IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

Real-time Emotion Recognition using CNN-based Facial Analysis and Emoji Display

Naveen G[1], Devendra Sai Krishna M[2], Praneeth K[3], Rajesh K[4], Mahesh K[5]
Asst. Professor, Department of Computer Science and Engineering[1]
Students, Department of Computer Science and Engineering[2,3,4,5]
Raghu Institute of Technology, Visakhapatnam, AP, India

Abstract: *Deep learning approaches are currently having great success across several industries, including computer vision. Yes, it is possible to train a convolutional neural networks (CNN) model to analyse photos and recognise facial expression. In this study, we develop a system that can identify students' facial expressions of emotion. The three parts of our system are facial detection using Haar Cascades, normalisation, and emotion recognition using CNN on the FER 2013 database employing seven different expression kinds. The obtained results demonstrate that face emotion detection is practicable in education, and as a result, it can assist teachers in adapting their presentation to the emotions of the pupils.*

Keywords: Facial expression, Emotion extraction, Convolutional neural networks (CNN), Deep learning, Intelligent classroom management system

I. INTRODUCTION

The face is a person's most expressive and communicative feature [1]. with no words spoken, it can communicate a wide range of emotions. Facial expression recognition extracts emotion from a face image and uses it to determine an individual's mood and personality. The six universally recognised basic emotions—fear, sadness, anger, fear, happiness, surprise, and disgust—were established by psychologists of America Friesen and Ekman [2] in the 20th century. Due to its effects on clinical practise, sociable robotics, and education, facial expression recognition has received a lot of interest lately. Numerous studies have shown that emotion is crucial to education. Exams, questionnaires, and observations are currently used by teachers as a feedback of sources, however these traditional approaches are frequently ineffective. The teacher can modify their method and educational materials based on the students' facial expressions to support student learning. The aim of this paper is to employ Convolutional Neural Networks (CNN), a deep learning method that is extensively used in picture classification, to implement emotion detection in education by creating an autonomous system that analyses students' facial expressions. In order to extract feature representations, it consists a multistage image processing method. The three stages of our system are face detection, normalisation, and emotion recognition, which has to be one of the following seven emotions: neutral, anger, fear, sorrow, happiness, surprise, or disgust.

The remainder of this essay is organised as follows: The associated work is reviewed in Section 2. The system is described in Section 3. Section 4 presents the implementation specifics, while Section 5 contains the experimental findings and comments. We wrap up this essay with a discussion of potential future developments in our field.

II. RELATED WORK

The use of Face Emotion Recognition (FER) to enhance the learning environment is of great interest to researchers. A method that can analyse students' facial expressions and assess the effectiveness of classroom instruction was proposed by Tang et al. [3]. Data gathering, face detection, face recognition, facial expression recognition, and post-processing are the system's five steps. In this method, classification is done using K-nearest neighbours (KNN), while pattern analysis is done using uniform local gabor binary pattern histogram sequence (ULGBPHS). A online programme that analyses the emotions of students taking part in active face-to-face classroom education was proposed by Savva et al. [4]. The programme gathers real-time recordings using webcams installed in schools, and then analyses the data using machine learning techniques. Whitehill et al. suggested a strategy in [5] that uses students' facial expressions to

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

determine their level of interest. The technique uses Gabor features and the SVM algorithm to monitor how students engage with cognitive skills training software. The writers gathered labels from videos that the judges had annotated.
When students were interacting with an educational game designed to convey essential concepts of classical mechanics from a school computer lab, the authors in [6] employed computer vision and machine learning approaches to identify the effect of the pupils there.
The authors of [7] suggested a system that recognises and tracks students' emotions while providing feedback in real-time in order to improve the online learning environment for better content delivery. In an online learning environment, the technology uses head and eye motions to infer important information about students' moods. Facial Emotion Recognition System (FERS) was created by Ayvaz et al. [8] to identify students' emotional states and motivations during video conference style online learning. Four machine learning algorithms with the greatest accuracy rates are used by the system: SVM, KNN, Random Forest, and Classification & Regression Trees. employing KNN and SVM algorithms, results were produced. Kim et al. [9] proposed a system which is able of producing real-time recommendation allowing the teacher to alter their non-verbal behaviour, such as body language and facial expressions, in real-time, to improve the memorability and quality of their lesson.
In order to detect emotions, the authors of [10] suggested a model that uses facial emotion recognition with the Haar Cascades approach [14] to identify the lips and eyes on the JAFF database. In [11], Chiou et al. developed an intelligent classroom management system that helps teachers quickly change instruction modes to avoid wasting time. They did this by utilising wireless sensor network technology.

