Computers and Electrical Engineering 93 (2021) 107277

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Application of Deep Learning on Student Engagement in e-learning environments☆

Prakhar Bhardwaj [a], P.K. Gupta [a], Harsh Panwar [a], Mohammad Khubeb Siddiqui [b,*], Ruben Morales-Menendez [b], Anubha Bhaik [a]

[a] Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Solan, HP, 173 234, India
[b] School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, N.L, Mexico

ARTICLE INFO

Keywords:
Digital learning
Deep learning
COVID-19
Engage Detection
Emotion recognition
Engagement detection

ABSTRACT

The drastic impact of COVID-19 pandemic is visible in all aspects of our lives including education. With a distinctive rise in e-learning, teaching methods are being undertaken remotely on digital platforms due to COVID-19. To reduce the effect of this pandemic on the education sector, most of the educational institutions are already conducting online classes. However, to make these digital learning sessions interactive and comparable to the traditional offline classrooms, it is essential to ensure that students are properly engaged during online classes. In this paper, we have presented novel deep learning based algorithms that monitor the student's emotions in real-time such as anger, disgust, fear, happiness, sadness, and surprise. This is done by the proposed novel state-of-the-art algorithms which compute the Mean Engagement Score (MES) by analyzing the obtained results from facial landmark detection, emotional recognition and the weights from a survey conducted on students over an hour-long class. The proposed automated approach will certainly help educational institutions in achieving an improved and innovative digital learning method.

1. Introduction

Social distancing due to COVID-19 has propelled students into a new paradigm of online education. This has severely impacted the education sector and has disturbed the traditional method of in-person learning. This disturbance is also pushing policymakers to develop new ways of keeping students more engaged, while guaranteeing comprehensive digital learning arrangements and overcoming the difficulty in this transition. As stated in Sahu [1], the outbreak of COVID-19 has had a blowing impact on the education and mental health of students and academic staff. It has also been highlighted by them that the academic staff must take care of student's learning experience by making it more effective. Now, almost all educational institutions around the globe are focusing on conducting classes through the digital medium so that learning is not halted. Basilaia and Kvavadze [2] have focused on the success of the transition to online education in schools during the pandemic in Georgia. A case study has also been conducted wherein 'Google Meet' was used to measure the usage of this platform in the first week of transition to online education. It was conducted on 950 students belonging to that private school. The study confirmed that the transition to online learning had been successful. In [3], five major principles with context to improving the impact of online teaching have been stated. It has also been reported that measures should be taken to relieve anxiety and to ensure the active engagement of learners. This indicates that there

☆ This paper is for special section VSI-tei. Reviews processed and recommended for publication by Guest Editor Dr. Samira Hosseini.

* Corresponding author.
E-mail addresses: 171303@juitsolan.in (P. Bhardwaj), pkgupta@ieee.org (P.K. Gupta), harshpanwar@ieee.org (H. Panwar), khubeb@tec.mx (M.K. Siddiqui), rmm@tec.mx (R. Morales-Menendez), anubhabhaik@ieee.org (A. Bhaik).

https://doi.org/10.1016/j.compeleceng.2021.107277
Received 16 August 2020; Received in revised form 14 April 2021; Accepted 14 June 2021
Available online 22 June 2021

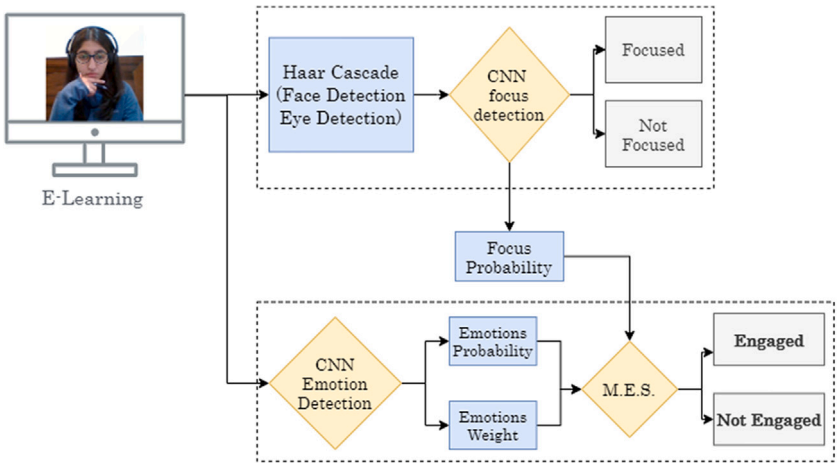


Fig. 1. Student engagement detection model.

has been a growth in the number of people who are adopting online education, and involvement of Artificial Intelligence (AI) can ease this process.

AI and Machine Learning (ML) techniques have been prominently applied to solve issues that are directly related to humans such as smart education, health sector, cybersecurity, consumer behavior, environment, etc., [4–8] and hence its applications have contributed in the betterment of our society. Similarly, in the present era, where the educational system is based on online teaching, smart classrooms and virtual blackboard teaching, AI and ML is playing a significant role in imparting quality education. AI-based learning provides innovative ways to teachers in evaluating the performance of their students. Intelligent tutoring systems have the potential to adapt to the learning styles and preferences of the student, which have immensely contributed in its progress. ML being a subset of AI has also seen a rise in finding its place in the educational sector.

