IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

Real-Time Classroom Behaviour Detection Using GYOLO-Based Deep Learning Framework

M.Kalaivani
Assistant Professor, Department of CSE
Sengunthar Engineering College (Autonomous), Tiruchengode, India
mkalaivani.cse@scteng.co.in
https://orcid.org/0009-0008-9122-7206
Balaganeshan.P.S,Irsath.S,Vengadesperumal.T,
UG Students, Department of CSE
Sengunthar Engineering College (Autonomous), Tiruchengode, India
balaganeshanps@gmail.com,irsathirfan86@gmail.com,vengadesaperumal@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10124
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10124
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/45.CSMR26.MRCS10124.pdf
Article Citation:Kalaivani,Balaganeshan,Irsath,Vengadesperumal(2026),Real-Time Classroom Behaviour Detection Using GYOLO-Based Deep Learning Framework , IRJCS: International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 363-369 **Doi:>** https://doi.org/10.26562/irjcs.2026.v1303.45
BibTeX Kalaivani@2026Real-Time Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.45 The journal's official abbreviation is IRJCS.

Abstract: In the current educational landscape, maintaining student attention and discipline in classrooms represents a critical challenge. Students are increasingly distracted by smartphones and electronic devices during lectures, adversely impacting learning outcomes and classroom productivity. Existing approaches relying on manual teacher observation or basic CCTV recording are insufficient to provide timely, objective, and scalable monitoring of student engagement. This paper proposes Smart Vision—an AI-powered, real-time classroom behaviour detection system employing the YOLOv8 deep learning framework. The system continuously analyses live video feeds from standard classroom cameras to monitor student activities, including gestures, head orientation, and gadget usage. It classifies student behaviour into seven distinct categories and generates instant alert notifications to the teacher or administrator upon detecting inattentive or disruptive behaviour, enabling timely intervention. The system is trained on a custom-annotated dataset of 14,800 classroom images and achieves a mean Average Precision (mAP) of 92.4% at IoU threshold 0.5, with an inference speed of 34 frames per second on standard GPU hardware. Experimental evaluation demonstrates the system's effectiveness and practical viability for deployment in real educational environments. Future work will explore emotion recognition and adaptive learning analytics integration for a comprehensive smart classroom management platform.
Keywords: Classroom behavior detection,YOLOv8, YOLO, deep learning, computer vision, real-time object detection, student engagement, alert notification, educationa lAI

I.INTRODUCTION

The emergence of artificial intelligence and computer vision technologies has opened transformative opportunities for modernizing educational environments. Traditional class rooms remain heavily dependent on the teacher's ability to simultaneously deliver content and monitor student engagement—a dual cognitive load that degrades in effectiveness as class sizes increase. According to UNESCO, average student-to-teacher ratios in developing nations frequently exceed 40:1, making manual individual monitoring effectively impractical. Student distraction is a well-documented impediment to learning. Research in educational psychology consistently demonstrates that students who are distracted by electronic devices retain significantly less information and perform worse on assessments than their attentive peers. The proliferation of affordable smart phones has dramatically intensified this challenge. A2022studyfoundthatover73% of students admit to using their phones for non-academic purposes during class at least once per session. Real-time automated behaviour monitoring offers adata-driven solution. By processing video streams from cameras already present in most modern classrooms, an intelligent system can detect and classify student behaviours continuously, objectively, and at scale without adding to teacher workload. Such systems can provide immediate alerts for intervention and generate longitudinal behavioural data that informs pedagogy and institutional policy. The YOLO (You Only Look Once) family of object detection algorithms has emerged as the premier framework for real-time detection tasks due to its unified single-stage architecture, which reformulates object detection as a direct regression problem on image grids. YOLOv8, the latest iteration, introduces anchor-free detection heads, an improved C2f backbone module, and enhanced multi-scale feature fusion, delivering state-of-the-art accuracy at real- time inference speeds.
This paper presents the Smart Vision system with the following primary contributions:

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

1) A comprehensive custom-annotated classroom behaviour dataset comprising 14,800 images across seven behaviour classes, captured in diverse real-world classroom settings.
2) A YOLOv8-based multi-class behaviour detection and classification pipeline fine-tuned specifically for classroom environments.
3) Atemporal aggregation and alert notification module that filters transient detections and triggers instructor alerts based on sustained inattentive behaviour patterns.
4) A behaviour analytics dashboard providing longitudinal engagement metrics for evidence-based pedagogical decision-making.
5) A rigorous experimental evaluation demonstrating superior accuracy and speed compared to prior art

