# Deep Learning Based Hybrid Approach For Facial Emotion Detection

Dr.S.Suriya[1] and Babuvignesh C[2]
{suriyas84@gmail.com[1], babuvignesh.c@gmail.com[2]}

Associate Professor, Dept. of Computer Science and Engineering, PSG College of technology, Coimbatore, India[1], PG scholar, Dept. of Computer Science and Engineering, PSG College of technology, Coimbatore, India[2]

**Abstract.** Humans exhibit an essential set of feelings which might be exhibited via steady facial expressions. Emotions have an important role in teaching as well as other disciplines. It is understood that, while culture, the environment, the language and behavior of people vary, emotions are considered to be universal among the entire human population. Emotions are also important in the realm of education, because their power rivals that of the environment or the language used to communicate information. It's important for cognitive functions including learning and digesting new information.Student emotions during lectures play an important role whether it's in classrooms or in virtual learning environments. Unfortunately, due to the adoption of a virtual method of study or simply the incapacity of humans to keep track of all students in a classroom, teachers do not always keep students' emotions in check, making it difficult to adjust and keep the line of communication open.Thus automated facial emotion detection is very useful and brings the hidden indicators of students internal emotions to limelight. The suggested method aids in the detection of emotions and their classification in order to detect interest in a topic, which is then used as feedback to the course management staff in order to improve learner experience and aid in course material updates.

**Keywords:** Emotion Recognition, FaceDetection, Haar Cascade, Deep Learning, CNN and SVM, Computer vision, Image Processing.

## 1 Introduction

Emotions from face plays a vital role in getting true feedback from students during lectures. During the unprecedented time of covid pandemic everything has moved to virtual environments. The teaching and learning process is affected greatly due to the fact of lack of physical presence. It’s quite easier for an instructor to observe the emotions of students from their face during the physical lectures whereas in virtual environments it’s quite challenging. Without this feedback from emotions instructor’s tend to go in the same pace of the course which may often end up with the students being bored for the entire course/lecture. Thus Facial emotion detection of students will help in analysing the emotions of the students which in turn can be used as a feedback by the instructor to change the pattern or approach of the teaching. The Facial Emotion detection can be done in real time whereinstructors get the emotion recognition analysis data from students based on real-time image analysis and can alert the instructor when majority of students feel bored or can be used as a non-real time

ICCAP 2021, December 07-08, Chennai, India

DOI 10.4108/eai.7-12-2021.2314585

analysis tool and change the course content or pattern of teaching. Hence this study aims to provide a robust design of face emotion detection system with better accuracy and efficiency.

## 2 Literature Survey

Face Detection and emotion recognition is an interesting research area where many emerging technologies have been used.Jain et al.(2021) propose the use of emotion and speech recognition through Deep Learningas feedback from students [1]. From the video clips, using cluster based techniques the important key frames are extracted. Using HAAR Cascade Method the face is detected from which emotions are classified using SVM along with CNN. The demerit of such a system is that high resolution camera feeds are required for the system to work efficiently. The accuracy can be further increased by using LBP for feature extraction. Using the algorithm's facenet and facial landmarks[2], Saleem et al.(2021) investigated the suitability of these algorithms for face identification using facial features for crime detection, verification, and human tracking. The ROI is mapped to the person once the features have been extracted and pre-processed in order to identify them. The Euclidean distance from one eye to another, the anatomy of the nose, the distance between the two ends of the lips, and other factors were all taken into account.The dataset used hereduring training is not adequate and the performance can be increased if the training set is increased. Also for applications like crime detection, emotion analysis would be helpful.

