Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15648
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparative Evaluation of YOLOv8 and YOLOv11 for Student Behavior Detection in Classroom CCTV Environments

Maya Sofhia
Universitas Prima Indonesia, Indonesia
mayasofhia@unprimdn.ac.id

Submitted :Dec 1, 2025 | **Accepted** : Dec 11, 2025 | **Published** : Jan 02, 2026

Abstract: Monitoring student behavior during classroom learning is important for supporting learning quality and teacher performance. This study presents a pilot comparison between YOLOv8 and YOLOv11 for detecting student classroom behaviors from CCTV images. Six elementary behaviors are consistently defined and used throughout the work: lookup, raise-hand, read, stand, turn-head, and write. The available SCB dataset contains 4,934 labeled images, but this study deliberately uses a front-facing subset of 100 images that best represent clear posture and behavior. After augmentation, the dataset grows to 220 images, split into 180 training, 30 validation, and 10 testing images. Both models are trained for 25 epochs on a T4 GPU with comparable configurations. At the detector level, YOLOv11 achieves higher mean average precision (mAP) of 42.9% compared to 28.9% for YOLOv8. At the behavior level, overall classification accuracy on the test set is 43.3% for YOLOv8 and 37.5% for YOLOv11. These results indicate a trade-off: YOLOv11 provides stronger bounding-box detection performance, while YOLOv8 produces slightly more stable behavior-level predictions on this very small and imbalanced dataset. The study emphasizes that these findings are exploratory baselines rather than definitive benchmarks, because the dataset is small and no statistical significance testing is performed. Future work must use a larger portion of the SCB dataset, more balanced class distributions, repeated experiments, and statistical analysis to obtain more robust conclusion.

Keywords: Student behavior, CCTV monitoring, YOLOv8, YOLOv11

INTRODUCTION

Student engagement is a key factor in determining learning success. It involves at least three dimensions: behavioral engagement(persistence and participation), emotional engagement(interest or boredom), and cognitive engagement(sustained attention). These components jointly influence how effectively students interact with learning materials and teaching strategies(Dewan et al., 2019; Trabelsi et al., 2023). In large classrooms, it is difficult for teachers to continuously observe all students, which can cause early signs of disengagement to be overlooked and decrease academic performance(Zaletelj & Košir, 2017).

In this context, it is important to distinguish related concepts: Human action recognition refers to the automatic understanding of physical actions (e.g., standing, raising a hand, writing) from video sequences(Pham et al., 2022; Sun et al., 2023); Student behavior monitoring in this study focuses on visible classroom actions, such as lookup, read, write, raise-hand, stand, and turn-head; Student engagement monitoring is broader and may integrate facial expressions, gaze, posture, and contextual cues to infer underlying engagement(Dewan et al., 2019; Yin Albert et al., 2022).

This work lies mainly in student behavior monitoring through action recognition at the level of discrete classroom behaviors, which in the long term can support engagement analysis. Previous studies have shown that computer-vision-based systems can be used to monitor student behaviors, attention, and participation without interrupting the learning process(Anh et al., 2019; Ling, 2022; Zhou et al., 2023). Using CCTV in classrooms allows continuous observation of student behavior, but traditional manual analysis is time-consuming, subjective, and less scalable(Rao, 2023; Wang et al., 2022). Deep learning–based methods, particularly You Only Look Once (YOLO) variants, have become popular for real-time object detection due to their balance between accuracy and speed(Wang et al., 2022; S. Q. Yang et al., 2022).

*maya sofhia

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15648
e-ISSN : 2541-2019
p-ISSN : 2541-044X

YOLOv8 has been applied to various tasks, including student behavior recognition, achieving high accuracy for specific actions such as jumping (Bisla et al., 2024) and obtaining significant gains when combined with multi-modal fusion(Li et al., 2024). YOLOv11 introduces architectural enhancements such as the C3k2 block, improving feature extraction and parameter efficiency for object detection, segmentation, and pose estimation(Alif, 2024; Khanam & Hussain, 2024). However, there is still limited work directly comparing YOLOv8 and YOLOv11 for student classroom behavior detection using a common dataset. This study therefore performs a controlled, small-scale comparative evaluation of YOLOv8 and YOLOv11 on a subset of the SCB dataset(F. Yang, 2025), focusing on front-facing student views.

