World Journal of Engineering Research and Technology
WJERT
www.wjert.org
SJIF Impact Factor: 5.924

REAL TIME FACE AND EMOTION RECOGNITION USING CNN

Nazmin Begum*[1], Dr. Md Shoaibuddin Madni[2] and Dr. Ismath Unnisa[3]

[1]Assistant Professor, Department of Computer Science and Engineering, PES University, Bangalore, India.

[2]Associate Professor, Department of Electronics and Communication Engineering, Navodaya Institute of Technology, Raichur.

[3]Solution Architect- Java, Content Development Department, IIHT-TECHADEMY, Bangalore, India.

Article Received on 09/01/2023 **Article Revised on 30/01/2023** **Article Accepted on 20/02/2023**

***Corresponding Author**
Nazmin Begum
Assistant Professor, Department of Computer Science and Engineering, PES University, Bangalore, India.

ABSTRACT

One of the most exciting areas of research, facial expression recognition, has attracted the attention of many academics for many years. An emotion recognition system based on the concept of convolutional neural networks is introduced in our paper (CNN). Prior to convolution, the image of the facial expression is first normalised, and then the edges of each layer are recovered from the image. Each of the feature images has the retrieved edge information applied to it in order to preserve the edge structure information of the picture's texture. The max pooling method was employed in an effort to reduce the dimensionality of the recovered implicit features. A SoftMax classifier is then used to classify and identify the expression of the test sample image. In order to recognise face expressions, this research aims to learn and identify information representations from 2-dimensional gray-scale pictures. The learned features are provided via a constructed convolutional neural network (CNN). The developed CNN model facilitates fast learning of information from images by cascading different layers together. Because it does not contain a high number of layers and handles the overfitting problem at the same time, the developed model is computationally efficient. Using different datasets such as FER-2013, CK+, and image datasets, our suggested approach assists us in focusing on crucial aspects of human faces to detect emotion.

Index Terms: Convolutional Neural Networks, Deep Learning, Emotion Recognition, Fer2013, CK+, Machine Learning.

I. INTRODUCTION

Humans are one of the most intelligent beings who have an extraordinary ability to detect and identify other people's emotions. Humans are able to differentiate between various emotions, this skill plays an important role in communications and interactions which happens on daily basis. Identifying human expressions and sending those results to the computer systems is one of the main goals of emotion recognition systems. Human connection with such systems can become friendlier as a result of feedback from computer systems. Human emotion identification plays a critical role that has been used in areas as diverse as music recommendation, mental health, and education. Human expressions aim in the decision-making process and communication with others. Human expression recognition plays an important role in various research field such as psychology and social sciences. Facial expressions and bodily gestures are all part of non- verbal communication. It is taken into account while recognizing expressions that the main aim of this paper to propose a model which achieves higher accuracy rate for facial expression identification through facial images or in real time video. The model proposed here accepts images as input and this is passed to our CNN model which will further classify the images resulting in the output, the given output is in the form of text. In our model, we can recognize a facial expression based on the different emotion categories mainly angry, neutral, happy, sad, surprise, confident, confused, contempt, fear, sleepy, shy, crying and disgust. The process of emotion identification is happening in following threesteps:

1) Taking the images and feeding them into the CNN model
2) Get resulted output in the form of text
3) Average the results with the prior output and re-output.

Real- time facial recognition is performed by combining CNN architecture and the webcam in this fashion. Furthermore, when compared to results produced using merely a convolution neural network, the suggested strategy significantly improves the overall resilience and accuracy of facial emotion identification. In this model, we have used datasets such as FER2013, CK+ and a customized dataset to train and test the model.

II. RELATED WORK

A Deep Convolutional Neural Network (DCNN) model that primarily recognises five distinct

facial emotions was created by Pranav. E et al.[1] The manually collected image dataset is being used to train and test this model. Adam Optimizer is being used, which lowers the loss function and results in an accuracy of 78.04%. Additional applications for this work include video sequences.

Bilal T *et al*[2] have extracted two features from grayscale images namely Laplacian of Gaussian (LoG) and LBP. After extraction it would be passed to SVM classifier which is going to handle large number of features during optimization. These features will be compared in the CNN model. They have made use of Bosphours database for conducting the experiment. Overall the model was capable of giving better accuracy about 88.2%. The extraction of facial features was quite simple in emotions such as happy class or sad class, but the accuracy was lower when it comes to correlation between these emotional classes such as Anger class with the Fear class and Sad class.

M.A. Ozdemir *et al.*[3] have merged three datasets (JAFFE, KDEF and their own custom dataset) to train CNN model. Here they have used LeNet architecture for emotion classification which in turn resulted in accuracy of 91.81%. Also they have used Haar Cascade library which helped in removing the effect of unimportant pixels due to which training time and number of networks were reduced.

