PriMera Scientific Engineering
Volume 8 Issue 2 February 2026
DOI: 10.56831/PSEN-08-249
ISSN: 2834-2550

PriMera Scientific Publications

EnganchAI: Preliminary Work on Real-Time Classroom Engagement Analysis Using Computer Vision

Type: Research Study
Received: December 31, 2025
Published: February 03, 2026

Citation:
Mauricio Figueroa Colarte., et al. "EnganchAI: Preliminary Work on Real-Time Classroom Engagement Analysis Using Computer Vision". PriMera Scientific Engineering 8.2 (2026): 17-24.

Mauricio Figueroa Colarte*, Claudio Valdivia Parra, José Pablo Casas, Alison Bottinelli Thomassen, Vicente Rivas Urrutia and Cristian Molina Pedernera

School of Informatics and Telecommunications, Duoc UC Professional Institute Foundation, Viña del Mar, Chile

***Corresponding Author:** Mauricio Figueroa Colarte, School of Informatics and Telecommunications, Duoc UC Professional Institute Foundation, Viña del Mar, Chile.

Abstract

Student engagement is essential for academic achievement and emotional well-being, encompassing behavioral, emotional, and cognitive dimensions. In face-to-face education, fostering engagement helps create learning environments that are both effective—meeting curricular goals—and affective—nurturing motivation and belonging. Technology-assisted strategies are especially valuable for helping teachers detect disengagement in real time and personalize their responses. These adaptive actions support greater student focus and commitment, enabling a teaching process that integrates both effectiveness and emotional connection.

EnganchAI, derived from the fusion of "Engagement" and "AI" (Artificial Intelligence), introduces a proof of concept (PoC) for real-time student engagement analysis in physical classrooms using computer vision technologies. Designed for low resource environments, this platform employs a YOLO-based model trained on custom-labeled datasets to process live video feeds, classifying engagement levels (Engaged, Bored, Frustrated and Confused). By providing actionable insights, EnganchAI enables educators to adapt teaching strategies dynamically, fostering more effective and affective learning experiences.

The PoC was tested both in controlled environments and in real classroom scenarios, ensuring accurate performance measurements and validation of its real-time capabilities. These trials were conducted under strict confidentiality protocols to protect personal data and ensure compliance with privacy regulations. The platform demonstrated promising results, achieving a mean average precision (mAP) of 70.25% and an inference response time under 2 seconds. Future iterations will focus on refining datasets, enhancing model accuracy, and expanding functionalities to support broader adoption.

Keywords: Engagement Analysis; Artificial Intelligence; Affective Learning; Computer Vision; Classroom Technology

Introduction

The integration of artificial intelligence (AI) in education has opened new avenues for enhancing both the effectiveness and emotional impact of learning processes. EnganchAI, a novel project combining "Engagement" and "AI". addresses the growing need for real-time engagement analysis in classroom settings. By leveraging computer vision and AI, this project seeks to empower educators with actionable insights to dynamically adapt their teaching strategies, fostering an effective and affective learning environment.

Student engagement is commonly understood as a multidimensional construct that encompasses behavioral, emotional, and cognitive dimensions, each reflecting how students participate in and respond to academic experiences. According to Fredricks et al., behavioral engagement involves participation in academic and social activities; emotional engagement relates to affective reactions such as interest or boredom; and cognitive engagement concerns self-regulated learning and the willingness to exert effort in mastering complex ideas. These dimensions interact dynamically and are shaped by contextual factors such as teaching strategies and perceived support from educators [1-3]. Incorporating these distinctions is essential when designing AI models aimed at detecting engagement states, as each dimension may manifest through different observable cues (e.g., posture, attention, facial expression) in real-time classroom settings.

Previous research underscores the critical role of student engagement in academic success and emotional well-being. The "Dataset for Affective States in E-Environments" (DAiSEE) has been instrumental in developing benchmarks for engagement analysis, providing insights into emotional states and behaviors using convolutional neural networks (CNNs) and other advanced techniques [4]. However, these efforts have predominantly focused on virtual learning environments, often neglecting the unique challenges of physical classrooms, such as dynamic interactions, diverse environmental conditions, and the necessity for real-time responses [4, 5].