In the proposed MES (mean engagement scores), students' real-time data is taken as input to detect the engagement percentage, and then the emotional status is determined. Two parameters, 'engagement percentage' and 'emotional status' help in calculating the engagement level of students using the proposed state-of-the-art algorithms in live sessions as shown in Fig. 1. The obtained MES can be sent to the teachers to track their students' engagement during online classes.

The main contributions of the paper are summarized as follows:

- We have proposed two different algorithms. In the first, unique weights are assigned to different emotions of the student by detecting the face. In the second, the Mean Engagement Score (MES) is calculated for analyzing the students' engagement during live classes in the digital learning environment.
- We have performed a survey on students attending e-classes and have observed their real-time facial emotions and engagement concentration.
- In addition to engagement detection, this study also represents the importance of emotion detection for determining the engagement of students during online classes.

The remaining structure of this paper is as follows: Section 2 discusses a review of literature related to our study. Section 3 describes the Data and Methodology which was used in the implementation of our proposed model. Section 4 includes the detailed working of our proposed model. Section 5 consists of the Results obtained by implementing our model with a discussion of the relevance of this model. Finally, the paper is concluded in Section 6.

2. Literature review

There have been recent research studies that focus on improving the quality of digital learning. Studies related to *AI and ML* that have helped in automating the recognition of facial expressions have emerged. Schmidt and Kasiński [9] have studied the performance of Haar Cascade Classifiers, which are applied for the detection of face and eyes. These has also been used in our work.

2.1. Digital learning frameworks

Various frameworks have been proposed for digital learning that support technological advancements. A framework for adaptive e-learning through eye tracking and content tracking has been proposed by Barrios et al. [10]. They have focused on finely-grained user information such as the user's area of focus on various objects, frequently visited content, etc. Similarly, Arguel et al. [11] have proposed a conceptual framework for detecting confusion in interactive digital learning environments to help learners resolve

their doubts. In [12], three main approaches have been proposed to recognize student responses. The work mainly focused on the development of tools for effective recognition, interventions in response to students' effect, and emotionally animated agents. In another study as performed by Grafsgaard et al. [13] discusses about a tool for the automatic recognition of facial expressions and subsequently predicting engagement and frustration. A relation between facial expressions and the various parameters of engagement, frustration and learning has also been developed which is based on use of predictive models. However, some emotions like 'happiness' and 'surprise' have not been taken into consideration. In addition to using physiology-based and facial-feature based engagement detection, Monkaresi et al. [14] also remotely monitored the heart rate in video-based methods, but it did not result in improving outcomes obtained.

2.2. Engagement detection techniques

These techniques effectively considers the various factors like emotions experienced and surrounding environmental conditions during various classes in digital environment. In a detailed review of Engagement Detection during digital earning, Dewan et al. [15] have discussed engagement detection by categorizing the existing methods into three categories — automatic, semi-automatic, and manual. They have implemented various techniques of engagement detection using computer vision and machine learning by analyzing the comparison of the obtained results. In another study, Dewan et al. [16], have proposed a deep learning approach to detect the engagement of learners through their facial expressions. A two-level (not-engaged and engaged) and three-level(not-engaged, normally-engaged and very-engaged) classification has been done by them using Local Directional Pattern (LDP) and Kernel Principal Component (KPC) analysis. They have achieved an accuracy of 90.89% for the two-level engagement method and 87.25% for the three-level engagement respectively. Sharma et al. [17] have proposed a machine learning system for student engagement detection using emotion analysis, eye tracking, and head movement by using a web-camera. A digital learning scenario has been used for the testing of the system. Obtained results represent that students with the best scores have high concentration indices. Similarly, Frank et al. [18] have proposed an engagement detection framework that can be used in meetings by analyzing the mental state of people to increase the effectiveness. This framework also integrates sound and information from 2D and 3D images. Chang et al. [19] have proposed an ensemble model using face and body tracking for engagement detection. However, they achieved a low MES of 0.0813 using cluster-based frameworks and neural networks.

2.3. Convolutional Neural Networks (CNN) based methods

CNN based methods have been used by several researchers for improving the engagement detection. Murshed et al. [20] have proposed an engagement detection technique for digital learning environments which is based on Convolutional Neural Networks(CNN). They have proposed three different models known as — all CNN, network-in-network CNN and very deep CNN, for classification of students' engagement during their classes in a digital environment. Three level decisions were made that resulted in the following accuracy levels: 91.74% for not-engaged, 89.55% for normally engaged, and 95.69% for highly engaged. In another study [21], experimental result of five different CNN models have been compared for the purpose of student engagement detection. CNN provides the highest accuracy results. They have also stated that the proposed system provides reliable results when emotions and behavior were mapped to two states, namely 'engaged' and 'not-engaged'.

