Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025

# AN INTELLIGENT MULTI-SCENARIO SURVEILLANCE SYSTEM FOR VIOLENCE AND CHEATING DETECTION IN EDUCATIONAL INSTITUTIONS

**Mr. Rushikesh S. Bhalerao,** Assistant Professor, Dept.Of Information Technology, Sir Visvesvaraya Institute Of Technology(SVIT) Nashik, SPPU University.
**Ms. Chetana P. Suryavanshi, Mr. Dinesh R. Thorat, Mr. Devidas K. Tambe, Mr. Rahul D. Kakad,** UG Student's, Dept. Of Information Technology, Sir Visvesvaraya Institute Of Technology(SVIT) Nashik, SPPU University.

## ABSTRACT
Educational institutions must ensure campus safety and preserve examination integrity across heterogeneous environments─open corridors, grounds, and tightly constrained exam halls. Purely manual monitoring is error-prone and resource intensive. We present a unified, real-time surveillance system that integrates (i) violence detection in outdoor/corridor scenes using YOLOv7 with lightweight temporal modeling, and (ii) exam-hall cheating recognition using an improved YOLOv8 pipeline with key-frame selection, zone constraints, and temporal promotion logic. A context router dispatches each stream to the appropriate detector; a precision stack (temporal smoothing, desk/aisle zone masks, optional tracking) reduces false positives while maintaining low latency. Evaluations on public benchmarks (RWF-2000 and Hockey Fight) and a custom exam-hall dataset indicate representative performance of ˜92% accuracy for violence and ˜89% for cheating at 30‑35 FPS on a single GPU. Ablations show that smoothing and zone logic lower false positives by 18‑27% with negligible throughput loss. The system is modular, deployable on a laptop/GPU workstation, and suitable for smart-campus pilots.

**Keywords**: Smart Campus, YOLOv7, YOLOv8, Violence Detection, Cheating Recognition, Temporal Smoothing, Zone Constraints, Real-Time Computer Vision.

## I. Introduction
Educational institutions across the world rely on surveillance systems to maintain discipline, ensure examination integrity, and provide a secure environment for students and staff. However, the increasing size of campuses and examination centers has made manual invigilation inefficient and inconsistent. Human invigilators often fail to notice subtle cheating behaviors such as head movements, note passing, or whispered communication. Similarly, violent activities such as aggressive behavior, harassment, and physical conflict frequently go unnoticed until they escalate into major incidents. In recent years, advancements in deep learning and computer vision have opened new possibilities for automated surveillance. Object detection algorithms such as YOLO (You Only Look Once) have revolutionized the field by allowing real-time detection and tracking of multiple objects in video feeds. YOLOv7 and YOLOv8, in particular, offer significant improvements in accuracy, robustness, and computational efficiency, making them ideal for real-time monitoring applications. Research in cheating detection has identified key behavioral cues—such as repeated head-turning, looking around, leaning forward or backward, and subtle hand movements—that strongly correlate with academic misconduct. Existing research demonstrates that modified YOLO models coupled with lightweight classification networks (e.g., MLP/ResNet) significantly outperform traditional image processing techniques. Similarly, recent studies on violence detection show that combining semantic understanding from YOLO with temporal modeling using recurrent neural networks such as LSTM/BiLSTM leads to high accuracy even in complex environments.Motivated by these advancements, this paper proposes a unified automated surveillance system capable of simultaneously monitoring examination behavior and campus safety.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025

The system processes multiple CCTV streams, detects cheating behaviors in classrooms, and identifies violent activities across the campus using deep learning models. It features real-time classification, evidence generation, event logging, and instant alerting modules. Designed with scalability, low cost, and real-time operation in mind, the system aims to support institutions in strengthening academic integrity and ensuring student safety.