The main contributions of this paper are: A consistent definition of six basic classroom behaviors lookup, raise-hand, read, stand, turn-head, and write and a corresponding annotation protocol based on CCTV frames; A pilot experimental framework that trains and evaluates YOLOv8 and YOLOv11 under comparable conditions on a curated subset of the SCB dataset; A balanced analysis that contrasts detector level metrics(mAP) with behavior-level accuracy, explicitly acknowledging the limitations arising from the very small dataset. Rather than claiming strong novelty at high-tier levels, this work positions itself as an exploratory comparative baseline that can be extended toward larger-scale, more rigorous studies in the future.

LITERATURE REVIEW

In education, human action recognition has been used to evaluate student attitudes, behavior, attention, and engagement during learning activities(Anh et al., 2019; Ling, 2022; Rao, 2023; Zhou et al., 2023). Classroom behavior monitoring systems can generate real-time or summary reports about the learning process, which can inform teacher interventions, classroom management, and instructional design (Anh et al., 2019; Chen et al., 2023). Recent works on student attention and behavior monitoring generally rely on computer vision and deep learning. Non-verbal signals such as body posture, head pose, and gaze direction play a critical role in estimating engagement and behaviors (Filipa et al., 2021; Ling, 2022; Yin Albert et al., 2022). Computer-vision-based approaches are attractive because they can be deployed passively using existing cameras and reduce the burden of manual observation (Hussain et al., 2018; Dewan et al., 2019).

Several studies have proposed classroom monitoring systems using variants of YOLO and related architectures. Wang et al., 2022improved YOLOv5 for multi-student learning behavior recognition, while S. Q. Yang et al., 2022 integrated attention mechanisms (CBAM) into YOLOv5 to detect in-class behaviors. Trabelsi et al., 2023 developed a real-time attention monitoring system, and Parkavi et al., 2024 proposed a deep learning based exam monitoring system to detect cheating behaviors.

YOLOv8 has been reported to achieve strong performance in various detection tasks, including student behavior monitoring and multi-modal fusion(Bisla et al., 2024; Li et al., 2024). YOLOv11 introduces further improvements in feature extraction and inference efficiency, supporting real-time applications at high frame rates (Alif, 2024; Khanam & Hussain, 2024). However, comparative studies between these two versions in education-specific scenarios are still rare. This study fills a small part of that gap by providing a direct comparison of YOLOv8 and YOLOv11 on the same student classroom behavior dataset, focusing on a front facing subset of SCB images. Given the limited dataset size, the results are interpreted as preliminary evidence rather than definitive benchmarks.

METHOD

Research Design

This research uses CCTV recordings to monitor student behaviors during class and to estimate behavioral engagement. The overall pipeline for developing and comparing YOLOv8 and YOLOv11 is summarized in Fig. 1:

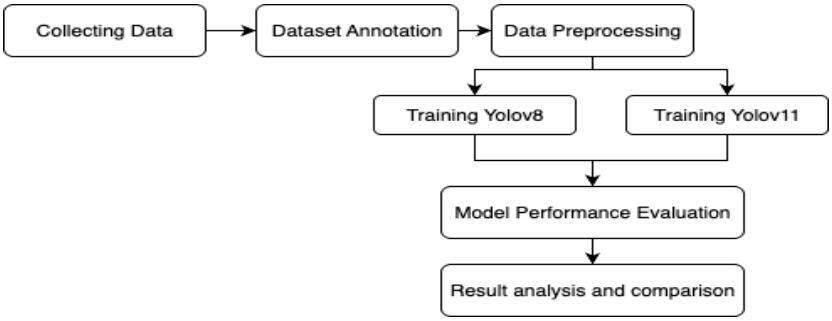

Fig. 1 Research design diagram for student classroom behavior detection using YOLOv8 and YOLOv11
The diagram is presented once to avoid structural duplication and represents the full experimental pipeline.

*maya sofhia

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15648
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Dataset and Subset Selection

The primary data source is the SCB-Dataset(F. Yang, 2025), which contains 4,934 labeled images of students and teachers in real classroom environments with various camera positions and lighting conditions. The dataset includes views from the front, back, front-left, front-right, back-left, and back-right, and covers classes from elementary to upper levels.

In this study, we consciously select a subset of 100 images that satisfy the following criteria: The camera faces the students (front-facing viewpoint); Student bodies are sufficiently visible to distinguish the six behaviors; The images are representative of realistic classroom settings but remain manageable for a pilot study.

This design choice is made explicit to avoid inconsistency between dataset availability(4,934 images) and dataset used(100 images). The remaining images are reserved for future work that will extend the experiments to a larger scale.