EnganchAI bridges this gap by introducing a proof-of-concept (PoC) platform that employs a "You Only Look Once" (YOLO)-based model, a real-time object detection system originally designed for general computer vision tasks [6]. The platform has been tested in both controlled and real classroom environments, ensuring accurate performance measurements while respecting strict confidentiality protocols to protect personal data. Initial results demonstrate the system's feasibility, achieving a mean average precision (mAP) of 70.25% and an inference response time under two seconds.

Building on prior implementations like EmoAI Smart Classroom, EnganchAI showcases the potential of lightweight YOLO variants to balance accuracy and computational efficiency in resource-constrained educational settings [6, 7]. This paper presents the conceptual foundation, methodology, and results of EnganchAI, highlighting its role as an innovative tool to transform classroom experiences. The subsequent sections explore related work, methodology, results, and the roadmap for scaling the platform toward a fully functional Minimum Viable Product (MVP).

State of the art

The study of student engagement using artificial intelligence (AI) has advanced significantly, leveraging computer vision models and specialized datasets. Among the most prominent contributions is the Dataset for Affective States in E-Environments (DAiSEE), introduced by Gupta et al. [4]. DAiSEE provides over 9,000 video fragments annotated at four levels of engagement (very low, low, high, very high) and has been instrumental in setting benchmarks for engagement analysis in virtual environments. Techniques such as convolutional neural networks (CNNs) and Long- Term Recurrent Convolutional Networks (LRCNs) applied to DAiSEE have demonstrated accuracies up to 57.9%, showcasing the dataset's potential for analyzing affective states in educational settings [4].

Beyond its quantitative benchmarks, DAiSEE offers a strong methodological foundation for modeling engagement. Its annotations were generated using a combination of crowdsourced inputs and expert evaluation by psychologists, ensuring both intersubjective validation and theoretical relevance [4]. This hybrid labeling process allows for a reliable operationalization of engagement states. For instance, Engaged behavior typically includes an upright posture, forward-facing gaze, and active facial expressions indicating focus or interest. Bored students may display a slouched posture, gaze aversion, yawning, or minimal facial responsiveness—consistent

with under-stimulation or low perceived task value [8]. Confused behavior often involves furrowed brows, squinting, or frequent gaze shifts, indicating cognitive disequilibrium or lack of comprehension [3]. Frustrated behavior may include frowning, lip pressing, or tense facial muscles, commonly associated with perceived obstacles or failure to meet task demands [9]. In EnganchAI, these annotated examples serve as the basis for supervised training, enabling the model to classify states effectively.

Building on DAiSEE, Malekshahi et al. [5] proposed a general and lightweight model optimized for engagement detection. Their work achieved an accuracy of 68.57%, surpassing many state-of-the-art methods. This highlights the feasibility of efficient algorithms for engagement analysis, even when working with datasets that exhibit inherent imbalances. However, these approaches primarily target virtual environments, where variables such as lighting and perspective are more controlled.

The introduction of YOLO ("You Only Look Once") models revolutionized real-time object detection by combining speed and accuracy [6]. In educational contexts, YOLO-based architectures have demonstrated their potential to process video feeds in real time, enabling immediate insights into student engagement. For instance, the EmoAI Smart Classroom project leveraged YOLO models to detect emotional and behavioral cues in students, demonstrating the practicality of computer vision in STEM classrooms [7]. Similarly, Alkabbany et al. [10] explored the use of YOLO in real-time engagement measurement platforms, achieving high precision and recall rates in controlled classroom environments.

Beyond YOLO, hybrid architectures like ResNet combined with Temporal Convolutional Networks (TCNs) have significantly enhanced engagement classification accuracy. Abedi and Khan [11] demonstrated how these architectures could capture both spatial and temporal features in online learning datasets. Meanwhile, Selim et al. [12] introduced a hybrid model combining EfficientNet-B7 with Bi-LSTM and TCN, achieving superior performance in engagement classification for e-learning platforms.

