Vidhyayana - ISSN 2454-8596

An International Multidisciplinary Peer-Reviewed E-Journal

www.j.vidhyayanaejournal.org

Indexed in: ROAD & Google Scholar

A Framework Of Students Facial Emotion Recognition Using Convolutional Neural Network For Different Articulations

Kejal Chintan Vadza

Research Scholar, Department of Computer Applications

Sabarmati University, Ahmedabad, Gujarat

Prof. Amit Kumar Shastri

Professor, Department of Computer Applications

Sabarmati University, Ahmedabad, Gujarat

ABSTRACT

These days, deep learning techniques know a major accomplishment in different fields including Computer vision. Without a doubt, a convolutional neural organizations (CNN) model can be prepared to examine pictures and recognize facial feelings. In this paper, we make a framework that perceives understudies' feelings from their faces. Our framework comprises three stages: face discovery utilizing Haar Cascades, standardization, and feeling acknowledgment utilizing CNN on FER 2013 information base with seven sorts of articulations. Acquired outcomes show that face feeling acknowledgment is plausible in training, thus, it can assist educators with changing their show as indicated by the understudies' feelings.

Keywords: facial expression, Emotion recognition, Convolutional neural networks (CNN), Deep learning, Intelligent classroom management system.

Introduction:

The face is an individual's most expressive and open part [1]. It is capable of expressing a wide range of emotions without saying anything. Look acknowledgment distinguishes sensation from a picture of a face; it is a sign of a person's movement and personality. In the twentieth century, American therapists Ekman and Friesen [2] identified six basic emotions (outrage, dread, disdain, pity, shock, and satisfaction) that are universal. Because of its impact on clinical practise, friendly mechanical technology, and training, look recognition has gotten a lot of attention in recent years. Feeling has an important role in training, according to many studies. Tests, surveys, and perceptions are currently used by educators as sources of criticism, however these old tactics are frequently associated with low productivity. Using the appearance of understudies, the instructor can alter their approach and educational materials to assist understudies in learning. The goal of this piece is to implement feeling recognition in education by comprehending a programmed framework that assesses understudies' looks using a Convolutional Neural Network (CNN), which is a deep learning calculation often used in image grouping. It consists of a multistage image processing system that eliminates include representations. Face identification, standardisation, and feeling acknowledgment are the three stages of our framework, which should be one of the seven feelings: impartial, wrath, dread, problem, bliss, astonishment, and disgust. The remainder of this document is organised as follows: The second section examines the linked work. The suggested system is described in Section 3. In part 4, the details of the implementation are described, followed by the experimental findings and comments in section 5. We finish this paper with prospective developments of our work in the concluding part.

II. Related Works

Face Emotion Recognition has piqued the interest of a number of academics who want to improve the learning environment (FER). Tang et al. [3] developed a framework for analysing the impact of study hall teaching by dissecting the looks of understudies. The framework is divided into five stages: data gathering, face recognition, face acknowledgment, gaze acknowledgement, and post-handling. For order, K-nearest neighbour (KNN) is employed, and for design analysis, Uniform Local Gabor Double Pattern Histogram Sequence (ULGBPHS) is used. Savva et al. [4] created a web software that simulates a study of understudies' feelings about active up close and personal homeroom direction. The tool captures live accounts using webcams located in study halls, which are subsequently exposed to AI algorithms.

Whitehill et al. presented a way for detecting commitment from the looks of understudies in [5.] To separate commitment while understudies communicated with intellectual capacities prepared to programme, the method uses Gabor elements and SVM calculation. The creators received feedback from human adjudicators on their recordings. The makers in [6] then used computer vision and AI ways to identify the influence of understudies in a school PC research facility, where the understudies were interacting with an instructive game aimed at clarifying fundamental notions of old-style mechanics.

The designers suggested a framework in [7] that recognises and screens understudies' feelings and provides feedback gradually in order to improve the e-learning atmosphere for better material delivery. To get understudies to settle in an e-learning environment, the framework uses moving images of eyes and heads to reason important info. Ayvaz et al. [8] developed a Facial Emotion Recognition System (FERS) that detects understudy' enthusiastic states and inspiration in videoconference-style e-learning. The system employs four machine learning algorithms (SVM, KNN, Random Forest and Arrangement, and Regression Trees), with the KNN and SVM algorithms providing the best exactness rates.

