Indian Journal of Modern Research and Reviews

This Journal is a member of the *'Committee on Publication Ethics'*

Online ISSN:2584-184X

Research Article

Towards Intelligent Road Safety: A Comprehensive Survey on Multi violation Traffic safety system using deep learning on accident reports

[ID]**Sundari V ¹***, [ID]**Dhana Laxmi A²**, [ID]**Ramya Devi R ³**, [ID]**Sanjana M ⁴**

¹ Assistant Professor, Department of Computer Science and Engineering
Meenakshi Sundararajan Engineering College, Chennai, India
²⁻⁴ Student, Department of Computer Science and Engineering
Meenakshi Sundararajan Engineering College, Chennai, India

Corresponding Author: * Sundari V [ID]

Abstract

The persistent rise in road traffic fatalities continues to pose a grave public health and socioeconomic challenge globally. Accident documentation—while rich in contextual information—exists predominantly in free-form textual formats that resist scalable automated processing. Contemporary advances in Artificial Intelligence (AI), Natural Language Processing (NLP), and deep learning now afford promising pathways for converting such unstructured narratives into actionable safety intelligence. This survey investigates how Multi-Task Learning (MTL) architectures can be applied to concurrently address two intertwined objectives: identifying multiple co-occurring traffic violations (e.g., overspeeding, signal non-compliance, helmet absence, wrong-lane usage, and unauthorised parking) and estimating the severity of the resulting collisions. Looking across the studies we reviewed, no single model family dominates – transformer encoders handle text well, graph networks handle road topology well, and probabilistic models handle uncertainty well, but combining all three remains an open engineering challenge. Beyond raw prediction accuracy, the harder problem turns out to be making these systems trustworthy enough for real enforcement use-explainability tools out to be making these systems trustworthy enough for real enforcement use – explainability tools help, but they are not yet adequate for legal contexts. Beyond raw prediction accuracy, the harder problem turns out to be making these systems trustworthy enough for real enforcement use – explainability tools help, but they are not adequate for legal contexts. We close by identifying what the field most urgently needs: shared benchmark datasets, standardised evaluation metrics, and honest engagement with the fairness and accountability questions that deployment will inevitably raise.

How to Cite this Article

Access this Article Online

www.multiarticlesjournal.com

KEYWORDS: Traffic Safety, Violation Detection, Crash Severity Estimation, Multi-Task Learning, Natural Language Processing, Deep Learning, Intelligent Transportation Systems, Explainable AI.

1. INTRODUCTION

More The intersection of rapid urban growth and exponentially increasing vehicle density has transformed road safety into one of the most pressing infrastructure governance challenges of the 21st century. Globally, crashes attributable to traffic violations—ranging from excessive speed and signal disregard to impaired and distracted driving—result in millions of preventable deaths and disabilities annually, imposing immense economic burdens on healthcare and emergency systems [10],[13].

Every crash generates an accident report—a documentary artefact that encodes causative factors, environmental conditions, vehicle characteristics, and outcome severity. Despite their richness, these reports are predominantly authored in unstructured, narrative prose whose volume and linguistic variability make manual large-scale analysis impractical and prone to inconsistency [4],[12]. Extracting reliable, reproducible insight from such data demands systematic computational methods.

Machine Learning (ML) and, more recently, deep learning have reshaped the analytical landscape for traffic safety research. Early statistical approaches—logistic regression, decision trees, support vector machines—provided a foundational understanding of crash determinants but were ill-suited to high-dimensional, heterogeneous inputs [1],[11]. Contemporary neural architectures, by contrast, can jointly process textual narratives, geospatial coordinates, temporal sequences, and environmental covariates to produce nuanced crash severity estimates and violation profiles [3],[5],[7].

A particularly important methodological evolution is the shift from single objective to Multi-Task Learning (MTL) paradigms, which allow a unified model to simultaneously identify multiple violations and estimate crash severity, exploiting shared representational structure to improve both efficiency and generalisation [9],[15]. Further, Graph Neural Networks (GNNs) and spatiotemporal hybrid models capture non-local spatial dependencies and evolving temporal crash patterns that flat tabular models cannot represent [6],[14].

