

Smart Teaching Using Human Facial Emotion Recognition (Fer) Model

Prof. Divya MN¹, Kashinath N², Bhagyashree S³, Chandini NS⁴, Shourya Jha⁵

¹School of ECE, REVA University (India)

²Shool of ECE, REVA University (India)

³School of ECE, REVA University (India)

⁴School of ECE, REVA University (India)

⁵School of ECE, REVA University (India)

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 10 May 2021

Abstract: Emotion recognition has attracted most of numerous technical/ non-technical fields due to its variety of applications such as entertainment, surveillance, psychology, marketing and few tech domains are some of examples. Emotion recognition is done based on some changes in human face, which we call them as regions of interest such as eyes, eyebrows, forehead, cheeks, mouth etc. On the other side, Online mode of education has also been blooming from nowhere especially after the Covid-19 pandemic, all institution including universities and training centers have been adapted to online mode of education, but on a contrary we have to agree to the fact that this mode of education is not as effective as traditional mode of education such as face-to-face [Classroom] teaching. So in this paper we try to provide a software solution to improve the online mode of education. We use Facial Emotion Recognition (FER) model to deploy into one of testing web application which streams live video from the students camera and can able to detect the emotions of students/attendees. We use FER2013 datasets to train our model and have used the google colab platform for testing of the model. we have obtained the good accuracy compared to our previous works. Presently we have developed our application for single device and have been using only LAN protocol, in future this mode of technology can be implemented at multiple device level using WAN protocol. With the help of already available app named "IP webcam" and a web application developed by us using FLASK we were successfully able to recognize the emotion of students in the application.

Keywords: Facial Emotion Recognition(FER), Smart-teaching, Convolution Neural Networks(CNN), Rectified Linear units(ReLU), FER2013.

1. Introduction

Facial Expressions are important carriers for human to convey emotions in communications. Study on non-verbal communication [1] says that almost 55% of the human communication might be intentional or non intentional is conveyed through Facial Expressions. The main aim of Facial expression Recognition is to successfully classify seven different emotions such as Happy, sad, anger, disgust, neutral, surprise and fear. Therefore the primary task in our paper is to classify these seven different emotions accordingly.

Face recognition technology has a slight edge on other technologies like finger-print technology, palm-print, iris reading technologies due to its non-contact process. Face recognition method is able to recognize the person from far distance without any interaction with the person. Additionally, due to the dramatically increased chip processing abilities (e.g., GPU units) and well-designed network architecture, studies in various fields have begun to transfer to deep learning methods, which have achieved state-of-the-art recognition accuracy and exceeded previous results by a large margin [2]. The challenge in facial facial expression recognition is to effectively locate and understand the facial regions of interest. In the past few years, these tasks were performed by traditional computer vision Methods, such as landmark detection, and object modelling. However, it was tacitly assumed that the recognition was performed in a controlled environment. Recently, several breakthroughs in image classification have been achieved by using deep Convolution Neural Networks (CNNs). these architectures consists of two main components: an automatic feature extractor and a classifier. The former produces low-level, Mid-level and high-level featured describing simple, moderate and complex textures respectively for the object of interest. Generally a strong classifier learns the target from a large number of high level features and thus a large amount of data should be used to train the network. Despite the powerful feature learning ability of deep learning, problems remain when applied to FER. First, deep neural networks require a large quantity of training data to avoid over fitting [3].

During the past few decades online mode of education has been growing rapidly at universities, schools, colleges, and even at training institutes, which aims the way for best use of Facial Emotion Recognition (FER) technology. We have to agree on the fact that, online mode of education isn't as effective as traditional mode of education. But wise utilization of technology can be used in improving the online mode of education(e-learning). When these FER models are deployed in online education platforms such as Zoom, Microsoft Teams

and streaming platforms like You tube etc. we think that teching mode will be more effective compared to normal streaming or teaching as the streamer or teacher will be able to see the overall performance of his/her teaching .By seeing overall emotion ratios of the students, teacher / streamer can change the way of his/her teaching. It is undeniable that the rapid growth of online education can effectively provide the convenience and flexibility for more students, so it also has broad development space in the future, therefore how to ensure that students keep the same level of concentration and learning efficiency as the traditional courses during the online education is critical to promote the further development of online education [4].

