Facial Emotion Recognition of Students using Convolutional Neural Network

Imane Lasri
Laboratory of Conception and Systems
Faculty of Sciences Rabat, Mohammed V University
Rabat, Morocco
imanelasri95@gmail.com

Anouar Riad Solh
Laboratory of Conception and Systems
Faculty of Sciences Rabat, Mohammed V University
Rabat, Morocco
anouarriadsolh@yahoo.fr

Mourad El Belkacemi
Laboratory of Conception and Systems
Faculty of Sciences Rabat, Mohammed V University
Rabat, Morocco
mourad_prof@yahoo.fr

Abstract— **Nowadays, deep learning techniques know a big success in various fields including computer vision. Indeed, a convolutional neural networks (CNN) model can be trained to analyze images and identify face emotion. In this paper, we create a system that recognizes students' emotions from their faces. Our system consists of three phases: face detection using Haar Cascades, normalization and emotion recognition using CNN on FER 2013 database with seven types of expressions. Obtained results show that face emotion recognition is feasible in education, consequently, it can help teachers to modify their presentation according to the students' emotions.**

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

I. Introduction

The face is the most expressive and communicative part of a human being [1]. It's able to transmit many emotions without saying a word. Facial expression recognition identifies emotion from face image, it is a manifestation of the activity and personality of a human. In the 20th century, the American psychologists Ekman and Friesen [2] defined six basics' emotions (anger, fear, disgust, sadness, surprise and happiness), which are the same across cultures.

Facial expression recognition has brought much attention in the past years due to its impact in clinical practice, sociable robotics and education. According to diverse research, emotion plays an important role in education. Currently, a teacher use exams, questionnaires and observations as sources of feedback but these classical methods often come with low efficiency. Using facial expression of students the teacher can adjust their strategy and their instructional materials to help foster learning of students.

The purpose of this article is to implement emotion recognition in education by realizing an automatic system that analyze students' facial expressions based on Convolutional Neural Network (CNN), which is a deep learning algorithm that are widely used in images classification. It consist of a multistage image processing to extract feature representations. Our system includes three phases: face detection, normalization and emotion recognition that should be one of these seven emotions: neutral, anger, fear, sadness, happiness, surprise and disgust.

The rest of this paper is structured as follows: Section 2 reviews the related work. Section 3 describes the proposed system. The implementation details are presented in section 4, followed by the experimental results and discussion in section 5. In the last section we conclude this paper with the future extensions of our work.

II. Related Work

Many researchers are interested in improving the learning environment with Face Emotion Recognition (FER). Tang et al. [3] proposed a system which is able to analyze students' facial expressions in order to evaluate classroom teaching effect. The system is composed of five phases: data acquisition, face detection, face recognition, facial expression recognition and post-processing. The approach uses K-nearest neighbor (KNN) for classification and Uniform Local Gabor Binary Pattern Histogram Sequence (ULGBPHS) for pattern analysis. Savva et al. [4] proposed a web application that performs an analysis of students' emotion who participating in active face-to-face classroom instruction. The application uses webcams that are installed in classrooms to collect live recordings, then they applied machine learning algorithms on its.

In [5] Whitehill et al. proposed an approach that recognizes engagement from students' facial expressions. The approach uses Gabor features and SVM algorithm to identify engagement as students interacted with cognitive skills training software. The authors obtained labels from videos annotated by human judges. Then, the authors in [6] used computer vision and machine learning techniques to identify the affect of students in a school computer laboratory, where the students were interacting with an educational game aimed to explain fundamental concepts of classical mechanics.