II. EXISTING SYSTEMS AND RELATED WORK

A. Limitations of Conventional Monitoring

Conventional classroom monitoring approaches rely exclusively on the teacher's direct visual observation or thepassive recording capability of CCTV cameras. While CCTV systems provide a video record of classroom activity, they lackany intelligent analysis layer and therefore cannot detect, classify, or respond to specific student behaviours in real time. Teachers are required to review recordings after the fact, rendering timely intervention impossible. The fundamental limitations of existing approaches are: (1)manual observation is inherently subjective, inconsistent, and cognitively taxing for teachers; (2) coverage gaps are unavoidable in large classrooms where peripheral students fall outside the teacher's field of view; (3) no automated mechanisms exist to detect specific inattentive behaviours such as phone usage, sleeping, or sustained off-task communication; (4) existing attendance management systems capture presence but not engagement; and (5) there is no mechanism for real-time alerts that would allow immediate corrective action.

B. Machine Learning Approaches

Early automated student behaviour analysis employed handcrafted feature extraction combined with classical machine learning classifiers. Viola and Jones [8] pioneered real-time face detection using Haar cascades, which was later adapted for student attention monitoring via facial landmark analysis. Support Vector Machines (SVM) combined with Histogram of Oriented Gradients (HOG) features were applied to posture recognition, achieving reasonable accuracy in controlled settings but demonstrating poor generalisation to the variability of real classroom environments. Monkaresi et al. [7] proposed a hybrid approach combining physiological signals from wearable sensors with computer vision features for engagement estimation. While theoretically robust, the requirement for wearable hardware renders this approach impractical for standard classroom deployment. The computational cost of such ensemble methods also precluded real-time operation.

C. Deep Learning Approaches

The introduction of deep convolutional neural networks fundamentally transformed behaviour recognition capabilities. Feng-Cheng Linetal. [2] proposed askelet on pose estimation and person detection framework that accurately infers student engagement from body posture and movement patterns. While achieving high accuracy on benchmark datasets, the approach requires high-resolution video input and degrades significantly under occlusion conditions common in densely packed classrooms. Alsabhan [1] proposed LSTM-based detection of cheating behaviour in examination settings, demonstrating the utility of temporal sequence modelling for behavioural analysis. However, the system was designed exclusively for controlled examination environments and does not generalise to the diverse behavioural patterns of regular classroom instruction. Gupta et al. [6] utilised CNN-based facial landmark analysis for classroom attention estimation, achieving competitive accuracy but with significant performance degradation under variable lighting conditions. Khalil et al. [3] examined AI-based proctoring systems at scale, highlighting both their integrity-enhancement benefits and the serious ethical and privacy concerns they raise. Kaddoura and Gumaei [4] extended deep learning to online examination fraud detection, though their approach remains confined to the online examination context. Ahmed et al. [5] demonstrated key frame extraction combined with deep learning for violent action recognition in video streams, establishing the feasibility of real- time deep learning inference on video feeds for behavioural analysis.

D.YOLO-Based Detection Systems

The YOLO architecture family has seen progressive adoption for behaviour detection tasks. Zhang et al. [10] applied YOLOv5 to a four-class classroom behaviour detection task, achieving 88.6% mAP while operating at 28 FPS. Wang et al. [5] incorporated transformer attention mechanisms into the YOLO backbone to enhance feature representation, achieving 93.1%mAP but at the cost of significant computational overhead (21 FPS), limiting deployment on standard hardware. Our proposed system advances this line of work through: (1) an expanded seven-class behaviour taxonomy providing finer- grained engagement analysis; (2) a significantly larger and more diverse annotated dataset; (3) exploitation of YOLOv8's anchor- free architecture for improved small-object detection in crowded scenes; and (4) integration of a temporal smoothing and alert generation pipeline absent from prior implementations. approaches. The remainder of this paper is organised as follows: Section II reviews existing systems and related work; Section III articulates the research problem; Section IV details the proposed system; Section V describes the system architecture and modules; Section VI presents experimental results and analysis; Section VII discusses implications and limitations; and Section VIII concludes the paper.