In this paper we use app named “IP Webcam” for our model testing and by using FLASK packages , we were successfully able to determine the emotions of students, even though we might not be able to detect all 6-7 emotions completely but we have worked in a way to provide software solution for the betterment of online mode of education..

2. Literature survey

In a research field of emotion detection, there is a contribution of several domains like machine learning, natural language, neuroscience etc. In previous works, they individually rummaged facial expressions, voice features, and textual data as universal indicator of emotions. One of the most used model is of Ekman's six basic classifications of all emotions [5] : Happy, Surprise, sad, Fear, Anger, Disgust however there's another category named “Neutral”. Ekman's model says that multiple emotions can coexist and this model is understandable to the beginners in this Facial Emotion Recognition (FER) Domain.. In later works are improved by combining the image, voice and textual data. The fusion can be done in three ways early, late and hybrid. Other ethos features the elements of emotion and the collaboration between emotional processes and other intellectual procedures[6]. According to the current research approach, experimental results shows that accuracy of the model might be higher but the overall time taken by the model to complete the trained task is high, such as Salman et al [7] who used decision trees to identify Facial expressions. Whereas some other solutions to the same experiment in FER had some limitations such as Chu et al [8] he has achieved less accuracy in less time as he has only extracted only lip feature on the face.

Computer Vision (CV) more often used as Open CV is one of the fields of study in machine Learning which helps the computer to study using different techniques and methods so that it can observe and capture what exist in an image or video being used. Most of the day to day applications are done by the help of Open CV such as Face Recognition technology in security areas, while grouping similar faces in some platforms like Google Photos, driverless cars as the Car has to drive autonomously by managing the traffic rate, surrounding cars etc. Even this Open CV is used in medical diagnostics [9]. We have agree to the fact that not all the faces looks similar , most of the faces have different shapes of their faces, so in such cases Face edging can also be used to complete the facial recognition by determining some points on the edges of the face. In addition, Facial texture is more often used for facial Localization ; Dai et al [10] proved facial localization using grey scale.

Dispite huge number of researches on Facial Emotion Recognition (FER) , still there are no comprehensive literature reviews on the topic of FER[11].some of the research papers have focused mainly on the conventional methods rather than approaching by deep-learning methods.recently, Ghayoumi[12] introduced a quick review of Deep learning in FER. But only reviews has been provides based on the differences between the deep leaning based approach and conventional approach. Most of the present day models are using Convolution Neural Network(CNN) for building the models of FER. As we can see from the past researches that CNN has achieved better accuracy even by using the small-sized dataset (EmotiW). CNN's application is not only limited to the facial expression recognition (FER) field, but when it is fine-tuned, it can be used to detect only some parts of faces instead of full face, we call it as regions of interest such as eyes,lips,mouth,eyebrows,etc. Moreover, the CNN works even more better over the already existing models on the multiple datasets such as FER2013, CK+, FERG and JAFFE [13][14]. It is observed that among most other advanced machine learning techniques being used convolution Neural Network(CNN) proved to be more efficient in terms o automated feature extraction, lesser input and classification accuracy[15]. On the other side, Liu et al [16] have proposed a model built using Convolution Neural Network(CNNs) it consisted of three Subnets with different structures. The model proposed by him had the accuracy of 65.03% on the Facial Emotion Recognition (FER) dataset.

In another approach, Haar cascades are being used for detecting eyes and mouth features combined with a neural network to provide much better emotion detection models[17]. Most of the methodologies of face recognition have used this labelled faces in the wild dataset for the training of their proposed solution . Shallow methods are based on the hand crafted features decided by humans on a local face image descriptor such as

scale-invariant feature transform , local binary patterns, histogram of oriented gradients etc[18]. with the help of



a system Facial Action Coding System(FACS), Facial actions have been classified into different Action Units (AUs) and emotions are classified using the collection of AUs. Works of Bartlett[19] report robust and correct detection of AUs leading to accurate emotion Detection. On the other hand, Zhou [20] combined the Gobar Transforms and Local Binary Patterns(LBP)[21] for facial feature extraction and achieved good performance . However , his algorithm had problem of high computational cost in the processing of the Gobar Transform.

