

## A survey on automatic engagement recognition methods: online and traditional classroom

Ajitha Sukumaran<sup>1</sup>, Arun Manoharan<sup>2</sup>

<sup>1</sup>Department of Computing and Electronics Engineering, Middle East College, Oman  
Affiliated to Coventry University, UK, London, United Kingdom

<sup>2</sup>School of Electronics Engineering (SENSE), Department of Embedded Technology, Vellore Institute of Technology (VIT),  
Vellore, India

### Article Info

#### Article history:

Received Sep 13, 2022

Revised Jan 17, 2023

Accepted Jan 19, 2023

#### Keywords:

Computer-vision  
Facial expressions  
Gaze direction  
Gesture and posture  
Neurological  
Physiological

### ABSTRACT

Student engagement in a learning environment is directly related to students' perception and involvement of the educational activities in the class, along with their physical and mental health. This paper presents an extensive survey of the various automatic engagement detection approaches and algorithms based on computer vision, physiological and neurological signals analysis-based methods. The computer vision-based techniques depend on the traits captured by image sensors such as facial expressions, gesture and posture analysis, and gaze direction. The physiological and neurological signal based approach depends on the sensor data, like heart rate (HR), electroencephalogram (EEG), blood pressure (BP), and galvanic skin response (GSR). A brief analysis of the available datasets for Engagement Recognition and its features are also summarized. This study highlights a few commercially available wearables which provides the physiological signals that helps in student's attentivity recognition. Our study reveal that the accuracy of engagement recognition system will increase if we increase the number of modalities used. In this survey, we intend to support the upcoming researchers as well as tutors of smart education set up by providing an overview of existing or proposed approaches of automatic engagement detection techniques in different scenarios.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Ajitha Sukumaran  
Department of Computing and Electronics Engineering, Middle East College, Oman  
Affiliated to Coventry University  
London, United Kingdom  
Email: ajitha@mec.edu.om

## 1. INTRODUCTION

Online education has been gaining more attention in last two decades, in the form of distance education and massive open online courses (MOOCs). The restrictions imposed by the corona virus outbreak in March 2020 placed us, in a situation where physical attendance of lectures and interaction at university campuses was impossible. Educational institutions had to carry out their teaching and learning activities under the constraints imposed by the pandemic. The mode of teaching and learning activities had to transition rapidly from face-to-face (on campus) to fully online to ensure the continuity of the teaching and learning process. Although the constraints imposed at that time have been removed, considering the benefits of online learning, several universities have decided to continue a few of their courses online. Meanwhile, online learning approaches and methods have multiplied, including full-online, hybrid, blended, hyflex, asynchronous, and synchronous.

Student engagement level is directly related to their academic performance [1]. The biggest challenge encountered in online schooling is tracking the engagement of learners. This is a critical aspect of the educational process as it allows the instructors to understand or take decisions on alternative approaches which can promote engagement. Hence the central issue we are facing in an online class or even in a normal on-campus classroom with large cohorts is how to monitor the engagement of students during instruction. The term ‘engagement’ involves emotional and attentional immersion in a task. Engagement fluctuates throughout an interaction experience, as it cannot remain stable [2]. Thus, fostering learner engagement is essential, both in an online learning environment and in a traditional classroom or even in an intelligent teaching system [3].

Fredricks *et al.* [4], categorized engagement into three classes: emotional, behavioral and cognitive. Another study from Anderson *et al.* [5] categorized engagement in terms of behavioral, cognitive, academic and psychological dimensions. The behavioral type of engagement is the observable act of students being involved in the learning process by attending classes, involving in classroom activities, submitting tasks on time and following the teachers instructions. Emotional engagement covers the affective factors or students’ emotional attitude towards learning. This can be assessed by analyzing the facial emotions. Cognitive engagement is abstracted in the literature as the effort that students make towards deep understanding and learning using their cognitive abilities. Another type of engagement is agentic engagement which occurs when the student takes the initiative to dynamically enhance the knowledge and experience that contribute to learning and teaching. Hence in this type of engagement students are active participants rather than passive recipients [6].

In a traditional classroom with moderate student strength, the instructor can directly observe the attention level of each student and can take action if necessary. But in online learning platforms or traditional classrooms with large cohorts, the instructor normally does not have an idea about the emotional states of their students and or their engagement level in the class. If adequate information about the affective states of students is captured, then the tutor can provide the appropriate learning materials and required support to the students. This will have a considerably beneficial effect on student engagement and learning. Several works on engagement detection were performed by considering the image trait captured by camera [7]-[9]. Recent advances in technology brings to the market, different data capturing gadgets or wearables, which enhanced the possibility of research in engagement tracking by using human physiological signals [10], [11]. A few recent research articles [12]-[14] are based on the analysis of multimodal data for engagement detection and obtained a positive impact performance evaluation.

In this survey, we intend to support the upcoming researchers as well as tutors of smart education set up by providing an overview of existing or proposed approaches of automatic engagement detection techniques in different scenarios. This paper presents a review analysis of various automatic engagement detection methods and their algorithms in the context of online learning and traditional classrooms. Here we have considered the related works of engagement recognition by using facial expression, gestures and postures, eye tracking and by the analysis physiological and neurological signals during learning process in various scenarios. A brief discussion of the various available databases for the engagement recognition and commercially available wearables for monitoring the physiological data is also provided. The key finding are discussed in the result and discussion session.

## 2. METHOD

A systematic literature review on automatic engagement detection techniques in both online and traditional classroom is conducted from recent articles. The study includes most of the articles published between 2015 and 2022, along with the classical approach and concepts derived formerly. This survey is extended to the insight of real time automatic multimodal engagement detection methods because of the huge popularity and usage of video conferencing technologies in schools and universities. To support the researchers in this area, a brief analysis of freely available dataset information is also included.

### 2.1. Categorization of automatic engagement detection methods

The automatic approach of engagement recognition is divided into computer vision-based approaches and physiological and neurological signal-based methods. The computer vision based techniques depend on the traits captured by image sensors such as facial expressions, gesture and posture analysis, and gaze direction. The physiological and neurological signal-based approach depends on the sensor data, such as heart rate (HR), electroencephalogram (EEG), blood pressure (BP), and galvanic skin response (GSR).

#### 2.1.1. The computer vision-based engagement detection techniques

Most of the literatures on engagement detection was based on computer vision based methods, as it is easy and low cost to capture a person’s facial expression, posture, gesture or even eye gaze with a web camera or the camera available with laptop, cell phone or even tablets. Hence computer vision-based techniques are easy to use in a classroom environment without affecting learning activities of students. Eventhough there are

quite great number of literatures, but considerable less research work towards the real time implementation of automatic engagement recognition in an educational set up.

The Figure 1 shows a generic flow of computer vision based detection techniques. The video frames of students, who participate in the learning activity are captured by a web camera or surveillance camera. The next step is to separate the region of interest from the input video frame. The frequently used modalities in computer vision based approaches are facial expressions, eye gaze direction and gestures and postures. In segmentation process the image of the frames are divided into different regions based on the characteristics of pixels or boundaries. Preprocessing is performed for data cleaning, integration, reduction, and transformation. In feature extraction, the dimensionality reduction is done in such a way that large number of pixels are efficiently represented for the interesting parts of the image which is to be captured. The classification model is used to map input image patterns based on the input training data using the classification algorithm. Based on the algorithm used, a classification score is generated to evaluate the performance of the model. As the last step, a decision unit which combines the classification scores to output, the listed engagement levels of the students from the input video stream.

