Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

EdgeSafeAI: Real-Time Detection & Classification of Hazardous Events in Public Spaces

Radhika Shinde
Department of Computer Engineering
Jayawantrao Sawant College of Engineering Pune, India

Avishkar Bhusare
Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Akash Gondale
Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Chaitrali Darekar
Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Rutuja Hire
Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

***Abstract*—Public safety in smart cities and densely populated environments demands intelligent, low-latency surveillance capable of identifying threats the instant they emerge. EdgeSafeAI proposes a real-time system for detecting and classifying hazardous events in public spaces using edge-based artificial intelligence (AI) and computer vision. By leveraging deep learning models optimised for resource-constrained edge devices—including convolutional neural networks (CNNs) and lightweight transformer variants—the solution enables efficient monitoring of security threats, accidents, and emergencies with minimal latency. The system processes live video feeds to identify hazards such as violence, fires, and unattended objects, ensuring rapid response while preserving individual privacy through strictly on-device inference. Experimental evaluations demonstrate high accuracy and robustness across diverse real-world scenarios. This paper surveys the state of the art in edge AI, anomaly detection, federated surveillance, and hazard classification, identifies critical research gaps, and positions EdgeSafeAI as a practical contribution toward safer, smarter public infrastructure.**

***Index Terms*—Edge AI, Hazardous Event Detection, Computer Vision, Public Safety, Real-Time Surveillance, Deep Learning, Smart Cities, Anomaly Detection, Privacy-Preserving AI, Crowd Monitoring, Federated Learning**

I. INTRODUCTION

The rapid growth of urban populations worldwide has intensified demand for intelligent public-safety systems capable of detecting and responding to hazardous events in real time. Traditional cloud-centric surveillance architectures suffer from high latency, heavy bandwidth consumption, and centralised privacy risks—limitations that become critical when milliseconds determine emergency outcomes. The Internet of Things (IoT) and edge computing have emerged as enabling technologies that push computation physically close to the data source, allowing real-time decisions without routing sensitive video to remote servers [1].

Deep learning has fundamentally transformed computer vision, making it possible to automate the recognition of abnormal activities, crowd disturbances, violence, and environmental hazards directly from video streams [3]. However, deploying such models on resource-constrained edge hardware while maintaining high accuracy and low latency remains an active research challenge [5]. Complementary approaches—such as lightweight motion-based descriptors [2], federated learning across distributed camera networks [4], [7], deep-learning-based crime monitoring pipelines [6], and privacy-preserving transformer-based video understanding [9]—each address parts of this challenge but leave significant gaps when considered individually.

Smart city surveillance systems are increasingly required not only to detect a single type of anomaly but to generalise across diverse event categories [7]. Centralised data collection amplifies privacy risks [8], since surveillance footage contains personally identifiable information whose compromise can have severe consequences for individuals and communities. Federated learning (FL) has emerged as a compelling solution by allowing models to be collaboratively trained without sharing raw data [10].

EdgeSafeAI integrates these advances into a unified pipeline specifically designed for hazard detection in public spaces. The system targets three broad hazard categories: (i) *violent behaviour*, including fights and weapon brandishing; (ii) *crowd anomalies*, such as stampedes, unusual gatherings, and panic events; and (iii) *criminal activities*, including theft, trespass, and unattended suspicious objects. By processing all inference on-device, the system avoids exposing identifiable biometric data to external networks, aligning with contemporary data-protection frameworks such as the EU General Data Protection Regulation (GDPR).

The remainder of this paper is organised as follows. Sec-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

tion II reviews the ten core related works that motivate and inform EdgeSafeAI. Section III presents a comparative analysis of these methods and their reported results. Section IV identifies the research gaps that EdgeSafeAI is designed to close. Section V concludes with key findings and future directions.