The current day advancement in the technology has accelerated the work and research on Facial Emotion Recognition (FER). One of the best application of the FER technology is in e-learning which helps in personalized learning support. Whereby learning is being adapted to be personalized based on learner’s emotion to fit as per the learner suitability of learning[22].In regular classroom learning settings, emotion detection by head posture and facial expression detection can find the student’s attentiveness and synchronization rate [23]. While teaching over the internet, a teacher by using facial emotion detection technique can know the learner’s feedback remotely and thus adapting the teaching methodology likewise [24]

3. Proposed framework

CNN architecture

The networks are program on top of Keras, operating on Python, using the keras learn library. We extract images of size 48*48 from the dataset FER2013 and give them as input to the model. So the extracted images are read by the model and are being displayed as shown in fig 1. The figure 1 shows 7 different emotions. The network begins with an input layer of 48*48 which focuses on producing more accuracy and less loss at the end of the result.

This model contains convolution layer with 64 filters each with size of [3*3][7] followed by a local contrast normalization layer, maxpooling layer, followed by two more convolution layers, max pooling, flattening respectively .we have used the drop out of 0.4 to reduce overfitting and the activation function used is “ReLU”(Rectified Linear Units).

In our dataset , we pass the input images of size 3*3 to the first convolution layer.after that it passes through the normalization layer which helps in getting quality of feature maps by reducing the average of neighboring pixel, followed by ReLu activation function.and after the first layer, second layer is also added simultaneously to reduce the dimensionality of the CNN model, followed by softmax function. Even we use drop out , adam optimizer for building a more accurate model.

The structure of our convolutional model is shown in fig 2.

Fig 1. sample of images read by the model showing 7 sets of 7 different emotions

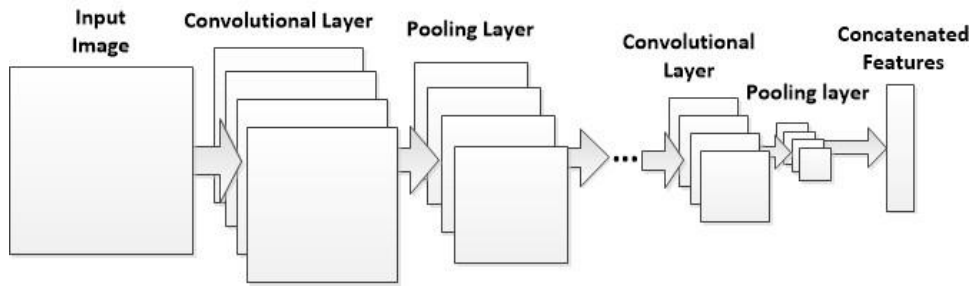


Fig2. Structure of Convolution layer

4. Experiment details

Datasets

Neural networks require large number of data to be trained , especially this type of model requires tens of thousands of images to be trained to get required accuracy of the model. And also type of images we use for the training is responsible for a larger part of the eventual models performance. So better quality and quantitative images needs to trained. So far so many datasets are available ranging from 1000 high quality images to tens of thousands of related images grouped under the datasets. So in this paper we have used Facial Emotion Recognition 2013 [FER2013] dataset for consideration which contains 35887 images in total with 7 different emotion like anger, surprise, happy,sad,disgust,fear,neutral.which are distributed as shown in fig 3. but the dataset is not uniformly distributed second emotion as we can see in the figure that “Disgust” is too less in number which makes the trained model difficult in identifying this particular emotion. The dataset used is available as cleaned only, so we don’t need to clean the dataset to check the presence of any null values. While the FER2013 set displays in the wild emotions . This makes it harder to interpret the images from the FER2013 dataset. But given the large size of the dataset a models robustness can be beneficial for the diversity.

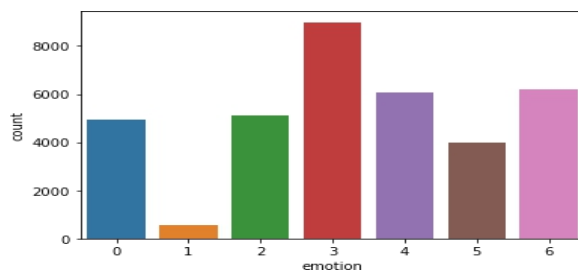


Fig 3. distribution of emotions in the database

1. Angry
2. Disgust
3. Fear
4. Happy
5. Sad
6. Surprise
7. Neutral

Online education platforms

The present day technology as advanced soo much that we have been able to overcome the traditional educational system like “Gurukul system”, current technology has introduced many platforms like Zoom, MS teams, DingTalk, Rain classrooms etc. But we have to agree on a fact that, current day online education system isn’t effective as classroom education. Teachers wont get to know the effectiveness of their teaching in the online platform but they can see visually in the classroom teaching. So our model helps in solving this issue, so far we haven't used any current platforms for consideration , we have used soft wares for testing our model, unfortunately we have succeeded to some extent in this prospect. We have used web application , and a already developed app named “IP webcam” which uses same IP address as the host and LAN network , and the IP address of the student is used to live stream the video, by doing so , teacher/presenter will be able to recognize any change in the emotion of the student/attendee.

