Deep Learning Based Emotion Detection in an Online Class

Conference Paper · May 2022

DOI: 10.1109/DELCON54057.2022.9752940

5 authors, including:

Prabhjot Kaur
Maharaja Surajmal Institute Of Technology
55 PUBLICATIONS **659** CITATIONS

Deep Learning based Emotion detection in an Online Class

Prabhjot Kaur Sidhu
Department of Information Technology
MSIT, GGSIP University, New Delhi
prabhjot.kaur@msit.in,

Aditya Kapoor,
Department of InformationTechnology
MSIT, GGSIP University, NewDelhi
2000aditya28@gmail.com

Yash Solanki
Department of Information Technology
MSIT, GGSIP University, New Delhi
solankiyash9099@gmail.com

Piyush Singh
Department of Information Technology
MSIT, GGSIP University, New Delhi
piyushsingh2104@gmail.com

Devansh Sehgal
Department of Information Technology
MSIT, GGSIP University, New Delhi
sehgal.devansh7@gmail.com

***Abstract* — Day to day research work in the field of human emotion recognition have played a crucial role in bringing certain changes in domain of online marketing, education etc. Keeping the transition to online mode of education in mind the current paper aims to provide insights to the ability of students to perceived things and their response in online class by analyzing the facial expressions. The paper also trends to provide a review of the recent wok in the field of Human Emotion recognition. Emotion recognition is certainly not an easy task; it certainly requires taking into account various gestures and eye movements. There lies a huge scope in the field of human emotion recognition and the results provided by them are certainly not restricted to one domain. Current study used deep learning technology to detect real time emotions of students in online class using publically available dataset on Kaggle with the name JonathanOheix.**

Keywords—Deep Learning, OpenCV, Emotion Recognition, Image Processing, Online Learning, Real-Time image analysis

I. INTRODUCTION

Since the onset of pandemic Covid-19, there has been a significant shift in the things to an online virtual platform. The main transformation has been seen in the education domain whereby education institutions be it schools or colleges switched over to e-learning. Certainly, this did not hamper the flow of knowledge from the educators to the students, but has posed some serious concerns in the effectiveness of the content delivered by the educators. In an offline mode of education, it was easy to detect the student's emotions and their state of mind. But in an online mode it became difficult to track the students' attentiveness and state of mind. This paper throws light on the same issue. We used Keras to import all deep learning libraries and to undergo preprocessing of the data. Model was trained on the publicly available dataset with the name Jonathan Oheix on KAGGLE.

II. BACKGROUNDWORK

In the year 2019, Rajat Mehta, Rashi Bhardwaj and Prakash Ramani proposed a decent approach to understand the human emotion with a focus on expressions like yawning activity and detecting closed eyes **[1]**.Work was done upon publicly available dataset FER2013 where the authors used HAAR cascade for detecting frontal face and local binary pattern for facial recognition. The paper aimed at getting better results for student surveillance as well. Frontal Face HAAR cascade removes redundant rigorous algorithms of capturing images with image augmentation where images are rotated, reflected and converted into grayscale. LBPH was further used to train models hence forth storing the weights in .ymlfile. To detect the sleep and yawning activity the EAR (Ear Aspect Ratio) is to be calculated to measure the changes between the variousfacepoints.Foryawndetectionthedistancebetween the upper lip landmark and lower lip landmark points is calculated.Laterontocategoriesthehumanemotionsitused a CNN model consisting of 17 layers with ReLU and Softmax as the activation functions. Accuracy encountered here was around 61.3% with a confidence of over85%.