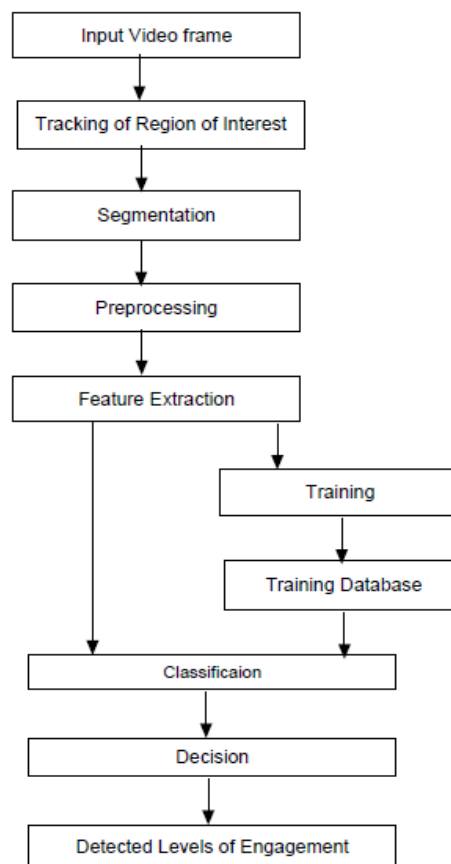


Figure 1. Generic flow of computer vision based detection techniques

#### A. Facial expressions

Human face reflects majority of the emotions, or the facial expressions convey affects, attitudes and intentions [15]. In order to automatically understand and detect the affective states of learners such as engagement, and frustration, it is essential to analyse the facial expressions. Facial expression are occurred due to the movements of facial muscles and features [16]. Research on facial expressions was began by Charles Darwin [17] more than a century ago, and to this point a large amount of works have done in recognizing basic facial expressions [15], [18]-[20].

The facial action coding system (FACS) is the set of movements of facial muscles corresponding to the displayed emotion or expression, was initially studied and created by Carl-Herman Hjortsjo by using 23 facial motion units in 1970 and is further enhanced by Ekman and Friesen. FACS describes all the visible facial movements as action units (AUs). FACS measures the facial muscle movements for the emotions or mixed

emotions associated with learning centered affect to aid in emotion detection task [21]. AUs can occur individually or in combinations with varying intensities. In various studies it is observed over 7000 AU combinations [22]. These combinations may be additive or nonadditive. In additive cases the appearance of constituents will not vary, while the appearance of the constituents will vary for nonadditive cases [23]. Action Unit measurements are only descriptive analysis of behaviour, whereas for an observational system like engagement is an inferential process about that which is being measured [24]. For many decades, the FACS are extensively used to study and analyse the facial movements [25]-[27]. In many literatures the mapping of AUs to respective emotions are well defined [28]. But the mapping of AUs to student-centered emotional states are still at its beginning phase.

McDaniel *et al.* [29] in their study, captured facial videos of students during their interaction with auto tutor on a complex topic regarding computer literacy. After completing the interaction, the engagement status of students was identified by a peer, learner and two experienced judges. The facial expressions of the students were recorded using Ekman's facial action coding system by two independent judges. The correlational analysis of action units and emotions were done and determined the extent to which each of the AUs or the group of AUs are mapped to the affective states of confusion, 'boredom', 'delight', 'frustration' and 'neutral'. As per the authors, the vastly animated affective states were 'delight' and 'confusion' which were easily recognizable from facial expressions, however 'frustration' and 'boredom' were not easily distinguishable. Since 'boredom' resembled an expressionless face and 'frustration' was associated with a physiological arousal and the emotion looks close to 'neutral'.

The computer expression recognition toolbox (CERT), is a software tool used in many research studies for engagement detection. CERT is used for fully automatic real-time facial expression recognition. From FACS, the CERT can code the intensity of nineteen distinct facial expressions and six classical expressions. In addition, it can evaluate the locations of ten facial features and the three dimensional orientation (pitch, roll, yaw) of the head. CERT obtained an average detection accuracy of 90.1% for evaluating the facial actions with extended Cohn-Kanade database and 80% for the spontaneous facial expression dataset [30].

An automated tracking of student fine grained movements of face such as eyebrow-raising (outer and inner), eyelid tightening, mouth dimpling and brow lowering during tutoring was proposed in the paper [24]. This analysis was done from FACS using CERT. The predictive models analyzed the correlation between frequency and intensity of facial movements with the outcomes of the tutoring session. The models emphasized the relations of facial expression with respect to the affective factors such as engagement, frustration, and learning. The validation study inveterate that, during the entire session of tutoring, CERT outrivals in the tracking of detailed facial movements.

Bosch *et al.* [31] in their paper used CERT for the tracking of facial features based on FACS. For classification models of affective states, they have used machine learning techniques. The detection accuracies of the affective states confusion and frustration were good, which was measured in Cohen's Kappa=0.22 and 0.23 (Cohen's kappa is used to assess the performance of a classification model) respectively, however for the detection accuracy for Boredom (Kappa=0.04), Flow/Engagement (Kappa=0.11) and Neutral (Kappa=0.07) were lower. Another study using CERT [32] for gathering the facial expressions of younger population was done by investigating the facial expressions of children.

In a traditional classroom, with a moderate number of students, an instructor is easy to understand the degree of engagement by analysing the student facial expressions directly. Hence in order to identify the students affective states in an online class environment, a promising non-verbal channel is needed to explore and research shows the most promising one is by using facial expressions [33]. A study conducted by Baker *et al.* [34] proved that boredom is very tenacious expression in the learning environment and is associated with poor understanding of learning materials. Confusion and engaged concentration were the most common states among learners.

Another method of recognizing facial expressions and then emotion recognition is by appearance-based methods. The feature extraction performs an important role in the engagement recognition task. By using the extracted features from face regions, we can generate patterns for engagement classification and hence extracting a rich number of features will make a successful classification model. Widely used methods for feature extraction techniques in engagement recognition are local binary patterns (LBP) and histogram of oriented gradients (HOG).

Monkaresi *et al.* [2] used computer vision methods for extracting the facial features and heart rate from remote videos. The study was conducted for the students who had given a writing task with moderate challenges. Microsoft kinect face tracker, a face tracking device was used to separate facial features and the classification tool box, WEKA was used to classify the features with updateable Bayes Net, Naïve Bayes, clustering logistic regression, rotation forest and dagging classifiers. The classifier accuracy was evaluated by the area under the ROC curve (AUC). They attained an AUC of 0.758 and 0.733 for concurrent and retrospective annotations respectively. The two levels of engagement detection as "not-engaged" and "engaged" was done by using feature level fusion of channels. Whitehill *et al.* [35] proposed an automated

method for tracking the engagement of students in real-time using facial expressions. The machine learning techniques such as Gabor features and linear support vector machines (SVMs) were used. They have classified the degree of engagement status in 4 levels: “not engaged at all”, “very engaged”, “nominally engaged” and “engaged in task”.

Nezami *et al.* [36] in their paper proposed an automatic engagement recognition system using deep learning approach. They have prepared a new dataset named engagement recognition (ER) dataset for the purpose to facilitate research on engagement recognition from images. They have annotated the data in the behavioral (“off task”, “on task” and “can’t decide”) and emotional (“Confused”, “bored”, “satisfied” and “can’t decide”) dimensions. Each student’s facial status was categorized as disengaged or engaged by combining both behavioral and emotional dimensions. At first the model was pre-trained using the existing FER 2013 dataset. The resultant model was employed to ER dataset by deep learning method. As per the authors, the model performance was good compared to the existing baseline methods.

## B. Eye movements

Our eyes are directed towards the objects of interest for capturing the visual information. Attention is a cognitive process which includes the concentration or the ability to actively process a specific information. Hence the tracking of the eye movement with the gaze direction is very important in cognitive psychology, since it records the information related to the attention during different tasks. In the past decade, a few research works were conducted in real-time engagement tracking of learners based on the eye gaze movements along with other modalities such as head movements and facial emotions [37]-[42].

The information enters to the brain through the visual pathway responsible for the conversion of light energy to electrical potentials to get interpreted by brain, Fovea is a tiny depression inside the neurosensory retina, where visual acuity is the highest and is processed in the cortical and subcortical parts of central nervous system. The saccadic movements of eye helped to shift the fovea from one point of visual field to another point of interest. Fixation is used to keep the fovea aligned on the target. Alternative the saccade and fixation process is repeated several lakhs times in a day and is important while doing concentration related tasks such as attending tutors, reading and driving. The saccades’ triggering can be ensued by the presence of visual stimuli which may be rousing to the object or instigated voluntarily by one’s attention to an object. During the cycles of visual fixation, the saccades can be repressed. An automatic saccade response will be inhibited by brain in these conditions [43]. Eye tracking equipment collects the metrics related to saccades, blinks and fixations. These tools are also used in the studies of other learning processes which includes marketing, and driving vehicles [44]-[46].