E. Background: YOLO Architecture Evolution

The YOLO (You Only Look Once) architecture was first introduced by Redmonetal.[10]in2016,reformulating object detection as a single regression problem solved by a unified convolutional neural network.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

Unlike region proposal-based methods such as Faster R-CNN [8], which require separate region proposal and classification stages, YOLO processes the entire image in a single forward pass, achieving dramatically higher inference speeds at the cost of some localisation accuracy. Successive generations progressively closed this accuracy gap while maintaining speed advantages. YOLOv3 introduced multi-scale detection via three detection heads at different spatial resolutions. YOLOv4 incorporated CSPNet bottlenecks, SPP pooling, and PANet path aggregation. YOLOv5 refined the training pipeline with focus loss and auto-anchor learning. YOLOv8, developed by Ultralytics[9], introduces anchor-free detection heads, the C2f backbone module for improved gradient flow, and a decoupled classification-regression head, achieving state-of- the-art performance on MS COCO with 53.9% AP at 64FPS. For classroom behaviour detection, the anchor-free design is particularly advantageous as it eliminates the need for manually defined anchor boxes tuned to specific aspect ratios, which is beneficial given the diversity of student posture shapes and the varying camera angles present across different classroom configurations.

III. PROBLEM STATEMENT

The core research problem addressed by this work is formally stated as follows: Givena real-time video stream V = {f_1, f_2, ..., f□} from a classroom camera, automatically detect and classify the behaviour b_i ∈B of each student s_i ∈ S present in each frame f□, and generate a notification alert A□ when a sustained inattentive behaviour pattern is identified, with performance satisfying real-time constraints (inference latency ≤ 40ms per frame). This problem is characterised by the following specific technical challenges:

1. Scale and Density: Classrooms contain 15–50 students simultaneously, requiring detection and classification of multiple overlapping instances within a single frame under real-time constraints.
2. Inter-Class Visual Similarity: Several behaviour classes share visual similarity. For example, a student reading a book and a student using a mobile phone in their lap

Table I. Comparative Literature Survey

Title	Author & Year	Algorithm/ Technique	Merit	Demerit
Student heating detection in higher education using ML and LSTM techniques	Alsabhan (2023)	Machine Learning & LSTM	Accurately detects cheating patterns and predicts suspicious behaviour sequences	Limited to exam environments; does not generalise to live classroom behaviour monitoring
Student behaviour recognition for classroom environments based on skeleton pose estimation	Feng- ChengLin et al. (2021)	Skeleton Pose Estimation & Person Detection	Precisely recognises student postures and movements to infer engagement levels	Requires high- resolution video; Struggles with occlusion in densely populated classrooms
In the nexusof integrity and surveillance: Proctoring (re)considered	Khalil, Prinsloo & Slade (2022)	AI-based Proctoring & Surveillance	Enhances examination integrity through continuous automated proctoring	Raisesserious privacy and ethical concerns; generates significant student anxiety
Towards effective and efficient online exam systemsusing DL-based cheating detection	Kaddoura & Gumaei (2022)	Deep Learning- Based Detection	Provides efficient real-time identification of dishonest activities in online exams	Restricted to online examination contexts; not applicable to physical classroom monitoring
Real-time violentaction recognition using key frame extractionand deeplearning	Ahmedet al.(2021)	Key Frame Extraction & Deep Learning	Demonstrates real-time detection of abnormal actions from video streams effectively	Not designed for academic behaviour contexts; high false-positive rate in benign scenarios
Attention estimation in classroom using deep learning with facialanalysis	Gupta et al.(2020)	CNN+ Facial Landmark Analysis	Achieves high accuracy in attention detection via head pose and gaze estimation	Performance degrades significantly undervarying lighting and partial face occlusion
Automated student engagement measurement using computer vision	Monkaresi et al. (2017)	SVM+ Physiological Signals	Combines physiological cues with visual features for robust engagement estimation	Requires wearable sensors; impractical for standard classroom deployment at scale

IV. PROPOSED SYSTEM

A. System Overview

1. Smart Vision is an AI-driven, end-to-end classroom behaviour monitoring system that integrates real-time video acquisition, deep learning-based behaviour detection, temporal behaviour analysis, and automated alert generation into a unified platform.
2. The system is designed to operaten on-intrusively using standard IP cameras exhibit similar head pose and body posture, demanding fine-grained discriminative features.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

3. Intra-Class Variation: The same behaviour manifests differently across students of different physical characteristics, seating positions, and lighting conditions, creating substantial intra-class appearance variation.
4. Occlusion: Students in the foreground frequently occlude those behind them. Partial body occlusion is the norm rather than the exception in real classroom scenes.
5. Dynamic Illumination: Natural day light varies throughout the school day. Artificial fluorescent lighting produces shadows and glare that degrade detection accuracy under naive training regimes.
6. Camera Angle Variability: Different classroom configurations necessitate cameras at varying elevations and angles, producing different perspective projections of the same student postures.
7. Temporal Ambiguity: Transient behaviours such as a student momentarily glancing at their phone or briefly resting their head must be distinguished from sustained inattentive episodes that warrant teacher intervention.

