PT Mekar Research and Publishing
MIKIR : Mathematics, Informatics, Knowledge And Information Research
ISSN : 3110-7699 (Online)
Url : https://mikir.org/mikir

Student Engagement Detection Based on Visual Behavior Indicators Using YOLOv8 on a Public Classroom Dataset

Mohammad Bhanu Setyawan[1*], Angga Prasetyo[1], Fauzan Masykur[1]
[1]Informatic Engineering, Universitas Muhammadiyah Ponorogo, Ponorogo, Indonesia

*Corresponding Author Email: m.banu@umpo.ac.id

Doi : https://dooi.org/10.65475/sser8h18

Key-Words :
student engagement, YOLOv8, object detection, classroom analytics, behavioral engagement, deep learning.

Received : June 29, 2026.
Revised : July 4, 2026.
Published : July 8, 2026

Abstract

Student engagement is an important indicator for evaluating the quality of the learning process; however, its measurement in conventional classrooms still largely relies on subjective and labor-intensive manual observation. This study aims to establish a reproducible baseline for student engagement detection based on visual behavioral indicators using the YOLOv8 model on a public dataset. The working dataset was constructed from two subsets of the Student Class Behavior (SCB) Dataset and restructured into five behavioral classes: hand_raising, reading, writing, bowing_head, and turn_head, resulting in 9,274 image-label pairs split into 6,491 training, 1,854 validation, and 929 test samples. The experiment used YOLOv8n with an image size of 416, a batch size of 8, and 50 effective epochs in Google Colab. Performance was evaluated using precision, recall, mAP@0.5, and mAP@0.5:0.95. The results show that the model achieved a precision of 0.4429, a recall of 0.5393, mAP@0.5 of 0.4630, and mAP@0.5:0.95 of 0.3211. The best class-level performance was observed for writing (AP 0.635) and hand_raising (AP 0.597), while bowing_head (AP 0.288) and turn_head (AP 0.332) remained comparatively weak. These findings indicate that YOLOv8n is feasible as a reproducible baseline for visual student behavior detection, although annotation refinement, comparative experiments, and architectural optimization are still required to strengthen the scientific contribution and the feasibility of real-world classroom deployment.

1. INTRODUCTION

Student engagement is one of the most extensively investigated constructs in educational research due to its close association with student motivation, participation, interaction quality, and learning outcomes [1], [2]. Engagement is generally conceptualized as a multidimensional construct that comprises at least three principal aspects: behavioral engagement, emotional engagement, and cognitive engagement [3]. Among these dimensions, behavioral engagement is typically the most observable in classroom settings because it manifests as visible actions such as paying attention, taking notes, asking questions, or responding to teachers. Nevertheless, the measurement of behavioral engagement in conventional classrooms is still predominantly carried out through teacher observation or manual assessment, which is susceptible to subjectivity, time limitations, and low inter-rater consistency [4].

Recent advances in computer vision and deep learning have opened up opportunities to observe classroom activities in a more objective, measurable, and reproducible manner. A recent systematic review indicates that computer-vision-based approaches are increasingly being employed to detect engagement through facial expressions, gaze direction, body posture, and visible behavioral cues during instruction [1]. In parallel, studies on classroom behavior recognition have shown that the You Only Look Once (YOLO) family remains a relevant choice, as it effectively balances detection accuracy and inference speed in complex classroom scenarios that involve multiple objects and high occlusion [5][6][7][8].

In the national publication context, several studies have employed YOLO for various object detection cases, ranging from waste detection [9], [10] and human detection [11] to the recognition of student emotions during lectures [12]. These studies suggest that YOLOv8 is capable of achieving competitive performance across a broad range of scenarios at relatively low computational cost, making it an attractive choice for adaptation to student behavior detection in Indonesia. Nevertheless, national research has thus far concentrated mostly on object detection in industrial, transportation, and healthcare domains, while studies that explicitly link visual behavior detection to the student engagement framework remain limited[13].