Mishra et al. (2021) propose the use of multiscale parallel deep CNN (mpdCNN) Architecture for face recognition of low resolution images for monitoring purpose[3]. The accuracy of the proposed architecture is 88.6% on the SCface database for low resolution images which is a good improvement than other architectures being used. Also for high resolution datasets it has achieved an accuracy of above 99%.The proposed architecture is good for long distance images and can be utilized in commercial face recognition systems like CCTV surveillance. Using "Hybrid Robust Point Set Matching Convolutional Neural Network" (HRPSM_CNN) Tamilselvi et al. (2021) propose a face recognition system under unconstrained situations[4]. The traditional methods are capable to work only with specific directions of the image whereas HRPSM_CNN is capable of detecting faces in unrestricted situations. The accuracy is further increased by using robust point set matching along with CNN. The drawback of this paper is that the analysis is done with minimum dataset having closer shots of images which can be trained and tested for large datasets with reasonable distance. Emotion recognition is proposed by Aarika et al. (2020) as a technique for boosting student performance and learning approaches[5]. A comparison of various approaches to face emotion detection is explored, as well as a basic architecture of facial expression recognition. Future work can include the various criteria in which the selection of techniques for extraction and classification is done.

Tonguç et al. (2020) propose the use facial expressions for automatic recognition of student emotions[6]. The "facial movements coding system" (FACS) was utilized to analyze facial expressions in this study, and these emotions were then translated to digital data using the Microsoft Emotion Recognition API. The research outcome shows that happy feeling decreased after first stage of the lecture whereas anger, fear and confusion increased. While during the closure section of lecture, Happiness increased rapidly. Using API's does not allow flexibility to change the underlying methodologies used to detect the facial emotions and it can be done using own model trained. Face recognition classifier based on deep convolutional networks[7] proposed by Goel et al.(2020) integrates the deep networks' feature extraction capabilities and the quicker speed of the “Kernel Extreme Learning Machine(KELM) classifier” allows it to find characteristics that are not affected by outlierslike lighting, posture, and so on.Overfitting and local minima are avoided by gradient descend approach . The suggested DC-OKELM provides improved classification results with the best accuracy while requiring the least amount of training time. For the ELM classifier, optimization of weights is required which need to be considered.

Patil et al.(2020) propose the use of image processing and machine learning for automatic student attendance marking system[8]. Using face recognition under controlled environments the attendance is marked on excel sheet once the face is recognized. The Viola-Jones object detection technique is utilized for face detection, whereas LDA, KNN, and SVM are employed for recognition. Analysis showed that LDA with KNN provides better results than LDA with SVM. Model trained and tested with a small dataset of 150 images and hence the accuracy would greatly vary in large datasets. Also the time taken for the facial detection is not discussed which must be considered as these systems would need near real-time classifications which is difficult with models considering many facial features. A method for recognizing facial emotions using machine learning [9] is proposed by Ninaus et al.(2019)in game-based learning. SVM which is multivariate and parameter free procedure is used and the sampling size is determined by using G*Power 3.1. It was discovered that in game-based learning, the emotional involvement rose significantly as a result of appealing visual aesthetics and virtual incentives. The approach's disadvantage is that video must be manually cut before analysis, adding to the system's workload.

Yang et al.(2018) propose facial emotion recognition in virtual learning environment[10]. Instead of considering only the optimal range of eyes, mouth features are also added using Haar Cascade, an object detection algorithm used for face recognition in real time videos or images to improve accuracy. Once face is detected in the image, eyes and mouth are cropped, edge detected using Sobel method followed by feature extraction. The emotions were classified into six categories using CNN. The posture of the face is not considered and the model is trained using the dataset with clear facial images, which will not be the case with commercial implementation where the lighting and clarity of images differ. Normal images taken by system would always be of different postures. Also the range of human eyes differ when the distance of the object to the camera changes when the image is taken. This must also be taken into account. An analysis of physiological, environmental, and geographical sensor data using deep learning[11] is proposed by Kanjo et al. (2018).

Emotion classification is done by iteratively adding sensor data’s and removing some sensor data which are not important from different sources like smartphones and other wearables. “Convolutional Neural Network and Long Short-term Memory Recurrent Neural Network (CNN-LSTM)” is used on the raw sensor data for emotion. Although utilizing multi-sensor data gives good results in the emotion classification it comes with the disadvantage of mandatory usage of wearables(for sensor data) which cannot be always ensured when it comes to systems like Emotion recognition during lectures. Also data from such sources has it’s own precision difference which further affects the accuracy of the system.

