IVORY

AI for Vision Zero in Road Safety

Methodologies for the detection, analysis and management of road risk through Artificial Intelligence (AI)

This project has received funding from the European Union's Horizon Europe Research and Innovation programme under the Marie Skłodowska-Curie grant agreement No 101119590. The opinions expressed in this document reflect only the authors' view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of ains.

Marie Skłodowska-Curie Actions

www.ivory-network.eu

Project identification

Grant Agreement No	101119590
Acronym	IVORY
Project Title	AI for Vision Zero in Road Safety
Start Date	01/11/2023
End-Date	30/10/2027
Project URL	www.ivory-network.eu

Document summary

Deliverable No.	**6.3**
Deliverable Title	Methodologies for the detection, analysis and management of road risk through Artificial Intelligence (AI)
Work Package	WP6
Contractual due date	31 October 2026
Actual submission date	31 October 2026
Nature	R – Document, report
Dissemination level	PU – Public
Lead Beneficiary	NTUA
Responsible Author	Apostolos Ziakopoulos (NTUA)
Contributions from	Akanksha Agarwal (UHasselt), Nimra Jabeen (UHasselt) Simone Paradiso (NTUA), Júlia Alves Porto (NTUA) Muhammad Shahid (FPZ), Göker Malik Altuntaş (UHasselt) Eleonora Papadimitriou (TU Delft), George Yannis (NTUA)

Please refer to the document as:

Ziakopoulos, A., Agarwal, A., Jabeen, N., Paradiso, S., Alves Porto, J., Shahid, M., Altuntaş, G. M., Papadimitriou, E., & Yannis, G. (2026, October 31). *Methodologies for the detection, analysis and management of road risk through Artificial Intelligence (AI).* Deliverable 6.3 of the EC Horizon Europe project IVORY.

Table of Contents

List of Figures

List of Tables

Executive Summary

IVORY - 'AI for Vision Zero in Road Safety' is a Marie Sklodowska-Curie Actions (MSCA) Industrial Doctorates Network project. The project aims to leverage AI methodologies in order to contribute towards the reduction of the global toll of road crashes and their consequences. Road crashes have plateaued to 1.19 million fatalities globally, with the added burden of serious injuries and property damage.

The present Deliverable (D6.3) is compiled within WP6 of the IVORY project, titled 'Proactive infrastructure safety management'. This document aims to present a comprehensive overview of methodologies for the detection, analysis, and management of traffic safety and road risk through Artificial Intelligence (AI). This report consolidates state-of-the-art approaches that leverage data-driven techniques, computer vision, and machine learning to enhance road safety assessment and proactive risk management. The deliverable builds on extensive literature reviews, methodological evaluations, and pilot applications to establish a framework for integrating AI into transport safety analysis across diverse use cases.

Firstly, systematically reviews for AI-supported road safety methodologies in several thematic areas are conducted, including infrastructure safety assessment, proactive network-level safety analysis, video-based image processing for trajectory and conflict detection, spatial scaling techniques, and generative advisory models such as Building Information Modelling (BIM). Afterwards, the detailed methodological concepts and frameworks of six Doctoral Candidates (DCs) are presented (DC8–DC13). These frameworks are illustrated with detailed flowcharts and a thorough description of present and future research stages. These methodological frameworks provide the roadmap for the execution of each doctoral research, but also provide the analytical approach layout for tackling each composite road safety problem, providing a paradigm for contemporary road safety research.

Subsequently, the exploitation, integration, availability and sensitivity of heterogeneous data sources are showcased. These data sources include telematics, crash data, traffic video images, LIDAR data, digital map data, driver simulator and microsimulation data, ensuring a well-rounded approach of modern road safety problems and a valuable foundation for the development of state-of-the-art models. Elaborations of the overall availability and replicability of the data source, as well as remarks on its respective sensitivity and overall compliance with data protection regulations (such as GDPR) are also provided.

Beyond methodological consolidation, Deliverable 6.3 demonstrates the practical contributions of AI to road risk management. These include automated infrastructure attribute extraction, multi-object trajectory generation, surrogate safety metric computation, and AI-based conflict analysis. The report also outlines innovative approaches for synthesizing findings across methodologies, enabling scalable and holistic safety pattern recognition. Ultimately, this deliverable provides a robust methodological foundation for the IVORY project's subsequent research tasks, while offering a roadmap for future integration of AI into evidence-based transport safety policymaking.

1 Introduction

1.1 About the IVORY project

IVORY - 'AI for Vision Zero in Road Safety' is a Marie Sklodowska-Curie Actions (MSCA) Industrial Doctorates Network aiming to develop a new framework for optimal integration of Artificial Intelligence (AI) in road safety research, and to train a new generation of leading researchers in the field. It addresses the UN Sustainable Development Goals target 3.6 and the EC Vision Zero strategy, of halving traffic fatalities by 2030 and eliminating them by 2050.

The incentive and deeper motivation behind the research at the wider core of IVORY, and WP6 specifically, is the fact that despite decades of methodological advancements, engineering progress and accompanying technological progress, global road fatalities have plateaued at 1.19 million deaths annually (World Health Organization, 2023).

Existing road safety approaches and methodologies have been to a very large extent reactive, based on static, contained datasets. It appears that they have largely explored the depth of context, complexity and adaptability that were available due to their scope. The related datasets have not facilitated the inclusion of all aspects of the road environment in rich, high-resolution detail, including road user behaviors and interactions, infrastructure conditions and other environmental attributes. Most importantly, the reactive nature of traditional road safety analyses implies the belief that, by analyzing past crashes, pertinent knowledge can be obtained to avoid future crashes and mitigate their consequences.

AI methodologies are poised to achieve considerable breakthroughs in the field of technologically-supported road safety research through several comparative advantages. Some of them include the capability to handle immense volumes of data in dynamic, often real-time configurations, the ability to examine a very large number of features/variables, some of which might also be intercorrelated, the ability to intergrate non-linear effects in model configurations and the ability to prioritize features based on contextual value, given the appropriate programming protocols. Most critically, AI algorithms shine when used to make predictions and forecasts given new data, and they can be scaled to broader study areas. These comparative advantages enable AI methodologies to provide insights beyond the traditional statistical and econometric approaches, thus achieving new road safety gains that were previously unobtainable. A parallel advantage is that AI analyitic methodologies are seamlessly integratable with other data-oriented tasks, from recording, transmition and storage to cleaning and pre-processing to feeding antecedent processing and post-processing algorithms.

Based on the aforementioned, IVORY addresses the persistent distance between available safety data and actionable insights for risk prevention. It aims to foster research that creates actionable and scalable AI tools capable of understanding, predicting, and mitigating infrastructure-related safety risks in a dynamic context. This objective is achieved by means of 4 research goals, aiming to develop:

(i) responsible, fair and impactful AI for road safety,
(ii) new ways of road user support and human-vehicle-environment interaction,
(iii) new scalable and equitable AI technologies for proactive infrastructure safety management,
(iv) a sustainable knowledge sharing network on AI for road safety.

IVORY outputs will not only provide more robust user support through AI in vehicle automation, but will also allow to responsibly and proactively manage the persistent problems of existing conventional, low automation transport systems, so that new opportunities for global road safety impact can emerge. Moreover, IVORY adopts a design for values approach for AI in road safety, operationalising the ethical principles of justice and explainability, and providing efficient AI solutions also for disadvantaged groups (e.g. vulnerable road users, low-to-middle-income countries).

IVORY consists of 4 academic and 8 non-academic beneficiaries, and 8 associated partners, joining from engineering, data science and ethics of technology disciplines, from 11 countries. 15 young researchers will receive high-level doctoral education as Doctoral Candidates (DCs), along with industrial exposure, local training, and 8.5 European Credit Transfer System (ECTS) of network-wide training on key advanced, core and transferable skills. As part of the activities and beyond, IVORY will create an online learning & networking platform for AI in road safety, to be available after the end of the project for future researchers in this field.

1.2 Overarching research goals for WP6

WP6, entitled 'Proactive infrastructure safety management', will develop new data fusion methods exploiting an array of hybrid road attributes datasets (e.g., video, sensor, geospatial, manual) (DC12) and translate them into road scanning and evaluation models of proactive risk mapping (DC8). This will be complemented by new scalable models for a seamless infrastructure safety management from micro to macro levels (DC9). It will improve AI methods for traffic conflicts detection (DC11) and incorporate subjective safety into safe routing advice tools for children (DC10), as well as BIM tools for the safe navigation of connected and Automated Vehicles (CAVs) in the complex urban setting (DC13); it will also create an AI-chat model to provide advice to road authorities for enhancing safety of VRU groups (DC15).

Specifically, WP6 contributes to achieving Research Goal 3 (RG3) of the project, i.e., to develop new scalable and equitable AI technologies for proactive infrastructure safety management.

The overarching problem addressed by WP6 is therefore the lack of scalable, real-time, and data-rich methods for proactive road safety management. Artificial Intelligence provides a transformative opportunity to overcome these limitations by integrating heterogeneous data sources—such as video, telematics, and geospatial information—into predictive models that identify risks before crashes happen and inform timely interventions. WP6 aims to operationalise this paradigm shift by developing, testing, and validating AI-based frameworks that enable continuous safety monitoring, dynamic risk mapping, and evidence-based design and management of safer transport infrastructure.

The WP6 goal is pursued from three different fronts: proactive safety assessment, safer routes suggestion and safe design. Leveraging video data, road user interactions can be tracked through computer vision algorithms that transform raw traffic video into 3D trajectories, surrogate safety indicators, and conflict matrices- This enables proactive detection and quantification of safety-critical interactions at urban intersections in near real time, as opposed to reactive or human-based conflict detection.

Moving on to network-level analysis, video data can also be integrated with LiDAR point clouds and satellite images for comprehensive road infrastructure analysis and proactive road safety assessments. This multimodal fusion captures the full spectrum of safety-critical road infrastructure features and produces a more accurate, robust, and holistic analysis of road infrastructure than any single source alone. The development of state-of-the-art algorithms using different data sources allows for novel, innovative safety performance functions based on automatically extracted attributes. These can be further combined with proactive surrogate safety measures, such as telematics-based features. The integration of telematics data with road networks remains a relatively unexplored area in the literature. Within this work package, telematics-informed networks that spatially reflect naturalistic driving behaviour will be built and explored using AI tools, enabling scalable analyses across large metropolitan areas and the processing of large datasets, a key limitation of non-AI approaches.

In the context of networks, numerous efforts have been made to incorporate safety into road navigation systems. This work package further leverages AI algorithms to identify complex patterns from existing data and to systematically characterise the subjective perception of safety, integrating these insights in real-time into the selection of optimal travel routes. Finally, this work package uses BIM for analysing digital road infrastructure, an area where robust methodologies are still lacking. Design and analysis remain separate processes, with the latter often taking place at later stages. Integrating AI into the design phase promises a standardised yet context-aware approach, in which rich contextual information is processed directly within the architecture of the machine-learning models.

The research in this Work Package aims to contribute to the following innovations and key exploitable results.

Innovations:

- Seamless AI-based methodologies for scaling safety assessment from micro to macro
- Novel data fusion techniques
- A new BIM framework for improved decision making in roadway design

Key exploitable results:

- Open-access backend data and codes
- AI software for road user profiling
- AI codes for automatic data collection
- AI-based scalable risk mapping tool
- AI-aided BIM-based framework
- AI-chat policy support model

Finally, by creating an open, collaborative knowledge-sharing network on AI for road safety by the end of the project in 2027, which will include the above innovations and results, IVORY and its WP6 will ensure that the methods and findings contribute to a measurable global reduction in fatalities and serious injuries, supporting the transition from reactive crash investigation to proactive agile crash prevention and mitigation of consequences.

1.3 About this Deliverable

1.3.1 Objectives

The rapid advancement of artificial intelligence (AI) has introduced new opportunities for enhancing road safety research and practice. In recent years, AI-driven methods have been increasingly adopted to tackle complex challenges such as driver behavior analysis, traffic monitoring, risk prediction, as well as the optimization of safety interventions. AI approaches are capable of processing vast, heterogeneous datasets and detecting patterns that are often not as easily detected by conventional methods. As road networks and mobility systems become more dynamic and the digital infrastructure provides a data-rich environment, methodological approaches that leverage AI tools are playing a pivotal role in shaping how safety risks are understood and managed.

Within this background, IVORY D6.3 aims to provide the current and future scientific context for the 6 doctoral dissertations conducted within WP6 of IVORY. The document focuses on the methodological strategies underpinning some of the newest AI-powered road safety analytical frameworks. It explores how AI tools are deployed in the wider state of the art, as well as the techniques used to adapt them to the specific complexities of road safety research. The methodological emphasis serves to provide clarity on how AI contributes to the design, execution, and refinement of studies in applied doctoral research. The strengths, limitations, and evolving practices that define the current landscape of AI-enabled road safety research are thus showcased, along the applied approaches as case studies within doctoral research.

1.3.2 Thematic areas

In order to better direct the state of the art findings and discussion, it is essential to structure the methodological discussion around thematic areas that capture the diversity of AI applications in road safety research. These areas represent distinct yet interrelated methodological lines of inquiry, each addressing different aspects of how data can be collected, processed, interpreted, and ultimately used to inform scientifically-supported decision-making. The broader thematic areas comprise common ground for subgroups of IVORY DCs, as their topics are included in them, and thus also present collaboration and knowledge exchange opportunities.

The thematic areas considered in this Deliverable span both infrastructure-focused and behavior-focused perspectives. They include the following:

A. Video image attribute extraction for infrastructure element labelling is examined, as well as video image processing techniques such as trajectory extraction for proactive conflict analysis.
B. Infrastructure safety assessment is also conducted, through evaluation and safety scoring by the analysts but also through subjective rating by the road users. Attribute selection processes for infrastructure safety assessment is discussed as well.

C. Proactive safety analysis is another core component of the report, with techniques being examined for infrastructure design (such as with BIM-supported approaches), conflicts and road user perception.
D. Generative advisory models and their capacity for informing road safety policy and design based on external documentation, but also road user feedback, are examined as well.
E. Spatial scaling approaches and how to treat the issues of locality in road safety are investigated.

Together, these themes define several critical fronts of advancement regarding the methodological landscape of AI in road safety and frame the subsequent analysis presented in this document.

1.3.3 Report outline

Based on the aforementioned, this report is structured as follows:

The Introduction (Section 1) places the document in view of the background and advances of the road safety science, highlighting its contents and necessity. Subsequently, the state-of-the-art is provided in the Review of Methodologies (Section 2), providing broader scientific context of AI advances and available methodologies. The Methodological approaches (Section 3) are then provided, which constitute the individual endeavors of IVORY DCs as case-studies within the aforementioned broader context. A parallel discussion of the essential data and their particularities for each dissertation is provided in Data acquisition (Section 4). Afterwards, the overall Wider Methodological Contributions of the IVORY DCs are showcased (Section 5), and the Deliverable concludes with key takeaways as well as Future steps (Section 6).

Specifically, the contributions of the following DCs are included in this deliverable, the topics of which find common thematic ground in novel proactive road safety approaches:

- DC 8: Júlia Alves Porto – Proactive risk mapping and infrastructure safety management
- DC 9: Simone Paradiso – AI for road safety monitoring and crash prediction from micro- to macro levels
- DC 10: Akanksha Agarwal – Using AI for the identification, monitoring and utilisation of a personalised self-learning safe route network for home-school trips
- DC 11: Nimra Jabeen – AI for proactive safety detection using conflict techniques
- DC 12: Muhammad Shahid – Harmonisation and hybrid application of AI datasets for road safety
- DC 13: Göker Malik Altuntaş – AI-aided BIM-based design for road infrastructure

2 Review of Methodologies for the detection, analysis and management of road risk through Artificial Intelligence

The present Section presents a broader scientific context of advances and available methodologies for AI in road safety. In particular, methodologies for the detection, analysis and management of road risk, based on the aforementioned thematic areas.

2.1 Overview of AI-supported infrastructure safety assessment methodologies

2.1.1 Attribute selection

Road Safety has become one of the most significant public safety issues for many countries in the world. According to the statistics of the World Health Organization (WHO), traffic accidents cause about 1.19 million deaths and 20 to 50 million non-fatal injuries in the world each year (World Health Organization 2023), (Karimi Monsefi et al. 2023). In addition to the tragic loss of lives, road crashes result in significant economic losses (Chen et al. 2019), including medical costs (Basili et al. 2024), lost productivity (Chantith, Permpoonwiwat, and Hamaide 2021) and damage to infrastructure (Stojanovic et al. 2023). Moreover, the impacts of road trauma are also profound (Harms and Talbot n.d.) which requires urgent and effective measures to improve road safety.

The United Nations (UN) has set 12 Voluntary Global Road Safety Performance Targets in the "Global Plan for the Decade of Action for Road Safety 2021-2030" (United Nations. Safe Roads for a Sustainable World 2023) to encourage worldwide initiatives aiming at lowering road traffic injuries and fatalities. Among these, targets 3 and 4 focus on improving road infrastructure. Figure 1 shows these targets in detail. Target 3 mandates that all new roads satisfy a minimum 3-star safety rating for all road users. While Target 4 aims to guarantee that at least 75% of travel occurs on roads meeting this quality by 2030. Although driving is a routine activity, it involves inherent risks for both drivers and other road users. These risks arise from various factors, including road design, surface conditions, the surrounding environment, and driver behavior (Zia et al. 2025).

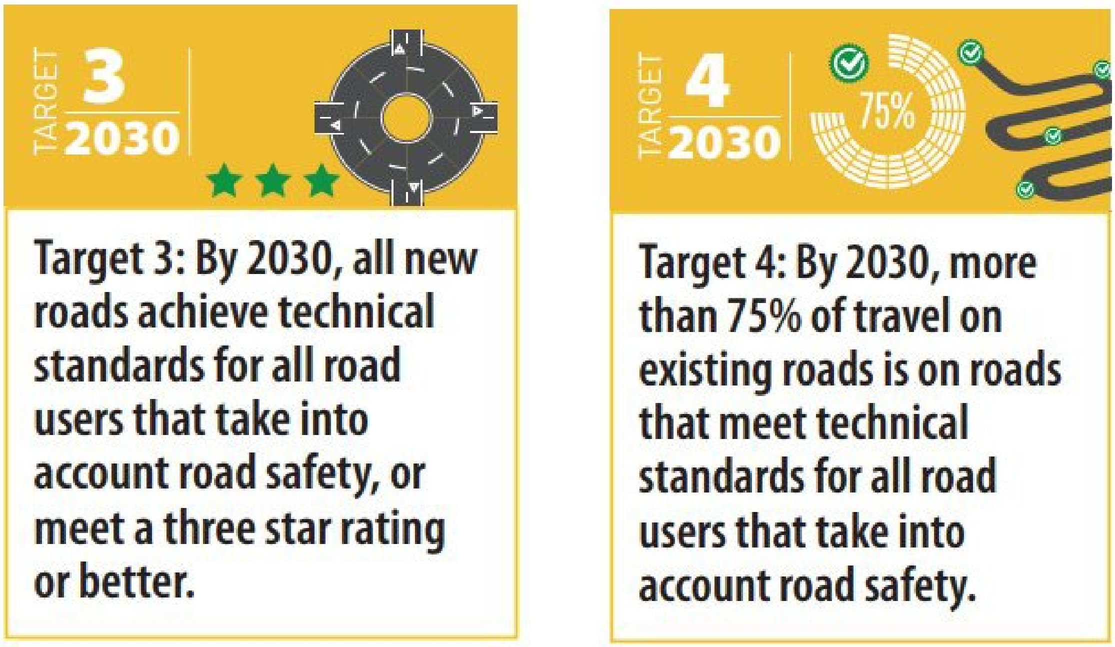

FIGURE 1. THE UN'S DECADE OF ACTION FOR ROAD SAFETY 2021-2030

Targets 3 and 4 require a minimum 3-star safety rating for all new roads and aim for 75% of travel on these roads by 2030 (United Nations. Safe Roads for a Sustainable World 2023)

Road safety assessment is a systematic process designed to evaluate road conditions and identify potential hazards, playing a vital role in preventing crashes. Traditional road safety approaches often rely on historical crash data to identify high-risk sections (Aziz and Ram 2022; He et al. n.d.). However, such methods are reactive, as they rely on past incidents to highlight high-risk zones (Ahmed, Puan, and Ismail 2013). In contrast, a proactive approach emphasizes safety by evaluating road infrastructure attributes before crashes occur (Ahmed et al. 2013)(Kačan, Ševrović, and Šegvić 2024). The International Road Assessment Programme (iRAP) has been working on assessing road infrastructure and identifying risk factors that contribute to crashes. iRAP employs a comprehensive framework to assess numerous road safety features (road attributes), which are crucial for maintaining safe driving conditions and proactive road safety assessments.

The iRAP Star Rating evaluates the safety level of road infrastructure for four primary road user types (International Road Assesment Programme (iRAP) 2019). This assessment involves analyzing more than 50 carefully chosen attributes related to road infrastructure elements and roadside objects within specific 100-meter or 10-meter road segments. Each attribute is assigned a category from its unique taxonomy, with the number of categories varying by attribute. Currently, these iRAP road safety attributes are manually classified by human coders from road survey videos. This manual road infrastructure assessment process is time-consuming, labor-intensive, and subject to human error (Zhou et al. 2022). With advancements in deep learning and computer vision, we can automate the extraction of road attributes from survey data, delivering more consistent, accurate, and frequent analysis than traditional manual methods. An AI-based road infrastructure analysis system can significantly outperform conventional approaches by providing faster, more scalable assessments, detecting road features with greater accuracy, and enabling timely interventions. Table 1 shows the attributes which can be extracted using computer vision.

TABLE 1: AUTOMATICALLY IDENTIFIABLE ROAD SAFETY ATTRIBUTES FROM ROAD SURVEY VIDEOS USING AI

Category	Attributes
Road Features	Roadworks, Sight distance, Delineation, Street lighting, Service road
Roadside Severity	Driver Side objects & their distance from road edge, Passenger Side objects & their distance from road edge Side objets & their distance from road edge
Shoulder	Paved shoulder - driver-side, Paved shoulder - passenger-side
Intersections	Intersection type, Intersection channelisation, Intersection quality, Property access points
Land Use	Land use - driver-side, Land use - passenger-side, Area type
Pedestrian Facilities	Pedestrian crossing - inspected road, Pedestrian crossing quality, Pedestrian crossing - side road, Sidewalk - driver-side, Sidewalk - passenger-side
Cyclist Facilities	Bicycle facility
School Zones	School zone warning, School zone crossing supervisor

2.1.2 Safety evaluation and infrastructure rating

In the pursuit of more precise and less resource-intensive models, the use of Artificial Intelligence for road safety purposes has increased, especially in two areas: (i) safety modelling and (ii) computer vision. AI in safety modelling is capable of making non-parametric connections and promote better predictions than traditional econometric models, while computer vision features enable the capture of infrastructure information with minimal human intervention.

For modelling approaches, Machine Learning (ML) applications have been extensively used (Ali et al., 2024; Silva et al., 2020; Sohail et al., 2023; Ziakopoulos & Yannis, 2024). Some difficulties when working with ML models for safety assessment are overfitting, when the data performs well on the training set but poorly when transferred to a new dataset, and lack of explainability (Ziakopoulos & Yannis, 2024). ML models are often addressed as a “black box”, but safety research is usually not only after prediction of the behavior, yet what features should be targeted to promote a safer environment.

As for which models are most used within safety research, there is an overall consensus on the popularity of neural network-based models for crash prediction (Ali et al., 2024; Silva et al., 2020; Sohail et al., 2023; Ziakopoulos & Yannis, 2024). A neural network refers to a machine-based algorithmic approach that mimics the human brain process by making intuitive associations through intermediate hidden layers, referred to as "neurons" (Haykin, 2009). For example, (Bustos et al., 2021) trained Residual Convolutional Neural Network model to calculate hazard levels for both pedestrians and vehicles based on street-level imagery, and used Class Activation Map (Grad-CAM++) to explain the relationship between the input attributes and the hazard level output. (Wang et al., 2019) used a back propagation neural network with simulated data from a Shanghai Expressway to predict crash risk. They included three types of outcomes: no-risk, high-risk and low-risk, and achieved a 0.75 AUC for the risk/no-risk test dataset, and 0.93 for the high-risk/low-risk dataset. (Zeng et al., 2016) developed a neural network crash prediction model with a multilayer perceptron structure, stratified by severity, and achieved superior results than their benchmark multivariate Poisson-lognormal model. (Karim et al., 2022) proposes an attention-based neural network, named Dynamic Spatial-Temporal Attention network (DSTA), for crash risk prediction in real time using dashcam videos.

In recent years, Decision Tree (DT)-based models have also been used frequently. Decision Trees work by an iterative process that begins with splitting the dataset into branches connected by nodes, which represents mathematical criterion, and then, for each data point, a decision rule is applied to direct it to a specific branch; with every iteration, the algorithm creates new nodes and branches, until the stopping conditions are met and the leaves contain the final predictions. Despite being known for being highly subjective to overfitting problems, ensemble models based on DT combine several weak predictors to output a strong predictor. The ensembled models can be themselves divided between bagging and boosting models, and have presented good results within the road safety field (Ali et al., 2024). (Ziakopoulos & Yannis, 2024) cite studies that compared different Machine Learning models and found that Random Forest outperforms several other architectures such as SVM, ANN, KNN, C5.0 and CART, the latter two also being under the DT umbrella, although highlighting that "algorithmic performance is largely data-driven", therefore, the results cited are not generalizable for all cases. Several studies have used XGBoost for crash (Chakraborty et al., 2023; Yue, 2024; Zhang et al., 2024) or surrogate (Ziakopoulos, 2024; Ziakopoulos et al., 2022) safety assessments.

Other model types, based on traditional probabilistic and linear statistics, that appear frequently in the literature are Support Vector Machines (SVM), Bayesian Networks (BN) and Naïve Bayes (Silva et al., 2020; Sohail et al., 2023; Ziakopoulos & Yannis, 2024). SVM models can be seen as an extension of linear classifiers, where the data separation between "positive" and "negative" examples is made by a hyperplane that maximizes the margin between classes. For non-linear relationships, SVMs use kernel functions to map the data into high-dimensional spaces where a linear separation is possible (Haykin, 2009). A Bayesian Network is a probabilistic model represented by directed acyclic graphs (DAGs), that is, a network composed of edges and nodes where edges represent conditional dependencies between variables (nodes) and their parent nodes. ML algorithms for Bayesian Networks iteratively learn the graph structure and/or the conditional probability distributions of the nodes, combining prior knowledge with observed data (Ben-Gal, 2007). The Naïve Bayes approach is a subset of Bayesian network estimation, where the DAG structure is learned by the "naive assumption" that all the variable nodes are conditionally independent given a single common parent node, and therefore, only the parameters need to be estimated. The parent node can either be assumed or defined by previous or expert knowledge (Ben-Gal, 2007). TABLE 2 presents application examples of the previously cited models. It should be noted that the Table is in no way exhaustive, and the papers were selected for having proximity with on-going research or for been highlighted by the cited review papers.

TABLE 2: EXAMPLES OF SAFETY RESEARCH WITH AI-BASED STATISTICAL MODELLING

Reference	Target variable	AI-based model used	Main results
(Zhu et al., 2022)	Pedestrian crash risk	Network-based kernel density technique and Bayesian spatial models	The presence of POIs is related to a higher risk of pedestrian-vehicle crash
(Zheng & Sayed, 2020)	Real-time crash risk, measured through conflict indicator MTTC	Bayesian hierarchical models	From the best-fitted Bayesian model, risk of crash was calculated indicated green cycles more prone to crash occurrences
(Y. Liu et al., 2025)	Perceptual indicators of wealth, life, boredom, depression, safety and beauty	Deep learning for indicator estimation and Bayesian multivariate Poisson-lognormal model for spatial effects	For all variables, distinct differences exist between the central zone and the surrounding areas
(Xiong et al., 2018)	Crash prediction	Support Vector Machine for scene classification and Gaussian-mixture-based Hidden Markov Model for crash pattern recognition	Over 95% accuracy for the scene estimation was achieved, and 85.4% correct recognition of crash occurrence by the overall model
(Chen et al., 2016)	Driver injury severity pattern from rollover crashes	Support Vector Machine models	Their best performing model achieved 94.20% accuracy with three injury severity levels

Regarding computer vision applications, although intensively reviewed for autonomous driving (Muhammad et al., 2022), broad transportation field (Gargoum & El-Basyouny, 2017; Ma et al., 2018; Outay et al., 2020) or geographic/urban studies (Cinnamon & Jahiu, 2021; Gui et al., 2024; L. Liu & Sevtsuk, 2024) applications, most of the methodologies used can be transferred for safety research as well.

(Ye et al., 2025) conducted a recent review of street view imagery (SVI) applications for safety research, and divided the models used for feature extraction from SVI into four types: semantic segmentation, object detection, depth estimation and image classification. However, there is still no specific review about how those features from SVI are integrated with safety research.

About the types of features that can be extracted through computer vision, different data sources excel in giving different outputs. TABLE 3 summarizes the main features cited by the aforementioned reviews, respecting the taxonomy given by the authors themselves, in a non-exhaustive way.

TABLE 3: MAIN FEATURES EXTRACTED THROUGH CV IN TRANSPORTATION RESEARCH

Reference	Data Source	Feature Type	Examples
(Ye et al., 2025)	Street-level imagery	Road factors	Median strips Traffic signals Intersections Barrier Sidewalk and bike lanes
		Streetscape factors	Green space Shadows cast by buildings Proportion of sky and road Trees
(Gargoum & El-Basyouny, 2017)	LiDAR	On-road	Road surface Lane markings and road edge
		Roadside	Traffic signs Roadside objects
		Geometric data	Cross section Vertical alignment Pavement condition Sight distance Vertical clearance
(Outay et al., 2020)	Unmanned aerial vehicle	-	Vehicle detection
(Gui et al., 2024)	Remote sensing	-	Vehicle
		-	Building

2.1.3 Subjective user rating

Assessing transportation infrastructure safety is demanding. While researchers have made several efforts to evaluate road infrastructures based on historical data, such as the number of crashes recorded, the rarity of such events still makes it challenging. Moreover, there is a widespread belief that evaluating infrastructure based only on objective metrics is insufficient due to the subjective nature of safety. People have different perspectives on what safety means to them. Hence, there is an increasing demand to incorporate subjective user feedback in the evaluation of transport systems. The core notion of subjective evaluation is to analyse the quality of a system based on the user's perception. Annett, (2002) argued that a genuinely holistic evaluation of systems requires not only acknowledging users' opinions but also ensuring that multiple users reach a shared interpretation.

Over the years, researchers have devised several strategies to quantify subjective perspectives. For instance, Likert Scales are used to record different degrees of human sentiment by offering a spectrum of responses instead of the usual binary choice (Koo, et.al., 2025). The System Usability Scale (SUS) is another prevalent tool that translates responses from a questionnaire into a single score, thus quantifying the subjective feedback. As many researchers consider SUS a technology-independent tool (Deshmukh, et.al., 2024), it demonstrates huge potential for evaluating road infrastructure.

Furthermore, the evolution of Artificial Intelligence (AI) has revealed new avenues for examining road infrastructure. Kim, et.al., 2023 explored the capability of text mining techniques to analyse user complaints and extract hazard categories, leading to efficient maintenance of roads. Several studies also employed techniques like sentiment analysis and topic modelling on social media data to gauge public perception about the road infrastructure issues (Torres, et.al., 2025). Melcher, et.al., (2001) utilized Bayesian methods to evaluate the effectiveness of road safety measures.

Despite the evident benefits of subjective evaluation, some challenges still remain associated with it. A key concern within the scientific community is the reliability and validity of subjective scales (Annett, 2002). Another significant challenge linked with subjective evaluation is the inherent bias in the crowdsourced data. Moreover, there is an ongoing debate about the correlation between objective and subjective safety. Various researchers claim that the Traffic risk perception of road users strongly impacts the user behaviour in traffic situations. It is observed that high perceptions of risk can lead to safer user behaviours. However, the relationship is complex, as many instances depict more reckless behaviour in high-risk situations (Boua, et.al., 2022). Consequently, the primary focus of this doctoral research would be to leverage the modern assessment technologies and balance the trade-off between subjective and objective safety.

2.1.4 User-based data exploitation

Road safety has been explored by using several types of data. Among them, crash databases are one of the primary data sources for road safety research, being the main externalities of transport systems (Imprialou & Quddus, 2019). Crash data have often been analysed using statistical models to estimate the impact of road design elements on safety, primarily as a function of characteristics related to traffic, geometry, and the environment (Hauer, 2004).

