

From Distraction to Interaction: Investigating Learner Engagement Challenges in Virtual Classrooms

Muhammad Irfan¹, Preeti Patel¹ and Bilal Hassan¹

¹ London Metropolitan University, 166-220 Holloway Road, London, UK
{mui0316@my.londonmet.ac.uk, p.patel@londonmet.ac.uk,
b.hassan@londonmet.ac.uk}

Abstract. Digital transformation in education, the associated advancements in technology and the global pandemic have led to an unprecedented acceleration of virtual learning and teaching, disrupting the educational environment irreversibly. COVID-19, in particular, necessitated a global and swift adaptation by all teaching institutions to virtual learning methods. For universities, the transition to online learning predominantly focused on migrating teaching content, leaving online pedagogy, social interactions and informal learning areas, largely unattended. Given that virtual classrooms are now used consistently in many educational settings, and given that there is a strong correlation between learner engagement and successful outcomes, the ability to effectively monitor, evaluate and modify that engagement is crucial. This paper investigates the current pedagogical and technological methods that have been proposed to enhance learner engagement. It identifies progress, opportunities, limitations and barriers that have been observed by researchers in this field through a systematic literature review from studies conducted between 2020 and 2024. Areas that have potential include hybrid combinations of human pose dynamics, facial and eye movement tracking; integrated with predictive modeling capabilities.

Keywords: Learner engagement, virtual classroom, education

1 Introduction

Advancements in technology have changed the traditional education environment into a digital online learning environment. This rapid digitization of education is accelerated due to the pandemic crisis that resulted in unprecedented changes to the traditional educational landscape [1]. The World Health Organization (WHO) declared the global pandemic in March 2020 [2], prompting universities worldwide to swiftly adapt to virtual teaching methods [3]. Schools and colleges had to shift instruction online, necessitating students to access their education primarily through computers [2]

For universities, the transition to online learning predominantly focused on migrating teaching content, leaving online pedagogy, social interactions, and informal learning areas largely unattended [4]. Throughout the pandemic, various online platforms were utilized to facilitate interactions between instructors and students, supporting higher education [4]. While terms such as distance learning and eLearning have been associated with digital transformation in education for over two decades (Crawford et al., 2020 [4]), the global shift towards digital education accelerated significantly during the pandemic [5]. Where digitalization of education provides ease and flexibility for instructors and students, there continue to be significant challenges facing educators. These include ensuring high levels of student engagement during online lectures, delivery of quality interactive course content, and minimizing distractions, whilst enhancing instructor-student communication. Pedagogic strategies to increase the effectiveness of online education have to address and understand the behavior of students during online lectures. Student engagement is characterized by students who are interested and involved for the entire duration of the session, and who are less likely to be distracted [6]. Engagement is directly proportional to student success. Behavior and movements of instructors can enhance participants' interest during online learning and some of the most motivating educators' practice is promoting active involvement, creating an educator's presence, clear expectations and educators positive body movements including facial expressions, gaze and eye movements [7].

Both higher education and corporate training sectors are increasingly transitioning to virtual delivery methods, making the assessment of online participant engagement critical [8]. Through a variety of techniques, such as facial expressions, text-based inquiries, interactive conversations, quizzes, and meeting replies, session educators try to determine the level of participant comprehension. The standard testing techniques that have worked successfully in traditional settings frequently show to be unworkable in the world of virtual education and training.

Previous studies show that behavior and movements of teacher/instructor enhance student engagement and different strategies are used to determine the interest level of participants in online learning. Studies also show that detecting student involvement levels can help us determine impact of online education and effectiveness of various online platforms like Zoom, Skype, Google Meet and MS Teams. The aim of this paper is to provide an overview of existing methods used for detection of student engagement in online learning or education. For this purpose, the we have systematically reviewed studies that address detection of learners' engagement and behaviors during e-learning or online education. We have analyzed the methodologies, objectives, frameworks and limitations in published methods in studies published in the period 2020 to 2024. The focus of the paper has been to review and analyse the available literature systematically to acquire insight into the behavioral analysis and facial factors that contribute significantly to student engagement.

