MATHEMATICAL MODELING AND COMPUTING, Vol. 13, No. 1, pp. 41–51 (2026)

Face-Based Engagement Detection Methods: A Review

Qarbal I.*, Sael N., Ouahabi S.

Faculty of Sciences, Hassan II University of Casablanca, Casablanca, Morocco
**Corresponding author: ikram.qarbal-etu@etu.univh2c.ma*

(Received 28 March 2025; Revised 7 November 2025; Accepted 5 December 2025)

The detection of student engagement in online learning environments has become increasingly important with the widespread adoption of e-learning platforms. This paper reviews current approaches for monitoring student engagement based on facial expressions, gaze tracking, fatigue and drowsiness detection, and multimodal systems. By analyzing facial expressions, systems can detect emotional states such as happiness, frustration, and boredom, offering real-time feedback to instructors. Gaze tracking provides insights into students focus, although challenges such as hardware costs and lighting conditions affect its accuracy. Fatigue and drowsiness detection, through blinking and yawning analysis, helps identify cognitive overload, while multimodal systems that combine facial, behavioral, and physiological data offer a more comprehensive picture of engagement. This review highlights the potential of these methods while addressing the need for more robust, scalable, and privacy-conscious systems for real-time engagement monitoring in di-verse e-learning contexts.

Keywords: *facial expressions; gaze tracking; drowsiness; fatigue detection; multi-modal approaches.*

2020 MSC: 68U10, 68T45, 97C80 **DOI:** 10.23939/mmc2026.01.041

1. Introduction

In recent years, the integration of technology in education has transformed the traditional learning environment, making student engagement a critical factor in both online and blended learning settings. Student engagement, broadly defined as the level of attention, curiosity, and involvement a learner demonstrates during a learning activity, is highly correlated with learning outcomes. When students are engaged, they are more likely to absorb and retain information, participate actively in discussions, and develop a deeper understanding of the material. Conversely, disengaged students tend to exhibit lower academic performance, higher dropout rates [1], and reduced overall satisfaction with the learning process.

Detecting student engagement, however, remains a challenge, especially in remote or large-scale learning environments where teachers cannot directly observe each student's behavior. To address this issue, a number of studies have explored engagement detection based on a variety of behavioral and physiological cues, such as body posture, gestures, and overall physical movements. Posture analysis, for instance, can provide insights into a student's level of attentiveness by monitoring how they sit, move, or lean, with certain patterns indicating distraction or cognitive overload. While these approaches have proven effective in traditional classroom settings, they are less applicable in remote learning environments where the full body of the student is often not visible.

The COVID-19 pandemic has significantly accelerated the adoption of distance learning [2], making face-based engagement detection systems more relevant and accessible than ever. In e-learning environments, where students primarily interact with their devices through webcams, the face becomes the most visible and reliable source of engagement data. Consequently, facial feature-based

https://orcid.org/0009-0009-8609-694X (Qarbal I.), https://orcid.org/0000-0002-8134-3886 (Sael N.), https://orcid.org/0000-0001-6478-7218 (Ouahabi S.)

engagement detection, which leverages the analysis of facial expressions, gaze direction, and signs of fatigue, has gained prominence. These methods offer a non-invasive, real-time solution that can be seamlessly integrated into virtual learning platforms, allowing educators to monitor and respond to student engagement more effectively in remote settings.

Facial feature-based engagement detection [3] focuses on visual indicators that correlate with cognitive and emotional states. Key facial cues include expressions that convey emotions like frustration or boredom, gaze direction that indicates focus or distraction, and signs of fatigue such as drooping eyelids or frequent blinking or yawning. By identifying these cues, engagement detection systems can provide valuable insights into student behavior, allowing interventions to reengage learners when necessary. This paper addresses three key research questions related to face-based student engagement detection: (Q1) What are the main taxonomies of face-based engagement detection? (Q2) What datasets are commonly used in face-based engagement detection? (Q3) What techniques and models are used for face-based engagement detection?

This review covers current methods for facial feature-based student engagement detection, focusing on facial expressions, gaze-related features, and fatigue detection. It also explores multimodal approaches that combine facial features with other data sources.

2. Taxonomy of face-based engagement detection methods

Engagement detection in educational settings is a complex task that varies depending on the learning environment [4], ranging from traditional classrooms to smart and distance learning setups. In traditional classrooms, engagement is typically assessed through teacher observations, aided by tools like self-reporting and observation checklists to gain a more structured understanding of student focus and participation. These approaches, while effective in small-group or face-to-face contexts, are often limited by subjectivity, teacher bias, and the inability to scale across large or diverse classrooms.

In contrast, smart classrooms utilize advanced technology, including camera-based systems and sensors, to capture data on full-body posture, gestures, and the environmental context, which can be analyzed to understand engagement in large-scale or technology-enhanced settings. These environments benefit from real-time analytics and automated engagement feedback systems, allowing for more dynamic and personalized learning experiences. The integration of artificial intelligence (AI) and Internet of Things (IoT) technologies further enhances the granularity and responsiveness of engagement monitoring in these environments.

