Multimodal Deep Learning Framework for Intelligent Traffic Signal Control with Emergency Vehicle Prioritization

P.T.H. Pathirana[1], R.J.wellassa[2], M.A.A. Karunarathna[3]

[1,3]Department of Electronics, Wayamba University of Sri Lanka

[2]Physical Education Unit, Wayamba University of Sri Lanka

DOI: https://doi.org/10.51584/IJRIAS.2026.11070083

Received: 18 July 2026; Accepted: 23 July 2026; Published: 04 August 2026

ABSTRACT

Urban traffic congestion and delayed emergency response times represent critical challenges in modern smart cities. This comprehensive review examines recent advances (2024-2025) in intelligent traffic management systems that integrate deep learning-based vehicle detection with adaptive signal control mechanisms, specifically focusing on emergency vehicle prioritization. We analyze 30 state-of-the-art systems that leverage YOLOv9 architectures, multimodal fusion techniques, and edge computing platforms to achieve real-time traffic optimization. Key findings reveal a paradigm shift toward attention-enhanced detection models (YOLOv9+CBAM), audio-visual fusion for robust emergency vehicle identification, and edge deployment on resource-constrained hardware (Raspberry Pi, Jetson platforms). Systems employing multimodal confirmation mechanisms demonstrate superior reliability, with reported accuracies exceeding 96% and response time reductions of up to 35%. However, standardized benchmarking for false positive rates remains limited. This review synthesizes architectural innovations, prioritization strategies, and performance characteristics to provide a comprehensive framework for researchers and practitioners developing next-generation intelligent transportation systems. We identify critical research gaps and propose future directions toward more reliable, scalable, and context-aware traffic management solutions.

Keywords: Traffic optimization, Machine learning, Emergency vehicle detection, YOLO, Computer vision, Traffic signal control

INTRODUCTION

Urban traffic congestion has emerged as one of the most pressing challenges facing modern cities, with profound implications for economic productivity, environmental sustainability, and public health. The World Health Organization estimates that delayed emergency medical response due to traffic congestion contributes to thousands of preventable deaths annually. Traditional fixed-cycle traffic signal systems, designed decades ago, are fundamentally inadequate for managing the dynamic, high-density traffic patterns characteristic of contemporary urban environments.

The convergence of artificial intelligence, computer vision, and Internet of Things (IoT) technologies has catalyzed a paradigm shift in traffic management. Recent advances in deep learning, particularly the YOLO (You Only Look Once) family of object detection models, have enabled real-time vehicle detection and classification with unprecedented accuracy and computational efficiency [1], [2], [3]. The release of YOLOv9 in 2024 introduced architectural innovations, including Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN), significantly improving detection performance while maintaining real-time processing capabilities [1], [4].

Simultaneously, the proliferation of edge computing platforms—ranging from resource-constrained single-board computers (Raspberry Pi) to powerful embedded GPUs (NVIDIA Jetson)—has made it feasible to deploy sophisticated AI models directly at traffic intersections, enabling low-latency decision-making without dependence on cloud infrastructure [5], [6], [9]. This distributed intelligence architecture represents a fundamental departure from centralized traffic management systems, offering improved resilience, reduced latency, and enhanced scalability.

Research Objectives

This comprehensive review addresses the following research objectives:

1. Systematically analyze the architectural innovations in YOLOv9-based traffic management systems developed during 2024-2025, with emphasis on attention mechanisms, multimodal fusion, and edge optimization techniques.

2. Characterize emergency vehicle prioritization mechanisms, including detection modalities (visual, audio, RFID, GPS), signal control strategies (green corridors, dynamic preemption, adaptive timing), and decision logic frameworks.

3. Evaluate system performance across multiple dimensions: detection accuracy, processing speed, response time improvements, and false positive management strategies.

4. Identify critical research gaps, practical deployment challenges, and opportunities for future innovation in intelligent traffic management systems.

Scope and Contributions

This review synthesizes findings from 30 peer-reviewed publications from 2024-2025 that specifically address intelligent traffic management with emergency vehicle prioritization using YOLOv9 and related deep learning architectures. Our analysis encompasses systems deployed across diverse geographic contexts (Asia, Middle East, North America) and hardware platforms (cloud, edge GPU, single-board computers).

