International Conference on Multidisciplinary Perspectives in Advanced Computing and Technology (IMPACT 2026)
G. B. Pant University of Agriculture and Technology, Uttarakhand, India. Jan. 10-11, 2026

DMPedia Lecture Notes in Computer Science & Engineering
Volume 1, pp. 335-345, 2026

Smart Traffic Management and Accident Detection System: A Review

Bijendar Tyagi, Ayush Agrawal, Harsh Jajaniya, Garvit Bhardwaj

Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida
tyagig100@gmail.com, ayushmaserati@gmail.com, ajjujajaniya@gmail.com,
garvitbhardwaj.1.1.9@gmail.com

Abstract: This paper shows us how deep learning and computer vision are changing the way we can handle our traffic and prevent accidents. With the rapid growth of A.I. and deep learning becoming faster and more powerful, it helps monitor traffic, detect accidents, and even predict them. This technology can do this in real time with impressive accuracy. In conditions like bad weather or low visibility, older traffic systems that relied on fixed rules or basic sensors often struggled to keep up. To solve the problem, researchers have been exploring models such as YOLO11-AMF, RT-DETR-EVD, and hybrid GAN-based frameworks. These models are now using convolutional and transformer-based networks to identify and classify objects more precisely. In this review paper, we use 10 different recent studies and compare them. The selected models primarily use deep learning for accident and emergency vehicle recognition. Some of these approaches even combine audio and video data to give better results. We can evaluate them based on their datasets, mean Average Precision (mAP), and how quickly they can make predictions. This paper also discusses the challenges in this field, including limited training data, high computational costs, and models that don't always perform well across different environments. Looking forward, we can clearly highlight trends such as federated learning, self-supervised learning, and model training, which are making it easier to run these systems on IoT and edge devices. Overall, this review shows just how powerful deep learning and AI models are in building safer, smart transportation systems and how close we are to fully automated, intelligent traffic management in the near future.

Keywords: Smart Traffic Management, Accident Detection, Deep Learning, Computer Vision, YOLO, DETR, Emergency Vehicle Detection, GAN, Multimodal Fusion, Edge AI

1. Introduction

We are all part of this problem. The Traffic jam problem. The bigger our cities are, the larger this problem becomes. Our roads get jammed, as a result of this, our daily commutes become a nightmare for us. Now, this is not just a problem; it has become a headache, causing a massive increase in daily road accidents. According to WHO reports, these road crashes kill about 1.3 million people worldwide every year, a problem that demands a solution.

For decades, we've relied on traditional methods to solve this problem. We are using those metal loops cut into the pavement, those rubber tubes we drive over, and those clingy, unclear CCTV recordings. The drawbacks of these methods are significant: they often fail under various conditions, such as bad weather, or simply can't keep up with the chaos on the roads. These are expensive to install and often break down in the middle. These methods are

Proceedings of IMPACT-2026, DMPedia Lecture Notes in Computer Science & Engineering. ISBN: 978-81-993813-9-1

so unreliable and they fail during their initial attempts whenever it gets dark, rainy, or the view becomes blocked. This is where the real revolution begins: Artificial Intelligence (AI).

By combining new technologies such as deep learning (AI models that learn from data) with computer vision, we have entered a new era. With powerful chips (GPU) and massive libraries we now have the technology to process it all instantly.

These new technologies can detect real-time patterns on a busy road, learn complex patterns, and automatically detect everything from cars, bikes, and buses to cyclists and pedestrians. And now, with the advancement of these technologies, we can even predict potential collisions before they occur.

Some of the key technologies with advanced new AI models that are now leading this incredible transformation are:

Convolutional Neural Networks (CNNs): This is the backbone of modern vision systems for hierarchical feature extraction in both static images and videos [3, 5, 9].

Object Detectors: Architectures such as YOLO (You Only Look Once) are expensive as they provide high-speed single-pass detection, making them ideal for real-time applications [1, 8].

Transformers: Models like DETR (Detection Transformer) help improve accuracy in complex scenarios by leveraging self-attention mechanisms to capture the global scene context [2].

Hybrid Frameworks: Researchers are combining models, such as using Generative Adversarial Networks (GANs) to generate synthetic accident data for training more robust CNNs [3] or fusing visual and acoustic data for multimodal detection [4].

