2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 **Research Article**

International Journal Research Publication Analysis

Page: 01-10

A SURVEY ON CLASSROOM ENGAGEMENT SYSTEMS USING ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

[*1]Dr. Abhishek Kulkarni, [2]Ananya R. Shastry, [2]Arpitha R., [2]Inchara Dhanaraj, [2]Shreya Basavaraju

[1]Associate Professor Department of Computer Science and Engineering Jyothy Institute of Technology Bengaluru, India.

[2]Department of Computer Science and Engineering Jyothy Institute of Technology Bengaluru, India.

Article Received: 28 March 2026
Article Revised: 18 April 2026
Published on: 08 May 2026

***Corresponding Author: Dr.Abhishek Kulkarni**
Associate Professor Department of Computer Science and Engineering Jyothy Institute of Technology Bengaluru, India.

DOI: https://doi-doi.org/101555/ijrpa.8665

ABSTRACT

In the landscape of modern education, accurately detecting student engagement has become a priority for improv-ing academic success and reducing dropout rates. Traditional methods, such as self-reports and subjective teacher assessments, are often limited by human bias and a lack of real-time resolution. This survey explores the shift toward automated systems that use computer vision and machine learning to analyze engage-ment through facial expressions and behavioral data. Current research highlights the effectiveness of deep learning models, particularly Convolutional Neural Networks (CNNs), in distin-guishing varying levels of student participation. These models have demonstrated near-human accuracy in binary classification tasks, such as identifying highly engaged versus disengaged states. Furthermore, studies show that automated engagement judg-ments correlate strongly with actual task performance and pre-test scores. The survey concludes that the integration of advanced architectures like hybrid CNNs and ensemble learning offers a robust framework for real-time monitoring. These technologies provide a scalable solution for diverse learning environments, including massive open online courses (MOOCs) and Affect-sensitive Intelligent Tutoring Systems (ITS). The findings suggest that automated tools can revolutionize how educators respond to student needs in dynamic classroom settings.

INDEXTERMS: Artificial Intelligence, Machine Learning, Com-puter Vision, CNN, Student Engagement, OpenCV.

1. INTRODUCTION

Student engagement is a multifaceted concept encompass-ing behavioral, emotional, and cognitive involvement in the learning process. Historically, interest in this field was sparked by high dropout rates and statistics showing that a large percentage of students felt chronically bored in the classroom. Successful education is now measured not just by knowledge retention, but by a student's appetite and capacity to learn.

With the rapid integration of digital technology in education, engagement strategies are evolving to include virtual learn-ing environments (VLEs) and educational games. Behavioral engagement refers to a student's willingness to participate, such as submitting work and following directions, while emo-tional engagement describes their attitude toward the material. Cognitive engagement is the most complex, involving focused attention and creative thinking to maximize learning abilities. Automated engagement detection addresses the subjectivity of manual observation by providing unbiased, real-time data. By predicting participation through interaction logs and visual cues, these systems allow instructors to adapt their teaching practices to shifts in student motivation. The ultimate goal is to create flexible, adaptive learning environments that foster growth and prevent academic failure at an early stage.

In addition to traditional classroom settings, the increasing adoption of digital learning platforms has further emphasized the need for automated engagement analysis. With the rise of online classes, hybrid learning models, and virtual classrooms, educators often lack direct interaction with students, making it difficult to assess their level of understanding and interest. AI-driven engagement detection systems provide a valuable solution by continuously monitoring student behavior through visual and interaction-based cues. These systems not only support teachers in real-time decision-making but also con-tribute to personalized learning by identifying individual stu-dent needs. As educational environments continue to evolve, the integration of intelligent engagement monitoring tools becomes essential for creating adaptive and student-centered learning experiences.

2. RELATED WORKS

Early research in the field focused on facial expression recognition and facial action unit classification to estimate perceived engagement. Researchers have developed teams of labelers to rate video clips and static images, finding that human observers show high reliability when discriminating between high and low engagement. These studies laid the groundwork for using static expressions as a primary data source for machine learning models.

Other significant contributions involve analyzing student data from Virtual Learning Environments (VLEs). Studies us-ing the Open University Learning Analytics Dataset (OULAD) have compared various machine learning algorithms, such as CATBoost, Random Forest, and XGBoost, to identify the best classifiers for handling heterogeneous log data. These models have proven effective in spotting academic failure by correlating student interactions with final exam scores.