Key contributions include:

- Comprehensive taxonomy of detection architectures, categorizing systems by model architecture, hardware platform, and technical innovations

- Structured comparison of emergency vehicle prioritization mechanisms and signal control strategies

- Quantitative synthesis of performance metrics, identifying best practices and performance benchmarks

- Critical analysis of reliability challenges, particularly false positive management and multimodal confirmation strategies

- Forward-looking recommendations for standardization, advanced integration, and scalable deployment

Background and Theoretical Foundations

Evolution of Intelligent Traffic Management

Traffic signal control has evolved through several distinct paradigms. First-generation systems employed fixed-cycle timing based on historical traffic patterns, offering no adaptability to real-time conditions. Second-generation adaptive systems introduced vehicle-actuated control using inductive loop detectors, enabling basic responsiveness to traffic demand. Third-generation systems incorporated centralized optimization algorithms (e.g., SCOOT, SCATS) that coordinated signals across networks based on real-time sensor data.

The current fourth-generation paradigm leverages artificial intelligence and computer vision to achieve unprecedented situational awareness and decision-making capabilities. These systems can simultaneously detect, classify, and track multiple vehicle types; predict traffic flow patterns; and dynamically optimize signal timing at millisecond resolution [2], [11]. The integration of emergency vehicle detection and prioritization represents a critical capability that previous generations could not effectively address.

Deep Learning for Object Detection

The YOLO architecture revolutionized object detection by framing it as a single-stage regression problem, enabling real-time performance without sacrificing accuracy. YOLOv9, released in 2024, introduced two key innovations [1], [4]:

1. Programmable Gradient Information (PGI): Addresses information bottleneck problems in deep networks by generating reliable gradients through auxiliary supervision, improving learning efficiency and model convergence.

2. Generalized Efficient Layer Aggregation Network (GELAN): Optimizes network architecture for both parameter efficiency and computational efficiency, achieving superior performance with fewer parameters than previous YOLO versions.

These architectural advances make YOLOv9 particularly well-suited for traffic management applications, where real-time processing (>30 FPS) and high accuracy (mAP >90%) are essential requirements [9], [12].

Emergency Vehicle Detection Challenges

Detecting and prioritizing emergency vehicles in urban traffic presents several unique challenges:

Visual ambiguity: Emergency vehicles share visual characteristics with civilian vehicles, requiring fine-grained classification capabilities. Occlusion, varying lighting conditions, and diverse camera angles further complicate visual detection [1], [29].

Temporal constraints: Emergency response effectiveness is highly time-sensitive. Detection and signal control must occur within seconds to meaningfully reduce response times [4], [10].

False positive consequences: Incorrect emergency vehicle detection can disrupt traffic flow, reduce system trust, and potentially create safety hazards. Balancing sensitivity and specificity is critical [6], [29].

Environmental variability: Systems must operate reliably across diverse weather conditions, lighting scenarios, and traffic densities [18], [30].

These challenges have motivated the development of multimodal detection approaches that combine visual, acoustic, and communication-based modalities to improve reliability [4], [29].

Detection Architectures and Technical Innovations

YOLOv9-Based Detection Systems

YOLOv9 has emerged as the dominant architecture for real-time vehicle detection in recent traffic management systems. Among the 30 systems analyzed, 23 explicitly employ YOLOv9 or YOLOv9-derived architectures as their primary detection engine [1], [2], [4], [12], [14], [17], [21]. The architecture's appeal stems from its optimal balance of detection accuracy (typically mAP 85-95%), processing speed (30-74 FPS on edge hardware), and computational efficiency suitable for resource-constrained deployment [9], [12].

Several systems report architectural modifications to standard YOLOv9 for traffic-specific optimization. Korade et al. developed a length-based metric integration that enhances YOLOv9's ability to distinguish vehicle types based on dimensional characteristics, improving classification accuracy for emergency vehicles [12]. The system achieves real-time traffic light optimization by coupling YOLOv9 detection with adaptive timing algorithms that consider both vehicle count and type.

Attention Mechanisms and Model Enhancements

A significant trend in 2024-2025 research is the integration of attention mechanisms with YOLOv9 to improve feature selection and reduce false positives. Shitole et al. pioneered the YOLOv9+CBAM (Convolutional Block Attention Module) architecture, which augments YOLOv9 with spatial and channel attention mechanisms [1]. CBAM enables the model to focus on salient emergency vehicle features (light bars, markings, vehicle shape) while suppressing irrelevant background information. This attention-enhanced architecture demonstrates improved robustness in cluttered urban scenes where multiple vehicles and complex backgrounds challenge standard detection models [1].