This review focuses on the top 10 recent deep learning models designed to enhance smart traffic management and accident detection. It provides a comparative analysis and summary of models such as YOLO11-AMF [1], RT-DETR-EVD [2], GAN-CNN hybrids [3], and multimodal fusion frameworks [4, 5]. We have also mentioned the challenges faced, the datasets, performance metrics, and research gaps that must be addressed to guide future innovations towards a safer, smarter, and more automated transportation infrastructure.

The paper is organized as follows: Section II provides a detailed overview of the deep learning methods. Section III presents the comparative analysis in tabular form and discusses the key findings from the reports. Section IV of this paper comprises the key challenges and research gaps in the field. Section V explores promising future directions, and Section VI concludes the paper.

2. Literature Survey

Deep learning frameworks have become an important part of modern traffic accident

detection systems. They are not bound to just the simple rule-based detection, they move beyond by learning complex real-time patterns and features from video streams, enabling them to identify the critical events in real time.

2.1. Convolutional Neural Networks (CNNs)

The foundational architecture for most computer vision tasks is CNNs. They are now proficient in learning the hierarchical representations. In traffic management, standard CNNs are often used for classification tasks. For example, Wu & Li [9] used a CNN to classify CCTV footage to determine whether it contained an accident. This is effective mainly for basic classification, with 88% accuracy, but it struggles with frame dependency and lacks temporal context to understand event progression. Some of the deeper CNNs are also used to detect vehicles and siren audio signatures, as proposed by [5].

2.2. Real-Time Object Detection (YOLO Family)

The YOLO family is known for its exceptional balance of speed and accuracy, which makes it a primary choice for real-time applications.

YOLOv8-EVD: Johny and Sharma [8] used YOLOv8 for Emergency Vehicle Detection. Their model achieved a 90% mAP but was not very stable for rural cases. This is a common generalization challenge.

YOLO11-AMF: Li et al. [1] proposed a new and updated version of YOLO, which uses a Mamba-like Linear Attention (MLLA) mechanism. This model helps in improving aspects like scaling and localization. This version is specifically designed to handle occluded and small objects in dense urban traffic. It achieved a very high 90.4% mAP but is not very effective for small custom datasets of 694 images, which is a notable limitation.

2.3. Transformer-Based Models (DETR)

Transformer, along with its detection variant DETR, was also introduced as a powerful alternative to CNN-based detectors.Instead of using hand-crafted components like anchors, it uses an

end-to-end architecture based on self-attention, allowing the model to weigh the importance of different parts of the image and capture the global dependencies. Hu et al. [2] also propose a much lighter version of DETR named RT-DETR-EVD, which is optimised for emergency vehicle detection. It achieves 88.7% mAP on the CityFlow dataset, with the only limitation being its performance under nighttime conditions.

2.4. Temporal and Predictive Models

Accidents are not static events. They are the processes, and recognising this requires models that can analyse temporal data.

CONVLSTM: Maneesh Kumar et al. [7] propose a 13D-CONVLSTM2D model. This architecture combines the CNNs with Long Short-Term Memory (LSTM) networks. This combination achieves 87% accuracy on the UA-DETRAC dataset in detecting the accident as it unfolds over time.

Predictive Models: Girija & Divya [10] takes a different approach. Instead of using vision data, they used the "Traffic Data Logs" (e.g., speed, flow, density). Their model simply predicts the likelihood of an accident before it occurs using these logs. This model achieves 82% accuracy. The only limitation of this model is its strong dependence on the availability and quality of long-term data logs.

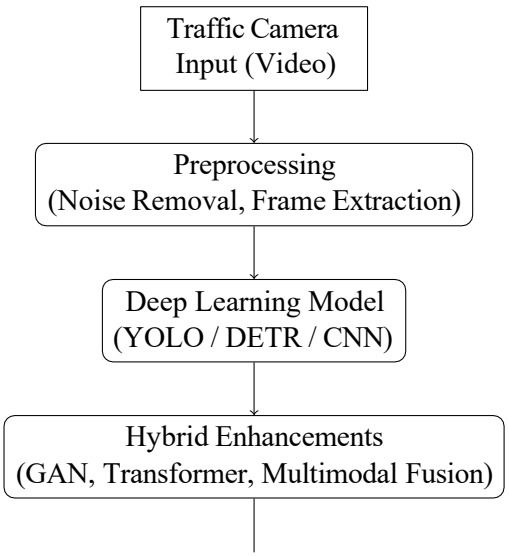
2.5. Advanced and Hybrid Frameworks

The researchers in this field are developing hybrid models. These models will be capable of overcoming problems like data scarcity and single-sensor failures.