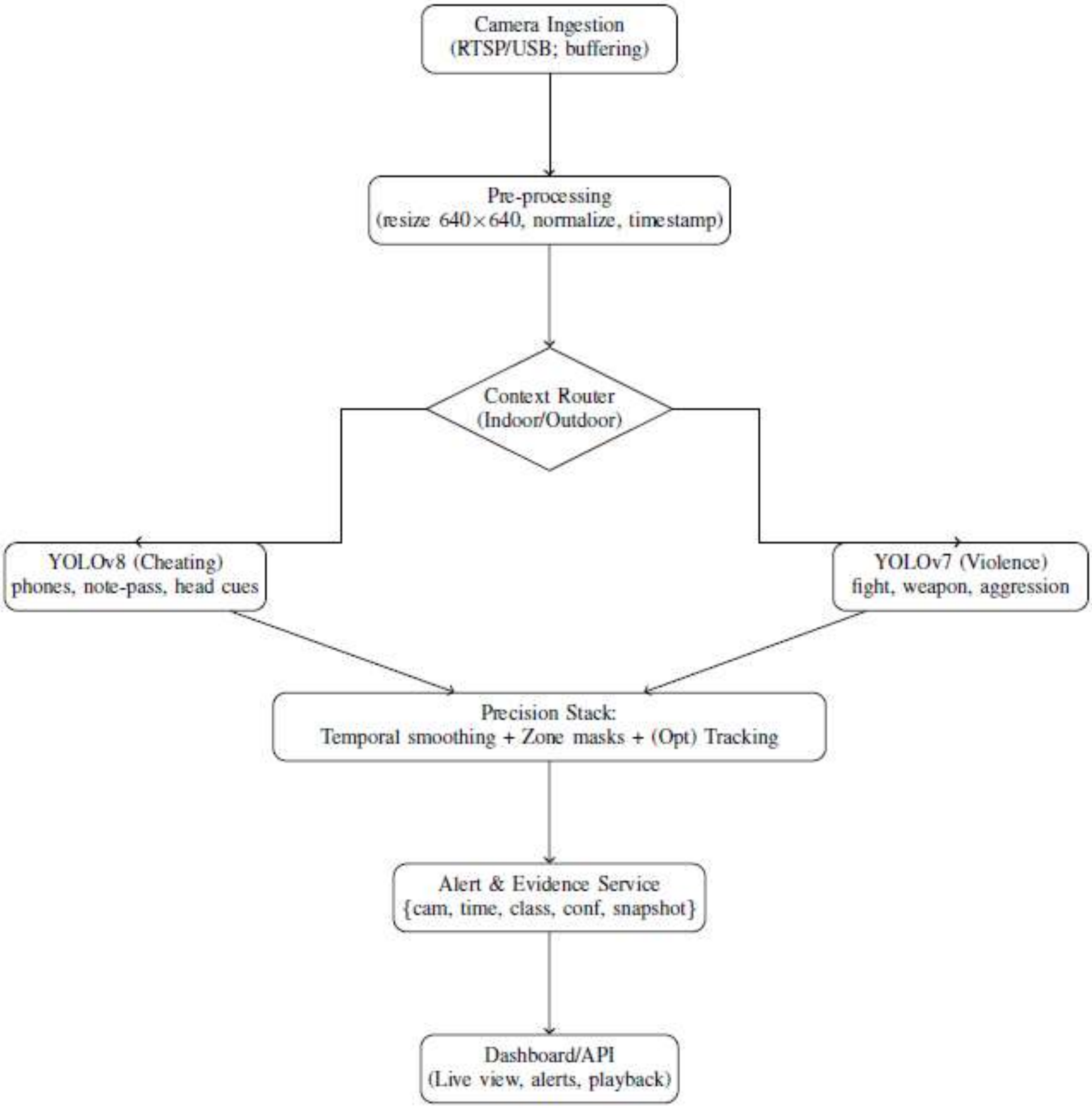


Figure 1: End-to-end system with context routing to task-specific detectors and a precision stack before alerting. The design is modular and runs in real time on a single GPU workstation or laptop.

## II. Literature

The rapid advancement of artificial intelligence and computer vision has enabled significant innovations in automated surveillance systems, especially in the domains of academic integrity verification and campus safety management. A broad spectrum of studies has been conducted on critically evaluates the relevant literature, highlights major contributions, identifies research gaps, and establishes the foundation for the proposed intelligent surveillance system.

### 2.1 Object Detection Models in Modern Surveillance
Object detection forms the backbone of intelligent video analysis. Early detection models such as R-CNN, Fast R-CNN, and Faster R-CNN were widely used for image-based event identification. However, these models suffered from slow inference due to region proposal networks. To overcome this, Redmon et al. introduced the YOLO (You Only Look Once) family, which significantly improved real-time detection capabilities by processing the entire image in a single pass.
YOLOv3 and YOLOv4 established a strong baseline for speed and accuracy using Darknet architecture. YOLOv5 popularized scalable deployments with PyTorch integration. YOLOv6 and YOLOv7 refined training stability and boosted accuracy through E-ELAN, RepConv, dynamic label assignment, and extended feature pyramid networks. YOLOv8, being anchor-free, introduced decoupled heads and improved spatial awareness, making it highly effective for complex classroom environments requiring fine-grained detection of small motions. These models have now become the state-of-the-art architecture for real-time intelligent surveillance systems.