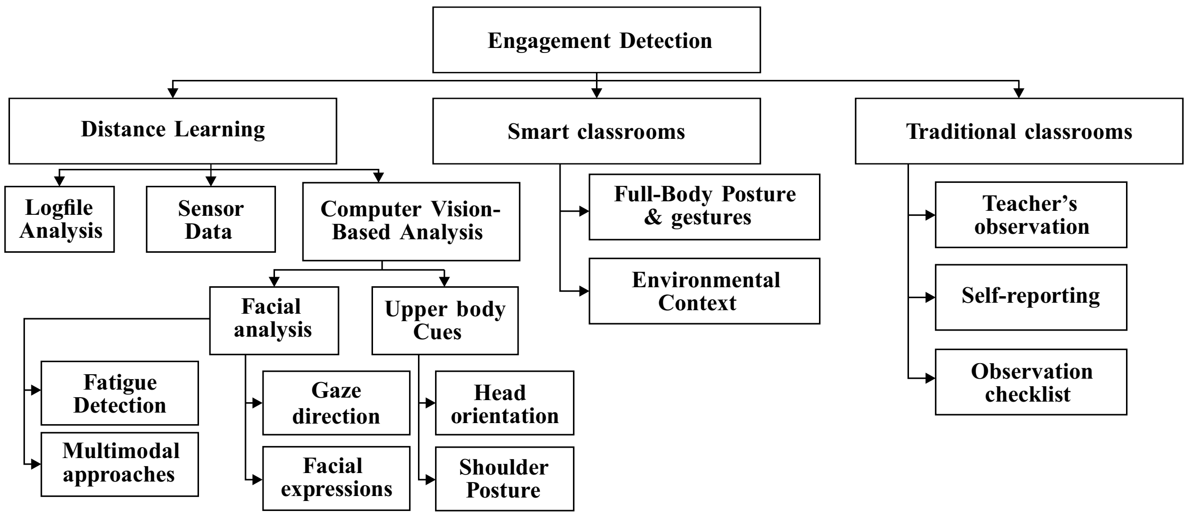

Figure 1. The proposed approach for student engagement detection in real time.

For distance learning environments, the assessment of engagement relies heavily on technologies that can analyze the limited view offered by a student's webcam. Methods such as logfile analysis (tracking platform interactions) and sensor data analysis [5] (using external wearable devices) provide some insight, but the primary approach is computer vision-based analysis. This involves evaluating facial features and upper body cues to infer engagement. Facial analysis includes methods like facial expression recognition, gaze tracking, fatigue detection, and multimodal approaches that combine various facial cues for more accurate assessment. These techniques help address the growing challenge of detecting disengagement in asynchronous and self-paced learning contexts, where traditional observational cues are unavailable.

Upper body cues, such as head orientation and shoulder posture, also play a significant role in understanding student attentiveness. Combining these with facial cues can enhance detection precision, especially in long learning sessions where cognitive fatigue can set in. Additionally, the use of engagement prediction models trained on domain-specific data is becoming more common, improving the system's ability to adapt to different learning behaviors and cultural expressions.

In this study, we focus specifically on facial region-based analysis, examining how facial expressions, gaze, and signs of fatigue can be effectively used to assess student engagement in distance learning environments. By leveraging facial features alone, we aim to provide insights into practical, non-intrusive methods for real-time engagement detection. Figure 1 illustrates the various methods used to assess student engagement across different learning environments, from traditional teacher observation to advanced computer vision-based analysis in distance learning and smart classrooms.

2.1. Facial expressions and emotional states

Facial expressions serve as a crucial window into a student's emotional and cognitive state, offering valuable insights into their level of engagement. These expressions, whether voluntary or involuntary, convey a broad spectrum of emotional cues that directly influence learning experiences. Engagement detection through facial expressions has been widely studied, with various approaches focusing on emotions such as happiness, sadness, anger, and surprise to assess student engagement. Bhardwaj et al. [6] applied Haar Cascade classifiers and Convolutional Neural Networks (CNNs), attaining 93.6% accuracy, where happiness and neutral expressions correlated with higher engagement. Similarly, Bouhlal et al. [7] used the VGG16 model to classify emotions linked to concentration levels, achieving 86% accuracy, with happiness indicating high concentration and sadness signifying lower engagement.

Among the various approaches to engagement detection, the analysis of facial expressions can be broadly categorized into macro-expressions, micro-expressions, and comprehensive emotional state analysis.