Alongside predictive accuracy, the demand for interpretable and accountable AI in high-stakes enforcement contexts has accelerated the adoption of Explainable AI (XAI) techniques. Methods such as attention visualisation and SHAP-based attribution provide traffic authorities with auditable reasoning that supports evidence-based policymaking [6],[12].

This survey is organised as follows: Section II establishes foundational concepts in AI-based accident analytics. Section III reviews deep learning methodologies applied to violation detection and severity estimation. Section IV offers a structured comparative analysis. Section V enumerates open research challenges, and Section VI synthesises conclusions with a forward-looking research agenda.

II. Conceptual Foundations of AI-Based Accident Analysis

AI-driven traffic safety systems draw from a layered hierarchy of methodological building blocks. Classical approaches grounded in statistical learning—logistic regression, support vector machines (SVMs), and ensemble methods such as Random Forests—established early benchmarks for crash severity classification. While computationally lightweight and interpretable, these techniques exhibit diminishing performance when confronted with the high dimensionality and severe class skew characteristic of real-world accident datasets [1],[11].

The introduction of sequential deep learning architectures—Recurrent Neural Networks (RNNs) and their Long Short-Term Memory (LSTM) variants—marked a paradigm shift by enabling models to learn temporal dependencies from sequential accident descriptions and time-stamped event logs. Transformer-based architectures further elevated capability by encoding long-range semantic dependencies across entire accident narratives through self-attention mechanisms, substantially improving violation extraction from complex textual descriptions [3],[12].

Spatiotemporal learning frameworks extend these capabilities into the geospatial domain, fusing sequential accident modelling with road-network topology to map crash risk across spatial segments. GNN-LSTM hybrid architectures, in particular, model interactions among locations, vehicle types, and environmental conditions—capturing relationships that neither purely textual nor purely tabular models can represent [6],[14].

Preprocessing pipelines play a foundational role in converting raw accident narratives into machine-readable representations. Standard steps include tokenisation, stop-word filtering, morphological lemmatisation, and multi-label binarisation. When textual features are further concatenated with structured attributes—GPS coordinates, meteorological data, time-of-day indicators—models gain richer contextual grounding for both violation identification and severity prediction [3],[4],[5],[8].

Multi-task and multi-label learning architectures close the gap between isolated subtask performance and end-to-end operational utility. By training violation detection and severity prediction within a unified computational graph, such systems minimise redundant feature learning and enable implicit cross-task regularisation, yielding stronger generalisation under distribution shift [9],[15].

III. Deep Learning Methodologies for Traffic Violation Detection and Severity Estimation

A growing body of empirical research has systematically evaluated AI-based frameworks for automated crash analysis. This section reviews key methodological threads and synthesises their contributions and limitations.

A. Multi-Label Violation Extraction from Accident Narratives

A cardinal feature of real-world crashes is their etiological complexity: most incidents involve overlapping violations rather than a single causal factor. Capturing this multiplicity requires multi-label classification frameworks capable of assigning simultaneous, non-exclusive violation tags to a single accident description [9].

Transformer encoders—pretrained on large corpora and fine-tuned on domain-specific accident texts—have demonstrated superior extraction accuracy by exploiting bidirectional

contextual embeddings. This enables the model to disambiguate between primary causative violations and secondary contributing conditions within the same narrative [9]. MTL architectures further enhance this capability by coupling violation detection with concurrent risk quantification under a shared encoder, reducing inference latency and improving consistency across related outputs [15].

Effective performance in multi-label settings depends critically on robust preprocessing and label-space management. Threshold calibration, class-frequency reweighting, and augmentation strategies are employed to address the long-tail distribution of rare violation categories, which can otherwise dominate false-negative error [12]. Attention mechanisms and contextual embedding alignment techniques add further discriminative capacity for distinguishing nuanced or co-occurring violations [9].