Furthermore, the literature emphasizes that engagement is not fully equivalent to the internal psychological states of students. Therefore, vision-based approaches are best regarded as an operational representation of observable behavioral indicators rather than as an absolute measure of affective or cognitive states [3], [14]. In educational contexts, such approaches remain relevant because they can provide consistent observational information that supports instructional evaluation and the development of intelligent classroom systems [15]. Most prior studies have focused on enhancing model architecture, for instance through attention mechanisms, lightweight modules, or multi-scale feature extraction optimization [6], [7], [8], [16].

Despite the considerable body of existing work, there remains a need to establish transparent and reproducible baselines that align with the data actually available, particularly at the early validation stage before progressing to more complex comparative experiments. Based on this gap, the present study addresses the scarcity of transparent and reproducible YOLOv8 baselines that explicitly map detection classes to the behavioral engagement framework in the Indonesian research context.

This study aims to build a baseline for student engagement detection based on visual behavior indicators using YOLOv8 on a public dataset. The scope is restricted to the behavioral engagement dimension, which is proxied through five visual student behaviors: hand_raising, reading, writing, bowing_head, and turn_head. The main contributions of this study are threefold: (1) constructing a five-class working dataset from publicly available SCB-Dataset subsets unified for the behavioral engagement scenario; (2) establishing a reproducible YOLOv8 baseline experiment in the Google Colab environment; and (3) providing a critical evaluation of the model's performance along with directions for future development.

2. METHODS

2.1 Research Design

This study adopts a quantitative-experimental approach within the object detection domain for educational applications. The primary objective is not to produce a final best-performing model, but rather to establish a valid baseline for detecting visual behavioral indicators of students as a proxy for behavioral engagement. The research workflow consists of four main stages: (1) dataset preparation, (2) model configuration and training, (3) performance evaluation, and (4) result interpretation. This approach is consistent with literature recommendations that early studies on classroom behavior recognition should prioritize pipeline reproducibility before further optimization [6], [16].

2.2 Dataset

The working dataset was built from two publicly available subsets of the Student Class Behavior (SCB) Dataset, which were then unified into the YOLO format. After verifying image-label pairs, organizing the directory structure, and aligning class names, five visual behavior classes were obtained: hand_raising, reading, writing, bowing_head, and turn_head. The final dataset comprises 9,274 image-label pairs, divided into 6,491 training samples, 1,854 validation samples, and 929 testing samples, following a 70:20:10 split. The class distribution is presented in Table 1.

In this study, the behavior labels are positioned as visual indicators of student engagement. The hand_raising, reading, and writing classes are treated as indicators of positive engagement (active engagement), whereas bowing_head and turn_head are regarded as indicators of declining or shifting attention (attention shift). This mapping is consistent with the behavioral engagement typology reported in recent studies on classroom behavior recognition [8],[16].

Table 1. Distribution of the Visual Behavior Dataset

No	Class	Train	Val	Test
1	hand_raising	~1,350	~390	~190
2	reading	~1,300	~370	~185
3	writing	~1,300	~370	~185
4	bowing_head	~1,270	~362	~182
5	turn_head	~1,271	~362	~187
Total		**6,491**	**1,854**	**929**

2.3 Model Configuration and Training

The base model employed in this study is YOLOv8n with pre-trained weights from Ultralytics. The selection of the nano variant was motivated by computational efficiency considerations suited to the Google Colab environment and the need for a compact baseline. The full experimental configuration is presented in Table 2. Training was conducted across multiple cloud sessions and resumed using the last.pt checkpoint. Therefore, this experiment should be understood as a single run that was carried out incrementally (resume training) rather than as one continuous and uninterrupted training session.