Kim et al. [9] suggested a framework capable of providing ongoing suggestions to the instructor in order to increase the memorability and character of their speech by allowing the educator to make gradual changes to their non-verbal behaviour, such as non-verbal communication and appearance. The authors of [10] suggested a model for detecting feelings in a virtual learning environment based on facial emotion recognition using the Haar Cascades approach [14] to separate mouth and eyes using a JAFF data set to identify feelings. Chiou et al. [11] used remote sensor network technology to create an intelligent classroom management system that instructs instructors on how to swiftly switch assistance modes to avoid wasting time.

III. Proposed Work

In this section, we show how we used Convolutional Neural Network (CNN) technology to deconstruct the looks of understudies. To begin, the framework recognises the face in the input image, and these identified countenances are clipped and standardised to a size of 4848 pixels. Then these photographs of people's faces are used as a donation to CNN. Finally, the glance acknowledgment results are the end outcome (outrage, joy, misery, repugnance, shock, or impartial). The construction of our proposed technique is depicted in Figure 1.

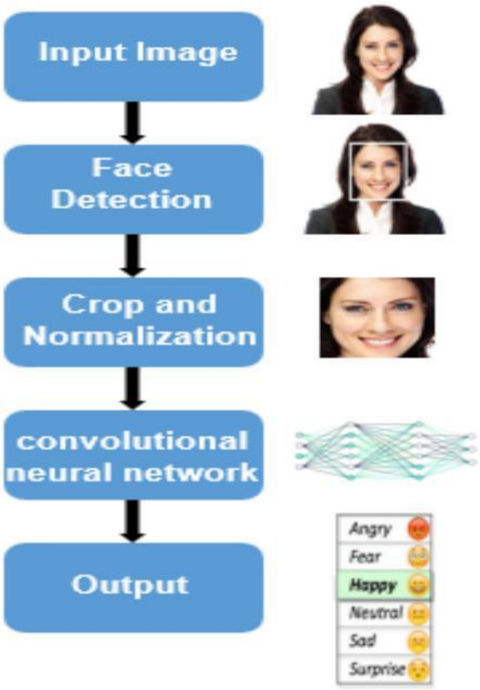


Fig. 1. The structure of our facial expression recognition system.

In comparison to previous picture arrangement calculations, a Convolutional Neural Network (CNN) is a powerful artificial neural network that can differentiate visual examples from input photographs with minimal pre-processing. This means that the network learns the pathways that were previously hand-engineered in traditional calculations [19]. A neuron is a significant unit inside CNN layers. They're linked together in such a way that the output of one layer's neurons becomes the contribution of the next layer's neurons. The backpropagation calculation is used to process the expense's incomplete subordinates. Convolution refers to the use of a channel or section of an information picture to construct a component map. Truth be told, the CNN model contains 3 sorts of layers as displayed in Figure 2:

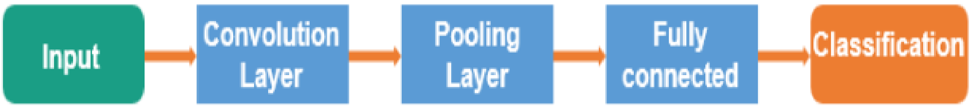


Fig. 2. CNN architecture.

Convolutional Network: is the key layer responsible for extracting features from input images. Convolution's primary function, in the case of a ConvNet, is to segregate items from the information picture. By learning picture highlights using small squares of information information, convolution saves the spatial link between pixels [21]. It is a tiny item that is played out between two frameworks, one of which is the picture and the other is a kernal. Equation 1 deals with the convolution recipe:

$$\mathbf{net(t, f) = (x * w)[t, f]} = \sum m \sum n \quad \mathbf{x[m, n]w[t - m, f - n]} \ (1)$$

Where net(t, f) is the output in the next layer, x is the input image, w is the filter matrix and * is the convolution operation. Figure 3, shows how the convolution works.