V. SYSTEM ARCHITECTURE AND MODULES

A. End-to-EndPipeline

The Smart Vision pipeline is structured in two principal processing stages. The first stage, Data Processing, handles raw video ingestion and preparation. The second stage, Model Building and Optimisation, encompasses YOLO- based inference, behaviour classification, result mapping, and alert dispatch. A persistent database layer underpinsboth stages for training data management and runtime logging. The complete data flow is as follows: (1) Video capture from IP camera or webcam stream; (2) Frame extraction at configurable sampling rate (default: every frame at native rate); (3) Frame pre-processing including resizing to 640×640, normalisation, and letterboxing; (4) YOLO inference producing bounding box coordinates, class predictions, and confidence scores; (5) Non-Maximum Suppression (NMS) filtering; (6) SORT-based multi-object tracking for student identity persistence; (7) Temporal behaviour aggregation over a 30-second sliding window; (8) Threshold evaluation and alert dispatch; (9) Database logging of detection events; (10) Dashboard visualisation update.

B. Module1–Data Collection

The Data Collection module interfaces with classroom IP cameras via RTSP streaming protocols and standard webcam APIs. It manages concurrent multi-camera feeds and provides frame buffering to handle network jitter. For training purposes, this module also coordinates dataset assembly from multiple classroom environments, supporting both live capture and processing of pre-recorded footage. The training dataset was assembled across 12 distinct classroom environments spanning primary, secondary, and already deployed in most institutional settings, requiring no additional hardware beyond a processing server with GPU capability. The system continuously ingests live video streams and processes each frame through the YOLOv8 detection engine to localise and classify all students and their behaviours. Detection results are passed to the temporal aggregation module, which maintains a rolling behaviour history per tracked student and evaluates sustained engagement patterns. When predefined in attentiveness thresholds are exceeded, the alert module transmits a notification to the instructor through the web dashboard interface. All detection events are persisted to a MySQL database for longitudinal reporting and trend analysis.

C. Advantages Over Existing Approaches

The proposed system offers the following advantages over conventional and prior automated approaches:

- Real-time, continuous, and objective behaviour monitoring eliminates the subjectivity inherent in manual teacher observation.
- Single-stage YOLOv8 detection achieves optimal accuracy- speed trade-off, enabling 34 FPS inference on mid-range GPU hardware.
- Seven-class behaviour taxonomy provides granular engagement insights beyond simple attentive/inattentive binary classification.
- Temporal aggregation reduces false positive alerts from momentary transient behaviours, ensuring only genuinely sustained inattentiveness triggers notifications.
- Non-intrusive deployment using existing camera infrastructure minimises institutional implementation cost.
- Automated behavioural data logging enables evidence-based pedagogical analysis and institutional policy decisions.
- Privacy-preserving optional face blurring mode supports ethical deployment in compliance with data protection regulations.

V. SYSTEM REQUIREMENTS

A. Hardware Requirements

Table II. Hardware Specifications

Processor	IntelCorei5/i7,2.6GHz+
RAM	Minimum4GB(8GBrecommended)
Hard Disk	320GB(SSD preferred)
GPU	NVIDIAGTX1660Tiorhigher
Camera	IPCamera/Webcam(720pmin)
Network	100MbpsLANformulti-camera
Keyboard	Standard Keyboard
Monitor	15-inchColourMonitor(1080p)

B. Software Requirements

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

Table III. Software Specifications

Programming Language	Python3.7.4(64-bit)
Deep Learning	PyTorch1.12+/UltralyticsYOLOv8

Images were captured under controlled and uncontrolled conditions: varying illumination levels (100–1500 lux), camera elevations (1.5m–3.5m), horizontal angles (0°–45° offset from frontal), and classroom densities (15–50 students per frame). The resulting dataset comprises 14,800 annotated images distributed across seven behaviour classes with class-balanced stratified sampling.