3. Data and methodology

3.1. Dataset information

Two datasets are used for the research. The first dataset "FER-2013" is an image dataset which is used to train the CNN model and the second dataset "MES dataset" is a tabular dataset used for calculating the weights, and the subsequent calculation of MES. The detailed description of these datasets is provided below:

3.1.1. FER-2013

To train the neural network, a publicly available facial image dataset by Goodfellow et al. [22] named 'FER-2013' from the Wolfram data repository have been used. This dataset consists of 35,887 images. Here, all the images are 48 × 48 gray scaled pixels. The faces occupy most of the pixels in every image. The objective is to first detect engagement in each image, and then to classify them into seven different categories namely 'Neutral', 'Angry', 'Disgust', 'Fear', 'Happy', 'Sad', and 'Surprise'.

The training set consists of 28,709 images of faces, whereas, both test and validation sets include 3589 images. Moreover, the training data also consist of two columns, known as 'emotion' and 'pixels'. Here, the 'emotion' column consist a numeric code for each emotion in an image and represents values 0 = Angry, 1 = Disgust, 2 = Fear, 3 = Happy, 4 = Sad, 5 = Surprise, and 6 = Neutral. Whereas, the 'pixels' column contains pixel values in double quotation marks for each image. The test set contains only the 'pixels' column and emotions are predicted using the training data.

Table 1
Sample record of a primary dataset containing the results obtained from the survey during online live classes.

Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
2	1	1	3	–	5	3

3.1.2. MES dataset

MES dataset was obtained by the authors through a survey conducted on 1000 students over a time period of one week. The complete dataset is available online to the open-source community in the mode of a Github repository (see Appendix A). To carefully observe the behavior and emotions of these students, 100 observers marked every student on a scale of 0–5 to determine their engagement level and emotions. These were categorized into angry, disgust, fear, happy, sad, surprise and neutral. A student may not experience multiple emotions at a point of time during live sessions. Hence, the remaining emotions have been marked with a dash, i.e. a null value, by the observers. A sample row of the dataset is presented in Table 1.

3.2. Methodology

The proposed system consists of two models — engagement detection and emotion detection. Trained weights for both the models have been considered to make predictions in real-time. Here, for the engagement detection model, Haar Cascade Classifier has been used to detect eye regions within faces [23]. In this work, we have used the Haar-Cascade Classifier followed by a CNN Classifier for engagement detection. For the purpose of engagement detection, merely the eye region in a face is inspected, and a Haar-Cascade classifier aids this process. Haar-Features are manually determined and are effective at detecting edges and lines, as shown in Fig. 2. Besides, a Haar-Cascade Classifier has higher execution speed as compared to CNN classifiers. Moreover, the need to train these Haar-Features is inessential, and only the weights for each Haar-feature are trained. In this study, we have used trained Haar-facial features. Once the eyes region is segmented, we manually label the image into two categories - 'engaged' and 'not-engaged'. The CNN is then trained on these manually labeled data and the engagement probability is calculated using the CNN model.

For the emotion recognition CNN model, facial images from the FER-2013 Data set [22] are trained on the CNN to classify emotions into 'Anger', 'Disgust', 'Fear', 'Happy', 'Sad', 'Surprise', and 'Neutral'.

3.2.1. Haar cascade classifier

It is a basic machine learning classifier to train a cascade function from a large number of images with faces (i.e., positive), and images without faces (i.e, negative). Based on the training, it detects objects in other input images. This model is trained on different feature sets like full-body, lower-body, eye, frontal-face, etc., and are saved as *.xml* files. In this paper, we have used only cascade frontal-face and cascade eye detection. This classifier has four main stages which are listed as follows:

1. *Haar feature selection*: features are calculated in the subsections of the input image. The difference between the sum of pixel intensities of adjacent rectangular regions is calculated to differentiate the subsections of the image.
2. *Creating an integral image*: this is used to reduce the computation to only four pixels since a lot of computation will be done when operations are performed on all pixels. This increases the speed of the algorithm.
3. *AdaBoost*: all the computed features are not relevant for the classification purpose. AdaBoost learning algorithm is used to select the specific Haar-like features for classification. AdaBoost is used to make an overall strong classifier using a mix of all weak classification functions. It states that the strongest classifier uses the robust feature, that is, it best separates the positive and negative samples.
4. *Cascading classifiers*: relevant features can now be used to classify a face. Every region of the image is not a facial region, so it is not necessary to use all features on all regions of the image as shown in Fig. 2. Instead of using all features at a time, the features are grouped into different stages of the classifier.

