International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

Human Security: Intelligent Video Surveillance for Criminal Activity Detection

Sharnav Mahajan[1], Pratham Shewale[2], Soham Patil[3], Jai Jadhav[4], Deepali Hajare[5], Sunayna Sutar [6]
[1,2,3,4]Student, Dept. of AI&DS, Dr. D.Y. Patil Institute of Engineering Management and Research, Akurdi, Pune, India
[5,6]Assistant Professor, Dept. of AI&DS, Dr. D.Y. Patil Institute of Engineering Management and Research, Akurdi, Pune, India
Emails: *sharnav123@gmail.com[1], s.pratham0100@gmail.com[2], pactilsoham@gmail.com[3], jadhavjai7878@gmail.com[4], deepalihajare1204@gmail.com[5], sunayanasutar28@gmail.com[6]*

Abstract
As cities keep expanding, ensuring public safety will become increasingly difficult considering the complexity and unexpectedness of crime, the density, and dynamism of crowds. Traditional public safety surveillance technologies have been subject to high false alarm rates and lack contextual awareness, which makes monitoring inefficient and reactions slower. The evolution of AI in modern computing, especially video processing through deep learning, opens new directions for building adaptive and efficient crime-detection systems that will build capabilities of threat identification and behaviour analysis in real time. This review of relevant academic research underlines state-of-the-art work on deep learning and surveillance systems, with convincing evidence for their ability to reduce false alarms and improve accuracy in real public settings. The review points to the scalability and interpretability of AI systems as key to supporting security personnel, communities, and society with reliable and automated decision-making. The review discusses how contextual interpretation, body pose estimation, behavioural modelling, and multilayered threat detection advance public safety. Eventually, this work introduces a novel intelligent surveillance architecture capable of applying deep learning techniques, as mentioned above, to establish a reliable crime detection system for urban capacity. The larger goal is to adapt best-in-class safety infrastructure with next-generation AI capability and deploy these into practice for decision-making reliability.

Keywords: *Behavioral Modeling, Crime Detection, Deep Learning, False Alert Reduction, Pose Estimation, Public Safety, Real-Time Monitoring, Smart Video Surveillance, Threat Identification, Urban Security.*

1. Introduction

The fast-changing urban context today has shifted the paradigm for public safety approaches from discrete police enforcement activities toward city-wide, integrated networks of security vigilance. Urban areas are not only interconnected but also subject to increasing instability or disruption that represents genuine vulnerabilities, heightened during crises and circumstances of mass social disorder when earlier modes of monitoring fell apart [19]. A key challenge in these contexts is cascading escalation whereby, through the feeds of urban surveillance activities, minor changes in crowd behavior, unclear movements or a single event of suspicious behavior/activity led to an exaggerated level of alarm and severity across the entire monitoring ecosystem. When regarded as an operational inefficiency, the ramping up of a city-wide layered surveillance system seems wasteful, as in the unintended consequence of generating too many alerts, taxing or misusing resources, and distracting from enforcement capabilities [17]. In this light, the resilience of surveillance systems, or the ability of surveillance systems to absorb, filter, and adapt to disruptions, would be the first and foremost concern for any solution maker who has ambitions to create or maintain a safe and competitive urban space. Large cities are utilizing sophisticated video analytics and automated tools to respond to threats proactively,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

while smaller towns and municipalities or organizations lack the software, technical knowledge, and funds to implement or maintain full-on AI-centered surveillance solutions [9]. There is a disparity between large and small jurisdictions that leads to a significant gap in security experiences, even as community-level institutions are the foundation of urban public safety networks. A burdensome industry report makes note of this dilemma, as we see, "92% of local agencies believe AI-powered surveillance will revolutionize safety responses," yet "60% do not know what tools to use, 58% do not have the budget, and 55% do not have internal capabilities." The rationale for this review, then, is a need for research to consolidate AI-centric video analytics and multimodal crime detection methodologies that are modular, easily digestible, and focused on practitioners within public safety [11, 20]. To resolve these challenges, new studies have shifted from typical central monitoring paradigms (which are normally too static for the pace of today's urban environments) to decentralized, AI-based threat detection frameworks [10]. Smart surveillance architectures are now widely regarded as complex adaptive systems: dynamically interacting networks of independent models (i.e., cameras, edge devices and analytic nodes) whose collective behavior is often unknown and emergent. In turn, multi-model video analytics ecosystems have begun to emerge as "unparalleled and robust tools" for the modeling of urban security that observe as real or autonomous agents would within the surveillance ecosystem of a city's governance. Some significant technological advances in this space in the last few years are the use of:

- Deep learning improves computer vision and behavior pattern recognition, with models such as YOLO [2], MobileNet [10], and LSTM [3, 15], helping provide increased precision and capabilities for threat detection and contextual awareness.
- Large Language Models (LLMs) now function in the same surveillance ecosystem as autonomous units, allowing for automated assessment of video behavior detection, automated interpretation of behavior, and more explainable automated event detection with increasingly sophisticated reasoning.
- Sequential pattern recognition algorithms, including Temporal Convolutional Networks and Graph Neural Networks [1] which enable the system to explore large volumes of video streams, and determine robust alerting strategies that account for complex uncertainty and vagueness in the environment.