The study conducted by Sáiz-Manzanares *et al.* [47] analysed the data obtained from the eye tracking method with statistical test, supervised and unsupervised machine learning techniques for a puzzle solving task. For tracking eyes, the tools used were SMI BeGazeTM, SMI Experimenter Center 3.0 and iView XTM. This equipment recorded the movements of eye, the pupil diameter of both eyes and the coordinates. The parameters included were saccades, fixations, scan path and blinks. In their observation, for answering the crossword puzzle, there was no noteworthy differences among participants and but, identified difference in the values of minimum saccade velocity and minimum saccade amplitude. The study was conducted using the supervised machine learning approaches, offered the viable features for evaluation. A great match was noticed among the algorithms DBSCAN, k-means ++ and fuzzy k-means used for the clustering techniques. The learning profile of the participants provided by these algorithms in 3 types of learners: over 50 years old, learners and teachers less than 50 years.

In Daniel and Kamioka [48], proposed a method for identifying learners’ attention status during the session of distance-learning system. This system used the biological information associated to the movement of eye metrics such as fixation rate, duration and counts along with average saccade-length collected from the eye tracking equipment. The eye movements of students for the duration of the class session are tracked by the eye tribe software installed on the computer. Machine learning techniques are used for the classification of student’s attention status. They have used three classifiers, namely sequential-minimum-optimization (SMO), multilayer perceptron (MLP) and J48. By using the classifier multilayer perceptron, they obtained an accuracy of 90.7% in recognizing the attention level, however, the time of execution was more compared to other algorithms. As per the author the most vital eye metric is the fixation duration when compared to the other 4 metrics.

Mu *et al.* [49] in their work used Tobii X120 instrument for capturing eye movement. Based on the eye movement index and total fixation duration, they concluded that according to the learning style of students, the focal area and online path differs. Shalini *et al.* [50] proposed a method for real time student attention tracking from facial emotion, eye-ball and head movements. For the facial expression recognition, the feature-extraction was done by principal component analysis (PCA) and the pupil detection was from haar cascade. The head movements tracking is done by local binary patterns, the machine learning model is used for generation and comparison done by open CV.

Sharma [51] proposed the system which identify the real time engagement status of the learners in an e-learning scenario. Their models used the information provided by integral web-camera of the laptop. They have combined the information about the eye movement, head pose variations and facial-emotions to generate the concentration index. The analysis of eye/head movement was done by Haar-cascade algorithm. The binary classifier provided the final decision as “Distracted” or “Focused”. If the student categorized as “Focused”, then the facial-emotion analysis was carried out by CNN and the dominant emotion probability (DEP) score was generated. For the study, they have considered the 7 emotions (Neutral, Surprise, Happy, Sad, Disgust, Scare and Anger). A concentration index (CI) was calculated as the product of dominant emotion probability and emotion weight (EW). The emotion weight was in the range between 0 to 1, where the emotions corresponded to attention have highest weight value. The proposed system provided three classes of engagement: “not engaged at all”, “nominally-engaged” and “very engaged”. Additionally, the authors observed that the students with top scores had better concentration indexes.

The studies so far discussed on engagement recognition using eye tracking was effective, but at the same time we have some challenges in proper eye calibration for getting data precision. Another disadvantages of eye-tracking technology is that we cannot track all eyes as it depends on the ability of camera to record the movements of eye with glasses, contact lenses and pupil color [52]. It is difficult to calibrate for the participants with eye-glasses or any eye disorders and required to be excluded from the studies [53]. Another disadvantage is that participants must remain within an eye-tracker range which may not be possible always.

### C. Gestures and postures

The body language signals gestures and postures can indicate a person's emotions and attitudes. The movement of a body part is referred to as gesture (head and hand), whereas posture tells the way in which our body is positioned while sitting or standing. Many studies proved that gestures and the body actions are important emotional cues as these spontaneous body gestures replicate the real time expressions [54]-[57]. Considering the online learning set up, we have more importance for the emotions and actions on the upper body parts. Many recent works have done on e-learning systems by analysing learner's upper bodily gestures [14]. Several researches revealed that the hand movements during conversations and social interactions convey the reactions or even different emotions or the affective states [58], [59].

Behera *et al.* [60], proposed an automatic deep learning model of hand-over-face (HoF) gestures in images to study nonverbal behaviors of learners comprising of HoF gestures, eye movements and head variations along with facial-expressions during learning activity. These nonverbal behaviours of learners were captured using a webcam and the HoF gestures identification were done from the still images. The behaviour was analyzed from a 40 minute session which includes the reading task and the problem solving task. The exercise problems on the sessions were divided to easy, medium and difficult set. This study assessed the facial emotions, HoF gestures, eye and head movements with their relations to the classroom behaviour interms of the occurrence frequency for different learning activities by considering the complexity levels, and examined the impact of time duration. This 40 minute session, the HoF gestures occur on an average accuracy of 21.35% and the HoF recognition attained the accuracy of 86.87%. The authors also observed that the occurrence frequency of HoF gestures in problem-solving tasks are more when comparing with the reading tasks. They got the accuracy rate of 23.79% for easy levels, 19.84% for medium levels and 30.46% difficult levels.

Ekman [61] explicated technical basis of micro facial-expressions and macro facial-expressions. Based on their studies, the instinctive micro actions or gestures depict the person's actual emotions, compared to the macro actions. Khenkar and Jarraya [62] presented their work of the engagement detection for e-learning students based on the spatiotemporal features extracted from micro body gestures. The engagement detection model implemented by the authors, analyzed the spatial and temporal information all through the video frames using a 3-D convolutional neural network. They have used the video Sports-1M dataset for training the C3D model and followed the transfer learning method. The implemented C3D model was centered on two methods. In the first method, they have extracted spatiotemporal features of e-learner's micro body gestures with linear classifiers, after applying fine-tuning to the pretrained model. In the second approach, they have extracted the deep features by the pre trained model using two linear classifiers (SVM-Naïve Bayes). The authors obtained an accuracy of 94% for pretrained C3D model, 93% for pre-trained C3D+SVM and 77% for C3D +Naïve Bayes. As per their observation, the performance of 3D CNN model was quite good for detecting e-learner's engagement detection task based on micro body gestures.

Riemer *et al.* [63] studied the effect of reactions undergone by students during the interaction session with a multimedia learning system. To identify the features of bodily expression, the body actions of seventy undergraduate students while playing the serious game related to financial education was captured. A depth image sensor, Microsoft Kinect is used to record the data. As per the study and the validation, the participant keeping head more towards right is related to frustration and towards left indicate enjoyment. The result revealed that as frustration increases enjoyment decreases. Also, body of the participant positioned nearer to the gaming screen is related to frustration and as time passed leading to boredom [64].

Garger [65], the frequency of head tilting is associated with boredom, conforms in the findings of [63]. The conclusions of these studies help researchers to label their data. Satyanarayana *et al.* [66] introduced the dataset called SDMATH for detecting deictic gestures using machine learning and computer vision approaches. This includes graph based visual saliency (GBVS), binary morphology, histogram of oriented gradients for object detection and support vector machines (SVM) classification to detect deictic gestures in the context of a one to one mathematics teaching session. All the studies discussed so far proved that the gestures and postures of students during learning reflect the learner's interest and hence are important cues for an intelligent tutoring and for an adaptive learning system.

### 2.1.2. Physiological signals

The functioning of different physiological systems like nervous system, cardiovascular system, and muscular system of our body generate the physiological signals. Associated with human emotions, there could be an increase or decrease in heart rate, piloerection, cutaneous blood flow, gastrointestinal motility, and sweating. The physiological signal responses are not much sensitive to cultural differences [67], since they can be occurred through spontaneous emotional responses. The Table 1 discuss the general sensors used commonly for obtaining physiological signals which are used to detect the affective states of learners.