2 Methodology

The grounded theory approach has been used for this systematic literature re- view, following the theory proposed by [9]. This theory ensures breadth and depth and allows new themes, opportunities, challenges and inconsistencies to emerge. The acquired approach consists of five main phases (define, search, select, analyze, and present) as illustrated in Figure 1.

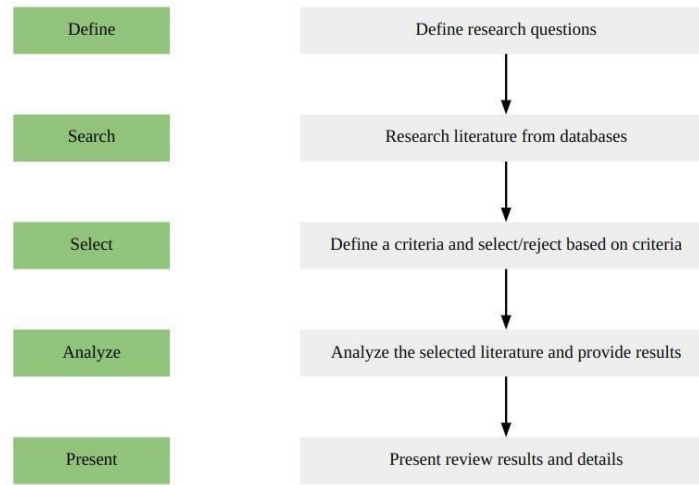


Figure 1. Literature review methodology

Firstly, research questions are formulated to guide the investigation [10].

- What specific learner features contribute significantly to the effectiveness of prediction models in identifying student engagement levels?
- How does integration of facial expressions, eye movements tracking and human pose dynamics enhance the accuracy of detection of student engagement in real time online education?
- What are the most effective strategies for mitigating privacy concerns while collecting and sharing video data for research purposes in the field of student engagement detection?

Secondly, the search scope is defined and restricted to computer science, social science, human computer interactions, education, computer vision, artificial intelligence and privacy. This study focused on automated and manual search processes to gain as many research articles as possible to fulfill the defined objectives. Queries and keywords (Table 1) were developed according to research objectives to search popular databases such as Scopus, IEEE Xplore, Science Direct, Web of Science, AIS Electronic Library and ACM Digital Library. The search engines Google Scholar, Yahoo! And RefSeek were also utilized during the additional manual search phase, in which reference lists of related articles were also accessed. Full texts and downloaded abstracts were thoroughly explored and directly relevant and closely related articles were selected.

Table 1. Keywords and queries

Keyword	Query
Face Detection	“face detection using CV” OR “face features extraction” OR “face soft metrics”
Emotion Detection	“emotion detection using deep learning” OR “emotion classification” OR “real time motion data collection”
Pose	“pose estimation using CV” OR “body dynamics extraction” OR “body movements”
Eye Tracking	“gaze detection” OR “eye movement” OR “eye tracking through machine learning”
Data Privacy	“securing data” OR “IoT platform privacy” OR “cloud data collection”
Engagement	“learner engagement” OR “digital learning engagement” OR “online education environment”

Thirdly, the selection and rejection criteria were defined.

Articles were selected for study if:

- English is the language of the article.
- Articles are related to learner engagement in online classrooms and learner behavior during digital education.
- Articles published in a journal between 2020 and 2024

Articles were rejected for review if:

- The article is related to learner engagement but not based on online education or virtual classroom environment.
- Articles related to student online presence detection but not engagement detection.
- Articles without any methodological evidence.
- A book, magazine, thesis, or a report.

The application of the above selection criteria resulted in 25 articles related to detection of student engagement in online learning, from which 18 articles were included and 7 articles were excluded.

3 A review of the contribution in the field

This section provides a review based on literature sources selected according to the criteria defined earlier. Table 2 presents a summary of the key information of each literature source. In general, researchers attempt to use computer vision to measure how engaged students appear or instructors try to ascertain whether students are engaged or not in a regular classroom.