Bah et al.(2019) propose to useLocal Binary Pattern (LBP) to enhance the face recognition accuracy along with contrast adjustment, bilateral filter, histogram equalization, and picture mixing and other image processing techniques. The issue of occlusion is not addressed in this study caused due to poor lighting, posture and masked face in face recognition. Student emotion recognition system(SERS)[13] is proposed by Krithika et al.(2016) for virtual learning environment based on learner concentration metric. The system will identify emotions of student and give live feedback to the instructor. It is implemented in MATLAB with Viola Jones and LBP algorithm functions. Abnormal head movements is considered  and the system understands that the person is no longer interested towards the lecture being taught. One good aspect of such a system is that there is no need for any physical wearables to be used by the students. It must be noted that only movement of head and eyes is considered as distractions and emotions are classified based on these movements. This always does not provide correct prediction of emotions and one should always consider more aspects in order to get further insights on the interest of the student on subject content.

Chickerur et al.(2015) proposed the use of 3-dimensional facial model dataset for facial emotion recognition[14]. The dataset is built by capturing a 3-dimensionalimageusing Kinect camera and is made available publicly. Facial expressions were broadly classified into 5 categorical emotions namely normal/neutral, happy, sad, surprised and angry. All the facial expressions were captured for a particular student in order to avoid distractions and then moved to next person. Capturing a 3D model with 3D camera's like Kinect has its disadvantage of cost as well as the positioning of the camera must be proper to ensure correct and relevant details are obtained. Shape and appearance features are used to classify facial expressions[15] as proposed by Routray et al.(2015). To create a hybrid feature vector, this study uses a hybrid combination of form and appearance features. Shape descriptors and appearance features are represented by the "Pyramid of Histogram of Gradients (PHOG)" and "Local Binary Patterns (LBP)", respectively. From the facial patches available, active patches are taken using localization which improves the performance of the system, lowers the cost of categorization and prevents over-fitting of characteristics. The LBP's recognition rate decreases for recognition under varying pose and illumination. Also for active facial patch detection, facial landmarks such as eyes, nose and mouth must be identified earlier which is an additional overhead.

## 3 Existing system

From the above literature survey we conclude that Viola Jones, AdaBoost and Eigenface values algorithms are used for face detection and for face emotion recognition LDA, PCA along with combinations of  CNN, SVM and KNN are used as shown in Figure 1.

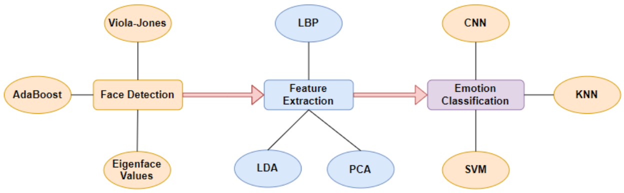

**Figure 1 – Existing System**

The Viola-Jones is an object detection framework which is proposed for Real-time face detection that may also be trained to recognize other objects. One shortcoming of this algorithm is thatfull frontal upright faces are required. To be identified, the full face must face the camera and must not be rotated to one side or the other. While this limitation appear to limit the algorithm's applicability, because the detection phase is usually followed by a recognition step, thisshortcoming on posture is tolerable in practice. This algorithm typically has 2 phases namely Training and Detection. Any image given to the algorithm is shrank to a size of 24 * 24 and during the training phase it requires large dataset in order to identify features in different forms. Also it requires considerable amount of non-facial images in training set to identify the negative class.