Table 1. Commonly used sensors for obtaining physiological signals and an overview

| Commonly used sensors for measuring physiological signals | Overview  |
|---|---|
| Electrocardiogram (ECG)                                   | ECG is the recording of the electrical activities of the heart muscle. The heartbeat is an electrical impulse that coordinate and causes the muscle to contract and pump blood through the heart. Normally a heartbeat on ECG, display the timing of upper and bottom chambers [68]. Monitoring and analyzing ECG patterns, helps in tracking the engagement status of the student in a learning environment.   |
| Electroencephalogram (EEG)                                | EEG is used to record and analyze the brain activities. Brain nerve cells (Neurons) communicate with each other through electrical impulses. An EEG tracks and records the brain activity patterns. By analyzing the wavelength band, the EEG can differentiate the different brain's active states. In a human life cycle, we learn new skills and also, we may require to modify or update the already learned one by enhancing the neural maps. As technology is changing day to day, brain require a rigorous re-adaptation to new interfaces [69]. An engaged user will always actuate learning in an optimal way by avoiding distractions [70], [71]. |
| Electromyogram (EMG)                                      | An EMG measures the response of muscle or electrical discharges in response to the stimulation of nerves from muscles. During the test a thin needle/needles are inserted into a muscle [72]. The emotion excitement of students during learning activity can be analysed by examining the EMG signals. The movement of facial muscles provides a good amount of information on student engagement with dynamic visual content.   |
| Blood pressure (BP)                                       | BP is the force of blood circulation against the arteries' walls. BP is taken as two measurements Systolic (the blood pressure, when the heart contracts) and diastolic pressure (the blood pressure, when the heart relaxes) [73]. The emotional state due to stress during the learning activity can be monitored by measuring BP.  |
| Skin temperature (ST)                                     | The skin temperature is measured on the outermost surface of the human-skin. With the use of skin sensors the peripheral temperature can monitor, which allow to sense the emotion stress and hence can help in the engagement detection of leaners [74].   |
| Galvanic skin response (GSR)                              | GSR is used to measure the conductance of skin. It depends on skin moisture level associated with sweat glands [75]. By monitoring the sweating gland activity, some researches have carried out in the engagement detection of students [76].  |
| Photoplethysmography (PPG)                                | This optical technique to used measure the variations of blood volume in microvascular bed of tissues [77]. In the education research PPG sensors have used for heart rate measurement for detecting student attentivity in classrooms.   |
| Respiratory pattern (RP) and Respiratory volume (RV)      | Our average breath pattern is 12 breaths per minute. According to our emotions the frequency of breaths may change [78]. The respiratory volume is the amount of gas present inside the lungs at a particular time. The lung volumes tidal volume, inspiratory reserve volume and expiratory reserve volume are measured using spirometry [79].   |

Belle *et al.* [80] proposed a system for engagement detection by ECG as the main physiological signal during a particular learning activity. They have identified the correlation between varying levels of attention and its effect on cardiac rhythm recorded in the ECG. For comparing the performance, they have analysed the EEG signals of a set of students. The feature extraction of ECG signals are done by Stockwell-transform and EEG signals by discrete wavelet transform (DWT). The machine learning algorithms used for classification are C4.5, a decision tree-based classification algorithm, classification via regression and Random forest for differentiating whether the participant is attentive or not attentive. The results obtained by using ECG were comparable with EEG.

Apicella *et al.* [81] proposed the engagement detection in learning 4.0 based on personalized wearable EEG system consist of a wireless-cap connected with dry-electrodes and eight data acquisition channels. The validation of the system is done for an experimental task involving cognitive and motor skills of 21 students during learning of functioning of a specific human-machine interface. The proposed method was based on filter bank, support vector machine and common spatial pattern. For the cross-subject case, the baseline removal by using TCA obtained an average accuracy of 72.8% for cognitive engagement and 66.2% for emotional engagement.

Awais *et al.* [11] for engagement detection was based on the wireless transmission of different physiological signals to the data-processing-hub through an integrated IoT frame work. The system used multimodal physiological sensors such as ECG, EMG SKT, GSR and BV. Three EMG sensors were used, in such a way that 2 on the face and 1 at the back. There are 30 people participated in the data-acquisition experimentation. They were displayed of 8 videos on four different emotions relaxing, amusing, boring and scary. The duration of the videos were 140-200 seconds, and the physiological signals of participants were recorded while they were watching the video. For training and classification they have used the special deep neural network, LSTM. The Iot framework permits the data-transmission and reception between the wearable sensors to Iot hub (data processing) and then to the cloud server. This work, also evaluated the performance of the combinations as per the Table 2. As per their observation  $C_1$  obtained the worst performance accuracy of around 70%, while  $C_2$  and  $C_3$ , both perform equally good by obtaining the  $f$  score above 90% with a difference of only 0.05%. This imply that the contribution of RP sensor is negligible. The best performance is achieved for  $C_4$ , but it is again comparable with  $C_2$  and  $C_4$ . It is not practical to use all the sensors. The study proved the best sensor combination as  $C_2$  with ECG, GSR, BV, and ST. This sensor combinations are available as wrist wearable devices in shimmer wrist bands [82].

Table 2. Sensor combinations [11]

| Combinations | Sensing modalities                      |
|--------------|---|
| $C_1$        | EMG1, EMG2, EMG3                        |
| $C_2$        | ECG, GSR, BVP, ST, RP                   |
| $C_3$        | ECG, ST, GSR, BVP                       |
| $C_4$        | EMG1, EMG2, EMG3, ECG, GSR, BVP, ST, RP |

Carroll *et al.* [83] analyzed the student engagement, for a sample of 45 students of unmanned-aircraft-systems (UAS) training program through physiological and behavioral inputs. They have used respiratory, cardiovascular, eye-tracking and electrodermal movement sensors. They were able to achieve the classification accuracy of 85% for different engagement levels and obtained 81% without including eye tracking features. Most of the studies in engagement detection of students using physiological signals, revealed the promising result when using EEG sensor, as EEG is feasible indicator for moment to moment changes of learner's attention [84]-[88]. The usage of other physiological-sensors like ECG, heart rate, skin-conductance and PPG are also proved their effectiveness in identifying the changes in students' affective state [89].

Normally ECG signals convey information on stress level, relaxation, and concentration surprise. At the same time PPG signals provide oxygen saturation, and blood volume. which are related to laughter, frustration, interest and stress. Currently many wearables for measuring these physiological signals are commercially available. Smartwatch from Apple, contain sensors such as electrical HR sensor, Oximeter, gyroscope sensor, optical HR sensor and accelerometer. Similarly smart watch from Empatica consist of peripheral temperature sensor, EDA sensor, gyroscope sensor and 3-axis accelerometer [90]. Other than smart watches, other gadgets like wrist band, vest, T-shirts, helmet, headphones, ring, and bracelet. Are also available in market with different combination of sensors. By making use of these gadgets, it is easy to measure physiological signals of people which could be useful for the engagement detection of students or other healthcare applications [79].

## 2.2. Dataset

Analysing the academic emotions have a great impact on learning effect. Students typically express their engagement status in learning externally by the facial expressions, speech, gestures and postures or behaviour. A good collection of datasets is pivotal for any work. Currently there are several datasets available for analysing the facial emotions and most of them are limited to the basic emotions. Academic emotions belong to the category of affective state and these academic emotions cannot be labelled only by basic emotions [91]. But still there are no concrete definitions for academic emotions. Many researchers have concluded the frequent feelings, students experiencing while attending classes are frustration, boredom, confusion, anxiety,



enjoyment, and confidence [92], [93]. Table 3 summarizes, the freely available database, features and the affective states recognized for engagement detection, which are used in some of the reviewed papers.