Recent advancements have shifted toward deep learning and computer vision techniques specifically for classroom analytics. Models like YOLOv4 and various ensemble learning approaches have been proposed to monitor engagement in large classroom settings. These newer works emphasize the use of micro-expressions and subtle facial cues to provide more granular insights into student attentiveness.

Recent studies have looked into ways to detect engagement that go beyond just faces. They mix in things like how students sit or move their hands, and even what they do with the lesson material. I think this makes sense because facial stuff alone can miss the point sometimes, like if a kid is tired but still paying attention through their body language.

Hybrid models seem like another big thing. You take deep learning for pulling out features from pictures, then pair it with older methods like SVM or Random Forest to make the final call. In classrooms, where lights change or kids block each other, this combo handles it better, I guess.

Online classes have pushed this further, especially since everything went virtual. They use webcams plus logs of clicks and how long someone stays on a task. It feels like adapting these tools to both in person and online spots is key, but not always easy.

Privacy issues come up a lot with these systems. Plus the computing power they need, and how they do not work well for all kinds of students. This part gets a bit messy, but it shows we need stuff that is more flexible overall.

One thing that stands out is the interaction patterns in multimodal setups. They help get a fuller picture of behavior. Sort of overcomes limits from just expressions.

Another important trend observed in recent literature is the shift towards multimodal engagement detection systems. Instead of relying only on facial expressions, researchers are combining multiple data sources such as body posture, hand gestures, and interaction patterns. This multimodal approach provides a more comprehensive understanding of student behavior, as engagement is not always accurately reflected through facial expressions alone. For example, a student may appear neutral facially but still be cognitively engaged through note-taking or interaction.

Several studies have also explored the role of temporal anal-ysis in engagement detection. Rather than analyzing individual frames, these approaches consider sequences of frames over time to capture dynamic changes in student behavior. This helps in distinguishing between short-term distractions and consistent disengagement. Temporal models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are often used to model these time-based patterns.

In addition, researchers have investigated the use of atten-tion mechanisms within deep learning models to improve per-formance. Attention-based models focus on the most relevant parts of an image, such as the eyes or mouth, which are critical indicators of engagement. This selective focus helps improve accuracy and reduces the influence of irrelevant background information.

Another area gaining importance is the use of explainable AI (XAI) techniques in engagement detection systems. These methods aim to make AI predictions more transparent and interpretable for educators. By understanding why a system classifies a student as disengaged, teachers can gain deeper insights into student behavior and take appropriate corrective actions.

Despite these advancements, challenges such as privacy concerns, data security, and ethical considerations remain significant. Many researchers highlight the importance of de-signing systems that respect student privacy while still provid-ing meaningful insights. This balance between technological advancement and ethical responsibility is a key focus in current research.

3. TECHNOLOGIES USED

The primary technology utilized in these systems is the Convolutional Neural Network (CNN), which is essential for extracting spatial features from facial images and body posture. Advanced architectures such as Xception, which uses depthwise separable convolutions, and NASNetMobile, opti-mized for mobile applications, are often hybridized to improve feature extraction accuracy. These deep learning frameworks enable the processing of complex visual data even in low-power educational environments.

In addition to deep learning, traditional machine learn-ing classification algorithms play a vital role in analyzing interaction logs. Ensemble techniques like LightGBM and CATBoost are favored for their ability to handle categorical variables directly and their efficiency in building symmetric, balanced decision trees. These algorithms are often optimized using frameworks like Optuna to find the best-performing hyperparameters for engagement prediction.

Computer vision toolboxes, such as the Computer Expres-sion Recognition Toolbox (CERT), are frequently employed for face registration and head pose estimation. These tools can automatically estimate the intensities of facial actions—like inner brow raisers or eye closures—and the 3-D pose of the head. By combining these automated estimates with multi-nomial logistic regression, researchers can identify the most discriminating features that humans use to judge engagement. The domain has reached a vital technological progress through transfer learning which allows deep learning models to build upon their original training for engagement detection tasks. The models VGGNet and ResNet and MobileNet have gained popularity because they can extract complex infor-mation from extensive image databases. Learning through transfer shortens training periods while it enhances model precision when students must deal with restricted training data to develop their skills.

The implementation of these systems depends on real-time processing frameworks which serve as their fundamental operational basis. The OpenCV library functions as a quick image processing tool which also captures video content to support classroom settings where students need immediate ac-cess. The frameworks help users detect multiple faces at once while tracking their movements which becomes necessary for monitoring all students in big classrooms.