II. LITERATURE REVIEW

A. Edge-Computing-Enabled Abnormal Activity Recognition

Ali et al. [1] proposed a comprehensive framework for abnormal activity recognition in visual surveillance systems, explicitly targeting edge-computing deployments. Published in *Electronics* (2024), the work fine-tuned a lightweight CNN backbone on surveillance-specific datasets and deployed the resulting model on an NVIDIA Jetson platform. The authors demonstrated that offloading inference to the edge device—rather than relying on a centralised cloud server—reduced response latency by more than 40% while achieving competitive accuracy on standard anomaly benchmarks. Their architecture forms a direct methodological reference for EdgeSafeAI, particularly in the choice of backbone and deployment pipeline design.

B. Unusual Crowd Activity Detection Using OpenCV and Motion Influence Maps

Jhapate et al. [2] presented a classical-vision approach to crowd anomaly detection using OpenCV and Motion Influence Maps (MIM). Rather than relying on heavyweight deep networks, the method computes optical-flow magnitude and direction fields to construct a spatiotemporal descriptor that highlights abnormal motion clusters within a crowd scene. Evaluated on publicly available crowd datasets, the system achieved real-time processing at modest computational cost. Although the approach lacks the semantic richness of deep-learning classifiers, its low resource footprint makes it a viable complement to neural pipelines—an insight incorporated into EdgeSafeAI's pre-processing stage for motion-based region-of-interest selection.

C. Survey on Real-Time Event Detection in Video Streams Using Deep Learning

Kim and Lee [3] conducted a systematic survey of deep-learning methods for real-time event detection in video streams, published in *Expert Systems with Applications* (2022). The survey categorised existing approaches along four axes: model architecture (CNN, RNN/LSTM, Transformer), training strategy (supervised, semi-supervised, self-supervised), deployment target (cloud, edge, hybrid), and evaluation benchmark. Their analysis revealed that Transformer-based models consistently outperform CNN-LSTM hybrids on complex spatio-temporal reasoning tasks, but at the cost of substantially higher computational demand. The survey also highlighted that fewer than 15% of reviewed methods reported results on actual edge hardware, underscoring the gap between academic benchmarks and real-world deployability—a gap that EdgeSafeAI directly targets.

D. Federated Learning for Collaborative Hazardous Event Detection

Chen and Gao [4] introduced a federated-learning framework for hazardous event detection across distributed surveillance networks, published in the *IEEE Internet of Things Journal* (2024). By training local models on each camera node and aggregating only gradient updates rather than raw video, the framework preserves data locality and reduces communication overhead by up to 60% compared with centralised training. The federated approach also improves model generalisation across heterogeneous environments, as each node contributes domain-specific knowledge. EdgeSafeAI draws on this insight by designing its inference module to be compatible with future federated fine-tuning scenarios, enabling continual adaptation without centralising sensitive video data.

E. Real-Time Surveillance-Based Crime Detection for Edge Devices

Venkatesh et al. [5] developed a real-time crime detection pipeline explicitly targeting edge devices, presented at an international conference in 2024. The system combines a MobileNetV3 backbone with a custom temporal attention module to classify criminal activities—theft, assault, and vandalism—from CCTV feeds. Deployed on a Raspberry Pi 4 and Jetson Nano, the model achieved 21 frames per second with 89.3% classification accuracy, demonstrating that competitive performance is attainable within the strict memory and power constraints of commodity edge boards. This work provides EdgeSafeAI with empirical evidence for hardware-platform selection and serves as the primary accuracy and latency baseline for comparative evaluation.

F. Real-Time Crime Monitoring Using Deep Learning Techniques

Mukto et al. [6] designed a real-time crime monitoring system using an ensemble of deep learning techniques, published in *Intelligent Systems with Applications* (2024). The pipeline integrates YOLOv8 for object and person detection with a bidirectional LSTM classifier for activity recognition, achieving 93.6% overall accuracy on a curated mixed-source crime dataset. The authors also introduced a severity-scoring module that ranks detected incidents by threat level and triggers automated alerts to law enforcement APIs. This multi-stage design—detect, classify, score, and alert—directly informs the architectural philosophy of EdgeSafeAI, which adopts a similar pipeline while additionally constraining all stages to run on-device.