Table 3. Engagement detection datasets and their features

| Database                             | Features   | Affective states recognized  | Modality  |
|--------------------------------------|--|--|---|
| DAiSEE [94]                          | The images of the dataset are created from 9068 videos and 112 subjects (80 males and 32 females). Available in the web link: <a href="https://people.iith.ac.in/vineethnb/resources/daisee/index.html">https://people.iith.ac.in/vineethnb/resources/daisee/index.html</a>  | Engaged, Frustration, Boredom and Confusion  | Facial Expression   |
| In-the-wild [95]                     | The dataset contain 725,000 frames are captured from 264 videos with 91 subjects about 16.5 hours of recording (25 males and 50 females).  | Engaged, Barely-engaged Highly-engaged, and Disengaged,  | Facial Expression   |
| HBCU [35]                            | HBCU dataset consists of images created from 120 videos and 34 subjects (9 males and 25 females).  | Very-engaged, Nominally-engaged Engaged, and Not-engaged,  | Facial Expression   |
| SDMATH [66]                          | The dataset has formed from 20 videos and with 20 subjects (It involved 210 females and 10 males).   | Deictic gestures   | Gestures, speech, eye gaze and facial expressions                               |
| WACV dataset [96]                    | WACV dataset contains total 4424 images in which 412-Disengaged; 2247-Partially Engaged and 1765-Engaged. Available in: <a href="https://github.com/e-drishti/wacv2016">https://github.com/e-drishti/wacv2016</a>  | engaged, partially engaged, and disengaged   | Facial Expression   |
| Engagement Recognition dataset [36]  | This database is prepared from the videos of 20 students of age group 14 to 15 years (11 girls and 9 boys). It consist of 4627 samples in which 2290 are for engaged samples 2337 are disengaged samples. Available in: <a href="https://github.com/omidmnezami/Engagement-Recognition">https://github.com/omidmnezami/Engagement-Recognition</a> .  | Engaged and Disengaged   | Facial Expression   |
| DEAP [97]                            | The dataset is prepared from EEG, peripheral-physiological signals GSR, respiration amplitude, GSR, ST, ECG, BV by plethysmograph, EOG and EMG of Trapezius and Zygomaticus muscles analysis of 32 participants were recorded. The frontal face video of 22 participants were also recorded. Available at: <a href="https://www.eecs.qmul.ac.uk/mmv/datasets/deap/">https://www.eecs.qmul.ac.uk/mmv/datasets/deap/</a>   | Valence, Arousal and like/dislike ratings  | Peripheral - physiological signals and EEG,                                     |
| Spontaneous expression database [98] | Dataset include 1274 video clips with 30184 images from 82 students (29 males and 53 females) are included in the database.  | Enjoyment, Confusion, fatigue, Distraction and Neutral   | Spontaneous facial expression   |
| Multimodal database [99]             | Database contain the recordings from 16 professional actors (8 males and 16 females) under studio conditions   | Fear, surprise, anger, sadness, happiness, disgust   | facial expressions, body movement and gestures                                  |
| EU Emotion Stimulus [100]            | 82 contextual social science specific scene clips, 249 clips of body gesture specific scenes, 87 clips of Speech stimuli from 2364 recordings by 19 professional actors. Totally 21 emotions and mental states are represented.  | Afraid, neutral, angry, bored, ashamed, disgusted, disappointed, excited, happy, frustrated, jealous, joking, hurt, interested, proud, sad, kind, sneaky, surprised, worried and unfriendly, | Facial Expression, Vocal Expression, Contextual social scene and body gestures, |
| HEU [101]                            | Dataset contains 19,004 video clips and 9951 subjects from face and body gestures. It is divided as two parts. HEU Part 1 include two modalities (face and body gesture) and ten emotions and contain the videos from Google, Tumblr and Giphy, including ten emotions and And HEU-part 2: include 3 modalities (facial expression, body gesture and emotional speech) with around ten emotions and contain the clips from movies, TV series and different shows | Bored, anger, confused, disgust, disappointed, fear, neutral, happy, sad, surprise   | (Facial Expression, Body Gesture, and Vocal Expression)                         |

### 3. RESULTS AND DISCUSSION

Student attentivity in a learning environment is directly associated with the perception of student's pedagogical activities in the class. Adopting proper teaching and learning approaches can always improve student engagement during learning activities. MOOCs are transforming the environment of how people learn, as it is providing good quality educational contents in huge range at a lower cost. However, the MOOCs also endure from less engagement from attendees, unidirectional information flow and lack of personal attention. Similarly, since the outbreak of corona virus, video conferencing technologies such as Microsoft Teams, Zoom, Google meet, and Skype have received a huge popularity because of their increased usage in schools and universities in the past two years. These online conferencing tools created a need for monitoring student engagement in real time. Very few works have been only reported in the real time cases. Hence such an online

environment it is very difficult to the instructor to track the involvement of students, hence here lies the importance of a proper engagement detection system. The same is applicable in traditional classroom with large number of students. Student engagement will increase their satisfaction level, enhances their motivation to learn and reduces the sense of isolation.

Because of the rapid advancement in intelligent systems, many researchers proposed and implemented systems for automatic measuring engagement in an educational domain. Some of the researchers adopted the computer vision-based methods to extract the information related to the cognitive, affective and behavioural states of learners during learning activities. Most of the works done so far in engagement recognition is in laboratory-controlled environment, hence analysis in natural scenarios is still more challenging as we need to take into consideration of many more aspects such as memory requirement and computational complexity. Most of the existing works are based on images or videos of facial expression, as human face reflects most of the emotions. But the images or videos in most of the existing databases are subjected to occlusions, head-pose changes, illumination variations and bias identity. Hence extracting features from such images or video segments are difficult and may cause loss of information. The video segmentation errors also cause difficulty in extracting features from the regions of interest.

As a result of fast-growing technology, the application of sensors in educational research, health monitoring has increased. When we integrate different sensors associated with different physiological signal measurements, more intelligent and more efficient detection of student attentivity is possible. More work needs to be carried out by using sensor data fusion. In earlier time, for collecting physiological signals, we need to connect electrodes and other things create discomfort to students. But now multiple sensors are available in different wearable gadgets like smart watch, wrist band, and T shirts. Hence obtaining physiological signals from learners is not a big task. But still, many of these gadgets lacks mechanisms in analyzing or processing data, as they allow to monitor only. By using embedded machine learning by IoT, the analysis of sensor data is possible. In addition, the technical aspect of wearable devices like battery performance, intercommunication for the transfer of data, and sensitivity should be improved.

Various wearables are used in recent research studies for facial expression, eye movement, emotion identification by physiological signal, gestures and postures. Smart glasses from oculus quest pro (Tracking of eye and face), glasses for eye tracking from iMotions, Google and Apple, Eye Tribe (Eye tracking device used in desktop computers and tablets), can be used for facial emotion recognition and eye movement tracking.

So far, we have discussed diverse works in different types modalities or the combination of multiple modalities. Most of the works shows facial expression based on images or videos is the prime modality for affect state identification. Studies show, the combination of multiple modalities improve accuracy of detection. The combination of facial expressions with different modalities will increase the exactness of student engagement detection.

The expressions corresponding to engagement are compound emotions. A comprehensive dataset exclusively for engagement recognition with natural expressions, are not available. As we discussed, many researchers attempted and come up with few datasets, with limited data. And in most of the datasets, the subjects are captured in wild environment.

#### 4. CONCLUSION

Automatic engagement detection methods are finding gravity in the current online and offline education system. The machine learning or deep learning approaches used for this application contributes a lot to the research area of human computer interaction. Different engagement recognition systems were proposed and implemented in real time based on facial expressions, eye movement, gestures, postures, and physiological signals. Still research is ongoing to increase the accuracy and reliability in providing the involvement status of a student in the learning environment (both online and traditional classroom). This paper presents a rigorous study of the various engagement detection approaches and algorithms used to substantiate and test the state-of-the-art methods for the identification of affective states of students during learning. In this study we have discussed and summarized the various available datasets in different modalities used in the reviewed papers for the engagement detection. One of the main conclusions drawn from our study is that the accuracy of engagement detection system improves with the increase in the number of modalities used.

#### ACKNOWLEDGEMENTS

We sincerely thank Middle East College for providing the space and facilities for doing this survey paper. We also thank Dr. Priya Mathew, Head of Center for Academic Writing, Middle East College for reviewing the paper.

