IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15506

AI-Based Video surveillance System

Afsa Saboo[1], B Sai Dikshitha[2], C Mohammad Athiq[3], K Sudeep Gouda[4], Nagateja P[5], Anita Patil[6]

6th Sem B.E.(CS&AI), Ballari Institute of Technology and Management (BITM), Ballari, Karnataka-583104, India[1-4]

Assistant Professor, Department of Computer Science and Artificial Intelligence Engineering.

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka 583104, India[5]

Professor, Department of Computer Science and Artificial Intelligence Engineering.

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka 583104, India[6]

Abstract: Rapid urbanisation and escalating public safety demands have created an urgent need for intelligent, automated surveillance solutions that can operate without continuous human oversight. Conventional CCTV infrastructure places excessive cognitive load on operators monitoring multiple feeds simultaneously, increasing the risk of missed incidents due to fatigue and delayed response. This paper proposes a dual-model AI surveillance framework that concurrently detects three real-world emergency categories — road accidents, fire incidents, and suspicious human activity — by combining YOLOv8 spatial object detection with ResNet-50 temporal activity classification in a unified processing pipeline. On emergency detection, the system autonomously assembles an alert payload containing a timestamped snapshot, GPS-tagged camera location, and event confidence score, dispatching notifications in parallel via SMS, email, and mobile push notification to relevant authorities. Experimental evaluation on publicly available benchmark datasets yields detection accuracies of 89.2%, 91.5%, and 86.0% for accidents, fire, and suspicious activity respectively, with per-frame inference latency of 0.8–1.2 seconds and end-to-end alert delivery within three seconds. The proposed framework significantly reduces reliance on manual monitoring and offers a scalable, deployable foundation for smart city infrastructure, transportation hubs, and public safety control rooms.

Keywords: Artificial intelligence; video surveillance; YOLO; CNN; emergency detection; computer vision; real-time alerts; smart cities; deep learning**.**

I. INTRODUCTION

Urban growth and increasing population density have significantly raised the demand for efficient and scalable public safety solutions. Traditional CCTV surveillance systems rely on continuous human observation, where operators are required to monitor multiple video streams simultaneously. Over time, this process becomes cognitively demanding and prone to human error. Studies have shown that prolonged monitoring leads to a noticeable decline in attention levels, often within 20–35 minutes, increasing the likelihood of missing critical incidents [1].

With the advancement of Artificial Intelligence (AI) and computer vision technologies, surveillance systems are gradually shifting towards automation. Modern deep learning models can identify and interpret visual patterns in video data with high accuracy and reduced latency. Object detection frameworks such as YOLO, along with convolutional neural networks (CNNs) for feature extraction, have demonstrated strong performance in recognizing complex real-world scenarios, making them suitable for deployment in large-scale smart city environments [2], [3].

In this work, a unified intelligent surveillance framework is proposed to detect multiple categories of emergency situations, including road accidents, fire and smoke events, and suspicious human activities. The system integrates YOLOv8 for spatial object detection with a ResNet-50-based model for temporal activity recognition, forming a cohesive multi-event detection pipeline. Additionally, an automated alert mechanism is incorporated to transmit notifications containing relevant contextual information such as captured images, timestamps, and location data to concerned authorities.

The main contributions of this paper are as follows: (i) development of a real-time multi-event detection architecture combining object detection and activity classification; (ii) implementation of an automated alerting system with contextual metadata delivery; and (iii) comprehensive evaluation of the proposed framework using publicly available datasets across diverse scenarios. The remainder of the paper is organized as follows: Section II presents related work, Section III describes the system architecture, Section IV outlines the methodology, Section V discusses experimental results, and Section VI concludes the paper with future research directions.

II. RELATED WORK

Research in automated video surveillance has advanced significantly across three key areas: road accident detection, fire

DOI: 10.17148/IJARCCE.2026.15506

and smoke recognition, and suspicious activity analysis. Despite these developments, most existing approaches focus on a single application domain, limiting their effectiveness in real-world multi-hazard environments.