One common thread in the research literature is the evolution toward integrated paradigms that have connected these frameworks into a multilayered surveillance system [13, 14]. These systems potentially enabled by technology leverage decentralized and autonomous characteristics of multi-unit platforms to provide resilient and adaptive monitoring, embedding artificial intelligence (AI) models to enable intelligent and interpretable decision making about alerts, interventions, and evidence collection [12]. This review attempts to synthesize the literature to evaluate the empirical literature on the potential for an AI-Native, multi-unit surveillance system for urban safety by differentiating between coordinated architectural units: Data Ingestion, Contextual Analytics, Behavioral Prediction, and Alert Management under a central AI Orchestrator [17]. The intended plan for this model is to deploy a low-cost, scalable, interpretable plan for automating surveillance functions, from data ingestion through real-time intervention and evidence management [15, 16]. This paper aims to systematically review the literature on the application of AI-powered smart surveillance in a variety of urban settings to ascertain the technological and operational factors impacting AI-assisted surveillance [19] and finally suggest future areas of exploration for reasoned and ethical application. This paper provides a roadmap to democratize the deployment of sophisticated crime detection tools for community safety, such as those utilizing 4D convolutional networks for long-range interactions [18], new levels of predictability for the stabilization of security management, and potentially long-term economic sustainability of urban safety infrastructure.

2. Literature Review

The current section encompasses a summary and critical assessment of prior research work that is linked to real-time crime analysis using the assistance of deep learning-enabled surveillance systems. The reviewed studies are assessed based on a set of key dimensions, such as dataset type, used learning model, target detection, evaluation method, system limitations, and relevance to the proposed study [19]. A shared theme of almost all literature is that from traditional CCTV frameworks, which mostly depend on the human element to carry out their tasks of monitoring, intelligent paradigms of surveillance are being developed that will automatically analyze videos for crime detection [12, 17]. Various real-world deep learning methods have been applied for the identification of suspicious objects, violent behaviors, and suspicious dynamic movements [10, 20]. A large part of the related literature uses a modular or component-based framework where different modules execute tasks like object detection, action recognition, tracking, and anomaly detection [13, 14]. For establishing real-time feasibility, performance metrics like accuracy, precision, recall, and processing time are commonly used [3, 12]. However, issues such as high computational requirements, degraded performance due to poor lighting, occlusion, and scalability of systems to metropolis settings are also identified [4, 17]. Despite these limitations, the overall evidence indicates that multicomponent deep learning integration results in more reliable and effective crime monitoring [16, 18]. The current trajectories of research emphasize the need to shift towards decentralized and edge-centric surveillance architectures that are both faster and more robust [8, 9]. This implies that much of the processing of the video is done near the source, reducing latency and network dependencies while reinforcing privacy [10, 15]. The literature has supported the development of an integrated deep learning-based framework that is interpretable, scalable, and can offer real-time crime monitoring; as such a system has substantial potential for enhancing public safety by allowing for quicker identifications, better situational awareness, and timely responses against criminal activity [11, 19]

Table 1 **Literature Review**

Sr	Paper Title	Dataset Used	Key Features	Models/ Algorithms	Evaluation Parameters	Research Gaps/ Limitations	Relevance to Proposed System
1	Weapon detection with FMR-CNN and YOLOv8 for enhanced crime prevention and security	Five-class annotated CCTV dataset (guns, blades, etc.)	Hybrid FMR-CNN + YOLOv8; integrates segmentation & detection; achieves ~98.7 % accuracy and AP ≈ 90.1; runs at ≈ 9 FPS	FMR-CNN (Faster + Mask R-CNN), YOLOv8, MobileNetV3	Detection accuracy, AP, FPS	Manual CCTV monitoring causes fatigue and bias; small-object detection still challenging; limited dataset	Provides architectural foundation for AI orchestration; integration of segmentation and detection aligns with contextual filtering requirements
2	An Efficient Deep Learning-Based Firearm Detection System Using Yolov8	Custom dataset (size not disclosed)	Combines CNN for person detection, YOLOv8 for weapons, pose estimation.	Custom CNN, YOLOv8, Amazon Recognition, VGG16	Detection accuracy (not specified)	Dataset unspecified; integrated pipeline is complex and resource-intensive	Demonstrates feasibility of complex multi-stage pipelines; supports use of computationally intensive model optimization