Fig. 1 Organisation of our suggested strategy.

III. PROPOSED APPROACH

In this part, we outline our suggested Convolutional Neural Network (CNN) architecture-based system for analysing students' facial expressions. Prior to cropping and normalising the faces to a dimension of 4848, the algorithm first distinguishes faces in the input image. These facial photos are then fed into CNN. The output, which includes findings for facial expression recognition (anger, happiness, sadness, disgust, surprise, or neutral), is the last step. Figure 1 shows the organisation of our suggested strategy.
In contrast to conventional image classification techniques, a convolutional neural network (CNN) is a deep artificial neural network that can recognise visual patterns from input images with little pre-processing. This implies that the filters, which were manually constructed in traditional techniques [19], are learned by the network. A neuron is a key component of a CNN layer. They are interconnected so that the output of one layer's neurons becomes the input of the following layer's neurons. The backpropagation algorithm is used to calculate the cost function's partial derivatives. Convolution is the process of creating a feature map from a source image by applying a kernel or filter. In reality, the CNN model has three different kinds of layers, as shown in Figure 2:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

Fig. 2 Three layers of CNN

3.1 Convolution Layer

The initial layer to feature-extract from a source image is the convolution layer. In the case of a ConvNet,Convolution's primary objective is to take the input image's features and extract them. Convolution learns visual features from tiny squares of input data, preserving the spatial relationship between pixels [21]. Between two matrices—one of which is an image and the other is a kernal—it does a dot product. Equation 1 illustrates the convolution formula:

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

Where input picture being x, w is the filter matrix, and is the convolution process, and net(t, f) being the result in the following layer.

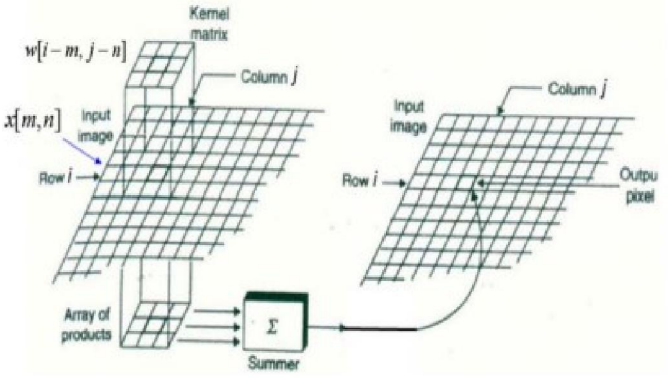


Fig 3 depicts the convolution in action

3.2 Pooling Layer

Minimises the size of each feature map while keeping the key details [21]. Max pooling, Sum pooling and Average pooling are three different types of pooling. The Pooling's goal is to gradually lower the input representation's spatial size and to render the network insensitive to minor distortions, translations, and transformations in the source image [21]. In our research, we used the block's maximum as the sole output to the pooling layer, as illustrated in Figure 4.