Table 2. YOLOv8n Training Configuration

Parameter	Value
Task	Object Detection
Model	YOLOv8n

Parameter	Value
Initial Weights	yolov8n.pt (pre-trained on COCO)
Image Size	416 × 416 px
Batch Size	8
Effective Epochs	50
Optimizer	Auto (default)
Patience	20
Augmentation	mosaic, fliplr, scale, HSV
Hardware	Google Colab GPU
Framework	PyTorch, Ultralytics

2.4 Evaluation

The model was evaluated using standard object detection metrics, namely precision (P), recall (R), mean Average Precision at an IoU threshold of 0.5 (mAP@0.5), and the average mean Average Precision over the IoU range of 0.5–0.95 (mAP@0.5:0.95). In addition, per-class Average Precision, training curves (loss and metric trajectories), and the F1-confidence curve were analyzed to identify the optimal confidence threshold. This multi-metric approach was adopted so that the interpretation would not be limited to a single aggregate score, but would also take into account the distribution of model errors across classes, as recommended in prior YOLO studies on classroom behavior recognition [5], [6], [8].

The overall research workflow is summarized in Figure 1, beginning with problem identification, dataset collection, data restructuring, model training, and concluding with the interpretation of results as a behavioral engagement detection baseline.

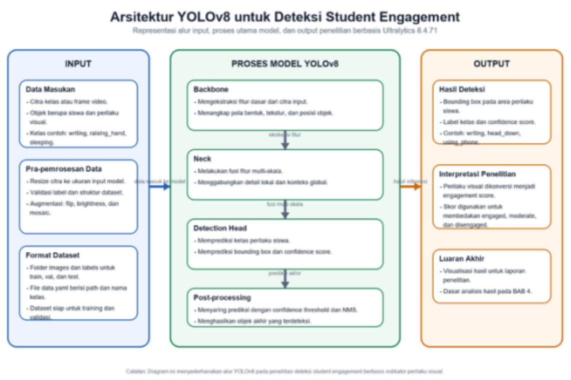

Figure 1. Research workflow for student engagement detection based on visual behavior indicators using YOLOv8.

3. RESULTS AND DISCUSSION

3.1 Overall Performance

The training results of the YOLOv8n model after 50 effective epochs reveal functional but suboptimal baseline performance. A summary of the evaluation metrics is presented in Table 3. The model achieved a precision of 0.4429 and a recall of 0.5393, with mAP@0.5 of 0.4630 and mAP@0.5:0.95 of 0.3211. The best F1-score was 0.48 at a confidence threshold of 0.225, indicating that the optimal trade-off between precision and recall is reached at a moderate confidence threshold.

Table 3. Overall Model Performance

Metric	Value
Precision	0.4429
Recall	0.5393
mAP@0.5	0.4630
mAP@0.5:0.95	0.3211
Best F1-score	0.48
Optimal Conf. threshold	0.225

The fact that recall exceeds precision suggests that the model is relatively aggressive in detecting objects, meaning that more true positives are recovered, but at the expense of additional false positives. This pattern is commonly observed in YOLO baselines trained on limited datasets or with annotations that are not fully consistent [9], [17]. For the purpose of engagement classification, a higher recall can be considered a desirable property because the priority is to minimize missed detections of student behaviors.

3.2 Per-Class Performance

Per-class analysis reveals an uneven performance distribution across the five behavior classes. Table 4 presents the Average Precision (AP) values for each class at an IoU threshold of 0.5.

Table 4. Average Precision per Behavior Class

No	Class	AP@0.5	Category
1	writing	0.635	Positive
2	hand_raising	0.597	Positive
3	reading	0.496	Positive
4	turn_head	0.332	Declining
5	bowing_head	0.288	Declining

The writing class (AP 0.635) and the hand_raising class (AP 0.597) emerged as the two best-performing classes. This pattern is consistent with the visual characteristics of each behavior. The act of writing typically involves a consistent body orientation (torso leaning toward the desk, hand holding a writing tool), while raising a hand displays distinctive spatial features that contrast strongly with the background.