B. Crash Severity Prediction
Accurate crash severity stratification—distinguishing minor injury, serious injury, and fatal outcomes—is foundational to effective emergency dispatch prioritisation and post-event policy calibration. Amiri et al. [1] demonstrated that ensemble learning strategies substantially outperform single-estimator classifiers on imbalanced crash severity datasets by reducing variance without proportional bias increase. Çeven and Albayrak [11] corroborated this finding with comparative evaluations across multiple classifier families.

Spatiotemporal deep learning frameworks advance severity prediction by incorporating location, temporal context, and environmental covariates jointly. These inputs allow models to learn context-sensitive severity signatures—for instance, recognising that high-speed crashes at rural night-time intersections in adverse weather carry systematically different severity profiles than urban daytime collisions [6],[14]. Integrating textual narrative features with structured environmental attributes has been shown to improve prediction.

robustness beyond what either modality achieves independently [5].

Despite strong aggregate performance metrics, interpretability remains a substantive barrier to deployment in regulatory contexts where decisions carry legal weight. This gap motivates increasing integration of XAI techniques as first-class components rather than post-hoc additions [6],[12].

C. Spatiotemporal Risk Modelling and Explainable AI
Moving beyond reactive crash classification, recent research has pivoted toward prospective risk modelling—identifying spatial clusters of elevated crash probability before incidents occur. GNN-based frameworks analyse road-network topology to propagate spatial crash risk across adjacent road segments, enabling fine-grained hotspot maps that inform infrastructure investment and patrol deployment [14]. GNN-LSTM hybrids further capture temporal drift in accident trends, supporting adaptive risk monitoring in dynamic urban environments [6].

Explainability mechanisms are increasingly treated as design imperatives rather than optional extensions. SHAP (SHapley Additive exPlanations) attribution provides per-prediction feature importance scores that align with domain knowledge—confirming, for instance, that weather severity, road curvature, and time-of-day are dominant severity predictors—thereby enhancing stakeholder trust and supporting audit trails [6]. Uncertainty-aware probabilistic GNNs add a further layer of epistemic transparency by producing calibrated confidence estimates alongside crash risk predictions, enabling risk-proportionate response planning [14].

IV. Comparative Analysis of Existing Frameworks
Systematic evaluation of existing AI accident-analysis systems reveals variation across seven principal dimensions. Understanding these distinctions is essential for identifying optimal architectural choices for specific operational contexts.

Table I. Comparative Overview of AI Frameworks

Dimension	Traditional ML	Deep Learning	Multi-Task
Feature Learning	Manual engineering	Auto hierarchical	Cross-task shared
Text Handling	BoW / TF-IDF	BERT embeddings	NLP + struct. fusion
Imbalance	SMOTE	Augmentation	Joint regularization
Spatial Model	Aggregate features	GNN / GNN-LSTM	Spatial-text fusion
Interpretability	High	Low (black-box)	Moderate + XAI
Scalability	High	GPU-intensive	Moderate
Multi-Violation	Sequential models	Multi-label heads	Unified MTL arch.

Table I summarises comparative positions across traditional ML, deep learning, and multi-task hybrid frameworks. Classical ML classifiers remain attractive where computational resources are constrained, and interpretability is paramount, but their inability to process raw text or model spatial topology limits their applicability for modern accident reporting environments [1],[11].

Deep learning architectures—LSTM, transformer encoders, GNNs—consistently achieve superior predictive performance. on complex, multi-modal datasets but require substantially greater compute budgets and careful hyperparameter

management. Their black-box character necessitates XAI augmentation before regulatory deployment [3],[6],[14].

Multi-task hybrid frameworks represent the most operationally promising trajectory, combining the representational power of deep learning with computational efficiency gains from shared encoder backbones. However, task interference and gradient dominance remain engineering challenges in multi-task training, particularly when violation frequencies are highly imbalanced across the label space [9],[15].