Nevertheless, the high penetration rate of smartphones or the development of On-Board Diagnostic (OBD) devices installed in vehicles offer possibilities of evaluating driver behaviour and new possibilities for faster, more accurate and low-price data collection (Boylan et al., 2024; Ziakopoulos et al., 2022), which may serve as an additional source of information for road safety analysis.

In-vehicle telematics data are dynamic data collected from single vehicles through onboard sensors, telematics devices, or smartphones. These data capture various aspects of vehicle movement and driver behaviour in real time or near real time. They have been used to improve the models used in the insurance industry, but also used as variables in various models to study driving behaviour and generally road safety key performance indicators (KPIs) (Boylan et al., 2024).

These data come from different sensors including accelerometers, which detect acceleration, vibration, and tilt to estimate the vehicle's speed and direction; Global Navigation Satellite Systems for location tracking; gyroscopes, which complement accelerometers by determining device orientation for better precision; and Engine Control Unit data accessed via the OBD-II module, providing metrics such as revolutions per minute and throttle position (Kirushanth & Kabaso, 2018). These raw data are pre-processed to generate meaningful inputs for road safety models.

The proliferation of large-scale in-vehicle data enabled by telematics technologies from diverse sources has enabled researchers to study driving behaviour within a spatial context, integrating vehicle dynamics with geographic information (Balsa-Barreiro et al., 2019).

Telematics data may be pre-processed obtaining inputs that will be consistently aggregated onto a road network based on the chosen spatial entity, such as road segments, intersections, or other specific point-based locations. Various algorithms have been proposed in order to map telematics to spatial segments (Stipancic et al., 2019; Ziakopoulos, 2021) and the use of a buffer around nodes has become a common approach for aggregating telematics in relation to point-based spatial features (Erramaline et al., 2022; Lee et al., 2025; Stipancic et al., 2018). This approach allows us to spatially account for the human factors, representing for approximately 90 to 95% of all road crashes, as previously mentioned. Consequently, it enables the analysis of data related to the majority of negative road safety outcomes and their spatial patterns across the road network.

Recently, AI tools have been increasingly used to analyse in-vehicle data in a spatial context, revealing patterns and insights into driving behaviour (P. Li & Abdel-Aty, 2022; Wang et al., 2025). The data are often used by integrating other types of data, including geometric features or road attributes derived from images enabling a more comprehensive understanding.

The growing use of technologies such as telematics, Internet of Things, image processing and AI have been increasingly applied to road safety monitoring and prediction. As a result, traditional econometric tools are being augmented by advanced digital technologies for analysis and for developing innovative countermeasures to improve road safety (Eskandari Torbaghan et al., 2022). Machine Learning (ML) and Deep Learning (DL) techniques, subfields of AI, have been increasingly explored in recent years. Studies found that these techniques improved the models' performances in comparison to statistical models in terms of predictive performance (Silva et al., 2020).

It might be argued that AI tools include a wide spectrum of techniques, including traditional econometric models, such as linear regression, which often emphasize statistical inference (James et al., 2013); machine learning models, which focus on data-driven pattern recognition and prediction (Pugliese et al., 2021); and deep learning models, characterized by their ability to learn representations of data with multiple levels of abstraction from large and complex datasets (LeCun et al., 2015). Regardless of the underlying methodology, all these approaches share the common objective of leveraging data to derive meaningful insights.

The models are applied both in an unsupervised manner for exploratory data analysis and in supervised manner for prediction and inference. Predicting crashes is the most common approach, though some studies focus on investigating Surrogate Safety Measures (SSM), such as harsh driving events, and predicting them. Regarding the inference, research has been done on interpreting the models by determining feature importance. When possible, speed consistently emerges as the most significant factor, along with other speed-related features, in both predicting crashes or harsh events.

Nevertheless, AI tools continue to face issues related to interpretability. In the case of ML models, this has been addressed using SHAP values (Lundberg & Lee, 2017), a novel approach that can help to provide insights into telematics when modelling crashes, for instance (Lundberg & Lee, 2017). Despite the growing use of these models, econometric approaches remain used, given their strength in their easy interpretability and demonstrating improved performance when incorporating spatial effects, however offering limited transferability to different contexts.

Advanced DL models remain largely unexplored in the analysis of in-vehicle telematics data for spatial applications, despite showing potential for model transferability through transfer learning techniques (Zhang & Abdel-Aty, 2022). Graph neural networks (GNNs) (Scarselli et al., 2009), which are well-suited for processing graph-structured data, have also seen limited application in telematics-based studies. Their potential to model complex spatial relationships makes them a promising tool for advancing the analysis of in-vehicle telematics data.

2.2 Overview of AI-supported proactive safety analysis methodologies

2.2.1 Network design

Although in road safety the mainstream understanding of risk management has focused on hotspot threats, alternative efforts have been made to approach safety as a matter of either the local or the broader road network (Lovegrove & Sayed, 2006; Marshall & Garrick, 2010; Moeinaddini et al., 2014). Consequently, AI-supported methods for analysing the safety of road networks have also emerged.

Research in this area has evolved in tandem with advances in handling network-like data structures within the field of AI. An initial line of work focused on adapting classical machine learning methods. For example, building on the success of word embeddings in natural language processing, Grover and Leskovec (2016) introduced the node2vec framework, which automatically learns meaningful representations of nodes within a network. In node2vec, the network is explored by simulating 'walks' that traversed from node to node. Flexible biases guided these walks—sometimes remaining close to the starting point and at other times moving further away—which enables the method to capture both local relationships and broader structural patterns. After generating these random walks, the researchers applied the Skip-gram model from natural language processing. Skip-gram produces low-dimensional vector representations—known as embeddings—that place similar items close to one another in the representation space. In a manner similar to how Skip-gram learns the meanings of words through embeddings, node2vec learns the functional 'roles' of nodes within the network by producing embeddings that encoded their structural context.

By adapting node2vec into the StreetNode2Vec framework, Huang et al. (2020) analysed traffic crashes on a road network based on road connections and traffic volume. In StreetNode2Vec, walks were not created randomly; instead, the researchers reconstructed actual travel routes by combining GPS trajectory recordings from taxis in Porto with OpenStreetMap (OSM) data. Andrew Kwok-Fai Lui (2021) introduced the concept of the "accident neighbourhood," developed to ensure consideration not only of the crash itself but also of the nearby road segment features in black spot identification. In this method, embeddings were first trained by applying the node2vec algorithm over the network three times, each time using a different property of the road segments: maximum speed, transit time, and length. Next, a deep neural network (DNN) classifier was designed to predict whether a road segment constituted a black spot.

A further advance in processing road network data with AI was the introduction of the self-attention mechanism (Vaswani et al., 2017), which enables data points within a sequence to attend to one another in parallel. Crucially, self-attention is permutation-invariant: it does not rely on a fixed order, but instead learns pairwise relations based on feature similarity. This property makes it particularly suitable for network-like data, where neighbourhoods are unordered sets of nodes. Building on this, Graph Attention Networks (GATs) (Veličković et al., 2017) extended self-attention to the graph domain by allowing each node to assign learnable importance weights to its neighbours. Concurrently, there has been growing interest in combining spatio-temporal correlations into the risk assessment of road networks, leveraging spatial information together with temporal attributes within time windows—such as rainfall, humidity, visibility, sunrise and sunset times, and holiday indicators, which vary daily.

Gao et al. (2023) developed a Zero-Inflated Tweedie Distribution (ZITD) framework. A key challenge was that most time windows contained no crashes at all. To address this, the model introduced a sparsity parameter, which enabled it to shift its focus away from periods where crashes were absent. The model encoder combined a Gated Recurrent Unit (GRU) and a Graph Attention Network (GAT), where the GRU accounted for temporal dynamics, including patterns influenced by attributes such as holidays, and the GAT accounted for spatial information. Mimi et al. (2025) introduced ST-GraphNet, a framework designed to predict the severity of crashes involving Automated Vehicles (AVs). Their dataset combined structured details—such as time, location, and automation level—with short descriptions of the event. These descriptions were converted into numerical embeddings using a pretrained language model. The authors then tested two setups: (1) a fine-grained graph where each crash event was a node connected to nearby crashes in space and time, and (2) a coarse-grained graph where crashes were aggregated into hexagonal regions, with each region represented as a node.

Nobin and Rifat (2025) developed STARN-GAT, a multi-modal GAT for predicting crash severity by combining spatial, temporal, and contextual information. They first transformed road networks into graphs, where each road segment was represented as a node, and edges represented not only physical links but also spatial proximity and functional similarity. Crash records were enriched with spatial features (e.g., curvature, width, slope), temporal features (e.g., time of day, weekday, month), and contextual features (e.g., weather, visibility, traffic density). Spatial relationships were processed using a GAT, temporal patterns were handled with a tailored deep network, and contextual data were processed with a smaller neural model. These outputs were then fused using an attention-based mechanism, enabling the model to prioritise the type of information most relevant for each case. The fused representation was passed into a classification network to predict injury severity (no injury, minor, moderate, severe).

Finally, several efforts have been made to publish suitable datasets to support further studies. Although not all these initiatives introduced a methodology to analyse safety, their contributions can be understood in three ways: (1) they introduced methods for verifying data quality using existing models; (2) they set a baseline for other researchers by providing reference datasets; and (3) some directly created models. For example, recognising the limitations of existing machine learning approaches, which focused on risky locations in isolation, Huang et al. (2023) developed the Traffic Accident Prediction (TAP) data repository, incorporating real-world road network topology and geospatial features. The dataset was organised under two main branches by scope: TAP-city and TAP-state. OpenStreetMap (OSM) was used to build the road networks, and nodes were marked with binary labels indicating whether a crash had ever been observed at the location, while crash types were categorised into eight severity levels. Furthermore, the authors developed the Graph Neural Network (GNN)-based Traffic Accident Vulnerability Estimation via Linkage (TRAVEL) framework, where the GNN's message passing was augmented with angular and directional information, encoding the turning angle and spatial direction, respectively.

In parallel, Nippani et al. (2023) addressed the challenge of fragmented accident data by building a unified dataset of over nine million crashes from eight U.S. states, enriched with traffic volume, road structure, and weather features. To gather the data, they investigated official Department of Transportation records published in several formats. They also collected the corresponding road networks from OpenStreetMap, annual average daily traffic counts (where available), and aligned local weather data from meteorological stations. Using this dataset, they evaluated a broad spectrum of models, from classical embeddings (node2vec, DeepWalk) to modern GNNs and spatio-temporal variants. For each edge, the models attempted to forecast both the number of crashes and to classify whether at least one crash would occur. To capture patterns across states, they also developed multitask learning models, where each state had its own prediction layer to capture state-specific patterns but also shared a common encoder. This approach introduced a new method where data-rich states could "teach" models useful structures that helped data-sparse states. Finally, to confirm that the shape of the road network itself carried critical predictive power, ablation studies were conducted in which graph-structural features such as connectivity and centrality were removed.

Overall, the review of the recent studies reveals that analysis methodologies based on individual elements of roads, such as iRAP's Star-Rating System (International Road Assessment Programme, 2019), have not reached their complete potential because (1) critical interdependencies between the elements are lost (Wang & Cao, 2021, as mentioned in Gao et al., 2023), (2) focusing on individual elements yields noisier data (Mimi et al., 2025), (3) road networks naturally form graphs (Bruna et al., 2014 as mentioned in Nobin & Rifat 2025) and (4) focusing on individual elements neglects network-specific tendencies (Nippani et al., 2023). Therefore, a network-level evaluation method is needed to enable higher-quality contextual information.

2.2.2 Conflict analysis

Traffic conflict analysis has transitioned from labour-intensive, observer-based techniques to sophisticated computer vision methods. Initial approaches—like the Swedish Traffic Conflict Technique (TCT) introduced in the 1970s—depended heavily on trained observers who manually identified conflicts, defined as instances where two road users were on a collision trajectory unless one took evasive action (Hydén, 1987). According to (Amundsen, 1977), a conflict is "an observable situation in which two or more road users approach each other in space and time to such an extent that there is a risk of collision if their movements remain unchanged." In these early methods, observers would classify conflict severity using parameters such as closing speed, angle of approach, and evasive maneuvers (Archer, 2005; Zheng et al., 2014). While foundational to traffic safety research, these methods were inherently subjective, time-consuming, and limited in scale.

Today, video-based analysis has become the norm, extracting road user trajectories and computing surrogate safety indicators—enabling proactive conflict detection without requiring human presence. For instance, (Samara, 2021) proposed a comprehensive video-based safety framework in which cameras record road activity and algorithms assess metrics like Time-to-Collision (TTC) and Post-Encroachment Time (PET). Conflict detection now relies on high-resolution, multi-modal video data, processed using deep learning and computer vision algorithms to track interactions between vehicles and pedestrians that meet specific risk criteria.

At the core of modern conflict analysis are surrogate safety indicators, which serve as proxies to assess risk in the absence of actual crashes. Among these, temporal and kinematic metrics are most widely used. TTC indicates the time remaining before a collision would occur if velocities were maintained, while PET measures the time gap between one user exiting a potential collision zone and another entering it. Other indicators include the Deceleration Rate to Avoid a Crash (DRAC) and delta-V—an energy-based measure traditionally used in crash reconstruction that estimates potential crash severity. Delta-V has since been integrated into automated tools like the FHWA's Surrogate Safety Assessment Model (SSAM) for conflict analysis (Shelby, 2011). Upon identifying a conflict trajectory, systems compute delta-V to approximate the severity of a potential crash.

Since surrogate indicators require predefined thresholds, extensive research has gone into calibrating and understanding their significance. For example, SSAM adopts a TTC threshold of 1.5 seconds and PET values grounded in Hydén's original work (Archer, 2005). Contemporary methods often integrate these surrogate metrics with machine learning for enhanced detection. (Jalayer, 2022) developed an AI-driven tool combining YOLOv5 object detection with DeepSORT tracking, extracting vehicle trajectories at intersections and calculating TTC and PET across all vehicle pairs. Their system reached over 92% detection accuracy and automatically generated risk metrics and counts of hazardous behaviors. Additionally, extreme-value theory was applied to TTC and PET distributions to estimate crash probability. Similarly, (Samara, 2021) utilized network-wide video tracking to develop conflict matrices and evaluate speed violations.

Automated conflict analysis presents several key advantages. Video data enables comprehensive monitoring across multiple sites, far exceeding the capacity of human observers (Tarko, 2018). These systems generate timestamped trajectory data, capture rare incidents like jaywalking or abrupt stops, and facilitate integration of multiple risk indicators (TTC, PET, DRAC, ΔV) for deeper safety insights. Crucially, surrogate indicators shift safety analysis from reactive to proactive—allowing early identification of high-risk locations and simulation-based intervention testing.

Data quality issues persist—such as occlusions, poor lighting, and sensor inaccuracies—can lead to noisy or incomplete trajectories (Tarko, 2018; Zheng et al., 2014). Object tracking algorithms may misidentify or lose track of road users in dense environments. Privacy is another concern, as video may inadvertently capture personal identifiers like faces or license plates, necessitating proper anonymization (European Data Protection Board, 2020). Methodological limitations also exist as fixed camera setups hinder scalability, many studies remain in the proof-of-concept stage, and there is limited benchmarking across systems or validation against real-world crash data.

Explainability in AI systems will play a vital role in future. Analysts need clear insights into why specific interactions were flagged, along with the contributing factors (e.g., speed, TTC, deceleration). Systems capable of annotating conflicts with interpretable descriptors—such as "vehicle A crossed in front of B with 0.5 s TTC and 6.0 m/s^2 deceleration"—will greatly enhance usability. Multimodal data integration also holds promise. Incorporating video with GPS, connected vehicle data, or radar can help mitigate issues like occlusion and improve near-miss detection (Li, 2020). Scaling up to city-wide or national-level deployments will likely depend on distributed or federated learning models that can adapt to different traffic layouts and cultural contexts. Early studies using graph neural networks suggest improved tracking and conflict detection through relational reasoning across multiple vehicles (Weng, 2020).

In conclusion, traffic conflict analysis has evolved from manual observation to AI-based systems. Surrogate safety measures like TTC, PET, and ΔV now offer quantitative insights into crash risk (Shelby, 2011), and computer vision enables these metrics to be computed on a large scale. While current applications demonstrate significant potential, challenges related to data integrity, trust, and scalability must still be addressed. Future research should prioritize transparent AI, multimodal integration, improved simulation tools, and scalable methodologies. These developments could enable truly proactive safety management—identifying risks and informing interventions well before accidents occur (Zheng, 2014).

2.2.3 User perception

The field of road safety has been fundamentally reactive. Although this approach has been fairly effective in identifying high-risk locations, it requires several tragic events to occur before any action is taken and measures are suggested. Moreover, the supporting data is often dated and insufficient, leading to inaccurate analysis (Vasudevan, et.al., 2022). Therefore, road safety professionals and researchers are shifting towards a proactive "Safe System" approach by designing systems that are more accommodating and human-centric, aiming to prevent crashes and reduce the severity of outcomes before incidents occur.

Road networks are complex environments due to the sheer number of people interacting with them on a daily basis. It is widely believed that user perception is a crucial component of traffic safety research due to its correlation with road user behavior and their perception of risks(Yu, et.al., 2024). Studies suggest that unsafe driving behaviour, often resulting from inadequate risk perception, is among the leading causes of accidents (Yang, et.al., 2025). Several studies demonstrate that users' risk perception is also driven by a multitude of factors. For instance, Iamtrakul, et.al., (2023) examined the impact of different modes of travel on the perceived risk. The study revealed that vulnerable road users, like cyclists and pedestrians, are more aware of risky behaviors than others. Pajković and Mirjana (2021) conducted an age-related comparison of driving behaviors and risk perceptions of different user groups and found that certain risky behaviors are more common in young users owing to several factors, such as lack of experience and overestimation of driving ability. Thus, a significant aspect of the proactive safety analysis framework is accounting for user perceptions in safety analysis.

Throughout the years, researchers have employed several techniques to measure this perceived risk, such as conducting extensive surveys to collect useful insights about public opinion on various safety measures (Alonso, et.al., 2022). Some studies have also investigated the potential of Virtual Reality (VR) technology to evaluate the acceptability of new infrastructural designs (Argota, et.al., 2024). However, the literature also highlights several complications associated with evaluating subjective perceptions. Useche, et.al., (2022) conducted a systematic review of behavioral studies in the domain of road safety and observed that road users often perceive themselves as "safer" than others, leading to an overestimation of their abilities and undervaluation of the skills of other road users. The study also highlighted that drivers often engage in risky behaviors like speeding, even while acknowledging the associated dangers, indicating a discrepancy between stated risk and action. Some studies illustrate that people's perception of safe systems is significantly influenced by public awareness and familiarity with the systems(Stokes, et.al., 2019). Artificial Intelligence(AI) has emerged as a powerful tool for evaluating subjective perspectives. The high computational power and reasoning capabilities of these algorithms can aid in identifying complex patterns and capturing cognitive biases.

Many researchers are examining the analytical capacity of artificial intelligence models to predict user perceptions in urban environments (Wu, et.al., 2025). Some studies have demonstrated the inherent capability of Large Multimodal Models (LMMs) to assess public perceptions about the safety of urban spaces (Beneduce, et.al., 2025). Researchers are also exploring the potential of AI to measure the perceived risks of road users under varied circumstances. Rasch, et.al., (2022) developed a Bayesian ordinal logistic regression model to measure the perceived risk of cyclists and drivers during an overtaking event. A fundamental requirement of a human-centered, proactive safety approach is developing clarity on the psychological and perceptual factors that motivate an individual to engage in risky behaviors. Many researchers studied the interactions between cyclists and drivers to gain a deeper understanding of factors influencing drivers' decisions to overtake(Farah, et.al., 2019). The study investigates the ability of Binary logistic regression models to predict the drivers' overtaking strategy and the lateral comfort distance from the cyclist.

Though the AI-driven proactive safety frameworks depict promising results, they also have some challenges that must be carefully navigated. In a study, Garcia, et.al., (2025) observed that there is a lack of accurate understanding of the limitations inherent with AI, leading to frequent overtrust in AI-based driving systems. Other crucial limitations associated with these systems include the lack of high-quality data, intrinsic human bias, and the generalizability of AI safety models (Artificial Intelligence (AI) for Intelligent Transportation Systems (ITS)). Based on the above discussion, it can be concluded that while Artificial Intelligence is a promising tool for advancing proactive safety, it is still an emerging field that demands continuous efforts in ethical design and curation of quality datasets.

2.3 Overview of video image processing methodologies for road safety analysis

2.3.1 Infrastructure attribute extraction (network element labelling)

The methodology developed by the International Road Assessment Programme – iRAP is currently the most complete method for network-level safety assessment without using crash data (Paliotto et al., 2024). Given street-level video recordings of the road network, it consists of manually labelling a series of elements for each 100 m segments, which are then used to calculate the safety rate of the segment on a scale of 1 to 5 stars (IRAP, 2024). Since 2019, iRAP has been working on the next step of their methodology: incorporating artificial intelligence and non-human based data sources for star rating the roads. The main goal is to shift from manual labour to computational-based features, so that humans can be better sorted in areas that artificial intelligence cannot reach, and to decrease network labelling costs.

Although the list of suppliers capable of providing infrastructure information keeps growing, it has not yet reflected into the academic community. Research-wise, open-source datasets and codes are preferred, and it has made street-level imagery-based safety assessment rely not on specific attribute extraction, but rather broader information on the environment.

Through a systematic review on Scopus database regarding SVI-based road safety and transportation research in the past decade, 32 papers that use the SVI or its features as input for a transportation-related assessment were found. Those papers can be divided into three regarding their main type of application: safety assessment (either crash risk or safety perception), vulnerable road user (VRU) accessibility assessment, and route/path proposal.

Figure 2 outlines a broad overview of methodological steps when working with SVI data for transportation assessment: after data collection, environment features can optionally be extracted from the SVIs either through segmentation, object detection or deep features (features extracted from the second-to-last layer of a deep learning encoder before applying those features to a second AI model). The second optional step would be to aggregate either those features or the raw images into indicators, such as greenery or complexity, or use them to cluster the input into types of scenes. Those steps are optional because either one of them, or combinations of them, can be used as input for the most crucial part of the transportation-related assessment, which in the workflow is described as “statistical modelling” to encompass all the different methodologies, regardless of whether they apply Machine Learning. Through the statistical modelling, the main research output is obtained, although in some cases it is preceded by some spatial analysis. For the last optional bit, when working with ML modelling, it is common approach to use an interpretability method to explain feature importance, for example, Class Activation Maps (CAM) or Shapely Additive Explanations (SHAP). Although this workflow focuses on SVI data, other types of data sources are mentioned since they are of the utmost importance for transportation research and are often used alongside the SVI features not only for statistical modelling, but also for aggregated indicator calculations and cluster scene labeling.

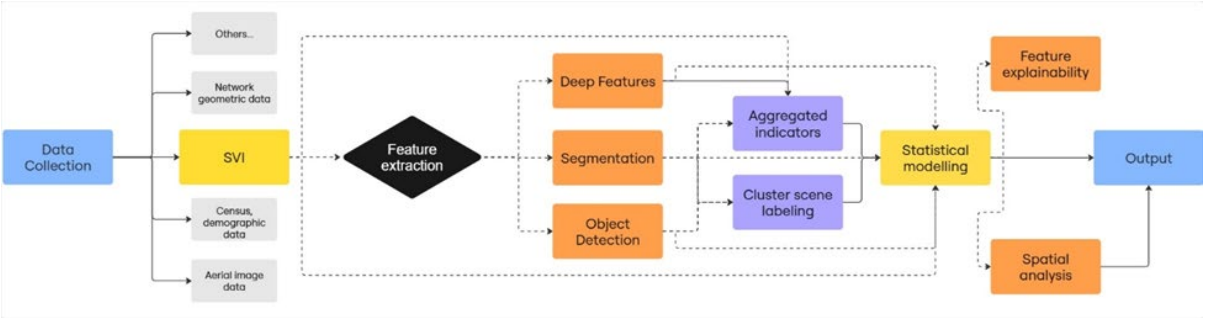

FIGURE 2: METHODOLOGICAL BASE WORKFLOW FOR SVI-BASED TRANSPORTATION RESEARCH

Only one reviewed paper process SVI data directly for their desired output. (Lin et al., 2024) trained a deep learning model (ResNet-50) to quantify cyclist stress level using Google Street View. The cycling stress labels were obtained following previously proposed methodologies based on detailed road network data. SVIs were also used directly for route planning by (Hosseini et al., 2024), whose team manually evaluated sidewalk context to feed this information into a graph relative to the road network designed to suggest best walking paths, taking into consideration specific pedestrian needs; and by (Bock & Verstockt, 2021), who applied Convolutional Neural Net Image Retrieval (CIR) to calculate the perceived complexity score from the images.

Deep features were also only found in one paper: (Guo et al., 2024) fused the last layers of two different deep learning models, one applied on SVI and other on aerial images, to identify crash-prone areas in the Xiaoshan District of Hangzhou, located in Zhejiang Province, China. They used a specific CAM model, Grad-CAM++ (Chattopadhay et al., 2018), to highlight what parts of the image were the most influential for the full model output.

The remaining 28 papers all used semantic segmentation features, and 8 of them also added object detection features. Although many of these papers use aggregated features and/or road scene clustering from the extracted features, there is not one specific rule of how those steps are integrated into the flowchart. There is, however, a clear prevalence of specific models for features extraction. For segmentation, the most used models are PSPNet (Bustos et al., 2021; Doiron et al., 2022; S. Hu et al., 2023; G. Huang et al., 2025; Y. Huang et al., 2024; Kwon & Cho, 2020; Rita et al., 2023; Stiles et al., 2022; Toohey et al., 2022; Y. Wang et al., 2022; Yue, 2024, 2025) and DeepLabV3 or DeepLabV3+ (He & He, 2024; C. Hu et al., 2023; Jeon & Woo, 2024; Y. Liu et al., 2025; Mishra et al., 2023; Qin et al., 2025; M. Wang et al., 2022; Yap et al., 2023; Zhang et al., 2024). For object detection, the most used models are YOLO (Doiron et al., 2022; Kwon & Cho, 2020; Rita et al., 2023) and R-CNN (Toohey et al., 2022; Yue, 2024, 2025), although the R-CNN models have slightly different architecture among each other. Usually, model is directly transfered with pre-trained weights of big databases, such as

Cityscapes (Cordts et al., 2016) and ADE20k (Zhou et al., 2017) and occasionally fine-tuned with specific data relative to the case study area.

For an overlook of methodological steps on SVI-based road safety research, Table 4 provides a summary of the steps taken, the main model used and the research field, where the main prevalence of semantic segmentation, often used for building aggregated indicators can be more evident.

TABLE 4: METHODOLOGICAL CHARACTERISTICS OF SVI-BASED ROAD SAFETY RESEARCH

Reference	Semantic segmentation	Object detection	Deep features	Aggregated indicators	Scene clustering	Main model(s)
(Kwon & Cho, 2020)	✓	✓			✓	Negative binomial
(Bustos et al., 2021)	✓			✓		ResNet50
(Stiles et al., 2022)	✓				✓	Negative binomial
(M. Wang et al., 2022)	✓			✓		Spatial lag
(Y. Wang et al., 2022)	✓	✓		✓		Squeeze-and-Excitation Network (SENet)
(S. Hu et al., 2023)	✓			✓		Geographically Weighted Poisson Regression (GWPR)
(Mishra et al., 2023)	✓					CNN
(Rita et al., 2023)	✓	✓				Pearson correlation
(C. Hu et al., 2023)	✓			✓		Gradient Boosting Decision Tree (GBDT)
(L. Liu et al., 2023)	✓			✓		Polynomial regression fitting
(Yue, 2024)	✓	✓				XGBoost
(Hamim & Ukkusuri, 2024)	✓					Random Forest
(Zhang et al., 2024)	✓			✓		XGBoost
(Jeon & Woo, 2024)	✓				✓	Two-level negative binomial regression
(Iamtrakul et al., 2024)	✓				✓	Statistical correlation
(Guo et al., 2024)			✓			Two-Branch Contextual Feature-Guided Converged Network (TCFGC-Net)
(Nguyen et al., 2024)	✓	✓				Poisson regression
(Y. Liu et al., 2025)	✓			✓		Bayesian multivariate spatial-varying coefficients model
(Yue, 2025)	✓	✓		✓		Zero-inflated Negative Binomial (ZINB)

As can be seen by Table 4, the association of features driven from semantic segmentation models as input for road safety assessment has become quite popular on the field, perhaps incentivized through open-source and ease-access models and databases. More often than not, the main recurring categories are aggregated into indicators, for example: greenness or green-looking ratio (proportion of any green-related pixels in the image, including trees, grass, plants); openness (proportion of sky pixels); enclosure (proportion of enclosing-space pixels such as buildings, walls and trees over ground-related pixels such as road, sidewalk, fence and earth) (Huang et al., 2025; Liu et al., 2023; Toohey et al., 2022; Zhang et al., 2024). Supporting what was brought on the beginning of this session, Table 4 also highlights the use of additional data from object detection and scene clustering for safety assessment with less but yet meaningful frequency, and the lack of a general consensus on what type of statistical analysis best fits this type of data.

2.3.2 Infrastructure attribute extraction (infrastructure data fusion)

Data fusion is becoming more and more important for road safety analysis, as single data sources often cannot capture all necessary attributes of a road environment. Combining multiple sensors or data types provides complementary perspectives and redundancy, thereby improving the robustness and completeness of safety assessments (Micko, Papcun, and Zolotova 2023). For example, visual cameras offer rich semantic information but lack direct depth measurement, while LiDAR provides precise 3D geometry but limited semantic detail. By fusing both data sources, we can overcome each sensor's deficiencies (Tan et al. 2025). Existing studies (Brkić et al. 2022; Kačan et al. 2020; Kacan, Sevrovic, and Segvic 2024) show that not all road attributes can be automatically identified using computer vision by a single sensor. This is why multi-sensor fusion for road safety features extraction is necessary to identify road safety-related features more accurately and efficiently (Brkić et al. 2022). In short, data fusion allows for a more comprehensive infrastructure assessment by combining multiple sources of data into a cohesive analysis, which is crucial for identifying risk variables such as precise object distances and road geometry that would otherwise be ignored by a single data source.

Fusion strategies are commonly categorized as early (low-level), mid-level, or late (high-level) fusion, depending on the stage at which data integration occurs. In early fusion, raw data from multiple sensors (e.g., unprocessed camera pixels and LiDAR point clouds) are combined before feature extraction. This preserves fine-grained information from each modality (Panduru and Walsh 2025). For example, projecting LiDAR depth data onto camera images can create dense depth maps that make it much easier to find objects and estimate their distance (Dong et al. 2024). However, early fusion demands precise spatial and temporal calibration between sensors and imposes heavy computational loads, as large volumes of raw data must be synchronized and processed jointly (Panduru and Walsh 2025). On the other hand, mid-level fusion merges features extracted independently from each sensor (e.g., combining a camera's visual features with a radar or LiDAR's depth features) (Thakur and Mishra 2024).

This feature-level fusion can balance accuracy and efficiency: it retains important modality-specific features (like object edges from images and range profiles from LiDAR) while reducing data dimensionality, making real-time processing more tractable. Late fusion (decision-level) integrates outputs or decisions from separate sensor-specific analyses. For instance, combining the detected objects or classifications from a camera-based model and a LiDAR-based model (Jahn et al. 2024). Late fusion is modular and computationally lightweight (each sensor can be processed independently), and it adds fault tolerance since the system can fall back on other sensors if one fails. However, the problem with late fusion is that it could overlook important details by just looking at high-level outputs. Subtle cues available in raw data might be missed once decisions are made per sensor (Panduru and Walsh 2025). In summary, early fusion offers information richness (at the cost of complexity), mid-level fusion offers a compromise, and late fusion offers simplicity and flexibility. The choice depends on the requirements of accuracy, computing resources, and the nature of the infrastructure data being analysed.

