Computer Vision-based AI Models for Tracking Learners' Facial Expressions and Gaze Behavior in Online Education

Pichaya Promla[1]* Dr.Somkid Saelee**

(Received: *December 31, 2024*; Revised: *March 25, 2025*; Accepted: *March 27, 2025*)

ABSTRACT

Online learning faces a significant challenge: the high dropout rate caused by limitations in observing learners' behaviors. As a result, instructors often lack sufficient data to adapt teaching methods and effectively motivate learners. This research aimed to 1) develop computer vision-based AI models for analyzing online learners' facial expressions, 2) to develop computer vision-based AI models for analyzing online learners' gaze behavior, and 3) to study the potential application of the developed models in the context of online learning environments. To address the diversity of devices and computational constraints in online learning environments, a transfer learning approach was adopted to train the models. Four pre-trained models including Yolov8n, Yolov9t, Yolov10n, and Yolov11n were selected cause by their optimization for low computational resource usage, and their effectiveness was evaluated and compared. The results revealed that the Learner Emotion Detection model (LeEmo) detects academic emotions from learners' facial expressions, performed best with Yolov9t, achieving a mAP of 78.10% and an F1-Score of 76.05%. The Lerner Eye Tracking model (LeET), developed for learners' eye-tracking tasks, achieved its best performance with the Yolov11n, achieving a mAP of 94.51%, an F1-Score of 85.23%, and a performance of 93.45% for monitoring blink and closed eye activities. Lastly, the Learner Drowsiness Detection model (LeDro), which detects activeness or drowsiness from learners' facial expressions, also performs best with the Yolov11n, achieving a mAP of 85.96% and an F1-Score of 82.46%. These findings demonstrate the significant potential of computer vision-based models for detecting and monitoring online learners' behaviors, providing valuable data for instructors to enhance online learning outcomes. Furthermore, these data could be analyzed using other artificial intelligence technologies to explore various learning states of online learners, such as sustained attention levels, flow states, engagement levels, and stress levels.

Keywords: Computer vision, Gaze behavior, Facial expression

[1]*Corresponding author: pichaya.promla@gmail.com*

**Student of Doctor of Philosophy Program in Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Thailand.*

***Lecturer, Department of Computer Education, Faculty of Technical Education, King Mongkut's University of Technology North Bangkok, Thailand.*

Introduction

Online learning is an educational innovation that integrates the internet, smart devices, and software to support blended learning, enabling remote interaction between learners and instructors. It is grounded in principles such as learner-instructor interaction, teamwork, self-directed learning, timely feedback, and flexibility in time and location [1]. These features support individual differences and promote lifelong learning.

Online learning, whether asynchronous or synchronous, offers flexibility but faces challenges most notably high dropout rates due to the greater concentration, responsibility, and self-discipline required from learners compared to traditional settings [2–3]. Sustained attention is critical for engagement and deep learning [4–5] typically lasts only 10–30 minutes in conventional classrooms [6]. Limited real-time visibility in online environments makes it difficult for instructors to observe learner behavior and adjust teaching, accordingly, resulting in a lack of data to inform instructional decisions.

To address this challenge, researchers propose using artificial intelligence (AI) to analyze learners' behaviors and provide valuable insights for instructors. Recent studies have identified several learning states associated with learner behavior, including: **Engagement**, which reflects active involvement and motivation and enhances learning outcomes [7–20]; **Concentration and Flow**, immersive states where focus is intense and tasks feel effortless [21–22]; and **Attention**, the ability to suppress distractions and focus, though often limited by fatigue or task complexity [23–27]. Other relevant states include **Perception**, involving learners' self-assessment and attitudes toward tasks [28–32]; **Confusion**, a state of uncertainty that disrupts progress and increases dropout risk [33]; Mind Wandering, where thoughts drift from the task and reduce attentiveness [7,34–36]; **Boredom**, caused by monotonous or irrelevant tasks that lead to disengagement [21]; and **Anxiety**, which arises when challenges exceed skill levels, resulting in stress and reduced engagement [21]. Learning states can be extracted using two primary methods: sensor-based and computer vision-based approaches. **Sensor-based methods** analyze physiological signals such as heart rate, skin temperature, galvanic skin resistance (GSR), and brain activity [34–40]. While effective, they require specialized equipment, increasing both cost and implementation complexity. In contrast, **computer vision-based methods** analyze visual cues such as facial expressions, emotions, head and face posture, gaze behavior, and upper-body movement using standard webcams [21, 26, 41–51]. This approach is low-cost, user-friendly, and more practical for widespread use.