The computational requirements of AI-based systems have driven organizations to implement cloud computing and edge computing technologies for their operations. Cloud platforms deliver powerful processing capabilities together with storage solutions yet edge devices perform immediate data processing through their built-in hardware systems. The system achieves operational efficiency because it connects to servers which maintain its expanding capability during real-time application performance.

The process of data preparation requires two essential steps which include preprocessing and augmentation methods to strengthen model performance. The training process includes image rotation and scaling and noise addition to help models develop resistance against various classroom conditions which include changing light levels and different camera positions and student looks.

Another important trend observed in recent literature is the shift towards multimodal engagement detection systems. Instead of relying only on facial expressions, researchers are combining multiple data sources such as body posture, hand gestures, and interaction patterns. This multimodal approach provides a more comprehensive understanding of student behavior, as engagement is not always accurately reflected through facial expressions alone. For example, a student may appear neutral facially but still be cognitively engaged through note-taking or interaction.

Several studies have also explored the role of temporal anal-ysis in engagement detection. Rather than analyzing individual frames, these approaches consider sequences of frames over time to capture dynamic changes in student behavior. This helps in distinguishing between short-term distractions and consistent disengagement. Temporal models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are often used to model these time-based patterns.

In addition, researchers have investigated the use of atten-tion mechanisms within deep learning models to improve per-formance. Attention-based models focus on the most relevant parts of an image, such as the eyes or mouth, which are critical indicators of engagement. This selective focus helps improve accuracy and reduces the influence of irrelevant background information.

Another area gaining importance is the use of explainable AI (XAI) techniques in engagement detection systems. These methods aim to make AI predictions more transparent and interpretable for educators. By understanding why a system classifies a student as disengaged, teachers can gain deeper insights into student behavior and take appropriate corrective actions.

Despite these advancements, challenges such as privacy concerns, data security, and ethical considerations remain significant. Many researchers highlight the importance of de-signing systems that respect student privacy while still provid-ing meaningful insights. This balance between technological advancement and ethical responsibility is a key focus in current research.

4. REAL-TIME APPLICATION

One of the most promising real-time applications is the development of Affect-sensitive Intelligent Tutoring Systems (ITS). These closed-loop systems use affective sensors to detect real-time engagement signals, allowing the software to adjust its teaching strategy automatically, much like a human tutor would. This ensures that the material remains challenging enough to keep the student engaged without causing frustra-tion.

In distance-learning and MOOC environments, real-time dashboards provide human teachers with immediate feedback on their audience’s level of interest. These dashboards can visualize disengaged segments of educational videos, enabling instructors to modify content to better hold student attention. This application is crucial for maintaining accountability and proving learning efficacy in increasingly digital higher educa-tion settings.

The implementation of these systems on low-power devices through edge computing is another emerging application area.

By utilizing efficient models like MobileNet, automated en-gagement detection can be deployed on tablets and smart-phones used in the classroom. This allows for widespread, unobtrusive monitoring that can help educational institutions intervene before a student falls too far behind.

Smart classroom environments use real-time engagement detection systems which allow teachers to create instant in-structional decisions. Teachers should change their teaching methods when most students show no interest in the lesson while they need to add interactive learning activities and review essential information. The system creates an ongoing feedback loop which strengthens the entire educational expe-rience.

The system uses engagement data for personalized learning systems to deliver customized content which meets each student’s distinct learning requirements. The system shows students their attention patterns and emotional reactions which lead to personalized learning content that helps them learn better and remember information longer.

These systems work together with learning management systems (LMS) to deliver extended analytics which track stu-dent performance and their engagement patterns. The revealed information allows teachers to detect students who need extra assistance while they can implement early support methods to stop students from falling behind in their studies.

The educational sector has started using these engage-ment detection systems which now also receive interest from corporate training programs and online meeting platforms and human-computer interaction technology developers. The system shows that AI-based engagement analysis tools operate successfully in multiple actual situations which exist in the real world.

Another emerging application of real-time engagement de-tection systems is in adaptive classroom environments, where the system can automatically adjust teaching aids based on student responses. For instance, smart boards or presentation systems can modify the pace of content delivery, highlight key concepts, or introduce interactive elements when a drop in engagement is detected. This creates a more responsive learn-ing environment where teaching strategies are continuously optimized according to student behavior.