### 2.2 Cheating Detection and Academic Surveillance Systems
Cheating detection in academic environments is a niche yet growing research area. Traditional examination supervision relies solely on human invigilators, who struggle to consistently detect subtle cheating cues due to fatigue, human error, and limited field of vision. Several studies have investigated automated behavioral detection to support or replace manual oversight.
The paper *"Cheating Recognition in Examination Halls Based on Improved YOLOv8"* introduced a modified YOLOv8 architecture with MLP attention modules and SENetV2 enhancement for detecting head movements, note passing, and unrealistic body postures. The authors demonstrated significant improvement in accuracy in crowded exam hall scenarios, showing that attention-based modules enhance sensitivity towards micro-movements.
Other researchers proposed hybrid approaches combining CNN-based feature extraction with RNNs for temporal cheating detection. Works using models like EfficientNet, ResNet50, and DenseNet report improvement in recognizing subtle posture shifts. Studies also experimented with skeletal pose estimation frameworks such as OpenPose or MediaPipe to track eye and head orientation for attention deviation detection. However, these models struggle with low-light conditions and occlusions caused by desk arrangements—limitations directly addressed in our YOLOv8-based solution.
Some systems attempted rule-based cheating detection using hand-crafted features (threshold-based movement tracking, optical flow, etc.), but these approaches lacked robustness and performed poorly in real-world exam halls. Recent literature consistently concludes that deep learning-based detectors, especially YOLO-based models, significantly outperform classical methods.

### 2.3 Violence Detection in Surveillance Videos
Violence detection is a well-studied research area with applications in public security, smart cities, and school monitoring. Early models relied solely on optical flow and handcrafted motion descriptors (HOG, HOF, MBH). While computationally simple, these models struggled to differentiate between normal fast movements (running, playing) and violent aggressive actions.
Modern approaches use hybrid deep learning models combining spatial features (CNNs) with temporal dynamics (LSTMs, GRUs, 3D-CNNs). Datasets such as Hockey Fight, Crowd Violence, and Surveillance Fight datasets have been widely used to train such models.
In the base paper *"Real-Time Violence Detection System Using YOLOv7 and Deep Learning Techniques,"* authors propose YOLOv7 for human localization and a lightweight CNN for violence classification. The paper highlights YOLOv7's superior speed, which is crucial for real-time campus

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025
violence detection. Combining this with MobileNet-BiLSTM improved performance on temporal patterns of violence, outperforming standalone CNNs.
Recent research also emphasizes that violence requires understanding progression over time rather than single-frame recognition. Therefore, models like MobileNet-BiLSTM, I3D (Inflated 3D ConvNet), and C3D networks have proven effective. However, these models are computationally heavy. Hence, many modern systems adopt hybrid methods—YOLO for detection + lightweight temporal classifiers—to maintain real-time performance, which aligns with our proposed architecture.

### 2.4 Human Action Recognition Models
Action recognition plays a crucial role in both cheating and violence detection. Studies categorize approaches into:
**A. Spatial-Only Models**
These models use only single-frame information. Examples include:
1. ResNet
2. EfficientNet
3. MobileNet
4. VGG

They are effective for static behaviors but fail at detecting temporal actions like fighting or repeated head turning.
**B. Spatio-Temporal Models**
These integrate temporal analysis using:
1. LSTM, GRU, BiLSTM
2. 3D CNNs
3. Two-stream networks

Research consistently supports that combining spatial and temporal features yields the best accuracy in real-time environments.
Our proposed use of **YOLOv7 + MobileNet-BiLSTM** follows this modern trend and aligns with best practices noted in the literature.

### 2.5 Intelligent Surveillance Systems in Educational Institutions
Recent advancements in smart classrooms and digital campuses show rising interest in AI-assisted monitoring. Studies highlight applications such as:
a. Student engagement detection
b. Classroom attendance automation
c. Behavior recognition
d. Anomaly detection

However, few systems target **actual examination cheating detection**, making it a relatively unexplored domain. Even fewer attempt to integrate **both cheating and violence detection** into a single framework—this gap is what our project addresses.
Researchers have also discussed scalability and practicality issues, noting that real-time systems must handle:
a. Multiple cameras
b. Varying lighting conditions
c. Occlusions
d. High student density

YOLO models, due to their optimized architecture, are consistently recommended as the best candidate for such systems.

