NEURAL NETWORK TECHNIQUE FOR DETECTING FACIAL EXPRESSIONS OF EMOTION

Koppula Prakash
AIT-CSE, AIML
Chandigarh University,
Punjab, India
21bcs8824@cuchd.in

MLK Subrahmanyam
AIT-CSE, AIML
Chandigarh University,
Punjab, India
21bcs8803@cuchd.in

Nikhil Kumar
AIT-CSE, AIML
Chandigarh University,
Punjab, India
20bcs6845@cuchd.in

Arya Chackraborthy
AIT-CSE, AIML
Chandigarh University,
Punjab, India
21bcs8814@cuchd.in

Abstract— Emotion Recognition is a task to process a human facial expression and classify it into certain emotion categories. Such task typically requires the feature extractor to detect the feature, and the trained classifier produces the label based on the feature. The problem is that the extraction of feature may be distorted by variance of location of object and lighting condition in the image. In this project, we address the solution of the problem by using a deep learning algorithm called Conventional Neural Network (CNN) to address the issues above. By using this algorithm, the feature of image can be extracted without user-defined feature-engineering, and classifier model is integrated with feature extractor to produce the result when input is given. In this way, such method produces a feature-location-invariant image classifier that achieves higher accuracy than traditional linear classifier when the variance such as lighting noise and background environment appears in the input image. The evaluation of the model shows that the accuracy of our lab condition testing data set is 94.63%, and for wild emotion detection it achieves only around 37% accuracy.

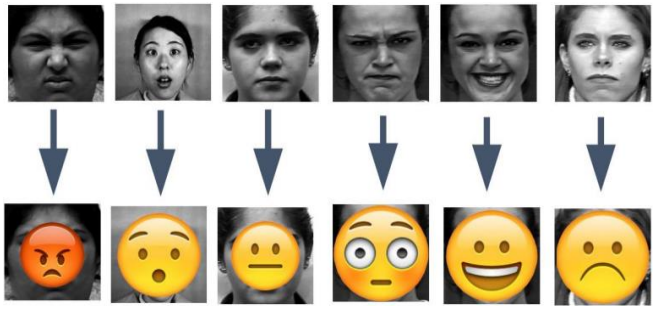
Keywords—Deep Learning, Convolutional Neural Network (CNN) (key words)

I. INTRODUCTION

Emotion Recognition is an important area of work improve the interaction between human and machine. Complexity of emotion makes the acquisition task more difficult. Quondam works are proposed to capture emotion through unimodal mechanism such as only facial expressions or only vocal input. More recently, inception to the idea of multimodal emotion recognition has increased the accuracy rate of the detection of the machine. Moreover, deep learning technique with neural network extended the success ratio of machine in respect of emotion recognition. It is very complicated to detect the emotions and distinguish among them. Before a decade or two emotions started to become a concern as an important addition towards the modern technology world. Rises the hope of new dawn for intelligence apparatus. Imagine a world where machines do feel what humans need or want. With the special kind of calculation then that machine could predict the further consequences and by which mankind could avoid serious circumstances and lot more. Humans are far stronger and more intelligent due to the addition of the emotion but less effective than machines.

The facial expression of human emotion is one of the major topics in facial recognition, and it can generate both technical and everyday application beyond laboratory experiment. This projection constructs a system of deep learning model to classify a given image of human facial emotion into one of the seven basic human emotions. The approach we take to build the model is through transfer learning of an existing pre-trained model, and the testing result will be evaluated based on accuracy of the model. The tasks in this project include preprocessing the image data, augmentation to enlarge the existing small dataset, test before training the model, training process, and predication with evaluation. The visual demonstration of the result will be similar to the figure below. The baseline of the test will be

around 14.7% (1 in 7), and our objective is to achieve a result better than baseline accuracy.



A demonstration: The emojis are correctly imposed on their corresponding faces. There are seven different emotions: happy, angry, sad, fear, surprise, neutral, and disgust. The input will be raw image of the expression, and output will be shown as above. The demonstration does not include disgust as its emoji is like angry.