Behavior Classes and Annotation

To maintain consistency, we define six behavior classes:

1. Lookup: student looks forward or slightly upward toward the front of the classroom or teacher
2. Raise-hand: student raises one hand above shoulder level to ask or answer a question.
3. Read: student looks down at a book or paper for a sustained period.
4. Stand: student stands up from the seat, partially or fully.
5. Turn-head: student turns head sideways away from the front (left or right).
6. Write: student writes on paper, notebook, or desk, with observable pen/pencil motion.

These definitions are used consistently in the methods, results, tables, and discussion. Static frames are extracted from classroom videos(using OpenCV at a sampling rate of 1 frame per second). Each image is then annotated using Roboflow: Each student performing a target behavior is enclosed in a bounding box; The bounding box is labeled with one of the six behaviors above; Images with severe blur, strong occlusion, or ambiguous behavior are discarded to maintain label quality.

Data Preprocessing and Augmentation

Data preprocessing includes:

1. Cleaning: Removing blurry, noisy, duplicate, or mislabeled images.
2. Normalization: Scaling pixel values from [0,255] to [0,1].
3. Resizing: Matching the model input resolution. YOLOv8 uses 800×800 pixels, while YOLOv11 uses 640×640 pixels, following recommended defaults.

Data augmentation is used to enrich the small dataset: Random horizontal and vertical flips; Rotation between −45° and +45°; Blur up to 2.5 pixels and noise up to 1.96% of pixels. After augmentation, the dataset contains 220 labeled images, which are split into: 180 images for training (82%), 30 images for validation (14%), 10 images for testing (5%).

YOLOv8 Training Configuration

YOLOv8 is trained on Google Colab with a T4 GPU. The main configuration is: Epochs 25, Image size 800*800, Optimizer AdamW with initial learning rate 0.001 and momentum 0.9, Weight decay: 0.0005 for selected parameter groups, Losses monitored: box regression, objectness, and classification. The best-performing model checkpoint is saved as best.pt.

YOLOv11 Training Configuration

YOLOv11 is also trained on a T4 GPU under comparable conditions: Epochs 25, Image size 640*640, Optimizer AdamW with initial learning rate 0.001 and momentum 0.9, Weight decay: 0.0005 for selected parameter groups, Model initialization: pre-trained yolo11s.pt weights. The same dataset split and label definitions are used as for YOLOv8. Thus, both models are evaluated on exactly the same test images.

Evaluation Metrics

We evaluate model performance using: Mean Average Precision(mAP) at the bounding-box level; Per-class precision and recall computed from the confusion matrix; Macro-average precision, recall, and F1-score across the six behaviors; Overall classification accuracy at the behavior level.

Formally, for a given class c:

$$Precision_c = \frac{TP_c}{TP_c+FP_c} \quad (1)$$

$$Recall_c = \frac{TP_c}{TP_c+FN_c} \quad (2)$$

*maya sofhia

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15648
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Where TP_c, FP_c, FN_c are true positives, false positives, and false negatives for class c. Macro-average precision and recall over C classes are:

$$Precision_{macro} = \frac{1}{C}\sum_{c=1}^{C} Precision_c,\ Recall_{macro} = \frac{1}{C}\sum_{c=1}^{C} Recall_c \quad (3)$$

The overall behavior-level accuracy on the test set is:

$$mAP = \frac{1}{C}\sum_{c=1}^{C} AP_c \quad (4)$$

where AP_c is the area under the precision–recall curve for class c

Experimental Setup

The experimental setup for both models is summarized in Table 1

Table 1. Experimental setup for YOLOv8 and YOLOv11

Component	Yolov8	Yolov11
GPU	NVIDIA T4	NVIDIA T4
Epochs	25	25
Image Size	800*800	640*640
Optimizer	AdamW	AdamW
Initial LR	0.001	0.001
Momentum	0.9	0.9
Weight decay	0.0005	0.0005
Pretrained model	Yolov8s.pt	Yolo11s.pt
Dataset split	180/30/10 (train/val/test)	180/30/10 (train/val/test)

This table clarifies that both models are trained under comparable conditions, with the main architectural and input-resolution differences inherent to YOLOv8 and YOLOv11.