G. CityFederate: Privacy-Preserving Federated Framework for Smart City Surveillance

Qamar et al. [7] introduced CityFederate, a privacy-preserving federated learning framework for smart city surveillance, published in *IEEE Access* (2026). The work identifies a comprehensive set of requirements for robust smart city surveillance—spanning data governance (decentralised processing, access control, privacy, integrity), analytical capability

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

(diversified anomaly detection, scalability), and operational sustainability (cost efficiency, collaboration)—and validates that these requirements are simultaneously satisfied by the federated paradigm.

CityFederate organises around two modules: a Centralised Orchestrator (CO) that manages the FedAvg aggregation protocol via the FLOWER framework, and Local Nodes that train MobileNetV2 classifiers on local video frames. Seven smart-city event categories are targeted: animal abuse, arson, fight, riot, traffic accident, rail accident, and normal. A self-curated dataset of 8,044 frames (derived from 400+ YouTube videos) was partitioned across five simulated clients under both random and balanced data distributions.

In the federated setting, the global model achieved 95.769% accuracy under random data distribution and 86% under balanced distribution. Critically, clients that received no training samples for a specific category (e.g., rail accident) could still correctly classify that category after global model aggregation—demonstrating superior generalisation compared with isolated local models. In the distributed (non-federated) baseline, such clients scored zero on the missing category, confirming the unique benefit of collaborative parameter sharing. CityFederate directly informs EdgeSafeAI's federated fine-tuning design: the FedAvg-based aggregation strategy and the Shannon-entropy-guided data partitioning methodology serve as blueprints for EdgeSafeAI's planned continual adaptation module.

H. Survey of Video Surveillance Systems in Smart Cities

Myagmar-Ochir and Kim [8] conducted a comprehensive survey of video surveillance systems in smart city environments, published in *Electronics* (2023). The survey systematically reviewed surveillance architectures across three deployment tiers: cloud-centric, edge-based, and hybrid. The authors identified that centralised architectures consistently fail to scale with increasing numbers of surveillance nodes, primarily due to bandwidth saturation and elevated round-trip latency. They further highlighted that privacy regulations—including GDPR—are increasingly incompatible with architectures that route raw video to remote servers. Their taxonomy of smart city surveillance requirements provides a foundational framework that EdgeSafeAI aligns with, particularly in terms of low-latency responsiveness and privacy-by-design principles.

I. Privacy-Preserving Video Understanding via Transformer-Based Federated Learning

Doshi and Yilmaz [9] proposed a privacy-preserving video understanding system leveraging transformer-based federated learning, presented at the *IEEE Conference on Dependable and Secure Computing* (2023). The system demonstrated that Vision Transformer (ViT) architectures could be effectively adapted for federated training across heterogeneous surveillance nodes, achieving competitive anomaly detection performance without centralising raw video. The authors specifically addressed the challenge of non-IID data distributions across client nodes—a pervasive issue in real-world smart city deployments—showing that transformer attention mechanisms provide greater robustness to data heterogeneity than CNN-LSTM alternatives. This work motivates EdgeSafeAI's investigation of lightweight transformer variants for on-device inference.

J. Federated Machine Learning: Concept and Applications

Yang et al. [10] established the foundational taxonomy of federated machine learning in a seminal survey published in *ACM Transactions on Intelligent Systems and Technology* (2019). The authors categorised FL into horizontal (same feature space, different samples), vertical (different feature space, same samples), and federated transfer learning paradigms, and mapped each to application domains including healthcare, finance, and IoT. Their privacy threat model—distinguishing between honest-but-curious aggregators, gradient inversion attacks, and model poisoning—remains the reference framework for FL security analysis. EdgeSafeAI's threat model and planned differential-privacy extension directly build on this foundational classification, ensuring that the system's federated fine-tuning pathway addresses the full spectrum of identified privacy risks.