Training details:

We train the model for more than 60 epoches to get the more accuracy, the model is trained on large set of data(images), we gave used google colab platform for training our model. We use the GPU for faster processing of the data. So once the testing of the model is done, it shows us the best accuracy and validation loss for the particular epoche, the accuracy of our model is slightly higher than the any previous models which are

built using this dataset, in turn which shows us that, using deep learning convolution neural networks for training improves the overall performance of the model.

Now once the training is done, we dump this code into the web application which is designed by us for testing of our model , we save the model during the training of model, so that it can be used in the web application of ours, we load the saved model which contains the trained data ,is now being used to recognize the real time human emotions in the live video.

Summary of our model is as follows:

dropout_3 (Dropout)	(None, 2, 2, 512)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 512)	1049088
dropout_4 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_5 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_6 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 7)	903
=====		
Total params: 5,905,863		
Trainable params: 5,902,151		
Non-trainable params: 3,712		

5. Experimental results and comparison

We have trained the model and successfully able to detect some of emotions, and obtained the validation accuracy of 65% And validation loss around 1.08% which isn't satisfactory. The dataset we have used had some cons, such as data isn't distributed equally, some missing Images in between, not clear images etc etc.. Still we have managed to get the accuracy of 65%, and once we had obtained the output of the model, we have Used the same for further web application which we had stated earlier Which involved smart teaching method, here also, we are able to Successfully determine the emotions from the students, teacher/presenter is Able to determine the emotions of the students. Below figures Fig 4. Graph of training accuracy vs validation accuracy Fig 5. Graph of training loss vs validation loss , explains the Results of accuracy and loss of our model which is better compared to previous other models. Fig 6 shows the demo of web application window which shows us the real time emotions of the students or attendees [Loss: 1.0873197317123413 , Accuracy: 0.6500418186187744 (65%)].