3.2.2. Convolutional Neural Network (CNN)

CNN is a prominent classifier of deep learning that is used to analyze images and extract features from them in a less computational time. It takes images as input, assigns learnable weights, biases them into different objects, and classifies them into different classes. The basic CNN architecture is shown in Fig. 3. Its application has been widely seen in different areas [24]. In this work, we have trained our CNN on manually labeled eye images that were obtained after applying the Haar Cascade model on facial images obtained from the FER13 Dataset [22]. Subsequently, binary classification was performed for classification into 'Focused' or 'Not engaged'. This network is also trained on FER13 data [22]. The output layer predicts the emotion classifying it into seven categories i.e. Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. The CNN architecture used in our proposed model is stated as follows:

- In the sequential model, Input layer 32×32 holds the pixel values of input images;
- Convolutional layers with a set of 3×3 filters computes the low-level features from the input images;
- Max Pooling layer of pool size 2×2 reduces the spatial size of convolved features;
- Classification through a Fully connected layer computes the two class scores using Relu as the activation function.

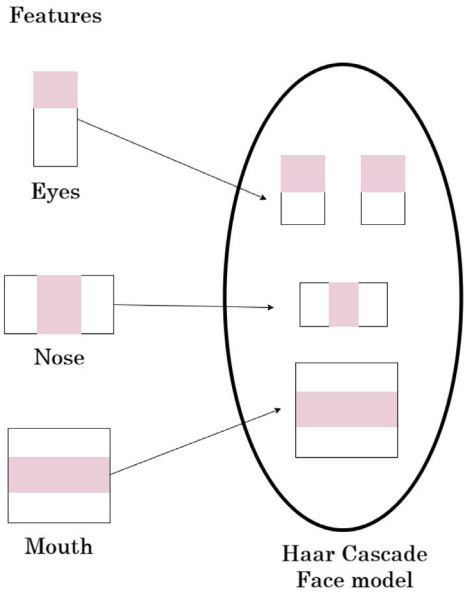


Fig. 2. Facial features cascading classifier.

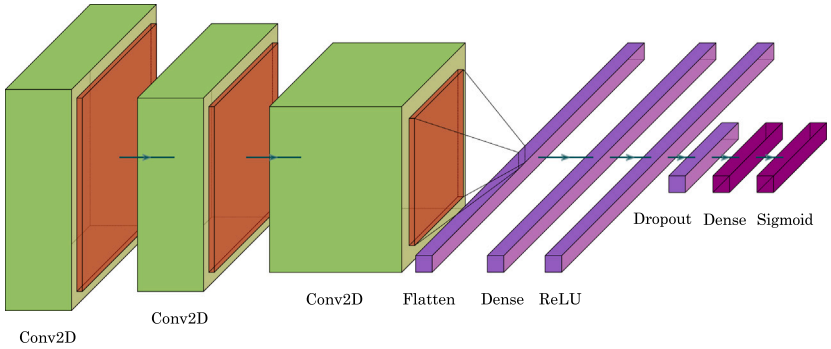


Fig. 3. The CNN architecture.

4. Proposed approach for engagement detection

The designed system is instrumental in maintaining the standards of an efficient learning system. Collecting real-time data from the web-camera of students, the system will automatically evaluate the engagement level of students. If the Mean Engagement Score (MES) defined by the concerned teacher falls under the threshold, it will raise a red flag. MES for each student is sent to the teacher for further assessment at the end of the each class. This system consists of two proposed algorithms—Algorithm 1 for calculating weights matrix of emotions and Algorithm 2 for estimating the MES and detecting engagement. For a better understanding, the main steps of these algorithms are as follows:

4.1. Steps for Algorithm 1

- Here, Algorithm 1 takes S_{ID} (unique student ID assigned to every student), E_{label} (unique label of every emotion), μ (learning rate hyper parameter used to define the amount of change in our model each time the weights are updated in response to the RMSE) and G_{truth} values (obtained from the survey in Table 1) as inputs.
- In the first two steps of the algorithm, the weight matrix corresponding to every student ($\omega_{student}$) and the seven emotions ($\omega_{emotion}$) are initialized randomly using the function *initialize_rand()*. This generates a random decimal number ranging from 0 to 1 every time the function is called.
- For calculating $\omega_{prediction}$, the dot product of $\omega_{student}$ and $\omega_{emotion}$ is taken, which is the multiplication of all the cells in the $\omega_{student}$ matrix with the corresponding cells in $\omega_{emotion}$.

Algorithm 1: Calculating weights matrix of emotions

Input : S_{ID} ⟵ unique student id
E_{label} ⟵ label for every emotion
μ ⟵ learning rate
G_{truth} ⟵ A matrix of actual truth values
Output: $\omega_{student}$ ⟵ student weights
$\omega_{emotion}$ ⟵ emotion weights
$\omega_{prediction}$ ⟵ prediction weights
begin;
1. $\omega_{student}$ = initialize_rand();
2. $\omega_{emotion}$ = initialize_rand();
for $i = 1$ **to** $len(S_{ID})$ **do**
  for $j = 1$ **to** $len(E_{label})$ **do**
    3. $\omega_{prediction}[i][j] = \sum_{i=0}^{len(s\_id)} \sum_{i=0}^{len(e\_label)} \omega_{student}\ \omega_{emotion}$;
  end
end
4. RMSE $= \sqrt{\frac{\sum_{i=0}^{len(S_{ID})} \sum_{i=0}^{len(E_{label})} \omega_{prediction}[i][j] - G_{truth}[i][j])^2}{len(S_{ID}) \times len(E_{label})}}$;
5. $\omega_{emotion} \longleftarrow GRG\_Nonlinear(minimize = RMSE, variables = \omega_{student}, \omega_{emotion}, learning\_rate = \mu)$;