S. Begaj *et al.*[4] have used iCV MEFED (Multi-Emotion Facial Expression Dataset) as their main dataset in their CNN model. It is a relatively new dataset and quite a challenging one, hence, giving 74.3% accuracy. Certain emotions such as contempt and sadness were underpredicted, whereas angry, disgust, surprise and fear were overpredicted. The authors suggest using the same dataset but trying a different approach into consideration.

L. Zhang *et al.*[5] have proposed a model which predicts only infant expressions, they have used IFER as well as self-built data. Since their own-built dataset was quite small, the images were extracted from the horizontal and vertical axes by enhancing the LBP and Sobel edge detection operators. This step was crucial because it allowed the network to efficiently solve the problem of the baby's facial contour and obtain the most striking portion from each image. Using a combination of custom features and a multistream CNN fusion network, the model had an accuracy of 91.67%.

A. Ghofrani et al.[6] 's usage of MTCNN (Multi-Task Convolutional Neural Network) and

ShuffleNet V2 architecture, which allows for a trade-off between accuracy and speed of the running CNN model, allowed them to determine the boundaries of the face with fewer residual margins. The combination of these two yields a real-time facial expression prediction model with an accuracy of 71.19%.

R. Subramanian *et al*[7] have created a CNN model to detect emotions in real time by using Keras, OpenCV and TensorFlow. They have made use of the FER2013 dataset. By keeping the SoftMax layer as the model's final layer has led to more accuracy. The model has attained an overall test accuracyof 89% in detection and classification of emotion in real-time.

A. Kumar *et al.*[8] compared their CNN based model with other machine learning classification algorithm such as Support Vector Machine (SVM), K Nearest Neighbours (KNN) and Naive Bayes classifier by evaluating them with the presence of various kinds of noises such as Gaussian Noise and Salt and Pepper Noise. This paper showed how CNN model is able to extract various facial features from the images with its performance not being affected significantly by the presence of large amount of noise. Whereas other ML classification models failed to recognise the emotions due to the presence of noise.

K.C Liu *et al.*[9] has proposed Average Weighting Method to reduce the potential errors while capturing real-time facial expression for recognition. External factors such as influence of light can change the characteristic of image while capturing in real time, so this problem is being overcome by this model. The method refers to previously taken images, later averages the weights between the prior and the current recognition, based on this the errors are reduced, hence making the CNN architecturemore robust.

III. METHODS

When a network is designed in order to identify and manage multi-dimensional data like picture or image or frames of images in such a case, we use the concept of convolutional neural network (CNN). The main idea of CNN is for extracting important features and properties, apart from this it can also be used in computation or finding of weights and biases at the time of training.

This neural network model is one of the common methods used for analyzing images and frames of images from video or from a live image stream. It provides a different perspective as compared to other models as it has hidden layers called convolutional layers.

The proposed method also uses CNN for producing the results; first the data set is cleaned and also, we get the images from the dataset, which is passed on to the model for the training process. The training process is explained in detail below. After this the model training is completed for the thirteen-emotion set and the testing process can be carried on in this case which can be used for getting the accuracy, thereby providing an insight for scope of improvement in the model. The compilation of these two processes then we can obtain the results using one of the two models based on the requirement, the options for both are provided.

A. Live webcam and image upload

In this case the image input is given from the web-camera then the image will be fed to the model that has been trained. This model forms a box around the face of the person for better feature reading. The model has many hidden layers. The features that are taken into account are eyes, cheek, mouth, eyebrows and nose.

In case of the image upload the size of image upload is restricted to 48 x 48, the training set used here is in greyscale and of the same size as the upload. The images that are uploaded can be of grayscale or RGB format but with restriction on the resolution. The box or boundary are not used in case of the image upload. Here also the image emotions are predicted using the same features and have the same number of hidden layers as the previous model. The predicted result will be displayed on the web page for the convenience of the user.