Conversely, the bowing_head (AP 0.288) and turn_head (AP 0.332) classes exhibited lower accuracy. This is presumed to be caused by three main factors: (1) ambiguous poses that may resemble reading when the student is bending toward a book; (2) the often small object size in multi-object classroom scenarios; and (3) variations in camera viewing angles that alter the appearance of the head. Similar patterns have also been reported by Sheng et al. [8], [16] and Feng et al., who note that behavior classes related to head movements generally require attention modules or multi-scale strategies to achieve balanced accuracy.

3.3 Detection Visualization

In addition to quantitative metrics, the model's performance was visually inspected using validation batch predictions (val_batch0_pred, val_batch1_pred, val_batch2_pred). Figures 2 to 4 show the detection outputs on classroom images that contain multiple students simultaneously.

Figure 2. Prediction results on the first validation batch.

In Figure 2, the model produces relatively diverse detections, particularly for the reading and writing behaviors. This indicates that the model can reasonably recognize learning activities with sufficiently clear visual features, such as students leaning toward a book or writing at a desk. However, several predictions show medium confidence values, suggesting that the separation between certain behaviors is not yet fully stable.

Figure 3. Prediction results on the second validation batch.

In Figure 3, hand_raising predictions appear dominant and occur with relatively higher confidence. This reinforces the quantitative finding that hand_raising is among the easiest classes for the model to recognize, owing to its distinctive and visually contrasting posture. On the other hand, several reading, writing, and bowing_head predictions remain visually close to one another, which may lead to classification ambiguity.

Figure 4. Prediction results on the third validation batch.

In Figure 4, the model again demonstrates strong performance on the hand_raising and reading classes, while also detecting a portion of the writing and bowing_head objects. Taken together, the three visualizations support the quantitative finding that the model performs better on behaviors with distinct spatial features and reasonably well on common learning activities, but still struggles with more ambiguous behaviors in dense classroom settings.

3.4 Comparison with Previous Studies

To position the baseline performance within the existing literature, Table 5 presents a brief comparison with several relevant studies that have applied YOLO to student behavior detection in classrooms.

Table 5. Comparison with Previous Studies

Study	Model	Classes	mAP@0.5
Chen et al. [5]	YOLOv8 mod.	3	0.776
Sheng et al. [8]	YOLOv8s mod.	6	0.820
Feng et al. [16]	IMRMB-Net	5	0.830
Zhang et al. [6]	YOLOv8+Att.	4	0.790
Helmiyah et al. [12]	YOLO (emotion)	3	0.917*
This study	**YOLOv8n base.**	**5**	**0.463**

**Accuracy, not mAP*

Table 5 shows that the mAP achieved in this study is lower than in previous studies, which typically employ modified YOLO architectures such as attention modules, multi-scale feature extraction optimization, or advanced training strategies. This gap is reasonable given that the present study explicitly positions itself as a baseline without architectural modification, so that the resulting value can serve as a reference point for subsequent experiments.

3.5 Implications and Limitations

The findings suggest that YOLOv8n is feasible as a reproducible baseline for detecting visual behavioral indicators of students in classrooms. This is consistent with prior studies that have positioned YOLO as a practical architecture for classroom behavior recognition because of its efficiency and adaptability in multi-object environments [5], [7], [8]. The strongest performance on writing and hand_raising confirms that the model more readily recognizes visual patterns with distinctive postural and morphological structures, whereas the weaker performance on bowing_head and turn_head illustrates the classic challenges of object detection in classrooms, namely ambiguous poses, small object size, and inter-student occlusion [16], [18], [19].

Conceptually, the findings should be interpreted with caution. Visual behavior detection cannot be directly equated with the actual psychological state of engagement. The literature emphasizes that engagement is multidimensional and can involve behavioral, emotional, and cognitive aspects [1], [3]. Therefore, the system developed in this study is more accurately understood as a detection system for behavioral engagement indicators based on visual observation. Its strengths lie in objectivity and reproducibility, while its limitation is that it does not directly capture students' affective and cognitive dimensions.