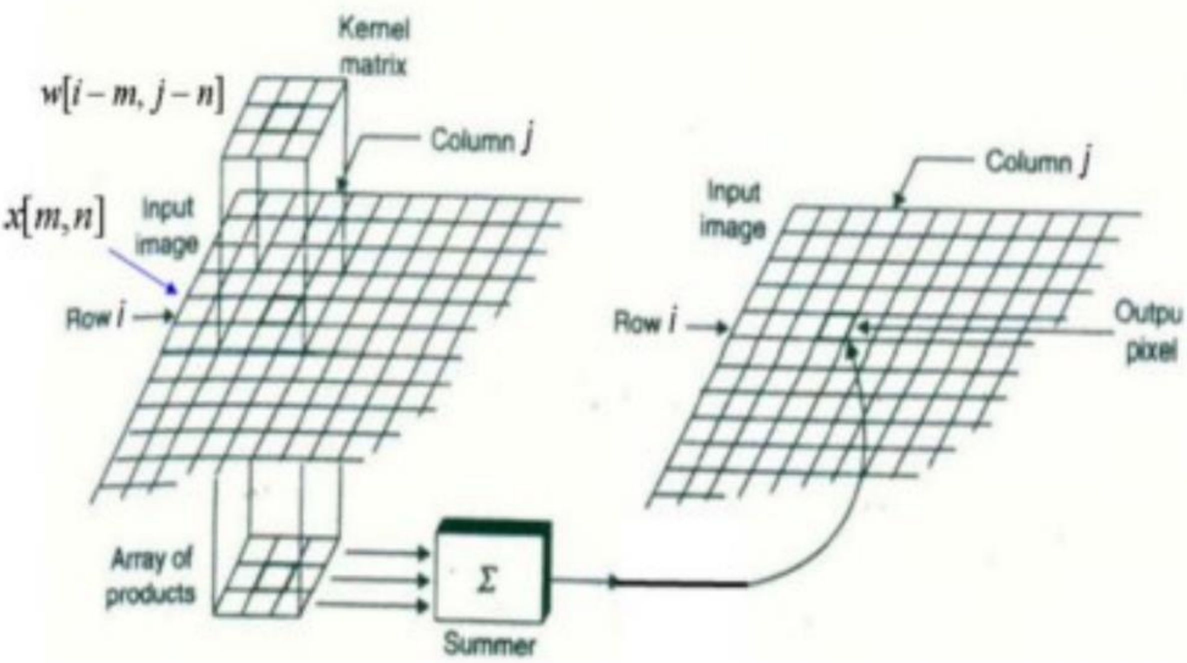
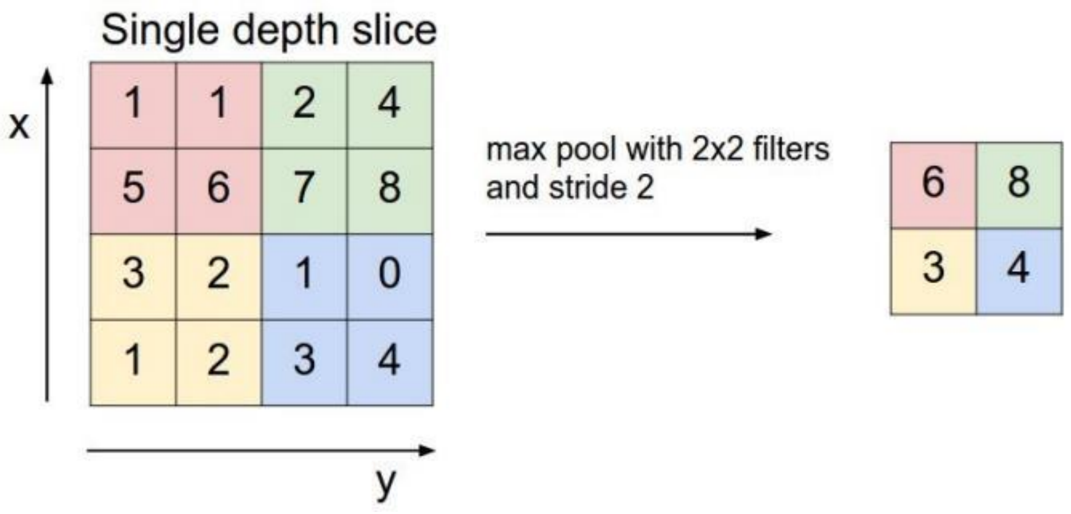


Fig. 3. Details on Convolution layer.

Pooling Layer: reduces the dimensionality of each element map while still retaining the primary data [21]. There are three types of pooling: maximum pooling, average pooling, and total pooling. Pooling's capability is to reduce the spatial size of the information depiction over time and make the organisation insensitive to minor changes, twists, and interpretations in the data [21]. In our work, we used the square limit as a single result for the pooling layer, as shown in Figure 4.