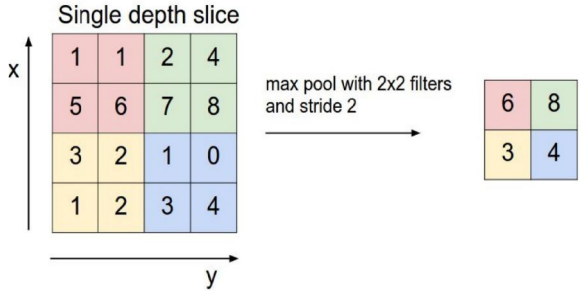


Fig 4 Maximum of the block as sole output

3.3 Fully Connected Layer

A conventional Multi Layer Perceptron that uses an active function in the output layer is the fully connected layer. The phrase "Fully Connected" suggests that all the neurons in the layer below is linked to every neuron in the layer above. The Fully Connected layer's function is to use the pooling and convolutional layers' output to classify the input image into different groups 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]

Figure 5 shows how our CNN model works. It consists of 2 fully connected layers, 2 pooling layers to extract information, 4 convolutional layers, and a softmax layer with 7 emotion classes. The input image is a 4848 grayscale

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

facial image. We utilised 33 filters with stride 2 for each convolutional layer. We employed the maximum pooling layer and 22 kernels with stride 2 for the pooling layers. As a result, we used the Rectified Linear Unit (ReLU), specified in Equation 2, which is currently the most popular activation function, to inject nonlinearity into our model.

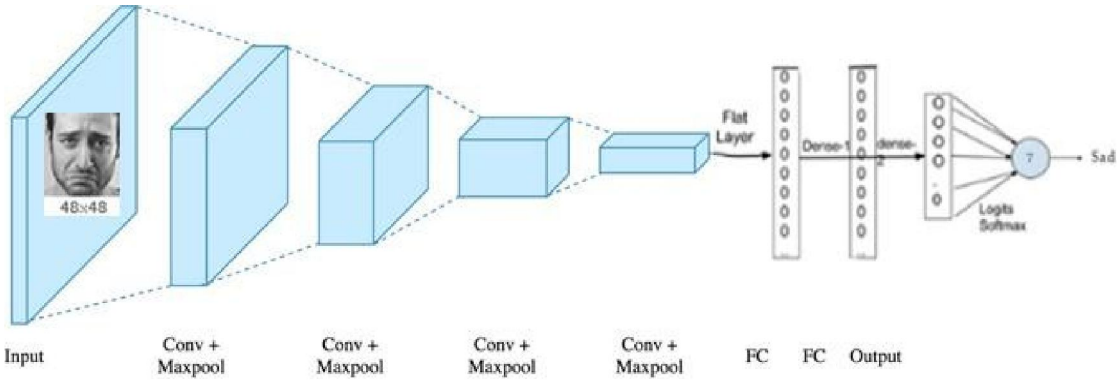


Fig 5Working of the CNN model

Figure 6 illustrates that when z < 0, R(z)=z, and when z is more than or equal to zero, R(z) is equal to z. The network configuration of our model is shown in Table I.

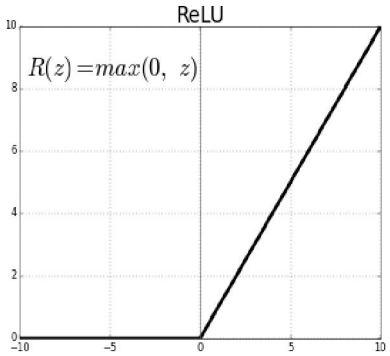


Fig 6 When z < 0, R(z) = z

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	-	-

Table I. CNN Configuration

IV. IMPLEMENTATION DETAILS

4.1 Data Acquisition

As seen in Figure 7, we used the FER2013 [12] database to train our CNN architecture. It was produced with the aid of the Google image search API and presented at the 2013 ICML Challenges. Automatic normalisation to 4848 pixels has been applied to all of the database's faces. 35887 images with seven emotion labels totaling 28709 training photos, 3589

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

validate images, and 3589 test images are included in the FER2013 database. Table II shows how many photos there are for each emotion.