A. Road Accident Detection

Recent studies have explored deep learning techniques for accident detection in traffic scenarios. Muskan and Sethu Madhavi [4] employed a YOLOv5-based approach to identify accidents in real time, reporting satisfactory performance under relatively simple traffic conditions. However, their model exhibits reduced reliability in complex environments involving multiple vehicles, where occlusion affects detection accuracy. Arifeen et al. [5] adopted convolutional feature-based classification methods that provide accurate post-event analysis, but their approach lacks real-time processing capability, restricting its applicability in live surveillance systems. Similarly, Sabitha et al. [13] utilized ResNet-based architectures to enhance feature representation; however, their work does not incorporate an automated alert mechanism, which is essential for timely emergency response.

B. Fire and Smoke Detection

Fire detection using deep learning has gained considerable attention due to its importance in safety-critical systems. Carletti et al. [6] introduced a YOLOv8-based model that achieves high detection accuracy, particularly in outdoor environments with clear visual cues. Xu et al. [7] contributed the TAD benchmark dataset, which provides detailed temporal annotations and supports standardized evaluation across different models. While these approaches demonstrate the effectiveness of deep learning in identifying fire and smoke, they are typically designed as standalone solutions and do not integrate with other emergency detection tasks, limiting their scalability in comprehensive surveillance systems.

C. Suspicious Activity and Multi-Event Systems

Detecting anomalous human behavior remains a challenging task due to variations in motion patterns, crowd density, and camera perspectives. Tyagi et al. [8] proposed a lightweight convolutional neural network for violence detection, highlighting challenges such as viewpoint variation and occlusion. Other works have attempted similar approaches but often lack mechanisms for real-time alert generation. In the context of multi-event systems, Krishnan et al. [10] combined AI, deep learning, and IoT technologies to develop a smart surveillance framework, emphasizing the role of edge computing in reducing latency. However, their system provides only partial alert functionality and does not evaluate performance across multiple emergency scenarios. Likewise, Thosar et al. [14] and Kaushik et al. [15] proposed integrated surveillance architectures, but their evaluations treat each event category independently rather than within a unified detection pipeline.

D. Research Gap

From the existing literature, it is evident that current systems either focus on individual event detection or lack complete real-time alerting capabilities. Very few approaches attempt to combine multiple emergency categories into a single cohesive framework. Furthermore, the absence of cross-domain evaluation and context-aware notification mechanisms limits practical deployment. To address these challenges, the proposed system introduces a unified real-time pipeline capable of detecting road accidents, fire incidents, and suspicious activities simultaneously, along with an automated alert system that delivers rich contextual information to authorities.

III. SYSTEM ARCHITECTURE

The proposed system is structured as a four-stage pipeline: video acquisition, data preprocessing, AI-based event detection, and automated alert generation.

A. Video Acquisition. Live video is captured from CCTV or IP cameras deployed at monitored locations. Cameras stream H.264-encoded footage continuously to the processing server. The module handles stream reconnection on network interruptions, ensuring uninterrupted coverage.

B. Data Preprocessing. Each video stream is decoded and split into individual frames using OpenCV. Each frame undergoes: (i) resizing to 640×640 pixels (YOLO input resolution); (ii) pixel normalization to the [0, 1] range; and (iii) Gaussian noise filtering to suppress sensor artifacts. These steps improve input signal quality and stabilize detection accuracy across varying camera conditions.

C. AI-Based Event Detection. Preprocessed frames are passed to two complementary deep learning modules. The YOLOv8 object detector localizes objects of interest—people, vehicles, fire/smoke regions, and abandoned objects. A ResNet-50 activity classifier then examines temporal sequences of detected objects to recognize behavioral patterns indicative of emergencies: sudden deceleration (accident), rapid flame expansion (fire), or irregular motion clusters (fighting or intrusion). Rule-based confidence thresholds determine final emergency classification.

D. Alert Generation. Upon emergency classification, the system captures a high-resolution snapshot and a five-second video clip. An alert payload is assembled containing: (i) captured media; (ii) event type and confidence score; (iii) UTC timestamp; and (iv) camera GPS coordinates. Alerts are dispatched via SMS (carrier gateway API), email (SMTP), and mobile push notification in parallel, ensuring redundant delivery.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026

DOI: 10.17148/IJARCCE.2026.15506

IV. METHODOLOGY

A. **Dataset Collection**–Training and evaluation data were sourced from three publicly available repositories. For road accident detection, the TAD benchmark [7] was used alongside dashcam footage collections from IEEE Dataport. Fire and smoke training data were drawn from the VisiFire dataset and the MIVIA Fire Detection dataset. Suspicious activity samples were obtained from the UCF-Crime benchmark, a widely adopted surveillance anomaly dataset containing videos across multiple real-world incident categories [21]. A combined training set of approximately 42,000 annotated frames was constructed and partitioned using an 80/10/10 train/validation/test split. To improve model generalization across varying lighting, angle, and camera quality conditions, augmentation techniques were applied including horizontal flipping, random cropping, brightness jitter, and mosaic augmentation.