Macro-expressions. Macro-expressions are overt and deliberate facial expressions that typically last longer than a second, making them easily recognizable. These expressions reflect emotions such as happiness, sadness, anger, and frustration, which are frequently studied in educational contexts. Happiness and curiosity are often associated with heightened motivation and active participation, whereas frustration and boredom can indicate cognitive overload or disengagement. The analysis of macro-expressions has been significantly enhanced by deep learning techniques, particularly CNNs and pre-trained models such as VGG-16 and ResNet-50. Gupta et al. [8] explored an ensemble of VGG19 and ResNet50 models for detecting cognitive states, achieving 93.11% accuracy. Similarly, Lasri et al. [9] employed CNNs trained on the FER-2013 dataset to classify student emotions, reaching 70% accuracy. These studies highlight the potential of macro-expression analysis in monitoring engagement, though achieving high generalizability remains a challenge.

Micro-expressions. Micro-expressions, in contrast, are brief and involuntary facial expressions that last less than half a second. These subtle emotional indicators often reveal suppressed or concealed feelings that students may not consciously express. A fleeting frown or raised eyebrow may suggest confusion or frustration, while a brief lip press might indicate discomfort or disengagement. Detecting these expressions requires advanced computational techniques such as Spatio-Temporal CNNs (ST-CNNs) and LSTM-based models. Hsia et al. [10] integrated macro- and micro-expression analysis using ResNet-18 and CNN-LSTM, improving engagement detection by analyzing both emotional and behavioral cues.

Recent research has further refined micro-expression recognition for real-time educational settings. Zhao et al. [11] proposed a temporal–spatial correlation and graph attention-guided network (TSG–MER–ELL) to enhance micro-expression recognition in online learning environments. This approach leverages spatio-temporal graph convolution and transformer encoders to improve accuracy in low-resolution streaming conditions, making it particularly valuable for tracking student engagement in virtual classrooms. The model demonstrated state-of-the-art performance across multiple datasets, highlighting the growing role of deep learning-driven micro-expression detection in education.

Comprehensive emotional state analysis. Beyond discrete emotions, recent advancements focus on analyzing complex and mixed emotional states to gain a deeper understanding of student engagement. Emotional experiences often involve multiple dimensions, such as curiosity mixed with frustration during problem-solving. To capture this complexity, valence and arousal analysis map emotions along positive/negative valence and high/low arousal dimensions. Ayvaz et al. [12] applied Support Vector Machines (SVM) and k-Nearest Neighbors (kNN) algorithms for emotion classification in e-learning environments, with SVM achieving 98.24% accuracy, demonstrating the potential of machine learning techniques for engagement detection.

Additionally, attention-based models like Vision Transformers (ViTs) are being explored for their ability to capture subtle emotional expressions in real-time. ViTs have demonstrated exceptional accuracy in image classification by leveraging self-attention mechanisms to analyze complex facial patterns [13]. Recent research successfully integrated ViTs with CNNs and LSTM networks, enhancing engagement detection performance. A ViT-CNN hybrid model achieved 100% classification accuracy in an educational context, showcasing its potential for precise and dynamic engagement tracking.

Facial expressions and emotional states are highly context-dependent [14]. The same expression may convey different meanings depending on the specific learning scenario. For instance, a neutral expression could indicate disengagement in one context but deep concentration in another. Additionally, cultural and demographic differences influence how emotions are both expressed and interpreted, emphasizing the need for adaptive models that account for diverse student populations.

By integrating AI-driven emotion recognition with contextual awareness, educational technologies can personalize learning experiences, ensuring more responsive and effective engagement monitoring in both physical and virtual classrooms.

Overall, while facial expression-based engagement detection systems offer valuable insights into students' emotional states, they work best when integrated with other modalities, such as gaze tracking or fatigue detection, to create a more comprehensive understanding of student engagement.

2.2. Gaze-based detection of engagement

Gaze tracking is a powerful tool for understanding student attention during online learning. By monitoring where students direct their visual attention, gaze-based systems can infer whether students are focused on the learning material or distracted by external stimuli. Several studies have explored how eye movements can be used to assess engagement, with varying degrees of accuracy depending on the technology and methods employed.

Jamil et al. [15] combined eye-tracking and EEG to monitor cognitive states such as attention, demonstrating the effectiveness of these combined methods for assessing engagement. Deng and Gao [16] reviewed eye-tracking research in video-based learning and highlighted key metrics like fixation count and dwell time as critical indicators of student interaction with content. Burch et al. [17] emphasized the potential of eye-tracking to personalize learning but pointed out challenges such as cost, data interpretation, and privacy issues. Lin et al. [18] applied line-of-sight tracking and deep learning models to improve accuracy, particularly in dynamic environments with head movements. Lastly, Haataja et al. [19] explored the role of teacher-student eye contact, finding that mutual gaze fostered greater student engagement during group work.

Gaze-based detection offers significant potential in tracking student engagement, especially with other methods such as facial expressions and head movements. However, the high cost and technical complexity of dedicated eye-tracking systems limit their scalability, suggesting that future research should improve the accuracy of more accessible, webcam-based solutions.