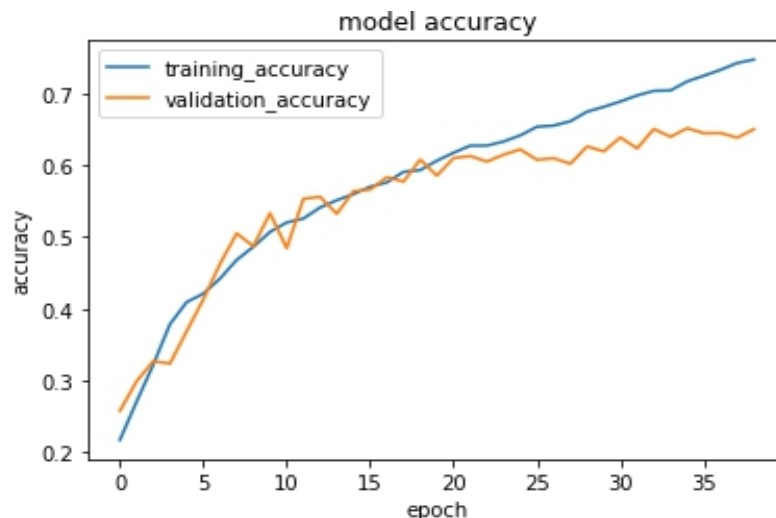


Fig4. Graph of Validation Accuracy and Training Accuracy

Fig5. Graph of Training Loss and Validation Loss

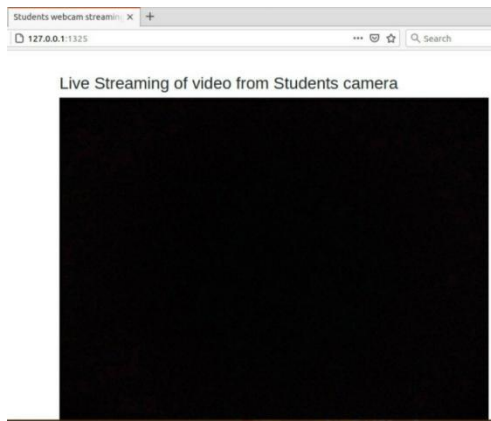
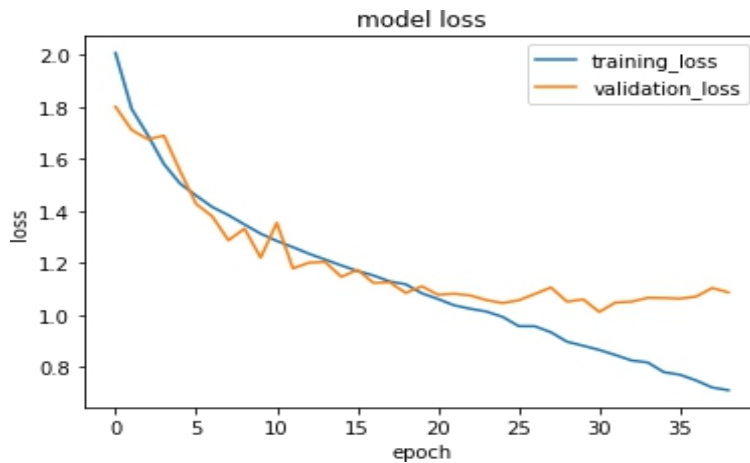


Fig 6. Demo of blank screen of web application

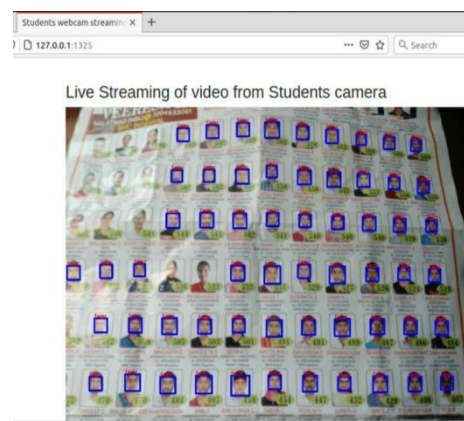


Fig7. Demo of web application Screen when it is in use

6. Conclusion

With the advancement in technology, education system has seen a tremendous improvement, with the help of online education system, students can learn from any corner of the world and in the same way, teachers can Teach from any corner of the world. But online teaching isn't as effective as classroom teaching, it has some drawbacks like efficiency of the teaching will be less. Teachers can't monitor students all the time, teachers wont get to know the amount of interest the students are paying to their teaching, so in order to eradicate or reduce such problems we have developed a system which lets the teachers to know the emotions of individual student and they can change or modify the way of teaching according to the interest of the student in order to increase the effectiveness of the student.

There are totally 7 standard emotions namely anger,sad,happy,disgust,fear,surprise,neutral, oursystem or proposed model might not able to determine exact all 7 emotions, but when this method isbeing Developed in a more professional way, we are sure this will be a tremendous achievement in thefield of online education. Currently we haven't used the actual platforms which are being used for online teaching but we have created our own web application for testing of the data, through which, student will be installing a app named "IP webcam" in his/her mobile/tablet, and as for the teacher's side, we have created application which helps in seeing the student's live video along with the emotions of involved students being shown on their face, so to conclude, this is one the best way of using our technology in making online education system more and more effective than classroom teaching, especially during the times of pandemics, online teaching effectiveness plays very important role.

7. Future work

Proposed model focuses only on a single student or device, this model can be improved for multi device or for a classroom with many students present in it so that the teacher can get to know the overall performance of his/her teaching. Even when comes to the dataset used in our project, it has some drawbacks which we need to work on to make it a better dataset , it can be done through feature extraction, data augmentation methods etc.

Major setback in our project is , we used the LAN network between the teacher and student , which poses the question what to do when they are miles away, at that point of time we make use of WAN network or server facilities to improve our work. Further works include implementing this model or application in actual streaming platforms like ZOOM, MS teams etc so that we can take best use of our application in current online teaching platforms. Our best interests aim for the betterment of the effectiveness of the online education system.

8. Acknowledgements

We thank the Director of ECE , REVA university , Dr R.C Biradar, Prof. Divya MN and everyone who helped our team.