1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

6	8
3	4

Fig. 4. Details on Pooling layer [20].

Fully Connected Network: is a standard Multi-Layer Perceptron with an initiation work in the result layer. The term "Fully Connected" implies that each neuron in the previous layer is linked to each neuron in the layer after it. The Fully Connected layer was created with the goal of using the results of the convolutional and pooling layers to organise the information picture into different classifications based on the preparation dataset. As a result, the Revolution and

The Pooling layers act as Feature Extractors from the information picture, whilst the Fully Connected layer acts as a classifier. [21].

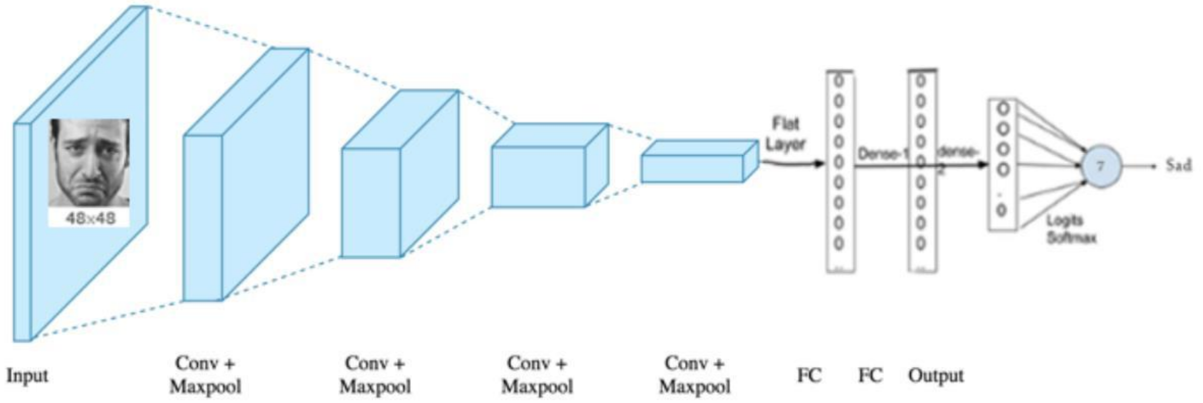


Fig. 5. Our convolutional neural network model.

Our CNN model is depicted in Figure 5. It has four convolutional layers, four pooling layers for extracting features, two fully associated layers, and a softmax layer with seven sensation classes. The input image is a grayscale facial image with a size of 4848 pixels. With step 2, we used 33 channels for each convolutional layer. We used the max-pooling layer and 22 pieces with step 2 for the pooling layers. As a result, we used the Rectified Linear Unit (ReLU), which is the most often used actuation approach as of late, to illustrate the non-linearity in our model.

R(z) = max(0, z) (2)

As shown in Figure 6, R(z) is zero when z is less than zero and R(z) is equal to z when z is above or equal to zero. Table I presents the network configuration of our model.