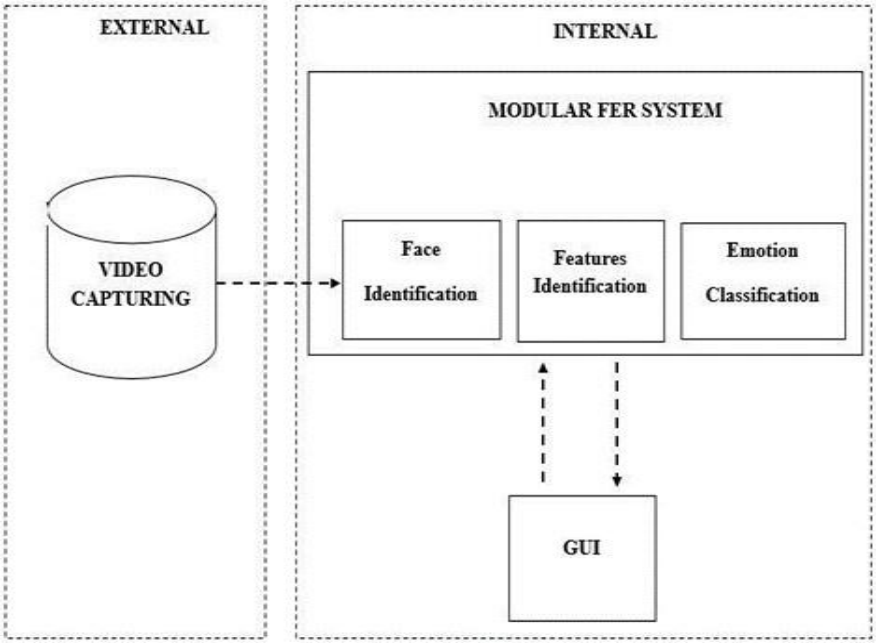

Fig. 1: System Architecture.

B. Extraction of face

A person's face is captured using a webcam PC or external webcam. In that living smoke the face is removed and all other unwanted parts are not considered. So we have achieved this efficiency and understanding that we have used the OpenCV library.

C. pre-processing

Common name for activities with pictures in very low output rate for both input and output are available image stabilization. The purpose of the preliminary processing enhancement of unwanted image data distorts or enhances other important aspects of the input for processing.

i. Face detection and retrieval
ii. Conversion of grayscale
iii. Normalization for the image
iv. Augmentation operations like rotation, flip, shear, zoom etc.

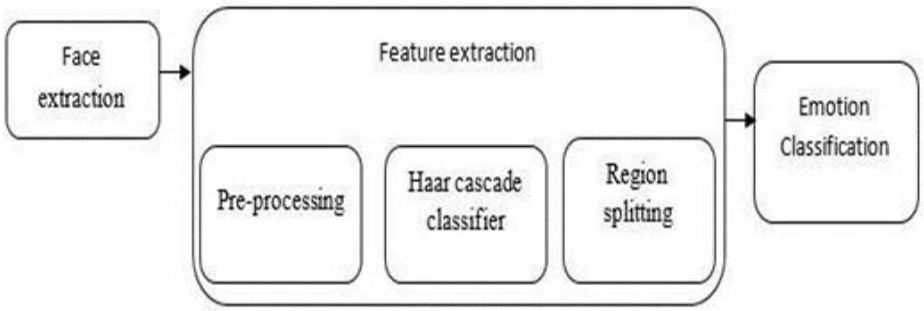

Fig. 2: Modules used.

D. Stratified K-Fold

Stratified k-fold cross-validation is the same as just k-fold cross- validation, But Stratified k-fold cross-validation does stratified sampling instead of random sampling. The idea of stratified sampling is chosen over random sampling in machine learning and deep learning models because it considers the imbalance in classification and thereby making predictions more accurate.

E. Haar Cascade classifier

HaarCascade is a used separator. HaarCascade is repaired with a top installation with a good image over many negative images. I training is usually done on the server and in separate categories. Better results are needed using a higher note image and increasing the amount of sections to do filter is adjusted. The HaarCascade classifier depends on the HaarCascade wavelet process for pixel analysis on its image into squares with function. This makes use of

entire photo thoughts for separate outstanding registrations. HaarCascades use Ada-help mathematical learning, choosing a few highlights from the main set to give the effect of the separator and use cascading techniques to get a face in a photo.

F. Splitting of Regions

To get emotional attention, the main area of the face below considerations are eyebrows and mouth. And the division of the mouth and eyebrows are named as regional divisions. These two will enable the model to be trained in a way to identify the expression by extraction of the above-mentionedfeatures.

G. Classification of emotions

After a small feature removal function has been completed, human reactions are produced simultaneously with their percentage rate. The flow of the model process is shown in four main steps starting with database preparation and model validation. These steps are described as A, B, C, and D.

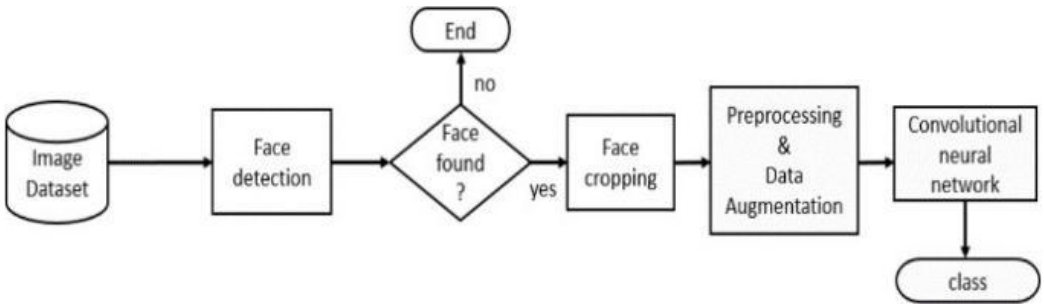

Fig. 3: Work flow model of classification.

