation of Model.

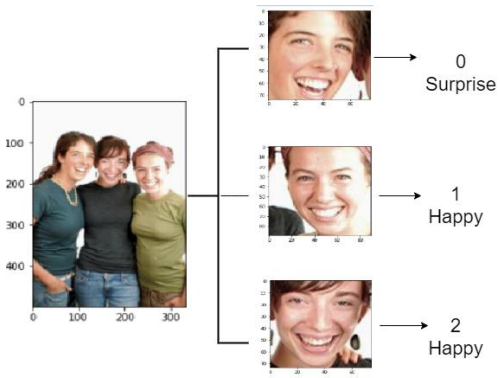

Fig. 8. Emotion Detection.

V. Conclusion

The proposed model is used for recognizing facial feelings using computer vision along with AI calculations which characterize the given input facial image into eight distinct feelings. On implementation using multiple algorithms and convolutional neural network, CNN outperforms with best

accuracy for image dataset. Model results suggest that computer can have human like capacity to recognize the feelings, at least for applications in which front facing perspectives can be expected utilizing the webcam. Proposed model performs well on classifying the given image into seven different classes of emotions irrespective of number of faces present in the image.

The proposed model consists of six convolutional layers attaining the accuracy of almost 85%. The model uses softmax activation function in the last layer of a fully connected convolutional neural network which makes the model more unique and more efficient in terms of performance. Visualization of the model is done based on the parameters, layers used and results obtained are also shown in terms of graphs and plotted confusion matrix as well for better understanding.

Emotion recognition will be exceptionally valuable soon in the exploration field of advanced mechanics and man-made reasoning for instance if a robot can detect the feelings of any human and that robot can act appropriately without any human aid. This model can be further extended for emotion recognition, can likewise be extended to the identification of different feelings other than these eight all-inclusive feelings.

ACKNOWLEDGMENT

We thank M S Ramaiah Institute of Technology for their continuous support in providing us the required assistance in conducting research.

REFERENCES

[1] Anithadevi, N., Gokul, P., Nandan, S. M., Magesh, R., & Shiddharth, S. (2020). Automated Speech Recognition System For Speaker Emotion Classification. 2020 5th International Conference on Computing, Communication and Security (ICCCS). Patna, India: IEEE.

[2] Chavan, S., Malkan, E., Bhatt, D., & Paranjpe, P. (2014). XBeats-An Emotion Based Music Player. International Journal For Advance Research In Engineering And Technology.

[3] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE.

[4] Ekman, P. F. (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 124–129.

[5] Engoor, S., SendhilKumar, S., Sharon, C. H., & Mahalakshmi, G. (2020). Occlusion-aware Dynamic Human Emotion Recognition Using Landmark Detection. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE.

[6] Fratesi, A. (2015). Automated Real Time Emotion Recognition Using Facial Expression Analysis. Carleton University Research Virtual Environment.

[7] Gupta, S. (2018). Facial emotion recognition in real-time and static images. 2018 2nd International Conference on Inventive Systems and Control (ICISC), DOI: 10.1109/ICISC.2018.8398861, 2018. IEEE.

[8] Iyer, A. V., Pasad, V., Sankhe, S. R., & Prajapati, K. (2017). Emotion based mood enhancing music recommendation. 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). Bangalore, India: IEEE.

[9] J. Nicholson, K. T. (2000). Emotion Recognition in Speech Using Neural Networks. Neural Computing & Applications, 290–296.

[10] Jaiswal, A. (2020). Facial Emotion Detection Using Deep Learning. 2020 International Conference for Emerging Technology (INCET), DOI: 10.1109/INCET49848.2020.9154121.

[11] Kauser, N., & Sharma, J. (2016). Automatic facial expression recognition: A survey based on feature extraction and classification techniques. International Conference on ICT in Business Industry & Government (ICTBIG). Indore , India: IEEE.

[12] Kolodziej, M., Majkowski, A., Rak, R. J., Tarnowski, P., & Pielaszkiewicz, T. (2018). Analysis of Facial Features for the Use of Emotion Recognition. 19th International Conference Computational Problems of Electrical Engineering. IEEE.

[13] Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACMVolume 60Issue 6, 84–90.

[14] Li, S., & Deng, W. (2020). Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, IEEE.

[15] Lv, Y., Feng, Z., & Xu, C. (2014). Facial expression recognition via deep learning. 2014 International Conference on Smart Computing. Hong Kong, China: IEEE.

[16] M. Mudrov´a, A. P. (2004). Principal Component Analysis in Image Processing . Institute of Chemical Technology, Prague, Department of Computing and Control Engineering.

[17] Mishra, P. (2018). HMM Based Emotion Detection in Games. 3rd International Conference for Convergence in Technology (I2CT) 2018. Pune, India: IEEE.

[18] P. Lucey, J. C. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE.

[19] Patel, T. B., & Patel, J. T. (2017). Occlusion Detection and Recognizing Human Face Using Neural Networks. 2017 International Conference on Intelligent Computing and Control (I2C2 2017) (pp. 904-908). Coimbatore, Indai: IEEE.

[20] Pathar, R., & Adivarekar, A. (2019). Human Emotion Recognition using Convolutional Neural Network in Real Time. 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), DOI: 10.1109/ICIICT1.2019.8741491.

[21] Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. (2020). Facial Emotion Recognition Using Deep Convolutional Neural Network. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). Coimbatore, India: IEEE.

[22] Rabhi, Y., Mrabet, M., Fnaiech, F., & Sayadi, M. (2018). A real-time emotion recognition system for disabled persons. 2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Sousse, Tunisia: IEEE.

[23] Reney, D., & Tripathi, N. (2015). An Efficient Method to Face and Emotion Detection. 2015 Fifth International Conference on Communication Systems and Network Technologies, DOI: 10.1109/CSNT.2015.155.

[24] Saaidia, M., Zermi, N., & Ramdani, M. (2014). Facial Expression Recognition Using Neural Network Trained with Zernike Moments. 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology. IEEE.

[25] Sergeeva, A. D., Savin, A. V., Sablina, V. A., & Melnik, O. V. (2019). Emotion Recognition from Micro-Expressions: Search for the Face and Eyes. 2019 8th Mediterranean Conference on Embedded Computing (MECO). IEEE.

[26] Tian, Y.-I., Kanade, T., & Cohn, J. (2001). Recognizing action units for facial expression analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[27] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. Proceedings of the IEEE, 1550 - 1560.

[28] Castellano G, De Carolis B, Macchiarulo N (2021) Automatic emotion recognition from facial expressions when wearing a mask. In: CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter. CHItaly '21. Association for Computing Machinery. 10.1145/3464385.3464730.

[29] Castellano G, De Carolis B, Macchiarulo N. Automatic facial emotion recognition at the COVID-19 pandemic time(2022). Multimed Tools Appl. 2022 Oct 22:1-19. doi: 10.1007/s11042-022-14050-0. Epub ahead of print. PMID: 36313484; PMCID: PMC9589699.