Other limitations of this study include: (1) the use of the YOLOv8 nano variant without comparison against the s, m, or l variants; (2) the dataset being sourced from a single domain (the public SCB-Dataset), which means generalization to Indonesian classrooms still needs to be verified; and (3) the absence of real-time inference testing in actual classroom settings. These limitations also indicate directions for further development, including auditing annotation quality, benchmarking against larger or newer YOLO variants such as YOLOv10 and YOLOv11 [13], incorporating attention modules as proposed by Zhang et al. [6] and Sheng et al. [8], and conducting real-time camera-based deployment tests in local classroom contexts.

4. CONCLUSION

This study successfully established a baseline for student engagement detection based on visual behavior indicators using YOLOv8n on a public five-class dataset. The model achieved a precision of 0.4429, a recall of 0.5393, mAP@0.5 of 0.4630, and mAP@0.5:0.95 of 0.3211, with the strongest performance on the writing (AP 0.635) and hand_raising (AP 0.597) classes. These results indicate that the YOLOv8 approach is feasible as an initial foundation for student behavior detection in classrooms, particularly in the behavioral engagement dimension.

Nevertheless, the model's performance remains moderate and is not yet strong enough to be positioned as a final solution. The primary weaknesses appear in the ambiguous classes, namely bowing_head and turn_head, and in the tendency to produce false positives that lower precision. Future work is therefore recommended to prioritize annotation cleaning and auditing, comparisons with larger and newer YOLO variants, the addition of ablation experiments based on attention mechanisms, and validation on classroom data more representative of local Indonesian contexts. Through these steps, the present baseline has the potential to be developed into a more accurate and deployable behavioral engagement detection system.

ACKNOWLEDGMENT

The authors would like to thank for the academic support and facilities provided during the conduct of this research, as well as the providers of the public SCB-Dataset, who made this experiment reproducibly feasible.

REFERENCES

[1] I. Qarbal, N. Sael, and S. Ouahabi, “Student’s Engagement Detection Based on Computer Vision: A Systematic Literature Review,” *IEEE Access*, vol. 13, pp. 140519–140545, 2025, doi: 10.1109/ACCESS.2025.3596885.

[2] N. Mahmood, S. M. Bhatti, H. Dawood, M. R. Pradhan, and H. Ahmad, “Measuring Student Engagement through Behavioral and Emotional Features Using Deep-Learning Models,” *Algorithms*, vol. 17, no. 10, p. 458, Oct. 2024, doi: 10.3390/a17100458.

[3] M. Bustos-López, N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, N. A. Cruz-Ramos, and G. Alor-Hernández, “Wearables for Engagement Detection in Learning Environments: A Review,” *Biosensors*, vol. 12, no. 7, p. 509, Jul. 2022, doi: 10.3390/bios12070509.

[4] B. Díaz and M. Nussbaum, “Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence,” *Computers & Education*, vol. 217, p. 105071, Aug. 2024, doi: 10.1016/j.compedu.2024.105071.

[5] H. Chen, G. Zhou, and H. Jiang, “Student Behavior Detection in the Classroom Based on Improved YOLOv8,” *Sensors*, vol. 23, no. 20, p. 8385, Oct. 2023, doi: 10.3390/s23208385.

[6] J. Zhang, L. Guo, and X. Wang, “Student Classroom Behavior Recognition Based on YOLOv8 and Attention Mechanism,” *Information*, vol. 16, no. 11, p. 934, Oct. 2025, doi: 10.3390/info16110934.

[7] Z. Trabelsi, F. Alnajjar, M. M. A. Parambil, M. Gochoo, and L. Ali, “Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student’s Behavior

Recognition," *BDCC*, vol. 7, no. 1, p. 48, Mar. 2023, doi: 10.3390/bdcc7010048.