V. Research Challenges and Open Problems

Despite substantial progress, AI-based traffic accident analysis faces a constellation of interrelated challenges that impede practical deployment at scale. Data Heterogeneity and Reporting Inconsistency: Crash datasets accumulated from diverse jurisdictions exhibit incompatible terminologies, incomplete field coverage, and region-specific reporting conventions. Textual narratives compound this challenge: semantically equivalent violations may be expressed through entirely different lexical patterns across officers, jurisdictions, or time periods [4],[13]. Domain adaptation and cross-lingual transfer learning remain nascent research areas within this domain.

Model Opacity and Trustworthiness: When an AI system flags a driver for a violation, someone needs to be able to explain why. That explanation cannot be 'the model assigned a probability of 0.87. Courts, regulators, and the public expect reasoning they can follow and challenge. SHAP scores give a rough approximation of which features mattered, but they do not reconstruct the actual decision path. For low-stakes applications, this is acceptable. For enforcement, it is not [6],[12].

Evidentiary and Legal Compliance: Automated enforcement pipelines must produce legally admissible records. This requires cryptographically secured audit trails, tamper-proof metadata embedding, structured logging architectures, and encrypted transmission protocols that preserve chain of custody for violation evidence [11],[12],[18]. Integration with legal frameworks governing data ownership, system liability, and institutional accountability is an under-addressed area that spans technical and policy domains [16].

Severe Class Imbalance: Here is a number that should concern anyone working in this area: a model can score 98% accuracy on a typical crash dataset while having zero recall on fatal crashes. This happens because fatalities represent 1-2% of most datasets, and the easiest way to maximise accuracy is to never predict them at all. Research knows this. Research knows this. Reporting accuracy as the primary metric anyway is, at best, misleading. F1 on the fatal class, or precision-recall AUC, would expose the problem immediately, and the community should adopt these as minimum reporting standards [1],[11].

Computational Scalability: Spatiotemporal GNN-LSTM architectures require substantial GPU resources for training and serving at the data volumes generated by large metropolitan areas. The best-performing spatiotemporal models require 30-40 GB of GPU memory just for inference. Roadside enforcement units do not have that. Until model compression research catches up with safety-critical accuracy requirements, real-time deployment will remain out of reach [2],[14].

Cross-Regional Generalisation: Models trained on one region's crash dataset frequently fail to generalise to regions with different road infrastructure, driver demographics, legal frameworks, and reporting standards. Federated learning and domain-adaptive pretraining strategies represent nascent solutions, but standardised benchmarks for cross-regional transfer remain absent from the literature [12],[14].

Absence of Unified Multi-Task Benchmarks: Many contemporary studies evaluate violation detection, severity prediction, and risk mapping as isolated subtasks on non-standardised datasets. Every paper in this area uses a different dataset with different preprocessing and different metrics. You cannot compare results across papers. This is the single biggest obstacle to scientific progress here, and fixing it requires community coordination, not just better algorithms [9],[15].

Fairness and Algorithmic Bias: Crash datasets may underrepresent certain road types, geographic regions, or demographic groups. Models trained on such corpora risk encoding structural biases that produce inequitable safety interventions—disproportionately flagging certain road users or systematically underserving underrepresented communities. Bias auditing frameworks and fairness-constrained training objectives are insufficiently explored in this domain.

VI. CONCLUSION

This survey has examined the evolving landscape of AI-driven traffic safety analysis with particular attention to the dual objectives of multi-violation extraction from accident narratives and crash severity estimation. The evidence base supports a clear trajectory: purely statistical models, while accessible and interpretable, are insufficient for the representational demands of modern heterogeneous crash datasets; deep learning architectures—transformers, LSTMs, GNNs—substantially expand predictive capability but introduce opacity and computational costs that complicate operational deployment.