2.3. Fatigue and drowsiness detection

Fatigue and drowsiness detection play a critical role in maintaining student engagement, particularly in long online learning sessions where attention may wane over time. By analyzing facial features such as eye blinking patterns, yawning, and other facial cues, several systems have been developed to monitor signs of fatigue in real time.

Detecting student fatigue involves tracking Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) to identify signs of drowsiness, such as eye closure and yawning. Li et al. [20] developed a real-time fatigue detection system that achieved 94.47% accuracy by monitoring yawning and blinking patterns. Dedhia et al. [21] used Dlib for facial landmark detection to analyze EAR, triggering alerts when drowsiness is detected. Their system also employed interactive tasks to re-engage students. Lahoti et al. [22] leveraged ResNet-50 to track EAR and MAR, achieving 97% training accuracy and 94% validation accuracy. Similarly, Abdulkader et al. [23] proposed an edge-based attentiveness analysis system that used VGG-19, achieving 95.29% accuracy. Safarov et al. [24] extended drowsiness detection with CNN and Haar-Cascade classifiers, achieving 95.8% accuracy by monitoring blinks, yawns, and head movements.

While these systems demonstrate high accuracy in detecting drowsiness and fatigue, the challenge lies in maintaining this accuracy across different hardware setups, lighting conditions, and student behaviors, ensuring they remain reliable and effective across a wide range of educational settings.

2.4. Multimodal detection systems

Multimodal approaches represent a significant innovation in engagement detection by integrating data from various sources such as facial expressions, eye-tracking, and physiological signals. This combination allows for a more comprehensive and accurate assessment of student engagement, capturing nuances that single modality systems might miss. By leveraging multiple data streams, these methods can detect subtle indicators of cognitive load, focus, or drowsiness, which are critical for timely interventions in e-learning environments.

For instance, Li et al. [25] introduced a system that combines blink detection, yawn detection, and head pose estimation for dynamic monitoring of engagement in real time. This approach allows the system to adapt quickly to shifts in student focus, offering a highly responsive engagement assessment. Similarly, Uçar and Özdemir [26] used Local Binary Patterns (LBP) and head pose estimation, achieving a classification accuracy of 72.4%. Though straightforward, this system demonstrates the potential of combining multiple visual cues for accurate, real-time monitoring of engagement levels.

Gupta et al. [8] advanced the field further by using deep learning models like VGG-19 and ResNet-50, integrating facial emotion recognition with blink detection and head movement analysis to reach an impressive 92.58% accuracy. This innovative combination of deep learning with multimodal inputs highlights the strengths of using advanced algorithms to capture complex engagement patterns. Kawamura et al. [28] took a different approach by incorporating physiological data including heart rate variability and seat pressure alongside facial recognition to detect signs of drowsiness. Their system achieved high accuracy with an F1 score of 0.82, demonstrating the potential of multimodal analytics to detect both physical and cognitive engagement states.

These studies illustrate how multimodal systems contribute to more nuanced and adaptive engagement detection in online learning. By integrating diverse data sources, these methods offer richer insights and a promising foundation for future developments in real-time, responsive e-learning environments.

3. Datasets used in face-based engagement detection

Several datasets are widely used for training models in face-based engagement detection, spanning different modalities like facial expression recognition, gaze tracking, fatigue detection, and multimodal systems. Datasets such as FER-2013, CK+, JAFFE, and RAF-DB are frequently utilized for recognizing basic and com-pound emotions that reflect engagement and concentration levels in students. For more subtle emotional cues, CASME II and SAMM focus on micro-expressions, capturing brief facial movements that indicate hidden emotions like frustration. In the domain of gaze tracking, GazeCapture and MPIIGaze provide critical eye-tracking data, assessing where students direct their attention during online sessions. Fatigue detection models benefit from the NTHU Drowsy Driver Detection Dataset, which tracks signs of drowsiness such as eye closures and yawning. Lastly, multimodal engagement systems often rely on datasets like DAiSEE, which integrate facial expressions, gaze data, and behavioral cues to offer a comprehensive view of student engagement. Table 1 presents dataset details.

Table 1. Datasets used in face-based engagement detection.

Modality	**Dataset**	**Size**	**Classes**
Facial Expression	FER-2013	35 887 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral
	CK+	593 sequences	
	RAF-DB	30 000 images	
	CASME II	255 videos	Micro-expressions like happiness, surprise, disgust, fear
	SAMM	159 videos	
	JAFFE	213 images	Anger, Disgust, Fear, Happiness, Sadness, Surprise
Gaze-Based Detection	DAiSEE	9 068 video snippets	Engagement, Boredom, Confusion, Frustration
	GazeCapture	2.5 million frames	Gaze metrics
	MPIIGaze	213 659 images	Gaze direction
Fatigue & Drowsiness Detection	NTHU Drowsy Driver Detection	18 900 video sequences	Awake, Drowsy, Yawning, Blinking

While publicly available datasets such as FER2013, AffectNet, and DAiSEE provide a foundation for engagement detection research, they often lack context-specific engagement labels, leading researchers to develop custom datasets tailored for online learning environments. These datasets aim to bridge the gap between emotion recognition and real-world engagement monitoring by capturing dynamic, long-term engagement fluctuations in virtual classrooms.