In [7] the authors proposed a system that identifies and monitors student's emotion and gives feedback in real-time in order to improve the e-learning environment for a greater content delivery. The system uses moving pattern of eyes and head to deduce relevant information to understand students' mood in an e-learning environment. Ayvaz et al. [8] developed a Facial Emotion Recognition System (FERS), which recognizes the emotional states and motivation of students in videoconference type e-learning. The system uses 4 machine learning algorithms (SVM, KNN, Random Forest and Classification & Regression Trees) and the best accuracy rates

978-1-7281-0003-6/19/$31.00 ©2019 IEEE

were obtained using KNN and SVM algorithms. Kim et al. [9] proposed a system which is able of producing real-time recommendation to the teacher in order to enhance the memorability and the quality of their lecture by granting the teacher to make modification in real-time to their non-verbal behavior like body language and facial expressions. The authors in [10] proposed a model that recognizes emotion in virtual learning environment based on facial emotion recognition with Haar Cascades method [14] to identify mouth and eyes on JAFF database in order to detect emotions. In [11] Chiou et al. used wireless sensor network technology to create an intelligent classroom management system that aids teachers to modify instruction modes rapidly to avert wasting of time.

III. Proposed Approach

In this section, we describe our proposed system to analyze students' facial expressions using a Convolutional Neural Network (CNN) architecture. First, the system detects the face from input image and these detected faces are cropped and normalized to a size of 48×48. Then, these face images are used as input to CNN. Finally, the output is the facial expression recognition results (anger, happiness, sadness, disgust, surprise or neutral). Figure 1 presents the structure of our proposed approach.

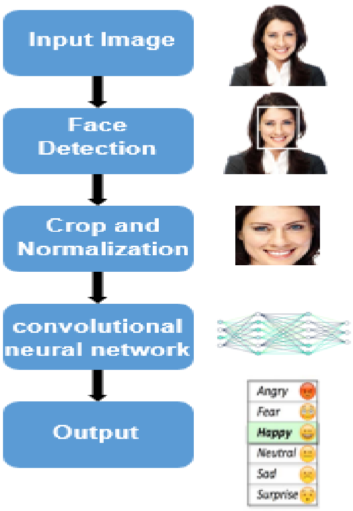


Fig. 1. The structure of our facial expression recognition system.

A Convolutional Neural Network (CNN) is a deep artificial neural networks that can identify visual patterns from imput image with minimal pre-processing compared to other image classification algorithms. This means that the network learns the filters that in traditional algorithms were hand-engineered [19]. The important unit inside a CNN layers is a neuron. They are connected together, in order that the output of neurons at a layer becomes the input of neurons at the next layer.

In order to compute the partial derivatives of the cost function the backpropagation algorithm is used. The term convolution refers to the use of a filter or kernel on the input image to produce a feature map. In fact, CNN model contains 3 types of layers as shown in Figure 2:

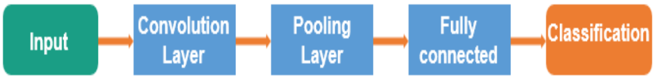


Fig. 2. CNN architecture.

Convolution Layer: is the first layer to extract features from an input image. The primary purpose of Convolution in case of a ConvNet is to extract features from the input image. Convolution preserves the spatial relationship between pixels by learning image features using small squares of input data [21]. It performs a dot product between two matrices, where one is the image and the other is a kernal. The convolution formula is represented in Equation 1 :

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

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.

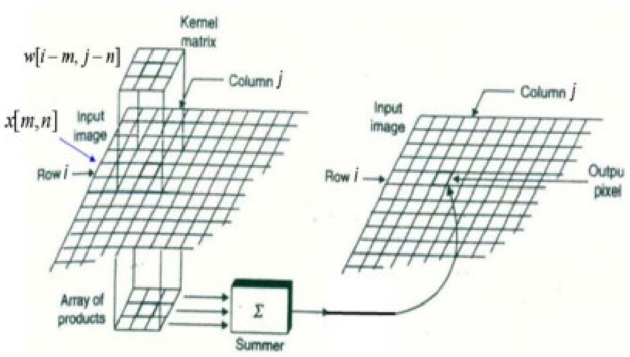


Fig. 3. Details on Convolution layer [20].