Having an automated system that provides this estimation would be helpful for online classrooms. In this regard, several researchers such as [11] have proposed a Convolutional Neural Network (CNN) based framework to detect engagement in real-time.

In [11] a cost effective approach of using a webcam to capture video and feed it to CNN model is provided, which represents the level of engagement of the students. Whilst [15] performed a survey and collected their own dataset, in which the calculated Mean Engagement Score (MES) based on facial and emotion detection using CNN was employed. Their method outperforms previous methods due to usage and collection of real time data. The [16] study prioritizes the importance of collecting both emotional and cognitive components of a learner in order to measure engagement level. They use computer vision technology, particularly the OpenPose CV library, to detect the pose of students, then using deep-learning models along with a transfer learning approach which is used to classify different poses of students to show engagement level. In [21] six basic emotions are studied and their effect on student engagement. Online and offline videos were analyzed using computer vision methods and pre-trained CNN was used to extract the six basic emotions. Subsequently engagement levels classification was deployed by using emotion recognition and concentration index. However, providing engagement detection based solely on emotions detection of the individuals is not enough. While doing student engagement detection using computer vision other features such as posture detection, eye tracking and hands movements should also be considered.

Student engagement detection can be enhanced by using multiple human behavior data or detection. [18] perform experiments using ResNet and Temporal Convolutional Networks (TCN) to enhance student engagement detection systems. They introduce a hybrid method based on ResNet and TCN, which improves student engagement detection by 3.9%. [20] purports that the fusion of affective and attention features of learners improves the estimation of engagement levels. Head poses and facial expressions are detected using trained Attention-Net and Affect-Net, which are further used to classify engagement levels. Here the engagement classifier showed area Under the Curve (AUC) values of 0.620 and 0.720. Due to the presence of multiple human behavior detection, preserving privacy for individuals becomes difficult. Also large data storage is needed to store all types of data for further large scale studies.

Humans believe that facial expressions give us good information that helps in engagement detection of individuals. Cameras provide us a best way to capture faces of learners in real time and help in detection of engagement levels. Many different methods have been proposed to automate this detection process by analyzing the face images. [6] shows real time detection of student engagement using webcam data, they used a haar-cascade algorithm for eye tracking and Convolutional Neural Networks (CNN) for engagement classification. They analyze that students with high concentration achieve high scores. [23] presented a lightweight Facial Expression Recognition (FER) model. Multitask CNN was used to detect faces, EfficientNet was used to extract facial features and Lightweight CNN was used to classify emotions. However, only facial expressions based detection is not sufficient because face video or images can sometimes be in low resolution which results in less facial expressions information and limited video based data.

Table 2: Systematic literature review of the contributions in the field

Year	Ref	Research Method	Technique Used	Model Used	Performance//Output
2023	[27]	Quantitative	Computer Vision	3D CNN	video based data used and improve student engagement detection (91%)
2023	[25]	Qualitative	Deep Learning	CNN	Provided AI-based video annotation model to (82%)
2023	[26]	Quantitative	Deep Learning	VGG-19	Achieves 95.29% accuracy of model
2022	[24]	Mixed	Interviews	Diary-Interview	Explored Swedish teachers' perceptions of student engagement
2022	[23]	Quantitative	Deep Learning	Efficient Net & CNN	Provide Lightweight FER
2022	[22]	Quantitative	Hybrid	RetinaNet, VGG16, K-means, TCN	Obtain presentation style model accuracy 86% and student engagement model accuracy 76%
2021	[15]	Survey	Computer Vision	CNN	Calculate MES and classify engagement level based on it
2021	[17]	Quantitative	Computer Vision	LGBP, LBP-TOP & FHOHO	Improve 5% identification rate of micro-expressions
2021	[21]	Mixed	Computer Vision	OpenCV	Classify student engagement levels with six basic emotions
2021	[20]	Quantitative	Deep Learning	AttentionNet & AffectNet	Engagement classifiers achieved AUCs of .620 and .720
2021	[18]	Quantitative	Hybrid	ResNet & TCN	Improve 3.9% accuracy in student engagement detection
2021	[16]	Case Study	Hybrid	CNN & Clustering	Address behavior recognition challenges
2021	[14]	Quantitative	Hybrid	LSTM & TCN	Engagement level classification accuracy of 67.4% and MSE 0.0508 on EmotiW dataset
2020	[11]	Case Study	Computer Vision	CNN	Uses web cameras for real time student engagement detection via eye tracking.
2020	[12]	Quantitative	Computer Vision	CNN	Obtain 62% to 65% accuracy
2020	[19]	Mixed	Deep Learning	IoT & LSTM	Obtain 1 ms latency and 95% accuracy
2020	[13]	Qualitative	API based	MS Emotion API	Social robots enhances classroom engagement
2019	[6]	Scenario	Computer Vision	CNN	Improves video rendering results accuracy for 93.2 %