Several studies have introduced custom datasets to address these limitations. Watanabe et al. [29] (EnGauge dataset) collected engagement data from online meeting participants, labeling engagement levels as high, middle, or low through a role-acting approach. Their deep learning model, based on MobileNetV2, achieved 89.5% accuracy, demonstrating the feasibility of real-time engagement tracking using webcam-based facial analysis. Similarly, Sassi et al. [30] developed a dataset that maps student emotions to educational states using the Plutchik Wheel of Emotions, integrating self-assessment, tutor evaluation, and AI-based classification to improve engagement monitoring. Their approach highlights the need for more contextual emotion labels in engagement datasets.

Beyond static engagement classification, Abdellaoui et al. [31] proposed a dataset that tracks engagement fluctuations over time. By employing Non-Negative Matrix Factorization (NMF) and Random Walk Models, their study models how students transition between emotional states throughout a learning session, offering predictive insights into disengagement patterns. This shift toward longitudinal engagement tracking emphasizes the importance of temporal analysis rather than relying on single-frame engagement snapshots.

Despite their advantages, custom datasets facing challenges related to generalizability, reproducibility, and privacy restrictions. Many remain institutionally restricted, limiting their use for benchmarking engagement detection methods. However, with the rise of synthetic data generation, self-supervised learning, and federated AI approaches, researchers are finding new ways to overcome data accessibility challenges while preserving privacy.

4. Models used for face-based engagement detection

Face-based engagement detection leverages a range of learning approaches to analyze facial expressions, gaze direction, and physiological signals for accurate, real-time assessment of student engagement. These models can be grouped based on the type of learning approach:

Machine learning techniques. Support Vector Machines (SVM) and k-nearest Neighbors (kNN) are effective machine learning algorithms, particularly for smaller datasets where deep learning may be impractical due to computational complexity. Ayvaz et al. [12] employed SVM and kNN to classify emotions such as happiness and sadness, achieving 98.24% accuracy with SVM. These methods provide an efficient, lower-complexity solution for emotion detection in e-learning contexts.

Deep learning techniques. Convolutional Neural Networks (CNNs) are among the most widely used techniques for analyzing facial expressions due to their ability to effectively classify images. For example, Bhardwaj et al. [6] and Lasri et al. [9] utilized CNNs to classify emotions like happiness and sadness, achieving accuracies of 93.6% and 70%, respectively. Long Short-Term Memory (LSTM) Networks are also used, especially in combination with CNNs, for tasks involving temporal data, such as micro-expression detection. Hsia et al. [10] employed this hybrid approach to identify subtle emotions like frustration and cognitive overload, each of which are critical for assessing student engagement in real time.

Transfer learning models. Transfer learning allows researchers to leverage pre-trained models, significantly reducing the time and computational resources needed to achieve high accuracy in engagement detection. VGG-16, VGG-19, and ResNet-50 are commonly used architectures for transfer learning in emotion analysis. Meriem Bouhlal et al. [7] used VGG-16 to achieve 86% accuracy in concentration level detection, while Gupta et al. [27] combined VGG-19 with ResNet-50, resulting in an accuracy of 93.11% for assessing cognitive states.

Hybrid and multimodal models. Hybrid models combine multiple deep learning techniques, while multimodal systems integrate different data streams such as facial expressions, gaze tracking, and physiological data to improve engagement detection accuracy. Gupta et al. [27] achieved 92.58% accuracy by integrating multiple data streams using an ensemble of VGG-19 and ResNet-50. Kawamura et al. [28] used a CatBoost classifier to combine facial data with physiological signals, achieving F1 scores of 0.82 for drowsiness detection. These multimodal approaches provide richer insights by simultaneously analyzing multiple aspects of student behavior.

Mathematical approaches for fatigue detection. Eye Aspect Ratio (EAR) and Mouth Aspect Ratio (MAR) are mathematical approaches that involve calculating geometric ratios based on facial landmarks to detect fatigue and drowsiness. EAR measures the ratio of vertical to horizontal distances around the eyes to assess eye closure, while MAR measures mouth-related distances to detect yawning. Dedhia et al. [21] and Lahoti et al. [22] achieved up to 94% accuracy in drowsiness detection using these metrics alongside models like ResNet-50. These approaches are computationally efficient and valuable for real-time engagement monitoring during extended learning sessions. Figures 2 and 3 show trends in the usage and performance of different model types for engagement detection.