Implementing data fusion in road infrastructure attribute extraction has many challenges and applications. A key challenge is achieving precise sensor calibration and alignment so that LiDAR point clouds and camera images can be properly overlaid. Errors in calibration or timing can lead to misidentified road locations, degrading the analysis results. Despite these difficulties, recent research demonstrates the importance of fusion for road safety. For instance, Ural et al. (2015) (Ural et al. 2015) fused aerial infrared orthophotos with mobile LiDAR scans to improve road detection. The LiDAR ground data helped filter out buildings of similar color to pavement, enabling the extraction of over 90% of roads in their study area. Similarly, Han et al. (2017) (Han et al. 2017) projected LiDAR onto monocular images and combined color and depth cues via an AdaBoost classifier for road segmentation. This early fusion approach improved detection performance (by leveraging depth to reduce false positives), although the authors noted that sensor noise at long range still caused some errors. Ozturk et al. (2023) (Ozturk et al. 2023) combined high-resolution satellite imagery with airborne LiDAR for extensive road mapping. Their approach, utilizing a U-Net model with feature-level fusion, fused CNN features from RGB images with LiDAR-derived geometric features. This strategy resulted in a continuous 1-5% improvement in road pixel prediction accuracy, which was especially useful in places hidden by dense woods or shadows where optical data was insufficient. Further evolving fusion techniques, Guo et al. (2024) (Guo et al. 2024) introduced DTRoadSeg, a Transformer-based dual fusion network that integrates camera and LiDAR data for enhanced road segmentation. Their model employs separate Transformer encoders for RGB images and LiDAR depth maps for extracting heterogeneous features. These features are then fused via a Heterogeneous Feature Fusion Module that leverages self-attention tokens. Their proposed methodology achieved approximately 97% accuracy and 96.3% recall on the KITTI benchmark, outperforming previous CNN-based methods while maintaining real-time performance. However, in the context of iRAP road attributes, multi-sensor fusion has not been applied yet, and the above studies are independent of iRAP standards.

2.3.3 Trajectory extraction

The extraction of vehicle (and pedestrian) trajectories from video has evolved from manual to automated techniques. In the earliest days, researchers manually digitized videos: analysts marked vehicle positions frame-by-frame on recordings or synchronized maps to reconstruct paths (Archer, 2005). (Shelby, 2011) notes that classic conflict studies were conducted with a team of observers manually identifying vehicle interactions (Shelby, 2011). This process was slow and limited to a few hours of footage per site. As computer vision emerged, semi-automated methods appeared. Early tracking relied on background subtraction, optical flow, or template matching (Lucas, 1981). These methods worked in controlled conditions but struggled in heavy traffic, shadows, or with camera motion.

The last decade brought a revolution with deep learning. Modern pipelines typically follow two stages: (1) detect objects of interest in each frame, and (2) associate detections across frames to form trajectories. CNN-based detectors such as Faster R-CNN and YOLO achieve high accuracy (Ren., 2015; Redmon et al., 2016). Once objects are detected, multi-object tracking (MOT) algorithms link them over time. The classic SORT algorithm inspired many improved variants. DeepSORT (Wojke, 2017) added CNN-based appearance descriptors for stronger associations, reducing identity switches by about 45%. DeepSORT can maintain IDs even through brief occlusions. ByteTrack (Zhang, 2022) improved tracking by associating nearly all detections, including low-confidence ones often discarded as false positives. By piecing together weak signals, it recovers objects lost during occlusion and significantly boosts IDF1 scores. ByteTrack excels in crowded urban scenes with frequent overlaps. FairMOT (Zhang, 2021) takes a different approach, combining detection and re-identification in one network. It uses an anchor-free detector (CenterNet) with dual heads for localization and embedding, balancing both tasks. FairMOT performs well in dense, overlapping scenes and offers efficient inference for real-time use, though it depends heavily on accurate detections and can falter in poor visibility. StrongSORT (Du, 2022) revisits DeepSORT with enhanced detectors, embeddings, and association strategies. Two new modules extend its robustness: AFLink (which re-links fragmented tracks without relying on appearance) and GSI (which interpolates missing detections with Gaussian processes). StrongSORT++ achieves state-of-the-art results on challenging MOT benchmarks by smoothing tracks and reconnecting fragments in crowded scenes. Comparing trackers highlights trade-offs such as DeepSORT's appearance features are strong when vehicles have distinct patterns, but it struggles with similar-looking objects. ByteTrack excels in dense occlusion but risks false associations. FairMOT balances detection and re-ID but is vulnerable in low-visibility conditions. StrongSORT handles fragmented tracks well but still relies on its base tracker for long occlusions. In practice, FairMOT and StrongSORT outperform DeepSORT in crowded intersections, while DeepSORT may be preferable for high frame-rate, simpler scenes.

Trajectory extraction requires accurate geometric calibration—specifically, mapping image coordinates to real-world positions using homography transforms. Early methods depended on manual camera calibration for each site, which is time-consuming and non-scalable. More recent automatic approaches infer scene scale by aligning detections with 3D vehicle models (Sochor, 2017), or by using the physical camera height and detecting vanishing points—work well for wide-angle lenses (Jain, 2020). Robust solutions have also been developed for low-quality CCTV streams, where vanishing point detection is combined with iterative refinement to achieve stable calibration even under poor imaging conditions (Dubská, 2014). State-of-the-art techniques leverage synthetic views and graph neural networks to learn homography matrices without manual intervention (D'Amicantonio, 2023). Even aerial imagery approaches effectively use known road features (e.g., circular lanes) for homography calibration (Riehl, 2025). Collectively, these advances enhance scalability and accuracy across diverse real-world traffic conditions.

Trajectory data supports broader safety frameworks. One application is automated conflict matrices or heatmaps, which classify and quantify interactions (e.g., rear-end, crossing, pedestrian) using TTC or PET. Automated systems scale beyond manual studies, covering weeks of data. Another application is simulation calibration: real-world trajectories and conflicts tune microsimulation models to better replicate actual behaviors. For example, car-following or gap-acceptance parameters can be adjusted so simulated conflicts match observed ones. Trajectories also play a role in Vision Zero, where continuous monitoring evaluates interventions—such as new crosswalks—by tracking changes in near-miss rates (Samara, 2021).

Despite progress their are some challenges which remain such as scalability is difficult: running deep detectors across hundreds of cameras requires significant computing power, and edge/cloud solutions are still evolving. Transferability is another issue: a tracker tuned for one city's road geometry or fleet may not perform equally well elsewhere. Domain adaptation and federated learning aim to improve generalization (Li, 2020). Extreme density is also problematic, especially in megacities where occlusion is near-total and many objects appear similar, causing fragmentation even for advanced trackers.

Graph neural networks (GNNs) represent traffic environments as evolving graphs, allowing for a holistic understanding of relationships among tracked entities (Li, 2020; Weng et al., 2020). This approach reduces ambiguities that pairwise matching misses, making it promising for crowded intersections. Federated learning is another frontier, enabling models to train across multiple cities without sharing raw video, thus preserving privacy. Reducing dependence on manual calibration is also a priority. Advances in unsupervised learning may eventually allow systems to infer ground-plane homographies directly from scene patterns.

Trajectpry extraction has moved from manual digitization to advanced AI-driven tracking. State-of-the-art algorithms like DeepSORT, ByteTrack, FairMOT, and StrongSORT reliably convert video into timestamped trajectories with consistent identities (Wojke, 2017; Zhang et al., 2022). These trajectories underpin modern safety analytics by enabling large-scale surrogate computation and integration into models. Each tracker has trade-offs—DeepSORT and StrongSORT excel with distinctive features, ByteTrack thrives in dense traffic, and FairMOT balances detection with re-ID. Often, ensembles are used to mitigate weaknesses. Future systems may be more robust to lighting, weather, and density, and may directly output conflict probabilities or learn from crash outcomes. Better trajectories enable better safety metrics, and safety needs in turn guide the design of next-generation trackers.

2.3.4 User perception

As road safety shifts towards proactive approaches, risk interpretation has evolved. Rather than being a mere representation of past events, risk is increasingly perceived as a characteristic of the ever-changing transportation systems. This shift demands rigorous systems for continuous monitoring and evaluation of these systems, facilitating prompt safety interventions. The complexity of the road transportation systems and the extensive volume of people interacting with these systems on a daily basis render this a highly demanding and labor-intensive process.

The advancement in the AI-driven video image processing techniques has revolutionized the field. A growing body of research highlights the capacity of AI-enabled systems to advance developments in the road safety landscape. For instance, Zdravković et.al., (2025) outline the competence of AI-supported computer vision frameworks in the early detection of high-risk behaviors like harsh braking, speeding, and route-adherence. Many researchers have identified the significant capabilities of modern technologies, like artificial intelligence and image processing, in advancing data collection, analysis, and road safety evaluations (Torbaghan, et.al., 2022). The robust processing power of these algorithms has streamlined the once labour-intensive process of data-gathering. Numerous studies emphasize deep learning's capacity to automate the detection of infrastructure damage, poor road conditions, and missing traffic signs from visual data (Ranyal, et.al., 2022).

There is a widespread belief among the scientific community that road safety is a multilayered subject, requiring inclusive infrastructure, human-centric designs, and data-driven policy considerations. Moreover, the ongoing transition towards proactive architectures highlights the necessity to gain a more holistic understanding of risk by accounting for both subjective and objective factors. Building on these foundations, several researchers are investigating methodologies to incorporate the subjective human perception in safety systems. For example, de Winter, et.al., 2023 illustrated the potential of computer vision models to predict the perceived risk of traffic scenes. Another study proposed a Video Analysis of Pedestrian movement (VAPM) to derive behavioural cues from video data of pedestrians navigating various environments. The study examined the relationship between pedestrians' self-reported perception of the walking infrastructure prior to walking, and their observed walking behaviour (Johansson, et.al. 2020). Some researchers leverage advanced machine learning algorithms to analyse video data from in-vehicle cameras to monitor driver behavior through observable physiological indicators, including the driver's facial expressions, hand placement on the wheels, and the driver's body postures, and alert them about potentially unsafe behaviors(Fangming, et.al., 2024).

Several studies in this domain also focus on strengthening safeguards for vulnerable road users. For example, a group of scholars proposed an advanced AI framework for interpreting pedestrian behavior at busy intersections(Zhou, et.al., 2024). The study employed sophisticated models like Graph Convolutional Networks (GCNs) and encoder-decoder architectures to examine video data and predict the intention of pedestrians to cross a street. In another study, Kchour, et.al., (2024) presented a systematic approach to exploit computer vision techniques for examining cyclists' eye movements and assessing their visual attention, workload, and hazard perception. In addition to the direct applications of Artificial

Intelligence(AI) in road safety, the scientific community is also exploring innovative approaches to incorporate subjective perception in AI-based decision-making. Chan et al. (2023) proposed a systematic methodology to define the cost of confusion between different categories, such as humans, dynamic objects, roads, and others, based on public opinion. The study emphasized the significance of incorporating public ethical opinions into the design of AI systems for automated driving.

Though the advancement in deep learning models and computer vision techniques has showcased remarkable potential, it is still associated with several challenges. Studies reveal that despite the significant performance of the advanced visual analysis frameworks, the accuracy of these models depends on various factors, as environmental conditions and local traffic policies (ElSahly, et.al., 2022). Another complication inherent in AI-powered systems is the possibility of biased training data and algorithms, which can hinder the reliability of these systems(Radanliev, 2024). There are also ethical and privacy concerns linked with the processing of sensitive visual data. In conclusion, the evolution of AI-based visual data processing techniques has demonstrated immense potential to transform the transportation research landscape. However, it is crucial to devote equal attention to developing AI-enabled proactive frameworks that are not only reliable but also transparent and ethically accountable.

2.4 Overview of AI-supported spatial scaling road safety analyses

Since road transport involves movement across space, spatial analysis naturally plays a crucial role in related research. Spatial analyses in road safety involve the examination of geographical patterns of road safety KPIs to identify high-risk hotspots or understand the factors contributing to crashes. However, the use of different areal unit levels in spatial road safety studies influences the observed spatial effects. In other words, the boundaries we choose for our analysis can shape the road safety outcomes (Ziakopoulos & Yannis, 2020).

Indeed, a common challenge in spatial analysis is to determine the appropriate size and scale of spatial units for analysis. This choice is far from trivial, since it has a direct impact on the analysis outcomes. Experience suggests that increasing the granularity, using a finer spatial resolution, can lead to a weaker correlation between variables aggregated at the output level. Moreover, finer spatial resolutions introduce spatial autocorrelation issues violating the assumption of independence often required by statistical and machine learning models (Loo & Anderson, 2015).

Literature often treats the selection of the spatial baseline as a straightforward or simplistic choice. Although various scales have been examined (Kumar et al., 2025) and research has combined multiple scales (Cai et al., 2019; Pervaz et al., 2023) exploring multi-scale models assuming that the study are is defined as a hierarchy of partitions (Zhai et al., 2025), telematics data have not been incorporated in the studies. There remains a lack of rationale and investigation into whether and how interactions between telematics and spatial structure exist.

When dealing with telematics data, the most common approach is to link them per segment to perform a segments-level analysis, where segments are either uniformly defined (Hu et al., 2020; X. Li et al., 2021; Nie et al., 2025; Paleti et al., 2017; S. Shen et al., 2024; Wang et al., 2025) or correspond to actual segments within the road network, as seen in several works (Gabaire et al., 2025; Nikolaou et al., 2025; Stipancic et al., 2018; Ziakopoulos, 2021).

Studies filtered the network before adopting a segment-based scale, focusing on highway segments (Xie et al., 2019) or specifically on freeway segments (Paleti et al., 2017; Zhang & Abdel-Aty, 2022) incorporating features from both upstream and downstream segments, as they have been shown to play a significant role in predicting crash potential on freeways (Abdel-Aty et al., 2008).

Overall, the large-scale data collection, carried out using OBDs and smartphones, enables to cover a wide spatial area. As a result, researchers often conduct analyses at network level, focusing on segments and intersections. For instance, in the present research, by using telematics data collected from a smartphone app, the study area was selected to include over 100,000 roads and over 60,000 intersections or endpoints of the road newtork.

Beyond intersections, point-based spatial scales have been used in various nuanced ways within road safety research with telematics data. For instance, studies have focused on specific points of interest such as Railway Level Crossings (Lee et al., 2025) or bus stops (P. Li et al., 2021). In both cases researchers involved a buffer-based methodology, building a defined radius around each point, to extract features from the surrounding area by aggregating telematics observations.

This work was made possible by the extensive coverage provided by in-vehicle telematics data, allowing for analysis of highly specific spatial entities. Previous research has utilized a grid-based approach to leverage telematics data at a country-wide scale (Ryder et al., 2019). In-vehicle telematics data associated with the spatial entities mentioned have been employed in supervised learning to model road safety KPIs such as crash frequency and crash occurrence. There have been efforts both to model harsh driving events as dependent variables and risk indices based on telematics data (Ziakopoulos, 2024).

Additionally, in-vehicle data have been used to derive risk indicators for descriptive analysis, using unsupervised learning methods such as PDFs, boxplots, or scatter plots to capture spatial patterns (Hou et al., 2020; Xiang et al., 2024). This descriptive approach is often used as a preliminary step prior to supervised modeling. Nevertheless, comparisons across spatial scales that integrate telematics data within multiscale models remain largely unexplored, highlighting the need for further research in this area.

2.5 Overview of Generative Advisory Models for road safety analyses

2.5.1 Building Information Modelling (BIM)

While Building Information Modelling (BIM) is still an emerging field, its applications have not yet been fully extended into infrastructure design. For example, Industry Foundation Classes (IFC), one of the most widely used BIM frameworks, have not yet been applied to the road infrastructure domain, although various organisations are working towards this goal (buildingSMART International). On the one hand, the use of advisory AI models for road safety in the BIM context remains largely unexplored. On the other hand, studies investigating the use of generative AI models in BIM more broadly exist, and their methodologies could be adapted to road safety with only modest adjustments.

Among these studies, a press release by the Korea Institute of Civil Engineering and Building Technology (2023) presents the closest area of interest to the focus of DC13's research. According to the release, KICT’s study first analysed data from a traffic crash big data system to determine the relationship between geometric aspects of road design and the occurrence of crashes. A classical Machine Learning (ML) model was then applied to generate optimised alternatives for a selected road segment in BIM by defining conditions and specifying certain variables. Although this solution was simple yet effective, it was limited to single road segments and lacked the ability to accommodate the more complex structures of road junctions.

While examples of integrating BIM with AI in infrastructure remain limited, there is growing interest in applying reinforcement learning (RL)–based generative AI within BIM. Much of this research, however, has focused on superstructures. For instance, Jiang et al. (2025) explored the management of collaborative models where design elements created by different stakeholders might clash. To resolve these clashes automatically, the researchers trained a Proximal Policy Optimisation (PPO) algorithm using an existing rule-based model checker, rewarding or penalising the model based on the checking results. Pan et al. (2024) employed a Deep Reinforcement Learning (DRL) approach to perform multi-objective optimisation for green building design. Their process began with identifying building performance objectives, followed by determining the related design parameters. Optimisation was controlled through a building performance simulation tool, which enabled the implementation of a DRL algorithm built on a Deep Neural Network. Zhang et al. (2024) applied DRL to automate architectural space composition. Their study framed the problem as a Markov Decision Process (MDP), where the next state of the environment depended only on the current state and the chosen action. This approach was tested on an architectural renovation problem, specifically the reconfiguration of a framed-structure shopping mall by readjusting public areas and store locations. The reward function was defined using rational performance metrics of space composition, such as lighting conditions and pedestrian flow simulation. Training was conducted in a multi-agent system, where each agent controlled the position of a room. The action space consisted of four directions, and agents were allowed to remain aware of each other's actions.

Another stream of research focuses on integrating Large Language Models (LLMs). To reduce the cognitive load of learning a software interface, Du et al. (2024) proposed an LLM-based multi-agent framework that interacts with BIM Authoring Software via Application Programming Interfaces (APIs). A rule-based model checker was included in the agentic workflow, supporting reasoning and problem-solving among agents. Inspired by the "Autocomplete" function commonly used in textual input interfaces, Du et al. (2025) introduced a command recommendation system for BIM Authoring Software. By applying comprehensive filtering and enhancement methods to real-world log data provided by Vectorworks, they trained their model to learn universal modelling patterns across different countries, disciplines, and projects.

Dynamo, a built-in graphical programming interface for programmatically creating or modifying 3D model elements, also enables writing custom blocks in Python. Ko et al. (2023) demonstrated that natural language prompts could be used to generate and iteratively refine Python scripts for BIM components, or to create geometric curves that could be lofted into architectural forms. While these experiments highlighted the potential for AI-assisted design, they remained exploratory, with limited practical impact and no training or domain adaptation involved.

Inspired by StructGAN (Liao et al., 2021), He et al. (2023) proposed Generative AIBIM, a diffusion-based framework for integrating BIM with generative AI in structural design. To build their end-to-end pipeline, the researchers first used Dynamo scripts to extract standardised 2D architectural layouts without structural elements, leaving structural components to be generated by the AI model. For model development, they adopted StructGAN's Original dataset and constructed a Modified dataset for training. At the core of their approach lay the Physics-based Conditional Diffusion Model (PCDM). Unlike conventional diffusion models that predict noise, PCDM directly predicted shear wall layouts while progressively denoising inputs. Conditioning was applied using three types of information: architectural layouts, diffusion time-step data, and simplified physical constraints (building height and seismic intensity, mapped to values of 1.0, 1.5, and 2.5 across the dataset groups). This conditioning enabled the model to adapt shear wall design to different structural requirements. Through this process, PCDM refined noisy inputs step by step into plausible, structurally consistent shear wall designs guided by both architectural context and physics-informed conditions.

The reviewed literature indicates that generative AI is rapidly transforming how designers and engineers interact with BIM. However, almost no generative AI models have been developed for BIM in the road network design domain, particularly with a focus on road safety analyses.

3 Methodological approaches of IVORY DCs

3.1 Overview of selected methodologies

The previous chapter presents a critical summary of state-of-the-art techniques used to tackle road safety problems, by assessing risk levels, analyzing risk factors, evaluating infrastructure features and conducting the respective management. The multifaceted nature of road safety developments through recent scientific and technical advances is evident.

Therefore, it is important to place the IVORY WP6 doctoral research within the broader context of these advances. In other words, within IVORY, each doctoral thesis is expressed as a case study applied within the previous thematic areas, contributing to their advancement. The following sections comprise the methodological framework of these theses, including the conceptual flowcharts of each entire process, as well as the individual steps that comprise them. It is crucial to note that the methodological frameworks are a snapshot of the moment of the present stage of each research, which can be further improved and updated as time progresses and each DC matures.

3.2 Methodological framework of DC8

Recent approaches towards road safety have been following the concept of Vision Zero (Tingvall & Haworth, 1999), where it is suggested that all stakeholders involved in building and managing road infrastructure are partially responsible for its safety and security. Therefore, the overall ambiance should provide a safe agreement between all its components – road infrastructure, roadside attributes and traffic regulation. However, those items do not exist in a fixed statement throughout time, since there are deteriorations in the infrastructure and social, economic and political changes that might influence the urban/rural distribution and how external agents interact with the road space, creating the need to recurrent assessment of road conditions and overall safety.

Historically, safety assessment has mainly relied on crash data (Nikolaou et al., 2023). However, there are three main downsides of using historic crash data as the main input for safety assessment: i) it is a reactive approach, i.e., attacking a problem after its occurrence; ii) crash data is often unreliable, especially in low and middle income countries; and iii) treating high crash occurrence locations (also called “blackspots” or “hotspots”) can result in the dislocation of these crashes within the road network instead of diminishing their occurrence altogether (Mahmud et al., 2016). Therefore, alongside the development of globalized data-based tools and computation algorithms, alternate methods for road safety research have become more advocated, such as the use of surrogate safety measures and proactive infrastructure assessment (Tarko, 2018).

Surrogate safety measures are non-crash occurrences that are correlated with crash occurrence, either by sharing the same causal events or by having proven statistical close relationship with crash occurrence. A type of data that can indicate the risk of a crash occurrence is telematics data (Boylan et al., 2024), referring to geo-positioning and monitoring of road users over time, usually by assessing mobile phone or in-vehicle sensors. This type of data enables driver behavior analysis and harsh events detection, such as harsh braking or acceleration, which can be associated with the need to perform evasive maneuvers. Although telematics data itself doesn’t give the researcher insights on the traffic environment at the moment a harsh event occurs, it can be an indicative of infrastructure flaws when harsh events are frequent at a specific location (Nikolaou et al., 2025; Ziakopoulos, 2024).

Regarding infrastructure assessment, the International Road Assessment Programme – iRAP has come up with a methodology to assess road infrastructure based on video data by rating attributes on a qualitative scale and measuring road network overall safety quality, and their methodology has been applied by private and public stakeholders throughout the entire globe (IRAP, 2024). To attribute risk factors for a specific infrastructure type, they rely on well established safety research, such as safety performance function indicators. Although iRAP’s methodology is based on the manual assessment of video data, recent developments on computer vision alongside with publicization of large public datasets is pushing the expansion of their method towards an AI-based assessment.

Vision-based network element labelling suffers from a common problem in the transportation field: imbalanced data (Dilek and Dener, 2023). Some elements of the road such as intersection presence or sharp curves are by themselves not frequent and are harder to identify by computer vision algorithms because of lack of data for reliable AI training, as opposed to common objects as traffic signs. A way to work through this problem is to fuse large multiclass segmentation models with element-specific image-based extraction algorithms to optimize data collection for safety research.

Although a lot of improvement towards Vision Zero has been made in different fronts, there is still a gap in combining them. Computer vision features are not entirely integrated to surrogate safety measures in a way that can be directly used to estimate safety. Combining all aforementioned items regarding safety research, we reach the research question of DC8: "how can we use AI vision-based algorithms for an optimized proactive road safety assessment"? Thus, in pursuit of the answer, some secondary objectives shall be achieved:

- Fit proactive risk statistical models using infrastructure attributes to identify understudied road attributes in the computer vision domain.
- Develop vision-based AI algorithms for the extraction of road infrastructure attributes.
- Perform feature engineering by integrating multiple data sources to support proactive road risk mapping.
- Design and implement a network-level framework for road safety assessment.
- Conduct quality assessment and performance evaluation of AI algorithms and proactive risk models.

Accomplishing those objectives will provide for a novel pipeline for safety assessment, based on open source data, that can be used for better resource allocation prior to relying on manual human labour. Data from IVORY partners is used to build a unified framework for a case study with the proposed DC's methodology. The next subitems provide the DC's methodology flowchart and explanations of each item within the flowchart.

3.2.1 Flowchart of DC8

Figure 4 presents the Flowchart of DC8 research. An overview of the theoretical framework conducted during this PhD research has already been addressed by the previous item, leading to the research question. The next subtopics provide additional explanation of each part of the methodology flowchart. The individual parts of the flowchart connect in the following order: collected data is used in two different stages, the computer vision algorithm training and fine-tuning; and statistical modelling of harsh events using either directly collected data or feature engineered data from the computer vision model as dependent variables; road attributes indicated as statistically significant by the models and understudied in the computer vision domain are selected for further algorithm development and training. Although the methodological framework by itself presents a linear shape as presented in FIGURE X, the path undertaken during the PhD research is at some extent cyclical, in order to provide more valuable and applicable results within the transportation field.

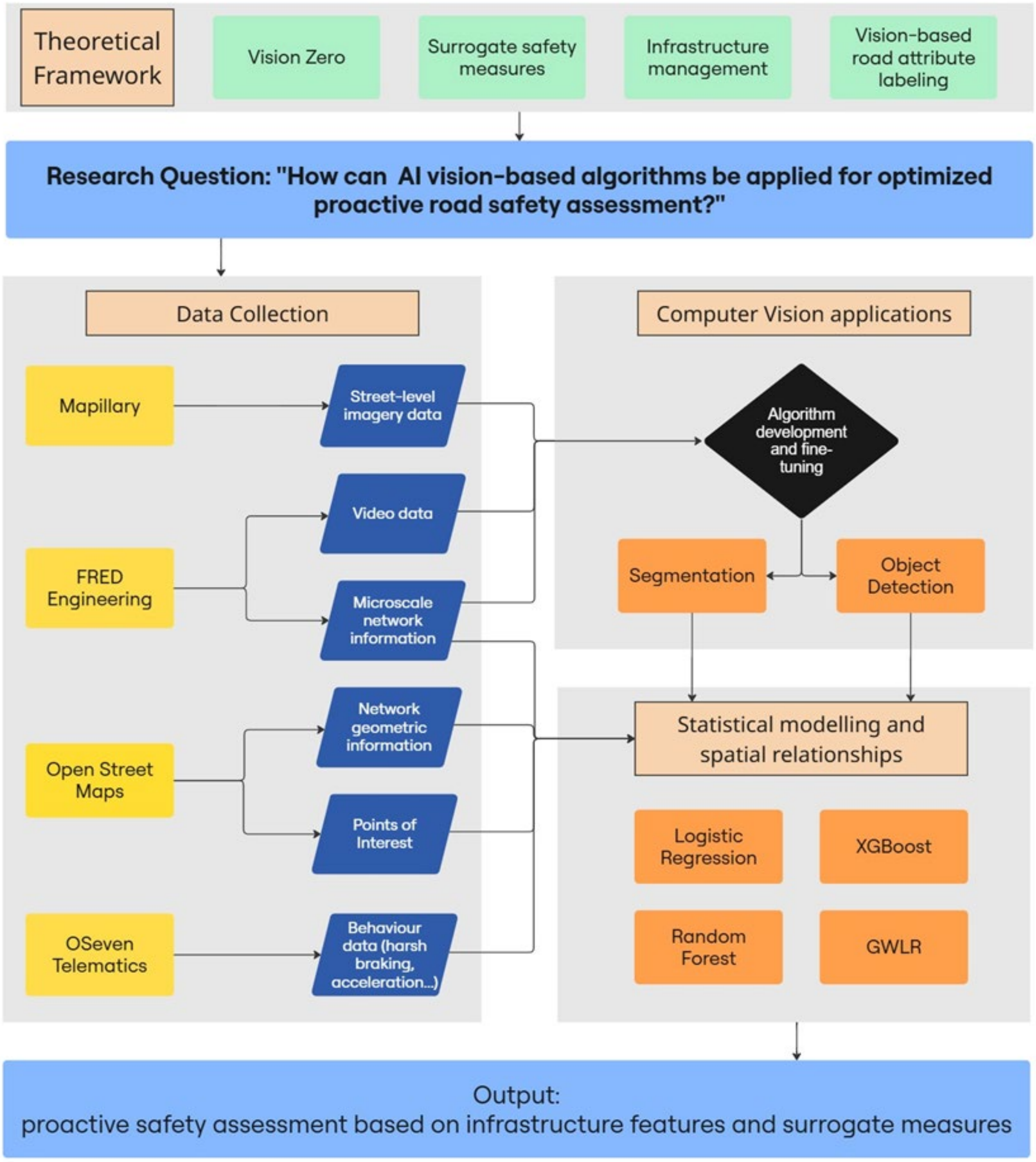

FIGURE 3: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC8

3.2.2 Data Collection

The data used for this research comes from two IVORY partners, FRED Engineering and OSeven Telematics, and two open-access crowd sourced platforms, Mapillary and Open Street Maps. The data is described in more detail on item 4. A summary of how each dataset is used is: FRED Engineering is used for computer vision algorithm training and fine-tuning and for statistical modelling; Mapillary data is used for feature extraction using trained computer vision algorithms; OSM data is used to provide additional dependent variables and granular information over the road network for the statistical models; and OSeven data provides the dependent variable analyzed on this framework, which are harsh events flagged by telematic monitoring.

3.2.2.1 FRED Engineering data

FRED Engineering data is the starting point of the proposed framework. Not only it defines the location of initial assessments, but it also provides a combination of geocoded labeled road attributes and video data that allows for a comprehensive understanding of the network environment.

The road attributes dataset is used in the first phase of statistical modelling, to explore relationships between infrastructure data and risk-prone areas based on telematics data. It is also used to evaluate the performance of AI computer vision algorithms tested on FRED Engineering video data. Video data is used to fine-tune pretrained segmentation models and to train end to end AI-based road attribute extraction algorithms, relying mainly on the road attribute dataset but also on manual labeling of some elements for a more granulated match, since video data is captured with a 29.97 fps rate while attribute data is geocoded for 100 meters segments.

3.2.2.2 Mapillary

Mapillary data is used to broaden image-based feature extraction. It has extensive coverage across European countries and is of free use upon login, making it suitable for collecting street-level imagery. Pretrained models fine-tuned on FRED Engineering can be directly applied on Mapillary data, since the images share similar locations and resolution and angle are compatibilized with the datasets used for training. This compatibility enables transferability, and features extracted from Mapillary data through computer vision algorithms are used as input for the statistical modelling stage.

3.2.2.3 Open Street Maps

Open Street Maps (OSM) data is used to further increase topological and contextual information. Geographic coordinates of the road network area assessed are extracted from OSM for compatibilization across the different datasets. From the road network, geometry and topology information can be achieved on a granular scale, as opposed to the fixed 100m segment length used in iRAP methodology, including curvature, centerline, number of lanes, travel directions, surface type. Furthermore, OSM includes building presence information, extending land use type information one step further and allowing for complexity estimations, which are also included in the initial statistical model tests.

3.2.2.4 OSeven Telematics Data

Finally, OSeven data provides the dependent variable: harsh events, identified through the company's methodology and categorized as harsh braking or harsh acceleration on a severity scale from 1 to 3. Due to the imbalance between harsh and non-harsh events, this severity scale is not considered in the present methodology; instead, the occurrence of a harsh event is treated as a binary outcome.

3.2.3 Computer Vision applications

The field in computational sciences that extracts data, features and key values from imagery inputs is called Computer Vision. On recent years, large datasets with multiple labeled semantic segmentation categories have been made public, such as Mapillary Vistas (Neuhold et al., 2017), ADE20K (Zhou et al., 2017) and Cityscapes (Cordts et al., 2016), allowing researchers to develop powerful Artificial Intelligence algorithms. In this research, multiclass semantic segmentation pre-trained algorithms outputs are combined with computer vision algorithms trained for specific tasks, such as lane delineation type segmentation and multiclass traffic sign object detection, to measure the performance of road safety models with: i) semantic segmentation vision-based features as dependent variables, and ii) micro-scale transportation features extraction. On the next subtopics, the models used for testing the full methodological pipeline are presented. Final selection of models will be made based on computational performance and final model improvement.

3.2.3.1 PSPNet

The Pyramid Scene Parsing Network (PSPNet), developed by Zhao et al. (2017), is amongst the most cited semantic segmentation models. It presented state-of-the-art performance on the Cityscapes and ADE20K datasets by the time of the publication and achieved first place in ImageNet scene parsing challenge 2016, with over 80% accuracy with their best model after ablation.