Instructor behaviors reflect on the student engagement in online learning. By analyzing teachers behavior and presentation style, one can also estimate student engagement during online education. Many studies have been conducted to understand the relation between teachers perceptions and student engagement. [22] presented how student engagement and presentation style correlate to each other. A combination of RetinaNet, Siamese Model, Character-Region Awareness For Text Detection (CRAFT), VGG16, and K-Means Clustering was used for presentation style detection and TCN was used to detect student engagement levels. At the frame level 86% accuracy was achieved and at the video level 76% accuracy was achieved by presentation style model. An accuracy of 76% was achieved and f1-score of 0.82. By using their method, they also found that 70% of students were engaged irrespective of presentation style used. The [24] study focuses on providing valuable insights on student engagement in a digital learning environment. They used Mixed Method Grounded Theory approach using Diary-interview with Swedish teachers to extract their perception about student engagement, disengagement and behavior during online classrooms. This study had limited demographics and can only be used for understanding the relationship between student and teacher engagement during online education.

IoT provides seamless connectivity between various components during online education. IoT devices enable remote monitoring of students' progress, engagement, and well-being in online learning environments. [19] studied IoT based framework for sensing engagement level of students in online classrooms. They collected wireless signals of emotions from physical devices and sent them to the IoT platform. Acquired data is then used to detect engagement levels of students using Long Short Term Memory (LSTM) deep-learning model. In the proposed method, IoT protocols achieve ultra-low latency of 1ms and deep-learning model attain high performance with 95% accuracy. The limitation in the provided method is that sometimes the information from signals can be mixed with unwanted interference, making it hard to see the real patterns. Signals also change depending on the environment they are in, which adds to the challenge. Achieving reliable signals consistently over time can be difficult.

Table 2 provides a summary of the various research reported between 2020-2024. The predominant techniques analyzed include CNN:Convolutional Neural Networks; MES:Mean Squared Error; LGBP:Local Gabor Binary Pattern; LBP-TOP:Local Binary Patterns from Three Orthogonal Planes; FHOHO:Fuzzy Histogram of Optical Flow Orientations; FER:Facial Expression Recognition System; IoT: Internet of Things; LSTM:Long Short Term Memory; VGG:Visual Geometry Group; TCN:Temporal Convolutional Networks.

Although previous studies provide some solutions for student engagement detection in an online environment, the main challenge for all solutions is the collection of accurate data and also the restricting of participants' movement to stay within the eye of camera range which is not feasible in a real-life educational environment. We presented an analysis of the studies related to supportive factors towards engagement detection, face detection, emotion detection, pose estimation, eye tracking and data privacy.

4 Open challenges in the field

In this era of science where advanced technology provides many solutions to tackle student engagement detection in an online learning environment, it also poses some challenges that demand attention and thoughtful considerations. Following are some open challenges that still require further research in student engagement detection.