Data set information

i) ***FER2013 dataset***

Fer2013 contains human facial images more than 30,000 with size restricted to 48×48. Emotion classes are mainly labelled into 7 types: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. While labels have nearly 5,000 samples each, the disgust expression has the minimal number of imageswhich is below 600.

ii) ***CK+ dataset***

The Extended Cohn-Kanade (CK+) dataset contains about 593 video sequences. Each video sequence is recorded at 30 frames per second (FPS) with a restricted resolution of 640x490 pixelsand depicts a shift in face expression from neutral to a targetedexpression.

iii) ***Custom dataset***

For each expression, there are over 400 images in the dataset. We have created dataset for various facial expressions (such as happy, fear, confused, confident, sad, disgust, contempt, angry,surprised, sleepy).

Construction of CNN model

With batch normalisation, the CNN model has two compatible layers. The output from this is then sent as inputs to other collections, one of which is managed by a convolutional layer and the other of which is treated with a visible convolutional layer. Following the conclusion of a SoftMax process, the findings from these layers are combined with size and earth scale compilation.

i) Convolutional 2D

When you start with a kernel that contains the weight matrix for all the smaller ones, convolution in two dimensions is simple. Each time the data enters, this kernel slides in one step to build a matrix multiplication by the matrix element, then condenses the entire output into a single output pixel. This uses the zero padding approach to produce an output that is the same size as the input image.

ii) Batch Normalization and Max 2D integration

Convolutional neural networks' performance, speed, and stability are all enhanced by the batch normalising technique, which lowers internal covariance changes in neural networks. The greatest value in each region is taken and stored in the corresponding area of the outgoing matrix when combining n characters of size n X n.

iii) Max Pooling

In a pooling process called max pooling, the filter-covered area of the feature map is used to select the largest element. As a result, the output following the max pooling layer would be a feature map that included the most noticeable features from the prior feature map.

iv) SoftMax

The SoftMax function assumes the actual N vector numbers as input and makes that vector normal values from 0 to 1 This willgenerally be placed as a last layer for output prediction.

v) Activation function

Reducing the overuse of data usage activities are used. In this model we used the ReLu function. The main role of ReLu is gradient is to continually same equal to 1. Negative values for matrix input are always transformed into zero and all the other good one's prices remain unchanged. F (x) = max (0, x)

Optimizer, Loss Function and Metric

The loss function is used to measure the total difference between our forecasts and the actual value available in the verification database. Loss function used in this model cross entropy phase. Cross-entropy loss shows the operation of an input output model with a number of chances between 0 and 1. Loss of cross-entropy the amount of work output varies as predicted opportunities differ from the actual output. Enhancements are used to minimize job loss by reviewing Neural network attribute values. Optimizer used in our area model Adam Optimizer. Adam represented Flexible Time Rate. Adaptive Moment Measurements are used to calculate variablelearning levels for each attribute.

The model

We have created two custom models one having (48,48,3) input and the other having (48,48,1) input. In the other words the model that is used for input has a depth of 3 whereas the model that is used for the webcam has a depth of 1 and the input image size in both cases is 48X48 pixels. The custom model is a sequential model has Conv2D, Batch normalization, Max-pooling and Dropout layers. There are 13 classes of emotions that are being classified with the batch size of 512 and input image dimensions of 48 x 48. Apart from these layers there are two dense and one flattened layer used in the custom model. In first dense layer the activation function used here is a ReLu activation function (Exponential linear unit) this aims to provide faster and accurate results, the final dense layer is the activation layer and the activation function used is a SoftMax activation function. We have used a 10-fold stratified function along with 50 epochs to provide the best possible results. The model is compiled with a categorical cross entropy loss function that is mainly used in case of multiclass classification. It uses the Adam optimizer which intends to provide high speed computation along with the advantage that it requires fewer parameters fortuning. The model is also built in such a way thatit provides the accuracy matrices.