GAN + CNN: Accident footage is rare and difficult to collect. To address this data scarcity, Xi et al. [3] use generative Adversarial Networks (GANs) to synthetically generate realistic accident data. This data is then used to train a more robust CNN, yielding a 7-10% gain in detection accuracy. The main drawback of this is the high cost of training GANs.

Multimodal Fusion: Depending on Videos as the sources of information may result in failure. Zohaib [4] proposes a multimodal framework that uses both audio and video as inputs. By using both these, the system achieves 89.2% accuracy, making it more reliable in noisy environments. The highlighted challenge here is the additional environmental noise.

Accident Severity: Chatterjee et al. [6] go far from simple detection to classify the severity of an accident. Their model trained on the CrashSim dataset achieves an accuracy of 85%. This helps in prioritizing resources, which is a critical step for emergency response. Its limitation lies in dataset imbalance as minor accidents are far more common than severe ones.



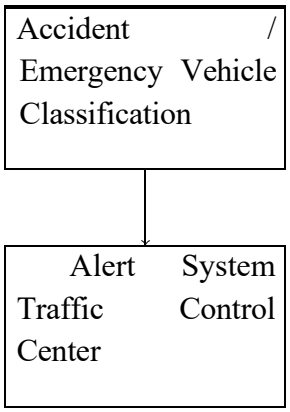


Figure 1: Pipeline of a Deep Learning-based Traffic Accident/Emergency Detection System

3. Comparative Analysis and Discussion

The Table 1 shows the reviewed studies of the 10 papers. This table shows the diversity in models, datasets, and taget tasks, from real-time EVD to predictive analytics. It also unfold several key trends from the comparative analysis of these studies.

3.1. Performance vs. Speed Trade-off

A clear domination of YOLO based architectures are seen in the tasks which requires

real-time inference. Both YOLO11-AMF [1] and YOLOv8-EVD [8] post high mAP scores (90.4% and 90%, respectively) while being designed for speed. The RT-DETR-EVD

model [2] is becoming a copetitor for these two in the field as it is designed as a lightweight transformer for real-time detection. This contrasts with models that prioritize accuracy or complexity over speed, such as the GAN-CNN hybrid [3] or the Deep CNN used by

Alaoui [5], which carry limitations of high computational cost and expensive GPU requirement. This exchange is central to practical deployment on edge devices.

3.2. Dataset Diversity and Specialization

The choice of dataset shows the model's capability. The reviewed papers use a wide array of data sources:

Large Public Benchmarks: Models trained on datasets like CityFlow [2], KITTI [5],

UA-DETRAC [7], and CityScape [8] are tested for generalization on common, complex urban scenes.

Custom or Simulated Data: Some of the paper also shows the lack of specific data by creating their own. The YOLO11-AMF [1] and Zohaib's multimodal model [4] raises

questions about their real-world generalization by using small custom datasets. The GAN-CNN [3] and Severity DL [6] models rely on simulated data (simulated accidents,

CrashSim). This is a very favourable way to overcome data scarcity, but the downside is that it may not capture the full complexity of real events.

Non-Vision Data: The work by Girija & Divya [10] proposed a new and different result, as it forgoes vision entirely in favor of traffic logs. This study highlights a separate branch of research that is focused on statistical prediction rather than real-time visual detection.

3.3. Task Specialization

The paper highlights a specialization of tasks. While accident detection is a broad term, the models target specific sub-problems:

Emergency Vehicle Detection (EVD): A significant portion of the research [2, 4, 5, 8] give attention to EVD. This is a crucial sub-task for clearing traffic. This also helps in reducing emergency response times.

Event vs. Scene Classification: A major difference that still exists between models that analyse temporal events (like the [7]) and those that perform static classification on single frames (like the CNN from Wu & Li [9]).The former is much stronger for accident detection but is more complex.

Beyond Detection: Not all papers follow the same pattern. Two of the papers move beyond simple detection. Chatterjee et al. [6] add a layer of analysis, and Girjia & Divya [10] focus on prediction.