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

							methods
3	RTVD-Net: A real-time violence detection method based on pre-training of human skeleton images	Kaggle dataset (~4 156 images of knives & handguns)	Real-time detection via YOLOv8; accepts images, videos and live webcam; reported ~64 % accuracy	YOLOv8 with Python/Flask front-end	Accuracy, precision, recall and F1-score	Limited to two classes (knife & handgun); moderate accuracy; generalization beyond the Kaggle dataset is uncertain	Highlights need for efficiency improvements in real-time operations; optimizing redundant evaluations in forecasting
4	A Smart Surveillance System to detect modern guns using YOLOv5 Algorithm: A Deep Learning Approach	Dataset of firearms and edged weapons (size not specified)	YOLOv8 model trained for real-time weapon detection; reports high precision, recall and F1 with strong map	YOLOv8	Precision, recall, F1, map	Dataset size and diversity not disclosed; lacks contextual filtering or behavior analysis	Serves as benchmark for real-time detection accuracy; supports modular design for scalable MSME-friendly application
5	Intelligent Weapon Detection System for Real-Time Surveillance Using Deep Learning	Custom dataset of 16 799 images of firearms & sharp weapons	Distributed IoT system; YOLOv5s detects weapons; 3D CNN classifies scenes as robbery or normal; reduces false positives	YOLOv5s, 3D CNN	map ≈ 0.87; FPS ≈ 4.43; 3D CNN accuracy ≈ 0.88	Dataset focused on robbery scenes; generalization to other crime types may be limited	Validates distributed IoT-based surveillance architecture; inspires incorporation of scene classification and false positive reduction techniques
6	Real-Time Weapon Detection Using YOLOv8 for Enhanced Safety	UBI Fight & UCF Cri me video datasets	Hybrid model transforms 2D kernels into 3D; uses transfer learning (ResNet-18 & Google Net) and bilinear LSTM to capture temporal features	3D CNN with ResNet-18/G oogle Net, bilinear LSTM	Accuracy metrics improved via 3D kernels & LSTM	Requires large video datasets and high computational cost; limited to fight detection contexts	Supports hybrid modeling approach and temporal feature extraction; relevant for scalable real-time monitoring
7	Distributed Intelligent Video Surveillance for Early	Dataset not specified (custom weapon	Uses YOLOv4 with data augmentation for real-time CCTV weapon	YOLOv4	Precision, recall, F1 (approx. 0.86)	Dataset details not disclosed; potential issues with occlusion and	Reinforces multi-agent decentralization concept; aligns with Industry 4.0

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

	Armed Robbery Detection	images)	detection; achieves F1 ≈ 0.86			class imbalance	urban safety paradigms
8	AI-Based Surveillance Framework for Physical Violence Detection	Dataset not specified	Integrates YOLOv5 with IoT & cloud computing via Software-Defined Networking (SDN); reports high map (≈97.3), precision (96.8), recall (95.6) and F1 (96.2)	YOLOv5, IoT & SDN framework	map, precision, recall, F1	Dataset not described; complex integration may hinder deployment and reproducibility	Provides framework for IoT and cloud-enabled distributed systems; supports multi-criteria alerting and behavioral analysis
9	Weapon Detection in Real-Time CCTV Using Deep Learning	UCF Crime & NTU CCTV fight datasets	CCTV-based violence detection system with feature extraction and abnormal-behavior recognition	Feature extraction & behavior recognition (model unspecified)	Evaluation metrics not reported	Limited dataset scope: performance metrics unspecified; system may not generalize beyond the studied datasets	Supports AI forecasting and Explainable AI objectives; informs design of interpretable threat detection modules
10	YOLOv5-Based AI-Enabled IoT and Cloud Computing with SDN for Real-Time Weapon Detection	Novel handgun dataset (details not provided)	YOLOv5 with data augmentation, transfer learning and test-time augmentation; shows high precision and recall and outperforms Faster R-CNN	YOLOv5 (with augmentation), compared to Faster R-CNN	Precision & recall improvements vs Faster R-CNN	Dataset details unavailable; results may not generalize to other weapons or environments	Strong market validation for accessible, low-cost AI solutions; aligns with focus on interpretable and scalable architectures

3. Methodology of the Review

This section of the article highlights to facilitate transparency and rigor for the review the systematic methodology for literature selection, organization, and review. To examine the core themes of AI-Native multi-model deep learning-based video surveillance systems, and their operational uses in dynamic urban environments (Table 1).