The literature review revealed that using computer vision to track learners' behaviors in online learning environments has significant potential to provide valuable data for instructors. This data can be utilized to adapt teaching methods, enhance learner motivation, and improve the effectiveness of online learning management, ultimately reducing dropout rates.

Objectives of the study

1. To develop computer vision-based AI models for analyzing online learners' facial expressions.
2. To develop computer vision-based AI models for analyzing online learners' gaze behavior.
3. To study the potential application of the developed models in the context of online learning environments.

Methodology

To address Objectives 1 and 2, three computer vision-based AI models were developed, including:

Learner Emotion Detection model (LeEmo): Designed to detect academic emotions from learners' facial expressions in online learning. It was trained using a supervised learning approach to classify seven emotions that influence learning: neutral, happy, surprise, angry, sad, fear , and disgust [52].

Lerner Eye Tracking model (LeET): Comprises two components: *1) Monitoring blink and closed-eye activities*, which use the Eye Aspect ratio (EAR) value to track blinked rate per minute and closed-eyes activity [53]. *2) Detecting whether the eyes are on-screen or off-screen*, by classifying learners' eyes direction and head pose into two classes: on-screen or off-screen. This component was trained using a supervised learning approach.

Learner Drowsiness Detection model (LeDro): Detects whether the learners' facial expressions indicate activeness or drowsiness. It was trained using a supervised learning approach.

Given the diversity of devices and computational constraints in online learning environments, transfer learning was used to train the models. Four pre-trained models including YOLOv8n, YOLOv9n, YOLOv10n, and YOLOv11n, optimized for low computational resource usage, were selected to evaluate and compare their effectiveness. Each model followed the development steps:

1. Data Collection and Preparation Three public datasets were used to train the proposed models.

The LeEmo model was trained using the EMOTIC dataset [54], originally labeled with 26 emotion categories. Emotions were re-categorized into 7 types (neutral, happy, surprise, angry, sad, fear, and disgust), with 3,100 images per category, totaling 21,700 images. The data were split into training (70%), validation (20%), and test (10%) sets.

The LeET model used the Columbia Gaze Dataset [55], which contains upper-body images of 56 individuals with various gaze directions and head poses. A total of 10,000 images (5,000 per class: on-screen/off-screen) were selected and divided into training (70%), validation (20%), and test (10%) sets.

The LeDro model was trained on the UTA-RLDD dataset [56], which includes 180 video clips from 60 participants. From these, 9,120 images were extracted and labeled as active or drowsy based on clip type proportions. Each class contained 4,560 images, split into training (70%), validation (20%), and test (10%) sets.

2. Training the Model. The four pre-trained models were used to train models using the same hyperparameter to evaluate their effectiveness and compare performance.

The LeEmo model was trained for 300 epochs, a learning rate of 0.01, Automatic Mixed Precision (AMP) enabled, and a batch size of 16 images. The input image resolution was set to 320 pixels.

The LeEt model consisted of two components. For the blink and closed-eye monitoring, the Eye Aspect Ratio (EAR) was calculated using the formula:

$$EAR = 2 \times \frac{|p2-p6|+|p1-p5|}{|p1-p4|}$$

where p1 to p6 = distances between specific eye coordinate points.

Euclidean distance was used to calculate these distances, with coordinate points extracted using MediaPipe's face mesh, which represents the face with 486 coordinate points. The EAR value ranged from 0 to 1, where values approaching 0 indicates closed eyes, and values approaching 1 indicates wide-open eyes. The average EAR from both eyes was calculated and compared to a default blink threshold of 0.2, which was automatically adjusted for individual learners. If the average EAR was below or equal to the threshold, the model identified closed eyes. The duration of closure was then evaluated: if the duration was less than or equal to 0.5 seconds, it was classified as a blink, while a duration exceeding 3 seconds, it indicated prolonged eyes closed.

For detecting whether the learner's eyes are on-screen or off-screen. The model was trained using transfer learning approach from four pre-trained models. It was trained for 500 epoch with a learning rate of 0.01, AMP enabled, and a batch size of 16 images. The input image resolution was set to 640 pixels.

The last model, LeDro, was trained for 500 epochs with a learning rate of 0.01, AMP enabled, and a batch size of 16 images. The input image resolution was set to 640 pixels.