Adaptive Boosting aka Adaboost comes under ensemble learning which takes a collection of weak classifiers along with the training set as input and combines them to produce a strong classifier. Adaboost supports multi-class classification but for face detection we stick on to binary classifier. The reason for choosing Adaboost is that it does not require prior knowledge of face structure and being an adaptive algorithm weights are adjusted at each iterations of the classifier to learn about harder examples and theoretically the training error exponentially converges to null. Drawback of this algorithm is that the training takes more time and the weak classifiers chosen must also be well enough to achieve good accuracy. The distance between the noticed picture and face space was determined utilizing projection coefficients and sign energy in the Eigenfaces technique to identify the presence of the face in a solitary picture. The spatial-temporal filtering approach was utilized to recognize the face in a picture sequence. The filtered picture was subjected to the thresholding approach, resulting in binary motion that assists in the study of motion blobs across time.For localization of the face, each motion blob represents a human head.The LBP is a handy process of extracting texture features. This procedure is every now and again used in face discovery and pattern identification. The LBP administrator changes a picture into a numerical labelled array or image that characterize the image's appearance at a tiny scale. Each central pixel is compared to its 8 neighbors.Neighboring pixels with a lower value than the central pixel are assigned the bit 0, while those with a value equal or greater than the central pixel are assigned the bit 1. A binary number may be created for each center pixel by appending all of these binary bits in a clockwise direction, starting with the top-left neighbor.The generated binary number's decimal value is used to replace the center pixel value. A texture descriptor for an image can be constructed using the histogram of LBP labels calculated over a region or an image.

PCA maximizes the data's intrinsic information after dimension reduction and evaluates the direction's relevance by assessing the variance of the data in the projection direction. Such projections, however, may not be enough to discriminate between different types of data. They could, instead, knead data points together until they're indistinct. PCA primarily seeks out improved projection algorithms based on feature covariance.LDA is a straightforward concept. LDA aims to project a training sample set onto a straight line with as few interclass sample projection points as feasible and as many intraclass sample projection points as possible. We projected a fresh sample onto the same line when classifying it. The position of the projected point determines the classification of this sample. Unlike PCA, which aims to preserve as much data as possible, the goal of LDA is to make data points more recognizable after dimension reduction.

The KNN classifier is a method for classifying objects based on the number k, which refers to the number of near neighboring samples in feature space. The neighbors are made up of a well-classified group of recognized things. This is used as the classifier's training set. The expression classes are classified using a support vector machine (SVM). SVM separates the two classes using a linear decision boundary after nonlinearly translating the feature vector to a higher-dimensional plane. SVM, on the other hand, is a two-class classification method. As a result, we must classify several expression classes using a one-on-one technique.In CV, CNN is the most commonly utilized ANN algorithm. CNN is made up of a sequence of convolutional layers, each of which has an output that is exclusively related to particular regions in the input. This is accomplished by sliding a weighted-matrix filter over the input and calculating the convolution product between the input and filter at each point.Thus finally the facial emotions can be classified by using one of the methods discussed in each phase.

## 4 Proposed system

The above systems are interesting, but the accuracy of emotions can be further improved by using a hybrid approach which is proactive and classify emotions with higher accuracy for better feedback to instructor during lectures. The architecture of the proposed method is as shown in Figure 2. The hybrid approach also takes into account of eye and mouth regions separately. We use both the SVM and CNN classifiers separately and combine its results for better accuracy of the emotion recognition which can be used by the instructors to understand how effective the lecture session has been and change the pattern of teaching or revisions of course contents whichever as appropriate.Images are captured using camera during the lecture. The following are the steps to be followed for emotion recognition:

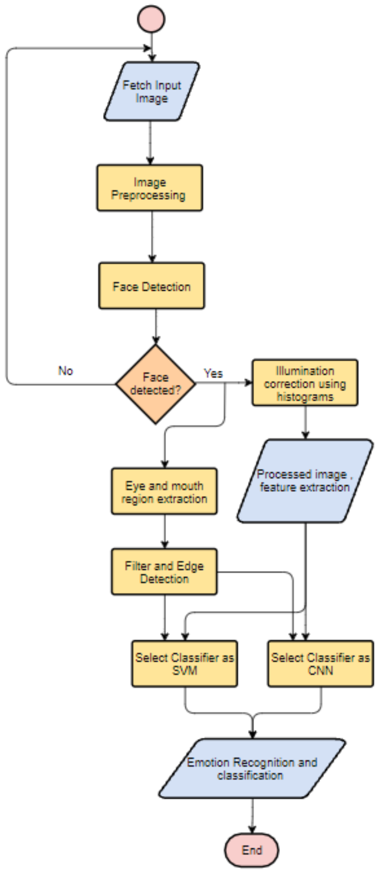
Fetch Input Image
Image Preprocessing
Face Detection
No
Face detected?
Yes
Illumination correction using histograms
Processed image , feature extraction
Eye and mouth region extraction
Filter and Edge Detection
Select Classifier as SVM
Select Classifier as CNN
Emotion Recognition and classification
End