In year 2020, Kangning Yang et al. made an effort to compare facial recognition in an ideal situation and real world scenarios **[2].** Their main idea was to compare the pictures clicked in ideal situation with ideal camera positioning versus the photos clicked under normal conditions with an inappropriate camera positioning. Authors compared five crucial systems for emotion recognition such as Amazon Recognition, Baidu Research, Face ++, Microsoft Azure and Affectiva. A comparative study of all these different systems with different conditions were done like when images were rotated, with partially visible face, and when brightness was hampered, blurred or noise was added. Parameters such as Matching Score, Precision and F1 score were primarily used for the comparison. The authors mentioned that fear and angry expressions were the most difficult to recognize. Noise here was particularly identified in terms of electronic media or

2022 IEEE Delhi Section Conference (DELCON) | 978-1-6654-5883-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/DELCON54057.2022.9752940

internal sensors of the camera. The authors categorized this noise to be particularly obscuring image details, degrading its quality or even making the image completely unusable. At the end, the paper concluded that Microsoft Azure and Face++ showed greater reliability in terms of processing images with medium and high level noise strength.

In year 2020, Ali et al. proposed a method to detect facial emotions using Neural Networks[8].Here they have laid the emphasis on how the CNN have played a vital role in helping to depict and correctly identify the human emotions. Authors have used Viola-Jones algorithm to detect the eyes and lips region from face . Here the emotions are detected with the help of frontal facial images. Authors used SVM for image classification and ReLU as an activation function.

In the same year, Awais Muhammad et al. emphasized upon there lation of human emotions with the mental and physical health of human beings **[3].** Keeping Covid-19 outbreak in mind, the authors gave an IoT based solution by integrated it with AI especially in the field of healthcare. Here LSTM based emotion recognition is performed on the signals transmitted with the help of IoT based framework. The solution proposed in the paper achieved 95% performance efficiency. In the same year, Jaiswal, Shruti and Nandi proposed a model to detect human emotions using CNN**[4].** They tested the model on eight different datasets. Their model gave an accuracy of around 74%, which was an improved accuracy from the state-of-art accuracy with reduced computation complexity. This CNN model operated with 9 layers and ReLU as an activation function. However, it was concluded that the model had an accuracy of around 74% with a loss of nearly0.712.

In 2021, Bhardwaj et al. carried out research work in order to perceive human's emotions from Human face [5]. Here the authors made an effort to work and identify only 6 emotions with the help of CNN and experiment conducted on people of age group 18-85 years of age. Here the authors focused primarily on the regions around the eyes and mouth. The area was particularly divided into various territories .These areas were significant enough to detect the changes in the human expressions and emotions. In the same year Mei and Weihong have done a survey on deep face recognition practice [6]. In their paper, they have laid emphasis onthose papers where the deep learning techniques were used for getting the facial expression. Deep learning have played a vital role to stitch together all the pixels into facial demonstration. In the paper, authors have also discussed about different network structures and loss functions proposed in rapid evolution of the deep face recognition. In the paper, they discussed about miscellaneous scenes in deep facial recognition primarily about the cross-factor, heterogeneous, multiple media and industrial scenes.

III. PROPOSED FRAMEWORK

In this paper, we proposed an emotion detection model using Keras and Convolution Neural Network. We used OpenCV to test the proposed model. The model classified students' emotions in an online class into seven groups, which are Happy, Angry, Disgust, Fear, Sad, Neutral andSurprise.

A public data-set from Kaggle is used to train the model. The proposed model workflow is shown in the Fig. 1. asbelow:

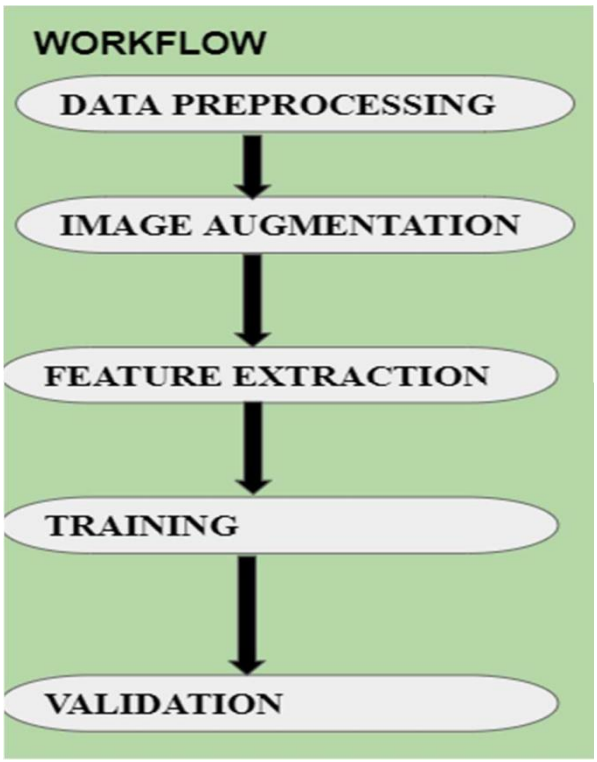


Fig. 1. WorkFlow of the proposed Model

A. Proposed Model

First step of the proposed model is data preprocessing whereby the data is prepared in a format to be accepted by the network [10]. This involves various processes from data extraction to data cleaning and loading. The main task is to make it readable by the network. After preprocessing the next step was image augmentation[11]. Image augmentation is done to generate ample amount of data from the existing one so as to train the model in a better way. This is done by either rotation or padding or various other techniques can be used for the same. Next important step towards facial expression is feature extraction[12].Here various numerical values are combined so as to reduce the data for computation and thereby making it convenient but describing the original data in the same way. In the Training Part the model is trained on the data collected so far [13]. After training the results are verified from the validation set.

Fig. 2. Classes showing different emotions

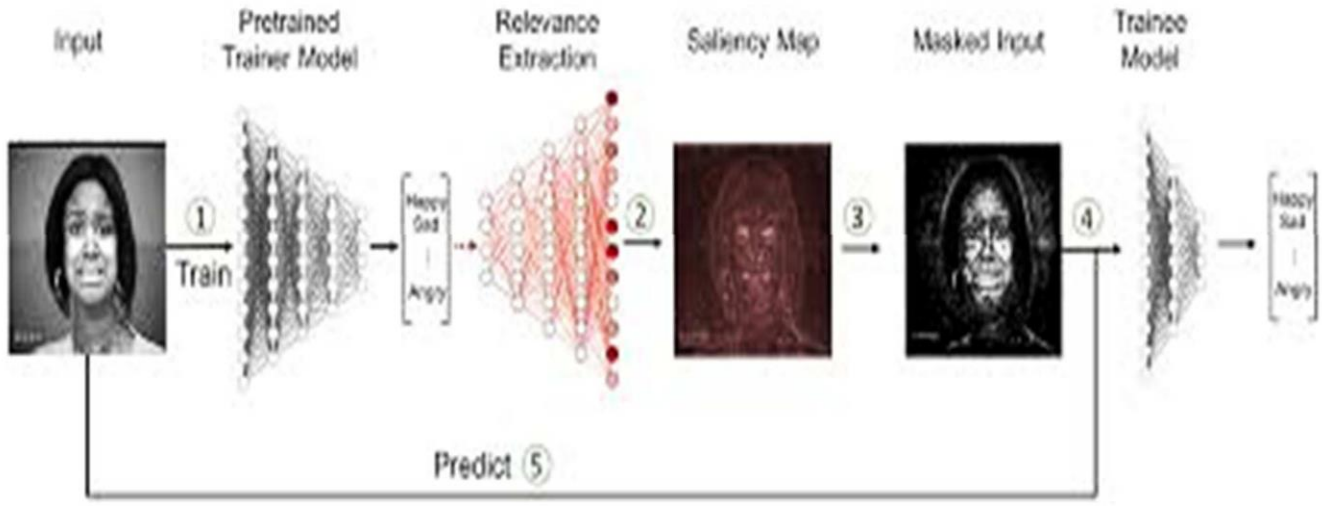


Fig. 3. Image processing through various layers

B. Dataset

The dataset of images with the name Jonathan Oheix was downloaded from Kaggle [7], which was provided by Jonathan Oheix**.** The dataset consisted of two separate folders for training and testing images having 7 classes of images as shown in the Fig. 2. Training set consists of 28821(Angry - 3993, disgust-436, happy-7164, neutral-4982, fear-4103, sad-4938, surprise-3205) and Testing set consists of 7066 images (Angry-960, disgust-111, happy-1825, neutral-1216, fear-1018, sad-1139, surprise-797). In total, the whole dataset has around 35.9k files with a size of around 56.51MB.