II. LITERATURE REVIEW

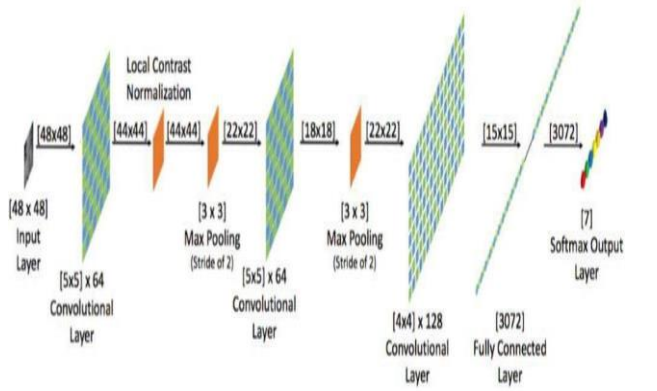
In recent years, facial emotion recognition has become a hot focus of research. To identify emotion from faces, most people utilize computer vision, machine learning, or deep learning technologies. This study gives a brief overview of FER research done over the last few decades. The traditional FER techniques are presented first, followed by a description of the typical FER system types and their major algorithms. The authors next describe deep-learning-based FER methods that use deep networks to enable "end-to-end" learning. This paper also looks at a new hybrid deep learning technique that employs a convolutional neural network (CNN) for spatial characteristics of a single frame and a Haar Cascade Classifier for temporal data of several frames. A brief overview of publicly accessible evaluation metrics is provided in the latter half of this work, as well as a comparison with benchmark findings, which constitute a standard for a quantitative comparison of FER investigations. Instead of minimizing the cross-entropy loss, learning reduces a margin-based loss. Study of multi-level features in a convolutional neural network for facial emotion identification by Hai-Duong Nguyen. They offer a model based on the data that purposely combines a hierarchy of characteristics to better the categorization job. The model was tested on the FER2013 dataset and found to be similar to existing state-of-the-art approaches in terms of performance. Using a feedforward learning model, the authors in developed an instructor's face expression recognition technique within a classroom. For successful high-level feature extraction, the face is first recognized from the obtained lecture videos and important frames are picked, removing all unnecessary frames. Then, using several convolution neural networks and parameter tweaking, deep features are retrieved and supplied to a classifier.

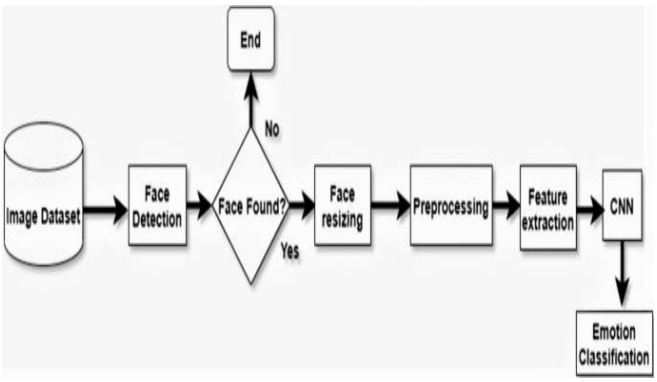
Table 1: Emotion recognition different approach and successes

Reference and year	Approach and Method	Performance
Wei-Long Zheng and Bao-Liang Lu (2016)	EEG-based affective models without labeled target data using transfer learning techniques (TCA-based Subject Transfer)	Positive (85.01%) emotion recognition rate is higher than other approaches but neutral (25.76%) and negative (10.24%) emotions are often confused with each other.
Zixing Zhang, Fabien Ringeval, Fabien Ringeval, Eduardo Coutinho, Erik Marchi and Björn Schüller (2016)	Semi-Supervised Learning (SSL) technique	Delivers a strong performance in the classification of high/low emotional arousal (UAR = 76.5%), and significantly outperforms traditional SSL methods by at least 5.0% (absolute gain).
Y. Fan, X. Lu, D. Li, and Y. Liu. (2016)	Video-based Emotion Recognition Using CNN-RNN and C3D Hybrid Networks	Achieved accuracy 59.02% (without using any additional Emotion labeled video clips in training set) which is the best till now.
A. Yao, D. Cai, P. Hu, S. Wang, L. Shan and Y. Chen (2016)	HoloNet: towards robust emotion recognition in the wild	Achieved mean recognition rate of 57.84%.
Yelin Kim and Emily Mower Provos (2016)	Data driven framework to explore patterns (timings and durations) of emotion evidence, specific to individual emotion classes	Achieved 65.60% UW accuracy, 1.90% higher than the baseline.