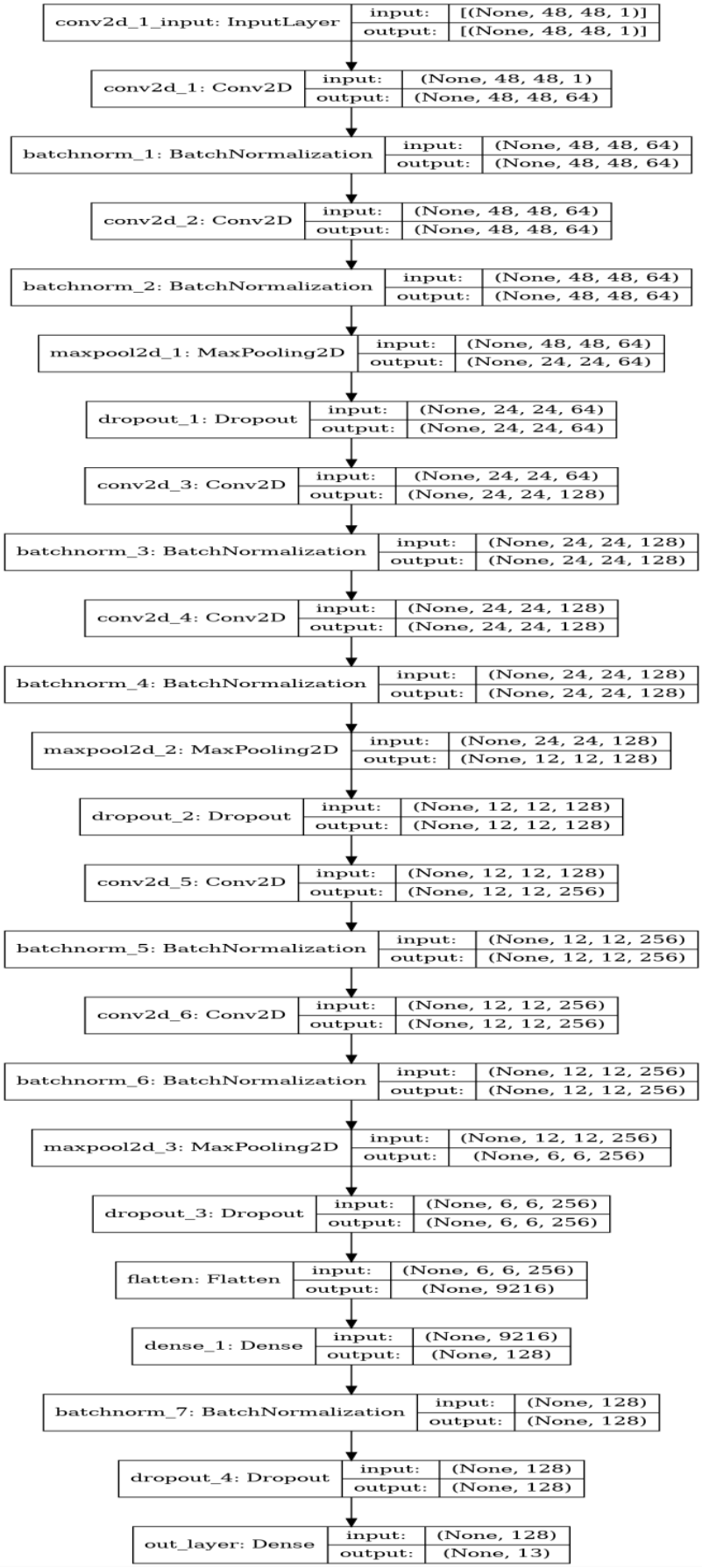

Fig. 4: Summary of the CNN model.

Model training

After all the steps of data collection, cleaned once which were processed previously needs to be added to the proposed CNN model. To train the following model, two steps are performed on encoded data:

1) Separating data by features and guidelines: First, training Our model requires a set of data training. Train data contains a complete set of flexible features (independent variable) and target variable (dependent variable). So, we need to distinguish all the features used for prediction targeted variables in our database. The class label is the only targeted variable predicted for all variables.
2) Separating data on a set of training and assessment: Where there are many ways to measure the separation of a train, validate and test data. In this paper the data set is divided into 8:1:1, 80% for training, 10% for validation and 10% for testing.

IV. RESULTS

The proposed model is trained in integrated databases, it has been successful in gaining 97.73% validation accuracy. This model succeeded in saving it a high level of recognition that is almost equal in each class as it can distinguish geometrically removed facial images. However, there is a slight variation in the level of recognition between thirteen classes, still better compared to the existing models. Proposed model, a convolutional neural network with data augmentation, has always been successfully achieving 97.73% verification accuracy i.e. the highest accuracy in our experiment of facial emotion recognition. This model has succeeded in keeping it high and almost equal the level of recognition of each class as it can differentiate.