Pooling Layer: reduces the dimensionality of each feature map but retains the most important information [21]. Pooling can be of different types : Max Pooling, Average Pooling and Sum Pooling. The function of Pooling is to progressively reduce the spatial size of the input representation and to make the network invariant to small transformations, distortions and translations in the input image [21]. In our work, we took the maximum of the block as the single output to pooling layer as shown in Figure 4.

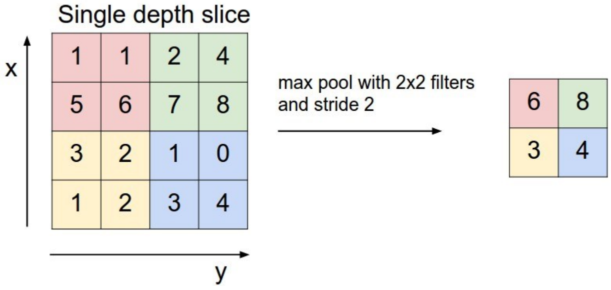


Fig. 4. Details on Pooling layer [20].

Fully connected layer: is a traditional Multi Layer Perceptron that uses an activation function in the output layer. The term "Fully Connected" implies that every neuron in the previous layer is connected to every neuron on the next layer. The purpose of the Fully Connected layer is to use the output of the convolutional and pooling layers for classifying the input image into various classes based on the training dataset. So the Convolution and Pooling layers act as Feature Extractors from the input image while Fully Connected layer acts as a classifier. [21].

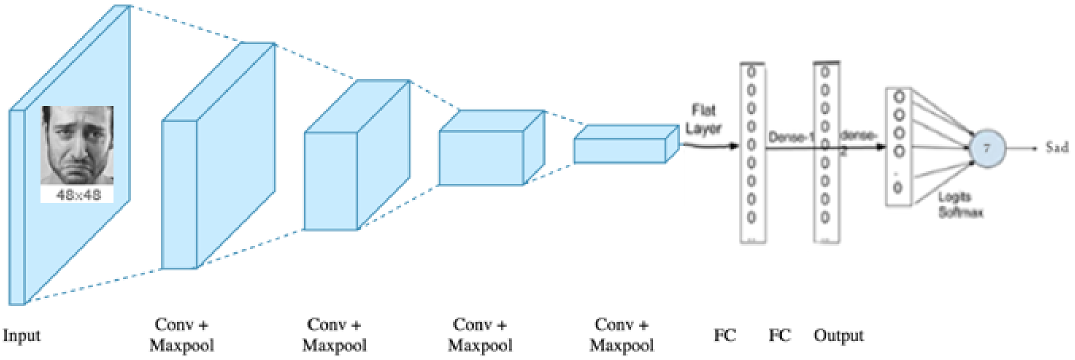


Fig. 5. Our convolutional neural network model.

Figure 5 represents our CNN model. It contains 4 convolutional layers with 4 pooling layers to extract features, and 2 fully connected layers then the softmax layer with 7 emotion classes. Input image is grayscale face image with a size of 48×48. For each convolutional layer we used 3×3 filters with stride 2. For the pooling layers, we used max pooling layer and 2×2 kernels with stride 2. Thus, to introduce the non linearity in our model we used the Rectified Linear Unit (ReLU), defined in Equation 2, wich is the most used activation function recently.

$$R(z) = \max(0, z) \tag{2}$$

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.

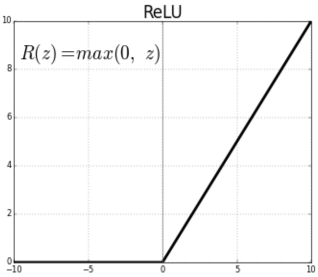


Fig. 6. ReLU function.

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

To train our CNN architecture, we used the FER2013 [12] database as shown in Figure 7. It was generated using the Google image search API and was presented during the ICML 2013 Challenges. Faces in the database have been automatically normalized to 48×48 pixels. The FER2013 database contains 35887 images (28709 training images, 3589 validation images and 3589 test images) with 7 expression labels. The number of images for every emotion is represented in Table II.