PSPNet is composed of two phases: feature map extraction through a convolutional neural network (CNN), followed by the pyramid pooling module. For the first part, state-of-the-art model ResNet50 is used. The pyramid pooling module proposed by Zhao et al. (2017) fuses the output of four different pyramid scales, thus reducing information loss between different sub-regions, which are later on concatenated after up-sampling the low-dimension features maps.

3.2.3.2 DeepLabV3+

DeepLabV3+ is another deep neural network model using the encoder-decoder format, with an atrous separable convolution layer, and is proposed as an extension of DeeplabV3 by (Chen et al., 2018). It has also achieved over 80% accuracy on Cityscapes dataset and is cited frequently in image-based transportation research. On the DeepLabV3+ architecture, the encoder module runs atrous convolutions between layers, i.e., it introduces a dilation rate between kernel weights, which increases the receptive field without increasing the number of parameters or reducing feature map resolution, thus allowing for spatial context capture while reducing computational cost. On the decoder part, the encoder features are bilinearly up-sampled by a factor of 4 and then concatenated with the corresponding low-level features of the encoder with the same spatial resolution, and finally convolution and up-sampling are applied until the output resolution is met.

3.2.3.3 Delineation type segmentation

Even on large datasets such as ADE20K that contain small specific annotated instanced (including knob, handle, seat cushion), specific transportation features aren't labelled. To train a segmentation model to recognize different types of delineation markings between road lanes (continuous, dashed and unmarked), a subsample of TUSimple dataset was re-labeled and a LinkNet-based model with a BYOL (Bootstrap Your Own Latent) module was trained.

TUSimple is an open-access dataset with images from US highways for lane detection challenges (Yoo et al., 2020). The dataset consists of approximately 6,000 folders, each folder containing at least 20 frames of images recorded with 1280x720 resolution size, taken from the driver's perspective of the road. For each frame, there is a corresponding json file with the lane delimitation description, which can be used to build the semantic segmentation masks. To avoid data leakage, only one image from each folder was selected to construct the TUSimple subset. Then, for each mask, a graph-object from the lane pixels was created, thus allowing the separation of each lane as a single object, and a script was built to manually assign a class (continuous, dashed or unmarked) for each object. Although some trials were made to ensure best separation between lane markings, some of them had a big overlap and separating them into single objects was not possible on all occasions. Additionally, the manual identification of lane types is subject to errors. Thus, the final dataset can be considered a noisy dataset, and benefit from weak supervision.

Weak supervision consists of a training method that combines full supervision with self-supervision. In this case, a segmentation fully supervised model, LinkNet, is used, and combined with self-supervised BYOL (Grill et al., 2020). The BYOL part aims to diminish the loss when two different augmented views of the same input image are given, trying to find similarities between them, using a contrastive loss function, while the LinkNet part is trained to achieve the best segmentation output, using cross entropy as the loss function and the aforementioned multiclass masks as the labelled target.

The LinkNet architecture consists of as encoder-decoder architecture, where the results of each encoder block is added to the corresponding layer on the decoder-block, allowing to maintain spatial information that otherwise would have been lost in the downsizing initial part (Chaurasia & Culurciello, 2017). Residual blocks are also used inside the encoder blocks. In this work, ResNet50 is used as the encoder of the LinkNet model.

For the BYOL part, features are extracted from the joint encoder (ResNet50) and are processed through two convolution layers, the projector (z) and predictor (q) layers, whose inputs are then used to calculate the loss between the online (view 1, θ) image and target (view 2, ξ) image, where both views are generated by randomly augmenting the input image.

3.2.4 Statistical and Spatial Modelling

After completing the previous steps of data collection and feature engineering, the selected data is used as input to model the safety surrogate measure, in this case, harsh events flagged by OSeven methodology, and Mapillary and OSM databases are used to create a multi-source feature map with coincident geographic locations. FRED Engineering data is used to fine-tune the selected computer vision models and guarantee transferability for the selected study area on northern Italy.

The objective of this step is to identify relationships between street and infrastructure variables and harsh-events-prone areas. We predict that harsh events will have a positive correlation to urbanization and complexity indicators, such as land use and Points of Interest density rate. We also predict that micro-scale data on the road segments, such as the lane delineation type, will increase the model's capacity to classify road segments into harsh-event-prone or not. Given the binary nature of the output, regression and machine learning models established in the literature shall be tested for adequacy, including Logistic Regression, XGBoost and Random Forest. Accuracy and interpretability of the models will be considered to evaluate suitability of the models.

It is well-established that transport and infrastructure attributes have significant spatial relationships, so studying spatial autocorrelations from Moran's I can improve statistical modelling for the binary flag outcome pursued in this research. Geographically Weighted Logistic Regression (GWLR) suitability will also be evaluated, depending on the outcomes.

3.2.5 Final output and methodological novelty

The final output of the proposed methodology is a model that classifies risk-prone areas using infrastructure and networks cues from imagery and map data. The abundant use of open access data sources makes the pipeline easily reproducible, although limited by the availability of the target output. Nevertheless, calibration of the model parameters should allow for easy transferability for different target datasets, and fine-tuning of the computer vision algorithms should allow for easy transferability for different location assessments.

The main novelty of this framework is the inclusion of specific computer vision algorithms for road attribute data extraction, given that said algorithms were first-hand developed and trained with the purpose of increasing the accuracy of binary models for road safety. The combination of street-level imagery-based features with telematic data on northern Italy is also novel, as far as the researchers involved in this project know. However, since available telematics data is scarce in the area, transferability of the framework pipeline to another location with more abundance of data is under investigation.

3.3 Methodological framework of DC9

Despite significant advances in individual areas, the intersection of in-vehicle telematics data, spatial analysis, and learning-based methodologies remains unexplored. While telematics systems are continuously evolving and generating high-resolution in-vehicle data, learning-based (AI tools) models rapidly evolve to process highly complex data, yet there is limited integration among these two areas along with spatial analysis.

The methodological approach of this research is indeed based on utilizing in-vehicle telematics data collected from a smartphone application developed by (OSeven, 2025), that records driver behaviour using smartphone hardware sensors. These data are then integrated onto a spatial network, together with geometry features from OpenStreetMap (OSM) (OpenStreetMap), a free, editable global map created by volunteers and released under an open-content license. OSM allows the extraction of a graph along with node and edge features stored in two separate datasets, with the considered nodes being “true” edge endpoints (i.e., intersections or dead-ends) (Boeing, 2024).

Aggregating telematics data at the edge level is straightforward, as it involves matching each data point to the nearest edge and then summarizing the data per edge. However, a novel method was introduced to enable aggregation at intersections using a buffer-based approach, enhanced by an additional constraint. Specifically, the refined method requires that telematics data points not only fall within the buffer zone created around an intersection but also lie on edges directly connected to the origin node. Preliminary results indicate that, with the novel method, the standard deviation of the telematics feature across all network nodes decreases by approximately 50%, accompanied by a reduction in spatial autocorrelation, suggesting a more promising representation.

The method does not guarantee that all edges and nodes will have associated telematics data, as coverage depends on the specific trip paths captured in the telematics dataset. Nevertheless, within the study area, there are approximately 100,000 edges and 63,000 nodes in total. Of these, 49,000 edges and 32,000 nodes contain telematics data and were used in the analyses. This indicates over 50% coverage and provides confidence in the network-level results and promising avenues for the next steps.

The research aims to partition the network in order to distinguish high-risk from low-risk nodes and edges. To achieve this, clustering techniques have been explored, which is type of unsupervised ML technique that groups objects into clusters, ensuring that elements within the same cluster are more alike than those in different clusters.

Clustering algorithms, such as K-Means (Steinley, 2006) or agglomerative hierarchical algorithms (Murtagh & Contreras, 2011) have been explored. Several indices can be used to evaluate clustering performance. In the case of K-Means, these they help determine the optimal number of clusters by running the algorithm across different values of K and selecting the one with the best performance. Among these, the Silhouette Score (Shahapure &

Nicholas, 2020) is one of the most commonly used, for instance. It provides information about whether the clusters are well-separated or overlapping.

While using raw features to partition a road network incorporating telematics is a first step in identifying different areas, this approach lacks information about the network's topology, about the relationship between the network entities and how they interact between each other is missing.

The AI field was boosted by advancements in artificial neural network architectures, leading the scientific community to develop Graph Neural Networks, which extend neural network methods to graph-structured data (Scarselli et al., 2009). A GNN model aims to encode the underlying graph structured data using the topological relationships among the nodes, instead of squashes the graph data into vectors. There has been an evolution of GNN which led the academic world to models such as Graph Convolution Networks (GCN) which involves the convolution operation (Kipf & Welling, 2017) or Graph Attention Networks (GAT) introduced by (Veličković et al., 2018).

This research has employed GNN models trained in a self-supervised way providing vectorial node representations, called embeddings, that incorporate neighbouring context and even edge features using an attention mechanism. Contrastive loss functions have been explored in order to defining one inspired by prior work in self-supervised learning such as the ones from (Chen et al., 2020; Oord et al., 2019; X. Shen et al., 2023).

Once the GNN model is trained, it produces node embeddings that can be used as input to a clustering algorithm, enabling the discovery of an improved partition of spatial entities, either nodes or edges. In the case of edge-based analysis, a dual-graph approach can be applied to represent edges as nodes, allowing the same clustering strategy to be used. Using embeddings for clustering leads to a loss of interpretability because embeddings are abstract vector representations. To gain meaningful insights from the identified clusters, the cluster labels can be mapped back to the original raw feature dataset. By averaging the raw features within each cluster, we can better understand the underlying meaning of each cluster found by using embeddings.

Preliminary results indicate that approximately 4% of the nodes exhibit hazardous behaviour, suggesting that interventions could be targeted to a small subset of nodes, enabling more efficient allocation of resources focusing on the small subset of risky intersections. The exploration of the variables may lead to conclusions related to how targeting those intersections.

3.3.1 Flowchart of DC9

The flowchart of Figure 4 outlines the overall methodological approach to be undertaken within the topic of exploring AI for road safety monitoring and crash prediction from micro- to macro levels. Its individual steps are outlined and analysed in the following passages.

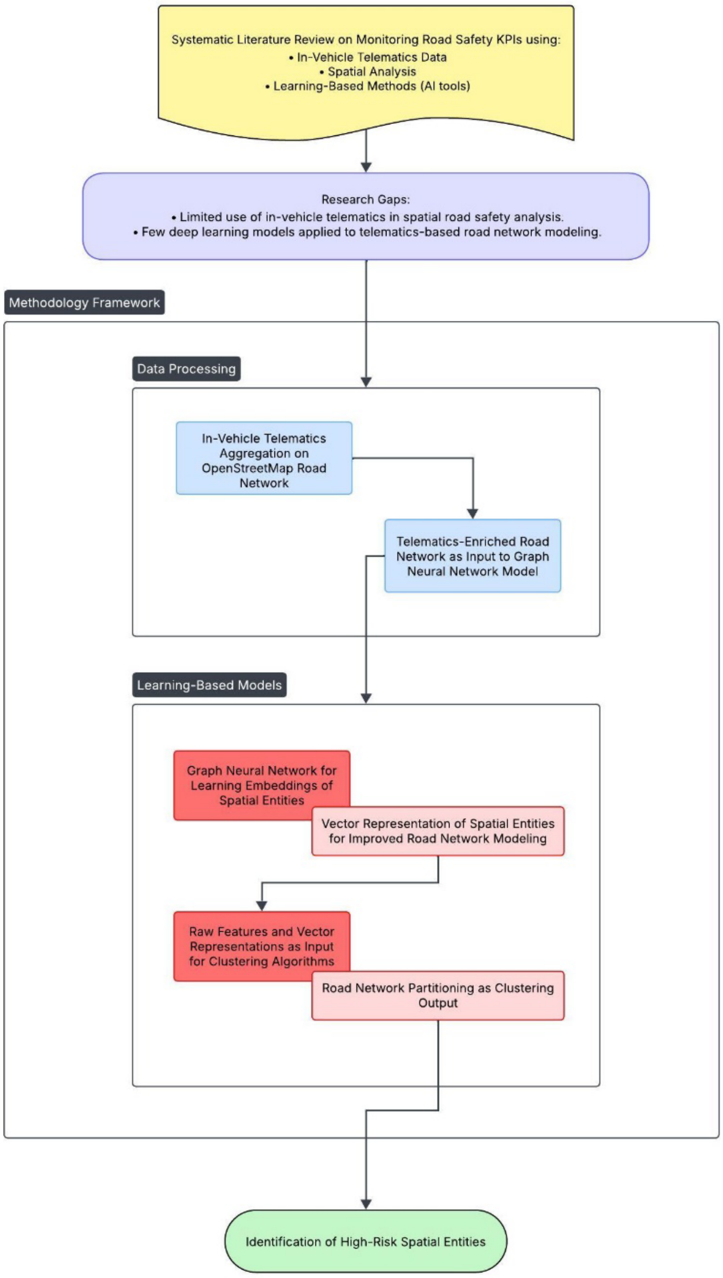

FIGURE 4: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC9

3.3.2 Systematic Literature Review and Research Gaps

A review of the literature at the intersection of in-vehicle telematics data, spatial analysis, and learning-based methods reveals a limited use of in-vehicle telematics data within spatial road safety analysis, with most research relying on tree-based machine learning algorithms and a recent trend towards deep learning methods. Furthermore, comparisons of telematics-informed spatial scales remain largely unexplored. More advanced deep learning approaches, particularly graph neural networks (GNNs), are also underutilized despite the proven transferability of deep learning models and the specific suitability of GNNs for spatial analysis, where data often take the form of graphs.

3.3.3 Data Processing

The gaps identified in the literature have motivated the exploration of graph neural networks while using telematics data integrated into a road network obtained through OSM. The data were pre-processed to ensure quality and then aggregated onto the network using two approaches.

In the edge-based approach, each observation was assigned to its nearest edge for aggregation. For node-based aggregation, two buffer methods were evaluated: one aggregated observations solely based on the buffer area, while the other added a constraint requiring points to lie both within the buffer and on edges intersecting at the node originating the buffer. To evaluate their effectiveness, statistical methods and AI tools were used to assess the reconstructibility of each dataset. Based on this, the second approach was selected, since its aggregated data proved more difficult to reconstruct, suggesting they contained contain more detailed, granular, and less redundant information. At this stage, the enriched graph with Telematics data can be used as input for a GNN models.

3.3.4 Learning-Based Models

This research explored the GNN to generate embeddings, which are a richer vector representation of spatial entities, enriched with information related to the topology network and neighbouring context. Various GNN models were explored, beginning with the Graph Convolutional Network (GCN) and extending to the Graph Attention Network (GAT) and its second version. These GNN models were trained in a self-supervised manner using a custom graph contrastive loss function designed to pull together nearby nodes, following to the homophily principle, while pushing apart distant nodes.

The GNN output served as input for an unsupervised ML model, a clustering algorithm. K-Means and Hierarchical Agglomerative Clustering algorithms were applied to partition the network and identify distinct areas with different characteristics, as shown in Figure 4. This approach was compared against a simpler baseline where raw features were directly fed into the clustering algorithms, demonstrating the superior performance of the GNN-based method.

3.3.5 High-and Low-risk Spatial Entities (Closing)

The use of GNN models to generate inputs for Clustering algorithms appears to be an innovative and promising approach. Further level of information is combined for each spatial entity when generating GNN outputs. Clustering performances are improved and the gap between interpretability and performance is bridged by mapping back the cluster labels to the original dataset. Additionally, leveraging a dual graph approach enables a shift from a node-centric to an edge-centric perspective, while preserving the same methodology used for node analysis.

3.4 Methodological framework of DC10

A prominent issue for safe navigation systems is accurately measuring or defining safety. The underlying objective of this research is to leverage data-driven learning to gain deeper insights into road safety and its interpretation by users. This section outlines the analytical steps and systematic design of this research. This study is structured into four phases, with each phase encompassing multiple tasks, as illustrated in the flowchart shown in Figure 5.

3.4.1 Flowchart of DC10

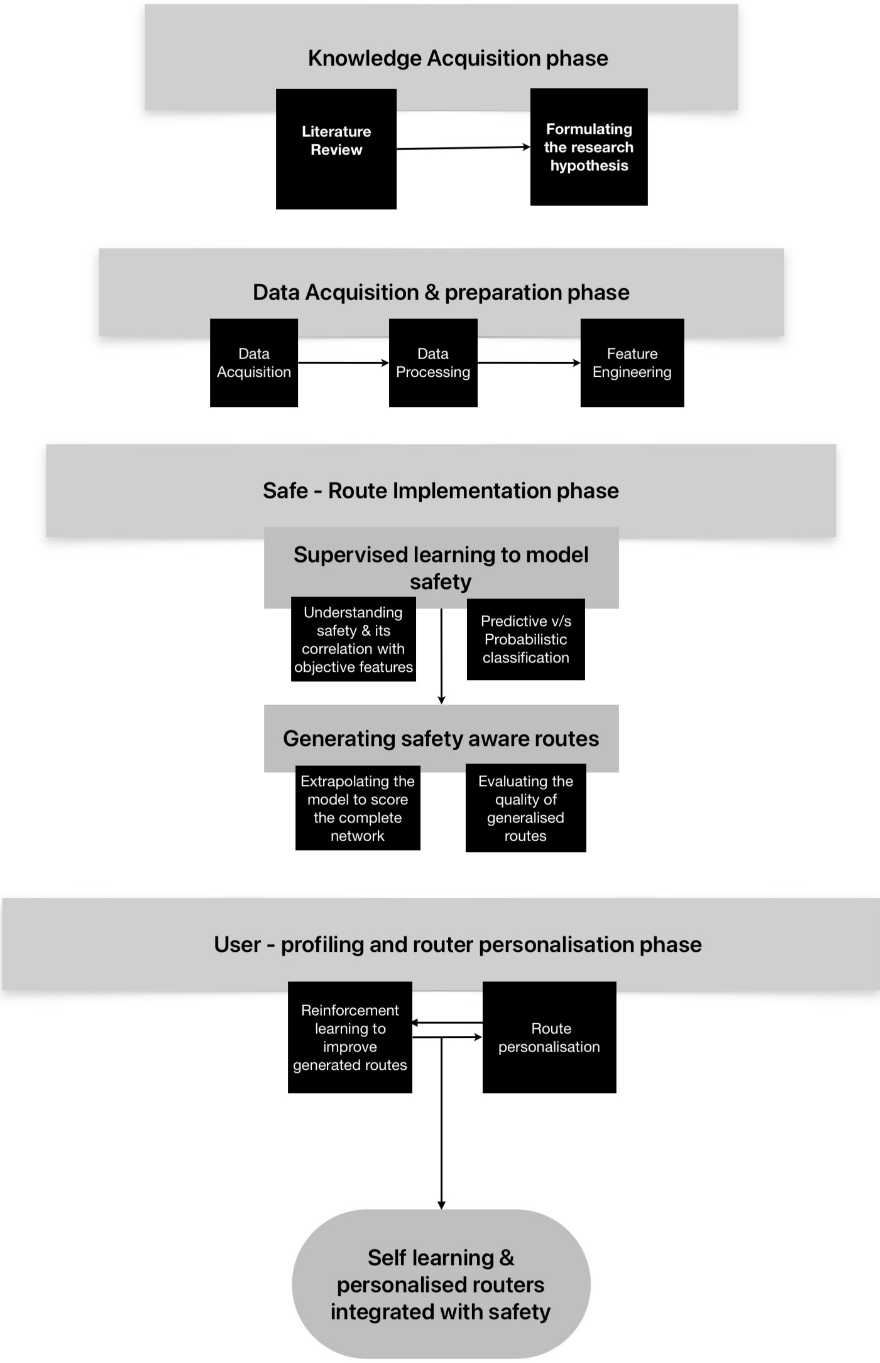

FIGURE 5: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC10

3.4.2 Knowledge Acquisition Phase

The objective of this phase is to study available literature in the domain to identify gaps and formulate the primary research questions. A critical task here is to define key focus areas for the research and to determine mainstream trends and methodologies used in the domain. This establishes a clear foundation and shapes the trajectory for the doctoral study. The field of safe routing is still evolving; hence, an ongoing literature review will be conducted throughout the research to stay aligned with the latest developments in the field.

3.4.2.1 Research Questions

The core questions this doctoral research aims to answer are summarized below:

1. To what extent does subjective safety differ from objective safety?
2. Is it possible to develop a safe-route network for home-to-school trips using existing data?
3. What can we learn about user behaviour by monitoring the safe route network, and how can this information enhance the safe routers?
4. How can machine learning techniques aid in designing a self-learning router which suggest routes most appreciated by the users?

3.4.3 Data Acquisition and Preparation Phase

This is the most important phase of any data-driven modelling. The objective here is to identify the key attributes affecting the safety of a link. In this phase, the focus will be on filtering data acquired from different sources. The emphasis would be on determining a well-defined and well-labelled feature set for safety modelling.

3.4.3.1 Data Acquisition

The first step is securing the relevant data. Considering that an essential aspect of this study is understanding the intricacies of both the subjective and objective nature of safety, it is vital to collect data reflecting both users' perceptions and quantitative safety indicators.

The Route2School dataset is the primary data representing subjective safety. The dataset, as described in the later sections, consists of a geocoded road network annotated with safety tags based on users' feedback. To enhance this dataset, infrastructural attributes must be sourced from OpenStreetMap. To incorporate objective safety, the Traffic Accidents dataset provided by the Belgian statistical office (Statbel) is obtained. The dataset contains information about the road accidents across Belgium.

3.4.3.2 Data Preprocessing

The next task is to prepare the gathered datasets for training the models. The expected outcome of this exercise is to derive structured, machine-understandable datasets from the semi-structured raw data. Another significant aspect is to conduct a comprehensive analysis of the datasets to derive key insights such as class imbalance ratios and the extent of missing data.

Standard preprocessing techniques, including data imputation, normalization, one-hot encoding, and label encoding, are systematically applied to handle missing data and produce efficient datasets for model training. Given the multiple sources used to procure the datasets, in addition to the usual data preprocessing steps, several geographical preprocessing strategies, like map-matching and spatial joins, must be employed to generate coherent data.

3.4.3.3 Feature Engineering

Selecting the features most relevant to the model is a vital step in supervised learning algorithms. The dataset resulting from the previous steps must be filtered to extract the most appropriate features for a safety model. Standard feature selection approaches like mutual information classifiers, chi-square tests, and model-based feature selection techniques must be employed to capture simple correlations between the attributes and safety. Encoder-based feature selection techniques are employed to rigorously capture complex interdependencies among features and to identify those most relevant to the target variable, safety.

3.4.4 Safe-Router Implementation Phase

This phase concentrates on evaluating several machine-learning models based on the dataset generated in the previous step. The primary aim of this experiment is to design a baseline safety model and evaluate its performance in a routing environment.

3.4.4.1 Supervised Learning to Model Safety

A fundamental component of this study is understanding an average road user's interpretation of safety and applying this knowledge to generate safe routes. Therefore, a supervised learning approach will be adopted to model safety as a function of the feature set resulting from the previous phase. As the safety model forms the core of the safe router, it is imperative to design a model that is generalizable and transferable across contexts.

This step involves training and comparing the performance of simple linear models, tree-based models, and more advanced deep learning models in predicting the safety of a route. A critical challenge to be addressed here is the significant class imbalance in the dataset. Oversampling and under sampling techniques will be explored to account for this and produce balanced results. To evaluate the performance of the models, techniques such as spatial cross-validation, confusion matrices, and precision and F1 scores will be employed.

Another essential task at this phase is determining the appropriate classification strategy between a predictive and probabilistic approach to account for the inherent uncertainty in accurately predicting the safety of a road segment. While predictive classification can achieve high accuracy scores, a probabilistic classification would allow quantifying the degree of prediction uncertainty and hence enable more informed decision-making.

3.4.4.2 Generating Safety Aware Routes

After developing a baseline model, the logical next step is to integrate the model with a router. This requires extrapolating the model to a broader region and generating a safe road network. It is also crucial to devise an appropriate strategy to add the safety component to routing algorithms. The emphasis at this stage would be to utilize the predictions from the model to generate a weighted safety score for each link. These weighted scores are vital for integration with the routing systems, as they can be used to modify the cost function of the routers. Since routes are ultimately the most relevant to the users, this approach would aid in evaluating the model performance at the route level instead of at the link level.

The subsequent step will be evaluating the quality of routes generated by the safe router and comparing these to the routes generated by a traditional router. A systematic framework would be adopted to sample origin-destination pairs and simulate real-time routing environments to precisely analyse the route quality and model performance.

3.4.5 User-Profiling and Router Personalization Phase

The final phase of the research is developing a preliminary application that allows users to interact with the router smoothly and gather feedback to improve the algorithms. A crucial task here is observing the user preferences to create a self-learning router powered by reinforcement learning to suggest routes most appreciated by the users.

3.4.5.1 Reinforcement Learning

The core idea of reinforcement learning is to continuously monitor the outcomes to make better decisions. The next task of this research is applying the reinforcement learning approach to develop a self-learning router, resulting in an evolving algorithm. This involves critically analysing the user-router interactions, ensuring data privacy, and utilizing the insights to enhance the router suggestions. Since the prime focus of this research is subjective safety, reinforcement learning could be a powerful algorithm to capture the perpetual bias of a user and integrate the subjective nature of safety in navigation systems.

3.4.5.2 Route Personalization

A substantial emphasis of this doctoral study is to suggest routes most relevant to the users. Therefore, the next task is to develop an experimental framework to train the model to stress different attributes while generating safe routes. The central objective of this phase is to develop multiple models stressing different attributes correlated with safety and enable users to adopt the router most suitable for them.

The current literature highlights several intriguing developments in the field of road safety and safe routing. However, it lacks a robust routing system incorporating safety, especially subjective safety. Throughout this research, the goal will be to explore the potential of machine learning algorithms to integrate safety in the routing systems. The research aims to investigate the similarities and/or differences between perceived safety and objective safety with a focus on enhancing existing datasets and designing more personalized routing solutions.

3.5 Methodological framework of DC11

This chapter presents the methodology for developing a multi-camera AI framework for proactive road safety analysis. The workflow begins with synchronized video collection and GDPR-compliant anonymization, followed by object detection and multi-object tracking to construct 2D trajectories. Using T-Analyst, camera calibration enables these trajectories to be projected into world coordinates, supporting 3D reconstruction of road user movement. From these trajectories, surrogate safety measures such as Time-to-Collision (TTC) and Post-Encroachment Time (PET) are derived and aggregated into conflict matrices. Finally, holistic analysis and AI-driven recommendations transform these indicators into actionable insights for road safety management.

3.5.1 Video Data Collection from Multiple Synchronized Cameras

In the first phase, cameras are deployed to monitor the intersection. Cameras are positioned at different angles, so their fields of view cover all relevant zones with minimal blind spots. Each camera captures high-resolution, high-frame-rate to resolve small objects and fast motion. Crucially, all cameras must be time-synchronized so that each captured frame can be exactly matched across views. Such synchronization ensures that moving vehicles or pedestrians appear in the correct temporal sequence across all cameras. The raw video streams from each sensor are then streamed or transferred (via wired or wireless networks) to cloud, along with timestamp metadata. This multi-camera capture provides overlapping perspectives on the scene, which is essential for handling occlusions and enabling the later 3D reconstruction steps.

3.5.2 Data Anonymization and GDPR-Compliant Storage

Raw video of public roadways inevitably contains personal data (faces, license plates, etc.), so privacy protection is mandatory. In accordance with GDPR's privacy by design principle, all identifiable information is removed or anonymized before analysis. In practice, this means automatically detecting sensitive details (e.g. pedestrians' faces, vehicle plates) and applying blurring or pixelation to make the subjects unrecognizable. GDPR explicitly defines anonymization as rendering personal data in such a way that the individual is not or no longer identifiable. Other measures include masking any regions outside the roadway or isolating only vehicle features. After anonymization, the processed videos are stored securely: common practice is to encrypt the files at rest and control access via authenticated accounts or keys. Data is usually kept in a secure server. GDPR also mandates retention limits, so a defined expiration policy ensures that data is deleted when no longer needed.

3.5.3 Object Detection

Next, each video frame is processed by a convolutional neural network (CNN) detector to locate road users (cars, trucks, motorcycles, pedestrians, etc.). State-of-the-art detectors like

YOLOv11 and Faster R-CNN are commonly used. YOLOv11 is a one-stage detector: it runs a single forward pass over the image to simultaneously predict bounding boxes and class probabilities. This single shot design makes YOLOv11 extremely fast and suitable for real-time analysis. In practice, YOLOv11 runs at tens of frames per second and achieves state-of-the-art accuracy among real-time detectors. It incorporates modern architectural optimizations (efficient backbone, loss functions, etc.) to balance speed and precision. By contrast, Faster R-CNN is a two-stage approach: it first uses a Region Proposal Network (RPN) to generate candidate object regions, then classifies and refines them. Faster R-CNN typically achieves very high precision (especially for small or overlapping objects) but at lower frame rates. Both methods are trained on large, annotated datasets (e.g. COCO or traffic-specific collections) and often fine-tuned to the camera views. In implementation, one might run YOLOv11 for efficiency, or Faster R-CNN for higher accuracy in complex scenes (Ren, 2015; Redmon et al., 2016). The output of this step is a set of 2D detections per frame: each detection is a bounding box plus a class label (vehicle, pedestrian, etc.). Typical toolkits (PyTorch, TensorFlow) provide pre-trained weights for these models. Practical considerations include setting confidence thresholds (to drop spurious boxes) and possibly using model pruning or TensorRT optimization if running on embedded GPUs. The goal is robustly identifying every road user of interest in each frame, yielding a sequence of object coordinates that feed the tracker in the next stage.

3.5.4 Multi-Object Tracking and 2D Trajectory Extraction

Following detection, a multi-object tracker links the detections over time to form object trajectories in each camera. This tracking-by-detection process typically uses a Kalman-filter-based motion model combined with data association. For example, the DeepSORT algorithm enhances the classic SORT tracker by also computing a deep appearance descriptor for each detection (Wojke, 2017). In DeepSORT, a small CNN extracts an embedding vector for each bounding box, encoding the object's appearance (Wojke, 2017). The tracker then matches detections to existing tracks by minimizing a cost matrix that combines motion (predicted by Kalman filter) and appearance similarity. This reduces identity switches during occlusions. ByteTrack (Zhang, 2022) adopts a two-stage association strategy: it first links all high-confidence detections to tracks using motion (IoU overlap) and re-identification features, then secondarily associates the remaining lower-confidence detections to any unmatched tracks. By valuing even low-score detections as part of the trajectory, ByteTrack recovers objects that standard thresholding would lose. Both methods ultimately output a set of 2D trajectories per camera: each trajectory is a time-ordered list of the object's bounding-box positions. Implementation parameters include the maximum age of a track (how long it survives without new detections) and the IoU threshold for matching. Optionally, trajectory smoothing or interpolation can be applied if detections are intermittent. The result is that every object in each camera view now has an identifier and a continuous 2D path (pixel coordinates over time).

3.5.5 Camera Calibration

Camera calibration determines how to map image pixels into real-world coordinates. Intrinsic calibration solves for each camera's internal parameters: focal length, principal point, and lens distortion. Extrinsic calibration finds each camera's rotation and translation relative to a common world frame. After 2D trajectories are constructed, camera calibration is performed using T-Analyst, which enables both intrinsic and extrinsic parameters to be derived directly from the roadway environment. Intrinsic calibration corrects for lens distortion and defines the internal geometry of the camera, achieved by referencing features of known dimensions such as lane widths or stop lines. Extrinsic calibration establishes the camera's spatial position and orientation relative to the real-world coordinate system, which is carried out by interactively selecting fixed points in the video (e.g., crosswalk corners, lane intersections, or curb edges) and assigning them their corresponding real-world coordinates obtained from maps or on-site measurements. Through this point-matching process, T-Analyst computes the transformations required to project every image-plane detection into world coordinates. For multi-camera setups, calibration is repeated for each view but referenced to the same roadway geometry, ensuring all trajectories are aligned within a single global frame. The result is a set of calibrated video streams where each tracked object can be consistently mapped into real-world space, forming the basis for 3D bounding box reconstruction and multi-view trajectory generation.