- **Diverse Dataset Acquisition:** In student engagement detection, a significant challenge lies in accessing datasets that sufficiently represent various demographics, learning styles, and cultural backgrounds. Analyzing such a dataset raises the need for advanced methodology to ensure fair and accurate assessments of engagement levels across diverse student populations and enhancing education strategies.
- **Handling Face Occlusion:** In previous studies it is found that face occlusion causes hindrance in facial emotion extraction. Occlusions can lead to inaccuracies or biases in facial emotion extraction, impacting the reliability and effectiveness of detection systems in understanding human emotions. There is a clear need of a system that could extract facial expressions from a 3D dataset.
- **Ensuring Student Privacy:** Preserving the privacy of students during facial expression detection is one of the concerns to providing secure detection. Sensitive biometric data is captured and processed during facial expression detection which leads to breaches of students' privacy rights.
- **Pedagogical Fairness Assurance:** Automated solutions may overlook student behavior and fail to capture qualitative aspects of engagement, such as critical thinking or creativity. Therefore, we should have to introduce a pedagogical method that is used by instructors to ensure they are inclusive, fair and considerate of students' needs and rights during the online learning environment engagement.
- **Need of Multimodal Data System:** Previous studies focused on behavioral and emotional engagement of learners, with less attention to cognitive engagement. There is the need of analyzing and measuring student engagement using multimodal data such as text, log, and voice. Researchers can consider negative emotions such as anxiety, tension, depression, and sadness for more precise student engagement.

5 Conclusion

The rapid digitization of education, accelerated by the COVID-19 crisis, has led to a significant shift towards online learning. Where digitization offers benefits, at the same time many challenges arise related to detecting student engagement and ensuring quality online education. In this paper, we provide a systematic literature review of available solutions to these challenges which utilise computer vision, machine learning, deep learning and IoT, that can help to detect student engagement in online environments. Consequently, further research is needed to investigate why learners disengage and how they can be re-engaged. Further research is required to find solutions that detect engagement levels whilst preserving the privacy of individual learners. Further research requires effort towards collecting more diverse video-based data for real time detection of student engagement. There is also the need of studying relationships and how engaged/disengaged behavior affects the learning outcomes of the learners.

References:

1. Han, Z., Tu, Y. and Huang, C., 2023. A framework for Constructing a technology-enhanced education metaverse: Learner engagement with human-machine collaboration. *IEEE Transactions on Learning Technologies*.
2. Anthony Jnr, B. and Noel, S., 2021. Examining the adoption of emergency remote teaching and virtual learning during and after COVID-19 pandemic. *International Journal of Educational Management*, 35(6), pp.1136-1150.
3. Klunder, S., Saab, N. and Admiraal, W., 2022. A teacher perspective on using a hybrid virtual classroom for students with a chronic illness in mainstream primary and secondary schools. *Technology, Pedagogy and Education*, 31(4), pp.493-508.
4. Crawford, J. (2020). COVID-19: 20 countries' higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, 3(1), pp.1–20. doi:10.37074/jalt.2020.3.1.7.
5. Krishnamurthy, S. (2020). The Future of Business Education: A Commentary in the Shadow of the Covid-19 Pandemic. *Journal of Business Research*, 117(1-5). doi:10.1016/j.jbusres.2020.05.034.
6. Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Filipe, V. and Reis, M.J., 1909. Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. *arXiv 2019*. arXiv preprint arXiv:1909.12913.
7. Timmis, S. and Williams, J., 2013. Students as co-researchers: a collaborative, community-based approach to the research and practice of technology enhanced learning. *The Student Engagement Handbook, Practice in Higher Education*, pp.509-525.
8. Wang, L., 2023. Students' psychology for teaching design with artificial intelligence approaches for enhancing teaching. *Soft Computing*, pp.1-18.
9. Joost F Wolfswinkel, Elfi Furtmueller & Celeste P M Wilderom (2013) Using grounded theory as a method for rigorously reviewing literature, *European Journal of Information Systems*, 22:1, 45-55, DOI: 10.1057/ejis.2011.51
10. Booth, A., Sutton, A., Clowes, M. and Martyn-St James, M., 2021. Systematic approaches to a successful literature review.
11. Wang, Y., Kotha, A., Hong, P.H. and Qiu, M., 2020, August. Automated student engagement monitoring and evaluation during learning in the wild. In 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom) (pp. 270-275). IEEE.
12. Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P.K.D., Dasgupta, N. and Choudhury, P., 2020, February. Facial emotion detection to assess Learner's State of mind in an online learning system. In *Proceedings of the 2020 5th international conference on intelligent information technology* (pp. 107-115).