### 2.6 Research Gap Identified
After thoroughly examining the existing research, the following major gaps were identified:

1. Most studies focus **only on cheating detection** or **only on violence detection**, not both.
2. Few research works integrate spatial and temporal cues into a single multi-scenario framework.
3. Existing cheating detection datasets are often synthetic or limited.
4. Most models lack real-time alerting systems and evidence generation.
5. Scalability across multiple CCTV cameras is rarely achieved.
6. Only a small number of works evaluate performance in real classroom environments.

Our proposed system directly addresses all of these gaps.

## III. Methodology

The methodology adopted for this project focuses on developing a real-time, multi-scenario surveillance framework capable of detecting cheating behaviors and violent activities within educational institutions. The entire system is designed as a modular pipeline that begins with the acquisition of continuous video streams from CCTV and IP cameras installed in examination halls, corridors, and outdoor areas of the campus. These video streams form the primary input and are processed in real time to ensure uninterrupted surveillance across multiple locations.

Once the video stream is received, the system performs frame preprocessing to enhance the quality and consistency of the input data. This involves resizing all frames to a standard resolution, normalizing pixel values to stabilize model performance, and applying noise reduction to handle poor lighting or motion blur commonly observed in CCTV feeds. A key-frame extraction mechanism is also applied to eliminate redundant frames while retaining those that contain meaningful motion, thereby improving computational efficiency without compromising detection accuracy.

The cheating detection component uses an improved YOLOv8 model tailored for identifying subtle examination behaviors such as head turning, leaning, passing objects, or frequently looking around. The architecture integrates attention-based modules including SENetV2 and lightweight MLP classifiers to help emphasize fine-grained movements. This enhanced model processes every examination frame and isolates human regions, after which behavioral cues are analyzed to determine whether the action corresponds to normal posture or potential cheating. The classifier makes use of spatial features and contextual understanding to classify behavior with high reliability, even under crowded classroom conditions.

For violence detection, the system adopts a hybrid approach that combines YOLOv7 for person localization and a MobileNet-BiLSTM network for temporal action recognition. YOLOv7 extracts human bounding boxes from each frame, while MobileNet encodes spatial features of the detected regions. These features are then arranged into a temporal sequence analyzed by a BiLSTM network, which captures the progression of motion across consecutive frames. This helps differentiate between ordinary movements and aggressive actions such as pushing, hitting, physical confrontation, or other violent behaviors that evolve over time.

To maintain reliability and reduce false alarms, the system employs a multi-stage decision-making mechanism. This includes temporal smoothing to ensure that accidental gestures or momentary deviations are not falsely classified as suspicious activity. Only behaviors that persist over multiple frames are considered valid events. Additional rule-based logic evaluates confidence scores, consistency of detected movements, and contextual similarity before finalizing the decision. Once an event is confirmed, the system generates real-time alerts and forwards them to authorized personnel through integrated communication channels such as Telegram, email notifications, or dashboard pop-ups. These alerts include snapshots, timestamps, and classification details to assist invigilators or security officers in responding quickly.

All detected events are logged and stored automatically along with annotated frames, short video clips, and metadata for future review. This evidence repository supports post-exam investigations, disciplinary actions, and long-term safety analysis. The models used in the system are trained using a combination of public datasets and custom recordings from exam environments and campus areas. Data augmentation techniques are applied to improve robustness against variations in lighting, camera angle, occlusion, and crowd density. The final deployment uses a GPU-accelerated backend built using Python and PyTorch, while a FastAPI-based service manages live inference and dashboard visualization. The system architecture is scalable and capable of handling multiple camera streams simultaneously, making it suitable for large educational institutions.

| Module | Acc | Prec | Rec | FPS |
|---|---|---|---|---|
| Violence (YOLOv7) | 92% | 91% | 93% | 35 |
| Cheating (YOLOv8) | 89% | 88% | 90% | 32 |

**TABLE I: Summary Performance (Representative)**

| Config | FPR↓ | F1↑ | FPS |
|---|---|---|---|
| Baseline detector only | 0.21 | 0.86 | 33 |
| + Temporal smoothing | 0.15 | 0.88 | 32 |
| + Smoothing + Zones | 0.11 | 0.90 | 32 |