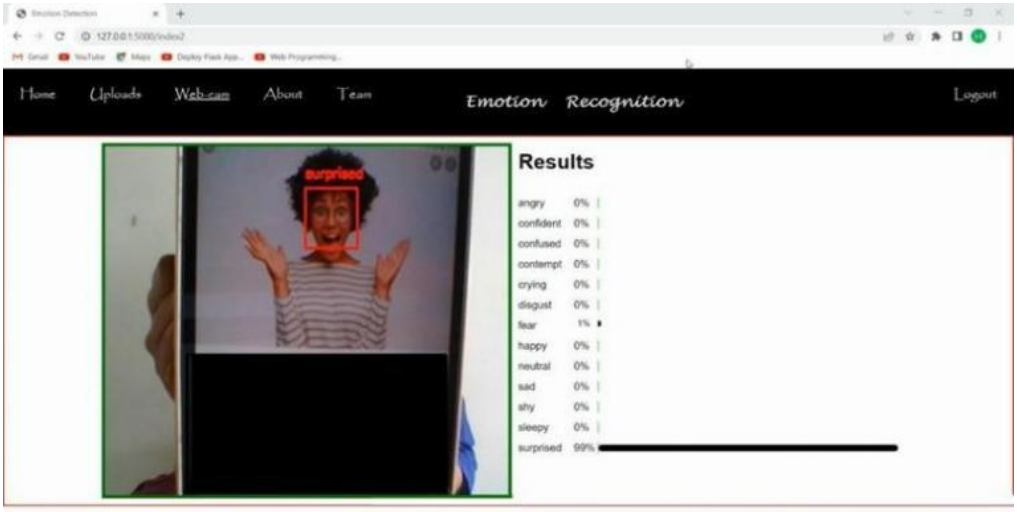

Fig. 5: Depicting surprised emotion in real time.

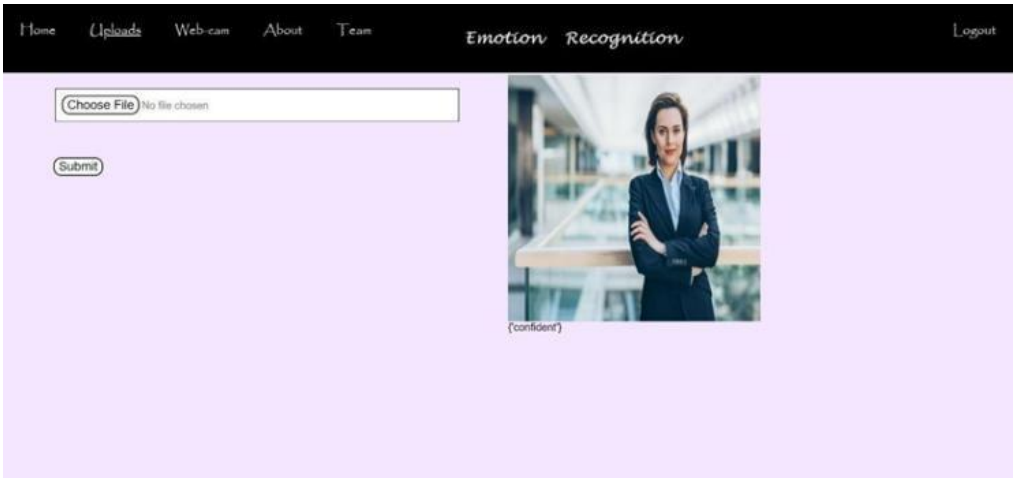

Fig. 6: Depicting confident emotion for image upload.

In the Fig 7 we can see that the accuracy is predicted for both testing and training dataset. It is observed that the model accuracy of the training data set is more as compared to the testing dataset.

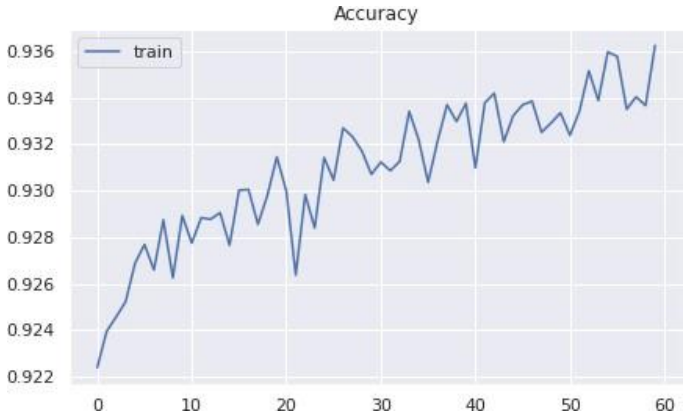

Fig. 7: The plot shows the training and testing accuracy of model for 60 epoch.

In the Fig 8 we can see that the loss is predicted for both testing and training dataset. It is observed that the model loss of the training data set is less as compared to the testing dataset.