RESULT

Detector-Level Metrics

Model evaluation on the validation split yields the following detector-level metrics: YOLOv8**:** mAP = 28.9%**;** YOLOv11**:** mAP = 42.9%**.** At the bounding-box detection level, YOLOv11 achieves higher mAP than YOLOv8 on this dataset, indicating stronger generic detection capability. Training-time measurements also show that YOLOv11 completes 25 epochs approximately 39.6 seconds faster than YOLOv8 under the tested configuration, although its postprocessing time per image is slightly longer.

Confusion Matrices and Per-Class Metrics

The confusion matrices for YOLOv8 and YOLOv11(Fig. 2 and Fig. 3) summarize behavior-level predictions on the test set. From these confusion matrices, we compute precision, recall, and F1-score for each behavior class.

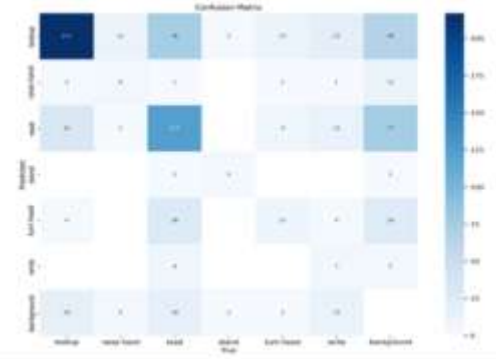

Fig. 2. Confusion matrix for student behavior detection with the YOLOv8 model.

*maya sofhia

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Issue 1, Januari 2026
DOI : https://doi.org/10.33395/sinkron.v10i1.15648
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Fig. 3. Confusion matrix for student behavior detection with the YOLOv11 model.

Per-class results are summarized in Table 2.

Table 2. Precision, Recall, F1 Score behavior with Yolov8 and YOLOv11

Behavior	Precision		Recall		F1 Score	
	Yolov8	YOLOv11	Yolov8	YOLOv11	Yolov8	YOLOv11
Lookup	78.6%	65.2%	53.19%	51.7%	15.8%	14.4%
Raise-hand	24%	8%	24%	22.2%	6%	2.9%
Read	47.86%	50.2%	47.67%	46.4%	11.9%	12.1%
Stand	50%	50%	62.5%	20.8%	13.9%	7.3%
Turn-head	30.6%	30.6%	17.74%	10.2%	5.6%	3.8%
Write	11.3%	18%	33.3%	44.7%	4.2%	6.4%

Macro-Average Performance and Accuracy

From the per-class metrics in Table 2, macro-average values are computed as simple averages across the six behaviors and summarized in Table 3.

Table 3. Macro-average precision, recall, and F1-score

Metric	Yolov8	YOLOv11
Precision	40.4%	37%
Recall	39.7%	32.7%
F1-Score	9.6%	7.8%

Behavior-level classification accuracy on the test set is: YOLOv8 43.3% and YOLOv11: 37.5%

Training Curves

The training curves (loss and mAP versus epoch) for YOLOv8 and YOLOv11 are shown in Fig. 4 and Fig. 5.

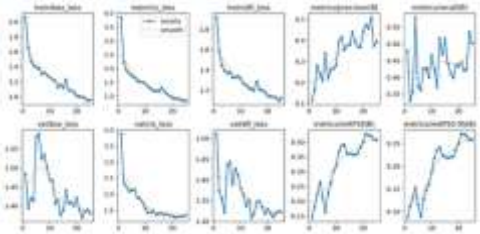

Fig. 4. Training curves (loss and mAP) for YOLOv8 model.

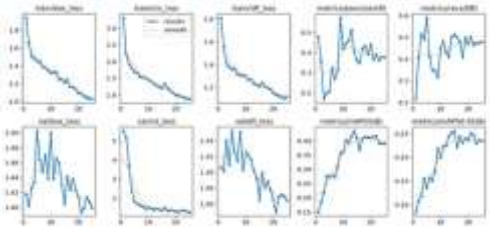

Fig. 5. Training curves (loss and mAP) for YOLOv11 model.

Qualitative Prediction Results

Qualitative examples of detection results show that: Both models reliably detect lookup as the most frequent behavior; Behaviors such as write and turn-head are more challenging, with lower confidence scores and more misclassifications; YOLOv11 tends to produce more bounding boxes and overlaps, sometimes with lower

*maya sofhia

confidence, while YOLOv8 provides fewer but slightly more stable detections. These qualitative observations support the quantitative metrics without contradicting them.