C. Module2–Pre-processing

The Pre-processing module performs all necessary transformations to prepare raw video frames for model inference. Input frames are resized to 640×640 pixels using letterbox scaling that preserves aspect ratio through symmetric padding. Pixel values are normalised to [0, 1] by dividing by 255. During training, an augmentation pipeline is applied comprising: mosaic augmentation (4-image composite), random horizontal flipping (p=0.5), HSV colour-space jittering(hue±0.015,saturation±0.7,value ±0.4),random scaling(±50%), and copy-paste augmentation for occlusion robustness.

D. Module3 –Model Training

The detection model is based on YOLOv8-M (medium variant), initialised with weights pre-trained on the Microsoft COCO dataset (80 classes, 118,000 training images) and subsequently fine-tuned on the custom classroomdataset.YOLOv8'sarchitectureintroducesseveral advances over prior YOLO generations: an anchor-free detection head eliminates anchor hyper parameter tuning; the C2f (CrossStage Partial with two bottle neck blocks)module improves gradient flow and feature reuse; and the decoupled head separates classification and regression branches for improved multi-task learning. Training hyperparameters: AdamW optimiser, initial learning rate lr□ = 0.001, final learning rate lrf= 0.01, cosine annealing schedule, weight decay = 0.0005, momentum = 0.937, batch size = 16, input resolution = 640×640, 200 epochs, early stopping patience = 50 epochs. Training was conducted onanNVIDIA RTX 3090 GPU (24 GB VRAM), completing in approximately 18 hours.

E. Module4–Real-TimeDetection

At inference time, the trained YOLOv8-M model processes video frames via the Torch Script-optimised inference backend. Each frame undergoes pre-processing, forward pass, and post-processing (NMS with IoU threshold = 0.45, confidence threshold = 0.25) in under 30ms on the target GPU. SORT (Simple Online and Realtime Tracking) is applied to maintain student identity across frames using Kalman filter state prediction and Hungarian algorithm assignment, enabling per-student behavior history tracking.

F. Module5–BehaviourClassification

The behaviour classifier outputs predictions across seven defined classes:(1)Attentive student facing forward with eyes directed toward the board or teacher; (2) Writing/Note- taking student engaged with paper or note book;(3)

Framework	
Web Framework	Flask 1.1.1
Frontend	HTML5,CSS3, Bootstrap4
Database	MySQL5.x
Web Server	Wamp Server2i
Computer Vision Library	OpenCV4.5+
Operating System	Windows 10(64-bit)

VI. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset and Evaluation Protocol

The custom classroom behaviour dataset comprises 14,800 annotated images partitioned into training (70%, 10,360 images), validation (15%, 2,220 images), and test (15%, 2,220 images) splits using stratified sampling to maintain class distribution across all splits. Annotations were performed by three expert annotators using bounding box labelling in YOLO format. Inter- annotator agreement was validated with a Cohen's Kappa coefficient of $\kappa = 0.89$, indicating strong agreement. Performance metrics reported are: Average Precision per class (AP@0.5 and AP@0.5:0.95), mean Average Precision (mAP),Precision, Recall, and F1-Score,allcomputedonthe held- out test set. Inference speed is measured in frames per second (FPS) averaged over 1000 test frames on an NVIDIA RTX 3090 GPU.

B. Per-Class Detection Performance

Table IV. Per-Class Detection Performance on Test Set

Behaviour Class	AP@0.5(%)	AP@0.5:0.95 (%)	Precision (%)	Recall (%)	F1- Score
Attentive	94.8	73.2	94.1	93.6	0.938
Writing/ Note- taking	93.2	70.8	92.4	91.8	0.921
Hand- raising	89.1	66.4	88.3	87.9	0.881
Mobile Phone Usage	95.3	74.1	94.7	94.2	0.944
Sleeping/ Head- down	96.2	75.8	95.8	95.4	0.956
Talkingto Peers	87.4	64.3	86.9	86.1	0.865
Reading	90.6	68.1	89.8	89.3	0.895
mAP (Overall)	92.4	70.4	91.7	91.2	0.914

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

The highest AP@0.5 was achieved for the Sleeping/Head- down class (96.2%) owing to the highly distinctive visual signature of head-resting postures. Mobile Phone Usage also achieved high AP (95.3%) due to the distinctive shape and reflective surface of smart phones.