- To calculate the standard deviation in our data, we use RMSE, which is calculated using the proposed Eq. (1). RMSE is a measure of how spread out our predicted and actual truth values are, and our aim is to reduce the RMSE value so as to optimize the weights.
- To reduce the RMSE value, we use Generalized Reduced Gradient (GRG) [25] function — GRG_Nonlinear() function. It is a method of optimizing 'n' number of non-linear equations. The weights ($\omega_{student}$), ($\omega_{prediction}$) are updated each time the RMSE value changes.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{len(S_{ID})} \sum_{i=0}^{len(E_{label})} \omega_{prediction}[i][j] - G_{truth}[i][j]^2}{len(S_{ID}) \times len(E_{label})}} \tag{1}$$

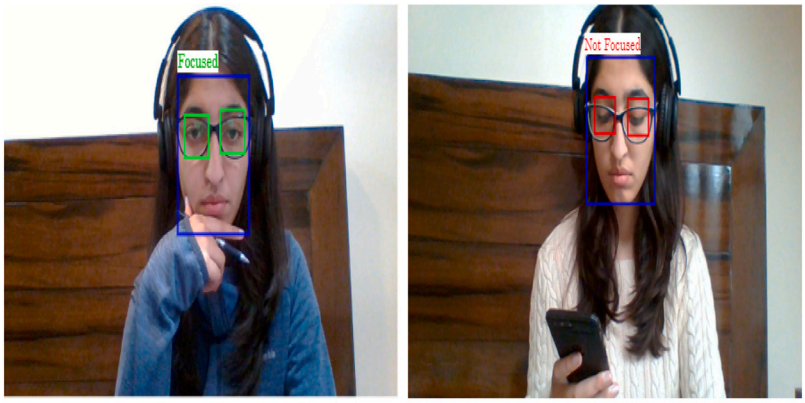
- The calculated values of $\omega_{emotion}$ are then used as an input to Algorithm 2, and further their outcomes are displayed in Table 2

Algorithm 2: Calculating MES and detecting engagement

Input : $\omega_{emotions}$ ⟵ matrix containing the emotion weights
τ ⟵ threshold
E_{label} ⟵ Labels for emotions
Output : MES ⟵ Mean Engagement Score
EC ⟵ Engagement Classification
Parameter: $P_{engaged}$ ⟵ probability of engagement detection
$P_{emotions}$ ⟵ probability of emotions detected
begin;
for $\epsilon = 1$ **to** t **do**
  1. detect_face ⟵ Haar Cascade Face Detector;
  2. detect_eyes ⟵ Haar Cascade Eye Detector;
  3. $P_{engaged,}$ ⟵ CNN(detected_eyes);
  4. $P_{emotions}$ ⟵ CNN(emotion_classifier);
  5. MES $= \sum_{i=1}^{t} (P_{emotion} \times \omega_{emotion}[E_{label}]) + P_{engaged}$;
  if $MES \geq \tau$ **then**
    6. EC ⟵ engaged;
  else
    7. EC ⟵ not engaged ;
  end
end

4.2. Steps for Algorithm 2

- In addition to $\omega_{emotion}$ generated from Algorithm 1, we take τ and E_{label} as inputs. Here, τ refers to the Threshold value. For the purpose of this study, we have used the value of the threshold as 3 which gives the best results in most of the cases. But



(a) Engaged (b) Not-Engaged

Fig. 4. CNN binary engage classification for prediction of image (a) engaged, (b) Not-engaged.

Table 2
Calculated weight of every emotion using the Algorithm 1.

Emotion	Calculated weight
Angry	2.533708926
Disgust	2.721086553
Fear	3.059092323
Happy	4.307270648
Sad	0.799157178
Surprise	3.482136208
Neutral	4.121795065

the threshold value can be changed according to the nature of digital learning classroom. For instance, in a theoretical lecture class, the threshold can be increased.

- During e-classes, faces of students are detected using the Haar Cascade Algorithm to extract the features from facial images in real time, and Haar features for detection of the eye region are set to two adjacent rectangles at a position defined relative to a detected face (see Fig. 2) [23]. Detected eye region is then fed into the Convolution Neural Network (CNN) for further classification into "engaged" or "Not engaged" in order to predict the interest of the student during e-classes, as shown in Fig. 4. The model will return the engaged probability (P_{engage}) for every instance of the classes and $P_{emotion}$, which is calculated using the CNN for emotion detection.
- To determine the proposed Mean Engagement Score (MES) of the students, we have used the proposed Eq. (2). As shown in Algorithm 2, MES is calculated over a time duration of t. This can be decided by the teacher or the administrator.

$$MES = \sum_{i=1}^{t} (P_{emotions} \times \omega_{emotions}[e\_label]) + P_{engaged} \tag{2}$$

- Further, on the basis of MES and τ, the engagement of students is classified into two class values– 'Engaged' and 'Not Engaged'.