C. ModelTraining

The model was trained using Python language. The various libraries used in training the model are Matplotlib, Numpy, Pandas, Seaborn. All these libraries play a vital role in data analysis and data visualization. Various deep learning libraries are also imported using Keras. The model in an array format reads all the images. The CNN model is used to create layers in the model. To make the size of images uniform, it is converted into 48*48 in this case. In sample display, all the pictures are of standard size of 48*48. So before training the model, we took a batch size of 128 images. Batch size defines the number of images under consideration during one iteration, which means in one iteration our model will take 128 images. Model get the data from train and test directories respectively. There are approximately 28821 images for the training set and 7066 in the testing set.

After this the process of model building is started. The base of the model is Artificial Neural Network (ANN) that will classify the images into 7 different emotions of the students. Here, we have used the sequential model. It is one of the easiest ways of model building in Keras. Here, we have built this model layer by layer (Fig.3.).

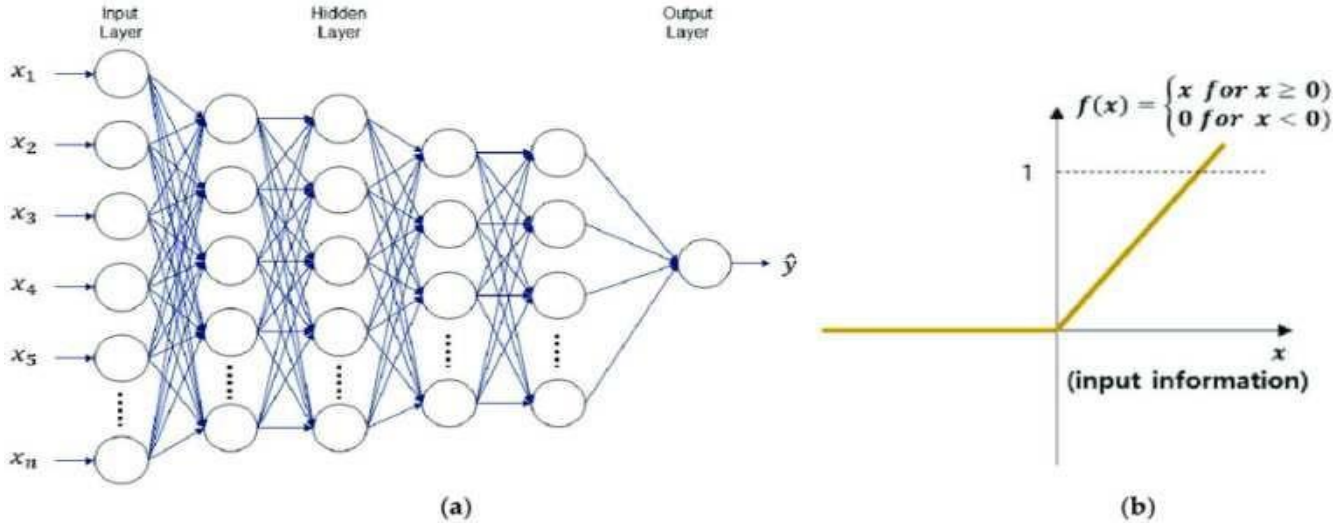


Fig. 4. Input Processing via layers and ReLU function

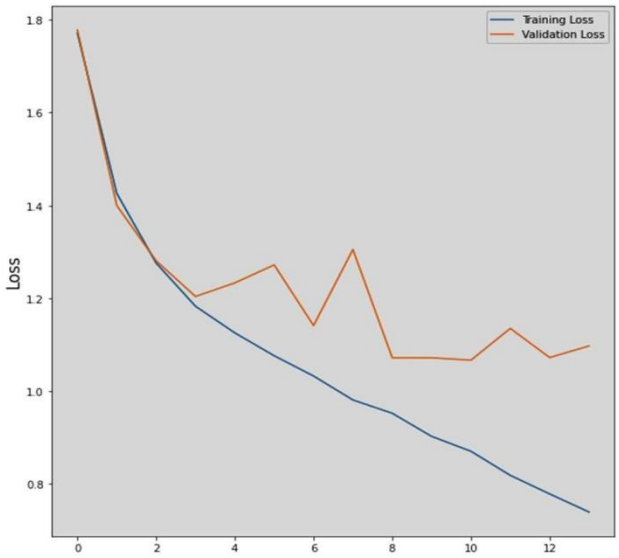
Every layer was defined with the function they perform in the model. All the layers combined with each other to form a deep neural network. In the first layer, we used 64 filters of 3*3 kernel size. Various other methods like padding have been used to manipulate the data so as to increase the data size. The convolution layer was followed up by the batch normalization, activation layer, MaxPooling2D and dropout layer. So the activation layer is basically involved in deciding what result should be shot out of neuron to be passed onto the next layer. The activation layer is present at the end. Here we used ReLU function as an activation function because it doesn't trigger all of them altogether (Fig. 4). Here we defined pool size as 2*2 for