Fig 7 FER2013 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

Table II shows how many photos there are for each emotion

4.2 Implementation of CNN

As illustrated in Figure 8, we used the OpenCV package [16] to capture live web camera frames and identify the faces of the students using the Haar Cascades approach [14]. The Adaboost learning method was developed by Freundetal. [15], who received the Gödel Prize in the year 2003, and is used by Haar Cascades. For the purpose of producing efficient classifier results, the Adaboost learning algorithm selected a few standout characteristics among a large number. Using the high-level APIs for TensorFlow [18] and Keras [17], we created a convolutional neural network model.

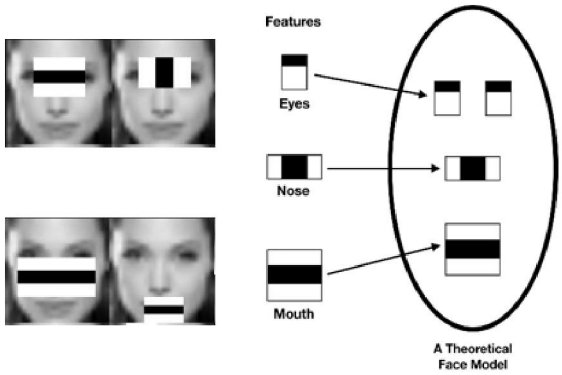


Fig 8 Face Detection using Haar Cascades

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

As seen in Figure 9, we enhanced images using Keras' ImageDataGenerator class. With the help of this lesson, The practise images may be rotated, shifted, sheared, zoomed in on, and flipped.
The settings are: horizontal_flip=True, rotation_range=10, width_shift_range=0.1, zoom_range=0.1, and height_shift_range=0.1.

Fig. 9. Image augmentation using Keras.

In the following step, we defined our CNN model with 4 convolutional layers, 4 pooling layers, and 2 fully connected layers. Then, to add nonlinearity to our CNN model, we applied the ReLU function, batch normalisation to make the activation of preceding layer consistent across batches, and L2 regularisation to impose penalties on the model's various parameters. As a result, we selected softmax as our final activation function; this function normalises a vector z of k numbers into a probability distribution.
We divided the database into training and test data for our own CNN model, then used stochastic gradient descent (SGD) optimizer to construct the model. Keras determines at each epoch if our model outperformed the models from the previous epochs. If so, a file is saved with the latest best model weights. If we want to utilise it in another circumstance, we won't have to retrain it because this will enable us to load the weights immediately.
For our CNN model, we split the database into training and test sets, and then we utilised the stochastic gradient descent (SGD) optimizer to build the model. At each iteration, keras assesses if our model outperformed the models from the prior epochs. If so, the new, ideal model weights are loaded in a file. Because this will facilitate us to save the weights right away, we won't need to retrain it if we want to utilise it in another situation.