The attention mechanism operates in two sequential stages: channel attention identifies "what" features are important (e.g., color patterns, light signatures), while spatial attention determines "where" to focus within the feature maps. This dual attention strategy has proven particularly effective for distinguishing emergency vehicles from visually similar civilian vehicles [1].

Beyond CBAM, several systems explore alternative attention architectures. Sabareesh et al. implemented a custom attention module optimized for traffic light adaptation, enabling the system to dynamically adjust focus between vehicle detection and traffic signal state monitoring [6]. This multi-task attention approach allows a single model to simultaneously handle detection and control optimization.

Multimodal Fusion Approaches

Recognizing the limitations of vision-only systems, recent research has increasingly adopted multimodal fusion strategies that combine visual detection with complementary sensing modalities. Zohaib et al. developed a sophisticated multimodal framework that integrates visual and acoustic information through ensemble learning [29]. Their system employs:

1. Multi-Level Spatial Fusion YOLO (MLSF-YOLO) for visual detection, incorporating multi-scale feature fusion to handle varying vehicle sizes and distances

2. Attention-based Temporal Spectrum Network (ATSN) for acoustic siren detection, using attention mechanisms to focus on characteristic emergency siren frequencies

3. Stacking ensemble learning to combine visual and acoustic predictions, requiring cross-modal agreement before confirming emergency vehicle presence

This multimodal approach achieved 96.19% accuracy with only 3.81% misdetection rate, significantly outperforming unimodal baselines [29]. The system demonstrates particular robustness in challenging scenarios where visual detection alone is unreliable (e.g., occlusion, poor lighting).

Rajesh et al. implemented a hierarchical multimodal fusion architecture combining YOLOv9 visual detection, ResNet-based audio classification for siren recognition, and RFID/GPS tags for redundant confirmation [4]. The system employs a rule-based decision module that requires at least two modalities to agree before triggering signal preemption, effectively reducing false positives while maintaining high sensitivity [4].

Fida et al. explored Ultra-Wideband (UWB) technology as an alternative to vision-based detection, deploying UWB readers at intersections and UWB tags on emergency vehicles [3]. This approach offers several advantages: immunity to visual occlusion, reliable operation in adverse weather, and precise distance estimation. However, it requires infrastructure investment and vehicle equipment, limiting immediate scalability [3].

Edge Computing and Hardware Platforms

The shift toward edge deployment represents a critical architectural trend, enabling low-latency decision-making and reducing dependence on network connectivity. Systems in our analysis employ three primary hardware categories:

Embedded GPU platforms: Raza et al. deployed their YOLO-based adaptive traffic control system on NVIDIA Jetson Xavier NX, achieving 74 FPS processing speed with approximately 90% mAP [9]. The Jetson platform's CUDA cores enable full-precision YOLOv9 inference without model compression, maintaining maximum detection accuracy. This approach is suitable for high-traffic intersections where processing throughput is critical [9].

Single-board computers: Multiple systems leverage Raspberry Pi 4B for edge deployment, particularly for control logic and decision-making rather than intensive vision processing [5], [6]. Fida et al. implemented their UWB-based system on Arduino Mega microcontrollers for traffic light operations, with a central cloud server handling data aggregation and analysis [3]. This hybrid edge-cloud architecture balances local responsiveness with centralized coordination [3].

Specialized edge devices: Al-Bahri et al. deployed a MobileNetV2-based detection model on ESP32-CAM, an ultra-low-cost edge platform with integrated camera and WiFi [30]. Using the FOMO (Faster Objects, More Objects) paradigm optimized for resource-constrained devices, they achieved real-time emergency vehicle detection suitable for developing regions with limited infrastructure budgets [30]. However, the ESP32-CAM's limited processing capability resulted in sporadic misclassification in low-light conditions, highlighting the performance-cost tradeoff [30].

The choice of hardware platform involves critical tradeoffs between detection accuracy, processing speed, power consumption, cost, and deployment complexity. High-performance embedded GPUs offer superior accuracy and speed but at higher cost and power consumption. Single-board computers provide adequate performance for many applications at moderate cost. Ultra-low-cost platforms enable widespread deployment but may sacrifice reliability in challenging conditions [9], [30].