DISCUSSIONS

Consistency Between Metrics and Conclusions

This study clearly distinguishes between: Detector-level metrics(mAP), where YOLOv11 achieves higher performance; Behavior-level metrics (macro-average precision, recall, F1-score, and accuracy), where YOLOv8 performs slightly better on this small test set. YOLOv11's higher mAP suggests that it is more capable of correctly localizing objects and assigning labels at the detection level. However, when predictions are aggregated at the behavior class level, YOLOv8 yields higher macro-average precision and recall and higher accuracy. On such a small dataset, these differences should be interpreted cautiously and not as statistically significant improvements.

Thus, rather than making a one-sided claim that YOLOv8 is strictly better than YOLOv11 or vice versa, we interpret the results as indicating a trade-off: YOLOv11 shows stronger generic detection capability (higher mAP); YOLOv8 provides slightly more stable behavior-level predictions and accuracy under the specific conditions and data distribution of this pilot experiment. This interpretation replaces previous inconsistent statements and aligns the narrative with the reported tables.

Methodological Limitations and Validity

A critical limitation is the very small dataset used for training and testing: Only 100 images are selected from the 4,934 available SCB images; After augmentation, there are 220 images, with only 10 images in the test set. This scale is insufficient for high-tier experimental standards (e.g., SINTA 1 or Q2 journals) and does not allow for robust statistical analysis or strong claims of generalization. The present work is therefore framed as a pilot study or comparative baseline.

Several steps are taken to strengthen methodological clarity: The subset selection criteria are stated clearly; The pipeline is described once and consistently, avoiding duplicated diagrams; Training parameters for YOLOv8 and YOLOv11 are described coherently; similarities and differences (such as input resolution) are explicitly mentioned; Augmentation is reported as part of the method, while its effects are interpreted cautiously. Even with these corrections, the study remains methodologically limited and must be extended in future work.

Behavior Categories and Terminology

The definitions of the six behavior classes are now formally given in the Method section and used uniformly in all sections, tables, and figures. This resolves earlier inconsistencies where labels such as "Focused" or "Unfocused" appeared without clear definition.

Terminology is also tightened: Student behavior monitoring is used consistently to refer to the task in this study; Action recognition describes the underlying computer vision problem; Engagement monitoring" is mentioned only when referring to broader literature and is not used interchangeably with behavior detection. This avoids conceptual confusion and improves logical coherence.

Implications and Future Work

Despite the limited dataset, several practical implications can be drawn: Both YOLOv8 and YOLOv11 can serve as starting points for classroom behavior monitoring systems based on CCTV, especially for detecting frequent behaviors such as lookup, read, and write; The dominance of the lookup behavior in predictions suggests that students spend much of their time facing forward, which may correlate with behavioral engagement.

For future research, the following steps are necessary: Use a larger portion of the SCB dataset and ensure more balanced class distributions; Perform multiple training runs with different random splits and hyperparameters to estimate variance; Include statistical significance tests (confidence intervals, hypothesis testing) to support claims of performance differences; Combine behavior detection with other modalities, such as facial expressions and gaze, to move from behavior monitoring toward deeper engagement analysis; Standardize reference formatting fully according to the journal's style and ensure all in-text citations match the reference list.

CONCLUSION

This paper has presented a comparative evaluation of YOLOv8 and YOLOv11 for student classroom behavior detection from CCTV images using a curated subset of the SCB dataset. Six behaviors: lookup, raise-hand, read, stand, turn-head, and write are defined consistently and used for annotation, training, and evaluation.

The main findings are: At the detector level, YOLOv11 achieves higher mAP than YOLOv8 (42.9% vs 28.9%); At the behavior level, YOLOv8 provides slightly higher overall accuracy (43.3% vs 37.5%) and better macro-average metrics. Training curves show that YOLOv8 learns more steadily across epochs, while YOLOv11 exhibits larger fluctuations but reaches higher mAP and faster total training time. These results indicate that no single model is universally superior in this setting; instead, there is a trade-off between detector-level performance and

*maya sofhia

behavior-level accuracy, heavily influenced by dataset size and composition. Because the study is based on a very small subset of images, the conclusions are exploratory and cannot be considered statistically robust.

Future work must expand the dataset, refine the experimental design, and incorporate rigorous statistical analysis to produce stronger evidence and integrate behavior detection into more comprehensive student engagement monitoring systems.