## REFERENCES

- [1] H. Lei, Y. Cui, and W. Zhou, "Relationships between student engagement and academic achievement: a meta-analysis," *Soc Behav Pers*, vol. 46, no. 3, 2018, doi: 10.2224/sbp.7054.
- [2] H. Monkarezi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated detection of engagement using video-based estimation of facial expressions and heart rate," *IEEE Trans Affect Comput*, vol. 8, no. 1, 2017, doi: 10.1109/TAFFC.2016.2515084.
- [3] M. Sagaljit, K. Sekhon, and S. Patil, "Student engagement in traditional learning vs online learning-a comparative study," *Palarch's Journal of Archaeology of Egypt*, vol. 18, 2021.
- [4] J. A. Fredricks, P. Blumenfeld, and A. H. Paris, "School engagement: potential of the concept, state of the evidence, review of educational research, 2004," *Rev Educ Res*, vol. 74, no. 1, pp. 59-109, 2004.
- [5] A. R. Anderson, S. L. Christenson, M. F. Sinclair, and C. A. Lehr, "Check & Connect: The importance of relationships for promoting engagement with school," *J Sch Psychol*, vol. 42, no. 2, 2004, doi: 10.1016/j.jsp.2004.01.002.
- [6] J. Reeve and C. M. Tseng, "Agency as a fourth aspect of students' engagement during learning activities," *Contemp Educ Psychol*, vol. 36, no. 4, 2011, doi: 10.1016/j.cedpsych.2011.05.002.
- [7] J. Shen, H. Yang, J. Li, and Z. Cheng, "Assessing learning engagement based on facial expression recognition in MOOC's scenario," *Multimedia Systems*, vol. 28, no. 2, 2022, doi: 10.1007/s00530-021-00854-x.
- [8] N. K. Mehta, S. S. Prasad, S. Saurav, R. Saini, and S. Singh, "Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement," *Applied Intelligence*, vol. 52, no. 12, 2022, doi: 10.1007/s10489-022-03200-4.
- [9] P. Goldberg *et al.*, "Attentive or not? toward a machine learning approach to assessing students' visible engagement in classroom instruction," *Educational Psychology Review*, vol. 33, no. 1, 2021, doi: 10.1007/s10648-019-09514-z.
- [10] V. L. Camacho, E. D. la Guia, T. Olivares, M. J. Flores, and L. Orozco-Barbosa, "Data capture and multimodal learning analytics focused on engagement with a new wearable IoT approach," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, 2020, doi: 10.1109/TLT.2020.2999787.
- [11] M. Awais *et al.*, "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Internet Things J*, vol. 8, no. 23, 2021, doi: 10.1109/JIOT.2020.3044031.
- [12] I. Dubovi, "Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology," *Comput Educ*, vol. 183, 2022, doi: 10.1016/j.compedu.2022.104495.
- [13] M. H. Taheri, D. J. Brown, and N. Sherkat, "Modeling engagement with multimodal multisensor data: the continuous performance test as an objective tool to track flow virtual reality view project GOET-game on extra time view project," *International Journal of Computer and Information Engineering*, vol. 14, no. 162, pp. 197-208, 2020.
- [14] A. Psaltis, K. C. Apostolakis, K. Dimitropoulos, and P. Daras, "Multimodal student engagement recognition in prosocial games," *IEEE Trans Games*, vol. 10, no. 3, 2018, doi: 10.1109/TGIAIG.2017.2743341.
- [15] A. Sukumaran, "A brief review of conventional and deep learning approaches in facial emotion recognition," *Artificial Intelligence for Internet of Things*, vol. 101, 2019.
- [16] B. Fasel and J. Luetin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, 2003, doi: 10.1016/S0031-3203(02)00052-3.
- [17] M. Ghiselin, P. Ekman, and H. E. Gruber, "Darwin and facial expression: a century of research in review," *Syst Zool*, vol. 23, no. 4, 1974, doi: 10.2307/2412481.
- [18] Y. Khaireddin and Z. Chen, "Facial emotion recognition: state of the art performance on FER2013 graphene charge density characterization by raman spectroscopy view project performance improvement on oxide thin film transistor view project," 2021, *arXiv: 2105.03588*.
- [19] Institute of Electrical and Electronics Engineers, *2017 IEEE 2nd International Conference on Signal and Image Processing, ICSIP: August 4-6, 2017*.
- [20] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors (Switzerland)*, vol. 18, no. 2, 2018, doi: 10.3390/s18020401.
- [21] S. K. D'Mello, S. D. Craig, and A. C. Graesser, "Multimethod assessment of affective experience and expression during deep learning," *International Journal of Learning Technology*, vol. 4, no. 3/4, 2009, doi: 10.1504/ijlt.2009.028805.
- [22] M. Khademi, M. T. Manzuri-Shalmani, M. H. Kiapour, and A. A. Kiaei, "Recognizing combinations of facial action units with different intensity using a mixture of hidden Markov models and neural network," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5997, LNCS, 2010, doi: 10.1007/978-3-642-12127-2\_31.
- [23] Y. L. Tian, T. Kanade, and J. F. Conn, "Recognizing action units for facial expression analysis," *IEEE Trans Pattern Anal Mach Intell*, vol. 23, no. 2, 2001, doi: 10.1109/34.908962.
- [24] J. F. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. C. Lester, "Automatically recognizing facial expression: Predicting engagement and frustration," in *Proceedings of the 6th International Conference on Educational Data Mining, EDM 2013*, 2013.
- [25] S. Polikovskiy, Y. Kameda, and Y. Ohta, "Facial micro-expression detection in hi-speed video based on facial action coding system (FACS)," *IEICE Trans Inf Syst*, vol. E96-D, no. 1, 2013, doi: 10.1587/transinf.E96.D.81.
- [26] E. A. Clark *et al.*, "The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review," *Frontiers in Psychology*, vol. 11, 2020, doi: 10.3389/fpsyg.2020.00920.
- [27] I. Alkabbany, A. Ali, A. Farag, I. Bennett, M. Ghanoum, and A. Farag, "Measuring student engagement level using facial information," in *Proceedings - International Conference on Image Processing, ICIP*, 2019, vol. 2019-September, doi: 10.1109/ICIP.2019.8803590.
- [28] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, 2019, doi: 10.1109/TAFFC.2017.2731763.
- [29] B. T. McDaniel, S. D'Mello, B. G. King, P. Chipman, K. Tapp, and A. C. Graesser, "Facial features for affective state detection in learning environments," in *Proceedings of the 29th Annual Cognitive Science Society*, 2007.
- [30] G. Littlewort *et al.*, "The computer expression recognition toolbox (CERT)," in *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, 2011, doi: 10.1109/FG.2011.5771414.
- [31] N. Bosch, Y. Chen, and S. D'Mello, "It's written on your face: Detecting affective states from facial expressions while learning computer programming," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8474, LNCS, 2014, doi: 10.1007/978-3-319-07221-0\_5.
- [32] G. C. Littlewort, M. S. Bartlett, L. P. Salamanca, and J. Reilly, "Automated measurement of children's facial expressions during problem solving tasks," in *2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops*, 2011, doi: 10.1109/FG.2011.5771418.