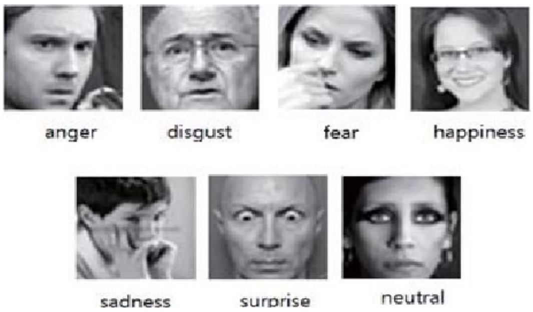


Fig. 7. Samples from FER 2013 database.

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

We used OpenCV library [16] to capture live frames from web camera and to detect students' faces based on Haar Cascades method [14] as shown in Figure 8. Haar Cascades uses the Adaboost learning algorithm invented by Freund et al. [15], who won the 2003 Gödel Prize for their work. The Adaboost learning algorithm chose a few number of significant features from a large set in order to provide an effective result of classifiers. We built a Convolutional Neural Network model using TensorFlow [18] Keras [17] high-level API.

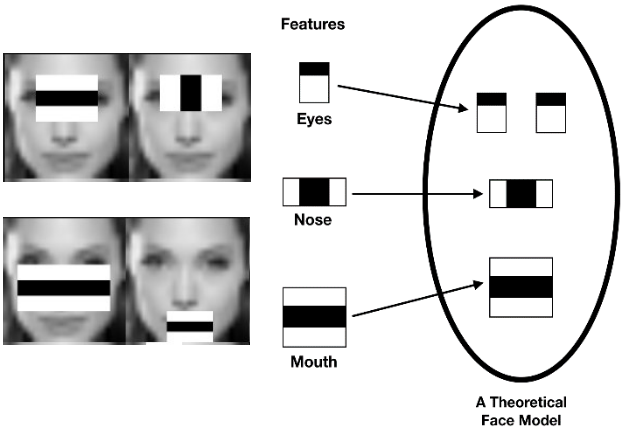
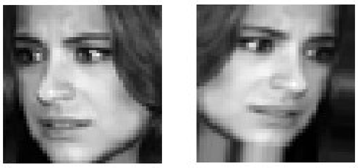


Fig. 8. Face Detection using Haar Cascades.

In Keras, we used ImageDataGenerator class to perform image augmentation as shown in Figure 9. This class allowed us to transform the training images by rotation, shifts, shear, zoom and flip. The configuration used is :
rotation_range=10, width_shift_range=0.1, zoom_range=0.1, height_shift_range=0.1 and horizontal_flip=True.



Original *Transformed*

Fig. 9. Image augmentation using Keras.

Then we defined our CNN model with 4 convolutional layers, 4 pooling layers and 2 fully connected layers. After that, to provide non linearity in our CNN model we applied the ReLU function and we used batch normalization to normalize the activation of the precedent layer at each batch and L2 regularisation to apply penalties on the different parameters of the model. Thus, we chose softmax as our last activation function, it takes as input a vector z of k numbers and normalizes it into a probability distribution. The softmax function is shown in Figure 10 :

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

Fig. 10. Softmax function.

To train our CNN model we splitted the database into 80% training data and 20% test data, then we compiled the model using Stochastic gradient descent (SGD) optimizer. At each epoch, Keras checks if our model performed better than the models of the previous epochs. If it is the case, the new best model weights are saved into a file. This will allow us to load directly the weights without having to re-train it if we want to use it in another situation.