**Figure 2 – Architecture of the Proposed System**

*Image Preprocessing*

After the capturing of the image, to avoid complications we bring the image under a single plane by converting the RGB-image into grayscale image using standard techniques. Image segmentation is the next crucial step. For isolating the facial region from the source image, image segmentation employs a variety of techniques. Thegray scale image of the original image is obtained by sampling the value of each pixel in order to obtain only the intensity information making it less complex under a single plane for further computations.

*Face Detection*

The second step is to detect face from the image. Face detection is done using combination of object detection algorithms like Viola Jones using Haar cascade algorithm and a rectangle box is drawn around the face. After face is detected illumination correction of the image is done using histogram equalization it improve image quality.

*Classification of emotions*

The pre-processed images are given as input to the “Convolution Neural Network(CNN)” and “Support Vector Machine(SVM)” classifiers and the results are combined to classify emotions. Also during the previous iteration using the mouth and eye region emotions are classified separately and these findings are also used as feedback to increase the accuracy of emotion recognition.

The working of the proposed system can be understood better from the Figure 3.

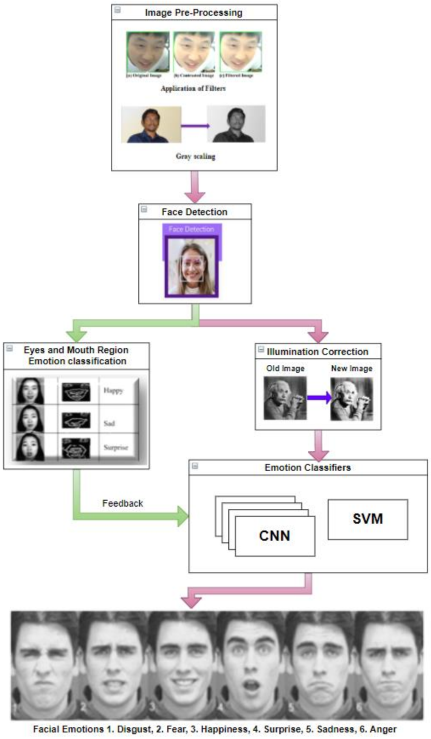

**Figure 3 – Working of the Proposed System[1][8][15][16]**

From the above figure we can visualize how each process of the architecture happens with example images. Once the image pre-processing is over, the face is detected and after illumination correction, the emotions are classified. The Eyes and mouth region are taken separately and classified and used as feedback to obtain better accuracy of the emotions classified.

## 5 Identification of dataset

For training the classifiers, 3 different datasets is identified which are openly available in Kaggle. The first dataset[17] named “Facial Emotion Detection Dataset” has around 35,000 images in .jpg format each of 48 *48 pixel size classified into 7 emotion classes. The next dataset[18] named “Natural Human Face Images for Emotion Recognition” consists of around 5500 images of .png format each of 224 * 224 pixel size classified into 8 emotion classes. Both these datasets are chosen to train the classifiers along with non-facial for the negative class training. The reason for choosing the above datasets are they are images focussed on the facial part and its easier since most images are already in gray scale and requires very less pre-processing. The third dataset[19] named “Google Facial Expression comparison triplet dataset” consists of 1,56,000 images as links(URL’s) in .csv format. The reason for not choosing this dataset is that links require time to fetch the image and not always reliable as hosting server may be down at times. Images are of RGB type and entire structure of person is considered, which requires more complex pre-processing and hence rejected. The dataset will be considered for future enhancement training of the hybrid model classifier.