3.5.6 3D Bounding Box Reconstruction Across Views

With multiple calibrated cameras covering the same intersection, 3D information is recovered using multi-view geometry. The basic technique is point triangulation: when an object is detected in at least two camera views at the same time, its 2D image positions are back-projected into 3D rays using the known intrinsic and extrinsic calibration parameters, and their intersection provides the object's 3D location. In practice, stable reference points within the bounding boxes—such as the bottom-centre or corners of vehicles and pedestrians—are selected for triangulation, yielding accurate ground-plane positions. Repeating this process frame by frame produces a time-stamped 3D trajectory for each object. Since the cameras are fixed and their fields of view overlap, the reconstruction is consistent and robust, and in addition to triangulating single points, it is also possible to generate 3D bounding cuboids that represent each road user's real-world dimensions and orientation. These 3D boxes capture position, scale, and heading in world coordinates, enabling precise measurement of distances, speeds, and interactions between road users. By reconstructing trajectories in 3D, the system overcomes occlusions that might affect single-camera views and provides the accurate spatial data required for conflict detection and safety analysis.

3.5.7 Multi-View Matching and 3D Trajectory Generation

After reconstructing 3D positions in each view, the pipeline must merge observations of the same object across cameras. This step is multi-camera data association: we match the 3D tracks from different camera views that correspond to the same real-world entity. In overlapping views, this can be done via spatial matching: if two tracks from different cameras share a consistent 3D location and time (given calibration), they are merged. In non-overlapping networks, matching relies on temporal continuity and appearance re-identification. In either case, the goal is to form a single multi-view track for each object. For example, a person entering from camera A's view and then appearing in camera B should receive one unified ID. This cross-camera linking is often handled by a cross-camera association module. In essence, one constructs multi-camera tracks that describe each object's movement across the entire sensor network. Once matching is done, we fuse the 3D point estimates: for each time step, we combine the corresponding 3D detections from each camera (often simply taking their average or trusting the most accurate view). The result of this stage is a unified 3D trajectory for each object. These trajectories integrate data from all cameras, providing a global view of each object's motion. Trajectories may be smoothed or interpolated to fill any brief gaps. In the end, every road user has a continuous 3D trajectory from entry to exit of the scene.

3.5.8 Surrogate Safety Metric Computation (TTC, PET, etc.)

With all objects tracked in 3D, the system computes surrogate safety metrics to quantify potential conflicts. Common measures include Time-to-Collision (TTC) and Post-Encroachment Time (PET). TTC is defined as the time remaining until a potential collision if both objects continue with their current velocity vectors. In practice, for any pair of objects on (nearly) intersecting paths, the algorithm projects their constant-speed motion forward and solves for the time at which their distance would reach zero. If they would collide, this time is TTC. A very small TTC (near zero) indicates an imminent collision risk. PET measures how close in time two trajectories come to crossing the same point. For example, if Object A leaves a conflict location at time t_A and Object B arrives at that location at time t_B, then PET = |t_B – t_A|. A small PET (e.g. under 1 second) means one object nearly followed another through the same spot, indicating a near-miss. Lower TTC/PET values indicate more severe conflicts. This data forms the basis for assessing safety: for instance, TTC is a widely used proxy for collision risk. In implementation, one loops over time and object pairs, calculates these metrics, and flags those below chosen thresholds as conflicts. The outcome is a time-indexed record of near-collision events with quantitative scores for each.

3.5.9 Conflict Matrices Generation

The analysis constructs conflict matrices which a tabular summary where rows and columns index categories of road users or movement directions. For example, rows might be vehicle-to-vehicle, vehicle-to-pedestrian, pedestrian-to-bicycle, etc. Each cell of the matrix accumulates statistics (count or average severity) of conflicts between those categories. As each surrogate conflict is detected, the system increments the appropriate cell: e.g. a near-miss between a car and a cyclist adds one to the (Car,Cyclist) cell. The conflict matrix thus highlights which interactions are most frequent or most dangerous. Visualizing this matrix helps identify hotspots: for instance, it may show that car–pedestrian interactions have many conflicts while bicycle–pedestrian is rare.

3.5.10 Holistic Safety Pattern Analysis and AI-Based Recommendations

In the final phase, the collected data and conflict summaries are analysed for higher-level patterns and used to generate safety recommendations. This involves statistical and machine-learning techniques to mine the trajectories and conflict records. Analysts might cluster conflict events by time of day, object speeds, or road layout to discover systemic issues (e.g. most conflicts occur at rush hour). Machine learning models can also be trained on the surrogate metrics to predict crash likelihood under different conditions. The outcome of this phase is a set of actionable recommendations. These may include engineering fixes (e.g. installing new signage, adjusting signal timing, redesigning lanes), policy changes (speed limit reductions), or even in-vehicle alerts. Some systems may also provide real-time warnings: if an AI model detects a rapidly developing conflict, it could trigger an instant alert or adaptive traffic control to mitigate the risk. In summary, holistic pattern analysis turns the raw conflict data into insight: it identifies recurring dangerous scenarios and uses AI-driven models to suggest targeted interventions, thereby closing the loop from observation to improvement.

3.5.11 Flowchart of DC11

The flowchart of DC11, comprising six phases, is shown on Figure 6 and outlined in the following passages.

3.5.11.1 Research Foundation

The first phase establishes the foundation of the research by integrating three critical activities: literature review, formulation of research questions, and initial data collection. The literature review surveys prior work in proactive traffic safety analysis, video-based trajectory extraction, surrogate safety indicators, and AI-driven risk assessment. This synthesis identifies the methodological gaps and technical challenges that the proposed research aims to address. Based on this review, research questions are formulated to guide the methodological design, focusing on how multi-camera AI-based systems can improve the detection and quantification of traffic conflicts. Parallel to this conceptual groundwork, the phase also incorporates the collection of raw multi-camera traffic video data at selected intersections. This early dataset not only grounds the research questions in practical reality but also provides the baseline for testing and validating subsequent methodological steps. Together, these activities ensure that the research is problem-driven, evidence-based, and oriented toward real-world urban safety challenges.

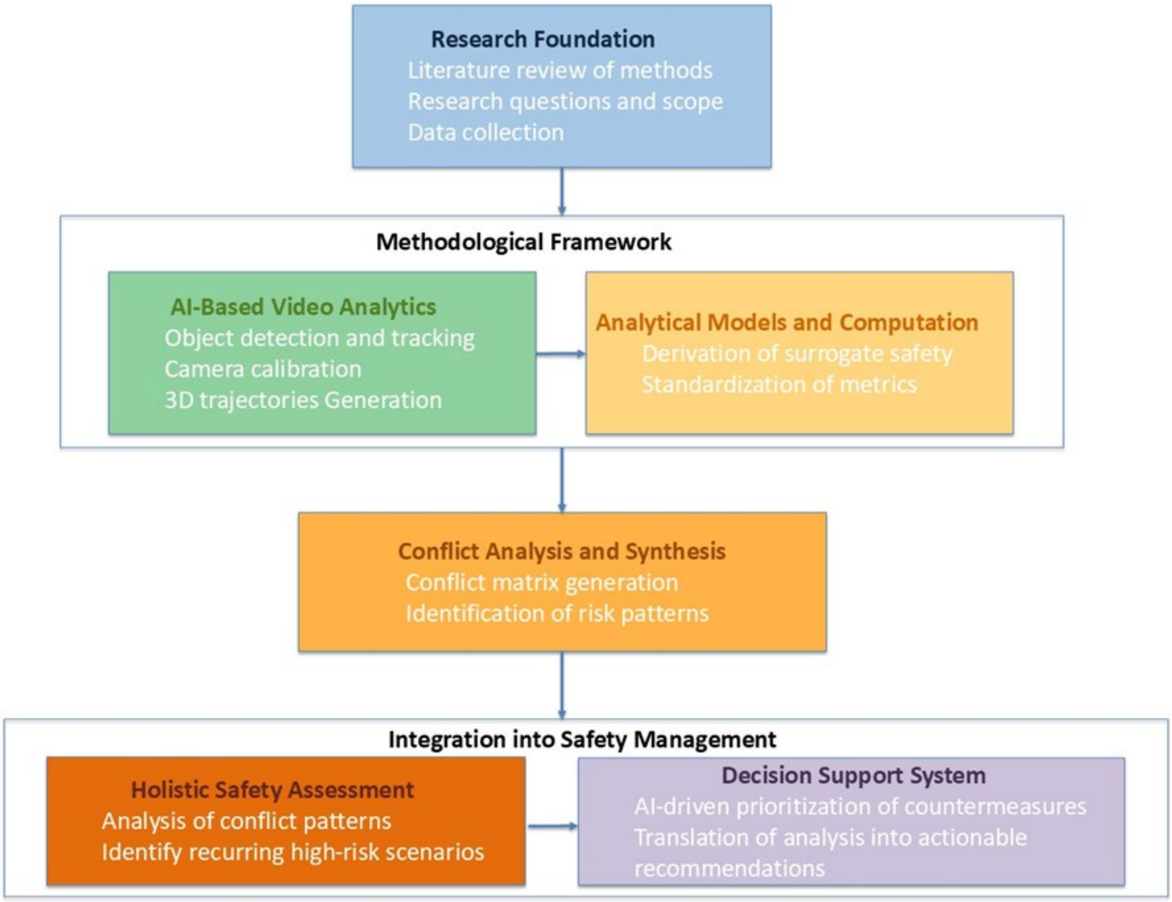

FIGURE 6: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC11

3.5.11.2 Methodological Framework

This phase outlines the overall framework through which the research questions are operationalized into a structured pipeline. The methodological framework acts as the bridge between conceptual research objectives and technical implementation. It defines the role of artificial intelligence in processing video data, the integration of computer vision techniques with calibration tools, and the design of analytical models for safety evaluation. The framework emphasizes modularity, ensuring that each technical component—detection, tracking, calibration, 3D reconstruction, and conflict analysis—fits into a coherent workflow. It also establishes evaluation criteria for accuracy, efficiency, and scalability, ensuring that the methodology is robust enough to handle large-scale urban deployments. It is divided into two parts.

3.5.11.2.1 AI-Based Video Analytics

At the core of the methodological framework lies the AI-based video analytics pipeline. This component is responsible for converting raw anonymized video data into structured spatiotemporal information. Using state-of-the-art object detection (e.g., YOLO, Faster R-CNN) and tracking algorithms (ByteTrack, DeepSORT), the system identifies and follows road users across frames to construct continuous trajectories. Camera calibration through TAnalyst then projects these trajectories into real-world coordinates, allowing precise measurement of distances, speeds, and paths. The output of this stage is a set of 3D trajectories for all road users, forming the empirical backbone for further safety analysis. This phase emphasizes scalability, GDPR compliance, and the use of robust multi-view fusion to minimize occlusion errors, thereby ensuring data quality and methodological soundness.

3.5.11.2.2 Analytical Models and Computation

The trajectories produced in the previous stage are processed through analytical models designed to quantify traffic safety risks. Surrogate safety indicators such as Time-to-Collision (TTC), Post-Encroachment Time (PET), and other metrics are computed from the 3D trajectories, enabling objective quantification of near-misses and conflict severity. Advanced statistical and computational approaches are used to ensure that the derived metrics are both precise and interpretable. This stage is crucial for transforming raw movement data into standardized indicators that can be compared across sites, times, and user groups. It operationalizes the theoretical concept of traffic conflicts into measurable, reproducible data, providing the analytical basis for subsequent aggregation and pattern detection.

3.5.11.3 Conflict Analysis and Synthesis

In this phase, the surrogate safety indicators are synthesized into structured conflict matrices. These matrices serve as an aggregation tool, categorizing conflicts by road user type (vehicle–vehicle, vehicle–pedestrian, cyclist–vehicle, etc.), movement patterns (crossing, merging, following), and severity levels. The conflict matrix provides a holistic snapshot of the safety performance at an intersection or roadway segment, allowing analysts to identify which interactions are most problematic and which areas or times of day are associated with elevated risk. By structuring individual conflict events into a systematic representation, the conflict matrix transforms continuous trajectory data into a management-friendly format. This stage marks the transition from micro-level trajectory analysis to meso-level systemic safety assessment.

3.5.11.4 Integration into Safety Management

This phase ensures that the analytical outputs are embedded into actionable safety practices. It is divided into two complementary sub-components:

- Holistic Safety Assessment
  The conflict matrices are analysed longitudinally and contextually to identify systemic safety issues. Trends are extracted across time, traffic volumes, and user categories to uncover recurring patterns of high-risk behaviour. AI-based clustering and anomaly detection techniques can be applied to highlight emerging risks. This stage provides urban planners and policymakers with an evidence-based understanding of where, when, and why unsafe interactions occur, thus supporting preventive safety strategies.

- Decision Support System
  The final stage translates analytical insights into decision-level recommendations. Using AI-driven scenario modelling and predictive analysis, the system evaluates the likely impacts of interventions such as adjusting traffic signal timings, redesigning intersections, or lowering speed limits. It provides stakeholders with a "what-if" analysis capability, helping them prioritize interventions with the greatest safety benefits. This decision support layer ensures that the methodology not only diagnoses safety problems but also provides actionable, data-driven solutions to improve road safety outcomes.

3.5.12 Methodological Contributions

The proposed methodology contributes a novel, end-to-end framework for proactive road safety analysis. Its originality lies in the integration of multi-camera AI-based trajectory extraction, surrogate safety computation, and conflict matrix synthesis into a single coherent pipeline. By embedding GDPR-compliant practices and leveraging advanced analytical models, it ensures that the research is both ethically responsible and technically robust. The multi-layered structure—from research foundation to decision support—demonstrates how raw video data can be transformed into actionable knowledge for safety management. This approach advances the state of the art by bridging the gap between technical computer vision outputs and the real-world needs of policymakers and practitioners, offering a scalable and transferable methodology for urban traffic safety improvement.

3.6 Methodological framework of DC12

This chapter outlines the methodology used to develop a computer vision model for the automated, scalable, and reliable extraction of road infrastructure attributes from heterogeneous data sources, in line with the standards of the International Road Assessment Programme (iRAP) (International Road Assesment Programme (iRAP) 2019) and the Network Road Assessment Program (NAWA) (National Technical University of Athens 2023). While iRAP defined more than 50 road safety attributes, this research focuses on a targeted subset listed in Table 1, encompassing geometric features (such as lane width, curvature, and grade), semantic elements (such as road surface type, signage, and pedestrian facilities), and contextual aspects (such as land use, area type, and roadside severity) of the road environment.

The approach addresses critical limitations in traditional road safety auditing methods, which are predominantly manual and thus incur high costs, extended update cycles, and inconsistencies due to human subjectivity. It also overcomes shortcomings in existing automated techniques, which often rely on a single data source (e.g., monocular video) and lack the required metric precision, network-wide contextual awareness, and robustness for comprehensive road safety assessments under diverse conditions. The proposed methodology consists of the steps outlined in the following.

3.6.1 Data Acquisition

The first step in developing the automated solution for road attributes identification using AI from geo-referenced road survey videos is data collection. We utilize a road-safety corpus that was collected along 2300 km of public highways in Bosnia and Herzegovina for our experiments and model development. We called this dataset 'iRAP-BH'. Although the iRAP Star Rating Score evaluates road segments in 100-meter increments, the corpus is annotated at a finer granularity of 10-meter segments to achieve more accurate estimates through averaging. The corpus comprises approximately 230,000 10-meter segments, each annotated by human iRAP coders with 52 distinct road safety attributes. These segments usually have about 30 to 40 video frames; however, this varies based on the road and traffic conditions at the time of data acquisition. All videos are captured in 2704x2028 RGB format at a rate of 25 frames per second. To construct the dataset for our recognition experiments, we utilize geographical data from the iRAP coding database. Each iRAP record pertains to a 10-meter road segment and includes the values for the 52 iRAP attributes, along with the GPS coordinates of the segment's two endpoints. We employ these GPS coordinates to identify the corresponding video frames for each endpoint. This involves locating the two nearest GPS references in the associated georeferenced road video, followed by interpolating the GPS position using the estimated vehicle speed. Ultimately, all frames between these endpoints are assigned to the respective segment. The work emphasizes classifying the road attributes as detailed in Table 1.

3.6.2 Data Pre-processing

The subsequent phase after data collection involves data pre-processing and cleaning. In our data pre-processing, we avoided employing horizontal flipping, as it would interfere with the detection of road attributes tailored to right-hand traffic scenarios (e.g., Paved shoulder - passenger-side, Land use- passenger side). Additionally, we also did not implement random cropping, since it eliminates visual information from the peripheral areas of the input images, which is essential for recognizing roadside attributes like Street lighting. All extracted frames are resized to a consistent dimension (e.g., 224x224 pixels or 336x336) to align with the input specifications of the pre-trained ViT model. Pixel values are normalized to a standard range to ensure stable training. Figure 7 shows sample data from the iRAP-BH dataset.

FIGURE 7: ROAD SEGMENTS EXTRACTED FROM THE IRAP-BH DATASET

3.6.3 Model Selection and Customization

In this section, we have explained our approach for selecting and adapting a hybrid (vision transformer + CNN) model for effective road attribute classification from video frames and addressing challenges such as data imbalance and temporal dependencies. We begin by leveraging a pre-trained backbone to extract generalized features, which form the foundation for our customized architecture. Subsequently, we introduce novel modules to enhance temporal processing and mitigate class imbalances, ensuring robust performance on a real-world road survey dataset.

3.6.3.1 Pre-trained Model Selection

We have employed a pre-trained hybrid model, MaxViT (Tu et al. 2022), as a backbone for our road attributes classification task; the model is pre-trained on the large-scale ImageNet-1k dataset. This pre-training on diverse natural images allows the model to capture rich, generalized visual features Click or tap here to enter text., which are subsequently fine-tuned for the downstream task of road attribute classification. This pre-training on diverse natural images enables the model to capture rich, generalized visual features, which are then fine-tuned for the road attribute classification task using fewer data and computational resources. The pre-trained weights serve as a robust initialization, speeding up convergence and enhancing performance by leveraging fundamental visual patterns. The backbone handles sequences of frames. For each 10-meter road segment, multiple frames are processed to generate feature maps from various stages, which are adaptively average-pooled and concatenated into a single, rich feature vector per frame. The resulting vectors are aggregated into a temporal sequence, enabling the incorporation of sequential information across the segment. Moreover, to adapt the model for multi-label road attributes classification, where a single image may contain multiple road attributes simultaneously, the original MaxViT classification head (a single linear layer designed for ImageNet tasks) is replaced with a custom head. The custom head is a fully connected layer with an output dimension matching the total number of distinct road attributes across all categories.

The pre-trained MaxViT backbone is incorporated into our custom modules designed to tackle the complexities of multi-attribute road classification, encompassing temporal information processing and enhancement of minority classes using a dynamic loss function. These custom modules are defined in the section Custom Modules .

3.6.3.2 Custom Modules

We have proposed novel custom modules for the pre-trained model to increase its accuracy on a highly imbalanced dataset. Pre-trained backbone processes multiple frames per input video segment to extract features. These features are flattened and concatenated across different backbone stages to form a single feature vector for each frame. After this, our custom temporal module processes these frame-level features. It uses a 1D convolution to model local temporal dependencies and a sigmoid-activated linear layer to detect "rare" or salient frames based on their features. A GRU (Gated Recurrent Unit) further enhances these temporal features, with the final output being a weighted average, emphasizing the contributions of the enhanced features and the original features explicitly weighted by their "rarity" scores, providing a concise summary representation for the entire input sequence.

3.6.4 Training Procedures

3.6.4.1 Dataset Splitting

Unlike traditional machine learning methods that generally utilize random dataset partitioni ng, we employ a systematic strategy to segment the annotated dataset into training, validati on, and test subsets for our experiments. Following the methodology outlined in (Kacan et al. 2024), the dataset is split into 214,073 training segments, 5,813 validation segments, and 6,563 testing segments (with the test set comprising approximately 15% of the total). This data distribution technique guarantees that all road segments associated with the same road section are allocated to the same subset, thereby preventing data leakage and enhancing the training of sequential and multi-frame models. Each road segment is represented by its core frame, scaled to dimensions of 224 × 224 pixels.

To address class imbalance, especially common with infrequent attributes, class weights are computed for each attribute before training. These weights, derived from the inverse frequency of classes within the training set, are incorporated into the loss function, thereby diminishing bias toward more dominant classes and promoting equitable learning across all attributes.

3.6.4.2 Fine-Tuning Model

During the fine-tuning process, we have trained the backbone on the iRAP-BH dataset with a frozen backbone initially for some epochs. Once the backbone learned some patterns from our dataset, we gradually unfroze the layers from the last with each epoch. We didn't unfreeze the complete backbone at once to avoid instability during the training. This gradual unfreezing starts with the final layers to adapt task-specific features initially, subsequently advancing toward the initial layers to refine more generalized representations. The AdamW optimizer (Loshchilov and Hutter 2017) is employed for its demonstrated efficacy in deep neural networks, particularly due to its decoupled weight decay mechanism that facilitates stable optimization. A cosine annealing learning rate scheduler with an initial warm-up phase is employed, wherein the learning rate incrementally increases to alleviate early training instability before smoothly decreasing according to a cosine function, thus facilitating accelerated convergence and enhanced generalization. Key hyperparameters are carefully selected and optimized through empirical experimentation or grid search, including batch size (e.g., 10, supplemented by gradient accumulation for larger effective batches), differentiated learning rates, number of epochs (e.g., 30), weight decay for regularization against overfitting (e.g., 1e-3), dropout rate to enhance model robustness (e.g., 0.7), scheduler T_max for annealing cycles (e.g., 10), and focal loss gamma to prioritize difficult examples (e.g., 2.0), to mitigate inherent data imbalances.

3.6.5 Handling Data Imbalance

During the analysis of our iRAP-BH dataset, we found that the dataset faces a severe imbalance problem. In these contexts, standard cross-entropy is predominantly influenced by majority classes, whereas straightforward solutions such as inverse-frequency weighting or over-/under-sampling may excessively penalize informative majority samples or lead to overfitting and information loss; even widely used alternatives like Focal Loss provide support but still depend on static class weights that do not adjust as learning advances (Cao et al. 2019; Johnson and Khoshgoftaar 2019; Lin et al. 2017).

To address this, we use a dynamic weighted focal loss across our multi-class road-attribute classification models. The approach keeps the primary idea of prioritizing hard examples while adjusting per-class weights throughout training according to recent class-specific recall. This results in a gradual down-weighting of well-learned majority classes and an increased focus on under-learned or minority classes, stabilized with smoothing techniques to mitigate early-epoch fluctuations. This approach addresses between-class imbalance through adaptive class weighting and within-class difficulty via targeted focusing, consistent with evidence that adaptive re-weighting and scheduling strategies enhance minority-class recall and macro-F1 on long-tailed data. This results in enhanced performance on infrequent road characteristics without compromising overall precision.

3.6.6 Model Evaluation & Results

We evaluate our multi-class road-attribute classifiers (in a multi-label setting) with metrics tailored to imbalanced data and mutually exclusive subclasses per attribute. For inference, the model produces a probability for each subclass of every attribute; we convert these to discrete predictions by selecting the highest-probability subclass per attribute (argmax). For attributes encoded as presence/absence indicators, this operates as a standard multi-class decision among mutually exclusive states. Performance is reported per attribute and per subclass using precision (share of predicted positives that are correct), recall/sensitivity (share of actual positives that are correctly found), F1-score (harmonic mean of precision and recall, emphasizing robustness under imbalance), and accuracy (share of all correct decisions). True Positives, True Negatives, False Positives, and False Negatives are obtained from per-attribute confusion matrices. To avoid majority-class bias, we use macro F1 as the primary aggregate by averaging F1 across subclasses so each class contributes equally. All metrics are computed on the held-out, unseen test set to assess generalization, and we summarize results by reporting per-class scores, macro F1, and overall accuracy for a balanced view of performance on rare and frequent road attributes alike.

3.6.7 Data Fusion

After finalizing and validating the camera-only multi-class classification AI models on the road-survey video dataset, we will extend the pipeline by incorporating a camera-LiDAR data fusion module. As discussed in section Infrastructure attribute extraction (infrastructure data fusion) single modality of data is not enough for precise classification of road attributes. Fusing camera-Lidar data will help in extracting road features that need exact geometric measurements, such as distances to roadside objects and widths of lanes, shoulders, or medians, where camera data alone is not enough. To achieve this, we will collect LiDAR point clouds and road survey videos/images from the same roads. After this, these heterogeneous data sources will be harmonized for calibration, timestamp alignment, and spatial registration. camera data will provide detailed visual appearance information for attribute recognition, while LiDAR will supply accurate 3D structural data, ensuring reliable measurements even under difficult conditions like low texture, glare, or partial occlusions. This approach aims to enhance recognition of underrepresented classes while providing high-accuracy geometric estimates, enabling advanced multi-modal road-attribute analysis in future project stages.

3.6.8 Flowchart of DC12

The flowchart of Figure 8 depicts the research approach for DC12.

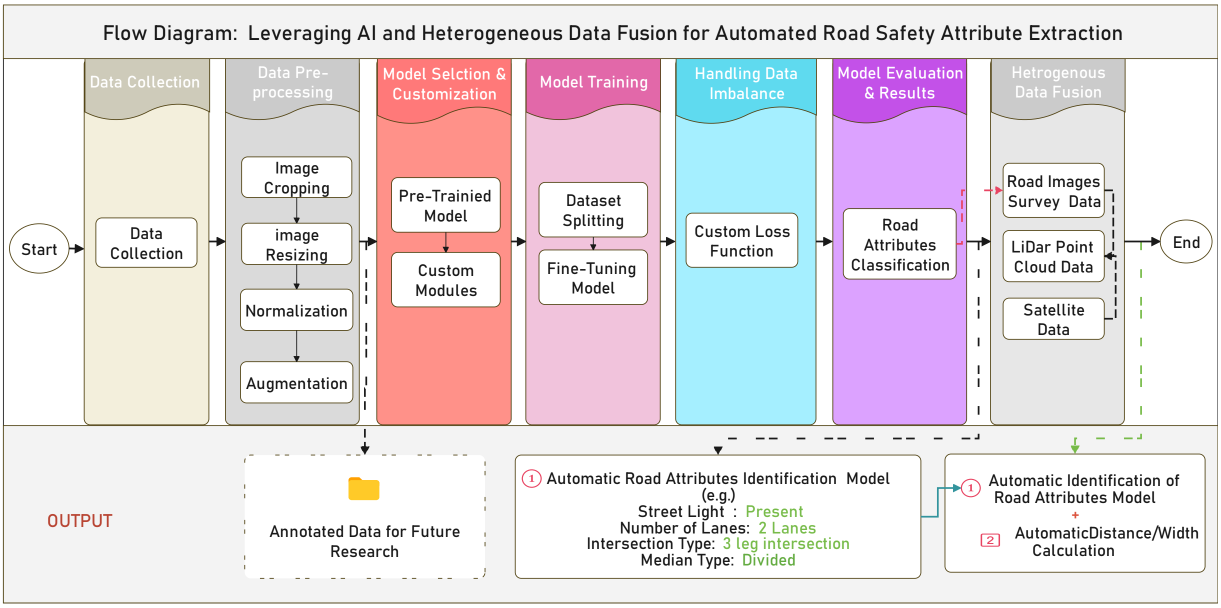

FIGURE 8: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC12

3.6.9 Research Novelty

The novelty of our approach lies in the development of a scalable, AI-driven framework for automated extraction of iRAP-aligned road infrastructure attributes from georeferenced road survey videos. Our approach addresses the key limitations present in traditional manual coding (e.g., time-consuming, non-consistent road coding) of road survey videos and single-modality automated systems through a multi-step pipeline by integrating advanced computer vision techniques. Our methodology utilises a pre-trained hybrid model as a backbone and further fine-tunes it on the iRAP-BH dataset. We have also developed custom temporal modules to improve temporal performance and loss functions to address the significant class imbalances present in the road survey dataset. Our approach improves the detection of minority classes without overfitting or compromising the accuracy of majority classes. Moreover, integration of camera-LiDAR data fusion will also improve the geometric accuracy of measurements that need to be precise, like distances between objects and lane widths.

Finally, the proposed AI based road safety attributes detection system will not only reduce road assessment costs but also speeds up the process, enabling real-time or frequent road safety assessments. With AI based road attributes identification system, cities can achieve higher survey frequency, more consistent and accurate data, and earlier identification of road hazards, leading to proactive safety measures. These benefits not only improve road safety by targeting high-risk areas more effectively but also reduce long-term infrastructure costs.

3.7 Methodological framework of DC13

3.7.1 Outline

This chapter introduces the methodology for developing a Building Information Modelling (BIM)-based framework enhanced with Artificial Intelligence (AI) to improve road safety. The workflow begins with data-driven identification of hazardous points, followed by the creation of conceptual BIM models. A generative adversarial approach based on Graph Attention Networks (GAT) is then explored for design optimisation, supported by a microsimulation environment for extracting Surrogate Safety Measures (SSMs). Finally, proposed road designs are evaluated through First-Person Perspective (FPP) simulations, providing complementary human-centred insights. Each component is presented in sequence, with attention to both its technical implementation and its role within the overall framework. The flowchart of Figure 9 depicts the research approach for DC13.

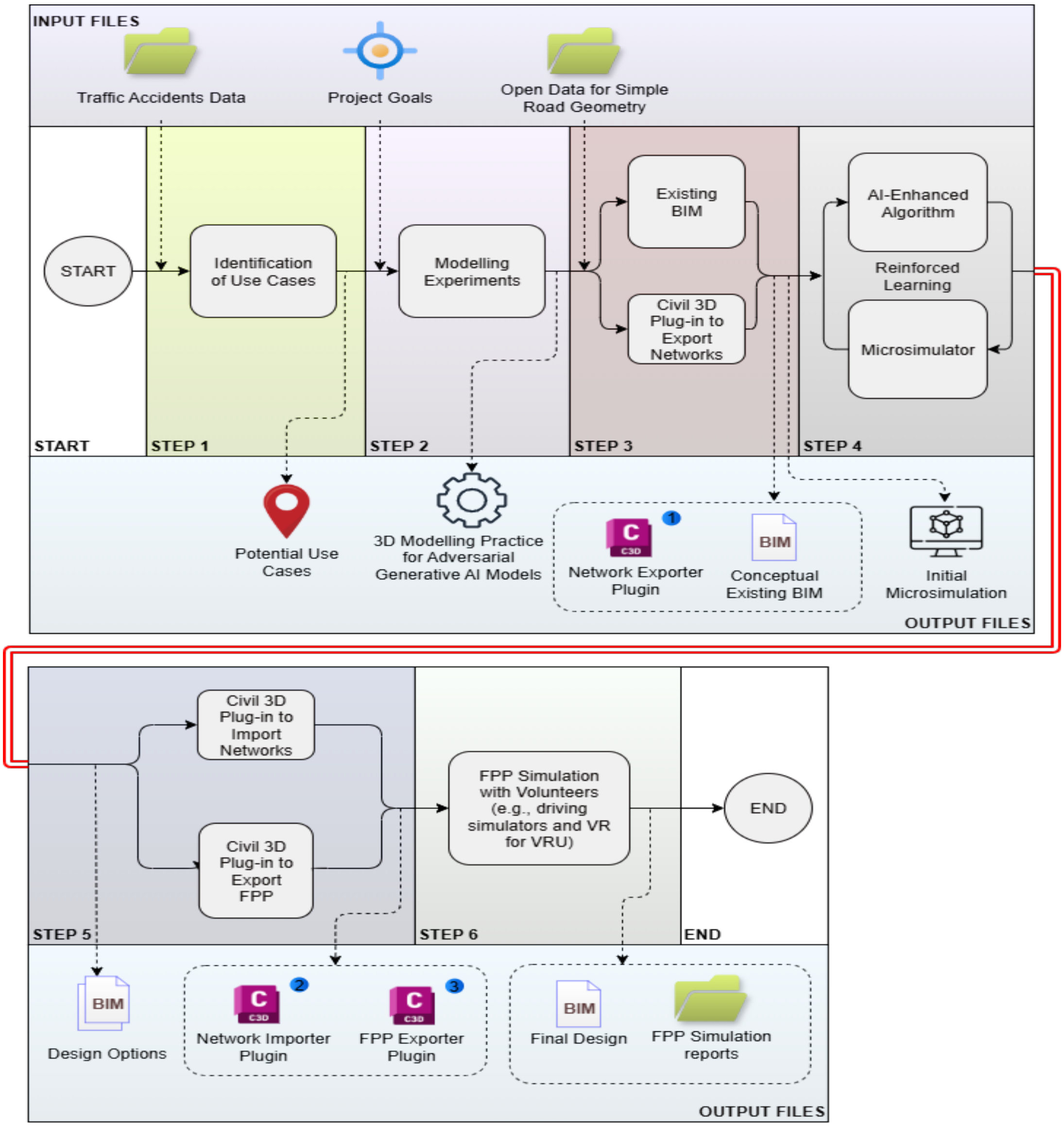

FIGURE 9: FLOWCHART OF THE METHODOLOGICAL APPROACH OF DC13

3.7.2 Identification of Points

The first step identifies hazardous locations in the road network by combining official accident records from STATBEL (2022) with the hotspots dataset of Agentschap Digitaal Vlaanderen (2024) (see step 1 in the flowchart). To focus on persistent risks, only points repeatedly classified as dangerous over the past seven years were retained, resulting in six problematic sites: three dominated by crashes among non-vulnerable road users and three involving vulnerable users.