Finally, Wu et al. [13] highlighted the potential of multi- modality datasets such as CMOSE, which integrate various data sources (e.g., video, audio, physiological signals) to improve engagement detection. These datasets set a benchmark for analyzing complex student behaviors and offer opportunities for holistic AI solutions in education.

EnganchAI builds on this body of research by addressing the challenges of applying these technologies to physical classrooms. Unlike previous works that focus on virtual or hybrid environments, EnganchAI employs an optimized YOLO-based model tailored for low-resource settings, achieving a mean average precision (mAP) of 70.25% with an inference response time under two seconds. The platform has been validated in both controlled and real classroom environments, ensuring its scalability and adaptability to dynamic conditions while prioritizing ethical considerations such as privacy and data security.

Methods

This section details the methodology implemented in the development and testing of the EnganchAI platform, leveraging advanced computer vision and artificial intelligence techniques.

Inference Pipeline

The platform employs a two-stage inference pipeline: (1) Face Detection: A pre-trained YOLOv8 model (WideFace) extracts face frames from the video stream and (2) Engagement Classification where the face frames are passed to a custom-trained YOLOv11 model that classifies engagement into categories such as Engaged, Bored, Frustrated and Confused.

Fig. 1 provides a visual representation of this pipeline, highlighting the interaction between the models and the real- time video stream.

System Architecture

The EnganchAI platform is structured around three main components:

- ***Frontend (Client):*** Built using ReactJS and the Next.js framework, the frontend provides an intuitive interface for educators. TailwindCSS is utilized for styling, ensuring a responsive and user-friendly design. The frontend fetches data from the backend via a RESTful API, displaying real-time video analysis and engagement metrics.
- ***Backend (Server):*** Developed with Node.js and Express.js, the backend handles API requests, user authentication, and data transfer between components. It is integrated with the inference module to process video streams.
- ***Inference Module:*** A Flask-based application processes video streams in real-time. The module implements a YOLO-based object detection model optimized for detecting engagement levels. It extracts features such as facial expressions and body posture from the live feed and sends the results to the frontend. Fig. 2 illustrates the complete system architecture, showcasing the interaction between components and the flow of data through the platform.

To ensure scalability and compatibility, the system is containerized using Docker. The client and server components are isolated within Docker containers, allowing for seamless deployment across different environments. The Flask-based inference module runs on CPUs to minimize resource constraints, with potential for future GPU acceleration.

Figure 1: A visual diagram of the inference pipeline.

Dataset and Preprocessing

The custom-labeled dataset used for training the YOLO model was constructed by combining publicly available classroom images with targeted image searches. These images were annotated to classify engagement levels into categories such as “Engaged”. “Bored, “Frustrated” and Confused. The dataset underwent augmentation techniques, including horizontal flips, scaling, and rotations, to enhance robustness. The data was split into training (80%), validation (12%), and testing (8%) subsets.

Model Training

The YOLO-based engagement classification model was trained using a custom-labeled dataset. Key training parameters, augmentations and partitioning are shown in TABLE I, Table II and TABLE III.

The training process employed a custom dataset, ensuring alignment with the operational context of the EnganchAI platform. The model was fine-tuned to balance accuracy and computational efficiency, achieving a mean average precision (mAP) of 70.25%.

User Interaction and Analytics

The platform provides real-time alerts when more than 10 students in the feed exhibit signs of disengagement. An aside panel tracks engagement levels over time, while post-class analytics offer a summary of classroom dynamics, enabling educators to adapt their teaching strategies.

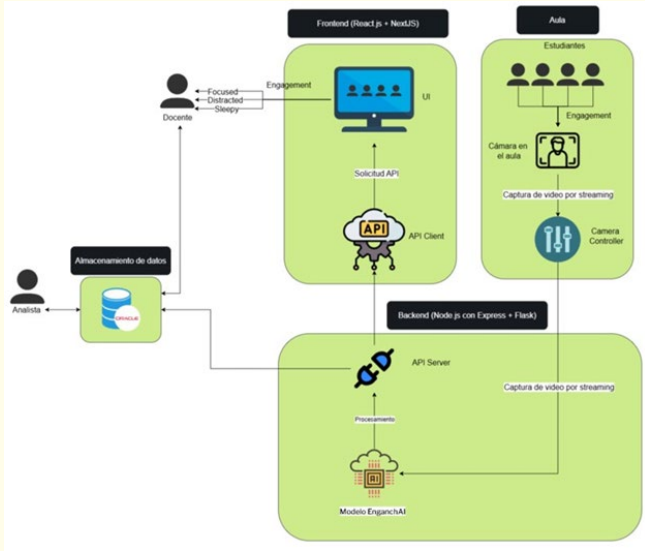

Figure 2: High level architecture diagram.