V. Experimental Results

We trained our Convolutional Neural Network model using FER 2013 database which includes seven emotions (happiness, anger, sadness, disgust, neutral, fear and surprise) The detected face images are resized to 48×48 pixels, and converted to grayscale images then were used for inputs to the CNN model. Thus, 9 youthful master's students from our faculty participated in the experiment, amoung them there were two wearing glasses. The Figure 11 shows the emotions' results of 9 students. The predicted emotion label are represented with red text, and the red bar represents the probability of the emotion.

We achieved an accuracy rate of 70% at the the 106 epochs. To evaluate the efficiency and the quality of our proposed method we calculated confusion matrix, precision, recall and F1-score as shown in Figure 12 and in Figure 13, respectively. Our model is very good for predicting happy and surprised faces. However it predicts quite poorly feared faces because it confuses them with sad faces.

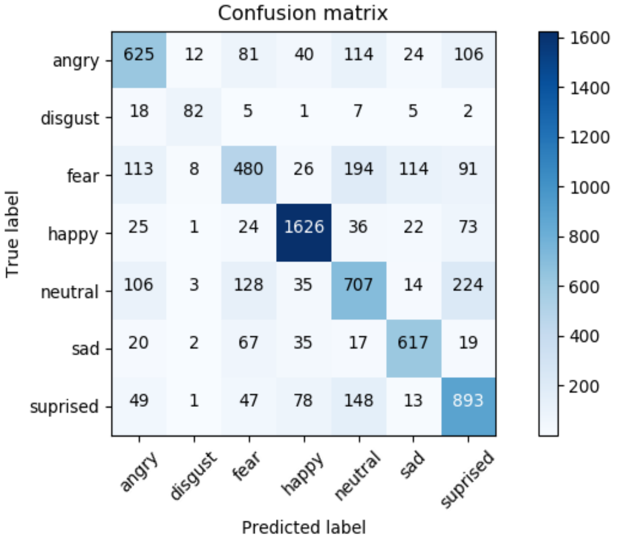


Fig. 12. Confusion matrix of the proposed method on FER 2013 database.

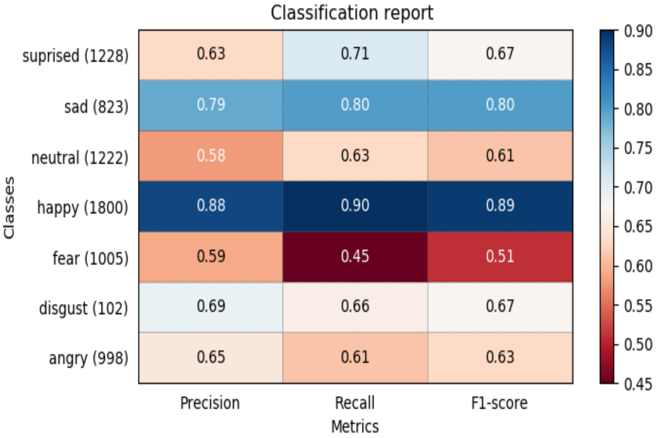


Fig. 13. Classification report of the proposed method on FER 2013 database.