To manage and explore this data, a custom application called HotMap was developed. Built with Flask and Plotly, it provides an interactive, platform-independent map that displays hazardous points and crash types. The tool also supports saving selections and exporting outputs, including OpenStreetMap (OSM) (OpenStreetMap Contributors) files and Excel datasets, ensuring both reproducibility and smooth integration into later workflow stages.

3.7.3 Conceptual BIM Models:

In developing conceptual BIM models, OSM is selected as the initial data source because it provides an open and flexible geospatial dataset that integrates seamlessly into infrastructure design workflows without licensing restrictions. The second step focuses on representing real-world scenarios within BIM environments. Experiments are conducted to establish best practices for interoperability, ensuring that BIM and the other environments used in this study share a common structure (see Step 2 in the flowchart). For example, intersections are modelled as separable corridor objects in BIM, which can then be projected into Eclipse's Simulation of Urban MObility (SUMO) (Krajzewicz et al., 2002) as junctions and into the GAT-based algorithm as custom-shaped nodes. These experiments form the methodological foundation for linking design, simulation, and AI.

For practical implementation, OSM files are first converted into CAD format with QGIS to ensure compatibility with Autodesk Civil 3D (Autodesk). The imported data is processed into 3D corridors, creating a three-dimensional representation of the road network. Auxiliary nodes are introduced to capture variations such as changes in alignment or lane count. These nodes connect successive road segments without being treated as intersections, thereby preserving geometric accuracy while maintaining topological coherence.

Finally, the conceptual BIM model is processed through the Civil 3D API using a custom plugin written in C# (.NET 8.0) (see Step 3 in the flowchart). The plugin exports the network as a structured JSON file of nodes and edges. This export provides a common framework for both SUMO and the GAT algorithm, ensuring a consistent and comparable interpretation of the road layout. The JSON file thus serves as the primary reference for constructing the virtual environment in which GAT-based generative AI models are trained with SSMs derived from SUMO.

3.7.4 Initial GAT Development (over a Simplified Task)

To manage the complexity of the GAT component, development began with a simplified task: adjusting preliminary road networks so they no longer intersected with buildings. Although this did not represent the real-world objective of safety optimisation, it provided a controlled proof of concept for applying GATs within a reinforcement learning framework. It was designed to share a key characteristic with the target problem: the absence of a derivable loss function, which makes reinforcement learning essential. For this reason, the initial GAT model is not part of the complete framework and does not appear in the flowchart.The testing framework was structured around three pillars: environment creation, training design, and model architecture.

3.7.4.1 Environment Creation:

Six progressively complex environments were developed. The simplest relied on base Python objects (e.g., dictionaries and lists), while the most advanced (Environment 6) employed custom classes inspired by the Gymnasium framework (Towers et al., 2024). From Environment 4 onwards, the concept of ghost edges was introduced: temporary connections between road segments and buildings that appeared whenever intersections occurred. This mechanism produced the most substantial performance improvement by providing a clear signal of conflicts. Environment 6 further incorporated advanced randomisation features, enabling controlled generation of varied network instances and thereby broadening the test space.

3.7.4.2 Training Design:

Seven training loops were implemented, ranging from a baseline loop that processed the simplest environment without gradient updates to more advanced loops that iteratively modified the environment based on model predictions. This incremental approach allowed systematic validation of the model’s learning capability across training styles, environments, and architectures.

In the most advanced setup, the reward function was defined as the difference in the state score before and after a move, where the state score quantified the total length of road segments intersecting with buildings. At each iteration, the node with the highest predicted potential was selected for processing. For that node, the four possible moves (up, down, left, right) were executed, and the resulting rewards were fed back into the model to compute gradients. The environment was then updated according to the move that produced the highest real reward. This design enabled the model to progressively align its predictions with actual outcomes, effectively linking node-level decision-making to global network optimisation.

3.7.4.3 Model Architecture:

Five different architectures were developed using the GATv2Conv module of PyTorch Geometric (Fey et al., 2025). Early versions primarily tested feasibility in the simplest environments, while later versions were designed to predict the contribution of individual actions to network optimisation. In the final architecture, the model was tasked with predicting the expected reward of each node for a one-unit move in each of the four directions. This structure allowed the model to operate at the node level while integrating directional decision-making into the learning process.

The experiments demonstrated that the GAT-based model successfully learned to adjust road layouts to eliminate road–building conflicts. Although only a simplified proof of concept, these results confirm the feasibility of using reinforcement learning for network optimisation. The next phase of the study (see step 4 in the flowchart) will extend this approach to Civil 3D–based environments, where the agent's task will evolve from avoiding buildings to improving SSMs derived from the microsimulator.

3.7.5 GAT model with a Microsimulation

In this step, an AI module will build on the earlier proof-of-concept GAT experiments by extending the learning task from simplified building-avoidance scenarios to safety-oriented optimisation. SUMO will play a central role in this integration, not only as a microsimulator but also as the reward mechanism for training. Each candidate network configuration generated by the AI will be evaluated in SUMO, where vehicle behaviour is simulated and safety indicators are extracted. These indicators will be used to calculate rewards that guide the agent's learning process. SUMO is open-source and can be launched directly as a Python package, which enables a high level of programmatic interaction with the simulation environment. It also offers a built-in SSM device (German Aerospace Center (DLR) & SUMO Contributors, 2024b), allowing the extraction of indicators such as time-to-collision.

In addition to the built-in features, SUMO provides a Python-based Traffic Control Interface (TraCI), which was used to monitor vehicle movement around specified locations. TraCI enables the development of custom safety measures tailored to this study, extending beyond the standard outputs of the SSM device (German Aerospace Center (DLR) & SUMO Contributors, 2024a). This flexibility makes SUMO particularly suitable for integrating with AI workflows, where programmatic access and real-time feedback loops are essential.

3.7.6 Optimised Design

Step 5 requires the development of a second plugin, the Network Importer, which brings the design alternatives generated by the AI module back into Civil 3D. This ensures interoperability and allows designers to review and refine AI-generated options within a familiar BIM environment. The imported designs must then be further detailed for the next stage, where they will be tested in FPP simulations. To enable this transition, a third plugin will be developed. Together with Step 6, enabling such a transition ensures that the outcomes of AI optimisation extend beyond abstract computational experiments and can be directly experienced and assessed by the road users who are likely to be non-experts.

3.7.7 First Person Perspective (FPP) Simulation

The final stage introduces human participants into the evaluation loop. FPP simulations, conducted through driving simulators or Virtual Reality (VR) environments tailored for vulnerable road users, are used to capture qualitative insights into the proposed design options. These assessments are combined with quantitative outputs from SUMO to evaluate both safety and usability. The final design is produced by consolidating these findings.

3.8 Methodological Synthesis

The six previously outlined approaches have clearly discernible methodological elements towards advancing proactive, data-driven road safety assessment and optimisation, combining AI, computer vision, spatial analysis, simulation, and human-centred evaluation. Collectively, the studies progress from extracting and integrating heterogeneous road and driving data (e.g., telematics, video, imagery, and infrastructure attributes) to developing predictive models that identify risk-prone areas and safety hotspots, leveraging both statistical and machine learning approaches (DC8, DC9). A related research task incorporates subjective perceptions and personalized preferences to generate safety-aware routing systems, bridging the gap between objective indicators and user experience through reinforcement learning and supervised/probabilistic modelling (DC10).

The methodology also advances towards with the deployment of multi-camera AI frameworks for precise 3D trajectory reconstruction, enabling surrogate safety metric computation (TTC, PET) and conflict matrix synthesis for systemic risk assessment (DC11). Large-scale AI-based infrastructure assessment pipelines are also developed using hybrid Vision Transformer & CNN models, temporal feature aggregation, and camera & LiDAR fusion to enhance attribute recognition, geometric precision, and minority-class detection across extensive road networks (DC12). These approaches demonstrate a consistent emphasis on multi-source data integration, rigorous preprocessing, spatiotemporal awareness, and performance evaluation under complex, real-world conditions. Finally, the methodology is extended into a design and human-centred optimization phase by integrating BIM, microsimulation, and generative AI via Graph Attention Networks (DC13). Reinforcement learning guided by Surrogate Safety Measures iteratively improves road network layouts, which are then evaluated in First-Person Perspective simulations to incorporate human feedback, particularly from vulnerable users.

Across all Doctoral Research projects of the WP6 of IVORY, the combined methodological synthesis highlights a multi-layered, iterative workflow: (i) systematic data acquisition and preprocessing, (ii) AI-driven modelling for safety assessment or optimization, (iii) multi-modal integration for enhanced precision, and (iv) human- and policy-centred evaluation to ensure actionable, user-informed insights for proactive road safety management.

4 Data acquisition

4.1 Data source description overview

The present section aims to provide an overview of the various data sources that will be exploited during the planning and execution of the IVORY WP6 doctoral theses. For each data source, a description of its origin and content is initially provided. This is followed by an exploration of the overall availability and replicability of the data source, as well as remarks on its respective sensitivity and overall compliance with GDPR regulations. The data sources that are shared between DCs are given a common introduction first, and their particular usage for each DC is explained subsequently.

4.2 OSeven Telematics Data (DC9 & DC8)

4.2.1 Description/Origin

OSeven provides datasets containing anonymized trip data with a variety of metrics. The data are collected with a minimum frequency of 1 Hz using smartphone hardware sensors and data fusion algorithms provided by Android (Google) and iOS (Apple). A variety of APIs is used to read sensor data and temporarily store them to the smartphone's database before transmitting them to the central backend database (Papadimitriou et al., 2019).

The data collected are highly disaggregated in terms of space and time. Once stored in the backend cloud server, they are converted into meaningful metrics, using signal processing, ML algorithms, Data fusion and Big Data algorithms. This process leverages inputs from the accelerometer as well as additional sensors (e.g., orientation, magnetometer, GPS, gyroscope), without relying on specific thresholds. These state-of-the-art technologies and procedures are protected Intellectual Property of OSeven and cannot be disclosed. However, the reliability of these methods has been extensively evaluated against literature data, OBD data, on-road experiments by certified experts and experiments on driving simulators (Kontaxi et al., 2021).

As previously mentioned, the determination of the metrics within the OSeven dataset is not based on specific thresholds. Some indicative rules are provided in (Kontaxi et al., 2021). For instance:

- a harsh accelerations event can be flagged when:
    - b) speed increase from 31 km/h to 40 km/h within one second
    - b) longitudinal acceleration 0.28 g
- a harsh braking event can be flagged when:
    - b) speed decrease from 51 km/h to 40 km/h within one second
    - b) longitudinal deceleration 0.30 g.

Nevertheless, the four indicative cases cannot be considered as thresholds, as the determination of the events is based on the coevaluation of several time series.

4.2.2 Availability/Replicability

The data can be shared for research purposes under a formal agreement, therefore exclusively available to the involved DCs. They cannot be published or shared with third parties without the explicit and direct consent of OSeven. External access is not permitted. This controlled access ensures the security of sensitive data and full compliance with applicable data protection regulations.

4.2.3 Sensitivity/GDPR

The data pre-processing steps are confidential and thus not disclosed. However, the steps are performed in compliance with standing Greek and European personal data protection legislation (GDPR) (Kontaxi et al., 2021).

Moreover, the data are stored in the OSeven backend system using advanced encryption and data security techniques. The APIs utilize support user authentication and encryption to prevent unauthorized data access (Papadimitriou et al., 2019).

4.2.4 DC8 Dataset

The provided dataset consists of anonymized trip data from around 200 trips within the northern regions of Italy passaging through motorways SS13, SS54 or SS55 on the regions of Udine and Gorizia over the years of 2020 to 2024. It contains trip coordinates along with per-second speed data, periodic binary flags (1 or 0) denoting the presence of harsh events, speeding or mobile use and also includes the intensity of harsh events on a scale from 1 to 3.

4.2.5 DC9 Dataset

The provided dataset consists of anonymized trip data from over 13,000 drivers within a central area of the Athens metropolitan region, collected over the last four months of 2024. It contains trip coordinates along with per-second speed data, periodic binary flags (1 or 0) denoting the presence of harsh events, speeding or mobile use and it also includes the intensity of harsh events on a scale from 1 to 3.

4.3 OpenStreetMap (DC8, DC9, DC10 & DC13)

4.3.1 Description/Origin

OpenStreetMap (OSM) is a collaborative project that creates a free, editable map of the world, contributed to by volunteers globally. It was launched in 2004 has now positioned itself as the most famous example of Volunteered Geographic Information (VGI) on the Internet (Mooney

& Minghini, 2017). OSM is essentially a spatial database containing geographic data and information from all over the world. OSM data can be obtained through OSMnx Python. OSMnx simplifies the analysis of street networks by providing a range of geometric and sociodemographic features.

4.3.2 Availability/Replicability

OSM data is freely available under the Open Database License (ODbL). It is highly accessible for use, modification, and redistribution. It is possible to download data via services like Geofabrik or use APIs, such as the Overpass API, for custom queries. Its open nature allows researchers to replicate or build upon the data without any restriction on usage for different regions, ensuring consistent applicability in multiple case studies.

4.3.3 Sensitivity/GDPR

OSM does not collect personal data in its geospatial content; it contains only infrastructure and geographic data, without any personal identifiers. Therefore, it does not raise GDPR concerns. However, contributor metadata (like usernames and edit history) can be considered personal under GDPR.

4.3.4 DC8 Use

OpenStreetMap (OSM) data was utilized to extract the road network and provide geographic reference for building queries in the imagery datasource APIs. Points of Interest (POI), such as building presence, and landuse data were also extracted to provide insights of urbanization, complexity and density and used as input features for the safety models.

4.3.5 DC9 Use

OpenStreetMap (OSM) data was utilized to extract the road network and represent it as a graph structure, enabling spatial analysis through the use of spatial entities. Moreover, selected road infrastructure features were used as input for the models to be integrated with telematics data, offering a richer contextual understanding.

4.3.6 DC10 Use

OpenStreetMap (OSM) data was used to extract the infrastructural characteristics of a road. The available datasets were enriched by incorporating OSM tags linked to the corresponding geometries. These contextual attributes—such as information about road type or quality of cycling infrastructure—were subsequently transformed into a structured feature set which served as the input for training predictive models designed to identify potentially unsafe locations.

4.3.7 DC13 Use

Road network geometries were obtained from OSM, where they are represented as horizontal alignment lines on the Earth’s surface. In their native form, OSM curvatures are segmented, consisting of multiple node connections. These geometries were refined in Autodesk Civil 3D by reducing segmentation and adjusting the number of nodes to an appropriate level. Following this refinement, Building Information Modelling (BIM) road corridors were generated by extruding predefined structural profiles (assemblies) along the Civil 3D alignments.

4.4 Geolocations of Traffic Accidents (DC10 & DC13)

4.4.1 Description/Origin

Origining the Federal Police, Statbel (the Belgian National Statistics Office) provides official crash datasets (STATBEL, 2024), including time, location (Belgian Lambert 72 coordinates), crash severity, involved parties, and contextual factors.

4.4.2 Availability/Replicability

Public datasets are available online from 2017 to 2024, either in an .XLSX or delimited, plain text format. Conclusions from the data are replicable within Belgium; for international replication, equivalent national crash databases are required.

4.4.3 Sensitivity/GDPR

The open data does not include sensitive personal data that could infringe on privacy.

4.4.4 DC 10 Use

The open-source dataset will be used to integrate the objective dimension of safety into the routers. The crash data will help identify risk-prone locations based on objective measures. A key emphasis of this doctoral study is to better understand the relationship between perceived safety and safety as defined by objective metrics. This dataset, when combined with the Route2School dataset, will also provide a basis for a comparative analysis between objective and subjective safety.

4.4.5 DC 13 Use

Combined with Data Dangerous Points by Flanders Digital Agency (mentioned below), these datasets help define the objectives of the "AI-supported, BIM-based design for road safety" framework. For example, in areas where vulnerable road users dominate the road users in crashes, VR-based simulation tests will be developed. Conversely, in areas where the majority of crash-involved parties are drivers, driving simulator tests will be designed.

4.5 Video Data-Uhasselt,Cegeka (DC11)

4.5.1 Data source overview

This research relies on traffic video data collected from real-world road intersections as the primary source of information for analysis. Video data provides rich observations of road user behavior (positions, movements, interactions) that are essential for detecting conflicts and assessing safety at a microscopic level. Initially, the project explored using an existing dataset from the MiA (Mobiliteit Innovatief Aanpakken) project, which consisted of multi-angle video recordings at five intersections in Belgium. While this provided a valuable starting point, the MiA data presented technical challenges (e.g. inconsistent timestamps between cameras) that hindered its use for precise multi-camera trajectory analysis. As a result, a new dedicated dataset tailored to the project's requirements was collected. This new data source is a video dataset from a busy intersection in Diepenbeek, Belgium, recorded in Spring 2025. The Diepenbeek site was selected for its representative traffic mix (including pedestrians, cyclists, and vehicles) and known safety concerns, making it ideal for conflict analysis research. Two cameras were installed to cover the intersection approaches from different angles, enabling 3D reconstruction of road user trajectories. Data gathered in batches between late March and mid-May 2025, many hours of traffic footage were recorded under various conditions. This large volume of raw video now serves as the foundation for developing and validating AI framework for object detection, tracking, and conflict analysis. In summary, the project's data sources comprise multi-camera traffic videos obtained in a real-world setting, offering an authentic and complex environment to test the methodology. By working with actual intersection video, the research ensures that the outcomes (models, algorithms, safety findings) are grounded in realistic scenarios and can be translated to practical road safety interventions.

4.5.2 Data Source #1 – Diepenbeek Traffic Video Dataset

4.5.2.1 Description/Origin

The primary dataset for this research is a custom-collected traffic video dataset from the Diepenbeek intersection in Belgium. This dataset was acquired in collaboration with UHasselt transportation researchers, using advanced camera equipment deployed on-site. The intersection is a medium-sized urban crossroads known to handle significant volumes of vehicular traffic as well as pedestrians and cyclists due to being close by Uhasselt, providing a rich environment for observing a variety of road user interactions. Two synchronized cameras were mounted at different angles to capture the whole area of intersection. The video recording campaign was conducted over multiple periods in 2025: approximately three weeks of data were collected in total, spread over late March, early April, and early May. This timing ensured inclusion of different traffic patterns (weekday rush hours, off-peak, weekend days) and environmental conditions (varying daylight and weather), making the dataset comprehensive. Each video is high-resolution (sufficient to detect and distinguish road users) and timestamped. In summary, the Diepenbeek video dataset originates from a field data collection effort by the project and is uniquely suited for the research objectives: it provides the raw visual data needed to develop and test the AI-based conflict detection and trajectory analysis methods under real traffic conditions.

4.5.2.2 Availability/Replicability

The Diepenbeek traffic video dataset is currently proprietary to the Uhasselt and is not publicly available, due to privacy and confidentiality considerations. Access to the raw videos is restricted to the research team and authorized project partners. However, the methodology used to collect this data is replicable, and similar datasets could be obtained by other researchers or road authorities with the proper equipment and legal/data permissions. Specifically, to replicate this data collection, one would need to set up video cameras at a target intersection and record sufficient footage of traffic flows, following the same privacy safeguards (see Section 6.2.3). The general characteristics of the dataset – multi-angle video of intersection traffic – are comparable to other traffic conflict video datasets reported in the literature, but what sets this dataset apart is its focus on enabling multi-camera trajectory analysis (hence the careful synchronization). In terms of availability of processed data, the project may derive and share *anonymized trajectory data* or aggregated conflict statistics from the videos in the future. For example, once personal identifiers are removed, the team could provide trajectory sets or conflict event logs for academic purposes. This would allow others to reproduce the conflict analysis without accessing sensitive raw footage. As of now, any external replication of results would require collecting a similar video dataset. The consistency of the data collection procedure (fixed cameras, defined time periods) ensures that the analyses performed are reproducible on this dataset, and they could be repeated on new data from other sites to verify the generality of the findings.

4.5.2.3 Sensitivity/GDPR

Traffic video data inherently contains sensitive personal information, since individuals (drivers, cyclists, pedestrians) can appear in the footage, potentially identifiable by their faces, vehicle license plates, or other attributes. As such, the Diepenbeek dataset is handled under strict GDPR compliance and ethical protocols. Prior to data collection, a Data Protection Impact Assessment was carried out to evaluate and mitigate privacy risks, in line with legal requirements. Several privacy-by-design measures were implemented during the data acquisition and processing: **(a)** Cameras were positioned and configured to focus on traffic movement rather than focus on pedestrian faces. **(b)** All video files are transferred only over secure channels with access limited to authorized personnel. Additionally, people in the area of the camera installation were informed of the ongoing traffic study through visible notices, satisfying the GDPR transparency requirement. No biometric or personal data beyond what is incidentally captured in public space is collected. In summary, the dataset is treated as highly sensitive: all handling of the video data abides by GDPR principles such as data minimization, purpose limitation, and security. These precautions ensure that the research can benefit from detailed traffic observations for safety analysis without compromising individual privacy rights.

4.6 iRAP-BH Dataset (DC12)

4.6.1 Description/Origin

iRAP-BH dataset comprises georeferenced road videos from 194 public road sections in Bosnia and Herzegovina, which is around 2,300 km. Videos were captured in 2704 × 2028 RGB and indexed into 10m segments. Each 10m road segment is coded by human coders for all iRAP road attributes.

4.6.2 Availability/Replicability

The dataset is not publicly available to anyone outside of the organization. The collection protocol (georeferenced forward-facing video, frame extraction, 10m segmentation) is documented in section Data Acquisition so that other teams can replicate the dataset in comparable settings. Pre-processing steps (frame sampling, segment indexing, label alignment) are deterministic and can be reproduced to obtain identical train/val/test splits when the same seeds and segment lists are used.

4.6.3 Sensitivity/GDPR

Our dataset and the AI model training techniques are fully compliant with the European Union's General Data Protection Regulation. The dataset does not contain any personal information of any individual or entity. However, some raw videos from the Bosnia dataset and the Croatian A1 motorway LiDAR dataset may contain visible components, like as license plates or human faces; these are explicitly ignored throughout the data preparation and model training workflow. These data points are not annotated, analyzed, or incorporated into the features used to train the AI system.

4.7 LiDAR Dataset of the A1 Motorway (Croatia) (DC12)

4.7.1 Description/Origin

This dataset covers nearly 570 km of the A1 motorway and consists of synchronized RGB imagery (1920 × 1080 at 10 fps) and LiDAR point clouds collected from a survey vehicle. The roadway is partitioned into 57,000 ten-meter segments, each annotated by human experts with 52 iRAP attributes.

4.7.2 Availability/Replicability

Access is restricted under agreements with the data owners.

4.7.3 Sensitivity/GDPR

Our dataset and the AI model training techniques are fully compliant with the European Union's General Data Protection Regulation. The dataset does not contain any personal information of any individual or entity. However, some raw videos from the Bosnia dataset and the Croatian A1 motorway LiDAR dataset may contain visible components, like as license plates or human faces; these are explicitly ignored throughout the data preparation and model training workflow. These data points are not annotated, analyzed, or incorporated into the features used to train the AI system.

4.8 Dangerous Points by Flanders Digital Agency (DC13)

4.8.1 Description/Origin

Agentschap Digitaal Vlaanderen publishes annual hotspot maps (Agentschap Digitaal Vlaanderen, 2024) that identify high-risk crash locations based on official statistics and safety models. These points guide the prioritisation of locations for design intervention.

4.8.2 Availability/Replicability

Freely accessible via WMS/WFS services and downloadable files. Replicability is specific to Flanders, though similar hotspot detection initiatives exist in other regions.

4.8.3 Sensitivity/GDPR

Contains aggregated location-based risk indicators, not individual crash records. GDPR risk is minimal as the dataset does not include personal identifiers.

4.9 Microsumulation (SUMO) (DC13)

4.9.1 Description/Origin

SUMO (Simulation of Urban MObility) (Krajzewicz et al., 2002) produces traffic microsimulation data, including vehicle trajectories, interactions, and surrogate safety measures (e.g., time-to-collision, post-encroachment time). It provides dynamic insights into how road layouts and design alternatives perform under traffic demand.

4.9.2 Availability/Replicability

SUMO is open-source and reproducible. Simulation scenarios can be rebuilt from OSM networks or Civil 3D exports, ensuring high replicability across contexts. Inputs (traffic demand, routes) must be calibrated with local conditions for validity.

4.9.3 Sensitivity/GDPR

Simulation data is synthetic and contains no personal information.

4.10 Human driver simulations (DC13)

4.10.1 Description/Origin

At the final stage of the project, optimised road designs will be evaluated through driving simulations with human participants. Testers will interact with a simulator environment to assess usability, safety perception, and behavioural responses to the proposed road layouts.

4.10.2 Availability/Replicability

This data is generated uniquely within the project and is not publicly available. Replicability is possible in other contexts provided that comparable driving simulator infrastructure and participant recruitment are in place.

4.10.3 Sensitivity/GDPR

As this involves human participants, GDPR applies. To ensure compliance, no personal recordings will be kept. Each participant will be assigned a pseudonymised code, and only coded performance and behavioural data will be stored. Informed consent will be obtained, and data will be handled in accordance with ethical approval procedures.

4.11 VR-based Pedestrian Simulations (DC13)

4.11.1 Description/Origin

In parallel to driving simulator tests, VR-based simulations will be conducted with vulnerable road users (e.g., pedestrians, cyclists) to evaluate the optimised road designs. Participants will experience scenarios through VR headsets, allowing the study of safety perception and behavioural responses in immersive environments.

4.11.2 Availability/Replicability

This data is generated specifically within the project. Replicability is possible in other studies provided that comparable VR infrastructure and participant recruitment are available.

4.11.3 Sensitivity/GDPR

GDPR is applicable, as participants are directly involved. No personal recordings will be stored. Each participant will be pseudonymised with a code, and only coded behavioural data (e.g., response times, gaze direction, trajectory choices) will be retained. Informed consent will be secured, and data processing will follow ethical approval protocols.

4.12 Route2School Dataset (DC10)

4.12.1 Description/Origin

This is the primary dataset that will be utilized throughout this research. The dataset comprises safety labels for home-to-school routes in approximately 80 cities across Belgium, based on user perception and expert opinions. It consists of a network of roads that are experienced as safe, less safe and downright unsafe. This Geospatial dataset is composed of several geocoded road segments and an accessibility label among - SAFE, OCCASSIONALY_UNSAFE, and UNSAFE associated with them.

4.12.2 Availability/Replicability

The dataset contains information about links spread across Flanders region in Belgium. The dataset is provided by the industrial partner of this doctoral project (AbeonaConsult, Belgium). Route2School Website: https://www.route2school.be/about/

4.12.3 Sensitivity/GDPR

The dataset does not contain any sensitive information about the users. It stores anonymous feedbacks collected from the users and hence does not contain any private data.

4.13 FRED Engineering Data (DC8)

4.13.1 Description/Origin

Video data corresponding to 9 street-level recordings, with approximately 10 minutes of duration each. They all have 29.97 frames per second (fps) rate, 2304 x 1296 pixel-size (width x height) and were surveyed on the 17th and 18th of September of 2024. It covers 64 km of motorways from the provinces of Udine and Gorizia. On the video, there is a logo on the top-right corner, and in the bottom of the frame it states the geographical position in degrees, the velocity and the time of the recording in the format hh:mm:ss YYYY/MM/DD.

The correspondent kml file includes GPS longitude and latitude data for each second of the video. Comma separated values are provided for the same area, with iRAP-style manual coding of road attributes every 100 meters for the recorded motorways.

4.13.2 Availability/Replicability

The dataset is owned by FRED Engineering and can only be made available under direct authorization from the owners. Replicability is not possible.

4.13.3 Sensitivity/GDPR

Video data might contain sensitive information such as license plate numbers or human faces. Sensitive information needs to be previously processed and blurred, as well as explicitly ignored throughout the data preparation and model training workflow. These data points are not annotated, analysed, or incorporated into the features used to train the AI system.

4.14 Mapillary Data

4.14.1 Description/Origin

Mapillary is a crowd-sourced platform with street-level imagery all around the globe, with a CC BY-SA (Creative Commons Attribution-ShareAlike) license and, although of free access for research purposes, user must register to gain access to it. The Mapillary team have also worked on a subset of 25,000 labeled images called Mapillary Vistas dataset, with instance segmentation labeled data for 124 semantic object categories, divided into object, construction, human, nature, marking, void and animal types. Although they comprise images with different ratios and resolution, all images are at least Full-HD, and the most common ratios are 4:3 and 16:9.

To use Mapillary data, there is a provided API with standard OAuth authorization, as explained in details by the website managers, that allows images to be selected within a specified geolocation and data range, among other queries. Some specific type of data can also be requested, like points of interest or traffic signs, based on their annotations. As for Mapillary Vistas, the data can be downloaded via a research application issued by a registered user.

4.14.2 Availability/Replicability

Mapillary data is freely available under an open-content license. It is highly accessible for use, redistribution and upload. It is possible to download the data through command line and API. Its open nature allows researchers to replicate or build upon the data without any restriction on usage.

4.14.3 Sensitivity/GDPR

Image data might contain sensitive information such as license plate numbers or human faces. Sensitive information needs to be previously processed and blurred, as well as explicitly ignored throughout the data preparation and model training workflow. These data points are not annotated, analysed, or incorporated into the features used to train the AI system.

5 Wider methodological contributions

5.1 Overview of methodological contributions

In the previous, the outline of the contemporary advancements and methodologies for AI-supported road safety applications, as well as the placement of the IVORY WP6 doctoral theses as case studies within the broader state-of-the-art and their respective data sources were all provided. Advancing further, the present section aims to showcase the methodological contributions that will stem from these doctoral projects. Three main fields are defined as forefronts of AI applications in road safety, namely (i) AI-supported road risk detection, (ii) AI-supported road risk analysis and (iii) AI-Supported road risk management, in which the six doctorates contribute. Moreover, methodological contributions in these fields will not only address academic gaps, but will also be taking into account validation and transferability processes. This will enable them to provide practical relevance for policy, planning, and road safety practice.

5.2 AI-supported road risk detection

The contribution of IVORY WP6 for road risk detection relies mostly on proactive analysis and it is significantly enhanced by AI technologies, placing the DCs' methodologies on the front end of transportation safety research. Instead of relying solely on historical crash data, the presented risk detection methodologies proposed are based on in-vehicle telematics data and video or image-based surveillance systems which enable proactive and preventive risk detection. Furthermore, the logic of assessing safety levels for existing infrastructures is inverted by DC13's research, where the outcome of crash-based safety investigations are used to propose design and planning procedures that minimize the risk of crash occurrence.

Based on in-vehicle telematics data mapped onto a road network, DC9 used a GNN model to generate a representation of spatial entities (nodes and edges) capturing both local contextual information and the topological structure. These new representations, named embeddings, have been used as input for a clustering algorithm aiming to identify different areas within a road network, while also improving the clustering accuracy. The clustering analysis identified two main areas: one corresponding to locations where drivers exhibit more aggressive behaviour and another where driving behaviour tends to be more cautious. The presented methodology enhances risk detection by applying AI models to road networks characterized by in-vehicle data, enabling the creation of telematics-based risk maps.

DC8 methodology also relies on in-vehicle telematics data. By using harsh events as the dependent variable and infrastructure data labelled using iRAP methodology, DC8 is going to identify key elements that contribute to road risk and have been understudied within the computer vision field. Specific AI-based semantic segmentation and object detection algorithms are trained to extract additional data from street-level imagery datasets, as has already been made with lane marking types. Therefore, DC8 proposes a framework for identifying risk-prone areas based on different detail levels of road attribute features extracted from imagery datasets. The research also explores the transferability of these algorithms with open-source data, supporting low-cost and continuous mapping of risky locations.