Modules of the proposed system

The various modules of the proposed system are discussed below. The output of each module is taken as input to the upcoming module to finally classify the emotions from the facial expressions

*Fetch Input Image*

First step is to obtain the input face image to be classified. The input can either be fetched using uri(uniform resource identifier)or directly obtained from the webcam of the device being used. The fetched image is resized to a fixed size preferably to 48 * 48 pixels as the training set majorly contains the above said size and is given to the pre-processing module.

*Image Pre-processing*

The fetched input image is pre-processed by applying set of filters namely median blur and bilateral filter so that the noise in the image is reduced considerably and better results of edge detection is obtained.The median filter is a non-linear digital filter that removes noise from pictures. Noise reduction is a popular a strategy for enhancing the outcomes of subsequent processing by pre-processing.A bilateral filter is a non-linear optical smoothing filter that keeps edges sharp while reducing noise. It replaces each pixel's value with a weighted average of intensity data from adjacent pixels. A Gaussian distribution may be used to calculate this weight.To reduce the complexity, the image captured/ fetched is converted into gray scale and then given to the Face detection module.

*Face Detection*

The face is recognized using the Viola Jones face detection framework, which is fed the pre-processed picture. The approach examines multiple relatively smallsegments of an image (grayscale pictures are used) and attempts to locate a face by examining certain characteristics in each segment. Once the face is identified further processes is done. Even if the face is turned, the algorithm is capable of identifying the face. But the model is trained such that it recognizes only the vertical facing faces as emotion recognition works better in this case. Else the procedure is done from beginning to fetch a new input image

*Feature Extraction*

In facial expression identification, geometric characteristics, appearance-based face features, and a mix of these two features are most commonly used. It is proposed that the face be divided into numerous tiny grids, with the characteristics from all of these grids being used to detect facial emotions. However, because to the extraction of characteristics from inconvenient areas, a little misalignment of the face lowers identification accuracy. Emotions classification is done based on these features extracted.

*Emotion Recognition*

The final step is to classify the emotions. For the usage of pictures as inputs, Convolutional Neural Networks (CNNs) are highly successful. These networks can improve the accuracy of emotion identification by further feature engineering the input pictures.We can train networks with 5 and 6 convolutional layers to investigate deeper CNNs, but it has been found that this does not improve classification accuracy. As a result, the model with four convolutional layers and two FC layers was chosen as the elite network.The pre-processed pictures are fed into CNN and SVM classifiers, and the results are merged to identify emotions. During the previous iteration, emotions were classified independently using the mouth and eye regions, and these findings were incorporated as input to improve the accuracy of emotion detection.

## 6 Conclusion

In this work, a facial emotion detection system which usesdeep learning is presented. This provides a feedback to the instructor on how effective the lecture sessions were and if majority of the students were distracted then the way of teaching can be changed. The efficiency and accuracy of the emotion recognition system is achieved by using a hybrid approach of using Haar cascade algorithm for face detection and also for mouth and eye detection which is used as a feedback for the actual classifier of the entire face image. The model should be trained on a custom data set consisting of images of various postures and illumination levels so that the illumination correction using histograms works efficiently, and each of the newly trained model layers act as feedback to the next layer since it is appended with the existing model.

This is the first time an attempt has been made to employ deep learning-based detection paradigm, and a combination of classifiers with feedback for emotion detection. The use of emotion recognition in applications of e-learning is a highly researched area and during this covid-19 pandemic it's importance has further increased. Since we have also corrected the illumination of images using histograms the system will be stable and can be widely used for real time applications. In the future, the work could be improved for diverse indoor and outdoor conditions and introduction of "Local Binary Pattern(LBP)" for feature extraction can be considered and high resolution images given to the system provides better results.