- [33] S. K. D'Mello and R. A. Calvo, "Significant accomplishments, new challenges, and new perspectives," in *New Perspectives on Affect and Learning Technologies*, 2011, doi: 10.1007/978-1-4419-9625-1\_19.
- [34] R. S. J. D. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments," *International Journal of Human Computer Studies*, vol. 68, no. 4, 2010, doi: 10.1016/j.ijhcs.2009.12.003.
- [35] J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," *IEEE Trans Affect Comput*, vol. 5, no. 1, 2014, doi: 10.1109/TAFFC.2014.2316163.
- [36] O. M. Nezami, M. Dras, L. Hamey, D. Richards, S. Wan, and C. Paris, "Automatic recognition of student engagement using deep learning and facial expression," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11908 LNAI, 2020, doi: 10.1007/978-3-030-46133-1\_17.
- [37] Z. A. T. Ahmed, M. E. Jadhav, A. M. Al-madani, M. Tawfik, S. N. Alsubari, and A. A. A. Shareef, "Real-time detection of student engagement: deep learning-based system," in *International Conference on Innovative Computing and Communications*, 2022, pp. 313-323, doi: 10.1007/978-981-16-2594-7\_26.
- [38] F. K. Oluwalola, "Effect of emotion on distance e-learning — the fear of technology," *International Journal of Social Science and Humanity*, vol. 5, no. 11, 2015, doi: 10.7763/ijssh.2015.v5.588.
- [39] J. Yi, B. Sheng, R. Shen, W. Lin, and E. Wu, "Real time learning evaluation based on gaze tracking," in *Proceedings - 2015 14th International Conference on Computer-Aided Design and Computer Graphics, CAD/Graphics 2015*, 2016, doi: 10.1109/CADGRAPHICS.2015.13.
- [40] C. J. Huang *et al.*, "Implementation and performance evaluation of an intelligent online argumentation assessment system," in *Proceedings - International Conference on Electrical and Control Engineering, ICECE 2010*, 2010, doi: 10.1109/ICECE.2010.632.
- [41] N. Ghatashah, "Knowledge level assessment in e-learning systems using machine learning and user activity analysis," *International Journal of Advanced Computer Science and Applications*, vol. 6, no. 4, 2015, doi: 10.14569/ijacsa.2015.060415.
- [42] C. Calvi, M. Porta, and D. Sacchi, "eLearning, an e-learning environment based on eye tracking," in *Proceedings - The 8th IEEE International Conference on Advanced Learning Technologies, ICALT 2008*, 2008, doi: 10.1109/ICALT.2008.35.
- [43] D. P. Munoz, I. Armstrong, and B. Coe, "Using eye movements to probe development and dysfunction," in *Eye Movements: A Window on Mind and Brain*, 2007, doi: 10.1016/B978-008044980-7/50007-0.
- [44] M. Moghaddasi, J. Marín-Morales, J. Khatri, J. Guixeres, I. A. C. Giglioli, and M. Alcañiz, "Recognition of customers' impulsivity from behavioral patterns in virtual reality," *Applied Sciences (Switzerland)*, vol. 11, no. 10, 2021, doi: 10.3390/app11104399.
- [45] L. Qin, Q. L. Cao, A. S. Leon, Y. N. Weng, and X. H. Shi, "Use of pupil area and fixation maps to evaluate visual behavior of drivers inside tunnels at different luminance levels—a pilot study," *Applied Sciences (Switzerland)*, vol. 11, no. 11, 2021, doi: 10.3390/app11115014.
- [46] Y. I. Giraldo-Romero, C. Pérez-De-Los-Cobos-Agüero, F. Muñoz-Leiva, E. Higuera-Castillo, and F. Liébana-Cabanillas, "Influence of regulatory fit theory on persuasion from google ads: An eye tracking study," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 5, 2021, doi: 10.3390/jtaer16050066.
- [47] M. C. Sáiz-Manzanares, I. R. Pérez, A. A. Rodríguez, S. R. Arribas, L. Almeida, and C. F. Martín, "Analysis of the learning process through eye tracking technology and feature selection techniques," *Applied Sciences (Switzerland)*, vol. 11, no. 13, 2021, doi: 10.3390/app11136157.
- [48] K. N. Daniel and E. Kamioka, "Detection of learner's concentration in distance learning system with multiple biological information," *Journal of Computer and Communications*, vol. 05, no. 04, 2017, doi: 10.4236/jcc.2017.54001.
- [49] S. Mu, M. Cui, X. J. Wang, J. X. Qiao, and D. M. Tang, "Learners' attention preferences and learning paths on online learning content: An empirical study based on eye movement," in *Proceedings - 2018 7th International Conference of Educational Innovation through Technology, EITT 2018*, 2018, doi: 10.1109/EITT.2018.00015.
- [50] S. K. Shalini, A. K. Philip, and D. R. Ravindra, "European journal of molecular and clinical medicine students attention and engagement prediction using machine learning techniques," *European Journal of Molecular and Clinical Medicine*, vol. 7, no. 4, pp. 3011-3017, 2020.
- [51] P. Sharma, S. Joshi, S. Gautam, S. Maharjan, V. Filipe, and M. C. Reis, "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning," *arXiv preprint arXiv:1909.12913*, 2019.
- [52] H. Khachatryan and A. L. Rihn, "Eye-tracking methodology and applications in consumer research," *EDIS*, vol. 2014, no. 7, 2014, doi: 10.32473/edis-fe947-2014.
- [53] A. Clemotte, M. Velasco, D. Torricelli, R. Raya, and R. Ceres, "Accuracy and precision of the tobi X2-30 eye-tracking under non ideal conditions," in *NEUROTECHNIX 2014 - Proceedings of the 2nd International Congress on Neurotechnology, Electronics and Informatics*, 2014, doi: 10.5220/0005094201110116.
- [54] I. Rodríguez-Moreno, J. M. Martínez-Otzeta, B. Sierra, I. Rodríguez, and E. Jauregi, "Video activity recognition: State-of-the-art," *Sensors (Switzerland)*, vol. 19, no. 14, 2019, doi: 10.3390/s19143160.
- [55] J. Carreira and A. Zisserman, "Quo Vadis, action recognition? A new model and the kinetics dataset," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, doi: 10.1109/CVPR.2017.502.
- [56] J. Wang, A. Cherian, F. Porikli, and S. Gould, "Video representation learning using discriminative pooling," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, doi: 10.1109/CVPR.2018.00126.
- [57] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," *IEEE Trans Affect Comput*, vol. 4, no. 1, 2013, doi: 10.1109/T-AFFC.2012.16.
- [58] Y. Gao *et al.*, "Vision-based hand gesture recognition for human-computer interaction—a survey," *Wuhan University Journal of Natural Sciences*, vol. 25, no. 2, 2020, doi: 10.19823/j.cnki.1007-1202.2020.0020.
- [59] A. Pease and B. Pease, "The definitive book of body language," *Bantam*, vol. 1, 2006.
- [60] A. Behera, P. Matthew, A. Keidel, P. Vangorp, H. Fang, and S. Canning, "Associating facial expressions and upper-body gestures with learning tasks for enhancing intelligent tutoring systems," *Int J Artif Intell Educ*, vol. 30, no. 2, 2020, doi: 10.1007/s40593-020-00195-2.
- [61] P. Ekman, "Lie catching and microexpressions," in *The Philosophy of Deception*, 2011, doi: 10.1093/acprof:oso/9780195327939.003.0008.
- [62] S. Khenkar and S. K. Jarraya, "Engagement detection based on analyzing micro body gestures using 3D CNN," *Computers, Materials and Continua*, vol. 70, no. 2, 2022, doi: 10.32604/cmc.2022.019152.
- [63] V. Riemer, J. Frommel, G. Layher, H. Neumann, and C. Schrader, "Identifying features of bodily expression as indicators of emotional experience during multimedia learning," *Front Psychol*, vol. 8, no. Jul, 2017, doi: 10.3389/fpsyg.2017.01303.
- [64] H. G. Wallbott and K. R. Scherer, "Cues and channels in emotion recognition," *J Pers Soc Psychol*, vol. 51, no. 4, 1986, doi: 10.1037/0022-3514.51.4.690.
- [65] S. Garger, "Is there a link between learning style and neurophysiology?," *Educational Leadership*, vol. 48, no. 2, 1990.