Emergency Vehicle Prioritization Mechanisms

Detection Modalities

Emergency vehicle detection systems in our analysis employ four primary modalities, often in combination:

Visual detection remains the most common approach, leveraging YOLOv9 and related architectures to identify emergency vehicles through visual features (light bars, color schemes, vehicle markings, body shape) [1], [2], [4], [12], [14], [17], [21], [30]. Visual detection offers the advantage of leveraging existing traffic camera infrastructure without requiring additional sensors or vehicle equipment. However, it is susceptible to occlusion, lighting variations, and visual ambiguity [1], [29].

Acoustic detection identifies emergency vehicles through characteristic siren sounds. Systems employ deep learning classifiers (ResNet, ATSN) trained on siren audio to distinguish emergency sirens from ambient traffic noise [4], [29]. Acoustic detection provides complementary information to visual detection and can detect approaching emergency vehicles before they are visible to cameras. However, urban acoustic environments are complex, with multiple sound sources and reflections that can challenge detection reliability [29].

RFID/GPS communication enables direct vehicle-to-infrastructure (V2I) communication, with emergency vehicles broadcasting their position and status to intersection controllers [4], [5], [7]. This approach offers high reliability and precise positioning but requires equipping emergency vehicles with communication devices and deploying compatible infrastructure [3], [4].

Ultra-Wideband (UWB) ranging provides precise distance and direction estimation through time-of-flight measurements between UWB tags on vehicles and UWB readers at intersections [3]. UWB offers centimeter-level positioning accuracy and operates reliably in non-line-of-sight conditions, but requires dedicated infrastructure and vehicle equipment [3].

Signal Control Strategies

Systems employ three primary signal control strategies for emergency vehicle prioritization:

> Green corridor creation establishes a continuous path of green signals along the emergency vehicle's route by coordinating multiple intersections [3], [7], [16]. Fida et al. implemented dynamic green corridors using UWB-based vehicle tracking, with their system calculating optimal signal timing sequences to minimize emergency vehicle stops while managing impacts on cross-traffic [3]. This approach is most effective in coordinated signal networks but requires network-wide communication and control [3], [7].
>
> Dynamic signal preemption immediately overrides normal signal timing when an emergency vehicle is detected, turning the relevant approach green and conflicting approaches red [1], [2], [4], [5], [10], [18], [21], [30]. Al-Bahri et al. implemented a straightforward preemption strategy where emergency vehicle detection triggers immediate green signal activation for the corresponding lane, with automatic reversion to normal operation after vehicle passage [30]. This approach offers simplicity and rapid response but may cause significant disruption to cross-traffic [30].
>
> Adaptive timing optimization integrates emergency vehicle priority into broader traffic optimization frameworks, using reinforcement learning or optimization algorithms to balance emergency response with overall traffic flow [2], [11], [20]. Samiha et al. coupled YOLOv9 detection with deep reinforcement learning for dynamic signal timing that considers both emergency vehicle priority and network-wide traffic efficiency [2]. This sophisticated approach minimizes negative impacts on general traffic but requires more complex algorithms and computational resources [2], [11].

Decision Logic and Confirmation Systems

To mitigate false positive risks, advanced systems implement multi-stage decision logic with confirmation mechanisms:

Multimodal agreement: Systems require confirmation from multiple detection modalities before triggering signal preemption. Rajesh et al. implemented a hierarchical decision module that requires at least two of three modalities (visual, audio, RFID) to agree before activating priority control [4]. This redundant detection approach significantly reduces false activations while maintaining high sensitivity for genuine emergency vehicles [4], [6].

Temporal consistency: Systems track detections across multiple frames to confirm persistent emergency vehicle presence rather than responding to single-frame detections that may represent false positives [1], [18]. This temporal filtering improves reliability at the cost of slight detection latency [18].

Confidence thresholding: Detection systems employ confidence score thresholds, only triggering priority control when detection confidence exceeds specified levels (typically 85-95%) [1], [12], [30]. Higher thresholds reduce false positives but may miss some genuine emergency vehicles, requiring careful calibration [30].

Spatial validation: Systems verify that detected emergency vehicles are approaching the intersection and within relevant lanes before activating priority control [3], [7], [10]. This spatial logic prevents inappropriate responses to emergency vehicles on perpendicular roads or moving away from the intersection [3].