  (i) *Engaged*: If the mean engagement score is greater than or equal to the teacher's predefined threshold, then the student falls into the Engaged category.
  (ii) *Not Engaged*: If the mean engagement score is found to be below the predefined threshold by the teacher, then the student falls into the Not-Engaged category.

5. Results and observations

Using Algorithm 1, the optimized and reliable weights for each emotion were calculated, resulting in an RMSE value of 0.105597. The final weight matrices of each emotion were then added and used in Algorithm 2. These are displayed in Table 2. It was also observed from the survey that people displaying 'Sad' emotions were the least engaged, whereas, the students who displayed 'Happy' and 'Neutral' emotions were the most engaged. P_{engage}, $\omega_{emotion}$, $P_{emotion}$ were fed into the proposed algorithm 2 and MES was calculated, which was then used to detect the engagement of the student in e-classes. As seen in Fig. 5, the model was correctly able to detect that the student is Engaged with a MES of 3.726 (Fig. 5(a)) and successfully able to detect that the student is Not Engaged with MES of 1.011 (Fig. 5(b)).

Fig. 5. A sample output of the proposed model where the Mean Engagement Score of a student is calculated and displayed in a live online classes (a) Student is 'engaged' as MES (3.726) is greater than the threshold τ (3), (b) Student is 'not engaged' as MES (1.011) is less than the threshold τ (3).

Table 3
Confusion matrix to visualize the performance of algorithm.

		Predicted class	
		Engaged-positive	Not-engaged-negative
Actual class	Engaged-positive	$TP = 194$	$FN = 29$
	Not-engaged-negative	$FP = 3$	$TN = 274$

From the experimental survey conducted on a group students, observers were asked to record the engagement level of students at half time for an accurate and realistic result. At the same time, we recorded the readings of engagement probability, emotional probability and its corresponding emotion weights(see Table 2) from our automated engagement detection system. We then used our state-of-the-art algorithm, and calculated the MES and engagement level for all students. A confusion matrix is then calculated (see Table 3) to clearly depict the performance obtained from our engagement detection system.

The main observation from the experiments are stated below:

- Most of the students showed 'Neutral' as a dominant emotion
- Only students showing 'Disgust' emotion were not found engaged, and all students showing 'Neutral' and 'Happy' were found engaged
- At a given time, most of the students were found to have an engagement probability of more than 50%.
- Even though engagement probability of some students is more than 50%, they are still found to be 'not engaged'. This concluded that along with engaged, emotions of the students at that moment plays a vital role in determining their engagement level.
- Here, τ is representing the Geometric Mean which is calculated by Eq. (3). It is the threshold for the MES value to determine the engagement level to ≈ 3.0 and get the best results at that time of the survey. In real-time scenarios, teachers or administrators are recommended to decide the threshold according to their own understanding and requirement of engagement.

$$\tau = \sqrt{MES[S_{ID} = 1] \times MES[S_{ID} = 2] \times \cdots .. \times MES[S_{ID} = n]} \tag{3}$$

5.1. Performance evaluation

The performance evaluation has been calculated using a confusion matrix between the predicted values obtained from Algorithm 2 and the Ground Truth values obtained from the survey mentioned in Appendix B.

There are four main values that have been calculated in confusion matrix:

- *True Positive*: When a student was actually 'engaged' and our algorithm also detected the student correctly as 'engaged'.
- *False Positive*: When a student was actually 'not engaged' but our algorithm detected the student falsely as 'engaged'.
- *True Negative*: When a student was 'not engaged' and our algorithm detected the student correctly as 'not engaged'.
- *False Negative*: When a student was 'engaged' and our algorithm detected the student falsely as 'not engaged'.

These four calculated values(see Table 3) can be used to find accuracy, precision and recall using the Eqs. (4), (5), and (6) respectively. Precision is a measure of how many selections which we have made are relevant, whereas, recall is a measure of

how many relevant selections are made. The proposed model was able to achieve 93.6% accuracy, 98.48% precision and 87% recall.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{4}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{5}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{6}$$

After a disruption in the normal life-style due to COVID-19, the continuation of education has only been possible with e-learning. With the rapid increase in learning through online platforms, it is also necessary to identify the level of engagement of students while attending these classes. In any digital learning environment, our model can be used to analyze the engagement of students. Teachers can modify their teaching strategies using our model. The academic staff can also understand the factors which increase or decrease the student's engagement level. Based on the engagement scores of a student, teachers can even assess the performance of students and award attendance to students. A student not paying attention in e-classes can be identified by a low MES and warned by the teacher immediately. Thus, through this study, we were able to attain our objective of automatically detecting the engagement of students in e-classes.