## References

[1] Jain, A., & Ram Sah, H. (2021). "Student's Feedback by emotion and speech recognition through Deep Learning". 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). doi:10.1109/icccis51004.2021.9397. IEEE

[2] Saleem, S., Shiney, J., Priestly Shan, B., & Kumar Mishra, V. (2021). "Face recognition using facial features". Materials Today: Proceedings. doi:10.1016/j.matpr.2021.07.402. Elsevier

[3] Mishra, N. K., Dutta, M., & Singh, S. K. (2021). "Multiscale parallel deep CNN (mpdCNN) architecture for the real low-resolution face recognition for surveillance". Image and Vision Computing, 104290. doi:10.1016/j.imavis.2021.104290. Elsevier.

[4] M. Tamilselvi, S. Karthikeyan, "An ingenious face recognition system based on HRPSM_CNN under unrestrained environmental condition", Alexandria Eng. J. (2021), https://doi.org/10.1016/j.aej.2021.09.043. Elsevier

[5] Bouhlal, M., Aarika, K., Abdelouahid, R. A., Elfilali, S., & Benlahmar, E. (2020). "Emotions recognition as innovative tool for improving students' performance and learning approaches". Procedia Computer Science, 175, 597–602. doi:10.1016/j.procs.2020.07.086. Elsevier

[6] Tonguç, G., & Ozaydın Ozkara, B. (2020). "Automatic recognition of student emotions from facial expressions during a lecture". Computers & Education, 148, 103797. doi:10.1016/j.compedu.2019.103797. Elsevier.

[7] Goel, T., & Murugan, R. (2020). "Classifier for Face Recognition Based on Deep Convolutional - Optimized Kernel Extreme Learning Machine". Computers & Electrical Engineering, 85, 106640. doi:10.1016/j.compeleceng.2020.10. Elsevier.

[8] Patil, V., Narayan, A., Ausekar, V., & Dinesh, A. (2020). "Automatic Students Attendance Marking System Using Image Processing And Machine Learning". 2020 International Conference on Smart Electronics and Communication (ICOSEC). doi:10.1109/icosec49089.2020.9215. IEEE.

[9] Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Moeller, K. (2019). "Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data". Computers & Education, 103641. doi:10.1016/j.compedu.2019.103641. Elsevier.

[10] Yang, D., Alsadoon, A., Prasad, P. W. C., Singh, A. K., & Elchouemi, A. (2018). "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment". Procedia Computer Science, 125, 2–10. doi:10.1016/j.procs.2017.12.003. Elsevier.

[11] Kanjo, E., Younis, E. M. G., & Ang, C. S. (2018). "Deep Learning Analysis of Mobile Physiological, Environmental and Location Sensor Data for Emotion Detection". Information Fusion. doi:10.1016/j.inffus.2018.09.001. Elsevier.

[12] Bah, S. M., & Ming, F. (2019). "An improved face recognition algorithm and its application in attendance management system". Array, 100014. doi:10.1016/j.array.2019.100014. Elsevier .

[13] Krithika L.B, & Lakshmi Priya GG. (2016). "Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric". Procedia Computer Science, 85, 767–776. doi:10.1016/j.procs.2016.05.264. Elsevier.

[14] Chickerur, S., & Joshi, K. (2015). 3D face model dataset: "Automatic detection of facial expressions and emotions for educational environments". British Journal of Educational Technology, 46(5), 1028–1037. doi:10.1111/bjet.12325.

[15] Happy, S. L., & Routray, A. (2015). “Robust facial expression classification using shape and appearance features”. 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). doi:10.1109/icapr.2015.7050661. IEEE.

[16] Simple Face Detection in Python - Web Article on How to detect faces in an image using OpenCV libraryhttps://towardsdatascience.com/simple-face-detection-in-python-1fcda0ea648e

[17] Facial Emotion Detection Datasethttps://www.kaggle.com/chiragsoni/ferdata

[18] Natural Human Face Images for Emotion Recognitionhttps://www.kaggle.com/sudarshanvaidya/random-images-for-face-emotion-recognition

[19] Google Facial Expression comparison triplet dataset

[20] https://www.kaggle.com/sohelranaccselab/google-facial-expression-comparison-dataset