Table 1: Comparative Summary of AI Models for Smart Traffic Accident Detection

Ref.	Model / System Proposed	Technique Used	Findings	Limitations
[1]	YOLO11–AMF	MLLA Attention + CNN	Handles occlu- sion and small ob- jects	Small dataset
[2]	RT–DETR–EVD	Transformer + DETR	Lightweight real-time detection	Weak at night
[3]	GAN + CNN Hybrid	GAN-generated accident data	7–10% gain in accuracy	High trainin g cost

[4]	Multimodal AV Fusion	CNN + Audio features	Detects sirens and vehicles	Noise interfer- ence
[5]	Deep CNN for EVD	CNN + Siren extractor	Highest accu- racy (91%)	Requires ex- pensive GPU
[6]	Accident Severity DL	CrashSim + classifier	Predicts crash severity	Dataset imbal- ance
[7]	CONVLSTM2D	CNN + LSTM	Detects evolv- ing accidents	Moderate la- tency
[8]	YOLOv8–EVD	YOLOv8 detector	Strong urban EVD performance	Fails in rural scenes
[9]	CNN Accident Classifier	Image-based CNN	Accurate scene classification	Frame-dependent
[10]	Accident Predictor	Traffic log deep learning	Predicts acci- dent probability	Requires long- term data

4. Key Challenges and Research Gaps

While the paper shows progress. The review paper also highlights some constant challenges that must be addressed for these systems to become reliable.

4.1. Dataset Scarcity and Imbalance

This is the main challenge. Real-world accident data are very difficult to find in large labeled quantities. This event problem leads to dataset imbalance, a specific limitation noted by Chatterjee et al. [6]. Using small datasets (such as the 594 images for YOLO11-AMF [1]) makes it difficult to assess a model's true generalisation. While GANs [3] give us a promising solution for data augmentation, they also risk creating a sim-to-call gap. Models trained on synthetic data fail to recognise the nuances of real-world events.

4.2. Environmental and Real-World Conditions

Models often do not give correct results when facing difficulties that are not well represented in the training data. This includes:

Adverse Weather: Rain and snow, fog and sun glare can block camera lenses or degrade image quality.

Poor Illumination: Poorly lit tunnels and nighttime conditions remain major challenges. They has been precisely noted for the RT-DETR-EVD model. [2].

Occlusion: Dense traffic and poor road conditions can also become limiting factors at vehicle traffic or accident scenes. This is a problem YOLO-based models. [1, 8] are specifically designed to mitigate, but it remains unsolved.

Sensor Noise: For multimodal systems. Background city noise (construction, car horns) can be mistaken for sirens. This is a key limitation for audio-fusion models [4].

4.3. Computational Cost and Edge Deployment

There is a constant tension between the complexity of different models and their real-time performance.Models like GANs [3], which are complex and deep or heavy CNNs [5], require expensive GPUs, which made them unsuitable for large-scale deployment. The objective is to run inference on low-power edge devices such as camera processors and roadside units. This drives research into lightweight architectures such as RT-DETR-EVD [2], but these models may sacrifice accuracy for speed.

4.4. Model Generalization

Johny & Sharma highlighted a challenge: a model trained in one city may not perform well in another. [8], Their model is trained on urban CityScape data and struggles to produce results in rural areas. Differences in road markings, vehicle types, traffic laws, and even camera mounting angles can cause a well-performing model to fail catastrophically in a new environment.

5. Future Scope

The Limitations in the review paper pointed towards several future possibilities.

5.1. Federated and Self-Supervised Learning

To combat data scarcity and privacy concerns, new training paradigms are needed.

Federated Learning (FL): FL allows models to be trained locally on edge devices like each traffic junction instead of collecting each video data on a central server. This will help preserve privacy and build a more robust model from diverse, decentralised data.

Self-Supervised Learning (SSL): Labelling each video becomes a bottleneck. SSL enables models to learn from vast amounts of unlabelled video data. The model can build a rich understanding of normal traffic by learning to predict the next frame, speed of objects, or the relationships between video and audio, making it highly effective at spotting abnormal events like accidents.

5.2. Advanced Multimodal Fusion

The fusion of audio and video [4, 5] is just the beginning. The future system should integrate a wider array of sensors for true environmental understanding, including:

LiDAR and Radar: These two sensors are robust to the poor weather and lighting conditions that defeat the cameras. Fusing their 3D spatial data with 2D visual data can create a much more reliable detection system.

V2X Communication: Vehicle-to-everything connection provides multiple verticals like sudden braking, airbags deployment, etc. This information can help locate the accident, eliminating the need for visual interpretation entirely.