3. Model evaluation. After training, the effectiveness of each model was evaluated using a confusion matrix to calculate key performance metrics: Mean Average Precision (mAP), Precision, Recall, and F1-Score. The mAP reflects the model's overall ability to accurately detect objects across all classes. Precision indicates the proportion of correct positive predictions out of all positive predictions, while Recall measures the proportion of actual positives that were correctly identified. The F1-Score, as the harmonic means of Precision and Recall, provides a balanced assessment of model performance. The results were then compared to determine which model performed best for each specific task.

Results

1. Result of LeEmo Model Training: The performance comparison of the models, in terms of mAP, Precision, Recall, and F1-score is presented in Table 1.

Table 1 Result of LeEmo Model Training.

pre-trained models	mAP@0.5	Precision	Recall	F1-Score
Yolov8n	67.59%	69.51%	73.69%	71.57%
Yolov9t	78.10%	74.23%	77.95%	76.05%
Yolov10n	75.71%	71.60%	73.76%	72.66%
Yolov11n	77.81%	70.34%	77.03%	73.53%

Based on Table 1, the YOLOv9t achieved the best performance, with a mAP of 78.10% and an F1-Score of 76.05%. The YOLOv11n followed closely with a mAP of 77.81% and an F1-Score of 73.53%. The YOLOv10n recorded a mAP of 75.71% and an F1-Score of 72.66%. Lastly, the YOLOv8n demonstrated the lowest performance, achieving a mAP of 67.59% and an F1-Score of 71.57%.

2. Result of LeET Model Training: For blink and closed-eye monitoring. The performance evaluation was conducted using 5 randomly selected video clips sourced from the internet, each with a duration of 3 minutes. Each clip featured a single individual speaking in various contexts, with diverse eye properties, such as wearing glasses. The LeEt model was used to monitor blink and closed eye activities and was compared with manual monitoring. The results are shown in Table 2.

Table 2 Result of LeET Model Training for Monitoring Blink and Closed-Eye Activities.

Clip Video	Blinking Times $_{\text{LeET}}$	Blinking Times $_{\text{Manual}}$	Closed Times $_{\text{LeET}}$	Closed Times $_{\text{Manual}}$
Clip 01	41	47	3	3
Clip 02	27	29	0	0
Clip 03	33	31	0	2
Clip 04	14	14	1	2
Clip 05	32	34	6	6
Overall	147	155	10	13

From Table 2, the LeEt model detected a total of 147 blinks and 10 closed eye times, compared to 155 and 13 detected manually. The model's performance was calculated using the formular:

$$\text{performance} = \frac{100}{\text{Blinking Times}_{\text{Manual}} + \text{Closed Times}_{\text{Manual}}} \times (\text{Blinking Times}_{\text{LeET}} + \text{Closed Times}_{\text{LeET}}) \quad \ldots\%$$

$$\text{perfomance} = \frac{100}{155+13} \times (147+10) \qquad = 93.45\ \%$$

With a performance of 93.45%, this component of the LeET model demonstrates high accuracy and reliable results in monitoring blink and closed eye activities.

For detecting whether the learner's eyes are on-screen or off-screen. The performance comparison of the models, in terms of mAP, Precision, Recall, and F1-score, is presented in Table 3.

Table 3 Result of LeET Model Training for On-Screen and Off-Screen Detection

pre-trained models	mAP@0.5	Precision	Recall	F1-Score
Yolov8n	92.72%	81.40%	86.62%	83.94%
Yolov9t	93.80%	84.70%	85.71%	85.21%
Yolov10n	91.80%	79.20%	89.32%	83.95%
Yolov11n	94.51%	80.84%	90.14%	85.24%

From Table 3, the YOLOv11n model achieved the best performance, with a mAP of 94.51% and an F1-Score of 85.24%. The YOLOv9t model closely followed with a mAP of 93.80% and an F1-Score of 85.21%. The YOLOv8n model recorded a mAP of 92.72% and an F1-Score of 83.94%. Lastly, the YOLOv10n model demonstrated the lowest performance, achieving a mAP of 91.80% and an F1-Score of 83.95%.