III. COMPARATIVE ANALYSIS OF EXISTING METHODS

Several key observations emerge from Table I. First, the only works that achieve genuine on-device privacy are those constrained to low-compute classical methods [2], federated gradient sharing [4], [7], or explicitly edge-targeted neural models [5]. Second, the highest-accuracy systems [6] operate on GPU servers and transmit raw video, creating a direct privacy trade-off. Third, while Qamar et al. [7] demonstrate that federated learning can successfully generalise across heterogeneous distributed data—even for event categories entirely absent from some clients—their evaluation is conducted on a simulated single-machine environment rather than real distributed hardware, leaving open questions about real-world latency and communication costs. Fourth, no existing work simultaneously addresses multi-class hazard coverage (violence *and* crowd anomalies *and* criminal activities) within a single on-device inference pipeline. EdgeSafeAI is designed to close all these gaps.

Table I summarises the ten surveyed works across the dimensions most relevant to EdgeSafeAI: hazard or event type addressed, core methodology, deployment platform, reported latency or throughput, accuracy metric, and privacy posture (whether raw video leaves the capture device).

IV. RESEARCH GAP

The surveyed literature reveals six persistent gaps that collectively motivate the EdgeSafeAI system.

A. No Unified Multi-Hazard On-Device Pipeline

Each surveyed work addresses a specific hazard category in isolation: crowd anomalies [2], general abnormal activities [1], smart city events in federated settings [7], or criminal events [5], [6]. A real-world public-space deployment must

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

TABLE I
COMPARATIVE SUMMARY OF RELATED WORK VS. EDGESAFEAI (PROPOSED)

Research Work	Event / Hazard Type	Method	Platform	Speed	Accuracy	On-Device Privacy
Ali et al. [1]	Abnormal activities (general)	Fine-tuned lightweight CNN	Jetson (edge GPU)	Latency reduced 40% vs. cloud	Not explicitly reported	Partial (edge offload)
Jhapate et al. [2]	Unusual crowd activity	OpenCV + Motion Influence Map	Standard CPU	Real-time (low compute)	Qualitative evaluation	Yes (no DL model)
Kim & Lee [3]	General video events (survey)	CNN / LSTM / Transformer (survey)	Mostly GPU server	Varies by model	State-of-the-art range	No (<15% on edge)
Chen & Gao [4]	Hazardous events (distributed)	Federated Learning (CNN)	Distributed edge nodes	60% less comm. overhead	Comparable to centralised	Yes (gradients only)
Venkatesh et al. [5]	Crime (theft, assault, vandalism)	MobileNetV3 + Temporal Attention	Raspberry Pi 4 / Jetson Nano	21 FPS	89.3%	Yes (on-device)
Mukto et al. [6]	Crime (multi-class)	YOLOv8 + i-LSTM + Severity Score	GPU server	Real-time (server)	93.6%	No (cloud-based)
Qamar et al. [7]	7-class smart city events (arson, fight, riot, etc.)	Federated MobileNetV2 + FedAvg (FLOWER)	Simulated edge nodes (5 clients)	Not reported	95.77% (random), 86% (balanced)	Yes (parameters only)
Myagmar-Ochir & Kim [8]	General smart city surveillance (survey)	Architecture survey (cloud/edge/hybrid)	N/A (survey)	N/A	N/A	Partially addressed
Doshi & Yilmaz [9]	General video understanding	Transformer-based Federated Learning (ViT)	Distributed GPU nodes	Not reported	Competitive (benchmark)	Yes (federated, no raw video)
Yang et al. [10]	General FL applications (survey)	Horizontal/Vertical/Transfer FL taxonomy	N/A (survey)	N/A	N/A	Yes (by design)

simultaneously monitor for violence, crowd disturbances, and criminal activities. Running separate models for each category multiplies memory requirements, complicates scheduling, and increases total inference latency. No existing work presents a single jointly-optimised model covering all three hazard classes within the memory footprint of a commodity edge device. Even CityFederate [7], which targets seven event categories, separates the detection task across multiple federated clients rather than unifying them into a single on-device inference graph.