5.3. Explainable AI (XAI)

The automatic system cannot be trusted by city officials and emergency responders. XAL techniques are needed to make models interpretable. The ‘why’ question should be answered, that is, why it flagged an event as an accident, perhaps by highlighting the specific vehicles involved or the spatio-temporal frames that triggered the alert. For legal and regulatory compliance, this is very crucial for debugging the false alarms.

5.4. Integration with Smart City Ecosystems

The future accident detection system should be deeply integrated into the large smart city network to automate responses. A detected accident or emergency vehicle [2, 8] should automatically trigger:

Smart Traffic Light Control: Altering signal timing to clear a path for emergency responders. **Automated Dispatch:** Sending incident data (location, severity [6]) directly to police, fire, and medical services. **Dynamic Rerouting:** Updating public-facing map services and digital road signs to redirect traffic away from the incident, preventing secondary congestion. These post-accident routing and response strategies can be optimised using Reinforcement Learning.

6. Conclusion

So, this paper basically looked at 10 of the latest AI methods people are using for smart traffic control and accident detection. It is clear that we have moved far beyond the old, unreliable sensor systems. We can predict that we are in the middle of a shift toward intelligent, AI-powered systems. We have also compared models and technologies, including

YOLO [1, 8], DETR [2], GAN-CNN hybrids [3], temporal models [7], and multimodal systems [4, 5]. All this to show how AI is just completely changing the game for city driving and safety. The point to note here is that, even though these new AI models are impressively accurate, they are not perfect. There are also some major hurdles.

The first point here is that the data they use to learn is limited. And mostly, these systems won't work when it's raining, foggy, and dark out. Another big issue here is that they require significant computing power, making it hard to run them on small devices. Another factor is that an AI trained in one city might not be as useful in another. If we want to build an automated, reliable system, we have to solve these problems.

Our focus here should be on a few key things. We should have simpler, lighter AI models that can run on those small, on-site devices. Researchers also need to explore more about combining different types of sensors, such as pulling data from cameras and radar, and the Vehicle-to-Everything (V2X).

A huge percentage of this is also about building AI we can actually trust, such as models that can explain why they made a decision rather than just spit out an answer. And if they get all of this right, the end goal is to create a fully automated city traffic management system that doesn't just react to accidents. They will turn out to be smart and reliable enough to actually help prevent accidents, get emergency help there automatically, and do so way more efficiently.

References

[1] W. Li, L. Huang, and X. Lai, "A Deep Learning Framework for Traffic Accident Detection Based on Improved YOLO11," *Vehicles*, vol. 7, no. 3, p. 81, 2025. [Online]. Available: https://www.mdpi.com/2624-8921/7/3/81

[2] J. Hu *et al.*, "RT-DETR-EVD: An Emergency Vehicle Detection Method," *Sensors*, vol. 25, no. 11, p. 3327, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/25/11/3327

[3] Z. Xi, X. Liu, and Y. Cai, "Integrating GANs and CNNs for Enhanced Traffic Accident Detection," *arXiv preprint arXiv:2506.16186*, 2025. [Online]. Available: https://arxiv.org/abs/2506.16186

[4] M. Zohaib, "Enhancing Emergency Vehicle Detection: A Deep Learning Multimodal Framework," *Mathematics*, vol. 12, no. 10, p. 1514, 2024. [Online]. Available: https://doi.org/10.3390/math12101514

[5] A. O. Alaoui, "Advancing Emergency Vehicle Systems with Deep Learning,"

ScienceDirect, 2025.

[6] T. Chatterjee *et al.*, "Deep Learning Model for Detection and Severity Analysis of Car

Accidents," *Foundations of Computing and Decision Sciences*, vol. 49, pp. 201–231, 2024. [Online]. Available: https://www.sciendo.com/article/10.2478/fcds-2024-0012

[7] S. M. Kumar and B. Tahseen, "Ensemble Deep Learning Framework for Traffic Accident Detection in Smart Cities," *IJERST*, 2025.

[8] C. Johny and A. Sharma, "YOLOv8-Based Emergency Vehicle Detection for Smart City Solutions," *JES*, 2024.

[9] X. Wu and T. Li, "A Deep Learning-Based Car Accident Detection Approach in Video-Based Traffic Surveillance," *Journal of Optics*, 2024.

[10] M. Girija and V. Divya, "Deep Learning-Based Traffic Accident Prediction: An Investigative Study," *VISTAS Journal*, 2024.