Fig. 8: The plot shows the training and testing loss for 60 epoch.

The below Fig 9 represents the confusion matrix of the given model that classifies emotions in 13 classes namely class 1 as angry, class 2 as confident, class 3 as confused, class 4 as contempt, class 5 as crying, class 6 as disgust, class 7 as fear, class 8 as happy, class 9 as neutral, class 10 as sad, class 11 as shy, class 12 as sleepy and class 13 as surprised.

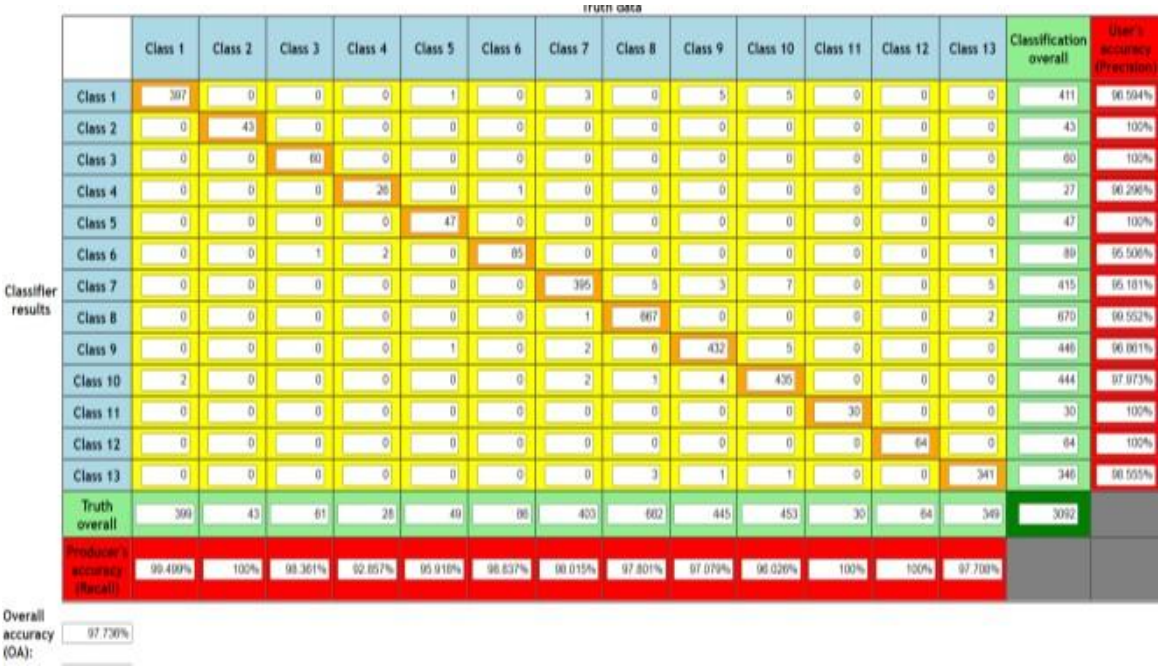

Fig. 9: Confusion Matrix.

V. CONCLUSION

The proposed model has obtained higher verification accuracy than any other available model. No conflicts between groups as our model fits well with the data. However, plan to work on a complex type or mixed group of emotions, such as a happy shock, surprised by frustration, dissatisfaction with anger, surprise with grief, and so on. As illustrated in our Confusion Matrix Fig 9, the system performed the best when it came to detecting confident, shy and sleepy emotions. The work done by us improved the model's performance, allowing it to distinguish distinct emotions in various lighting and postures. We began by developing and training the model so that it is capable of recognizing and classifying facial expression in realtime. The model was accurate; however, it took a long time to train the data.

Future research includes exploring the many types of human variables such as personality traits, age, and gender that affect the functioning of emotional detection. The increasing availability of large medical data has necessitated the use of machine learning methods revealing hidden health care patterns. In addition, the focus on the psychological or emotional

factors of the man or woman who serve as a mentor as well he leads the situation depending on the moral. Apart from this, we will try to refine the model more so accurately that a natural way to recognize facial expressions can be given.