B. **Model Training**–The YOLOv8-medium backbone, pre-trained on the COCO dataset, was fine-tuned over 50 epochs using the AdamW optimizer. A learning rate of 0.001 and weight decay of 0.0005 were selected based on preliminary experiments that showed faster convergence compared to SGD on our imbalanced training distribution. Training was conducted on an NVIDIA RTX 3060 GPU. The ResNet-50 activity classifier was trained independently for 30 epochs using cross-entropy loss. Class-weighted sampling was applied to mitigate the effect of category imbalance, particularly for the underrepresented suspicious activity class.

C. **Inference Pipeline**–Each video frame is processed independently by the YOLOv8 detector, with detections retained above a confidence threshold of 0.45. A sliding window of 10 consecutive frames is then passed to the ResNet-50 activity classifier to capture temporal context. An emergency alert is triggered only when the classifier probability exceeds 0.70 across at least three consecutive windows. This multi-window confirmation step was introduced specifically to suppress false positives caused by transient visual artifacts such as lighting changes or partial occlusions.

D. **Alert Dispatch**–The alert dispatcher operates as an asynchronous thread, deliberately decoupled from the inference pipeline so that network latency during alert delivery does not affect detection throughput. To prevent notification floods during prolonged incidents, a deduplication mechanism suppresses repeat alerts for the same camera zone within a configurable 60-second cooldown window.

V. EXPERIMENTAL RESULTS

A. Experimental Setup -

Experiments were conducted on a workstation with an Intel Core i5-12400, 16 GB RAM, and NVIDIA RTX 3060 (12 GB VRAM). The software stack comprised Python 3.10, PyTorch 2.1, Ultralytics YOLOv8, and OpenCV 4.9. All evaluation used the held-out test split; no test data was used during training or hyperparameter tuning.

B. Detection Performance -

Table I summarizes detection performance across the three emergency categories. TABLE I. DETECTION PERFORMANCE ACROSS EMERGENCY SCENARIOS

Scenario	Accuracy	Precision	Recall	F1-Score
Accident Detection	89.2%	87.4%	85.1%	86.2%
Fire Detection	91.5%	90.1%	88.3%	89.2%
Suspicious Activity	86.0%	84.2%	83.0%	83.6%

Fire detection achieved the highest accuracy (91.5%) owing to visually distinctive color and motion signatures of flames and smoke. Accident detection performed well but exhibited lower recall in complex multi-lane traffic scenarios where vehicle occlusion obscures collision cues. Suspicious activity detection showed the lowest scores, reflecting the inherent variability of human behavior.

C. Real-Time Performance-

The system processed video streams at an average throughput of 28 frames per second, with per-frame inference latency of 0.8–1.2,seconds including preprocessing. Alert generation introduced an additional latency

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 5, May 2026
DOI: 10.17148/IJARCCE.2026.15506

under 2 seconds, yielding end-to-end emergency notification delivery within approximately 3 seconds of the triggering event.

D. Comparison with Prior Systems-

Table II compares the proposed system with representative prior work.

TABLE II. COMPARISON WITH EXISTING SYSTEMS

System	Multi-event	Auto Alert	Avg. Acc.	Real-Time
Muskan & Madhavi [4]	No	No	87.0%	No
Aryan et al. [9]	Partial	No	85.0%	No
Krishnan et al. [10]	Yes	Partial	88.0%	Partial
Proposed System	Yes	Yes	88.9%	Yes

Limitations:
The following limitations were observed:

(i) detection accuracy degrades under poor illumination (below 10 lux);
(ii) highly crowded scenes increase false detection rates for suspicious activity;
(iii) alert delivery depends on network connectivity, with intermittent connections potentially delaying SMS and email dispatch;
(iv) the current model is trained on RGB data and requires re-training for infrared cameras

VI. CONCLUSION

This work demonstrated that a unified deep learning pipeline can reliably detect three distinct emergency categories — road accidents, fire incidents, and suspicious human activity — from live video feeds without requiring separate, independently deployed systems. The combination of YOLOv8 for spatial object localization and ResNet-50 for temporal activity classification proved effective across diverse real-world scenarios, achieving F1-scores between 83.6% and 89.2% and sustaining end-to-end alert delivery within three seconds.