6. Conclusion and future work

Due to the COVID-19 pandemic, digital learning and e-classes have been more prevalent in the education system. Maintaining the same learning standards is going to be challenging for both teachers and students. In this study, we have contributed to support the education system and have tried to provide some authenticity to digital learning platforms. We have applied deep learning-based approaches to detect the student's engagement, and then subsequently combined it with the facial emotion recognition. Various weights obtained with the proposed algorithms are used to obtain the MES of the student in live e-classes. Finally, the proposed system can be used by teachers or school administrations to find whether students are actively engaged or not by analyzing their MES. This will help teachers in an improved assessment of students during live classes, comparable to the traditional offline classrooms. However, issues related to environmental constraints of a student such as head poses, illumination variations, and health beat monitoring can also be taken into consideration in the future. Other factors such as the learner's geographic locations, age, demographic variability, teaching style, course design and course concepts can be investigated too. Further effort can be given to examine 'when' and 'why' students engage and disengage, and how these are linked to digital learning.

CRediT authorship contribution statement

Prakhar Bhardwaj: Data collection, Experiments. **P.K. Gupta:** Neural network model, Methodology. **Harsh Panwar:** Model design. **Mohammad Khubeb Siddiqui:** Supervision in AI based deep learning algorithms, Result analysis and interpretation. **Ruben Morales-Menendez:** Review. **Anubha Bhaik:** Assisted in designing tables, figures, and proofreading.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the contribution of Dr. Allauddin Siddiqi (DINY), Ph.D an Australian Citizen and a Faculty at School of Dentistry and Oral Health, Griffith University, Brisbane, Australia in carefully proofreading the paper.

All authors approved the version of the manuscript to be published.

Appendix A. Survey conducted for obtaining MES dataset

A survey was conducted by the authors where 1000 students and 10 observers participated. All the students were university level students and the observers were chosen randomly. The parameters for the survey were as follows:

- The observers will observe an ongoing class in a digital learning classroom environment and note their observations.
- Each observer will observe one student at a time and note down the engagement level of the student on a scale of 0–5 while displaying a certain emotion and record their engagement for that instance.
- Only 7 recognized emotions are allowed to be noted down by the observers. These emotions are 'Neutral', 'Angry', 'Disgust', 'Fear', 'Happy', 'Sad', and 'Surprise'.
- While taking note of the observations, if the observers find that an emotion is not displayed by the student, then it is marked by a dash (or a null value) instead of rating from 0–5. And thus, it is not necessary for any student to display all the 7 emotions. The complete survey dataset includes 1000 rows and 7 columns, and is available at the GitHub Repository https://github.com/Harsh9524/MES-Dataset/blob/main/MES_dataset.csv.

Appendix B. Survey conducted for evaluation

A survey was conducted by 10 observers where 500 students were observed in over 50 digital classroom sessions. The parameters of the survey were as follows:

- At half time, the observers will note the engagement of the students.
- If the student looks engaged, then the observer will note down 'engaged' and if the student does not look engaged, then the observer will note down 'not engaged'. This will be known as Ground Truth Value.
- Each student is given a student ID for mapping the Ground Truth result with the Predicted result. The complete data with 500 rows and 8 columns is available on the Author's Github Repository. https://github.com/Harsh9524/MES-Dataset/blob/main/MES_evaluation.csv