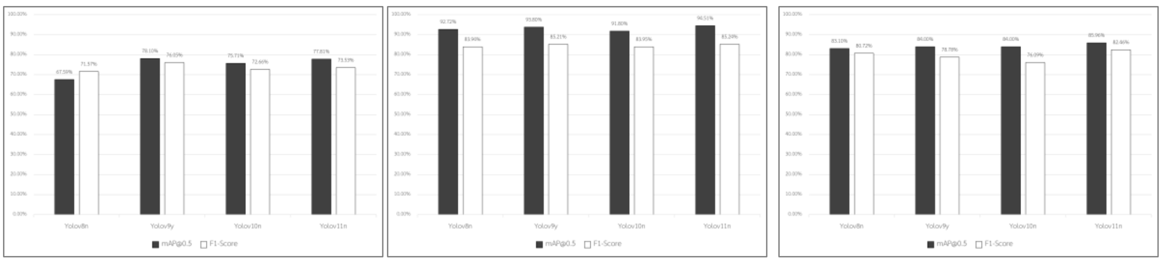
3. Result of LeDro Model Training: The performance comparison of the models presented in Table 4.

Table 4 Result of LeDro Model Training.

pre-trained models	mAP@0.5	Precision	Recall	F1-Score
Yolov8n	83.10%	78.50%	83.07%	80.72%
Yolov9t	84.00%	78.05%	79.51%	78.78%
Yolov10n	84.00%	78.82%	79.34%	79.08%
Yolov11n	85.96 %	78.74%	86.55%	82.46%

Based on Table 4, the YOLOv11n model achieved the best performance, with a mAP of 85.96% and an F1-Score of 82.46%. The YOLOv10n model closely followed, achieving a mAP of 84.00% and an F1-Score of 79.08%. The YOLOv9t model also achieved a mAP of 84.00% but recorded a lower F1-Score of 78.78.09%. Lastly, the YOLOv8n model demonstrated the lowest performance, with a mAP of 83.10% and an F1-Score of 80.72%.

The performance comparison results for all models are presented in Figure 2.



(A) LeEmo Model (B) LeET for On-Screen/Off-Screen Detection (C) LeDro Model

Figure 1 The performance comparison results for all models

Discussion and Conclusions

Based on the objectives of this study, we developed three computer vision-based models for detecting and monitoring online learner behavior including the LeEmo model for detecting learners' emotions from facial expressions, the LeET model for monitoring blink and closed eye activities and detecting whether eyes are on-screen or off-screen, and the LeDro model for detecting activeness or drowsiness based on learners' facial expressions.

All models were trained using a transfer learning approach with pre-trained models to accommodate diverse devices and computational constraints. Four pre-trained models: Yolov8n, Yolov9t, Yolov10n, and Yolov11n were selected to train each model and compare their performance. The LeEmo model achieved its best performance with the Yolov9t, which had a mAP of 78.10% and an F1-Score of 76.05%. The LeET model achieved its best performance with the Yolov11n, with a mAP of 94.51%, an F1-Score of 85.23%, and a performance of 93.45% for monitoring blink and closed eye activities. Lastly, the LeDro model also achieved its best performance with the Yolo11n, which had a mAP of 85.96% and an F1-Score of 82.46%.

These findings align with the study of Shaikh et al. [57], which demonstrated that newer YOLO versions, such as YOLOv5, achieved higher accuracy in detecting faces and emotions. Similarly, Zheng et al. [58] demonstrated that computer vision-based model for monitoring driver fatigue showed high accuracy and robust real-time performance. Therefore, this study demonstrates the potential of computer vision-based models for tracking and analyzing online learners' behaviors, providing valuable insights for instructors to enhance teaching strategies and improve online learning outcomes. Furthermore, in other studies, the behavioral data obtained in this research could be analyzed using other artificial intelligence technologies to explore various learning states of online learners, such as sustained attention levels, flow states, engagement levels, and stress levels.

Beyond its technical contributions, this study supports dropout reduction in online learning through two key approaches: instructional strategies, by using AI-generated data to adjust teaching methods; and learner behavior monitoring, through real-time attention tracking using computer vision. Together, these elements help create a more adaptive and engaging learning environment.

Limitations

This study demonstrates the potential of computer vision-based AI models for monitoring learners' behaviors in online environments; however, several limitations remain.

1. The models were trained on publicly available datasets, which may not reflect the diversity of real-world learners, potentially limiting generalizability across different contexts.

2. This approach relies solely on visual data, which may not capture deeper cognitive or emotional states. Multimodal inputs such as audio, text, or physiological signals could enhance interpretation.

3. Although YOLO-based models were chosen for efficiency, real-time classroom deployment was not tested. Further evaluation under real-world conditions is needed.

