Assessing Vehicle Collision Prevention based on Machine Learning and V2X Communication

Andreia A. Felix*, Joahannes B. D. da Costa†, Helder M. N. da S. Oliveira‡
*Federal University of ABC (UFABC), Santo André, Brazil
†Federal University of São Paulo (UNIFESP), São José dos Campos, Brazil
‡Universidade de São Paulo (USP), São Paulo, Brazil
Contact: andreia.alexandre@ufabc.edu.br, https://joahannes.github.io, helderoliveira@ime.usp.br

***Abstract*—Road safety has become increasingly efficient by incorporating technologies such as Vehicle-to-Everything (V2X) communication and Machine Learning (ML) algorithms. These solutions enable real-time data exchange between vehicles and infrastructure, helping predict and prevent accidents. The latency in transmitting this information is a critical factor that impacts the system's effectiveness. In this context, this work investigates vehicle collision prevention and uses simulations with different collision scenarios to train predictive ML models. As a result, trained models were capable of predicting collisions up to five seconds in advance and accurately classifying different risk scenarios. This demonstrates the potential of these technologies to enhance safety and efficiency in traffic.**

***Index Terms*—VANETs, Collision prevention, Machine learning**

I. Introduction

Despite significant advances in the application of Artificial Intelligence (AI) and Vehicle-to-Everything (V2X) communication for collision detection and prevention, current models still face notable challenges [1]. Numerous approaches demonstrate limited adaptability to highly dynamic and nonlinear scenarios and frequently depend on large volumes of centralized data, thereby compromising scalability and applicability in real-world traffic environments [2]. These restrictions directly affect the effectiveness of collision prevention applications and hinder vehicles' ability to make timely decisions.

Moreover, in Vehicular Networks (VANETs), node mobility imposes significant challenges to message transmission and reception. VANET operates in highly dynamic environments, where network topology constantly changes due to the high-speed and multidirectional movement of vehicles [3], [4]. High mobility directly affects the stability of communication links. Mobility also introduces issues related to time synchronization and spatial alignment, which are critical for safety applications such as collision warnings and emergency braking. The mobility of nodes negatively affects the reliability, latency, link stability, and efficiency of communication protocols. This highlights the need for robust and adaptive solutions to ensure effective information dissemination in highly dynamic environments [5].

To address these limitations, several studies aim to apply Machine Learning (ML) techniques to predict adverse road conditions and make decisions based on this information [6]. A significant portion of existing research relies on limited or unrealistic datasets, which hinders the generalization of models to diverse traffic patterns, weather conditions, and infrastructure scenarios. Even in simulated environments, experiments often fail to replicate complex situations involving multiple vehicle interactions accurately. This highlights the need for more robust and adaptable solutions validated in contexts that closely resemble real-world traffic conditions.

In this context, this work presents an approach based on ML algorithms and V2X communication to predict vehicle collisions. The proposed method addresses the limitations of previous models by analyzing interactions between vehicles in realistic urban scenarios, allowing rapid responses and reducing risks to road safety. The main innovation lies in integrating predictive modeling, traffic simulations incorporating multiple dynamic variables, and systematic validation of the results. Based on a realistic mobility trace, the simulation results highlight the benefits of the proposed approach and demonstrate its effectiveness in predicting future behavior and identifying risk situations. The model achieved a coefficient of determination (R^2) of 93.96% for predicting vehicle behavior 5 seconds in advance, indicating high accuracy in trajectory estimation. Furthermore, the accuracy in classifying situations as risky or not reached 97.31%, reinforcing the system's ability to anticipate critical events reliably. The average processing time per prediction was only 1.09 seconds, underscoring the solution's potential for near real-time applications in dynamic urban scenarios.

In summary, the contributions of this work are:

- an approach that integrates ML algorithms and V2X communication to predict vehicle collisions in highly dynamic and nonlinear traffic scenarios;
- validation in realistic urban environments with mobility traces and dynamic simulations; and
- the achievement of high accuracy in trajectory prediction and risk classification, with low processing time, enabling near real-time application in vehicular networks.