References

[1] Sahu P. Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff. Cureus 2020;12(4).
[2] Basilaia G, Kvavadze D. Transition to online education in schools during a SARS-CoV-2 coronavirus (COVID-19) pandemic in Georgia. Pedagog Res 2020;5(4):1–9.
[3] Bao W. COVID-19 and online teaching in higher education: A case study of peking university. Hum Behav Emerg Technol 2020;2(2):113–5.
[4] Hastings P, Hughes S, Britt MA. Active learning for improving machine learning of student explanatory essays. In: International conference on artificial intelligence in education. Springer; 2018, p. 140–53.
[5] Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319(13):1317–8.
[6] Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Singh V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. Chaos Solitons Fractals 2020;109944.
[7] Siddiqui MK, Morales-Menendez R, Gupta PK, Iqbal H, Hussain F, Khatoon K, Ahmad S. Correlation between temperature and COVID-19 (suspected, confirmed and death) cases based on machine learning analysis. J Pure Appl Microbiol 2020;14.
[8] Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Sharma S, Sarker IH. Aquavision: Automating the detection of waste in water bodies using deep transfer learning. Case Stud Chem Environ Eng 2020;100026.
[9] Schmidt A, Kasiński A. The performance of the haar cascade classifiers applied to the face and eyes detection. In: Computer recognition systems, Vol. 2. Springer; 2007, p. 816–23.
[10] Barrios VMG, Gütl C, Preis AM, Andrews K, Pivec M, Mödritscher F, Trummer C. Adele: A framework for adaptive e-learning through eye tracking. Proc IKNOW 2004;609–16.
[11] Arguel A, Lockyer L, Lipp OV, Lodge JM, Kennedy G. Inside out: detecting learners' confusion to improve interactive digital learning environments. J Educ Comput Res 2017;55(4):526–51.
[12] Woolf B, Burleson W, Arroyo I, Dragon T, Cooper D, Picard R. Affect-aware tutors: recognising and responding to student affect. Int J Learn Technol 2009;4(3–4):129–64.
[13] Grafsgaard J, Wiggins JB, Boyer KE, Wiebe EN, Lester J. Automatically recognizing facial expression: Predicting engagement and frustration. In: Educational data mining 2013. 2013.
[14] Monkaresi H, Bosch N, Calvo RA, D'Mello SK. Automated detection of engagement using video-based estimation of facial expressions and heart rate. IEEE Trans Affect Comput 2016;8(1):15–28.
[15] Dewan MAA, Murshed M, Lin F. Engagement detection in online learning: a review. Smart Learn Environ 2019;6(1):1.
[16] Dewan MAA, Lin F, Wen D, Murshed M, Uddin Z. A deep learning approach to detecting engagement of online learners. In: 2018 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE; 2018, p. 1895–902.
[17] Sharma P, Joshi S, Gautam S, Filipe V, Reis MJ. Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. 2019, arxiv preprint arxiv:1909.12913.
[18] Frank M, Tofighi G, Gu H, Fruchter R. Engagement detection in meetings. 2016, arxiv preprint arxiv:1608.08711.
[19] Chang C, Zhang C, Chen L, Liu Y. An ensemble model using face and body tracking for engagement detection. In *Proceedings of the 20th ACM international conference on multimodal interaction*, 2018, p. 616–22.
[20] Murshed M, Dewan MAA, Lin F, Wen D. Engagement detection in e-learning environments using convolutional neural networks. In: 2019 IEEE intl conf on dependable, autonomic and secure computing, intl conf on pervasive intelligence and computing, intl conf on cloud and big data computing, intl conf on cyber science and technology congress (DASC/PiCom/CBDCom/CyberSciTech). IEEE; 2019, p. 80–6.
[21] Dash S, Dewan MAA, Murshed M, Lin F, Abdullah-Al-Wadud M, Das A. A two-stage algorithm for engagement detection in online learning. In: 2019 international conference on sustainable technologies for industry 4.0 (STI). IEEE; 2019, p. 1–4.
[22] Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Cukierski W, Tang Y, Thaler D, Lee D-H, et al. Challenges in representation learning: A report on three machine learning contests. In: International conference on neural information processing. Springer; 2013, p. 117–24.
[23] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Comput Vis Pattern Recognit 2001;511–8.
[24] Panwar H, Gupta P, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-scan images. Chaos Solitons Fractals 2020;109944. http://dx.doi.org/10.1016/j.chaos.2020.110190.
[25] Lasdon LS, Fox RL, Ratner MW. Nonlinear optimization using the generalized reduced gradient method. Rev Franç Autom Inform Rech Opér Rech Opér 1974;8(V3):73–103.

Prakhar Bhardwaj is a student of B.Tech. in Computer Science & Engineering at the Jaypee University of Information Technology. His research interest includes the applications of Computer Vision and Machine Learning. He is an experienced team lead with a demonstrated history of working in the education management industry.

P. K. Gupta is Post-Doctorate from the University of Pretoria (South Africa-2015–16) in the Department of Electrical, Electronic and Computer Engineering. He is currently working as a Associate Professor in the Department of Computer Science and Engineering at Jaypee University of Information Technology (JUIT). He has 20+ years of extensive experience in the IT industry and Academics in India and abroad.

Harsh Panwar is an undergraduate student at the Jaypee University of Information Technology pursuing bachelors of technology in Computer Science and Engineering. His research focuses on human-centered AI and deep learning in the context of Educational Data Science and Medical Image Processing. He is one of the co-founders of Edustage, a government-backed EdTech startup.

Mohammad Khubeb Siddiqui received the Ph.D. degree from Charles Sturt University, Australia, in 2018 and Post-Doctorate in Machine Learning from Tecnologico de Monterrey, Monterrey, NL. He is currently working as a Assistant Consultant at TCS, Australia. He has over 10 years of research experience in the field of applications of Machine Learning/Data Mining on different real-world problems like disease diagnosis. He is a reviewer of various reputed journals like Nature Scientific Report.

Ruben Morales-Menendez received his M.Sc. in Process Systems and Automation, and Ph.D. in Artificial Intelligence from Tecnologico de Monterrey and University of British Columbia. His research areas are Artificial Intelligence, Control systems and Educational systems. He is a member of the National Researchers System of Mexico Level II, the Mexican Academic of Sciences and the Engineering Academic of Mexico.

Anubha Bhaik is pursuing her B.Tech. in Computer Science & Engineering from Jaypee University of Information Technology. She is currently enrolled in the CISE Senior Certificate program at the University of Florida, USA. Her interests lie in Deep Learning, Computer Vision and Data Science. She is an experienced team lead with a demonstrated record of contributing to the research industry.