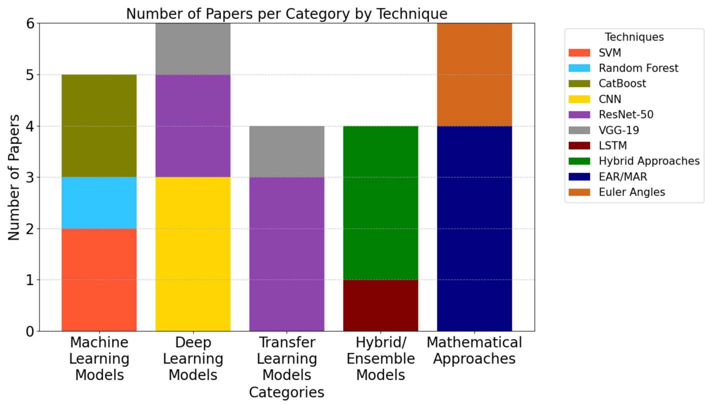

Figure 2. Distribution of Techniques Used across Modeling Approaches in Engagement Detection.

The analysis of modeling techniques shows different performance levels across categories. Machine learning models such as SVM, Random Forest, and CatBoost deliver consistent accuracy, typically between 85–93%. Deep learning models, including CNNs, including pre-trained CNNs like VGG-19 and ResNet-50, vary in performance. CNNs range from 70–95% in accuracy, while pre-trained models offer more stability (85–95%). Hybrid/ensemble models stand out with consistently high accuracy (94–98%), emphasizing the benefit of combining methods for superior results. Mathematical approaches, such as EAR/MAR calculations, often exceed 90% accuracy, especially effective for simpler tasks like fatigue detection. Overall, hybrid and pre-trained deep learning models show the most potential, while traditional machine learning models remain reliable for consistent results.

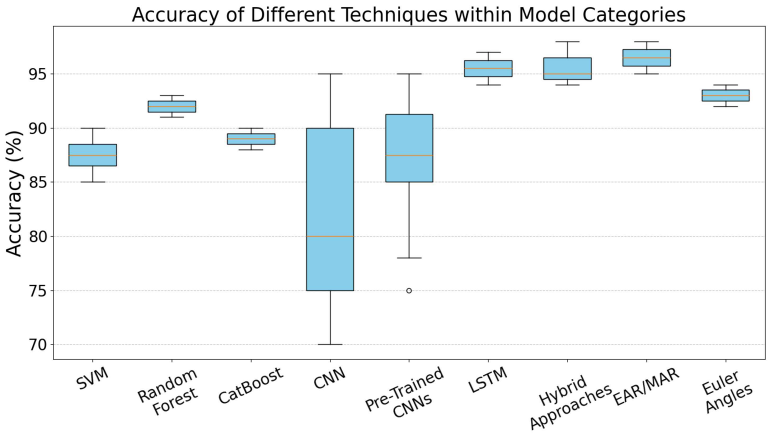

Figure 3. Performance Accuracy of Techniques in Student Engagement and Fatigue Detection.

The study of student engagement detection in online learning highlights a range of promising methods, each offering unique insights and capabilities. Facial expression analysis, gaze tracking, fatigue detection, and multimodal systems each contribute valuable information about student behavior and cognitive state. Together, these methods underscore the potential of engagement monitoring while revealing important limitations to consider as the field advances.

Facial expression analysis is one of the most direct ways to interpret engagement, as emotions like boredom, frustration, and happiness often correlate with a students' level of focus. High accuracies have been achieved in controlled settings with models trained on datasets like FER2013 and CK+, where clear expressions are easier to capture and classify. However, relying solely on facial expressions can present a partial view of engagement. Students may display neutral expressions despite being highly engaged or, conversely, appear attentive while mentally disengaged. Additionally, real-world variability such as differences in lighting or camera quality can reduce accuracy, emphasizing the need for systems that are robust in diverse environments.

Gaze tracking adds another layer of understanding by directly monitoring where a students visual attention is focused, providing real-time information on whether they are following the learning material. Dedicated eye-tracking hardware offers precise data but is costly, making it challenging to scale. Webcam-based gaze tracking, while more accessible, often lacks the precision needed to consistently gauge attention, especially in varied settings where factors like lighting and camera position affect performance. As such, while gaze data is valuable, it often serves best in combination with other indicators to create a richer, more complete picture of engagement.

Monitoring fatigue and drowsiness provides yet another dimension, addressing a key factor that can hinder sustained attention in online learning. Metrics like blinking rate and yawning frequency are useful indicators of cognitive and physical exhaustion, helping identify when a student may need a break to re-engage effectively. Although deep learning models trained on custom datasets show high accuracy in detecting fatigue, their reliance on high-resolution video input and real-time processing can be resource-intensive. Moreover, distinguishing between genuine tiredness and cognitive overload remains a challenge, signaling an area for further refinement.