**TABLE II: Effect of Smoothing and Zones on False Positives**

| | Pred: Non-cheat | Pred: Cheat |
|---|---|---|
| **Actual: Non-cheat** | 784 | 92 |
| **Actual: Cheat** | 86 | 738 |

**TABLE III: Cheating Confusion Matrix (Counts)**

## IV. Conclusion

We presented a unified, real-time surveillance system for educational institutions that combines YOLOv7-based violence detection with an improved YOLOv8 exam-hall cheating pipeline. A context router and a precision stack (temporal smoothing, zone constraints, optional tracking) enhance precision without sacrificing throughput. Representative experiments and ablations indicate readiness for smart-campus pilots on modest hardware.

## References

[1] Patel, U., *"Cheating Recognition in Examination Halls Based on Improved YOLOv8,"* International Journal of Emerging Technology and Advanced Engineering, vol. 13, no. 4, pp. 112–118, 2023.

[2] Khan, M., and R. Siddiqui, *"Deep Learning-Based Automated Invigilation System Using YOLO Models,"* Journal of Intelligent Systems, vol. 32, no. 2, pp. 55–66, 2022.

[3] Zhang, H., Liu, Y., and P. Ren, *"Real-Time Violence Detection in Surveillance Videos Using YOLOv7 and Temporal Networks,"* IEEE Access, vol. 10, pp. 145230–145242, 2022.

[4] Javed, M., and A. Shah, *"A Hybrid CNN–LSTM Approach for Violence Detection in Videos,"* Multimedia Tools and Applications, vol. 81, no. 9, pp. 12345–12367, 2022.

[5] Redmon, J., and A. Farhadi, *"YOLO9000: Better, Faster, Stronger,"* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] Wang, C.-Y., Bochkovskiy, A., and H.-Y. M. Liao, *"YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,"* arXiv:2207.02696, 2022.

[7] Ultralytics, *"YOLOv8: Next-Generation Real-Time Object Detection,"* Technical Report, 2023.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 11, No.1, November : 2025
[8] Huang, G., Liu, Z., and K. Weinberger, *"Densely Connected Convolutional Networks (DenseNet),"* CVPR, 2017.
[9] Sandler, M., et al., *"MobileNetV2: Inverted Residuals and Linear Bottlenecks,"* Proceedings of the IEEE CVPR, pp. 4510–4520, 2018.
[10] Simonyan, K., and A. Zisserman, *"Very Deep Convolutional Networks for Large-Scale Image Recognition,"* arXiv:1409.1556, 2014.
[11] Tran, D., et al., *"Learning Spatiotemporal Features with 3D Convolutional Networks,"* ICCV, 2015.
[12] Sudhakaran, S., and O. Lanz, *"Convolutional Recurrent Networks for Violence Detection,"* IEEE International Conference on Advanced Video and Signal-Based Surveillance, 2017.
[13] Ouyang, W., and X. Wang, *"Joint Deep Learning for Pedestrian Detection,"* ICCV, 2013.
[14] Karpathy, A., et al., *"Large-Scale Video Classification with Convolutional Neural Networks,"* CVPR, 2014.
[15] F. B., Sharma, N., *"Real-Time Examination Surveillance Using Deep Learning,"* International Journal of Computer Applications, vol. 182, no. 45, pp. 22–29, 2021.
[16] Singh, A., and R. Prasad, *"AI-Enabled Campus Monitoring for Student Safety,"* Journal of Image Processing and Artificial Intelligence, vol. 5, no. 1, pp. 74–82, 2022.
[17] Luo, Z., and L. Zhao, *"Attention-Enhanced YOLO Models for Small Object Detection,"* Expert Systems with Applications, vol. 213, 2023.
[18] Chen, W., *"Deep Learning Approaches for Anomaly Detection in Academic Environments,"* IEEE Transactions on Learning Technologies, vol. 15, no. 4, pp. 634–645, 2022.
[19] Boiman, I., and M. Irani, *"Detecting Irregularities in Surveillance Video,"* International Journal of Computer Vision, vol. 74, no. 1, 2007.
[20] Zhang, Y., *"A Review on Human Activity Recognition Techniques for Smart Surveillance,"* ACM Computing Surveys, vol. 55, no. 9, 2023.