3.1. Identifying Literature and Criteria

The literature review consists of several significant white papers and reports from industry sources focused on artificial intelligence related to the public safety sector. The inclusion criteria were:

- **Timeframe**: Works published from 2018 to 2025, capturing the more recent developments associated with intelligent systems of surveillance, such as 4D convolutional blocks for long-range

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

interactions **Error! Reference source not found.** and high-accuracy integrated Bi-LSTM frameworks **Error! Reference source not found.**.

- **Inclusion Criteria**: This includes peer-reviewed journals and conference papers published in English that focused on AI, deep learning, multi-model systems, real-time video analysis **Error! Reference source not found.**, anomaly detection **Error! Reference source not found.**, behavioral modeling **Error! Reference source not found.**, and urban security applications **Error! Reference source not found.**.
- **Exclusion Criteria**: Non-peer reviewed literature; articles not published in English; theoretical papers without examples of practical implementation; and computer vision studies. were completely unrelated to AI or intelligent surveillance systems.

3.2. Categorization and Analytical Approaches

To organize the results of selected articles, results were Grouped into categories, which were analytically useful in providing a clear, systematic framework in which the different findings can be highlighted: across various aspects of crime detection.

- **Foundational Frameworks and Reviews**: These articles provide holistic perspectives on smart video deep learning integrated surveillance systems **Error! Reference source not found.** and multi-model frameworks applied for the improvement of urban security and sustainable eco-systems [11, 14].
- **Specific AI Models and Algorithms**: This category encompasses studies that develop and evaluate state-of-the-art algorithms, including Convolutional Neural Networks (CNNs) for spatial features **Error! Reference source not found.**, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for temporal dependencies [3, 15], and Graph Neural Networks (GNNs) for skeleton-based recognition **Error! Reference source not found.**.

Section 4 includes a suitably condensed and comparative table of the synthesized literature review findings. The table compares all articles related to participant methodology, evaluative criteria of innovativeness, and gaps in research literature.

3.3. Proposed System Architecture/Model

This paper presents an advanced generation AI-Native multi-component system architecture for city-wide crime detection and public safety management. In this architecture, the major constituent is the AI Orchestrator that coordinates a set of specialized autonomous modules. Each module operates its decision-making loop, which feeds into the resilient, flexible, and distributed surveillance capability of the architecture [17, 21] While some of the key modules constituting the architecture are as follows:

- **Data Ingestion**: It captures live videos from city surveillance. This can enable investigators to have updated information relevant to the case using edge-centric data ingestion in order to cut latency [9, 10].
- **Context analysis**: Applies modern AI to interpret scenes and behaviors. This module transforms unactionable data into actionable insights using models like YOLO for person detection [2, 10] and MediaPipe for structural skeleton extraction (Figure 1).