Multimodal detection systems integrate these various data streams facial expressions, gaze direction, and physiological cues to provide a more holistic view of engagement. By capturing a wider range of behavioral and physiological signals, these systems achieve higher accuracy and offer nuanced insights into both cognitive and physical aspects of engagement. However, this increased precision comes with additional complexity and computational demands, which can limit their applicability in large educational settings. The use of multimodal data also raises privacy concerns, as these systems often require the collection of sensitive information that must be managed carefully to ensure ethical standards.

These findings reveal that while each method has strengths, their effectiveness often depends on context. The most promising path forward may lie in developing models that are adaptable to real-world variability, cost-effective, and respectful of students' privacy as the continuous monitoring of visual and physiological data necessitates transparent, secure datahandling practices.

5. Conclusion

This review highlights the growing importance of detecting student engagement in online learning environments through facial expressions, gaze tracking, and fatigue detection, with multimodal approaches proving particularly effective. While many systems achieve high accuracy, especially in controlled environments, challenges remain in deploying these solutions at scale in real-world settings. Issues such as variability in lighting, camera quality, and student behavior complicate the generalization of models trained on standard datasets. Additionally, concerns about privacy and the technical requirements for multimodal systems pose barriers to widespread adoption. To move forward, future research should focus on developing models that are more robust to environmental factors and require fewer computational resources, ensuring that these systems can be used in diverse e-learning settings. Overall, facial feature-based engagement detection offers a promising avenue for enhancing online education, but continued innovation and refinement are necessary in real-world applications.

[1] Rothkrantz L. Dropout Rates of Regular Courses and MOOCs. Computers Supported Education. 25–46 (2017).

[2] Mishra L., Gupta T., Shree A. Online teaching-learning in higher education during lockdown period of COVID-19 pandemic. International Journal of Educational Research Open. **1**, 100012 (2020).

[3] Villaroya S. M., Gamboa-Montero J. J., Bernardino A., Maroto-Gomez M., Castillo J. C., Salichs M. A. Real-time Engagement Detection from Facial Features. 2022 IEEE International Conference on Development and Learning (ICDL). 231–237 (2022).

[4] Dewan M. A. A., Murshed M., Lin F. Engagement detection in online learning: a review. Smart Learn. Environ. **6** (1), 1 (2019).

[5] Pal S., Mukhopadhyay S., Suryadevara N. Development and Progress in Sensors and Technologies for Human Emotion Recognition. Sensors. **21** (16), 5554 (2021).

[6] Bhardwaj P., Gupta P. K., Panwar H., Siddiqui M. K., Morales-Menendez R., Bhaik A. Application of Deep Learning on Student Engagement in e-learning environments. Computers & Electrical Engineering. **93**, 107277 (2021).

[7] Meriem B., Benlahmar H., Naji M. A., Sanaa E., Wijdane K. Determine the Level of Concentration of Students in Real-Time from their Facial Expressions. International Journal of Advanced Computer Science and Applications (IJACSA). **13** (1), 159–166 (2022).

[8] Gupta S., Kumar P., Tekchandani R. EDFA: Ensemble deep CNN for assessing student's cognitive state in adaptive online learning environments. International Journal of Cognitive Computing in Engineering. **4**, 373–387 (2023).

[9] Lasri I., Solh A. R., Belkacemi M. E. Facial Emotion Recognition of Students using Convolutional Neural Network. 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS). 1–6 (2019).

[10] Hsia C.-H., Chiang B., Ke L.-Y., Ciou Z.-Y., Lai C.-F. Student Engagement Analysis Using Facial Expression in Online Course. 2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA). 1–2 (2022).

[11] Zhao H., Kim B.-G., Slowik A., Pan D. Temporal–spatial correlation and graph attention-guided network for micro-expression recognition in English learning livestreams. Discover Computing. **27** (1), 47 (2024).

[12] Ayvaz U., Gürüler H., Devrim M. O. Use of Facial Emotion Recognition in e-Learning Systems. Information Technologies and Learning Tool. **60** (4), 95–104 (2017).

[13] Albarrak K. M., Sorour S. E. Web-Enhanced Vision Transformers and Deep Learning for Accurate Event-Centric Management Categorization in Education Institu-tions. Systems. **12** (11), 475 (2024).

[14] Abbas R., Ni B., Ma R., Li T., Lu Y., Li X. Context-based emotion recognition: A survey. Neurocomputing. **618**, 129073 (2025).

[15] Jamil N., Belkacem A. N., Lakas A. On enhancing students' cognitive abilities in online learning using brain activity and eye movements. Education and Information Technologies. **28** (4), 4363–4397 (2023).

[16] Deng R., Gao Y. A review of eye tracking research on video-based learning. Education and Information Technologies. **28** (6), 7671–7702 (2023).

[17] Burch M., Haymoz R., Lindau S. The Benefits and Drawbacks of Eye Tracking for Improving Educational Systems. ETRA '22: 2022 Symposium on Eye Tracking Research and Applications. 53, 1–5 (2022).