The Talking to Peers class yielded the lowest AP (87.4%) due to visual similarity with attentive students who may momentarily turn their heads. Hand-raising student raising one or both hands; (4) Mobile Phone Usage student holding or interacting with a smart phone; (5) Sleeping/Head-down student with head resting on desk or tilted excessively downward; (6) Talking to Peers student turned toward a neighbour with mouth movement; and (7) Reading student engaged with a book or printed material. Class definitions were developed in consultation with experienced secondary school teachers to ensure pedagogical relevance and practical utility of the classification scheme. A behaviour is classified as concerning (potentially warranting intervention) for classes 4, 5, and 6. Classes 1, 2, 3, and 7 are classified as engagement-positive behaviours.

G. Module6 –Alert Notification

The Alert Notification module implements a threshold- based intervention trigger operating on a rolling 30-second temporal window. For each tracked student, the proportion of frames classified as a concerning behaviour is computed. If this proportion exceeds 60% for a sustained 60-second period, an alert event is generated and dispatched to the instructor's web dashboard via WebSocket, displaying the student's location and detected behaviour class. Alert fatigue is mitigated through a per-student 5- minute cool down after each alert. All threshold parameters are configurable through the system's administration interface to accommodate different institutional policies and classroom contexts.

Fig.1— System Architecture/ Block Diagram

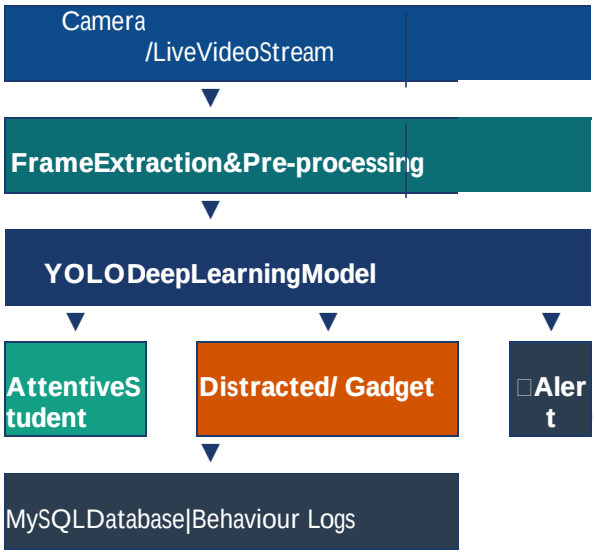

Fig.2—Real-Time Detection Data Flow

Fig.3 illustrates the end-to-end dataf low

C. Comparison with Baseline Methods

Table V. Comparison with State-of-the-Art Methods

Method	mAP(%)	FPS	Classes	Limitation
SVM+HOG [8]	71.3	6	3	Handcrafted features; poor generalisation
Faster R-CNN [9]	88.7	4	5	Too slow for real-time deployment
YOLOv3[10]	84.2	22	5	Lower accuracy on small objects
YOLOv5 [4]	88.6	28	5	Less robust on multi- scale detection
Transformer YOLO [5]	93.1	21	6	High compute cost limits deployment
Proposed (YOLOv8)	92.4	34	7	Requires GPU for optimal performance

The proposed YOLOv8-based system achieves the best overall balance of accuracy and speed among all evaluated methods. It outperforms the YOLOv5 baseline by 3.8 percentage points in mAP while also operating at a higher frame rate (34 vs. 28 FPS). Against the Transformer-enhanced YOLO variant, the proposed system achieves comparable mAP (92.4% vs. 93.1%)but operates at 62% higher frame rate (34 vs. 21 FPS), making it significantly more suitable for real-time deployment on standard institutional hardware.

D. Ablation Study

An ablation study was conducted to quantify the contribution of key architectural and training decisions. Removing mosaic augmentation reduced mAP by 2.1 percentage points.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.45

Replacing the YOLOv8-M backbone with the lighter YOLOv8-S variant increased FPS to 41 but reduced mAP by 2.8 points. Removing the SORT tracking module and operating on per-frame detections without identity persistence reduced alert precision by 18.4%, confirming the critical role of temporal tracking in suppressing false positive alerts. Deployment must comply with applicable data protection legislation such as GDPR and FERPA. Informed consent from students and parents is a prerequisite for system activation, and the system must be positioned as an educational support tool rather than a surveillance mechanism to maintain institutional trust.