REFERENCES

Alif, M. A. R. (2024). YOLOv11 for Vehicle Detection: Advancements, Performance, and Applications in Intelligent Transportation Systems. http://arxiv.org/abs/2410.22898

Anh, B. N., Son, N. T., Lam, P. T., Chi, L. P., Tuan, N. H., Dat, N. C., Trung, N. H., Aftab, M. U., & Dinh, T. Van. (2019). A Computer-vision based application for student behavior monitoring in classroom. *Applied Sciences (Switzerland)*, *9*(22). https://doi.org/10.3390/app9224729

Bisla, T., Shukla, R., Dhawan, M., Islam, M. R., & Horio, K. (2024). Jumping behavior Analysis after Identification of Daycare Children's Activities utilizing YOLOv8 Algorithm. *Proceedings of CONECCT 2024 - 10th IEEE International Conference on Electronics, Computing and Communication Technologies*. https://doi.org/10.1109/CONECCT62155.2024.10677280

Chen, Y., Li, W., Weng, H., Zheng, J., Qian, Z., Huang, J., Lun, J., Chen, Q., Zhang, Q., & Wu, Y. (2023). Research on Intelligent Teaching Mode Based on Classroom Behavior Monitoring of Body Motion. *Journal of Modern Educational Research*, *2*, 15. https://doi.org/10.53964/jmer.2023015

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: a review. *Smart Learning Environments*, *6*(1). https://doi.org/10.1186/s40561-018-0080-z

Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements*. http://arxiv.org/abs/2410.17725

Li, P., Bo, S., Zhang, H., & Bin, X. (2024). Student classroom behavior detection based on YOLOv8 multi-modal fusion. *2024 7th International Conference on Computer Information Science and Application Technology, CISAT 2024*, 931–935. https://doi.org/10.1109/CISAT62382.2024.10695405

Ling, W. (2022). Automatic Recognition of Students' Classroom Behavior Based on Computer Vision. *Academic Journal of Computing & Information Science*, *5*(2). https://doi.org/10.25236/ajcis.2022.050205

Parkavi, A., Pushpalatha, M. N., & Alex, S. A. (2024). Deep Learning Classroom Management System for Exam Monitoring and Student Behavior Detection. *8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings*, 2004–2009. https://doi.org/10.1109/I-SMAC61858.2024.10714864

Pham, H. H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2022). Video-based Human Action Recognition using Deep Learning: A Review. http://arxiv.org/abs/2208.03775

Rao, M. V. P. (2023). Student Behaviour Monitoring System. *International Journal for Research in Applied Science and Engineering Technology*, *11*(5), 779–787. https://doi.org/10.22214/ijraset.2023.51458

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2023). Human Action Recognition From Various Data Modalities: A Review. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (Vol. 45, Issue 3, pp. 3200–3225). IEEE Computer Society. https://doi.org/10.1109/TPAMI.2022.3183112

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. *Big Data and Cognitive Computing*, *7*(1). https://doi.org/10.3390/bdcc7010048

Wang, Z., Yao, J., Zeng, C., Wu, W., Xu, H., & Yang, Y. (2022). YOLOv5 Enhanced Learning Behavior Recognition and Analysis in Smart Classroom with Multiple Students. *IEIR 2022 - IEEE International Conference on Intelligent Education and Intelligent Research*, 23–29. https://doi.org/10.1109/IEIR56323.2022.10050042

Yang, F. (2025). SCB-Dataset: A Dataset for Detecting Student and Teacher Classroom Behavior. http://arxiv.org/abs/2304.02488

Yang, S. Q., Chen, Y. H., Zhang, Z. Y., & Chen, J. H. (2022). Student in-class behaviors detection and analysis system based on CBAM-YOLOv5. *2022 7th International Conference on Intelligent Computing and Signal Processing, ICSP 2022*, 440–443. https://doi.org/10.1109/ICSP54964.2022.9778630

Yin Albert, C. C., Sun, Y., Li, G., Peng, J., Ran, F., Wang, Z., & Zhou, J. (2022). Identifying and Monitoring Students' Classroom Learning Behavior Based on Multisource Information. *Mobile Information Systems*, *2022*. https://doi.org/10.1155/2022/9903342

Zaletelj, J., & Košir, A. (2017). Predicting students' attention in the classroom from Kinect facial and body features. *Eurasip Journal on Image and Video Processing*, *2017*(1). https://doi.org/10.1186/s13640-017-0228-8

Zhou, H., Jiang, F., Si, J., Xiong, L., & Lu, H. (2023). Stuart: Individualized Classroom Observation Of Students With Automatic Behavior Recognition And Tracking. https://github.com/hnuzhy/StuArt.

*maya sofhia