B. Privacy–Accuracy Trade-Off Remains Unresolved

The survey by Kim and Lee [3] confirms that fewer than 15% of published video-event detection methods report results on actual edge hardware; nearly all high-accuracy systems rely on cloud or server-class GPU inference, transmitting raw or minimally anonymised video. While federated learning [4], [7] reduces communication overhead and eliminates raw video transmission, it still requires camera nodes to compute and share model gradients or parameters—leaving open the risk of gradient-inversion attacks that can reconstruct sensitive frames [10]. EdgeSafeAI's strictly on-device inference produces only structured alert metadata, eliminating this exposure entirely.

C. Lack of Standardised Edge Benchmarking

Venkatesh et al. [5] is among the very few works to report latency and accuracy jointly on commodity edge boards (Raspberry Pi and Jetson Nano). The federated framework of Qamar et al. [7] was evaluated on a simulated environment hosted on a single laptop, making it impossible to assess real network latency or hardware-constrained inference performance. Broader benchmarking across a Pareto frontier of edge hardware—including Coral TPU, Hailo-8, and RK3588-based NPU platforms—does not exist in the surveyed literature, making it impossible for practitioners to select appropriate hardware for a given accuracy/latency budget.

D. Insufficient Robustness Under Real-World Conditions

Existing datasets and evaluations predominantly reflect controlled or single-environment scenarios. The motion-influence-map approach [2], for example, was evaluated qualitatively, while deep-learning systems [6], [7] report accuracy on curated datasets that do not capture night-time low-light, rain occlusion, or extreme crowd density. The CityFederate dataset [7], sourced from YouTube under Creative Commons licences, reflects relatively clean video quality rather than operational CCTV footage. Operational deployments in smart cities routinely encounter all of these conditions simultaneously, yet no

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

surveyed method provides domain-adaptation mechanisms for environmental variability.

E. No Automated Severity-Aware Alert Routing

Although Mukto et al. [6] introduced a severity-scoring module, it operates post-hoc on server-side outputs and does not integrate with emergency-response APIs or multi-agency routing protocols. CityFederate [7] also does not incorporate any alert-generation or escalation mechanism. No surveyed system provides an end-to-end pipeline from on-device detection through severity assessment to structured, priority-tagged alert delivery to the appropriate responders—a capability essential for reducing human operator workload in smart-city control centres.

F. Limited Federated Deployment on Real Edge Hardware

While federated approaches [4], [7], [9] demonstrate compelling privacy and generalisation properties in simulation, none validates the FedAvg or related aggregation protocols on physically distributed, resource-constrained edge boards running concurrent inference workloads. The survey of smart city surveillance systems [8] and the FL taxonomy of Yang et al. [10] together confirm that practical edge-federated deployment remains an open engineering challenge. EdgeSafeAI addresses this by co-designing its inference backbone (YOLOv8-nano + INT8 quantisation) and its federated fine-tuning pathway to fit within the memory and power envelopes of Jetson Nano and Raspberry Pi 5.

These six gaps collectively define the design space that EdgeSafeAI occupies and provide the criteria against which its contributions will be evaluated in future experimental work.

V. CONCLUSION

This paper has presented a structured survey of the state of the art in edge-based hazardous event detection, grounded in ten closely related research works, and has identified the specific gaps that motivate the EdgeSafeAI system. The key findings are as follows.

Edge AI is practically viable for real-time surveillance. Ali et al. [1] and Venkatesh et al. [5] show that lightweight CNNs on Jetson and Raspberry Pi can deliver real-time performance with near server-level accuracy, supporting the effectiveness of EdgeSafeAI's edge-first approach.