MaxPooling2D. This is done to extract important information from the area where we are putting that pool. The size of the pool is taken as 2*2. To prevent overfitting of the model, we used dropout. All the other convolution layers are designed in the same way. For better model design, the value of kernels and number of filters are changed in the consecutive layers.

The flatten layer basically collapses all results in one dimensional array so as to give input to the model. The dense function was used to connect all these layers together. The current model has used Adam's optimizer with a learning rate of 0.0001. Accuracy matrix is used to check the accuracy of proposed model. The three important functions like Model Checkpoint, Early Stopping, Reduced Ron plateau are used. Checkpoint basically checks our model for its functioning. Early stopping is used in the proposed model to reduce the convergence timing. The advantage of early stopping is that it stops the model in case the accuracy is not changing much. But it may result in overfitting. To monitor it, the model look upon the while loss part. It has also defined the reduced learning rate. Here, the learning rate is reduces to a smaller value by a factor of 0.2. Total 48 epochs are defined. After running the model, the early stopping at11 epoch was observed.

IV. Result And Simulation

Post the model training was done and when it was run through the data it gave an accuracy of around 67% with an early stopping at 11 epoch as there wasn't much change in the accuracy after 11 epoch. It has given the validation accuracy of 53%. Loss functions are plotted in Fig. 5. It is observed from the figure that accuracy increased gradually with the decreasing error. It was observed that the loss was decreasing with an increasing accuracy simultaneously.

To test with the online class, the model integrated with HAAR cascade and web camera using open CV file [9]. The model was tested with the class of 40 students. Emotions detected with the model are shown in Fig. 6. The results of the students' classification based on their emotions are shown in Table I.