Comparative Analysis of System Performance

Detection Accuracy and Speed

Detection performance varies significantly across systems based on model architecture, hardware platform, and validation methodology. The highest reported accuracy comes from Zohaib et al.'s multimodal system at 96.19%, demonstrating the effectiveness of audio-visual fusion [29]. However, this system's processing speed is not reported, leaving questions about real-time feasibility. Raza et al.'s edge-deployed system achieved the highest reported processing speed at 74 FPS with approximately 90% mAP, demonstrating that edge GPU platforms can meet both accuracy and speed requirements [9].

Several systems report "real-time" performance without specific FPS metrics, making direct comparison difficult [1], [12], [30]. This lack of standardized reporting represents a significant gap in the literature, hindering objective performance assessment.

Response Time Improvements

Few systems provide quantitative data on emergency response time improvements, a critical metric for evaluating real-world impact. Fida et al. report a 35% reduction in ambulance response time in their UWB-based system deployment, representing one of the few concrete measurements of operational impact [3]. This substantial improvement demonstrates the potential of intelligent traffic management to meaningfully enhance emergency response effectiveness.

This reports accelerated emergency vehicle detection and intelligent signal control but does not quantify response time improvements [5]. Similarly, Sasikala et al. describe priority-based dynamic route optimization but lack quantitative validation of time savings [7]. The scarcity of response time data likely reflects the challenge of conducting real-world deployments with emergency vehicles, as most systems are validated through simulation or controlled testing [7], [10].

Reliability and False Positive Management

False positive management represents a critical challenge, as spurious emergency vehicle detections disrupt traffic flow and erode system trust. However, standardized false positive rate reporting is notably absent from most publications.

Zohaib et al. report a 3.81% misdetection rate for their multimodal system, representing the most explicit false positive metric in our analysis [29]. Their multimodal fusion approach, requiring agreement between visual and acoustic modalities, demonstrates an effective strategy for reducing false activations [29].

Several systems describe qualitative approaches to false positive reduction without quantitative validation:

- Shitole et al. propose that CBAM attention mechanisms improve robustness to cluttered scenes, theoretically reducing false positives, but provide no empirical false positive rates [1]
- Rajesh et al. describe redundant detection pathways requiring multimodal agreement, claiming reduced false activations, but do not quantify the improvement [4]
- Al-Bahri et al. acknowledge sporadic misclassification in low-light conditions with their ESP32-CAM system but do not provide false positive rates [30]

The lack of standardized false positive benchmarking represents a significant limitation in current research. Without consistent metrics, it is difficult to objectively compare system reliability or identify best practices for false positive management. Future research should prioritize establishing standardized evaluation protocols that include false positive rates under diverse operating conditions [1], [4], [29].

DISCUSSION

Key Trends and Convergence

Analysis of 2024-2025 research reveals several converging trends in intelligent traffic management:

Attention-enhanced architectures: The integration of attention mechanisms (particularly CBAM) with YOLOv9 represents a clear trend toward more sophisticated feature selection [1], [6]. Attention mechanisms enable models to focus on discriminative emergency vehicle features while suppressing irrelevant information, improving both accuracy and robustness. This architectural pattern is likely to become standard in future systems [1].

Multimodal fusion: There is strong convergence toward multimodal approaches that combine visual, acoustic, and communication-based detection [4], [29]. The superior reliability of multimodal systems, particularly in challenging conditions, makes this trend likely to accelerate. However, multimodal systems introduce additional complexity, cost, and potential failure modes that must be carefully managed [29].

Edge deployment: The shift from cloud-based to edge-based processing reflects both technological advances (more powerful edge hardware) and architectural preferences (lower latency, improved resilience, reduced bandwidth) [9], [30]. Edge deployment is becoming feasible even for sophisticated deep learning models, enabling truly distributed intelligence in traffic networks [9].

Reinforcement learning integration: Several systems couple detection with reinforcement learning for adaptive signal optimization, representing a move toward more sophisticated control strategies that balance multiple objectives [2], [11]. This integration of perception and control through end-to-end learning may represent the next frontier in traffic management [2].

Limitations and Challenges

Despite significant progress, several limitations and challenges persist:

Evaluation methodology gaps: The lack of standardized evaluation protocols, particularly for false positive rates and response time improvements, hinders objective comparison and identification of best practices [1], [4], [29]. Most systems are validated through simulation or limited field testing rather than extensive real-world deployment [7], [10].

Scalability questions: While individual intersection systems demonstrate promising performance, questions remain about network-wide scalability. Coordinating hundreds of intersections, managing communication overhead, and ensuring system-wide reliability present significant challenges not fully addressed in current research [3], [7].