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

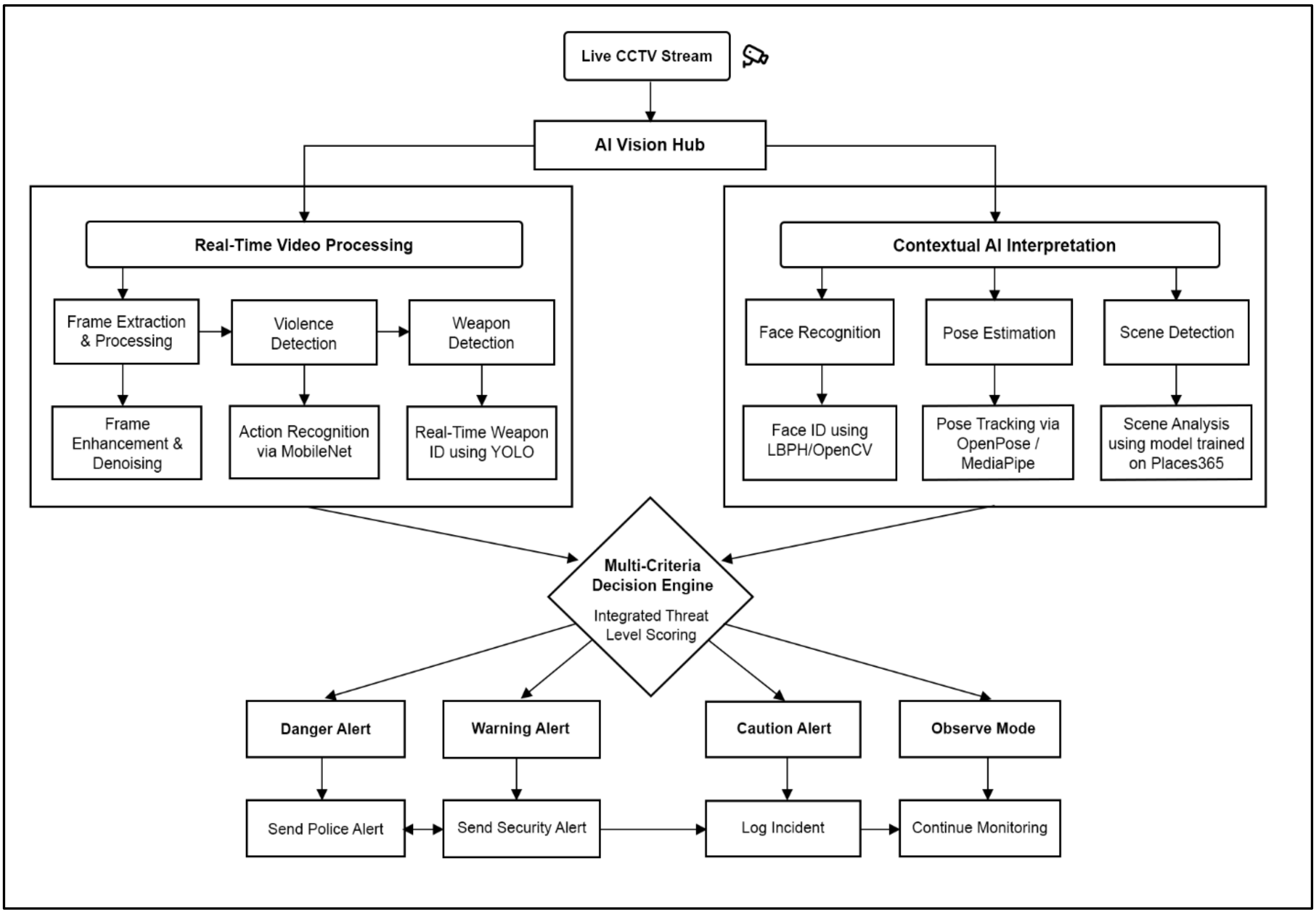

Figure 1 **AI-Native Multi-Model Architecture for Smart Crime Monitoring System**

- **Behavioral prediction**: Applies modern AI to interpret scenes and behaviors. This module transforms unactionable data into actionable insights using models like YOLOv3 for person detection [2, 10] and OpenPose for structural skeleton extraction [15, 17].
- **Alert handling**: Ranks and prioritize alerts to cut false alarms and suggest resource deployment. This is achieved through hybrid deep learning models, such as CNN-BiLSTM frameworks, which improve the precision of violence detection in real-time CCTV feeds [17, 21].
- **Mobilize response**: Begins automated reports to police and private security, compiles evidence, recommends intervention and keeps first responders apprised of situational awareness and asset/resource availability.
- The proposed framework is designed to bridge the gap between high-end research and practical, resource-efficient deployment. Its core functionality is defined by the following characteristics:
- **Resource-Efficient Scalability**: The framework is developed to be inexpensive to grow and interpretable, fulfilling the need for adoptable technology in resource-constrained municipalities [9, 15]. These design considerations ensure a viable, AI-assisted solution for crime-related decision-making that does not require massive high-end server farms, making it suitable for decentralized edge-centric

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

architectures **Error! Reference source not found.**.

- **Pose-Based Behavioral Analysis**: The system utilizes MediaPipe or OpenPose to extract human skeletal data [10, 15]. This approach allows the model to focus on human kinematics rather than raw pixel data, enabling it to distinguish between high-risk behaviors (e.g., brandishing a weapon) and non-threatening postures through precise joint-angle analysis [16, 17].
- **Multi-Condition Alert Logic**: To drastically reduce "alert fatigue," the system only notifies police when multiple conditional requirements are fulfilled simultaneously. These include detected weapon presence in a context where an individual is displaying threatening skeletal postures and a crowd is present, ensuring high-priority response for mass-casualty threats [13, 17].
- **"Double-Check" Verification** – Time-Based: The design also incorporates the aspect of waiting for a few seconds before taking any necessary actions to alert police officers. Time windowing in the design helps the algorithm to identify whether the suspicious body motion is a case of violence or a "fluke" situation (for example, a random non-violent body motion) [18, 21].
- **Actionable Intelligence & Evidence Management**: The notifications that go out to law enforcement are not simple pings; rather, they include precise GPS coordinates and real-time image snapshots coupled with high-fidelity timestamps **Error! Reference source not found.**. Precisely, this complete evidence management enables the first responders to achieve maximum situational awareness and verify the severity of the incident before arrival at the scene [11, 20].