With a different approach yet building on proactive vision-based infrastructure assessment, the methodology presented in Section Methodological framework of DC12 introduces a significant advancement in AI-supported road risk detection by leveraging computer vision techniques for automated extraction of road infrastructure attributes from georeferenced road survey videos, aligned with iRAP standards. DC12 has leveraged a pre-trained hybrid Vision Transformer (MaxViT) model as the backbone and fine-tuned it on the iRAP-BH dataset comprising approximately 230,000 annotated 10-meter road segments from 2300 km of highways in Bosnia and Herzegovina. Novel custom modules proposed by DC12 enhance the model's capability to handle temporal sequences across multiple frames per segment, incorporating 1D convolutions, GRU layers, and rarity-based weighting to emphasize salient features. To address severe class imbalances in the dataset, where infrequent attributes like specific roadside objects or school zones are underrepresented, a dynamic weighted focal loss is applied, which adjusts weights based on class-specific recall during training. This mechanism ensures unbiased learning across all attributes, improving detection of minority classes without sacrificing performance on dominant ones.

While working with real time video data, DC11's research significantly improves the detection of road safety risks by deploying AI systems capable of identifying near-miss incidents and conflict situations. The developed solution automates the process of risk detection through continuous video monitoring and hazardous interactions identification. The system integrates high-precision object detection for vehicle and pedestrian recognition with multi-camera tracking techniques. This enables the detection of subtle and potentially dangerous behaviours that may precede collisions, such as a pedestrian stepping into a crosswalk while a vehicle approaches at unsafe speed, or vehicles approaching an intersection with ambiguous right-of-way. A key innovation lies in the use of multi-angle visual data. By synchronizing inputs from multiple cameras, the system overcomes visual occlusions and blind spots inherent in single-view setups, thereby reducing false negatives in conflict identification. Furthermore, the system incorporates privacy-preserving design elements, including on-device processing and data anonymization, making it suitable for deployment in public spaces without compromising individual privacy. This system equips traffic safety agencies with an essential tool for monitoring critical safety events as they happen, enabling earlier and more targeted interventions.

Finally, the framework developed by DC13 enables a form of design-phase risk detection. Hazardous locations are predefined using official datasets, such as the Dangerous Points from the Flanders Digital Agency (Agentschap Digitaal Vlaanderen, 2024), complemented by crash information from Statbel (STATBEL, 2024). Once a proposed design is modelled in BIM, alternative layouts are generated by the AI and evaluated in SUMO, where unsafe designs can be identified through indicators such as frequent conflicts or poor surrogate safety measures

(SSMs). These weaknesses become visible through the contrast between the original proposal and the AI-optimised outcome. In this way, the focus of risk detection shifts from locating hazards in the existing road network to uncovering risky design decisions during the planning process itself.

With the outcome of these different AI-driven research approaches, IVORY contributes directly to proactively detecting road risk in different phases of the road network lifespan and based on different data source types, allowing stakeholders to evaluate amongst the proposed methods what best suits their resources and objectives towards road safety.

5.3 AI-supported road risk analysis

An essential element of proactive road safety is recognizing and analyzing its multiple dimensions, including both objectively measurable indicators (e.g., crash statistics, traffic speed, road design features) and subjective perceptions of safety as experienced by road users. While the projects in WP6 pursue different facets of safety, their collective work serves as a foundation for developing more comprehensive and data-driven strategies to assess and address risks in complex traffic environments. The methodologies outlined in this report aspire to serve as a leading step in the advancement of AI-supported risk analysis for road safety.

For instance, the methodology of DC9 presents an innovative framework based on clustering techniques to identify node- and edge-based high-risk areas on a map characterized by telematics data. In order to improve the clustering performance, GNN models were employed to learn enhanced representations of both nodes and edges. The new spatial entity representations lack interpretability due to the abstraction involved in their generation. Therefore, to interpret the clustering results, the labels produced by the embedding-based clustering are mapped back to the raw feature dataset. Feature values are averaged within each cluster, allowing for the identification of the high-risk cluster as one characterized by higher average speeds per spatial entity and a greater frequency and intensity of hazardous events, such as harsh driving maneuvers and mobile phone usage. Moreover, a simple econometric regression model may be trained on the data using the clustering-derived labels as the dependent variable. By examining its coefficients, it is possible to gain insight into which features are most influential in determining cluster membership.

The primary focus of DC10’s research is to improve existing routing solutions by incorporating safety, particularly subjective user perceptions, into them. The research aims to investigate approaches to generate a safe road network by employing AI techniques to identify unsafe (risk-prone) locations. It would explore the possibility of combining objective attributes with subjective user feedback to train models to classify road segments as safe or unsafe.

The methodology presented by DC10 also attempts to develop a proactive framework by analyzing various attributes associated with a road segment to generate a safety score linked with each segment. The intention here is to modify the travel cost of a route by adding a weight for unsafety such that the router suggests safer routes. As subjective safety is a critical component of this research, it also aims to utilize the Route2School dataset to identify a correlation between human perceptions and infrastructural characteristics. Given that the dataset represents public opinion on the safety of roads across Belgium, the study seeks to develop a model that can reveal nuanced patterns within the data, thereby providing insights into the subjective dimensions of road safety. This approach extends traditional safety analysis, which primarily emphasizes objective indicators, by adding a component of public perception. The research aims to set a foundation for AI-powered smart routing solutions that can evolve over time and suggest safer, more user-oriented routes.

Beyond detection, the research of DC11 advances the analysis of road safety risks by introducing a comprehensive framework for understanding traffic conflicts in depth. Once conflicts are detected, the framework evaluates them using a fusion of surrogate safety measures such as Time to Collision (TTC), Post-Encroachment Time (PET), and Delta-V. These indicators collectively assess proximity, timing, and severity of road user interactions, providing a comprehensive evaluation of risk. For example, a near-miss involving a small PET and a large Delta-V (indicating sudden braking) would be classified as highly severe, whereas conflicts with larger PETs and lower speed changes would be considered less critical. This multi-indicator fusion surpasses traditional analyses that rely on conflict frequency alone or on isolated metrics. The methodology also enables pattern discovery by analysing conflict characteristics over time, location, and contextual conditions (e.g., weather, time of day, traffic volume). It highlights not only where and when risks occur, but also which types of movements (e.g., left turns, pedestrian crossings) are most problematic. Importantly, because it operates independently of crash records, the system can identify high-risk locations even in the absence of historical accidents. Another important contribution is in the accessibility and presentation of results. The conflict analysis is visualized through interactive dashboards, complete with animated replays of events based on tracked trajectories. These animations, built from data rather than raw video, ensure privacy while allowing stakeholders to visually interpret the dynamics of each near-miss. This approach enhances both academic insight and the practical usability of findings by planners and decision-makers.

The methodology provided by DC12 advances AI-supported road risk analysis by providing a robust, data-driven framework for interpreting extracted infrastructure attributes and quantifying safety risks in alignment with iRAP protocols. Unlike traditional methods, which rely on historical crash data or manual road evaluations, which frequently miss subtle risk patterns and lack scalability, our AI-based model allows for granular, proactive analysis at 10-meter road segments. The model produces precise classifications of attributes such as roadside severity, pedestrian facilities, and intersection quality by processing multi-frame video sequences through a finely tuned hybrid computer vision model augmented with temporal and dynamic weighted focal loss to handle data imbalance. These extracted road attributes will help in identifying the non-safe road segments by calculating probabilistic risk scores per segment by aggregating attribute impacts on vehicle occupants, pedestrians, cyclists, and motorcyclists. The use of rarity-based weighting mechanism used during AI-model training ensures that infrequent but critical risks (e.g., poor delineation or school zone warnings) are not overlooked, allowing for comprehensive pattern analysis across large networks.

Lastly, DC13's methodological framework does not invent new safety metrics; it relies on SUMO's existing SSMs (e.g., TTC, PET, conflict counts) to train the generative AI. For each case, the designer's initial SSM outcomes are reported alongside the AI-optimised outcomes, making the safety effect visible as before/after deltas. Its main contribution is therefore to render the analysis comparative and design-centred.

Within the scope of this research, the framework of DC13 is limited to improving the safety performance of selected cases, rather than training the model to gain a general understanding. In principle, the model could also be extended to provide targeted feedback on which design modifications improve specific safety metrics in any given case. Realising such functionality, however, would require a broad collection of use cases to capture sufficient variability.

5.4 AI-supported road risk management

Each DC contributes risk management either through their work or with the workframe they establish during the study. Therefore, there are several aspects of AI-supported road risk management that the project suggests improvements from improving data management to recommendation of targeted actions.For example, the framework of DC8 contributes to road risk analysis by studying the relationship between infrastructure attributes extracted from imagery datasets and surrogate safety measures (harsh events). The attributes extracted are both general and transportation-specific, such as road markings, allowing for different types of analysis. Sinceeffective road safety risk analysis requires overcoming key challenges in data-driven methods, particularly ensuring data quality and reliable ground-truth labels (Sohail et al., 2023), and road features are unevenly distributed by nature, presenting a comprehensive proactive road safety framework, where so many factors contribute to the outcome (infrastructure, environmental, social, behavioural), not only manual data processing is a overwhelming task, but maintaining data balance is very challenging. During the research of DC8, however, attention is placed on specific road attributes, enabling micro-level analysis not only of their presence but also their quality and type, such as lane delineation, directly linked to safety outcomes. By applying established statistical models, such as logistic regression, the relationship between these features and risk measures will be rigorously examined considering contribution to the outcome and causality relationships based on transportation theory. Ultimately, the research aims to deliver an AI-driven framework capable of detecting and flagging poorly managed road attributes, thus supporting targeted risk management interventions.

On the other hand, the work has been done by DC9 at this stage contributes to AI-supported road risk management by providing actionable insights. The proposed framework may be used to inform on where to focus safety efforts and resources, aiming to improve overall traffic management and public safety. For instance, risky cluster areas can be targeted for interventions to enhance road safety, such as infrastructure improvements, awareness campaigns, or enforcement measures. Furthermore, insurers can use this clustering to create risk profiles by identifying patterns of risky and safe driving. Drivers in high-risk clusters might face higher premiums, while those in safer clusters could receive discounts. Overall, this approach allows insurers to deliver more precise location-based pricing and tailored advice, thereby encouraging safer driving behaviours.Similar to DC9, DC11's contribution emerges around identification of safety improvements. In this regard, a key innovation is the development of an AI-powered Decision Support System (DSS) that leverages conflict data to suggest targeted safety improvements. Acting as a digital assistant for urban planners and traffic engineers, the DSS processes trends in conflict data and recommends context-specific countermeasures. For instance, if recurring pedestrian-vehicle conflicts are detected at a specific crosswalk, the system might propose changes such as introducing a dedicated pedestrian signal phase or improved signage. In cases of frequent hard braking and rear-end near-collisions, it may suggest traffic calming features or speed limit adjustments. Crucially, the DSS supports simulation and forecasting: it can estimate how specific interventions would impact safety metrics like TTC and PET, enabling predictive evaluation of crash risk reduction. As interventions are deployed and new data collected, the system evaluates whether conflict frequencies and severities have improved, refining future recommendations in a continuous learning loop.This adaptive, AI-assisted approach brings agility to traffic safety management. Rather than waiting years for crash data to accumulate, authorities can identify and address emerging safety concerns in a matter of weeks or months, with real-time feedback. In a broader context, this system supports strategic road safety initiatives—such as Vision Zero—by identifying high-risk locations and tracking progress through reductions in conflict indicators.

The proposed methodology in section by DC12 also focuses at safety improvements and contributes to AI-supported road risk management by converting analysed infrastructure data into actionable insights for targeted interventions, thereby fostering long-term improvements in road safety management systems. Traditional management approaches are frequently reactive and resource-intensive, relying on post-incident responses or infrequent audits, whereas our AI-based system provides real-time, scalable decision support via automated road attribute extraction and road risk profiling. The handling of data imbalances ensures that vulnerable user groups (e.g., pedestrians and cyclists) receive equitable attention, while the planned multi-sensor fusion will provide precise metrics for cost-benefit analyses of upgrades, such as optimising roadside object distances to reduce severity risks.

Finally, the framework by DC13 contributes to risk management by moving beyond a single optimised outcome and instead generating multiple alternative designs that satisfy safety improvements. Each option is derived from the same hazardous location but represents a different balance of geometric configuration and traffic performance. This creates a design space where safety is not a constraint added at the end of the process, but a criterion actively shaping the available choices. Final selection is not made by the AI but by stakeholders, supported through First-Person Perspective (FPP) simulations. Human participants experience the shortlisted alternatives and indicate which they find most appropriate. In this way, risk management becomes a shared process: AI ensures that all options presented are demonstrably safer than the baseline, while human judgment incorporates practical, social, and contextual considerations that cannot be captured by metrics alone.

6 Conclusions

The present section aims to summarize the topics and progress of individual DC, and outline the next steps involved in each doctoral research project.

6.1 Recapitulation

6.1.1 DC 8

The research of DC8 aims to develop a proactive framework for road safety assessment by integrating multiple data sources that have not previously been combined, namely street-level imagery and telematics data. The proposed framework is a step towards shifting safety assessment from human-based to computationally-based road attribute data collection, thus allowing for better resource allocation when performing network-level assessments. Video data from FRED Engineering, supported by iRAP safety ratings, is being used to train and fine-tune advanced computer vision models for detecting and segmenting road features, with Mapillary and OpenStreetMap providing additional, crowd-sourced imagery and network data to extend coverage. These enriched datasets will then be linked with OSeven Telematics records on driver behavior, particularly harsh acceleration, which serves as a surrogate safety measure. By applying both conventional and spatially sensitive models such as Random Forest, XGBoost, Logistic Regression, and Geographically Weighted Logistic Regression, this doctoral research seeks to create a novel methodological framework that associates infrastructure characteristics with driver behaviour for comprehensive network-level safety assessment.

6.1.2 DC 9

The research of DC9 focuses on bridging the gap between in-vehicle telematics and AI-based methods by developing a framework that integrates smartphone-recorded driver behavior data with network features from OpenStreetMap, in order to achieve spatial scaling of models. The approach involves aggregating telematics data at both edge and node levels, refining node-based aggregation through buffer constraints to enhance data quality. Local attributes and topological relationships are captured by Graph Neural Networks (GNNs). The produced embeddings serve to enrich subsequent clustering analyses. The research aims to combine GNN outputs with clustering algorithms (K-Means and hierarchical methods), in order to identify high- and low-risk nodes and edges across the road network. Mapping clustered embeddings back to raw features restores interpretability while retaining the predictive power of deep learning. Overall, the methodology introduces an innovative means of using GNN-driven representations to improve road safety monitoring and crash prediction across different spatial scales. This would be achieved by transitioning from traditional reactive analyses based on recorded crashes, which are usually underreported, to proactive monitoring that can identify high-risk locations before crashes occur. The proactive approach through smartphone telematics data can reach over 50% coverage within a metropolitan area of an European capital, such as Athens, allowing interventions to focus on a small subset of nodes (e.g., 4%) for more efficient resource allocation.

6.1.3 DC 10

The research of DC10 involves developing a data-driven framework with the aim to integrate both subjective and objective road safety perspectives into routing algorithms. The methodological approach comprises four phases: knowledge acquisition, data preparation, safe-route implementation, and user-profiling and personalization. Initially, the doctoral research defined research questions around differences between perceived and objective safety, the feasibility of safe-route networks, and the potential role of reinforcement learning. Data from the Route2School survey, OpenStreetMap, and Belgian crash records are exploited. The data are harmonized and processed through feature engineering to capture and generate key safety attributes. Subsequently, supervised learning models are trained in order to predict safety levels. Model outcomes are integrated into routing systems to generate safety-aware paths. Their quality against traditional routes is evaluated in parallel. Finally, user profiling and reinforcement learning are employed to develop a self-learning router capable of adapting to individual preferences, thus personalizing route recommendations and advancing the integration of safety into navigation systems.

6.1.4 DC 11

The research of DC11 involves the creation and implementation of a multi-camera AI framework for proactive road safety analysis, integrating technical, analytical, and decision-support components into a single pipeline. The framework begins with synchronized, high-resolution video collection, with the entire process respecting standing GDPR anonymization demands. This is followed by object detection and multi-object tracking to construct 2D and 3D trajectories of road users through calibration and reconstruction techniques. Surrogate safety measure (SSM) indicators such as Time-to-Collision (TTC) and Post-Encroachment Time (PET) are then computed. From the SSMs, conflict matrices are calculated, serving to identify critical patterns of risk across different road user groups. In parallel, on the systemic-scale, AI-driven models are employed in order to detect systemic safety issues, uncover temporal and spatial trends, and generate predictive scenario-based recommendations. The novelty of this research lies in the fusion of multi-camera computer vision, SSM computation and conflict synthesis into a scalable and GDPR-compliant framework. The outputs can offer policymakers and practitioners a robust, evidence-based tool for transforming raw traffic video into actionable urban mobility strategies.

6.1.5 DC 12

The research of DC12 introduces a methodological approach for developing an AI-based system that automates the extraction of road infrastructure attributes, aligned with iRAP and NAWA standards. The iRAP-BH dataset serves as foundation for training, containing over 230,000 annotated 10-meter road segments. Meticulous data preprocessing ensures that critical road and roadside features are faithfully recorded in the dataset. A hybrid vision transformer–CNN backbone (MaxViT) is fine-tuned with custom temporal modules and dynamic weighted focal loss to handle temporal dependencies and severe class imbalances. To train and validate the model, stratified dataset splits are implemented with several performance metrics used as benchmarks. Additionally, the methodology incorporates data fusion with LiDAR to enhance geometric precision for features such as lane width and roadside distances. The integration of multi-frame temporal modeling, adaptive loss re-weighting, and multi-modal fusion into a single research endeavor is a very novel aspect of the research. This scalable and transferable pipeline improves the recognition of minority classes while maintaining overall accuracy, providing robust and context-aware road safety assessments.

6.1.6 DC 13

The research of DC13 proposes a BIM- and AI-driven framework to enhance road safety through an integrated, multi-stage process. The initialization concerns the identification of hazardous locations using official crash records and hotspot datasets, supported by HotMap, a custom interactive tool for managing and exporting geospatial data. Subsequently, conceptual BIM models are built from OpenStreetMap and processed through Civil 3D to create structured files that ensure interoperability across design, simulation, and AI components. A proof-of-concept of Graph Attention Networks (GAT) with reinforcement learning successfully demonstrates the feasibility of optimising road layouts. As a first step by avoiding building conflicts and by integrating safety-focused optimisation afterwards. The process is executed in SUMO microsimulations, where surrogate safety measures serve as rewards. The AI-generated design alternatives are reintroduced into Civil 3D through custom plugins for further refinement. First-Person Perspective simulations are conducted to incorporate human-centred insights alongside quantitative safety metrics. This concept combines data-driven hazard detection, generative AI, microsimulation, and immersive user evaluation. This innovative research aspect harnesses the power of generative AI, aiming to produce more user-oriented road designs.

6.2 Future steps

6.2.1 DC8

The research to date has centred on developing AI-based computer vision algorithms and evaluating state-of-the-art models for semantic segmentation and object extraction from both street-level and aerial imagery. The next phase will involve using these extracted features as inputs for logistic regression models to analyse their association with harsh driving events (e.g., harsh braking and acceleration, as flagged by the OSeven methodology). This step will establish the foundation for the continuation of the study. Subsequently, iRAP-coded data provided by FRED Engineering will also be incorporated into the logistic regression models. From these analyses, the most influential features for risk classification will be identified and used to guide the development of specialized AI models. The research will follow an iterative cycle (feature selection, attribute extraction, and model fitting), testing various combinations of algorithms, features, and models to optimize the risk assessment framework while minimizing computational cost.

6.2.2 DC9

Future work will focus on advancing the current approach by integrating Transformer-based architectures related to Graph Neural Networks (GNNs), enabling global attention mechanisms rather than relying on local attention coefficients, which may allow the model to capture deeper and more complex underlying pattern. Additionally, incorporating crash data could lead to a shift from a self-supervised framework to a supervised learning setting, allowing for regression or classification tasks. Further exploration of additional telematics features will be considered to enrich the model's input space. Another promising direction involves developing a hierarchical GNN architecture capable of aggregating small subgraphs into vector representations, which can then be used for a more macro spatial analysis.

6.2.3 DC10

The experiments conducted thus far establish a clear link between objective features (infrastructural attributes) and subjective safety. Building on this foundation, the next step is to address the substantial missingness in the OSM dataset using advanced deep learning frameworks such as denoising autoencoders. Efficient validation methodologies must be developed to evaluate the performance of these models. Moreover, a probabilistic classification approach will be explored to generate a weighted safety score for each link, improving the model’s sensitivity to attribute changes and enabling more granular decisions. These weighted scores are critical for integration with routing systems, as they directly inform modifications to router cost functions. To further refine the system, an experimental framework will be devised to incorporate user feedback, combining it with the safety model to generate routes tailored to user requirements. Designing this methodology involves considering factors like feedback collection, intended application, and effect on the model, while also ensuring ethical data collection and handling.

6.2.4 DC11

Next phase of this work will focus on scaling the framework from intersection-level case studies to city-wide and eventually network-wide deployments, supported by federated and distributed learning to accommodate diverse geometries and traffic cultures. Advances in multimodal data fusion such as integrating video with connected vehicle telemetry, radar, and crowdsourced mobility data, will improve robustness under occlusion and adverse weather conditions, while reducing dependence on single-sensor inputs. The integration of explainable AI will be essential to build stakeholder trust, with models that not only detect conflicts but also annotate the causal factors behind risk patterns. At the same time, continuous learning systems will enable the Decision Support System to adapt dynamically as new data and interventions are introduced, ensuring that recommendations remain context-specific and timely. Ultimately, this progression envisions a proactive, adaptive, and transparent AI-enabled safety ecosystem that bridges real-world monitoring, predictive modelling, and policy implementation in support of long-term Vision Zero goals.

6.2.5 DC12

Future work will focus on advancing multi-sensor fusion by integrating monocular video data with LiDAR point clouds to improve geometric precision for critical attributes like roadside object distances, lane widths, road curvatures, and obstacle detection. Different fusion strategies will be explored to address the limitations of single-modality systems, such as depth estimation inaccuracies in monocular vision or LiDAR's limitations in adverse weather conditions. These efforts aim to enhance robustness against challenges like occlusions, data imbalance, and environmental noise. The primary contribution will be the development of a scalable, high-performance AI framework for automated, proactive road risk detection, optimized for diverse road environments. This framework will prioritize real-time processing, high accuracy on imbalanced datasets, and seamless integration into multi-modal road safety auditing systems.

6.2.6 DC13

During the subsequent stages of the study, an intermediate layer between Civil3D and the SUMO-driven GAT-based GAN will be established to manage the training data. The GAN model will first be trained to optimise the nodes within a real-world network at the intersection of Evence Coppéelaan and Westerring in the City of Genk, Belgium. At the initial stage, the model will only adjust the central points of node connections, while the final geometry of the intersection will be determined by SUMO's automated processing. Once improvements in SUMO metrics are observed, supported by a positive learning curve, alternative design BIMs will be generated. These design BIMs will then be employed to conduct FPS-based simulation tests. Nevertheless, further refinement of the pipeline will be required, including the incorporation of a wider range of use cases and road attributes, rather than relying solely on the central points of node connections.

References

Abdel-Aty, M., Pande, A., Das, A., & Knibbe, W. J. (2008). Assessing Safety on Dutch Freeways with Data from Infrastructure-Based Intelligent Transportation Systems. Transportation Research Record, 2083(1), 153–161. https://doi.org/10.3141/2083-18

Agentschap Digitaal Vlaanderen. (2024). Gevaarlijke Punten 2024

Ahmed, Ishtiaque, Othman Che Puan, and Che Ros Ismail. 2013. "A Comparative Review of Road Safety Audit Guidelines of Selected Countries." Jurnal Teknologi (Sciences & Engineering) 65(3).

Ali, Y., Hussain, F., & Haque, M. M. (2024). Advances, challenges, and future research needs in machine learning-based crash prediction models: A systematic review. Accident Analysis & Prevention, 194, 107378.

Alrassy, P., Smyth, A. W., & Jang, J. (2023). Driver behavior indices from large-scale fleet telematics data as surrogate safety measures. Accident Analysis and Prevention, 179. Scopus.

Amundsen, F. &. (1977). Traffic conflict technique: Proceedings of the First Workshop on Traffic Conflicts. Oslo: Institute of Transport Economics (TØI).

Andrew Kwok-Fai Lui, Y.-H. C., Ka-Ho Lo, Wang-To Cheng, Hang-Tak Cheung. (2021). Predictive Screening of Accident Black Spots based on Deep Neural Models of Road Networks and Facilities | Proceedings of the 2021 5th International Conference on Computer Science and Artificial Intelligence.

Archer, J. (2005). Indicators for traffic safety assessment and prediction and their application in micro simulation modelling: A study of urban and suburban intersections. Stockholm: Royal Institute of Technology (KTH).

Autodesk. About Civil 3D .NET Core Development. Retrieved 19/08/2025 from

Aziz, Shawon, and Sewa Ram. 2022. "A Meta-Analysis of the Methodologies Practiced Worldwide for the Identification of Road Accident Black Spots." Pp. 790–97 in Transportation Research Procedia. Vol. 62. Elsevier B.V.

Balsa-Barreiro, J., Valero-Mora, P. M., Berné-Valero, J. L., & Varela-García, F.-A. (2019). GIS Mapping of Driving Behavior Based on Naturalistic Driving Data. ISPRS International Journal of Geo-Information, 8(5), Article 5.

Basili, Elsa, Carla Caschili, Bruno Dalla Chiara, and Michela Pellicelli. 2024. "The Impact of Road Accidents on Hospital Admissions and the Potential of ADAS in Containing Health Expenditure: Evidence from Piedmont Data." Transportation Research Interdisciplinary Perspectives 25. https://doi:10.1016/j.trip.2024.101125

Ben-Gal, I. (2007). Bayesian networks. In F. Ruggeri, F. Faltin, & R. Kenett (Eds.), Encyclopedia of statistics in quality & reliability. Wiley & Sons. Retrieved August 30, 2025, from

Bock, J. D., & Verstockt, S. (2021). SmarterROUTES-A Data-driven Context-aware Solution for Personalized Dynamic Routing and Navigation. ACM Transactions on Spatial Algorithms and Systems, 7(1). Scopus.

Boeing, G. (2024). Graph Simplification Solutions to the Street Intersection Miscount Problem (No. arXiv:2407.00258). arXiv.

Boylan, J., Meyer, D., & Chen, W. S. (2024). A systematic review of the use of in-vehicle telematics in monitoring driving behaviours. Accident Analysis & Prevention, 199, 107519.

Boylan, J., Meyer, D., & Chen, W. S. (2024). A systematic review of the use of in-vehicle telematics in monitoring driving behaviours. Accident Analysis & Prevention, 199, 107519.

Brkić, Ivan, Mario Miler, Marko Ševrović, and Damir Medak. 2022. "Automatic Roadside Feature Detection Based on Lidar Road Cross Section Images." Sensors 22(15). https://doi:10.3390/s22155510 .

buildingSMART International. Infrastructure Domain. buildingSMART International. Retrieved 23/08/2025 from

Bustos, C., Rhoads, D., Solé-Ribalta, A., Masip, D., Arenas, A., Lapedriza, A., & Borge-Holthoefer, J. (2021). Explainable, automated urban interventions to improve pedestrian and vehicle safety. Transportation Research Part C: Emerging Technologies, 125. Scopus.

Cai, Q., Abdel-Aty, M., Lee, J., & Huang, H. (2019). Integrating macro- and micro-level safety analyses: A Bayesian approach incorporating spatial interaction. Transportmetrica A: Transport Science, 15(2), 285–306.

Cao, Kaidi, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. 2019. "Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss." Advances in Neural Information Processing Systems 32.

Chakraborty, M., Gates, T. J., & Sinha, S. (2023). Causal Analysis and Classification of Traffic Crash Injury Severity Using Machine Learning Algorithms. Data Science for Transportation, 5(2), 12.

Chan, R., Dardashti, R., Osinski, M., Rottmann, M., Brüggemann, D., Rücker, C., Schlicht, P., Hüger, F., Rummel, N., & Gottschalk, H. (2023). What should AI see? Using the public's opinion to determine the perception of an AI. AI and Ethics, 3(4). https://doi.org/10.1007/s43681-022-00248-3

Chantith, Chaturaphat, Chompoonuh K. Permpoonwiwat, and Bertrand Hamaide. 2021. "Measure of Productivity Loss Due to Road Traffic Accidents in Thailand." IATSS Research 45(1):131–36. https://doi:10.1016/j.iatssr.2020.07.001 .

Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 839–847.

Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. 2017 IEEE Visual Communications and Image Processing (VCIP), 1–4.

Chen, C., Zhang, G., Qian, Z., Tarefder, R. A., & Tian, Z. (2016). Investigating driver injury severity patterns in rollover crashes using support vector machine models. Accident Analysis & Prevention, 90, 128–139.

Chen, Simiao, Michael Kuhn, Klaus Prettner, and David E. Bloom. 2019. "The Global Macroeconomic Burden of Road Injuries: Estimates and Projections for 166 Countries." The Lancet Planetary Health 3(9):e390–98. https://doi:10.1016/S2542-5196(19)30170-6 .

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations (No. arXiv:2002.05709). arXiv.

Cinnamon, J., & Jahiu, L. (2021). Panoramic Street-Level Imagery in Data-Driven Urban Research: A Comprehensive Global Review of Applications, Techniques, and Practical Considerations. ISPRS International Journal of Geo-Information, 10(7), 471.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., … Schiele, B. (2016). The Cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3213–3223). IEEE.

D'Amicantonio, G., Bondarev, E., & De With, P. H. N. (2023). Automated camera calibration via homography estimation with graph neural networks. arXiv preprint arXiv:2311.02598.

Doiron, D., Setton, E. M., Brook, J. R., Kestens, Y., McCormack, G. R., Winters, M., Shooshtari, M., Azami, S., & Fuller, D. (2022). Predicting walking-to-work using street-level imagery and deep learning in seven Canadian cities. Scientific Reports, 12(1). Scopus.

Dong, Hao, Weihao Gu, Xianjing Zhang, Jintao Xu, Rui Ai, Huimin Lu, Juho Kannala, and Xieyuanli Chen. 2024. "Superfusion: Multilevel Lidar-Camera Fusion for Long-Range Hd Map Generation." Pp. 9056–62 in 2024 IEEE International Conference on Robotics and Automation (ICRA). IEEE.

Du, C., Deng, Z., Nousias, S., & Borrmann, A. (2025). Predictive Modeling: BIM Command Recommendation Based on Large-scale Usage Logs.

Du, C., Esser, S., Nousias, S., & Borrmann, A. (2024). Text2BIM: Generating Building Models Using a Large Language Model-based Multi-Agent Framework.

Du, Y., Song, Y., Yang, B., & Zhao, Y. (2022). StrongSORT: Make DeepSORT great again. arXiv preprint arXiv:2202.13514v1. Retrieved from

Dubská, M., Herout, A., Juránek, R., & Sochor, J. (2014). Fully automatic roadside camera calibration for traffic surveillance. IEEE Transactions on Intelligent Transportation Systems, 16(3), 1162-1171.

Durai, A., Vempati, S., & Bindschaedler, L. (2023). Accurate automatic camera calibration on low-quality CCTV traffic video streams. Unpublished manuscript.

Erramaline, A., Badard, T., Côté, M.-P., Duchesne, T., & Mercier, O. (2022). Identification of Road Network Intersection Types from Vehicle Telemetry Data Using a Convolutional Neural Network. ISPRS International Journal of Geo-Information, 11(9), Article 9.

Eskandari Torbaghan, M., Sasidharan, M., Reardon, L., & Muchanga-Hvelplund, L. C. W. (2022). Understanding the potential of emerging digital technologies for improving road safety. Accident Analysis & Prevention, 166, 106543.

European Data Protection Board. (2020). Guidelines 3/2019 on processing of personal data through video devices (Version 2.0). European Data Protection Board.

Farah, H., Bianchi Piccinini, G., Itoh, M., & Dozza, M. (2019). Modelling overtaking strategy and lateral distance in car-to-cyclist overtaking on rural roads: A driving simulator experiment. Transportation Research Part F: Traffic Psychology and Behaviour, 63.

Fey, M., Sunil, J., Nitta, A., Puri, R., Shah, M., Stojanovič, B., Bendias, R., Barghi, A., Kocijan, V., Zhang, Z., He, X., Lenssen, J. E., & Leskovec, J. (2025). PyG 2.0: Scalable Learning on Real World Graphs.