Additionally, real-time engagement analysis can support institutional decision-making by providing aggregated insights across multiple classrooms. Educational institutions can analyze trends in student engagement over time, identify subjects or teaching methods that result in lower attention levels, and implement improvements accordingly. Such data-driven approaches enable better curriculum design and teaching methodologies, ultimately enhancing the overall quality of education.

5. CHALLENGES

A significant challenge in automated engagement recogni-tion is achieving generalization across diverse populations. Modern face detectors often struggle to accurately detect individuals with dark skin, making it difficult for systems trained on one demographic to perform reliably on another. This issue necessitates the collection of diverse datasets to ensure that the AI remains fair and robust across different races and classroom environments.

Technical hurdles also include the high temporal resolution required to distinguish between meaningful disengagement and brief, natural events like eye blinking. Determining the correct timescale for data annotation—whether it be 10-second clips or individual frames—is essential for ensuring that the training labels are both reliable and valid. Inconsistent human labeling of what actually constitutes ”engagement” can also lead to noise in the training data.

Finally, the use of physiological and neurological sensors, such as EEG or heart rate monitors, presents practical diffi-culties for large-scale classroom implementation. While these sensors provide high-quality data, their specialized nature and the potential for occlusions, such as hand-on-mouth gestures, make them less practical than non-invasive computer vision methods. This drives the ongoing research into more reliable visual-only detection systems.

6. CONCLUSION AND FUTURE SCOPE

In conclusion, the research demonstrates that automated systems using machine learning and deep learning can suc-cessfully detect student engagement with accuracy comparable to human observers. Hybrid CNN architectures and ensemble models have proven to be the most effective tools for capturing the intricate cues indicative of participation. These technolo-gies validate the potential for creating responsive and dynamic learning environments that can significantly improve overall learning outcomes.

The strong correlation found between automated engage-ment judgments and student task performance underscores the practical value of these systems. While traditional machine learning methods like CATBoost excel at log-based analyt-ics, deep learning models are superior for real-time visual monitoring. Together, these approaches offer a comprehensive framework for educators to monitor and enhance student participation.

Looking forward, the future of the field lies in integrating multimodal data, including eye gaze, posture analysis, and physiological signals, to further refine detection accuracy. There is also a growing interest in ”Explainable AI” (XAI) to help educators understand the reasoning behind an engagement score, providing deeper insights into student behavior. As datasets continue to expand with more diverse classroom recordings, these systems will become increasingly robust and universally applicable.

REFERENCES

1. Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., Movellan, J. R. ”The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions.” Transactions on Affective Computing, 2014.
2. Gopal, V. P. H., Babu, P. A., Anays, S. M., Hussian, P. J., Hemanjali, M., Charan, M. S. ”AI-based Student Engagement Detection: Leveraging Convolution Neural Network Models for Classroom Analytics.” Inter-national Journal of Intelligent Systems and Applications in Engineering, 2024.
3. FER2013 Dataset, Kaggle.
4. Alruwais, N., Zakariah, M. ”Student-Engagement Detection in Class-room Using Machine Learning Algorithm.” Electronics, 2023.
5. Z. Zhou, “Artificial Intelligence and Student Engagement in Online Learning: A Literature Review,” *American Journal of Distance Edu-cation*, 2025. contentReference[oaicite:0]index=0
6. A. Singh, “The Impact of AI Tools on Student Engagement and Learning Outcomes,” *ScienceDirect*, 2026. :contentReference[oaicite:1]index=1
7. A. S. Pillai, “Student Engagement Detection in Classrooms through Computer Vision and Deep Learning: A Novel Approach Using YOLOv4,” 2024. contentReference[oaicite:2]index=2
8. N. Alruwais and M. Zakariah, “Student Engagement Detection in Classroom Using Machine Learning Algorithm,” *Electronics*, 2023. contentReference[oaicite:3]index=3

9. Y. Chen, “A Novel Student Engagement Analysis of Real Classroom Using AI,” *Sensors*, 2025. :contentReference[oaicite:4]index=4
10. D. Y. Long, “Artificial Intelligence in Higher Education: A Systematic Review,” *Frontiers in Education*, 2025. :contentRefer-ence[oaicite:5]index=5
11. R. Das, “Optimizing Student Engagement Detection using Facial and Behavioral Features,” *Neural Computing and Applications*, 2025. :con-tentReference[oaicite:6]index=6
12. T.Fu¨tterer,“Artificial Intelligence in Classroom Management: A Systematic Review,” *ScienceDirect*, 2025. :contentRefer-ence[oaicite:7]index=7