B. Limitations and Future Work

The system has several limitations motivating future research. Performance degrades under extreme illumination variation and wide-angle camera distortion. The seven-class taxonomy does not capture nuanced cognitive engagement indicators such as confusion or boredom that are not visually distinct from

VII. DISCUSSION

A. Pedagogical Implications

The Smart Vision system provides educators with an objective, data-driven lens on classroom dynamics unavailable at scale through conventional observation. By continuously measuring engagement across all students, the system enables teachers to identify persistent attention deficits, recognise behavioural patterns correlated with specific lesson content or teaching styles, and intervene proactively. Longitudinal behavioural data aggregated over a semester provides institutional administrators with evidence- based insights for curriculum design and teacher professional development. The real-time alert mechanism is designed to augment rather than replace teacher professional judgement. Alerts serve as a supplementary signal that directs teacher attention to specific students, reducing the cognitive load of comprehensive simultaneous monitoring without removing the human from the intervention process.

B. Ethical and Privacy Considerations

Continuous video monitoring of students raises important ethical and privacy considerations. The system is designed with the following safeguards: (1) all processing occurs on- premises with no external data transmission; (2) an optional automatic face blurring mode renders student faces unidentifiable; (3) raw video frames are not persisted only aggregated behavioural statistics are retained; and (4) no individual student identifiers appear in alert notifications. attentiveness. The system also requires GPU hardware, which may limit deployment in resource-constrained institutions. Future directions include: (1) integration of facial expression recognition and eye-gaze estimation for multi-modal engagement signals;(2)transformer-based temporal modeling via Video Swin Transformers; (3) federated learning to train on distributed classroom data without centralising sensitive video; (4) emotion recognition for adaptive instructional response; and (5) longitudinal field studies measuring impact on learning outcomes.

VIII. CONCLUSION

This paper presented Smart Vision, a real-time classroom behaviour detection system based on the YOLOv8 deep learning framework. The proposed system addresses the critical limitations of manual and passive monitoring by providing automated, objective, and scalable behaviour analysis of all students simultaneously. Operating at 34 FPS with a mean Average Precision of 92.4% across seven behaviour classes, the system demonstrates practical viability for real educational deployment. The system integrates video acquisition,YOLOv8 multi-class detection, SORT-based student tracking, temporal behaviour aggregation, automated alert notification, and longitudinal analytics into a cohesive end-to-end platform. Results demonstrate superior performance over prior state-of-the-art methods in both accuracy and inference speed. By providing timely, data-driven insights into student engagement, Smart Vision supports more responsive instruction while reducing teacher cognitive load. Future enhancements including motion recognition and federated learning will expand the platform toward a comprehensive AI- powered smart classroom management solution.

REFERENCES

1. W.Alsabhan,"Student cheating detection in higher education by implementing machine learning andLSTMtechniques,"Sensors,vol.23,no.8,p.4149,2023.
2. F.C.Lin et al., "Student behavior recognition system for the classroom environment based on skeleton pose estimation and person detection," Sensors, vol. 21, no. 16, p. 5314, 2021.
3. M.Khalil, P.Prinsloo, and S.Slade, "In the nexus of integrity and surveillance: Proctoring(re) considered," J. Computer Assisted Learning, vol.38,no.6,pp.1589–1602,2022.
4. S.Kaddoura and A.Gumaei, "Towards effective and efficient online exam systems using deep learning-based cheating detection," Intelligent Systems with Applications, vol. 16, p. 200153, 2022.
5. M.Ahmed et al., "Real-time violent action recognition using key frames extraction and deep learning," Proc. IEEE ICIP,2021.
6. S.Ren et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans.Pattern Analysis Machine Intell., vol. 39, no. 6, pp. 1137– 1149, 2017.
7. Y.Zhang, H.Wang, and Q.Liu, "Classroom student behaviour recognition using YOLOv5," Proc. IEEE ICCV Workshop, 2021, pp. 1–8.
8. G.Jocheretal.,UltralyticsYOLOv8,GitHub,2023.[Online].Available:https://github.com/ultralytics/ultralytics
9. J.Redmon et al., "You only look once: Unified, real-time object detection," Proc. IEEE CVPR, 2016, pp. 779–788.
10. A.Bewleyetal.,"Simple online and real time tracking," Proc. IEEE ICIP, 2016, pp. 3464–3468.