Environmental robustness: Systems must operate reliably across diverse weather conditions (rain, fog, snow), lighting scenarios (day, night, shadows), and traffic densities. While some systems acknowledge these challenges, comprehensive validation across environmental conditions is rare [18], [30].

Infrastructure requirements: Multimodal and communication-based approaches require significant infrastructure investment (cameras, microphones, RFID readers, UWB infrastructure) and vehicle equipment [3], [4]. The economic feasibility and deployment timeline for such infrastructure remain unclear, particularly in resource-constrained regions [30].

Privacy and security: Vision and audio-based systems raise privacy concerns about continuous surveillance of public spaces. Additionally, the potential for adversarial attacks (e.g., spoofing emergency vehicle signals) represents a security risk that requires attention [4], [16].

Practical Deployment Considerations

Economic scalability: Vision-only YOLOv9 systems generally offer the lowest deployment cost because they can utilise existing CCTV infrastructure. In contrast, multimodal systems integrating RFID, UWB, and acoustic sensors provide greater detection reliability but require substantially higher infrastructure investment and maintenance. Consequently, system selection should balance operational performance against available municipal budgets and long-term maintenance requirements.

Privacy and security: Future deployments should incorporate edge-based anonymisation techniques, automatic masking of personally identifiable information, encrypted vehicle-to-infrastructure communication, and authentication mechanisms to reduce cybersecurity risks. Robust defences against spoofed emergency signals and adversarial attacks should also be considered to improve system reliability and public trust.

Successful deployment of intelligent traffic management systems requires addressing several practical considerations:

Incremental deployment strategies: Given infrastructure costs and complexity, systems should support incremental deployment, starting with high-priority intersections and gradually expanding coverage [3], [30]. Hybrid approaches that combine advanced capabilities at critical intersections with simpler systems elsewhere may offer practical pathways to adoption [30].

Interoperability and standards: As systems proliferate, interoperability between different vendors and technologies becomes critical. Industry standards for communication protocols, data formats, and performance metrics would facilitate integration and scalability [3], [16].

Maintenance and reliability: Real-world systems must operate continuously with minimal maintenance. Hardware reliability, software robustness, and remote monitoring/updating capabilities are essential for practical deployment [5], [6].

Stakeholder engagement: Successful deployment requires engagement with multiple stakeholders: traffic management agencies, emergency services, city planners, and the public. Understanding operational requirements, addressing concerns, and demonstrating value are critical for adoption [3], [7].

Future Directions and Recommendations

Standardization and Benchmarking

Proposed Benchmarking Framework: Future studies should adopt a standardized evaluation framework including mean Average Precision (mAP), precision, recall, F1-score, false positive rate (FPR), false negative rate (FNR), processing latency (ms), frames per second (FPS), emergency response time reduction (%), computational resource utilisation, and energy consumption. These metrics should be evaluated under different weather, lighting, and traffic-density conditions using common benchmark datasets to enable objective comparison across intelligent traffic management systems.

The research community should prioritize developing standardized evaluation protocols and benchmark datasets for intelligent traffic management systems. Key elements include:

Standardized metrics: Establish consensus on essential performance metrics including detection accuracy (mAP, precision, recall), processing speed (FPS, latency), false positive rate, false negative rate, and response time improvement. Metrics should be reported consistently across publications to enable objective comparison [1], [9], [29].

Benchmark datasets: Develop diverse, publicly available datasets capturing various traffic scenarios, weather conditions, lighting situations, and geographic contexts. Datasets should include labeled emergency vehicles, traffic density variations, and challenging scenarios (occlusion, adverse weather) [10], [29].

Evaluation protocols: Define standardized testing procedures that assess system performance across multiple dimensions: accuracy under ideal conditions, robustness to environmental variations, false positive rates in challenging scenarios, and computational efficiency on target hardware platforms [9], [30].

Advanced Multimodal Integration

Future research should explore more sophisticated multimodal integration approaches:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Learned fusion: Rather than rule-based fusion, employ deep learning architectures that learn optimal fusion strategies from data. Attention-based fusion mechanisms could dynamically weight different modalities based on their reliability in current conditions [29].

Temporal modeling: Integrate temporal information more explicitly through recurrent architectures or temporal attention mechanisms. Tracking emergency vehicles across time could improve detection reliability and enable predictive signal control [11], [18].