4. Thematic Synthesis of the Literature

4.1. Theme 1: Multi-Model Systems Transformation in Urban Crime Detection

A common thread in recent studies is that urban surveillance systems are shifting from centralized, monolithic orientations toward decentralized, dynamic multi-unit systems -. These are composed of many autonomous parts, such as independent pieces of software, which interact as part of complex, adaptive networks all forming emergent behaviors that might be difficult to predict or manage by traditional approaches. MAS provides flexibility and scalability to surveillance by enabling so-called heterogeneous sensors, video analytics modules and decision models to cooperate, but independently [10, 11]. Real-world surveillance actors-cameras, analytic nodes, and law enforcement dispatchers -are counterparts of models that would reside within the MAS architectures, promoting bottom-up modeling approaches to emulate the decentralized realities of urban surveillance. Research accelerated since 2021, with enhancements by smart cities and ethical A.I. frameworks, positioning MAS as the foundation for resilient and adaptive infrastructures of crime detection [6, 7]. The traditional framework categorizes models by role, including data collectors, analyzers, response coordinators, which gives way to a modular framework that merges architecture for the real-time processing of data with strategic, predictive analyses. MAS frameworks are promising forms of adaptability, but there are research gaps in extending to citywide deployments, data integration of diverse real-time flows, and socio-technical complexity.

4.2. Theme 2: AI-Powered Behavior Prediction: The Shift from Traditional Analytics

The world of crime prediction is moving away from older statistical methods toward AI-based forecasting and behavior prediction. Current deep learning models such as LSTM [3, 15] and Transformer models are far better than conventional linear or rule-based methods at recognizing time and place patterns in suspicious behavior [12, 21]. Combining traditional statistical approaches and the capabilities of neural networks (i.e., hybrid models) is also better, particularly in messy, high-dimensional data, like problems with surveillance video images [13, 14]. However, deep learning systems are often not interpretable and operate as "black boxes," which makes trust in their decisions for important public safety decisions difficult. Therefore, active research exists today to develop the concept of explainable

IRJAEH
International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 02 February 2026
Page No: 385-396
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0053

artificial intelligence (XAI) to allow the maintenance of the prediction capability of the model while making better model decisions [17,19].

4.3. Theme 3: Advanced Optimization and Decision Algorithms

- Cost-effective decision support systems are required in complex urban setups to address optimization in the allocation of resources and alert systems. Reinforcement learning and the use of Monte Carlo Tree Search techniques are found to be effective alternatives in the allocation of resources in scenarios where the state space is complex and uncertain [17].
- These methods can help improve resource allocation by reducing false positive detections and allowing for actions/interventions to take place on time against dynamic situations [21]. Simulated results give performing algorithms great promise, but real-world testing of these algorithms to evaluate resource allocation remains limited, presenting an important opportunity for future empirical testing [19].

4.4. Theme 4: General Modular Pipeline of Reviewed Systems

- The system begins with live videos, including security cameras and typical streaming feeds [9, 10].
- We process the images next, so the computer recognizes them. We first take each frame from the videos; images are then rescaled to a fixed size and augmented with minor perturbations to ensure the system learns successfully [13, 14].
- Performance improvement, namely a reduction in false alarms, the system is paired-capable with auxiliary devices and correlates different cues. To provide more reliable reading [5, 12].
- To provide more reliable readings and reduce false alarms, the system is capable of correlating different cues [17, 21].
- We are able to analyze the position of individuals or what they are doing with their bodies through body tracking tools such as OpenPose or MediaPipe [10, 15, 16]. This enables us to understand interactions in real-time.
- The alert system is activated when the system detects something that warrants fast action; for example, alerting security or emergency personnel [11, 20].
- The system displays rectangular boxes around detected objects, showing alerts on the screen and recording each event [12, 16]. It is designed to handle distracting real-world situations and locate focal items while attenuating surrounding noise [18, 19].

4.5. Theme 5: Tabular Performance Analysis of Methods

Table 2 shows the Comparative Performance Study.