V. EXPERIMENTAL RESULTS

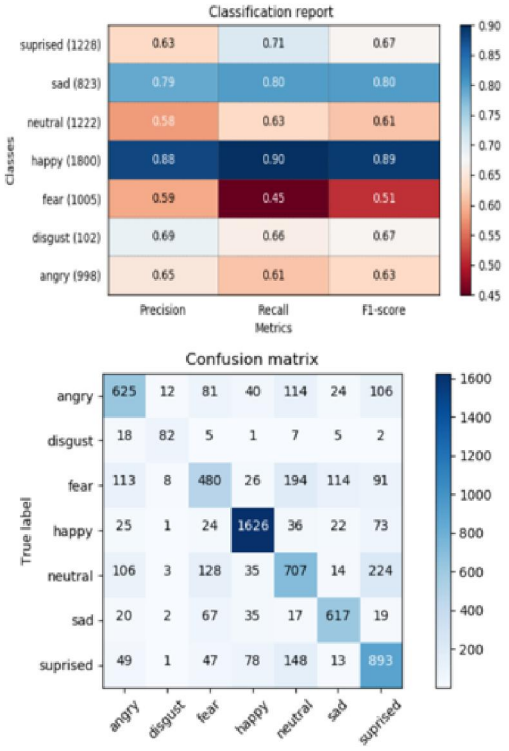
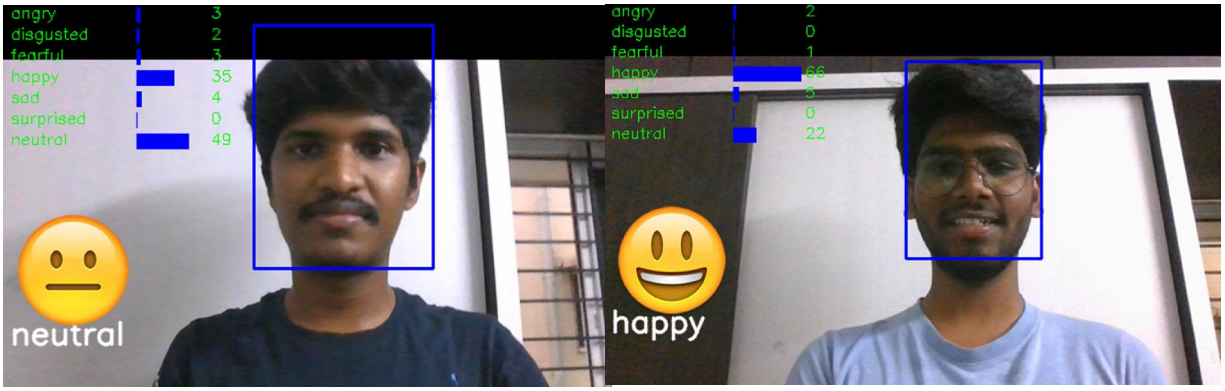


Fig. 10,11 Classification Report and Confusion matrix.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

In our pursuit of knowledge, we endeavored to train a CNN model using the FER 2013 database, which encompasses a spectrum of seven emotions: anger, disgust, happiness, neutral, sadness, fear and surprise. To prepare the images for analysis, we resized the detected faces to 48x48 pixels and converted them to grayscale.. After conducting an exhaustive 106 epochs, we accomplished an accuracy rate of 70%. To gauge the efficacy and caliber of our methodology, we meticulously computed a confusion matrix, precision metric, recall metric and F1-score as evidenced by Figure 10 and Figure 11. Our model demonstrated exceptional proficiency in predicting surprised and happy faces but was found wanting in its ability to predict fearful faces as it tended to confuse them with sad faces.



VI. CONCLUSION

Through the implementation of this particular manuscript, we have successfully crafted a Convolutional Neural Network (CNN) model using intricate design and thoughtfully chosen methodology. The primary objective of our proposed framework is to recognize the facial expressions of students. It comes with a robust design consisting of four convolutional layers, followed by additional four strata for max pooling and two fully connected layers. Our advanced technology uses a Haar-like detector to accurately differentiate faces in images obtained from the students. The system proficiently classifies these utterances into specific groupings, comprising Surprise, Fear, Disgust, Sadness, Happiness, Anger or Neutrality. Remarkably enough though, this categorization is achieved without any room for mistake or inaccuracy. Additionally, testing on the FER 2013 database provided an accuracy rate of 70%, which clearly demonstrates the efficacy of our approach. Our innovative model has paramount significance since it can reliably determine to what extent individuals comprehend novel information based on their current emotional state derived by analyzing their facial expressions meticulously. To advance our CNN approach further and needful desired precision with higher accuracy in identifying an array of emotions explicitly; we plan to use three-dimensional representation profiles. Thus conclusively speaking; it seems that employing this promising technology indeed matters in such an ever-changing world filled with cognitive possibilities making itself serve as a priceless perceptible outcomes assessment instrument

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]. 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.

[5]. 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.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 3, April 2023
Impact Factor: 7.301

[6]. 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.

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

[11]. Opencv. opencv.org.

[12]. Keras. keras.io.

[13]. Tensorflow. tensorflow.org .

[14]. aionlinecourse.com/tutorial/machine-learning/convolution-neuralnetwork. Accessed 20 June 2019

[15]. 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

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