III. PROPOSED WORK

In this project, we approach the problem by taking deep-learning method of Convolutional Neural Networks (CNNs), which integrates the step of handcrafted feature extraction with training of classifier. This system can achieve the relatively most optimal solution through the process of backpropagation in which the algorithm learns the weights through modified stochastic gradient descent that can find the directions that best minimize the loss from the ground truth. The numerical result of the algorithm will show a probabilistic result of each labeled class. In order to reduce computational expense, the technique of fine-tuning is applied so that a pre-trained model can adapt the variance of our local dataset with benefit of reducing computational expense. As results, such method best resolves the issues of lighting variations and different orientation of object in the image and thus achieves a higher accuracy.



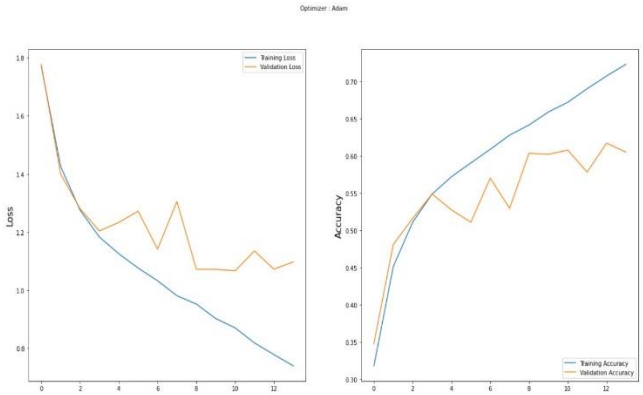


IV. RESULTS AND ANALYSIS

The first major problem was the limited amount of material available for developing a wide framework. Which must be defeated for the sake of nature's structure. The most common answer to this is move learning. This technique began with a pre-planned approach and calibrated this model using put-away information acquired from the real world. A series of preliminary studies confirmed the hypothesis that facial recognition would be more useful in highlight extraction. Some models make good use of such systems.

On datasets of a few hundred highlights or segments, machine learning methods operate effectively. The system accurately detects a picture, classifies its sentiment, and selects the appropriate emotion for the image. The choice of the deep learning classifier is based on the fact that it processes data across many layers. A deep learning algorithm, on the other hand, might be effective for less unexpected difficulties because it has access to a large amount of data. For photos, the standard benchmark for developing deep learning models for wide picture recognition is more than 2.1 million images. A decision tree was employed for a flawless display of sentiment analysis pattern analysis. The nodes and levels of the decision tree indicate the character, while the branch represents the outcome of the experiment. The decision tree has the advantage of making it extremely easy to see and comprehend the feeling and the outcome. A decision tree's operation is simple to comprehend. If the data has been categorized based on their movement, responses, and order, which ideally corresponds to different sorts of emotions. This has also been categorized into trees and sub-trees, which represent whether the individual is sad, furious, or pleased, and so on. If only there was a way to categorize their use of these approaches in a more straightforward manner. To do this, a retrain approach was utilized, which remembered the pattern and satisfied the criterion. Continue to the end of the tree when any of the conditions are met. It will cease checking and report "The