Table 2 Comparative Performance Study

Model/ Approach	Best Accuracy	Map	Speed /FPS	Key Advantages	Main Limitation
FMR-CNN + YOLOv8	98.7%	90.1	9.2 fps	High precision + segmentation	Complex architecture
YOLOv8 (Firearm)	98.4%	64%	Real- time	High accuracy on VGG classifier	YOLO Underperforms

YOLOv5 + Ensemble	98.0%	81.8	0.78s Detection	Lightweight, efficient	Limited dataset
YOLOv5s + 3D CNN	-	87%	4.43 fps	Scene classification (robbery)	Dataset specific
YOLOv5 + SDN	-	97.3	Fast	IoT-cloud synergy	Hardware dependent
RTVD-Net (YOLO-Pose)	-	-	-	Skeleton fusion	Evaluation missing
YOLOv4	86% (F1-score)	-	-	Simplicity	Outdated model
YOLOv5 (Handgun)	-	-	Fast	Augmented training	Dataset unknown

5. Future Directions and Research Gaps

We will also need systems that make use of smart software and easy-to-understand visual displays of information to create better crime-fighting tools [11, 20]. This should be accompanied by an important shift in the development of programs police can actually understand, building their confidence and demonstrating increased rates of utilization **Error! Reference source not found.**. Innovations today offer secure ways to share information between agencies and protect privacy when collaborating on these databases innovations that present enormous possibilities for collaboration in ways that improve the working conditions of professionals across agencies; it's about working together **Error! Reference source not found.**. In addition, increasing the diversity of people, places, and circumstances included in training data will improve accountability and fairness in AI detection systems **Error! Reference source not found.**. However, despite this knowledge, there are significant gaps in research in this rapidly evolving domain. For example, there are challenges around tackling algorithmic bias that can lead to unfair targeting, improving transparency and interpretability of the AI model to create trust in the automated decisions of human operators, and designing and implementing privacy-protective solutions for safeguarding sensitive personal data while people are surveilled [17, 19]. Further, ethical frameworks have not yet been incorporated in the design and deployment of these systems, and the applied functionality and interface customization for users using the identified distinct role development process is yet to be refined for real and effective use [15, 21]. Bringing together different data sources, including video observation, audio sources, geo-location data, and social media, enhances and speeds up understanding of the context at that moment and can act as a warning system to anticipate risk [4, 13]. User interfaces that provide different levels of access for various stakeholder's officers on the street, investigators, and command-level can help streamline the flow of data into aspects of actionable intelligence [10, 20]. Thoughts will need to be given to bias and ethical considerations to avert the possibility of technological adaptation simply reproducing social bias **Error! Reference source not found.**. As appropriate, it would be important to contextually integrate, marry, or cross-pollinate the technical considerations of data analysis and data-informed technology with the legal, social, and policy perspectives **Error! Reference source not found.**.

Interdisciplinary partnerships lend themselves to ensuring that technological adaptation is producing a beneficial step toward healthy communities and a better future.

Conclusion

The following article discussed how the use of deep learning systems has impacted crime detection and responses. Such advanced artificial intelligence, with timely visualizations using video, sensors, and other data, provides police and other decision-makers with clearer, faster, more actionable situational awareness compared to traditional crime prevention models [4, 13]. Despite these benefits, important challenges still remain. There must be normative methods for judging these systems' effectiveness, and new tools must be developed to increase the trust in understanding AI-driven decisions **Error! Reference source not found.**. Society still experiences such issues as data bias, a variety of privacy protections are in need, and there must be solutions with respect to what constitutes a user 'need' **Error! Reference source not found.**. Improvement was needed in normative methods to establish this process, whereby varied systems function properly, fairly, transparently, and ethically **Error! Reference source not found.**. The best way to advance from here is by working across disciplines including technology, law, ethics, and working with members of the community **Error! Reference source not found.**. We also need to compile comprehensive datasets that fairly represent diverse people and situations, design AI systems to be explainable by default, and keep ongoing conversations between technologists, practitioners, and affected communities [17, 19]. These collaborative efforts will facilitate the translation of promising ideas into trusted practices and tools that help improve safety in neighborhoods. Linked together, improved professional ethical and legal frameworks for safety along with more sophisticated crime-oriented detection systems serve to enhance public safety, justice, and the prospect of safer and just communities **Error! Reference source not found.**. In deploying these technologies, we build a future through lived experience in which technologies enhance human vigilance but also compassion to create more safety, more connectedness, and healthier communities.