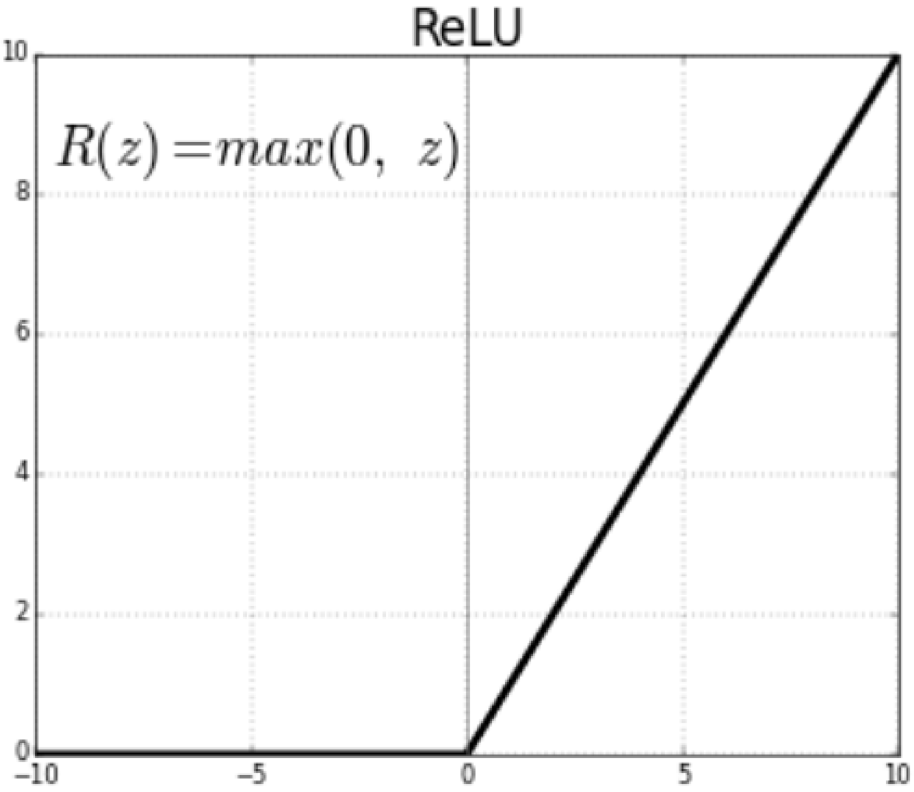


Fig. 6. ReLU function.

TABLE I. CNN CONFIGURATION

Layer type	Size	Stride
Data	48x48	-
Convolution 1	3x3	2
Max Pooling 1	2x2	2
Convolution 2	3x3	2
Max Pooling 2	2x2	2
Convolution 3	3x3	2
Max Pooling 3	2x2	2
Convolution 4	3x3	2
Max Pooling 4	2x2	2
Fully Connected	-	-
Fully Connected	-	-

IV. IMPLEMENTATION DETAILS

A. Data acquisition: We used the FER2013 [12] data set to create our CNN architecture, as shown in Figure 7. It was developed using the Google image search API and debuted at the ICML 2013 Challenges. Faces in the data base have been standardised to 4848 pixels by default. With 7 articulation names, the FER2013 information base has 35887 photographs (28709 preparing pictures, 3589 approval pictures, and 3589 test pictures). Table II shows the number of photos required for each emotion.

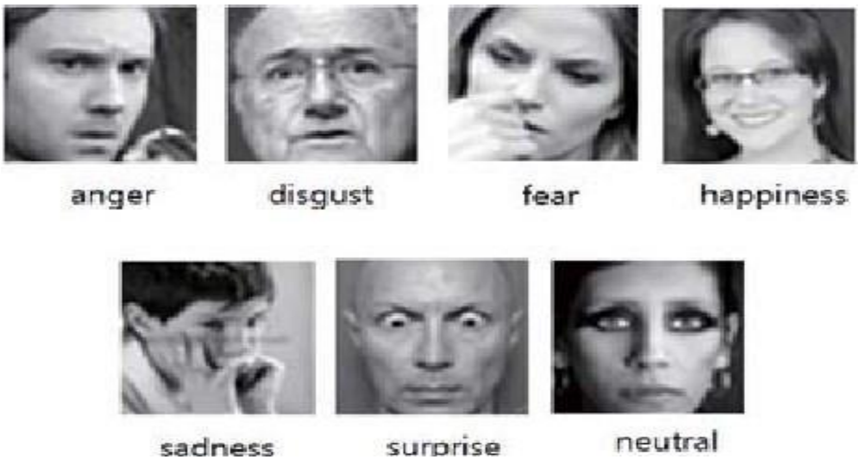


Fig. 7. Samples from FER 2013 database.

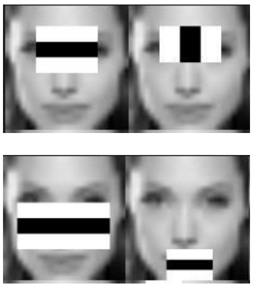
TABLE II. THE NUMBER OF IMAGE FOR EACH EMOTION OF FER 2013 DATABASE

Emotion label	Emotion	Number of image
0	Angry	4593
1	Disgust	547
2	Fear	5121
3	Happy	8989
4	Sad	6077
5	Surprise	4002
6	Neutral	6198

B. CNN Implementation: As shown in Figure 8, we used the OpenCV software [16] to capture live video from web cameras and to discriminate understudy' faces using the Haar Cascades technique [14]. Freund et al. [15], who earned the 2003 Gödel Prize for their work, invented the Adaboost learning computation, which is used by Haar Cascades. To produce a compelling result of classifiers, the Adaboost learning calculation selected a few of a large number of large elements from a large set. Using TensorFlow [18] and Keras [17] undeniable level API, we created a Convolutional Neural Network model.