Hyperparameter	*Value*
Epochs	50
Batch size	16
Image size	640px

Table 1: Yolo Hyperparameter Configuration.

Augmentations	*Value*
Horizontal flip	100%
Vertical flip	Disabled
Scale	±20%
Shear	±10°
Rotation	±10°

Table 2: Yolo Data Augmentation Configuration.

Dataset	*Partitioning Value*
Train	80%
Validation	12%
Test	8%

Table 3: Custom Dataset Partitioning.

Results

The developed YOLO-based system successfully performs inference on live video streams, classifying engagement levels (Engaged, Bored, Frustrated and Confused) in real time. While specific FPS measurements were not taken, the system demonstrated sufficient processing speed to handle continuous IP camera feeds without noticeable lag. A visual representation of the live platform is shown in Fig. 3.

Comparison with the YOLO implementation in EmoAI Smart Classroom [7] is challenging due to differences in datasets and system objectives. However, both systems highlight the feasibility of YOLO-based models for classroom engagement analysis. Unlike EmoAI, which utilizes the DAISEE dataset, this system relies on a custom dataset constructed using publicly available images, which introduces unique challenges and opportunities.

Figure 3: EnganchAI running platform screenshot.

Model	***mAP50***	***Precision***	***Recall***
EmoAI YoloV8	65.30%	74.50%	60.60%
EnganchAI YoloV11	70.25%	67.58%	66.46%

Table 4: Performance Comparison.

The comparison between our YOLO-based model and the one presented in EmoAI Smart Classroom [7] reveals interesting insights into the performance of both systems, as shown in Table IV. Despite having trained for fewer epochs, our model achieved a higher mAP50 (Mean Average Precision at 50% IoU), reaching 70.25%, compared to EmoAI's 65.30%. This suggests that our dataset and training pipeline, although resource constrained, were effective in creating a model capable of generalizing better to the task.

However, our model exhibits lower precision, with an average of 67.58%, in contrast to EmoAI's 74.50%. This discrepancy indicates that our model may be producing more positives that are false during inference. This could stem from the dataset's imbalances, or the broader diversity of images sourced from Google, which may include edge cases or ambiguous scenarios not encountered during EmoAI's training process.

On the other hand, our model demonstrates a significantly higher recall at 66.46%, compared to EmoAI's 60.60%. This indicates that our model is better at detecting the presence of engagement-related classes, even if some predictions are incorrect.

Conclusion

EnganchAI represents a promising step towards real-time engagement analysis in physical classrooms using computer vision technologies. The platform's proof of concept demonstrates the feasibility of applying AI-driven solutions in resource-constrained educational environments, achieving a mean average precision (mAP) of 70.25% and an inference time under 2 seconds. These results validate the system's capability to provide actionable insights to educators, fostering more adaptive and personalized teaching strategies.

While the initial results are encouraging, they also highlight several challenges. Dataset biases, particularly underrepresentation of specific physical traits, affected classification accuracy in some cases. Similarly, the current YOLOv11-based architecture, though efficient, could benefit from further optimization to enhance precision while maintaining recall. Despite these limitations, the modular

architecture and lightweight design of EnganchAI make it a scalable solution with potential for broad adoption. Looking ahead, future work will focus on:

1. ***Dataset Diversification:*** Expanding the dataset to include a wider range of physical traits, classroom environments, and cultural contexts will improve the system's generalization capabilities.
2. ***Real-Time Optimization:*** Introducing GPU acceleration and refining the inference pipeline can further reduce latency and enable multi-stream video analysis.
3. ***Ethical and Privacy Considerations:*** Ensuring data security and privacy remains a top priority. Future iterations will incorporate advanced anonymization techniques to protect students' identities, comply with international data protection regulations (e.g., GDPR), and address ethical concerns related to AI bias and transparency.
4. ***User-Centered Enhancements:*** Developing customizable analytics dashboards and real-time intervention tools will empower educators to make data- driven decisions in the classroom. Importantly, the platform is not intended as a tool for evaluating teacher performance, but rather as a supportive resource that enhances instructional strategies. By providing timely, actionable insights, it helps educators respond more effectively to student needs, ultimately improving engagement and learning outcomes.
5. ***Scalability and Community Engagement:*** Collaborating with interdisciplinary teams and stakeholders in education, cognitive science, and AI ethics will be essential to refining the platform and ensuring its broader impact. Beyond traditional educational settings, the system shows strong potential to scale into other face-to-face environments that involve real-time information delivery—such as seminars, corporate meetings, and professional training sessions—where maintaining audience attention is critical for effective communication and knowledge transfer.
6. ***Model Explainability:*** Explainability and traceability could also help reveal emerging patterns linked to students who may require additional support beyond the classroom context. While not intended for diagnostic use, such insights could inspire new applications of the technology, including early-alert tools or complementary support systems that assist educators in addressing diverse student needs with greater awareness.

In conclusion, EnganchAI lays a solid foundation for the integration of AI technologies into physical classrooms, bridging gaps identified in prior research. By prioritizing ethical considerations and data security, the platform aspires to become a transformative tool that enhances both effective and affective learning experiences. Collaboration and interdisciplinary research will be critical to refining and scaling this platform for broader adoption in global educational contexts.

Acknowledgments

This project was supported by a team of undergraduate students as part of their Capstone Project to obtain the professional title of Computer Engineer at Fundación Instituto Profesional Duoc UC. The authors would like to thank the School of Informatics and Telecommunications for their academic guidance, as well as the faculty advisors who mentored the development process. Special thanks to the students and educators who voluntarily participated in the classroom trials, enabling the real-time validation of the system under ethical and confidential conditions.

References

1. O Alrashidi, H Phan and B Ngu. "Academic Engagement: An Overview of Its Definitions, Dimensions, and Major Conceptualisations". (2016).
2. K Prananto., et al. "Perceived Teacher Support and Student Engagement Among Higher Education Students: A Systematic Literature Review". (2025).
3. B Harris and L Bradshaw. "Battling Boredom Part 2: Even More Strategies to Spark Student Engagement". 2nd Edition ed., New York: Routledge (2017).
4. A Gupta., et al. "DAiSEE: Towards User Engagement Recognition in the Wild". Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV) (2016).

5. S Malekshahi., et al. "A General Model for Detecting Learner Engagement: Implementation and Evaluation". arXiv preprint arXiv:2405.04251 (2024).
6. J Redmon., et al. "You Only Look Once: Unified, Real-Time Object Detection". Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016).
7. EmoAI Smart Classroom: The Development of a Student Emotional and Behavioral Engagement Recognition System (2023).
8. T Goetz., et al. "Academic Boredom". The Routledge International Handbook of Boredom, 1st Edition ed., London, Routledge (2024): 25.
9. R Pekrun., et al. "Educational Psychologist". in Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research, Berlin (2002).
10. I Alkabbany., et al. "An Experimental Platform for Real-Time Students Engagement Measurements from Video in STEM Classrooms". Proceedings of an academic conference (unspecified) (2023).
11. A Abedi and SS Khan. "Improving State-of-the-Art in Detecting Student Engagement with ResNet and TCN Hybrid Network". in Proceedings of the 18th Conference on Robots and Vision (CRV) (2021).
12. T Selim, I Elkabani and MA Abdou. "Students Engagement Level Detection in Online E-Learning Using Hybrid EfficientNetB7 Together with TCN, LSTM, and Bi-LSTM". IEEE Access 10 (2022): 99573-99583.
13. C-H Wu., et al. CMOSE: Comprehensive Multi-Modality Online Student Engagement Dataset with High-Quality Labels (2023).