References

1. G. Mehrabian, *Nonverbal Communication*. New Brunswick, NJ, USA: Aldine, 2007
2. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
3. Shan Li and Weihong Deng, "Deep Facial Expression Recognition: A Survey" DOI 10.1109/TAFFC.2020.2981446, IEEE Transactions on Affective Computing.
4. Weiqing Wang, Kunliang Xu, Hongli Niu, and Xiangrong Miao, "Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation" Volume 2020, Article ID 4065207, 9 pages <https://doi.org/10.1155/2020/4065207>
5. P. Ekman, W. V. Friesen, M. O'Sullivan, A. Chan, I. DiacoyanniTarlantzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712, 1987.
6. Md. Forhad Ali, Mehenag Khatun, Nakib Aman Turzo "Facial Emotion Detection Using Neural Network".*International Journal of Scientific & Engineering Research* Volume 11, Issue 8, August-2020 ISSN 2229-5518
7. Salman, Madaini, & Kissi (2016) " Facial Expression Recognition using Decision Trees". IEEE International Conference on Computer Graphics, Imaging and Visualization (CGiV), 120-130.
8. Chu, Chen, & Hsihe (2015) " Low-cost Facial Expression on Mobile Platform", IEEE International Conference on Machine Learning and Cybernetics (ICMLC), pp 586-590,.
9. <https://analyticsindiamag.com/my-first-cnn-project-emotion-detection-using-convolutional-neural-network-with-tpu/>
10. Dai, & Nakano (1996) "Face-texture Model Based on SGLD and its Application in Face Detection in a Color Scene", *Pattern Recognition*. 29(6), 1007-1017.
11. Byoung Chul Ko A Brief Review of Facial Emotion Recognition Based on Visual Information 2018 Feb; 18(2): 401. Published online 2018 Jan 30. doi: 10.3390/s18020401
12. Ghayoumi M. A quick review of deep learning in facial expression. *J. Commun. Comput.* 2017;14:34–38. [Google Scholar].
13. S. Minaee and A. Abdolrashidi, "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network," arXiv, no. 1902.01019, 2019.
14. M. M. Taghi Zadeh , M. Imani and B. Majidi , "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," in 5th Conference on Knowledge-Based Engineering and Innovation, Iran University of Science and Technology, Tehran, Iran, 2019.
15. P. R. Dachapally, "Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units," arXiv, no. 1706.01509, 2015.
16. T. D. Sanger, "Optimal Unsupervised Learning in Feedforward Neural Networks," *Neural Networks*, vol. 2, pp. 459–473, 1989.
17. D. Yanga, A. Alsadoona, P. Prasad, A. K. Singh and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," in 6th International Conference on Smart Computing and Communications, ICSCC 2017, Kurukshetra, 2017.
18. M. Z Khan et al: Deep unified model for face recognition based on cnn and edge computing.DOI:

- 10.1109/ACCESS.2019.2918275 (<https://ieeexplore.ieee.org/document/8721062>)
19. M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2005.
 20. CONFERENCE ON FAKE NEWS DETECTION TECHNIQUES, Dr. D. Kavitha, Mrs. Y. Padma, Mrs. J. Sirisha, International Journal Of Advance Research In Science And Engineering <http://www.ijarse.com> IJARSE, Volume No. 10, Issue No. 01, January 2021 ISSN-2319-8354(E).
 21. Q. Zhou and S. Zhang, "A Comparative Study of Geometry, Gabor Wavelets Representation and Local Binary Patterns for Facial Expression Recognition," *Adv. Biomed. Eng.*, vol. 11, p. 200, 2012.
 22. Mohd. AbdulMuqet "Local binary patterns based on directional wavelet transform for expression and pose-invariant face recognition. <https://doi.org/10.1016/j.aci.2017.11.002>
 23. Saurabh Pal ,Pijush Kanti Dutta Pramanik Anand Nayyar Prasenjit Choudhury "Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System DOI: 10.1145/3385209.3385231
 24. K. Fujii, P. Marian, D. Clark, Y. Okamoto and J. Rekimoto, "Sync Class: Visualization System for In-Class Student Synchronization," in 9th Augmented Human International Conference, Seoul, 2018
 25. A. Sun, Y.-J. Li, Y.-M. Huang and Q. Li, "Using facial expression to detect emotion in e-learning system: A deep learning method," in International Symposium on Emerging Technologies for Education (SETE 2017), Cape Town, South Africa, 2017.