A key finding is that multi-event unification does not come at the cost of per-category accuracy. Fire detection benefited from visually distinctive flame signatures, while suspicious activity detection highlighted the inherent challenge of behavioral variability — suggesting that future models must account for scene context, not just object appearance. The asynchronous alert dispatch design proved critical in ensuring that network communication overhead did not degrade detection throughput, a consideration often overlooked in single-task prior systems.

Taken together, these results position the proposed framework as a practical foundation for smart city safety infrastructure, where operators must monitor heterogeneous threats across large camera networks. Three directions remain open for future investigation: improving robustness under low-light and infrared imaging conditions through domain adaptation techniques; distributing inference to edge nodes co-located with cameras to reduce backbone network dependency; and embedding privacy-preserving mechanisms such as on-device face anonymization to meet evolving data protection regulations.

REFERENCES

[1] D. Donald, C. Donald, and A. Thatcher, "Work exposure and vigilance decrements in closed-circuit television surveillance," *Appl. Ergonom.*, vol. 47, pp. 220–228, 2015.
[2] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
[4] M. S. and R. S. Madhavi, "Accident detection and alert system using YOLO model," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 12, no. 3, pp. 45–52, 2024.
[5] Z. U. Arifeen, M. S. Islam, and M. A. Rahman, "Traffic accident detection and classification using deep learning," in

DOI: 10.17148/IJARCCE.2026.15506

Proc. IEEE Int. Conf. Adv. Inf. Commun. Technol. (ICAICT), 2019, pp. 112–117.

[6] Y. Xu *et al.*, "TAD: A large-scale benchmark for traffic accidents detection from video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 1, pp. 12–25, 2025.

[7] A. Aryan, R. Mehta, and S. Kulkarni, "Intelligent video surveillance system for activity recognition," *J. Comput. Vis. Image Process.*, vol. 8, no. 2, pp. 78–91, 2023.

[8] S. Krishnan, P. Anand, and M. Nair, "Smart safety surveillance using AI, deep learning and IoT," in *Proc. IEEE Int. Conf. Syst., Comput. Sci. (ICSCS)*, 2025, pp. 301–308.

[9] M. Latha, R. Devi, and K. Sundar, "AI-driven YOLO surveillance system with automated reporting," in *Proc. IEEE Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, 2025, pp. 189–195.

[10] S. Sharma, A. Gupta, and R. Tiwari, "AI-powered surveillance: A comprehensive survey," *IEEE Access*, vol. 13, pp. 55421–55445, 2025.

[11] D. Thosar, S. Patil, and A. Joshi, "SmartSurveil: An integrated surveillance system for urban environments," in *Proc. IEEE Int. Conf. Electron., Comput. Autom. (ICECA)*, 2025, pp. 210–217.

[12] A. Kaushik, R. Singh, and P. Verma, "AI-based detection and tracking system for public safety," in *Proc. IEEE Int. Conf. Artif. Intell. (ICCAI)*, 2025, pp. 325–332.

[13] L. Hu, W. Zhang, and Q. Chen, "Highway surveillance AI system with multi-modal sensor fusion," *IEEE Sensors J.*, vol. 24, no. 5, pp. 8901–8915, 2024.

[14] P. Suthahar, K. Rani, and S. Mohan, "AI-based road safety system with real-time hazard detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 16, no. 2, pp. 211–219, 2025.

[15] C. Thyagarajan, V. Selvan, and R. Mani, "Rapid crime response system using deep learning surveillance," in *Proc. IEEE Int. Conf. Smart Comput. Commun. Control (ICSCCC)*, 2025, pp. 89–96.

[16] V. Chundi, S. Reddy, and N. Rao, "Intelligent video surveillance systems: A review," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 9, pp. 8453–8467, 2021.

[17] M. Choubisa, N. Jain, and A. Sharma, "Object tracking in surveillance using deep learning," in *Proc. IEEE Int. Conf. Adv. Comput. Commun. Informat. (ICACCI)*, 2020, pp. 501–507.

[18] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 6479–6488.