The remainder of this paper is organized as follows. Section II presents the main studies related to the use of ML and V2X communication in collision prevention. Section III describes the methodology for modeling and simulating the scenarios. Section IV discusses the results obtained with the proposed models. Finally, Section V concludes the paper with the findings, study limitations, and future research directions.

102nd IEEE Vehicular Technology Conference (VTC2025-Fall)

II. Related Work

This section reviews recent research on vehicle collision prevention, highlighting the use of ML and V2X communication. Several studies explore predictive models to enhance traffic safety by analyzing sensor data and smart infrastructure using ML to anticipate risks in VANETs. For example, Parada *et al.* [7] proposed a collision prevention system between vehicles and vulnerable road users using 5G networks, Deep Learning (DL), and a Monte Carlo-based probability estimation algorithm. The method predicts vehicle trajectories using neural networks and calculates the probability of collision based on random samples of these trajectories.

Alagarsamy *et al.* [8] presented a study in which they analyze the main causes of accidents on urban highways through the use of ML algorithms, specifically Reinforcement Learning (RL) and Random Forest (RF), to identify factors associated with collisions. This approach aims to understand accident patterns better and classify them more precisely. However, the authors do not consider dynamic and nonlinear scenarios, which limits the applicability of their study.

In the same direction, Veluchamy *et al.* [9] presented a braking decision-making system in Advanced Driver Assistance Systems (ADAS). The study combines Generative Adversarial Networks (GANs) and Deep Convolutional Neural Networks (DeepCNNs) to process video captured by cameras, extracting relevant information for automated decision-making. However, the training of the models occurs using images that are compromised by different weather conditions on highways.

In Ribeiro *et al.* [10], a model based on V2X communication and Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, was developed to predict collisions involving vulnerable road users, such as motorcyclists. The study demonstrated that V2X communication can enhance collision detection, enabling more accurate predictions in situations where conventional sensors face limitations, such as line-of-sight obstructions. However, the applicability of the model is limited in certain contexts due to unmodeled factors such as complex urban traffic, diverse vehicle interactions, dynamic mobility, and city-specific conditions.

Farhat *et al.* [11] propose a collaborative collision avoidance system which integrates Mobile Edge Computing (MEC), V2X communication, and a DL model (YOLOv5) for visual vehicle detection and risk assessment. The system leverages onboard cameras and MEC servers deployed in Road Side Units (RSUs) to continuously monitor traffic, predict potential collisions, and issue alerts or trigger emergency braking.

In contrast to previous studies, this work offers a more robust, adaptive, and context-aware approach to collision risk classification. Although many existing models depend heavily on infrastructure (*e.g.*, MEC servers) and visual sensors (*e.g.*, LIDAR and cameras), which are costly to deploy, susceptible to environmental conditions or constrained by infrastructure availability, our approach relies solely on data generated internally by the vehicle, such as speed, direction, and position, in addition to its V2X communication capability.

III. System Overview

This section describes the methodology used to evaluate the ML techniques for vehicle collision prevention.

A. System model

The system model employed in this work is composed of vehicles, communication infrastructures (*e.g.*, RSUs), and a remote server in the Internet cloud. In this scenario, vehicles move around the city and can communicate with the RSUs. Each RSU r_i, denoted as $r_i \in R = \{r_1, r_2, \ldots, r_m\}$, has its coverage area in meters, can collect data from all vehicles within its coverage, and has wired communication with the remote server. Vehicles periodically send information to the RSU, including: position (x, y, z), speed, acceleration, direction, distance (distance to all surrounding vehicles, calculated based on the beacons received from those vehicles via V2V communication), and TTC (time to collision, computed using the collision documentation provided by SUMO simulator[1]). The city is divided into $|R|$ regions, where $|R| = m$ and represents the number of RSUs present in the scenario. In this case, we consider $m = 9$ and the placement of the RSUs ensures full coverage of the scenario.

B. Communication model

The proposed vehicular communication model, or message exchange model, is presented in Figure 1. Each vehicle $v_i \in V$ maintains and reads its own set of historical data locally. Using this information, each vehicle estimates its future state over a horizon of five time steps $t + 5$, predicting variables such as position, speed, and acceleration. This process allows anticipating risky situations and generating messages containing its predicted state $S_i(t+5)$, which are immediately transmitted to the RSU via Vehicle-to-Infrastructure (V2I) communication.