IV. Implementation Details

V. Experimental Results



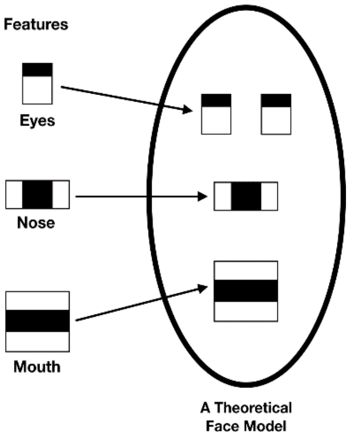


Fig. 8. Face Detection using Haar Cascades.

We used the ImageDataGenerator class in Keras to expand the size of the photo, as shown in Figure 9. This class allowed us to pivot, move, shear, zoom, and flip the prepared photographs. Rotation range=10, width shift range=0.1, zoom range=0.1, height shift range=0.1, and horizontal flip=True are the settings used.

Original *Transformed*

Fig. 9. Image augmentation using Keras.

After that, we defined our CNN model as having four convolutional layers, four pooling layers, and two totally associated layers. From then on, we'll add nonlinearity to our equations.

We used group standardisation to standardise the actuation of the point of reference layer at each clump in the CNN model, and L2 regularisation to apply punishments on the model's numerous boundaries. As a result, we chose softmax as our final enactment project; it accepts a vector z of k integers as input and standardises it into a likelihood circulation. Figure 10 depicts the softmax project:

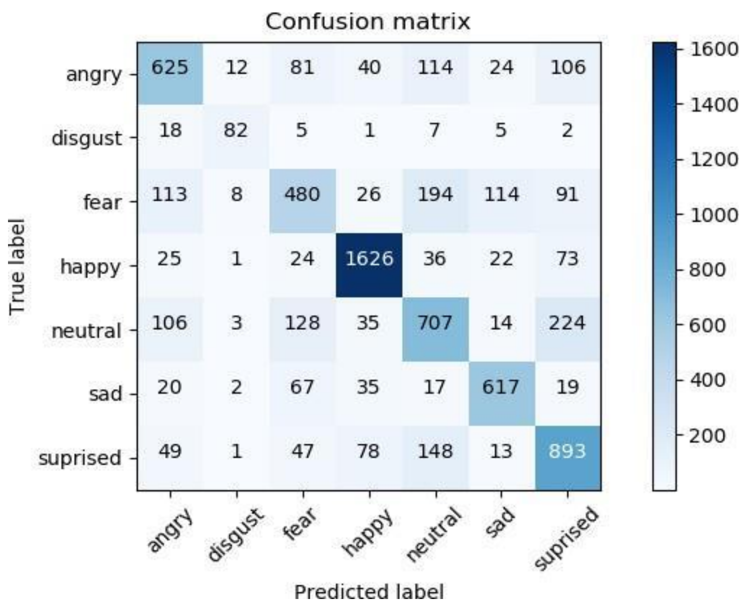
$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}} \; for\; j = 1, \ldots., k$$

Fig. 10. Softmax function.

We split the data set into 80 percent preparation information and 20 percent test information to create our CNN model, and then we gathered the model using the Stochastic angle drop (SGD) streamlining agent. Keras examines whether our model performed better than models from previous ages at each age. If this is the case, the new best model loads are saved to a file. This will allow us to stack the loads without having to retrain it in order to use it in a different situation.

V. EXPERIMENTAL RESULTS

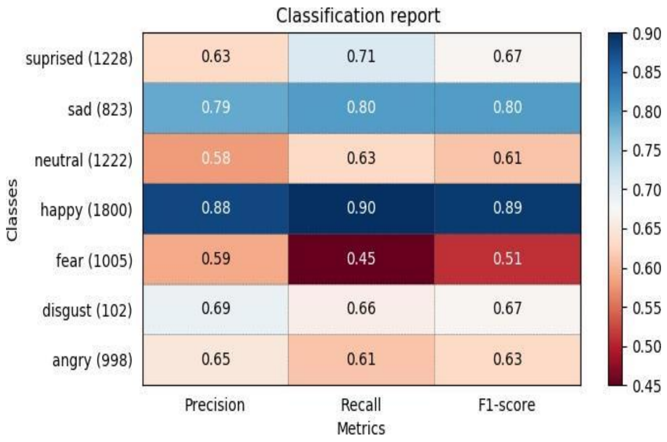
We built our Convolutional Neural Network model using the FER 2013 data set, which includes seven emotions (bliss, outrage, misery, disdain, nonpartisan, dread, and shock) The famous face images were reduced to 4848 pixels and converted to grayscale images, after which they were used as contributions to the CNN model. As a result, nine young skilled understudies from our staff participated in the test, two of whom were wearing spectacles. Figure 11 depicts the emotional outcomes of nine understudies. The red message addresses the predicted feeling mark, and the red bar addresses the likelihood of the feeling.