- [66] S. Sathayanarayana *et al.*, "Towards automated understanding of student-tutor interactions using visual deictic gestures," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2014, doi: 10.1109/CVPRW.2014.77.
- [67] P. D. Drummond and S. H. Quah, "The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians," *Psychophysiology*, vol. 38, no. 2, 2001, doi: 10.1017/S004857720199095X.
- [68] A. Dupre, S. Vincent, and P. A. Iaizzo, "Basic ECG theory, recordings, and interpretation," in *Handbook of Cardiac Anatomy, Physiology, and Devices*, 2005, doi: 10.1007/978-1-59259-835-9\_15.
- [69] M. Robertson and M. Robertson, "Book review: the brain that changes itself: stories of personal triumph from the frontiers of brain science," *Australasian Psychiatry*, vol. 17, no. 3, 2009, doi: 10.1080/10398560902721606.
- [70] L. Scott, "Engaged-Learning: Community Engagement Classifications," *Teachers College, Columbia University*, 2012.
- [71] H. Jang, J. Reeve, and E. L. Deci, "Engaging students in learning activities: it is not autonomy support or structure but autonomy support and structure," *J Educ Psychol*, vol. 102, no. 3, 2010, doi: 10.1037/a0019682.
- [72] J. Perdiz, G. Pires, and U. J. Nunes, "Emotional state detection based on EMG and EOG biosignals: A short survey," in *ENBENG 2017 - 5th Portuguese Meeting on Bioengineering, Proceedings*, 2017, doi: 10.1109/ENBENG.2017.7889451.
- [73] G. N. Levine *et al.*, "Meditation and cardiovascular risk reduction," *J Am Heart Assoc*, vol. 6, no. 10, 2017, doi: 10.1161/JAHA.117.002218.
- [74] K. M. S. Jamal and E. Kamioka, "Emotions detection scheme using facial skin temperature and heart rate variability," *MATEC Web of Conferences*, vol. 277, 2019, doi: 10.1051/mateconf/201927702037.
- [75] M. V. Villarejo, B. G. Zapirain, and A. M. Zorrilla, "A stress sensor based on galvanic skin response (GSR) controlled by ZigBee," *Sensors (Switzerland)*, vol. 12, no. 5, 2012, doi: 10.3390/s120506075.
- [76] K. S. McNeal, J. M. Spry, R. Mitra, and J. L. Tipton, "Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course," *Journal of Geoscience Education*, vol. 62, no. 4, 2014, doi: 10.5408/13-111.1.
- [77] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, 2007, doi: 10.1088/0967-3334/28/3/R01.
- [78] J. A. Kruse, "Clinical methods: the history, physical, and laboratory examinations," *JAMA: The Journal of the American Medical Association*, vol. 264, no. 21, 1990, doi: 10.1001/jama.1990.03450210108045.
- [79] M. Bustos-López, N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, N. A. Cruz-Ramos, and G. Alor-Hernández, "Wearables for engagement detection in learning environments: a review," *Biosensors*, vol. 12, no. 7. MDPI, Jul. 01, 2022, doi: 10.3390/bios12070509.
- [80] A. Belle, R. H. Hargraves, and K. Najarian, "An automated optimal engagement and attention detection system using electrocardiogram," *Comput Math Methods Med*, vol. 2012, 2012, doi: 10.1155/2012/528781.
- [81] A. Apicella *et al.*, "EEG-based measurement system for student engagement detection in learning 4.0," 2021, doi: 10.21203/rs.3.rs-944837/v1.
- [82] A. Burns *et al.*, "SHIMMERTM - A wireless sensor platform for noninvasive biomedical research," *IEEE Sens J*, vol. 10, no. 9, 2010, doi: 10.1109/JSEN.2010.2045498.
- [83] M. Carroll *et al.*, "Automatic detection of learner engagement using machine learning and wearable sensors," *J Behav Brain Sci*, vol. 10, no. 03, 2020, doi: 10.4236/jbbs.2020.103010.
- [84] A. B. Khedher, I. Jraidi, and C. Frasson, "Tracking students' mental engagement using EEG signals during an interaction with a virtual learning environment," *Journal of Intelligent Learning Systems and Applications*, vol. 11, no. 01, 2019, doi: 10.4236/jilsa.2019.111001.
- [85] M. S. Benlamine, J. Harley, C. Frasson, and A. Dufresne, "A. toward brain-based gaming: measuring engagement during gameplay," *EdMedia+ Innovate Learning*, 2015.
- [86] C. Berka *et al.*, "EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks," *Aviat Space Environ Med*, vol. 78, no. 5 II, 2007.
- [87] F. R. Lin and C. M. Kao, "Mental effort detection using EEG data in E-learning contexts," *Comput Educ*, vol. 122, 2018, doi: 10.1016/j.compedu.2018.03.020.
- [88] M. Chaouachi, I. Jraidi, and C. Frasson, "Modeling mental workload using EEG features for intelligent systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6787 LNCS, 2011, doi: 10.1007/978-3-642-22362-4\_5.
- [89] P. Pham and J. Wang, "Attentivelearner: Improving mobile MOOC learning via implicit heart rate tracking," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9112, pp. 367–376, doi: 10.1007/978-3-319-19773-9\_37.
- [90] F. Sabry, T. Eltaras, W. Labda, K. Alzoubi, and Q. Malluhi, "Machine learning for healthcare wearable devices: the big picture," *Journal of Healthcare Engineering*, vol. 2022. Hindawi Limited, 2022, doi: 10.1155/2022/4653923.
- [91] Q. Wei, B. Sun, J. He, and L. Yu, "BNU-LSVED 2.0: Spontaneous multimodal student affect database with multi-dimensional labels," *Signal Process Image Commun*, vol. 59, 2017, doi: 10.1016/j.image.2017.08.012.
- [92] N. Hara and R. Kling, "Students' frustrations with a web-based distance education course," *First Monday*, vol. 4, no. 12, 1999, doi: 10.5210/fm.v4i12.710.
- [93] S. Govaerts and J. Grégoire, "Development and construct validation of an academic emotions scale," *Int J Test*, vol. 8, no. 1, 2008, doi: 10.1080/15305050701808649.
- [94] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "DAiSEE: towards user engagement recognition in the wild," *arXiv preprint arXiv:1609.01885*, Sep. 2016, [Online]. Available: <http://arxiv.org/abs/1609.01885>
- [95] A. Kaur, A. Mustafa, L. Mehta, and A. Dhall, "Prediction and localization of student engagement in the wild," in *2018 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2018*, 2019, doi: 10.1109/DICTA.2018.8615851.
- [96] S. Batra *et al.*, "DMCNet: Diversified model combination network for understanding engagement from video screengrabs," *Systems and Soft Computing*, vol. 4, 2022, doi: 10.1016/j.sasc.2022.200039.
- [97] S. Koelstra *et al.*, "DEAP: a database for emotion analysis; Using physiological signals," *IEEE Trans Affect Comput*, vol. 3, no. 1, 2012, doi: 10.1109/T-AFFC.2011.15.
- [98] C. Bian, Y. Zhang, F. Yang, W. Bi, and W. Lu, "Spontaneous facial expression database for academic emotion inference in online learning," *IET Computer Vision*, vol. 13, no. 3, 2019, doi: 10.1049/iet-cvi.2018.5281.
- [99] T. Sapiński, D. Kamińska, A. Pelikant, C. Ozcinar, E. Avots, and G. Anbarjafari, "Multimodal database of emotional speech, video and gestures," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11188, LNCS, 2019, doi: 10.1007/978-3-030-05792-3\_15.
- [100] H. O'Reilly *et al.*, "The EU-emotion stimulus set: a validation study," *Behav Res Methods*, vol. 48, no. 2, 2016, doi: 10.3758/s13428-015-0601-4.




- [101] J. Chen *et al.*, "HEU Emotion: a large-scale database for multimodal emotion recognition in the wild," *Neural Comput Appl*, vol. 33, no. 14, 2021, doi: 10.1007/s00521-020-05616-w.

## BIOGRAPHIES OF AUTHORS



**Ajitha Sukumaran**    received B.Tech. Degree in Electronics and Communication from Mar Athanasius College of Engineering, Kerala, India in 2002 and M Tech in Digital System and Communication Engineering from National Institute of Technology, Kerala, India in 2005. She is currently pursuing the Ph.D. degree in the area of Digital Image Processing at Vellore Institute of Technology, Tamil Nadu, India. Currently she is working as a Senior Lecturer in the Department of Electronics and Communication at Middle East College. Her expertise includes nonlinear chaotic communication, cryptography, coding theory and Multiuser detection. Over the last 15 years, she has taught several topics and programmes related to Electronics and Communication to students in highly reputed higher education institutions in India and Oman. She is a fellow of the Higher Education Academy in recognition of attainment against the UK Professional Standards Framework for teaching and learning support in higher education. She can be contacted at email: [ajitha@mec.edu.om](mailto:ajitha@mec.edu.om).



**Arun Manoharan**    received the Ph.D. degree in High Performance Computer Networks from Anna University, Tamilnadu, India in 2011. He was a Post-Doctoral Researcher at Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Portugal. His present research work focuses on the Edge Level Security Standards for LoRaWAN networks. His research contributions are towards the development of High Performance Heterogeneous Computing Algorithms for compute intensive use cases like CFD, DNA Sequence Search and Space datasets. His research results in the publication of 40 plus research papers indexed in Scopus and Web of Science. He has 18 years of Engineering Academic and Research Experience. He secured funded projects worth of INR 3500000 from Indian Government Agencies like Department of Science and Technology (DST), Indian Centre for Medical Research (ICMR) and AICTE. He worked for industrial consultancy projects to provide vision based solutions to textile and manufacturing industries problems in collaboration with National Instruments, NVIDIA and Titan India. At present, he is working as Associate Professor and Head of the Department of Embedded Technology at Vellore Institute Technology, India. He can be contacted at email: [arunm@vit.ac.in](mailto:arunm@vit.ac.in).