4. Ethical and privacy concerns regarding facial data collection were beyond the scope of this study and should be addressed in future work to ensure responsible use of AI in education.

References

1. Wichuda R. Web-based: A new alternative to Thai educational technology. J Educ Stud. 1999;27(3):29–35. Thai.
2. Guo YR, Goh DHL, Luyt B, Sin SCJ, Ang RP. The effectiveness and acceptance of an affective information literacy tutorial. Comput Educ. 2015;87:368–84.
3. Rothkrantz L. An affective distant learning model using avatars as user stand-in. Proc 18th Int Conf Comput Syst Technol [Internet]. New York, USA: ACM; 2017. p. 288–95.
4. Immordino-Yang MH, Damasio A. We feel, therefore we learn: The relevance of affective and social neuroscience to education. Mind Brain Educ. 2007;1(1):3–10.
5. Marshall L, Rowland F. A guide to learning independently. 3rd ed. Open University Press; 1998.
6. Young MS, Robinson S, Alberts P. Students pay attention! Combating the vigilance decrement to improve learning during lectures. Act Learn High Educ. 2009;10(1):41–55.
7. Ma S, Zhou T, Nie F, Ma X. Glancee: An adaptable system for instructors to grasp student learning status in synchronous online classes. Proc CHI Conf Hum Factors Comput Syst [Internet]. New York, USA: ACM; 2022 [cited 2022 May 1]. p. 1–25.
8. Wang HY, Sun JCY. Exploring behavioral patterns of online synchronous VR co-creation: An analysis of student engagement. Proc 2021 Int Conf Adv Learn Technol (ICALT). 2021. p. 392–4.
9. Kodagoda N, Gamage A, Suriyawansa K, Jayasinghe B, Rupasinghe S, Ganegoda D, et al. Innovative use of collaborative teaching in conducting a large scale online synchronous fresher's programming course. Proc 2021 IEEE Glob Eng Educ Conf (EDUCON). 2021. p. 891–6.
10. Pozdniakov S, Martinez-Maldonado R, Singh S, Chen P, Richardson D, Bartindale T, et al. Question-driven learning analytics: Designing a teacher dashboard for online breakout rooms. Proc 2021 Int Conf Adv Learn Technol (ICALT). 2021. p. 176–8.
11. Altuwairqi K, Jarraya SK, Allinjawi A, Hammami M. A new emotion–based affective model to detect student's engagement. J King Saud Univ - Comput Inf Sci [Internet]. 2018; Available from: http://www.sciencedirect.com/science/article/pii/S1319157818309224
12. Shi Y, Tong M, Long T. Investigating relationships among blended synchronous learning environments, students' motivation, and cognitive engagement: A mixed methods study. Comput Educ. 2021;168:104193.
13. Raes A, Vanneste P, Pieters M, Windey I, Van Den Noortgate W, Depaepe F. Learning and instruction in the hybrid virtual classroom: An investigation of students' engagement and the effect of quizzes. Comput Educ. 2020;143:103682.
14. Almpanis T, Joseph-Richard P. Lecturing from home: Exploring academics' experiences of remote teaching during a pandemic. Int J Educ Res Open. 2022;3:100133.