	angry	disgust	fear	happy	neutral	sad	suprised
angry	625	12	81	40	114	24	106
disgust	18	82	5	1	7	5	2
fear	113	8	480	26	194	114	91
happy	25	1	24	1626	36	22	73
neutral	106	3	128	35	707	14	224
sad	20	2	67	35	17	617	19
suprised	49	1	47	78	148	13	893

At the 106 ages, we maintained an exact speed at 70%. Figures 12 and 13 show the jumbled grid, accuracy, review, and F1-score, which were used to evaluate the productivity and nature of our proposed technique. Our model is fantastic at predicting happy and great outcomes. In any case, it foresees insufficiently feared appearances since it mistook them for mournful faces.

Fig. 12. Confusion matrix of the proposed method on FER 2013 database.



	Precision	Recall	F1-score
suprised (1228)	0.63	0.71	0.67
sad (823)	0.79	0.80	0.80
neutral (1222)	0.58	0.63	0.61
happy (1800)	0.88	0.90	0.89
fear (1005)	0.59	0.45	0.51
disgust (102)	0.69	0.66	0.67
angry (998)	0.65	0.61	0.63

Fig. 13. Classification report of the proposed method on FER 2013 database.

VI. Conclusion and future work

As shown in Figure 8, we used the OpenCV package [16] to capture live images from a web camera and to differentiate understudy' faces using the Haar Cascades technique [14]. Freund et al. [15], who earned the 2003 Gödel Prize for their work, invented the Adaboost learning computation, which is used by Haar Cascades. To produce a compelling result of classifiers, the Adaboost learning calculation selected a few of a large number of large elements from a large set. Using TensorFlow [18] and Keras [17] undeniable level API, we created a Convolutional Neural Network model.

VII. References

[1]. R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: the state of the art. New York: Wiley, 1978.

[2]. P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” Journal of Personality and Social Psychology, vol. 17, no 2, p. 124 129, 1971.

[3]. C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, “Automatic Facial Expression Analysis of Students in Teaching Environments,” in Biometric Recognition, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan,

[4]. W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.

[5]. A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, “Recognizing student facial expressions: A web application,” in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.

[6] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” IEEE Transactions on Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014.

[7]. N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang and W. Zhao, “Automatic Detection of Learning-Centered Affective States in the Wild,” in Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI ’15, Atlanta, Georgia, USA, 2015, p. 379-388.

[8]. Krithika L.B and Lakshmi Priya GG, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” Procedia Computer Science, vol. 85, p. 767-776, 2016.

[9]. U. Ayvaz, H. Gürüler, and M. O. Devrim, “USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS,” Information Technologies and Learning Tools, vol. 60, no 4, p. 95, sept. 2017.

[10]. Y. Kim, T. Soyata, and R. F. Behnagh, “Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education,” IEEE Access, vol. 6, p. 5308-5331, 2018.

[11]. D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, “An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment,” Procedia Computer Science, vol. 125, p. 2-10, 2018.

[12]. C.-K. Chiou and J. C. R. Tseng, “An intelligent classroom management system based on

wireless sensor networks," in 2015 8th International Conference on Ubi-Media Computing (UMEDIA), Colombo, Sri Lanka, 2015, p. 44-48.

[13]. I. J. Goodfellow et al., "Challenges in Representation Learning: A report on three machine learning contests," arXiv:1307.0414 [cs, stat], juill. 2013.

[14]. A. Fathallah, L. Abdi, and A. Douik, "Facial Expression Recognition via Deep Learning," in 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), Hammamet, 2017, p. 745-750.

[15]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, vol. 1, p. I-511-I-518.

[16]. Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no 1, p. 119-139, août 1997.

[17]. Tensorflow. tensorflow.org .aionlinecourse.com/tutorial/machine-learning/convolution-neural- network. Accessed 20 June 2019

[18]. S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, p. 1-6.

[19]. Ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/. Accessed 05 July 2019