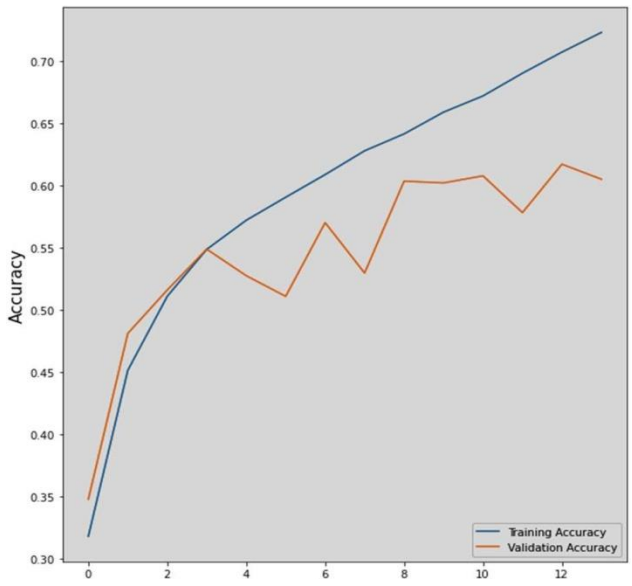


Fig 5. Decreasing Error curve and Increasing Accuracy Curve

Fig. 6. Emotion Classification of student

This experiment shows that in a class of around 40 students nearly 50% of the class had disgust feelings and 30% of class was sad. This interpretation can further be used to imply that the online mode of education had a significant impact on the students' learning ability and ability to grasp things. These results can be interpreted and motivated to change the teaching pedagogy and make the classes more interactive so that the output of the class shall increase and doesn't deteriorate the level of education.

TABLE I. STUDENTS' DATA CLASSIFICATION ON THE BASIS OF EMOTIONS

Emotion Identified	No of Students
Angry	0
Disgust	20
Fear	0
Happy	2
Neutral	6
Sad	12
Surprise	0

V. CONCLUSION

In this paper, a model based upon deep learning is proposed to detect the emotions of the students during online class. It was observed that the efficiency of students in an online mode of education was significantly low. There is a requirement of some relevant actions to be taken in order to make the online mode of education more productive and interactive. This further has a future scope of inclusion of more human emotions into consideration and integrating systems within online meeting platforms to give a real time analysis.

REFERENCES

[1] Mehta, Rajat, Rashi Bhardwaj, and Prakash Ramani. "Real-Time StudentSurveillanceSystemUsingMachineLearningAndComputer Vision." International Journal of Advanced Research in Computer Science, Vol. 10, Issue.4,2019.

[2] Yang, Kangning et al."Bench marking commercial emotion detection systems using realistic distortions of facial image datasets", The Visual Computer, pp. 1-20,2020.

[3] Awais, Muhammad, et al. "LSTM based Emotion Detection using Physiological Signals: IoT framework for Healthcare and Distance Learning in COVID-19." IEEE Internet of ThingsJournal,2020.

[4] .Jaiswal, Shruti, and Gora Chand Nandi. "Robust real-time emotion detection system using cnn architecture." Neural Computing and Applications, Vol 32, Issue15, pp. 11253-11262,2020.

[5] Bhardwaj, Shambhavi, Ashutosh Dixit, and Kshitij Yadav "A Study on Human Emotion Detection", 2021 International Conference on Advance Computing and Innovative Technologies in Engineering *(ICACITE)*. IEEE,2021.

[6] Wang, Mei, and Weihong Deng. "Deep face recognition: A survey." Neurocomputing, Vol. 429, pp. 215-244,2021.

[7] https://www.kaggle.com/jonathanoheix/face-expression-recognition dataset

[8] Ali, MdForhad, MehenagKhatun, and NakibAmanTurzo. "Facial Emotion Detection Using Neural Network." the international journal of scientific and engineering research,.2020.

[9] https://analyticsindiamag.com/my-first-cnn-project-emotion-detection-using-convolutional-neural-network-with-tpu/

[10] https://www.mathworks.com/help/deeplearning/deep-learning-data-management-and-preprocessing.html #:~: text=Preprocessing%20data%20is%20a%20common,that%20can% 20bias%20the%20network.

[11] https://www.analyticsvidhya.com/blog/2021/03/image-augmentation-techniques-for-training-deep-learning-models/

[12] https://deepai.org/machine-learning-glossary-and-terms/feature extraction

[13] https://www.techopedia.com/definition/33181/training-data#:~:text=Training%20data%20is%20an%20extremely,the%20training%20data%20is%20labeled.&text=The%20training%20data%20is%20an,learn%20and%20produce%20sophisticated%20results.

[14] Kaur, Prabhjot, Prakul Sharma, and Ankur Palmia."Fuzzy clustering-based image segmentation techniques used to segment magnetic resonance imaging/computed tomography scan brain tissues: Comparative analysis." International Journal of Imaging Systems and Technology 30, no. 4 (2020):1294-1323.