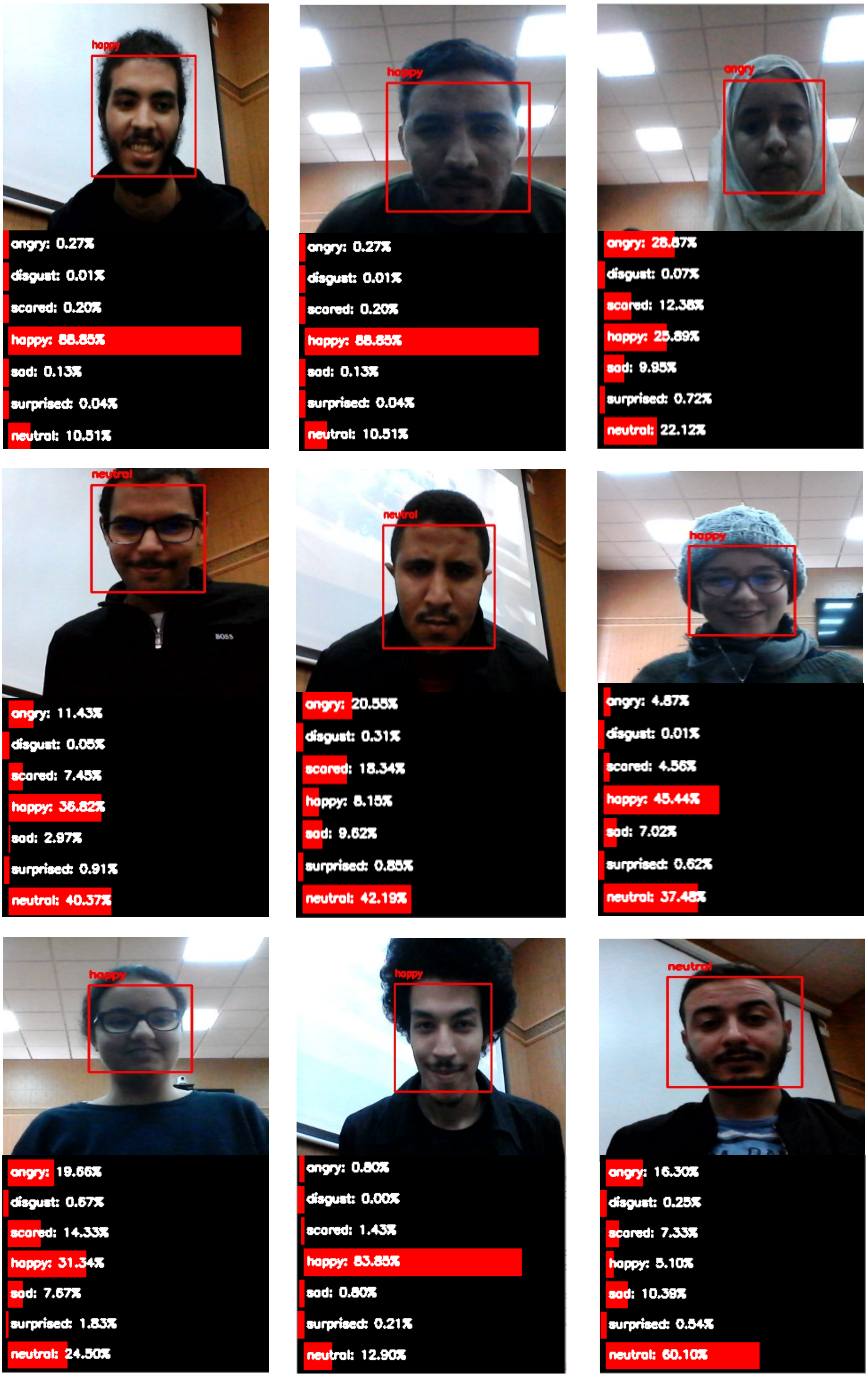


Fig. 11. Students' facial emotion recognition results.

VI. Conclusion and Future Work

In this paper, we presented a Convolutional Neural Network model for students' facial expression recognition. The proposed model includes 4 convolutional layers, 4 max pooling and 2 fully connected layers. The system recognizes faces from students' input images using Haar-like detector and classifies them into seven facial expressions: surprise, fear, disgust, sad, happy, angry and neutral. The proposed model achieved an accuracy rate of 70% on FER 2013 database. Our facial expression recognition system can help the teacher to recognize students' comprehension towards his presentation. Thus, in our future work we will focus on applying Convolutional Neural Network model on 3D students' face image in order to extract their emotions.

Acknowledgment

We wish to thank the 9 students for their participation in the experiment.

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] Opencv. opencv.org.

[17] Keras. keras.io.

[18] Tensorflow. tensorflow.org .

[19] aionlinecourse.com/tutorial/machine-learning/convolution-neural-network. Accessed 20 June 2019

[20] 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.

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