References

[1]. Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1).
[2]. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767.
[3]. Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., & Baik, S. W. (2018). Action recognition in video sequences using deep bi- directional LSTM with CNN features. IEEE Access, 6, 1155–1166.
[4]. Zhou, P., Ding, Q., Luo, H., & Hou, X. (2019). Violence detection in surveillance video using low-level features. PLoS ONE, 14(10), e0223247.
[5]. Carreira, J., & Zisserman, A. (2019). Quo vadis, action recognition? A new model and the kinetics dataset. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 6299–6308.
[6]. Li, C., Zhong, Q., Xie, D., & Pu, S. (2020). Collaborative spatiotemporal feature learning for video violence detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 214–222.
[7]. Singh, G., & Lee, Y. J. (2020). End-to-end localization and ranking for relative attributes. European Conference on Computer Vision (ECCV), 753–769.
[8]. Wu, Z., Wang, X., Jiang, Y. G., Ye, H., & Xue, X. (2020). Revisiting RCNN: On awakening the classification power of Faster RCNN. European Conference on Computer Vision (ECCV), 318–334.
[9]. Hossain, M. S., & Muhammad, G. (2021). Cloud-assisted Industrial Internet of Things (IIoT)– enabled framework for health monitoring. IEEE Internet of Things Journal, 8(5), 3578–3586.
[10]. Shindhe, D., Govindraj, S., & Omkar, S. N. (2021). Real-time Violence Activity

Detection Using Deep Neural Networks in a CCTV Camera. Uses Open Pose for pose estimation, YOLOv3 for person detection, CNN on skeleton data to classify actions. Attempts lightweight real-time operation.
[11]. Srividya M. S. & Anala M. R. (2021/2022). AI-Based Surveillance Framework for Physical Violence Detection. Uses hybrid deep learning (3D-CNN, LSTM, transfer learning) to detect physical abuse, physical violence in surveillance contexts.
[12]. Das, A., Panda, R., & Roy-Chowdhury, A. K. (2022). Detecting violence in videos using attention-based neural networks. IEEE Transactions on Image Processing, 31, 2207–2219.
[13]. Bin Jahlan, H. M., & Elrefaei, L. A. (2022). Detecting Violence in Video Based on Deep Features Fusion Technique. arXiv:2204.07443. Combines Alex Net + Squeeze Net each followed by ConvLSTM, fusing states for classification. Tested on standard benchmark datasets.
[14]. Elakiya, V., Aruna, P., Puviarasan, N., & Suresh Kumar, R. G. (2022). Video Based Violence Detection Using Deep Learning CNN-CHA-SPA Double Attention Mechanism with Mosaicking. Proposes attention-based feature extraction with mosaicking to improve discriminative feature capture for violent behaviour in videos.
[15]. Partika, F. B. De Moura. (2022). Simple Approach for Violence Detection in Real-Time Videos Using Pose Estimation with Azimuthal Displacement and Centroid Distance as Features. Uses pose estimation features (from Open Pose) transformed to simpler spatial features; then LSTM of these features for violence detection.
[16]. Detection of Violent Behavior Using Neural Networks and Pose Estimation (Kevin B. Kwan-Loo et al., 2022). IEEE Access. Extracts joint-angle based features using pose estimation; classifies behaviour; proposes dataset Kranok-NV, high accuracy.
[17]. Üstek, İ., Desai, J., López Torrecillas, I., Abadou, S., Wang, J., Fever, Q., Kasthuri, S. R., Xing, Y., Guo, W., Tsourdos, A. (2023). Two-Stage Violence Detection Using ViTPose and Classification Models at Smart Airports. Uses ViTPose for human pose estimation → CNN-BiLSTM classifier of key point sequences for violence detection, aiming to reduce false positives.
[18]. Violence 4D: Violence detection in surveillance using 4D convolutional neural networks (2023). Proposes “Violence 4D”, using four-dimensional blocks across video (inter-clip) to capture long-range spatio-temporal interactions, improving over 3D CNNs for real-world violence benchmarks.
[19]. Negre, P., Alonso, R. S., González-Briones, A., Prieto, J., & Rodríguez-González, S. (2024). Literature Review of Deep- Learning-Based Detection of Violence in Video. Sensors, 24(12):4016. This is a recent survey of methods including CNN, CNN+LSTM, 3D-CNNs, etc., highlighting datasets, accuracies, and challenges in real-world violence detection.
[20]. Jam al Kolaib, R., & Waleed, J. (2024). Crime Activity Detection in Surveillance Videos Based on Developed Deep Learning Approach. Uses pre-trained CNN, etc., on surveillance videos to detect crime-related activity, aiming for generalizable detection.
[21]. Sumeet Kothari & Mukesh Kumar. (2025). Deep Learning-Based High-Accurate Violence Detection for Intelligent Video Surveillance Systems. Integrates CNN and Bi-LSTM models; reported high accuracy; focuses on reliability and reducing false positives.