VI. REFERENCE

1. Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. H. (2020, March). Facial emotion recognition using deep convolutional neural network. In *2020 6th International conference on advanced computing and communication Systems (ICACCS)*, (317-320). IEEE.
2. Taha, B., & Hatzinakos, D. (2019, May). Emotion recognition from 2D facial expressions. In *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, (1-4). IEEE.
3. Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R., & Akan, A. (2019, October). Real time emotion recognition from facial expressions using CNN architecture. In *2019 medical technologies congress (tiptekno)*, (1-4). IEEE.
4. Begaj, S., Topal, A. O., & Ali, M. (2020, December). Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN). In *2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)* (58-63).IEEE.
5. Zhang, L., Xu, C., & Li, S. (2020, October). Facial Expression Recognition of Infants Based on Multi-Stream CNN Fusion Network. In *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)* (37-41). IEEE.
6. Ghofrani, A., Toroghi, R. M., & Ghanbari, S. (2019). Realtime face-detection and emotion recognition using mtcnn and minishufflenet v2. In *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)* (817-821). IEEE.
7. Subramanian, R. R., Niharika, C. S., Rani, D. U., Pavani, P., & Syamala, K. P. L. (2021, May). Design and Evaluation of a Deep Learning Algorithm for Emotion Recognition. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, (984-988). IEEE.
8. Kumar, A., Jani, K., Jishu, A. K., Patel, H., Sharma, A. K., & Khare, M. (2020, November). Evaluation of Deep Learning Based Human Expression Recognition on Noisy Images. In *2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI)* (pp. 187-191). IEEE.
9. Liu, K. C., Hsu, C. C., Wang, W. Y., & Chiang, H. H. (2019, July). Real-time facial

expression recognition based on CNN. In *2019 International Conference on System Science and Engineering (ICSSE)* (120-123). IEEE.

10. Verma, A., Singh, P., & Alex, J. S. R. (2019, June). Modified convolutional neural network architecture analysis for facial emotion recognition. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)* (169- 173). IEEE.
11. Pathar, R., Adivarekar, A., Mishra, A., & Deshmukh, A. (2019, April). Human emotion recognition using convolutional neural networks in real-time. In *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)* (1-7). IEEE.
12. Samadiani, N., Huang, G., Hu, Y., & Li, X. Happy Emotion Recognition From Unconstrained Videos Using 3D Hybrid Deep Features. *IEEE access*, 2021; *9:* 35524-35538.
13. Meryl, C. J., Dharshini, K., Juliet, D. S., Rosy, J. A., & Jacob, S. S. Deep Learning based Facial Expression Recognition for Psychological Health Analysis. In *2020 International Conference on Communication and Signal Processing (ICCSP)*, July, 2020; (1155-1158). IEEE.
14. Gory, S., Al-Khassaweneh, M., & Szczurek, P. Machine Learning Approach for Facial Expression Recognition. In *2020 IEEE International Conference on Electro Information Technology (EIT)*, July, 2020; (032-039). IEEE.
15. Abdullah, M., Ahmad, M., & Han, D. (2020, January). Facial expression recognition in videos: An CNN-LSTM based model for video classification. In *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, (1-3). IEEE.
16. Liu, S., Li, D., Gao, Q., & Song, Y. (2020, November). Facial Emotion Recognition Based on CNN. In *2020 Chinese Automation Congress (CAC)*, (398-403). IEEE.
17. Modi, S., & Bohara, M. H. (2021, May). Facial Emotion Recognition using Convolution Neural Network. In *2021 5th International Conference on Intelligent Computing and ControlSystems (ICICCS)*, (1339-1344). IEEE.
18. Kollias, D., & Zafeiriou, S. Exploiting multi-cnn features in cnn-rnn based dimensional emotion recognition on the omg in-the-wild dataset. *IEEE Transactions on Affective Computing*, 2020; *12*(3): 595-606.
19. Moravčík, E., & Basterrech, S. Image- Based Facial Emotion Recognition Using Convolutional Neural Networks and Transfer Learning. In *International Conference on Intelligent Information Technologies for Industry*, September, 2021; (3-14). Springer, Cham.

20. Pham, L., Vu, T. H., & Tran, T. A. (2021, January). Facial expression recognition using residual masking network. In *2020 25th International Conference on Pattern Recognition (ICPR)* (pp. 4513-4519). IEEE.
21. Liu, D., Zhang, H., & Zhou, P. Video- based facial expression recognition using graph convolutional networks. In *2020 25th International Conference on Pattern Recognition (ICPR),* January, 2021; 607-614). IEEE.
22. Antoniadis, P., Filntisis, P. P., & Maragos, P. Exploiting emotional dependencies with graph convolutional networks for facial expression recognition. In *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021),* December, 2021; (1-8). IEEE.
23. Marriwala, N. Facial Expression Recognition Using Convolutional Neural Network. In *Mobile RadioCommunications and 5G Networks*, 2022; (605-617). Springer, Singapore.
24. Talegaonkar, I., Joshi, K., Valunj, S., Kohok, R., & Kulkarni, A. (2019, May). Real time facial expression recognition using deep learning. In *Proceedings of International Conference on Communication and Information Processing (ICCIP).*
25. Sun, X., Zheng, S., & Fu, H. ROI-attention vectorized CNN model for static facial expression recognition.*IEEE Access*, 2020; *8:* 7183-7194.