13. Bourguet, M.L., Jin, Y., Shi, Y., Chen, Y., Rincon-Ardila, L. and Venture, G., 2020, December. Social robots that can sense and improve student engagement. In 2020 IEEE international conference on teaching, assessment, and learning for engineering (TALE) (pp. 127-134). IEEE.
14. Abedi, A. and Khan, S., 2021. Affect-driven Ordinal Engagement Measurement from Video. arXiv preprint arXiv:2106.10882.
15. Bhardwaj, P., Gupta, P.K., Panwar, H., Siddiqui, M.K., Morales-Menendez, R. and Bhaik, A., 2021. Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering*, 93, p.107277.
16. Vanneste, P., Oramas, J., Verelst, T., Tuytelaars, T., Raes, A., Depaepe, F. and Van den Noortgate, W., 2021. Computer vision and human behaviour, emotion and cognition detection: A use case on student engagement. *Mathematics*, 9(3), p.287.
17. Saeed, U., 2021. Facial micro-expressions as a soft biometric for person recognition. *Pattern Recognition Letters*, 143, pp.95-103.
18. Abedi, A. and Khan, S.S., 2021, May. Improving state-of-the-art in detecting student engagement with resnet and tcn hybrid network. In 2021 18th Conference on Robots and Vision (CRV) (pp. 151-157). IEEE.
19. Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S.U. and Rodrigues, J.J., 2020. LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19. *IEEE Internet of Things Journal*, 8(23), pp.16863-16871.
20. Sümer, Ö., Goldberg, P., D'Mello, S., Gerjets, P., Trautwein, U. and Kasneci, E., 2021. Multimodal engagement analysis from facial videos in the classroom. *IEEE Transactions on Affective Computing*, 14(2), pp.1012-1027.
21. Hasnine, M.N., Bui, H.T., Tran, T.T.T., Nguyen, H.T., Akçapınar, G. and Ueda, H., 2021. Students' emotion extraction and visualization for engagement detection in online learning. *Procedia Computer Science*, 192, pp.3423-3431.
22. Thomas, C., Sarma, K.P., Gajula, S.S. and Jayagopi, D.B., 2022. Automatic prediction of presentation style and student engagement from videos. *Computers and Education: Artificial Intelligence*, 3, p.100079.
23. Savchenko, A.V., Savchenko, L.V. and Makarov, I., 2022. Classifying emotions and engagement in online learning based on a single facial expression recognition neural network. *IEEE Transactions on Affective Computing*, 13(4), pp.2132-2143.
24. Bergdahl, N., 2022. Engagement and disengagement in online learning. *Computers & Education*, 188, p.104561.
25. Verma, N., Getenet, S., Dann, C. and Shaik, T., 2023. Designing an artificial intelligence tool to understand student engagement based on teacher's behaviors and movements in video conferencing. *Computers and Education: Artificial Intelligence*, 5, p.100187.

26. Abdulkader, R., Ayasrah, F.T.M., Nallagattla, V.R.G., Hiran, K.K., Dadheech, P., Balasubramaniam, V. and Sengan, S., 2023. Optimizing student engagement in edge-based online learning with advanced analytics. *Array*, 19, p.100301.
27. Khenkar, S.G., Jarraya, S.K., Allinjaw, A., Alkhuraiji, S., Abuzinadah, N. and Kateb, F.A., 2023. Deep analysis of student body activities to detect engagement state in E-learning sessions. *Applied Sciences*, 13(4), p.2591.