15. Schwenck CM, Pryor JD. Student perspectives on camera usage to engage and connect in foundational education classes: It’s time to turn your cameras on. Int J Educ Res Open. 2021;2:100079.
16. Capone R, Lepore M. From distance learning to integrated digital learning: A fuzzy cognitive analysis focused on engagement, motivation, and participation during COVID-19 pandemic. Technol Knowl Learn. 2021;1259–89.
17. Hjalmarson MA. Learning to teach mathematics specialists in a synchronous online course: a self-study. J Math Teach Educ. 2017;20(3):281–301.
18. Alamer A, Alharbi F. Synchronous distance teaching of radiology clerkship promotes medical students’ learning and engagement. Insights Imaging. 2021;12(1):41.
19. Händel M, Bedenlier S, Kopp B, Gläser-Zikuda M, Kammerl R, Ziegler A. The webcam and student engagement in synchronous online learning: visually or verbally? Educ Inf Technol [Internet]. 2022 Apr 18 [cited 2022 May 1]. Available from: https://doi.org/10.1007/s10639-022-11050-3
20. Olt PA. Virtually there: Distant freshmen blended in classes through synchronous online education. Innov High Educ. 2018;43(5):381–95.
21. Sun W, Li Y, Tian F, Fan X, Wang H. How presenters perceive and react to audience flow prediction in-situ: An explorative study of live online lectures. Proc ACM Hum-Comput Interact. 2019;3(CSCW):162:1–19.
22. Meriem B, Benlahmar H, Naji M, Filali S, Kaiss W. Determine the level of concentration of students in real time from their facial expressions. Int J Adv Comput Sci Appl. 2022;13:159–66.
23. Li J, Ngai G, Va Leong H, Chan S. Multimodal human attention detection for reading. Proc 31st Annu ACM Symp Appl Comput [Internet]. New York, USA: ACM; 2016. p. 187–92.
24. Bailey D, Almusharraf N, Hatcher R. Finding satisfaction: Intrinsic motivation for synchronous and asynchronous communication in the online language learning context. Educ Inf Technol. 2021;26(3):2563–83.
25. Rabinovich T, Berthon P, Fedorenko I. Reducing the distance: Financial services education in web-extended learning environments. J Financ Serv Mark. 2017;22(3):126–31.
26. Raca M, Dillenbourg P. Translating head motion into attention - Towards processing of student’s body-language. Proc 8th Int Conf Educ Data Mining. Madrid, Spain; 2015. p. 320–6.
27. Allison N. Students’ attention in class: Patterns, perceptions of cause and a tool for measuring classroom quality of life. J Perspect Appl Acad Pract. 2020;8:58.
28. Sharifrazi F, Stone S. Students perception of learning online: Professor’s presence in synchronous versus asynchronous modality. Proc 5th Int Conf Comput Technol Appl [Internet]. New York, USA: ACM; 2019. p. 180–3.

29. Xu M, Zeng S. Chinese EFL learners' perception of synchronous-computer-mediated communication in conducting online interactive tasks. Proc 2019 14th Int Conf Comput Sci Educ (ICCSE). 2019. p. 987–91.

30. McCarthy M, Kusaila M, Grasso L. Intermediate accounting and auditing: Does course delivery mode impact student performance? J Account Educ. 2019;46:26–42.

31. Mendoza AV, Díaz KP, Raffo FS. Perceptions of university teachers and students on the use of Blackboard Collaborate as a teaching tool during virtual learning due to the COVID-19 pandemic. Proc 2021 IEEE 1st Int Conf Adv Learn Technol Educ Res (ICALTER). 2021. p. 1–4.

32. Schwenck CM, Pryor JD. Student perspectives on camera usage to engage and connect in foundational education classes: It's time to turn your cameras on. Int J Educ Res Open. 2021;2:100079.

33. Atapattu T, Falkner K, Thilakaratne M, Sivaneasharajah L, Jayashanka R. What do linguistic expressions tell us about learners' confusion? A domain-independent analysis in MOOCs. IEEE Trans Learn Technol. 2020;13(4):878–88.

34. Pham P, Wang J. AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking. In: Conati C, Heffernan N, Mitrovic A, Verdejo MF, editors. Artif Intell Educ. Cham: Springer Int Publ; 2015. p. 367–76.

35. Xiao X, Wang J. Towards attentive, bi-directional MOOC learning on mobile devices. Proc 2015 ACM Int Conf Multimodal Interact [Internet]. New York, USA: ACM; 2015. p. 163–70.

36. Richardson DC, Griffin NK, Zaki L, Stephenson A, Yan J, Curry T, et al. Engagement in video and audio narratives: Contrasting self-report and physiological measures. Sci Rep. 2020;10(1):11298.

37. Al Machot F, Ali M, Ranasinghe S, Mosa AH, Kyandoghere K. Improving subject-independent human emotion recognition using electrodermal activity sensors for active and assisted living. Proc 11th PErvasive Technol Relat Assist Environ Conf. New York, USA: ACM; 2018. p. 222–8.

38. Di Lascio E, Gashi S, Santini S. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018 Sep 18;2(3):103:1–21.

39. Hassib M, Schneegass S, Eiglsperger P, Henze N, Schmidt A, Alt F. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. Proc 2017 CHI Conf Hum Factors Comput Syst [Internet]. New York, USA: ACM; 2017. p. 5114–9.

40. Kosmyna N, Maes P. AttentivU: An EEG-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning. Sensors. 2019;19(23):5200.