Gabaire, M., Ghomi, H., & Hussein, M. (2025). Investigating the contributing factors to autonomous Vehicle-Road user Conflicts: A Data-Driven approach. Accident Analysis & Prevention, 211, 107898.

Gao, X., Jiang, X., Zhuang, D., Chen, H., Wang, S., Law, S., & Haworth, J. (2023). Uncertainty-Aware Probabilistic Graph Neural Networks for Road-Level Traffic Accident Prediction.

Garcia, Katherine R., et al. "Human Perception of AI Capabilities at Classifying Perturbed Roadway Signs." IEEE Transactions on Human-Machine Systems (2025).

Gargoum, S., & El-Basyouny, K. (2017). Automated extraction of road features using LiDAR data: A review of LiDAR applications in transportation. 2017 4th International Conference on Transportation Information and Safety (ICTIS), 563–574.

German Aerospace Center (DLR), & SUMO Contributors. (2024a). Enabling Sender / Receiver devices. Retrieved 01/02/2025 from

German Aerospace Center (DLR), & SUMO Contributors. (2024b). SSM Device. Retrieved 01/02/2025 from

Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks.

Gui, S., Song, S., Qin, R., & Tang, Y. (2024). Remote Sensing Object Detection in the Deep Learning Era—A Review. Remote Sensing, 16(2), 327.

Guo, W., Xu, C., & Jin, S. (2024). Fusion of satellite and street view data for urban traffic accident hotspot identification. International Journal of Applied Earth Observation and Geoinformation, 130. Scopus.

Guo, Zhiyang, Xing Hu, Jiejia Wang, XiaoYu Miao, MengTeng Sun, HuaiWei Wang, and XueYing Ma. 2024. "A Duplex Transform Heterogeneous Feature Fusion Network for Road Segmentation." Scientific Reports 14(1):17438.

Hamim, O. F., & Ukkusuri, S. V. (2024). Towards safer streets: A framework for unveiling pedestrians' perceived road safety using street view imagery. Accident Analysis and Prevention, 195. Scopus.

Han, Xiaofeng, Huan Wang, Jianfeng Lu, and Chunxia Zhao. 2017. "Road Detection Based on the Fusion of Lidar and Image Data." International Journal of Advanced Robotic Systems

Harms, Louise, and Michelle Talbot. n.d. Harms and Talbot / The Aftermath of Road Trauma: Survivors' Perceptions of Trauma and Growth The Aftermath of Road Trauma: Survivors' Perceptions of Trauma and Growth. .

Hauer, E. (2004). Statistical Road Safety Modeling. Transportation Research Record, 1897(1), 81–87.

Haykin, S. (2009). Neural Networks and Learning Machines (3rd edn). Pearson Prentice Hall.

He, Songtao, Mohammad Amin Sadeghi, Sanjay Chawla, Mohammad Alizadeh, Hari Balakrishnan, and Samuel Madden. n.d. Inferring High-Resolution Traffic Accident Risk Maps Based on Satellite Imagery and GPS Trajectories.

He, X., & He, S. Y. (2024). Using mobile phone big data and street view images to explore the mismatch between walkability and walking behavior. Transportation Research Part A: Policy and Practice, 180. Scopus.

He, Z., Wang, Y.-H., & Zhang, J. (2023). Generative AIBIM: An automatic and intelligent structural design pipeline integrating BIM and generative AI.

Hosseini, R., Lim, S., Tong, D., Sohn, G., & Seyedabrishami, S. (2024). A specialized inclusive road dataset with elevation profiles for realistic pedestrian navigation using open geospatial data and deep learning. Computers, Environment and Urban Systems, 114. Scopus.

Hou, W., Lyu, N., Liu, Z., & Wang, X. (2020). Modelling large vehicles operating speed characteristics on freeway alignment based on aggregated GPS data. IET Intelligent Transport Systems, 14(8), 857–865.

Hu, C., Yang, W., Liu, C., Fang, R., Guo, Z., & Tian, B. (2023). An image-based crash risk prediction model using visual attention mapping and a deep convolutional neural network. Journal of Transportation Safety and Security, 15(1), 1–23. Scopus.

Hu, J., Huang, M.-C., & Yu, X. (2020). Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. Accident Analysis & Prevention, 144, 105665.

Hu, S., Xing, H., Luo, W., Wu, L., Xu, Y., Huang, W., Liu, W., & Li, T. (2023). Uncovering the association between traffic crashes and street-level built-environment features using street view images. International Journal of Geographical Information Science, 37(11), 2367–2391. Scopus.

Huang, B., Hooi, B., & Shu, K. (2023). TAP: A Comprehensive Data Repository for Traffic Accident Prediction in Road Networks.

Huang, G., Yu, Y., Lyu, M., Sun, D., Dewancker, B., & Gao, W. (2025). Impact of Physical Features on Visual Walkability Perception in Urban Commercial Streets by Using Street-View Images and Deep Learning. Buildings, 15(1). Scopus.

Huang, S., Shao, C., Li, J., Yang, X., Zhang, X., Qian, J., & Wang, S. (2020). Feature Extraction and Representation of Urban Road Networks Based on Travel Routes. Sustainability 2020, Vol. 12, Page 9621, 12(22).

Huang, Y., Song, Q., & Qiu, W. (2024). DO VISUALLY PERCEIVED DESIGN QUALITIES INFLUENCE DOCKLESS BIKESHARE CYCLING ROUTES?: A Case Study of Ithaca Using GPS Trajectories. 2, 99–108. Scopus.

Hydén, C. (1987). The development of a method for traffic safety evaluation: The Swedish traffic conflict technique. Lund: Lund University, Department of Traffic Planning and Engineering.

Iamtrakul, P., Chayphong, S., Kantavat, P., Nakamura, K., Hayashi, Y., Kijsirikul, B., & Iwahori, Y. (2024). Assessing Subjective and Objective Road Environment Perception in the Bangkok Metropolitan Region, Thailand: A Deep Learning Approach Utilizing Street Images. Sustainability (Switzerland), 16(4). Scopus.

Imprialou, M., & Quddus, M. (2019). Crash data quality for road safety research: Current state and future directions. Accident Analysis & Prevention, 130, 84–90.

International Road Assesment Programme (iRAP). 2019. IRAP Coding Manual Version 5.0—Drive on Right Edition.

J. Bruna, W. Zaremba, and Y. LeCun, "Spectral networks and deep locally connected networks on graphs," in Proc. Int. Conf. Learn. Represent. (ICLR), Banff, AB, Canada, Apr. 2014.

Jahn, Lennart Lorenz Freimuth, Seongjeong Park, Yongseob Lim, Jinung An, and Gyeungho Choi. 2024. “Enhancing Lane Detection with a Lightweight Collaborative Late Fusion Model.” Robotics and Autonomous Systems 175:104680.

Jain, A. G., & Saunier, N. (2020). Autocamera calibration for traffic surveillance cameras with wide-angle lenses. arXiv preprint arXiv:2001.07243.

Jain, N., & Mittal, S. (2022). A Computational Model for Driver Risk Evaluation and Crash Prediction Using Contextual Data from On-board Telematics. Recent Advances in Computer Science and Communications, 15(4), 620–635. Scopus.

Jalayer, M. P. (2022). A real-time proactive intersection safety monitoring system based on video data (Final report). Piscataway, NJ: Rutgers Center for Advanced Infrastructure and Transportation (CAIT).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 103). Springer.

Jeon, J., & Woo, A. (2024). The effects of built environments on bicycle accidents around bike-sharing program stations using street view images and deep learning techniques: The moderating role of streetscape features. Journal of Transport Geography, 121. Scopus.

Jiang, Y., Du, C., Wu, J., Nousias, S., & Borrmann, A. (2025). Towards automated BIM conflict resolution using reinforcement learning.

Johnson, Justin M., and Taghi M. Khoshgoftaar. 2019. “Survey on Deep Learning with Class Imbalance.” Journal of Big Data 6(1):1–54.

Kačan, Marin, Marin Oršić, Siniša Šegvić, and Marko Ševrović. 2020. “Multi-Task Learning for IRAP Attribute Classification and Road Safety Assessment.” Pp. 1–6 in 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE.

Kacan, Marin, Marko Sevrovic, and Sinisa Segvic. 2024. “Dynamic Loss Balancing and Sequential Enhancement for Road-Safety Assessment and Traffic Scene Classification.” IEEE Transactions on Intelligent Transportation Systems. https://doi:10.1109/TITS.2024.3456214 .

Kačan, Marin, Marko Ševrović, and Siniša Šegvić. 2024. “Dynamic Loss Balancing and Sequential Enhancement for Road-Safety Assessment and Traffic Scene Classification.” IEEE Transactions on Intelligent Transportation Systems.

Karim, M. M., Li, Y., Qin, R., & Yin, Z. (2022). A Dynamic Spatial-Temporal Attention Network for Early Anticipation of Traffic Accidents. IEEE Transactions on Intelligent Transportation Systems, 23(7), 9590–9600.

Karimi Monsefi, Amin, Pouya Shiri, Ahmad Mohammadshirazi, Nastaran Karimi Monsefi, Ron Davies, Sobhan Moosavi, and Rajiv Ramnath. 2023. “CrashFormer: A Multimodal Architecture to Predict the Risk of Crash.” Pp. 42–51 in Urban-AI 2023 - Proceedings of the

1st ACM SIGSPATIAL International Workshop on Advances in Urban-AI. Association for Computing Machinery, Inc.

Kchour, Fatima, Salvatore Cafiso, and Giuseppina Pappalardo. "Understanding Cyclists' Visual Behavior Using Eye-Tracking Technology: A Systematic Review." Sensors 25.1 (2024): 22.

Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks (No. arXiv:1609.02907). arXiv.

Kirushanth, S., & Kabaso, B. (2018). Telematics and Road Safety. 2018 2nd International Conference on Telematics and Future Generation Networks (TAFGEN), 103–108.

Ko, J., Ajibefun, J., & Yan, W. (2023). Experiments on Generative AI-Powered Parametric Modeling and BIM for Architectural Design.

Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021). Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. Traffic Injury Prevention, 22(6), 460–466.

Korea Institute of Civil Engineering and Building Technology. (2023). Newly Developed BIMbased Digital Design Workflow for Road Safety Improvement. Korea Institute of Civil Engineering and Building Technology (KICT). Retrieved 24/08/2025 from

Krajzewicz, D., Hertkorn, G., Feld, C., & Wagner, P. (2002). SUMO (Simulation of Urban MObility): An open-source traffic simulation. 4th Middle East Symposium on Simulation and Modelling, Sharjah (United Arab Emirates).

Kumar, P., Tiwari, G., & Paul, S. B. (2025). Road safety studies at micro, meso, and macroscopic levels: A systematic review. IATSS Research, 49(1), 10–26.

Kwon, J.-H., & Cho, G.-H. (2020). An examination of the intersection environment associated with perceived crash risk among school-aged children: Using street-level imagery and computer vision. Accident Analysis and Prevention, 146. Scopus.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Lee, S., Chen, T., Sze, N. N., Mao, T., Ou, Y., Mihaita, A.-S., & Chen, F. (2025). Analysing driver behaviour and crash frequency at railway level crossings using connected vehicle and GIS data. Travel Behaviour and Society, 39, 100957.

Li, J., Gao, X., & Jiang, T. (2020). Graph networks for multiple object tracking. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2020) (pp. 3560–3569).

Li, P., & Abdel-Aty, M. (2022). Real-Time Crash Likelihood Prediction Using Temporal Attention–Based Deep Learning and Trajectory Fusion. Journal of Transportation Engineering, Part A: Systems, 148(7), 04022043.

Li, P., Abdel-Aty, M., & Yuan, J. (2021). Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on GPS trajectory data. Accident Analysis & Prevention, 150, 105924.

Li, X., Mousavi, S. M., Dadashova, B., Lord, D., & Wolshon, B. (2021). Toward a crowdsourcing solution to identify high-risk highway segments through mining driving jerks. Accident Analysis & Prevention, 155, 106101.

Liao, W., Lu, X., Huang, Y., Zheng, Z., & Lin, Y. (2021). Automated structural design of shear wall residential buildings using generative adversarial networks. Automation in Construction, 132.

Lin, B., Saxe, S., & Chan, T. C. Y. (2024). AutoLTS: Automating Cycling Stress Assessment via Contrastive Learning and Spatial Post-processing. 38(20), 22222–22230. Scopus.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. "Focal Loss for Dense Object Detection." Pp. 2980–88 in Proceedings of the IEEE international conference on computer vision.

Liu, L., & Sevtsuk, A. (2024). Clarity or confusion: A review of computer vision street attributes in urban studies and planning. Cities, 150, 105022.

Liu, L., Gao, Z., Luo, P., Duan, W., Hu, M., Mohd Arif Zainol, M. R. R., & Zawawi, M. H. (2023). The Influence of Visual Landscapes on Road Traffic Safety: An Assessment Using Remote Sensing and Deep Learning. Remote Sensing, 15(18). Scopus.

Liu, Y., Chen, T., Chung, H., Jang, K., & Xu, P. (2025). Is there an emotional dimension to road safety? A spatial analysis for traffic crashes considering streetscape perception and built environment. Analytic Methods in Accident Research, 46. Scopus.

Loo, B. P. Y., & Anderson, T. K. (2015). Spatial Analysis Methods of Road Traffic Collisions. CRC Press.

Loshchilov, Ilya, and Frank Hutter. 2017. "Fixing Weight Decay Regularization in Adam." ArXiv Preprint ArXiv:1711.05101 5(5):5.

Lovegrove, G. R., & Sayed, T. (2006). Using Macrolevel Collision Prediction Models in Road Safety Planning Applications. Transportation Research Record, 1950(1).

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI '81) (Vol. 2, pp. 674–679). Morgan Kaufmann.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.

Ma, L., Li, Y., Li, J., Wang, C., Wang, R., & Chapman, M. A. (2018). Mobile Laser Scanned Point-Clouds for Road Object Detection and Extraction: A Review. Remote Sensing, 10(10), 1531.

Marshall, W. E., & Garrick, N. W. (2010). Street network types and road safety: A study of 24 California cities. URBAN DESIGN International 2010 15:3, 15(3).

Micko, Kristian, Peter Papcun, and Iveta Zolotova. 2023. “Review of IoT Sensor Systems Used for Monitoring the Road Infrastructure.” Sensors 23(9):4469.

Mimi, M. S., Islam, M. M., Tusti, A. G., Somvanshi, S., & Das, S. (2025). ST-GraphNet: A Spatio-Temporal Graph Neural Network for Understanding and Predicting Automated Vehicle Crash Severity.

Mishra, S., Rajendran, P. K., Vecchietti, L. F., & Har, D. (2023). Sensing Accident-Prone Features in Urban Scenes for Proactive Driving and Accident Prevention. IEEE Transactions on Intelligent Transportation Systems, 24(9), 9401–9414. Scopus.

Moeinaddini, M., Asadi-Shekari, Z., & Shah, M. Z. (2014). The relationship between urban street networks and the number of transport fatalities at the city level. Safety Science, 62. https://doi.org/10.1016/j.ssci.2013.08.015

Mooney, P., & Minghini, M. (2017). A Review of OpenStreetMap Data. In Mapping and the Citizen Sensor. Ubiquity Press.

Muhammad, K., Hussain, T., Ullah, H., Ser, J. D., Rezaei, M., Kumar, N., Hijji, M., Bellavista, P., & De Albuquerque, V. H. C. (2022). Vision-Based Semantic Segmentation in Scene Understanding for Autonomous Driving: Recent Achievements, Challenges, and Outlooks. IEEE Transactions on Intelligent Transportation Systems, 23(12), 22694–22715.

Murtagh, F., & Contreras, P. (2011). Methods of Hierarchical Clustering (No. arXiv:1105.0121). arXiv.

Nguyen, Q. C., Alirezaei, M., Yue, X., Mane, H., Li, D., Zhao, L., Nguyen, T. T., Patel, R., Yu, W., Hu, M., Quistberg, D. A., & Tasdizen, T. (2024). Leveraging computer vision for predicting collision risks: A cross-sectional analysis of 2019-2021 fatal collisions in the USA. Injury Prevention. Scopus.

Nie, Y., Zhang, M., Wang, B., Zhang, C., & Zhao, Y. (2025). Vehicle trajectory fractal theory for macro-level highway crash rate analysis. Accident Analysis & Prevention, 215, 107989.

Nikolaou, D., Ziakopoulos, A., Kontaxi, A., Theofilatos, A., & Yannis, G. (2025). Spatial analysis of telematics-based surrogate safety measures. Journal of Safety Research, 92, 98–108.

Nippani, A., Li, D., Ju, H., Koutsopoulos, H. N., & Zhang, H. R. (2023). Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis. https://doi.org/10.48550/arXiv.2311.00164

Nobin, P. R., & Rifat, I. A. (2025). STARN-GAT: A Multi-Modal Spatio-Temporal Graph Attention Network for Accident Severity Prediction.

Oord, A. van den, Li, Y., & Vinyals, O. (2019). Representation Learning with Contrastive Predictive Coding (No. arXiv:1807.03748). arXiv.

OpenStreetMap. (2025). About OpenStreetMap—OpenStreetMap Wiki.

OSeven. (2025). Oseven.io. Oseven.Io.

Outay, F., Mengash, H. A., & Adnan, M. (2020). Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. Transportation Research Part A: Policy and Practice, 141, 116–129.

Ozturk, Ozan, Mustafa Serkan Isik, Martin Kada, and Dursun Zafer Seker. 2023. “Improving Road Segmentation by Combining Satellite Images and LiDAR Data with a Feature-Wise Fusion Strategy.” Applied Sciences 13(10):6161.

Pajković, V., & Grdinić-Rakonjac, M. (2021). Road users’ attitudes and perception on selected road safety issues - Age-related comparison. European Transport - Trasporti Europei, 85.

Paleti, R., Sahin, O., & Cetin, M. (2017). Modeling the impact of latent driving patterns on traffic safety using mobile sensor data. Accident Analysis & Prevention, 107, 92–101.

Paliotto, A., Meocci, M., Terrosi, A., La Torre, F. (2024). Systematic review, evaluation and comparison of different approaches for the implementation of road network safety analysis. Heliyon, Volume 10 (7). https://doi.org/10.1016/j.heliyon.2024.e28391.

Pan, Y., Shen, Y., Qin, J., & Zhang, L. (2024). Deep reinforcement learning for multi-objective optimization in BIM-based green building design. Automation in Construction, 166.

Panduru, Krishna, and Joseph Walsh. 2025. “Exploring the Unseen: A Survey of Multi-Sensor Fusion and the Role of Explainable Ai (Xai) in Autonomous Vehicles.” Sensors (Basel, Switzerland) 25(3):856.

Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., & Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. Safety Science, 119, 91–97.

Pervaz, S., Bhowmik, T., & Eluru, N. (2023). An econometric framework for integrating aggregate and disaggregate level crash analysis. Analytic Methods in Accident Research, 39, 100280.

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. Data Science and Management, 4, 19–29.

Qin, J., Feng, Y., Sheng, Y., Huang, Y., Zhang, F., & Zhang, K. (2025). Evaluation of Pedestrian-Perceived Comfort on Urban Streets Using Multi-Source Data: A Case Study in Nanjing, China. ISPRS International Journal of Geo-Information, 14(2). Scopus.

Qu, Fangming, et al. "Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques." Journal of Big Data 11.1 (2024): 32.

Radanliev, P. (2024). Artificial intelligence: reflecting on the past and looking towards the next paradigm shift. Journal of Experimental & Theoretical Artificial Intelligence, 1-18.

Rasch, A., Moll, S., López, G., García, A., & Dozza, M. (2022). Drivers' and cyclists' safety perceptions in overtaking maneuvers. Transportation Research Part F: Traffic Psychology and Behaviour, 84.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016) (pp. 779–788). IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the 29th International Conference on Neural Information Processing Systems (NeurIPS 2015), Vol. 1 (pp. 91–99). MIT Press.

Riehl, K., Bregler, C., Hwang, J., Norrie, J., & Dey, D. (2025). Consistent vehicle trajectory extraction from aerial imagery. Scientific Reports, 15(1), 12301.

Rita, L., Peliteiro, M., Bostan, T.-C., Tamagusko, T., & Ferreira, A. (2023). Using Deep Learning and Google Street View Imagery to Assess and Improve Cyclist Safety in London. Sustainability (Switzerland), 15(13). Scopus.

Ryder, B., Dahlinger, A., Gahr, B., Zundritsch, P., Wortmann, F., & Fleisch, E. (2019). Spatial prediction of traffic accidents with critical driving events – Insights from a nationwide field study. Transportation Research Part A: Policy and Practice, 124, 611–626.

Samara, L. K. (2021). Video based network wide surrogate safety analysis to support Vision Zero policies. Transportation Research Board (TRB).

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The Graph Neural Network Model. IEEE Transactions on Neural Networks, 20(1), 61–80. IEEE Transactions on Neural Networks.

Shahapure, K. R., & Nicholas, C. (2020). Cluster Quality Analysis Using Silhouette Score. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 747–748.

Shelby, S. G. (2011). Delta V as a measure of traffic conflict severity. 90th TRB Annual Meeting. Transportation Research Board.

Shen, S., Zhang, F., Linwood, S. L., Lu, B., & Zhu, M. (2024). Using Aggregated Fine Geo-Resolution Vehicle Telemetric Data to Predict Crash Occurrence. Transportation Research Record, 2678(7), 835–844.

Shen, X., Sun, D., Pan, S., Zhou, X., & Yang, L. T. (2023). Neighbor Contrastive Learning on Learnable Graph Augmentation (No. arXiv:2301.01404). arXiv.

Silva, P. B., Andrade, M., & Ferreira, S. (2020). Machine learning applied to road safety modeling: A systematic literature review. Journal of Traffic and Transportation Engineering (English Edition), 7(6), 775–790.

Sochor, J., Juránek, R., & Herout, A. (2017). Traffic surveillance camera calibration by 3D model bounding box alignment for accurate vehicle speed measurement. arXiv preprint arXiv:1702.06451.

Sohail, A., Cheema, M. A., Ali, M. E., Toosi, A. N., & Rakha, H. A. (2023). Data-driven approaches for road safety: A comprehensive systematic literature review. Safety Science, 158, 105949.

STATBEL. (2024). Géolocalisation des accidents de la circulation 2017-2024

Steinley, Douglas. (2006). K-means clustering: A half-century synthesis. British Journal of Mathematical and Statistical Psychology, 59(1), 1–34.

Stiles, J., Li, Y., & Miller, H. J. (2022). How does street space influence crash frequency? An analysis using segmented street view imagery. Environment and Planning B: Urban Analytics and City Science, 49(9), 2467–2483. Scopus.

Stipancic, J., Miranda-Moreno ,Luis, Labbe ,Aurélie, & and Saunier, N. (2019). Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected GPS data. Transportation Letters, 11(7), 391–401.

Stipancic, J., Miranda-Moreno, L., Saunier, N., & Labbe, A. (2018). Surrogate safety and network screening: Modelling crash frequency using GPS travel data and latent Gaussian Spatial Models. Accident Analysis & Prevention, 120, 174–187.

Stojanovic, Nadica, Bojana Boskovic, Miroslav Petrovic, Ivan Grujic, and Oday I. Abdullah. 2023. “The Impact of Accidents during the Transport of Dangerous Good, on People, the Environment, and Infrastructure and Measures for Their Reduction: A Review.” Environmental Science and Pollution Research 30(12):32288–300.

Stokes, C., Raftery, S., & Woolley, J. (2019). Road user perception of safety at Safe System intersections. Journal of Road Safety, 30(2).

Tan, Shuyi, Yi Zhang, Yan Li, and Byeong-Seok Shin. 2025. “CrossModalSync: Joint Temporal-Spatial Fusion for Semantic Scene Segmentation in Large-Scale Scenes.” Scientific Reports 15(1):25362.

Thakur, Abhishek, and Sudhansu Kumar Mishra. 2024. “An In-Depth Evaluation of Deep Learning-Enabled Adaptive Approaches for Detecting Obstacles Using Sensor-Fused Data in Autonomous Vehicles.” Engineering Applications of Artificial Intelligence 133:108550.

Toohey, G., Nguyen, T. B., Vilppola, R., Qiu, W., Li, W., & Luo, D. (2022). DATA-DRIVEN EVALUATION OF STREETS TO PLAN FOR BICYCLE FRIENDLY ENVIRONMENTS: A CASE STUDY OF BRISBANE SUBURBS. 243–252. Scopus.

Towers, M., Kwiatkowski, A., Terry, J., Balis, J. U., De Cola, G., Deleu, T., Goulão, M., Kallinteris, A., Krimmel, M., KG, A., Perez-Vicente, R., Pierré, A., Schulhoff, S., Tai, J. J., Tan, H., & Younis, O. G. (2024). Gymnasium: A Standard Interface for Reinforcement Learning Environments.

Tu, Zhengzhong, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. 2022. “Maxvit: Multi-Axis Vision Transformer.” Pp. 459–79 in European conference on computer vision. Springer.

United Nations. Safe Roads for a Sustainable World. n.d. .

Ural, S., J. Shan, M. A. Romero, and A. Tarko. 2015. “Road and Roadside Feature Extraction Using Imagery and LiDAR Data for Transportation Operation.” ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2:239–46.

Useche, S. A., Faus, M., & Alonso, F. (2022). Is safety in the eye of the beholder? Discrepancies between self-reported and proxied data on road safety behaviors—A systematic review. In Frontiers in Psychology (Vol. 13).

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks (No. arXiv:1710.10903). arXiv.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2017). Graph Attention Networks.

Wang, J., Kong, Y., & Fu, T. (2019). Expressway crash risk prediction using back propagation neural network: A brief investigation on safety resilience. Accident Analysis & Prevention, 124, 180–192.

Wang, M., Chen, Z., Rong, H. H., Mu, L., Zhu, P., & Shi, Z. (2022). Ridesharing accessibility from the human eye: Spatial modeling of built environment with street-level images. Computers, Environment and Urban Systems, 97. Scopus.

Wang, S. and Cao, J. (2021). Ai and deep learning for urban computing. Urban Informatics, pages

815–844.

Wang, Y., Liu, D., & Luo, J. (2022). Identification and Improvement of Hazard Scenarios in Non-Motorized Transportation Using Multiple Deep Learning and Street View Images. International Journal of Environmental Research and Public Health, 19(21). Scopus.

Wang, Y., Wang, X., Wang, T., & Quddus, M. (2025). Freeway real-time crash prediction using floating car data. Transportation Research Part C: Emerging Technologies, 171, 105009.

Weng, X., Wang, Y., Man, Y., & Kitani, K. (2020). GNN3DMOT: Graph neural network for 3D multi-object tracking with multi-feature learning. arXiv preprint arXiv:2006.07327.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 3645–3649). IEEE.

World Health Organization. 2023. Global Status Report on Road Safety 2023. World Health Organization.

Xiang, J., Ghaffarpasand, O., & Pope, F. D. (2024). Mapping urban mobility using vehicle telematics to understand driving behaviour. Scientific Reports, 14(1), 3271.

Xie, K., Yang, D., Ozbay, K., & Yang, H. (2019). Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. Accident Analysis & Prevention, 125, 311–319.

Xiong, X., Chen, L., & Liang, J. (2018). A New Framework of Vehicle Collision Prediction by Combining SVM and HMM. IEEE Transactions on Intelligent Transportation Systems, 19(3), 699–710.

Yap, W., Chang, J.-H., & Biljecki, F. (2023). Incorporating networks in semantic understanding of streetscapes: Contextualising active mobility decisions. Environment and Planning B: Urban Analytics and City Science, 50(6), 1416–1437. Scopus.

Ye Dong, and Jonathan S. Wood. 2025. “Evaluation of Crash Contributing Factors.” Journal of Transportation Technologies 15(1):155–78. https://doi:10.4236/jtts.2025.151009 .

Ye, X., Li, S., Gong, W., Li, X., Li, X., Dadashova, B., Li, W., Du, J., & Wu, D. (2025). Street View Imagery in Traffic Crash and Road Safety Analysis: A Review. Applied Spatial Analysis and Policy, 18(2). Scopus.

Yue, H. (2024). Investigating the influence of streetscape environmental characteristics on pedestrian crashes at intersections using street view images and explainable machine learning. Accident Analysis and Prevention, 205. Scopus.

Yue, H. (2025). Investigating streetscape environmental characteristics associated with road traffic crashes using street view imagery and computer vision. Accident Analysis and Prevention, 210. Scopus.

Yu, X., Ma, J., Tang, Y., Yang, T., & Jiang, F. (2024). Can we trust our eyes? Interpreting the misperception of road safety from street view images and deep learning. Accident Analysis and Prevention, 197.

Zdravković, Saša, et al. "AI-Driven Safety Evaluation in Public Transport: A Case Study from Belgrade's Closed Transit Systems." Sustainability 17.18 (2025): 8283.

Zeng, Q., Huang, H., Pei, X., & Wong, S. C. (2016). Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. Analytic Methods in Accident Research, 10, 12–25.

Zhai, X., Sze, N. N., Lee, J. J., Xu, P., & Huang, H. (2025). Multi-scale approaches to cope with scale effect issues in macroscopic crash analysis. Accident Analysis & Prevention, 214, 107971.

Zhang, K., Tamakloe, R., Cao, M., & Kim, I. (2024). Exploring fatal/severe pedestrian injury crash frequency at school zone crash hotspots: Using interpretable machine learning to assess the micro-level street environment. Journal of Transport Geography, 121. Scopus.

Zhang, S., & Abdel-Aty, M. (2022). Real-time crash potential prediction on freeways using connected vehicle data. Analytic Methods in Accident Research, 36, 100239.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., ... & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. arXiv preprint arXiv:2110.06864.

Zhang, Y., Wang, C., Wang, X., Zeng, W., & Liu, W. (2021). FairMOT: On the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision, 129(11), 3069–3087.

Zhang, Z., Guo, Z., Zheng, H., Li, Z., & Yuan, P. F. (2024). Automated architectural spatial composition via multi-agent deep reinforcement learning for building renovation. Automation in Construction, 167. https://doi.org/10.1016/j.autcon.2024.105702

Zheng, L. I. (2014). Traffic conflict techniques for road safety analysis: Open questions and some insights. Canadian Journal of Civil Engineering, 633–641.

Zheng, L., & Sayed, T. (2020). A novel approach for real time crash prediction at signalized intersections. Transportation Research Part C: Emerging Technologies, 117. Scopus.

Zhou, W., Liu, Y., Zhao, L., Xu, S., & Wang, C. (2023). Pedestrian crossing intention prediction from surveillance videos for over-the-horizon safety warning. IEEE Transactions on Intelligent Transportation Systems, 25(2), 1394-1407.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ADE20K dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 5122–5130). IEEE.

Zhou, Y., Guo, X., Hou, F., & Wu, J. (2022). Review of intelligent road defects detection technology. Sustainability, 14(10), 6306.

Zhu, M., Li, H., Sze, N. N., & Ren, G. (2022). Exploring the impacts of street layout on the frequency of pedestrian crashes: A micro-level study. Journal of Safety Research, 81, 91–100. Scopus.

Zia, H., Hassan, I. U., Khurram, M., Harris, N., Shah, F., & Imran, N. (2025). Advancing Road Safety: A Comprehensive Evaluation of Object Detection Models for Commercial Driver Monitoring Systems. Future Transportation, 5(1), 2.

Ziakopoulos, A. (2024). Analysis of harsh braking and harsh acceleration occurrence via explainable imbalanced machine learning using high-resolution smartphone telematics and traffic data. Accident Analysis and Prevention, 207. Scopus.

Ziakopoulos, A., & Yannis, G. (2024). Key Artificial Intelligence and Digitalization Solutions Towards Vision Zero in Road Safety. In A. Sładkowski (Ed.), Using Artificial Intelligence to Solve Transportation Problems (pp. 1–26). Springer Nature Switzerland.

Ziakopoulos, A., Petraki, V., Kontaxi, A., & Yannis, G. (2022). The transformation of the insurance industry and road safety by driver safety behaviour telematics. Case Studies on Transport Policy, 10(4), 2271–2279.

Ziakopoulos, A., Vlahogianni, E., Antoniou, C., & Yannis, G. (2022). Spatial predictions of harsh driving events using statistical and machine learning methods. Safety Science, 150. Scopus.

Ziakopoulos, A. (2021). Spatial analysis of harsh driving behavior events in urban networks using high-resolution smartphone and geometric data. Accident Analysis & Prevention, 157, 106189.

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. Accident Analysis & Prevention, 135, 105323.