emotion cannot be recognized" if none of the requirements meet the intermediate condition. The emotion is a mystery. Emotions are difficult to comprehend. The same feeling can be expressed in a variety of ways. For the same feeling, various people have different ways of expressing it. Modern machine learning software can enhance law enforcement authorities in detecting emotion so that the computer can comprehend human emotion and behave and act more like humans. This emotional data was gathered from several online and offline sources. For example, Google's kaggel.com website. Friends and relatives, strangers, and so on. The Keras library was used to initially categorize and evaluate the emotion and obtain the data. The emotion is then identified using Haar features and Numpy. And with the assistance of Anaconda's platform. It creates output from raw data, and the outcome is displayed in real-time. The decision tree is a hierarchical data mining process that aids in the generation of probability judgments by calculating several characteristics that are originally used to define the emotion pattern. It also performed effective field research to gather more individuals and diverse types of people, as well as varied emotional deferent expressions and a variety of faces, in addition to online and offline data collection. The data set for online data gathering comes from kaggel.com. They deliver high-quality data sets. They used the numerical quantity of the photographs to transform the photos into pixel grayscale. Therefore, it provides excellent data as well as the best possible outcome. Both experts agreed that this sentiment analysis may assist detect emotions more precisely and help take proper actions in the name of accurate emotion identification. It would give additional information about the many sorts of expressions of their feeling, and also the proportion of each existing type of emotion. While working on this project, we discovered that a vast amount of testing datasets and keywords are required for improved accuracy. A scarcity of original data is also necessary to prolong the study activity. If you wish to handle a huge amount of test data in the quickest amount of time, you'll need a high-configuration graphics processing unit (GPU) certified computer. So, if this receives enough data and a powerful computer, it'll be much easier to increase the accuracy to more than 97 percent. It will also be able to apply that method to a new platform and get a different result, as well as assist in determining the emotion expression profile.



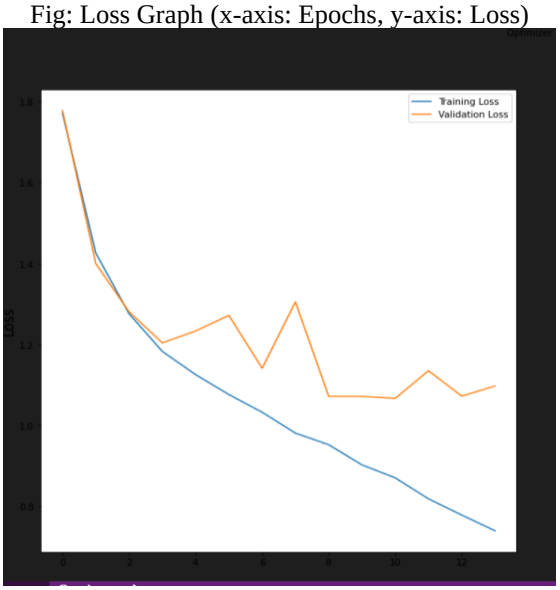

```
Epoch 1/48
225/225 [==============================] - 180s 799ms/step - loss: 1.7711 - accuracy: 0.3183 - val_loss: 1.7775 - val_accuracy: 0.3482
Epoch 2/48
225/225 [==============================] - 22s 99ms/step - loss: 1.4261 - accuracy: 0.4517 - val_loss: 1.3999 - val_accuracy: 0.4814
Epoch 3/48
225/225 [==============================] - 22s 97ms/step - loss: 1.2760 - accuracy: 0.5113 - val_loss: 1.2810 - val_accuracy: 0.5163
Epoch 4/48
225/225 [==============================] - 22s 100ms/step - loss: 1.1831 - accuracy: 0.5490 - val_loss: 1.2041 - val_accuracy: 0.5491
Epoch 5/48
225/225 [==============================] - 23s 100ms/step - loss: 1.1257 - accuracy: 0.5724 - val_loss: 1.2334 - val_accuracy: 0.5278
Epoch 6/48
225/225 [==============================] - 22s 98ms/step - loss: 1.0765 - accuracy: 0.5909 - val_loss: 1.2720 - val_accuracy: 0.5112
Epoch 7/48
225/225 [==============================] - 23s 101ms/step - loss: 1.0328 - accuracy: 0.6091 - val_loss: 1.1415 - val_accuracy: 0.5705
Epoch 8/48
225/225 [==============================] - 23s 103ms/step - loss: 0.9813 - accuracy: 0.6282 - val_loss: 1.3052 - val_accuracy: 0.5300
Epoch 9/48
225/225 [==============================] - 23s 103ms/step - loss: 0.9526 - accuracy: 0.6418 - val_loss: 1.0722 - val_accuracy: 0.6038
Epoch 10/48
225/225 [==============================] - 27s 120ms/step - loss: 0.9028 - accuracy: 0.6593 - val_loss: 1.0720 - val_accuracy: 0.6024
Epoch 11/48
225/225 [==============================] - 23s 104ms/step - loss: 0.8707 - accuracy: 0.6724 - val_loss: 1.0670 - val_accuracy: 0.6081
Epoch 12/48
225/225 [==============================] - 22s 98ms/step - loss: 0.8188 - accuracy: 0.6906 - val_loss: 1.1353 - val_accuracy: 0.5786
Epoch 13/48
225/225 [==============================] - 23s 104ms/step - loss: 0.7788 - accuracy: 0.7076 - val_loss: 1.0726 - val_accuracy: 0.6175
Epoch 14/48
225/225 [==============================] - ETA: 0s - loss: 0.7399 - accuracy: 0.7234Restoring model weights from the end of the best epoch.

Epoch 00014: ReduceLROnPlateau reducing learning rate to 0.000200000009494906.
225/225 [==============================] - 23s 102ms/step - loss: 0.7399 - accuracy: 0.7234 - val_loss: 1.0975 - val_accuracy: 0.6054
Epoch 00014: early stopping
```