[8] X. Sheng, S. Li, and S. Chan, "Real-time classroom student behavior detection based on improved YOLOv8s," *Sci Rep*, vol. 15, no. 1, p. 14470, Apr. 2025, doi: 10.1038/s41598-025-99243-x.

[9] B. Prasetio and N. Pratiwi, "Deteksi Sampah Organik dan Anorganik Menggunakan Model YOLOv8," *jipi. jurnal. ilmiah. penelitian. dan. pembelajaran. informatika.*, vol. 10, no. 1, pp. 494–506, Jan. 2025, doi: 10.29100/jipi.v10i1.5965.

[10] Visen and N. Charibaldi, "Penerapan Object Detection Menggunakan Deep Learning Yolov8 Untuk Mengidentifikasi Sampah Anorganik (Maksimal Sepuluh Objek) Dalam Satu Citra," *JTIIK*, vol. 12, no. 1, pp. 195–202, Feb. 2025, doi: 10.25126/jtiik.20251219012.

[11] V. Hendriko and D. Hermanto, "Performance Comparison of YOLOv10, YOLOv11, and YOLOv12 Models on Human Detection Datasets," *Brilliance*, vol. 5, no. 1, pp. 440–450, Jul. 2025, doi: 10.47709/brilliance.v5i1.6447.

[12] S. Helmiyah, "Perancangan Sistem Deteksi Emosi Mahasiswa Pada Jam Perkuliahan Menggunakan Metode Yolo," *JATISI*, vol. 12, no. 1, Mar. 2025, doi: 10.35957/jatisi.v12i1.10195.

[13] M. Adiastoro, Febry Putra Rochim, and S. Hidayat, "PERFORMANCE EVALUATION OF RECENT YOLO VERSIONS FOR CLASSROOM STUDENT BEHAVIOR DETECTION," *jitk*, vol. 11, no. 4, pp. 1061–1074, May 2026, doi: 10.33480/jitk.v11i4.7773.

[14] Q. Jia and J. He, "Student Behavior Recognition in Classroom Based on Deep Learning," *Applied Sciences*, vol. 14, no. 17, p. 7981, Sep. 2024, doi: 10.3390/app14177981.

[15] Resa Pramudita, Mochamad Rizal Fauzan, Ilyasa Nafan Faza, Jaja Kustija, Ibnu Hartopo, and Muhammad Adli Rizqulloh, "Student Behavior Detection Using YOLOv10 for Classroom Engagement Analysis," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 15, no. 2, pp. 91–98, May 2026, doi: 10.22146/jnteti.v15i2.24611.

[16] C. Feng, Z. Luo, D. Kong, Y. Ding, and J. Liu, "IMRMB-Net: A lightweight student behavior recognition model for complex classroom scenarios," *PLoS ONE*, vol. 20, no. 3, p. e0318817, Mar. 2025, doi: 10.1371/journal.pone.0318817.

[17] H. Achmad, A. Pramudwiatmoko, M. Satrio Gumilang, B. Al Karim, and H. Wiyono, "Analisis Kinerja Model Deteksi Objek Yolo, Ssd, dan Faster R-Cnn pada Citra Penglihatan Malam untuk Pengenalan Tindak Kejahatan," *JTIIK*, vol. 12, no. 1, pp. 145–152, Feb. 2025, doi: 10.25126/jtiik.2025128409.

[18] D. S. Mahardhika and M. A. I. Pakereng, "Deteksi Objek Secara Real-Time Berbasis YOLOv8 dan Algoritma DeepSORT," *jmcs*, vol. 5, no. 1, pp. 671–686, Jan. 2026, doi: 10.37676/jmcs.v5i1.10038.

[19] F. Masykur, E. Kumalasari, A. Prasetyo, and M. Arifin, "A Resource-Efficient YOLO Framework via Discrete Wavelet Transform-Based Image Compression for Plant Disease Detection," vol. 31, no. 2, pp. 447–454, 2026.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.