Multi-Task Learning frameworks represent the most coherent unifying paradigm, enabling concurrent violation detection and severity prediction within shared architectural backbones while reducing redundant computation. Their effectiveness is amplified by spatiotemporal modelling techniques that embed crash events within their geographic and temporal context, and by Explainable AI methods that render model decisions interpretable and auditable for regulatory stakeholders.

Critical gaps remain: the field lacks unified evaluation benchmarks, scalable fairness auditing tools, and legally compliant evidence management architectures. Addressing these limitations will require interdisciplinary collaboration spanning computer science, transportation engineering, legal informatics, and public policy. As intelligent transportation infrastructure continues to mature, the development of robust, transparent, and equitable AI systems for road safety governance becomes not merely a technical aspiration but a societal imperative.

REFERENCES

1. Amiri MA, Afshari S, Soltani A. Machine learning approaches to traffic accident severity prediction: Addressing class imbalance. **Machine Learning with Applications**. 2025.
2. Wicaksono D, et al. A novel approach to predict traffic accident assistance based on deep learning. **Neural Computing and Applications**. 2025.
3. Wicaksono D, et al. Learning spatial patterns and temporal dependencies for traffic accident severity prediction: A

deep learning approach. **Knowledge-Based Systems**. 2024.

4. Wicaksono D, et al. A systematic review of factors, data sources, and prediction techniques for earlier prediction of traffic collision using AI and machine learning. **Multimedia Tools and Applications**. 2025.

5. Wicaksono D et al. A data-driven deep learning framework for prediction of traffic crashes at road intersections. **Applied Sciences**. 2025.

6. Wicaksono D, et al. A computational intelligence GNN-LSTM framework for spatiotemporal prediction of traffic accident severity in smart cities using SHAP XAI. **International Journal of Computational Intelligence Systems**. 2025.

7. Wicaksono D, et al. Traffic accident duration prediction using multi-mode data and ensemble deep learning. **Heliyon**. 2024.

8. Wicaksono D, et al. Traffic accident risk prediction based on deep learning and spatiotemporal features of vehicle trajectories. **PLoS ONE**. 2025.

9. Wicaksono D, et al. A survey on traffic violations prediction with deep learning. **Transportation Research Procedia**. 2025.

10. Kotsyubynska O, et al. Machine learning and deep learning for predicting traffic crash injury severity: A systematic review and meta-analysis. **Journal of Road Safety**. 2026.

11. Çeven S, Albayrak A. Traffic accident severity prediction with ensemble learning methods. **Computers and Electrical Engineering**. 2024.

12. Hamdan N, Sipos T. Advancements in machine learning for traffic accident severity prediction: A comprehensive review. **Periodica Polytechnica Transportation Engineering**. 2025.

13. Shubenkova K, et al. Machine learning for predictions of road traffic accidents and spatial network analysis. **Elsevier**. 2024.

14. Gao X, et al. Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. **Accident Analysis and Prevention**. 2024.

15. Yang L, et al. Predicting Lane change manoeuvres and associated collision risks based on multi-task learning. **Accident Analysis and Prevention**. 2025.

16. Hou Q, et al. A deep transfer learning approach for real-time traffic conflict prediction with trajectory data. **Accident Analysis and Prevention**. 2025.

17. Ganesh S, Gomathi R, Kannadhasan S. Brain tumour segmentation and detection in MRI using convolutional neural networks and VGG16. **Cancer Biomarkers**. 2025.

18. Khan SW, et al. Anomaly detection in traffic surveillance videos using deep learning. **Sensors**. 2022.

19. Gao X, et al. Uncertainty-aware probabilistic graph neural networks for road-level traffic crash prediction. **Accident Analysis and Prevention**. 2024;196.

20. Zheng Y, Liu H, Zhang J. Deep learning-based spatiotemporal modelling for traffic crash risk prediction in intelligent transportation systems. **Transportation Research Part C: Emerging Technologies**. 2024;156.

21. Hels T, Reiff LK, Lahrmann H, Madsen TKO. Prevalence of drink driving in road traffic in Denmark. **Traffic Safety Research**. 2025;10: e000127.