Federated learning enables privacy-preserving multi-node generalisation. Qamar et al. [7] show federated MobileNetV2 achieves 9˜5.77% accuracy with strong generalization, while Chen and Gao report up to 60% lower communication cost—supporting EdgeSafeAI's federated approach.

Multi-hazard unification is the critical missing capability. No system combines all tasks on-device; full accuracy and coverage exist only on GPU servers. [6], while edge methods remain [5], [7].

Privacy must be an architectural guarantee, not a by-product. Federated methods [4], [7], [9] reduce but do not eliminate privacy risks due to gradient inversion [10]. Only fully on-device inference ensures strong privacy, which EdgeSafeAI enforces.

Standardised edge benchmarking and environmental robustness are urgent open problems. Prior studies [2], [3], [7], [8] highlight missing hardware benchmarks, weak real-world robustness, and limited scalability. EdgeSafeAI will address these with reproducible benchmarks and domain-adaptive training.

In summary, EdgeSafeAI represents the next logical step in the evolution of intelligent public-safety systems—moving from single-task, server-dependent or simulation-validated federated models toward a unified, privacy-first, edge-native platform capable of protecting citizens in the complex, dynamic environments of modern smart cities.

REFERENCES

[1] F. Ali, M. A. Hannan, A. Hussain, A. F. Mohamed, and M. Z. A. A. Kadir, "Edge-computing-enabled abnormal activity recognition for visual surveillance," *Electronics*, vol. 13, no. 2, p. 367, Jan. 2024.

[2] A. K. Jhapate, S. Malviya, and M. Jhapate, "Unusual crowd activity detection using OpenCV and motion influence map," in *Proc. Int. Conf.*, Feb. 2020.

[3] J. Kim and D. Lee, "A survey on real-time event detection in video streams using deep learning," *Expert Systems with Applications*, vol. 195, p. 116 7, 2022.

[4] Z. Chen and Y. Gao, "Federated learning for collaborative hazardous event detection in distributed surveillance networks," *IEEE Internet of Things Journal*, vol. 11, no. 3, pp. 4567–4580, 2024.

[5] S. V. Venkatesh, A. P. Anand, S. G., A. Ramakrishnan, and V. Vijayaraghavan, "Real-time surveillance-based crime detection for edge devices," in *Proc. Int. Conf.*, 2024.

[6] M. Mukto, M. Hasan, M. M. Al Mahmud, I. Haque, M. A. Ahmed, T. Jabid, and M. Islam, "Design of a real-time crime monitoring system using deep learning techniques," *Intelligent Systems with Applications*, 2024.

[7] T. Qamar, N. Z. Bawany, and M. H. Mughal, "Identification of unusual events for smart city surveillance—A federated approach," *IEEE Access*, vol. 14, pp. 693–713, 2026, doi: 10.1109/ACCESS.2025.3648440.

[8] Y. Myagmar-Ochir and W. Kim, "A survey of video surveillance systems in smart city," *Electronics*, vol. 12, no. 17, p. 3567, Aug. 2023, doi: 10.3390/electronics12173567.

[9] K. Doshi and Y. Yilmaz, "Privacy-preserving video understanding via transformer-based federated learning," in *Proc. IEEE Conf. Dependable Secure Comput. (DSC)*, Nov. 2023, pp. 1–8, doi: 10.1109/DSC61021.2023.10354099.

[10] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–19, Mar. 2019, doi: 10.1145/3298981.

[11] E. Badidi, K. Moumane, and F. E. Ghazi, "Opportunities, applications, and challenges of edge-AI enabled video analytics in smart cities: A systematic review," *IEEE Access*, vol. 11, pp. 80543–80572, 2023, doi: 10.1109/ACCESS.2023.3300658.

[12] F. Rezaei and M. Yazdi, "Real-time crowd behavior recognition in surveillance videos based on deep learning methods," *J. Real-Time Image Process.*, vol. 18, no. 5, pp. 1669–1679, Oct. 2021, doi: 10.1007/s11554-021-01116-9.