41. Cha S, Kim W. Concentration analysis by detecting face features of learners. Proc 2015 IEEE Pacific Rim Conf Commun Comput Signal Process (PACRIM). 2015. p. 46–51.

42. Dewan MAA, Lin F, Wen D, Murshed M, Uddin Z. A deep learning approach to detecting engagement of online learners. Proc 2018 IEEE SmartWorld, Ubiquitous Intell Comput, Adv Trusted Comput, Scalable Comput Commun, Cloud Big Data Comput, Internet People Smart City Innov (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2018. p. 1895–902.
43. D'Mello S, Olney A, Williams C, Hays P. Gaze tutor: A gaze-reactive intelligent tutoring system. Int J Hum-Comput Stud. 2012;70(5):377–98.
44. Hutt S, Mills C, Bosch N, Krasich K, Brockmole J, D'Mello S. "Out of the Fr-Eye-ing Pan": Towards gaze-based models of attention during learning with technology in the classroom. In: Proc 25th Conf User Model Adapt Pers [Internet]. New York, USA: ACM; 2017. p. 94–103.
45. Zeng H, Shu X, Wang Y, Wang Y, Zhang L, Pong T, et al. EmotionCues: Emotion-oriented visual summarization of classroom videos. IEEE Trans Vis Comput Graph. 2019;1–1.
46. Murali P, Hernandez J, McDuff D, Rowan K, Suh J, Czerwinski M. AffectiveSpotlight: Facilitating the communication of affective responses from audience members during online presentations. Proc 2021 CHI Conf Hum Factors Comput Syst [Internet]. New York, NY, USA: ACM; 2021. p. 1–13.
47. Ahuja K, Kim D, Xhakaj F, Varga V, Xie A, Zhang S, et al. EduSense: Practical classroom sensing at scale. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2019;3(3):71:1–26.
48. Bednarik R, Eivazi S, Hradis M. Gaze and conversational engagement in multiparty video conversation: An annotation scheme and classification of high and low levels of engagement. Proc 4th Workshop Eye Gaze Intell Hum Mach Interact [Internet]. New York, USA: ACM; 2012. p. 1–6.
49. Yang FY, Chang CY, Chien WR, Chien YT, Tseng YH. Tracking learners' visual attention during a multimedia presentation in a real classroom. Comput Educ. 2013;62:208–20.
50. Ahuja K, Shah D, Pareddy S, Xhakaj F, Ogan A, Agarwal Y, et al. Classroom digital twins with instrumentation-free gaze tracking. Proc 2021 CHI Conf Hum Factors Comput Syst [Internet]. New York, USA: ACM; 2021. p. 1–9.
51. Teevan J, Liebling D, Paradiso A, Suarez C, Veh C, Gehring D. Displaying mobile feedback during a presentation. 2012 Sep 21.
52. Pekrun R, Goetz T, Titz W, Perry RP. Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educ Psychol. 2002;37(2):91–105.
53. Smith BA, Yin Q, Feiner SK, Nayar SK. Gaze locking: Passive eye contact detection for human-object interaction. Proc 26th Annu ACM Symp User Interface Softw Technol [Internet]. New York, USA: ACM; 2013. p. 271–80.
54. Kosti R, Alvarez JM, Recasens A, Lapedriza A. Context based emotion recognition using EMOTIC dataset. IEEE Trans Pattern Anal Mach Intell. 2019;1–1.
55. Columbia Gaze Data Set | CEAL [Internet]. [cited 2024 Dec 20]. Available from: https://ceal.cs.columbia.edu/columbiagaze/#project-publications

56. Ghoddoosian R, Galib M, Athitsos V. A realistic dataset and baseline temporal model for early drowsiness detection [Internet]. [cited 2024 Apr 30]. Available from: http://arxiv.org/abs/1904.07312

57. Shaikh A, Mishra K, Kharade P, Kanojia M. Comprehensive study on emotion detection with facial expression images using YOLO models [Internet]. [cited 2024 Dec 22]. Available from: https://www.semanticscholar.org/paper/Comprehensive-Study-on-Emotion-Detection-with-Using-Shaikh-Mishra/1f17cf816c125e899599ee1fee8913336da9fb8d#citing-papers

58. Zheng Y, Chen S, Wu J, Chen K, Wang T, Peng T. Real-time driver fatigue detection method based on comprehensive facial features. In: Tari Z, Li K, Wu H, editors. Algorithms and architectures for parallel processing. Singapore: Springer Nature; 2024. p. 484–501.