Contextual awareness: Incorporate contextual information (time of day, traffic patterns, weather conditions, special events) to adapt detection and control strategies dynamically. Context-aware systems could adjust sensitivity, fusion strategies, and control policies based on current conditions [2], [11].

Explainable AI: Develop interpretable models that provide explanations for detection and control decisions. Explainability is critical for system trust, debugging, and regulatory acceptance [10].

Scalability and Network-Wide Optimization

> Future systems should investigate network-wide multi-agent reinforcement learning (MARL), where each intersection operates as an intelligent agent coordinating with neighbouring intersections to optimise emergency vehicle movement while simultaneously minimising congestion across the entire urban road network.

Moving beyond individual intersections to network-wide optimization represents a critical frontier:

Distributed coordination: Develop distributed algorithms that enable intersections to coordinate emergency vehicle priority across networks without centralized control. Distributed approaches offer improved scalability and resilience [3], [7].

Multi-agent reinforcement learning: Employ multi-agent reinforcement learning where each intersection is an agent that learns to coordinate with neighbors to optimize network-wide objectives (emergency response time, overall traffic flow, emissions) [2], [11].

Predictive routing: Integrate traffic management with emergency vehicle routing, using real-time traffic information to guide emergency vehicles along optimal paths while coordinating signal control along those paths [7], [16].

Resilience and fault tolerance: Design systems that degrade gracefully when components fail. Redundant detection modalities, fallback control strategies, and distributed decision-making can improve system resilience [4], [6].

CONCLUSION

This research reveals significant progress in intelligent traffic management systems that integrate deep learning-based detection with adaptive signal control for emergency vehicle prioritization. Key advances include attention-enhanced YOLOv9 architectures, multimodal fusion approaches combining visual and acoustic detection, and edge deployment on diverse hardware platforms ranging from embedded GPUs to ultra-low-cost microcontrollers.

The convergence toward multimodal, attention-enhanced, edge-deployed systems reflects both technological maturation and growing understanding of practical deployment requirements. Systems employing multimodal confirmation mechanisms demonstrate superior reliability, with the best-performing systems achieving over 96% accuracy and reducing emergency response times by up to 35%. However, significant challenges remain, particularly in standardized evaluation, false positive management, network-wide scalability, and real-world validation.

The core idea of intelligent traffic optimization with emergency vehicle prioritization has evolved from conceptual proposals to sophisticated, deployable systems. YOLOv9's architectural innovations provide the detection foundation, attention mechanisms enhance feature selection, multimodal fusion improves reliability, and edge computing enables low-latency decision-making. These technical advances, combined with increasingly sophisticated control strategies integrating reinforcement learning and optimization algorithms, position intelligent traffic management as a transformative technology for smart cities.

Future progress depends on addressing critical gaps: establishing standardized evaluation protocols and benchmark datasets, developing more sophisticated multimodal integration and temporal modeling approaches, scaling from individual intersections to coordinated networks, and conducting extensive real-world validation across diverse operating conditions. With continued research addressing these challenges, intelligent traffic management systems have the potential to significantly improve urban mobility, reduce emergency response times, and save lives.

The research synthesized in this review provides a comprehensive foundation for researchers and practitioners developing next-generation intelligent transportation systems. By building on the architectural innovations, prioritization strategies, and performance insights documented here, the community can accelerate progress toward more reliable, scalable, and effective traffic management solutions that realize the vision of truly intelligent cities.