[18] Lin Y., Zhang P., Song G. Study on Learning Status Information in Education Based on Line-of-Sight Tracking in Artificial Intelligence Environment. **2021** (8), (2021).

[19] Haataja E., Salonen V., Laine A., Toivanen M., Hannula M. S. The Relation Between Teacher–Student Eye Contact and Teachers' Interpersonal Behavior During Group Work: a Multiple–Person Gaze–Tracking Case Study in Secondary Mathematics Education. Educational Psychology Review. **33** (1), 51–67 (2021).

[20] Li X., Lin J., Tian Z., Lin Y. An Explainable Student Fatigue Monitoring Mod-ule with Joint Facial Representation. Sensors. **23** (7), 3602 (2023).

[21] Dedhia R., Dand M., Akhave A., Sonawane B., Scholar U. Drowsiness Detection in E-Learning. Journal of Emerging Technologies and Innovative Research. **6** (4), (2019).

[22] Lahoti U., Joshi R., Vyas N., Deshpande K., Jain S. Drowsiness Detection System for Online Courses. International Journal of Advanced Trends in Computer Science and Engineering. **9** (2), 1930–1934 (2020).

[23] Abdulkader R., Tayseer Mohammad Ayasrah F., Nallagattla V. R. G., Kant Hiran K., Dadheech P., Balasubramaniam V., Sengan S. Optimizing student engagement in edge-based online learning with advanced analytics. Array. **19**, 100301 (2023).

[24] Safarov F., Akhmedov F., Abdusalomov A. B., Nasimov R., Cho Y. I. Real-Time Deep Learning-Based Drowsiness Detection: Leveraging Computer-Vision and Eye-Blink Analyses for Enhanced Road Safety. Sensors. **23** (14), 6459 (2023).

[25] Li D., Cui Z., Cao F., Cui G., Shen J., Zhang Y. Learning State Assessment in Online Education Based on Multiple Facial Features Detection. Computational Intelligence and Neuroscience. **2022**, 3986470 (2022).

[26] Uçar M. U., Özdemir E. Recognizing Students and Detecting Student Engagement with Real-Time Image Processing. Electronics. **11** (9), 1500 (2022).

[27] Gupta S., Kumar P., Tekchandani R. A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. Multimedia Tools and Applications. **82** (18), 28589–28615 (2023).

[28] Kawamura R., Shirai S., Takemura N., Alizadeh M., Cukurova M., Takemura H. Detecting Drowsy Learners at the Wheel of e-Learning Platforms With Multimodal Learning Analytics. IEEE Access. **9**, 115165–115174 (2021).

[29] Watanabe K., Sathyanarayana T., Dengel A., Ishimaru S. EnGauge: Engagement Gauge of Meeting Participants Estimated by Facial Expression and Deep Neural Network. IEEE Access. **11**, 52886–52898 (2023).

[30] Sassi A., Chérif S., Jaafar W. Intelligent Framework for Monitoring Student Emotions During Online Learning. Engineering Applications of Neural Networks. 207–219 (2024).

[31] Abdellaoui B., Remaida A., Sabri Z., Abdellaoui M., El Hafidy A., El Bouzekri El Idrissi Y., Moumen A. Analyzing emotions in online classes: Unveiling insights through topic modeling, statistical analysis, and random walk techniques. International Journal of Cognitive Computing in Engineering. **5**, 221–236 (2024).

Методи виявлення взаємодії на основі обличчя: огляд

Карбал І., Саель Н., Уахабі С.

Факультет природничих наук, Університет Хасана II у Касабланці, Касабланка, Марокко

Виявлення залученості студентів в онлайн-навчальних середовищах стає дедалі важливішим із широким впровадженням платформ електронного навчання. У цій статті розглядаються сучасні підходи до моніторингу залученості студентів на основі міміки, відстеження погляду, виявлення втоми та сонливості, а також мультимодальних систем. Аналізуючи міміку, системи можуть виявляти емоційні стани, такі як щастя, розчарування та нудьга, пропонуючи викладачам зворотний зв'язок у режимі реального часу. Відстеження погляду дає уявлення про зосередженість студентів, хоча такі проблеми, як вартість обладнання та умови освітлення, впливають на його точність. Виявлення втоми та сонливості за допомогою аналізу моргання та позіхання допомагає виявити когнітивне перевантаження, тоді як мультимодальні системи, що поєднують дані про обличчя, поведінку та фізіологічні дані, пропонують більш повну картину залученості. Цей огляд підкреслює потенціал цих методів, водночас розглядаючи потребу в більш надійних, масштабованих та конфіденційних системах для моніторингу залученості в режимі реального часу в різних контекстах електронного навчання.

Ключові слова: *міміка; відстеження погляду; сонливість; виявлення втоми; мультимодальні підходи.*