V. CONCLUSION

In this project, we studied the what are the techniques used previously which mentioned in literature survey, what are the real time application for the proposed system which covered in objective and motivation.
By using the mentioned dataset, the neural network will be trained and by adding maximum hidden layer i.e with deep layer in convolutional network the result will be calculated. It covers the concept of facial expression recognition with aimed to classify images of faces into any of seven discrete emotion or face expression categories that represent universal human emotions.
Human face is being detected efficiently and analyzed the patterns of various kinds of expressions of emotions. Research suggests that the facial expressions accompanying anger, disgust, fear, happiness, sadness, and surprise are human universals, and the amygdala is central to emotion recognition.

VI. RESULTS

Fig: Loss Graph (x-axis: Epochs, y-axis: Loss)



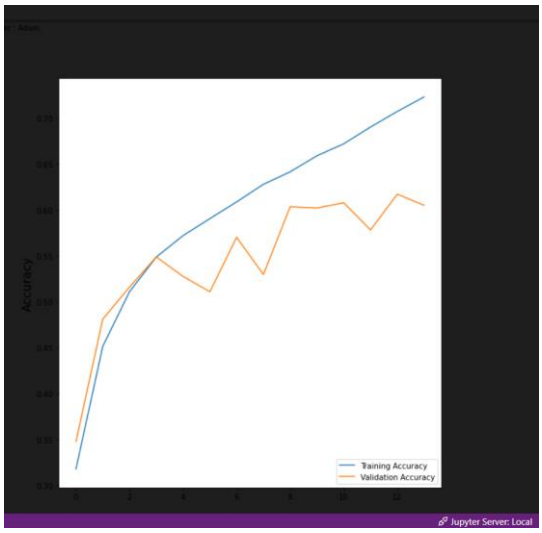
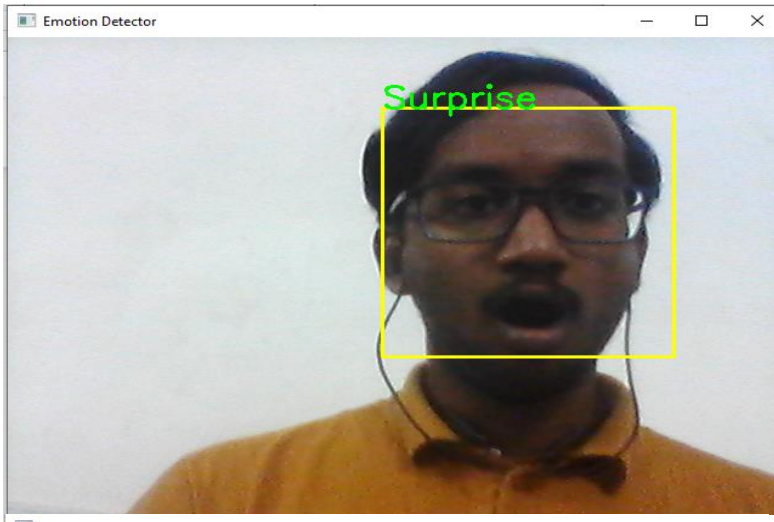
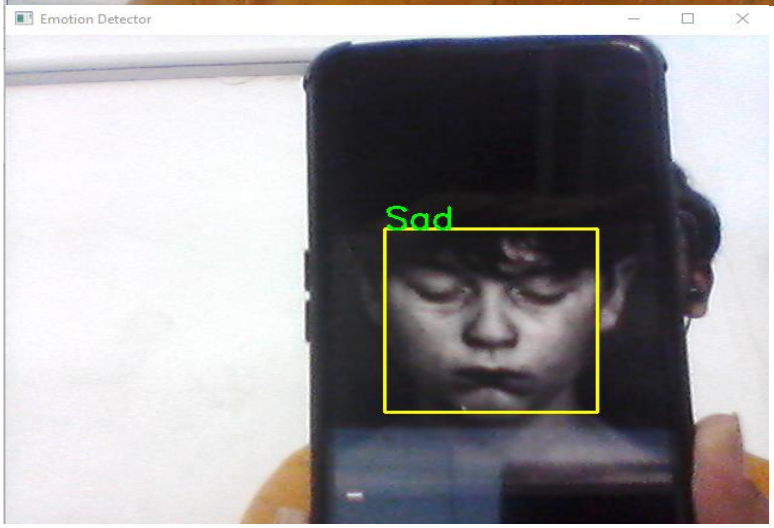
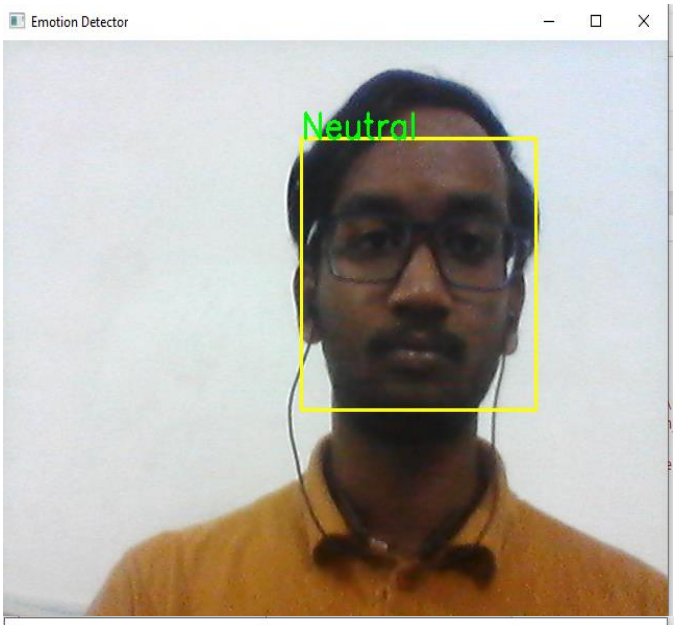


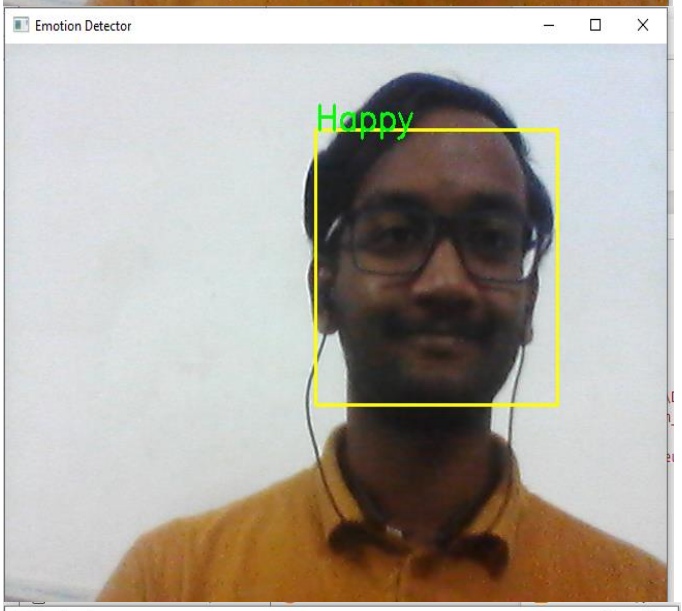
Fig: Accuracy Graph (x-axis: Epochs, y-axis: Accuracy)