REFERENCES

1. Shitole et al. (2024). YOLOv9 meets CBAM: A real-time smart traffic signal system for emergency response optimization.
2. Samiha et al. (2024). Adaptive AI-driven traffic control system for dynamic signal timing and enhanced urban mobility.
3. Fida et al. (2025). IoT-based intelligent traffic management system with dynamic green corridors for emergency vehicle prioritization. ASEAN Engineering Journal. https://doi.org/10.11113/aej.v15.23422
4. Rajesh et al. (2025). Traffic optimization and automated emergency vehicle response system. https://doi.org/10.1109/RAIT65068.2025.11089158
5. Pise. (2025). Accelerated emergency vehicle detection and intelligent traffic signal control. https://doi.org/10.1109/I2ITCON65200.2025.11208865
6. Sabareesh et al. (2024). A smart AI based solution for traffic management with adaptation of traffic light. https://doi.org/10.1109/ICCCSMD63546.2024.11015192
7. Sasikala et al. (2025). Emergency traffic prioritization system with priority-based dynamic route optimization and IoT-enabled dynamic signal control. https://doi.org/10.1109/CONIT65521.2025.11166819
8. T et al. (2025). Integrated AI model for traffic density management and emergency vehicle prioritization. International Research Journal on Advanced Science Hub. https://doi.org/10.47392/irjash.2025.101
9. Raza et al. (2025). An edge-deployed real-time adaptive traffic light control system using YOLO-based vehicle detection and PCE-aware density estimation. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3602844
10. Hanif et al. (2025). Ambulance detection and priority passage at urban intersections using transfer learning and explainable AI. International Journal of Advanced Computer Science and Applications. https://doi.org/10.14569/ijacsa.2025.0161030
11. Azfar et al. (2024). Traffic co-simulation framework empowered by infrastructure camera sensing and reinforcement learning. https://doi.org/10.48550/arXiv.2412.03925
12. Korade et al. (2025). Real-time traffic light optimization using YOLOv9 and length-based metrics. International Journal of Electrical & Electronics Research. https://doi.org/10.37391/IJEER.130212
13. Ahamad et al. (2024). An IoT-enabled system for prioritizing emergency vehicles and detecting accidents in smart traffic management. Indian Scientific Journal of Research in Engineering and Management. https://doi.org/10.55041/IJSREM39111

14. P et al. (2025). YOLOv9 driven Internet of Things enabled sustainable solution for intelligent traffic light management system for emergency vehicles in large scale urban traffic. https://doi.org/10.1109/ICMLAS64557.2025.10968719
15. C et al. (2024). YOLO-based traffic signal optimization for intelligent traffic flow management. https://doi.org/10.1109/I-SMAC61858.2024.10714806
16. (2024). AI-driven traffic management system with graph convolutional neural network, 5G-enabled V2X communication, and IoT for emergency vehicle prioritization. https://doi.org/10.5281/zenodo.14558086
17. Singh et al. (2024). Traffic management in India using YOLOv9 for emergency and regular vehicle detection. International Journal of Innovative Research in Engineering & Multidisciplinary Physical Sciences. https://doi.org/10.37082/IJIRMPS.V12.I5.231091
18. Nirmaladevi et al. (2024). Smart traffic management system for emergency vehicles using deep learning. https://doi.org/10.1109/ICCCNT61001.2024.10724936
19. Rajulapati et al. (2025). Smart traffic management system emergency vehicle. https://doi.org/10.38124/IJISRT/25MAY1592
20. Sasi et al. (2024). An intelligent traffic control system incorporating deep learning and computer vision with prioritized and dynamic timing. https://doi.org/10.1007/978-981-99-9486-1_29
21. Pagare. (2024). AI-driven traffic management system for emergency vehicle prioritization using YOLO. International Journal for Science Technology and Engineering. https://doi.org/10.22214/IJRASET.2024.65461
22. Tabassum et al. (2024). Density-based traffic control system using IoT and machine learning. https://doi.org/10.1109/ICKECS61492.2024.10617246
23. Khake. (2025). Intelligent traffic management systems leveraging IoT and machine learning. https://doi.org/10.5281/zenodo.16926495
24. A novel vehicle collision detection system: Integrating audio-visual fusion for enhanced performance.* (2024). Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2024.123828
25. Maureal et al. (2025). TrafficEZ: Optimizing traffic monitoring through a novel two-tier edge computing model integrating advanced computer vision algorithms. Mindanao Journal of Science and Technology. https://doi.org/10.61310/mjst.v22is1.2209
26. Nguyen et al. (2024). Development of the intelligent traffic light system based on image processing and fuzzy control techniques. CTU Journal of Innovation and Sustainable Development. https://doi.org/10.22144/ctujoisd.2024.302
27. Hu. (2024). Intelligent traffic monitoring and management based on computer vision technology. https://doi.org/10.1109/ICDSIS61070.2024.10594478
28. Sneha et al. (2024). IoT based traffic congestion management and accident detection system. Journal of Electrical Engineering and Automation. https://doi.org/10.36548/JEEA.2024.1.005
29. Zohaib et al. (2024). Enhancing emergency vehicle detection: A deep learning approach with multimodal fusion. Mathematics. https://doi.org/10.3390/math12101514
30. Al-Bahri et al. (2025). A smart AI-based system for emergency vehicle prioritization in urban traffic. International Journal of Electrical & Electronics Research. https://doi.org/10.37391/IJEER.130328