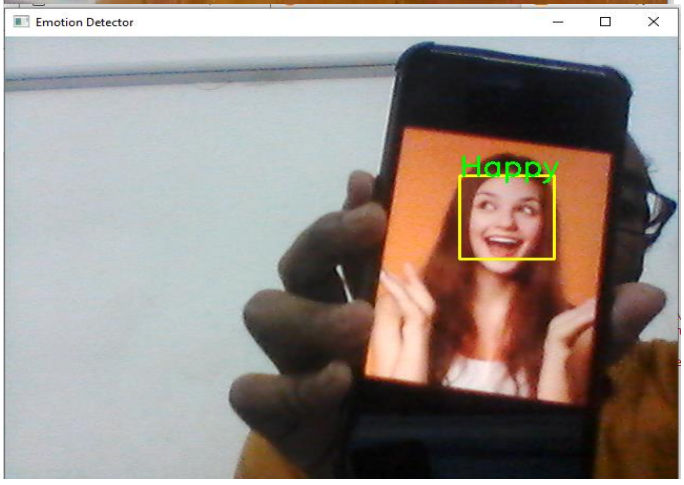
VII. RESULTS AND OUTPUTS

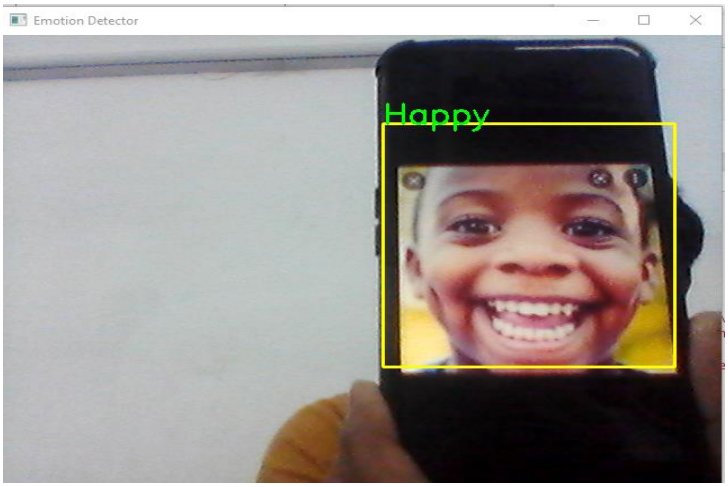












VIII. REFERENCES

[1] Guo, J., Zhou, S., Wu, J., Wan, J., Zhu, X., Lei, Z., & Li, S. Z. (2017).
Multi-modality Network with Visual and Geometrical Information for Micro Emotion Recognition. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 814–819.
https://doi.org/10.1109/FG.2017.103

[2] Liu, K., Zhang, M., & Pan, Z. (2016). Facial Expression Recognition with CNN Ensemble. Proceedings - 2016 International Conference on Cyberworlds, CW 2016, 163–166. https://doi.org/10.1109/CW.2016.34

[3] Nour, N., Elhebir, M., & Viriri, S. (2020). Face Expression Recognition using Convolution Neural Network (CNN) Models. International Journal of Grid Computing & Applications, 11(4), 1–11.
https://doi.org/10.5121/ijgca.2020.11401

[4] Liu, Y., Yuan, X., Gong, X., Xie, Z., Fang, F., & Luo, Z. (2018). Conditional convolution neural network enhanced random forest for facial expression recognition. Pattern Recognition, 84, 251–261.
https://doi.org/10.1016/j.patcog.2018.07.016

[5] Rajendra Kurup, A., Ajith, M., & Mart´ınez Ramon, M. (2019). Semi- supervised facial expression recognition using reduced spatial features and Deep
Belief Networks. Neurocomputing, 367, 188–197.
https://doi.org/10.1016/j.neucom.2019.08.029