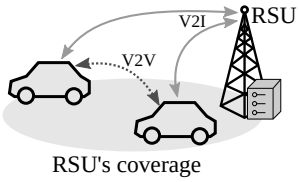


Fig. 1. Message exchange model considered in the study.

Upon receiving a set of messages $\{S_1, S_2, \ldots, S_n\}$ from the vehicles $v_i \in V$, the RSU individually analyzes each situation. If the received message indicates that the future state of the vehicle is classified as $situation(v_i) = SAFE$, the RSU only records the information in its local data repository. On the other hand, if the message indicates a $situation(v_i) = RISK$, the RSU generates an alert message and sends it to all vehicles within its coverage area. This decision ensures that nearby vehicles are informed and can react preventively.

In more critical scenarios, when the RSU identifies a predicted $situation(v_i) = COLLISION$, an emergency

[1]http://sumo.dlr.de/docs/Simulation/Output/Collisions.html

protocol is triggered. The RSU generates and transmits a high-priority message using dedicated high-priority communication channels, thus ensuring rapid dissemination of the information to surrounding vehicles. This hierarchical response process, based on the classification of the vehicles' future states, enables efficient action by the intelligent road infrastructure, increasing safety and reducing drivers' reaction time.

C. Prediction Time Window study

Given that the average speeds of the vehicles were recorded during the simulation, a quantitative analysis was conducted to determine the minimum stopping time required in hazardous situations based on each vehicle's respective speed. Table I presents the values observed in each RSU.

TABLE I
AVERAGE VEHICLE SPEED NEAR THE RSU, WHERE THE FIRST LINE INDICATES THE RSU IDENTIFIER, AND THE SECOND LINE SHOWS THE CORRESPONDING AVERAGE SPEED RECORDED.

#1	#2	#3	#4	#5	#6	#7	#8	#9
9,17	6,50	9,83	9,22	11,33	12,11	10,06	15,33	16.94

Based on the maximum speeds observed during the simulations, it is possible to estimate the time required for a vehicle to come to a complete stop, assuming constant deceleration [12]. To estimate the braking time, the following kinematic equation was considered:

$$t = \frac{v}{a} \tag{1}$$

where t represents the time required for the vehicle to come to a complete stop, v is the initial speed (in meters/second), and a corresponds to the average deceleration (in meters/second squared). Considering a deceleration of $6\,\text{m/s}^2$ [13], we have:

$$t = \frac{16{,}94}{6} \approx 2{,}82\,\text{seconds} \tag{2}$$

Thus, it is estimated that a vehicle traveling at the maximum speed observed in the simulations would require approximately $2{,}82$ seconds to come to a complete stop. For the system to effectively prevent collisions in connected vehicular environments, it is crucial that vehicle behavior predictions are made with adequate anticipation. This anticipation must consider not only the stopping time of the vehicle but also any delays introduced by communication between network nodes.

In this context, Equation (3) is designed to compute the total time required to perform predictions in cooperative vehicular systems, considering the main factors that influence the overall latency of the process.

$$f(x) = t_{v\to\text{RSU}} + t_{\text{RSU}\to v} + \left(\frac{v}{a}\right) + (j_{v\to\text{RSU}} + j_{\text{RSU}\to v}) + t_p \tag{3}$$

The function $f(x)$ represents the total system response time based on the input variable x. The terms $t_{v\to\text{RSU}}$ and $t_{\text{RSU}\to v}$ denote the communication times between the vehicle and the RSU in both directions, including transmission, propagation, and queuing delays. The expression $\left(\frac{v}{a}\right)$ corresponds to the total braking time, calculated as the ratio of the vehicle's speed v to its deceleration a; for simulation purposes, an average of 6 seconds per meter is assumed in controlled environments. The term $j_{v\to\text{RSU}} + j_{\text{RSU}\to v}$ captures the average communication jitter, referring to the statistical variation in packet transmission delays, which can impact stability in dynamic vehicular networks. Finally, t_p denotes the model's prediction processing time, *i.e.*, the duration required for a machine learning model or inference system to generate a response based on the input variable x.

Algorithm 1: Vehicle Risk Prediction

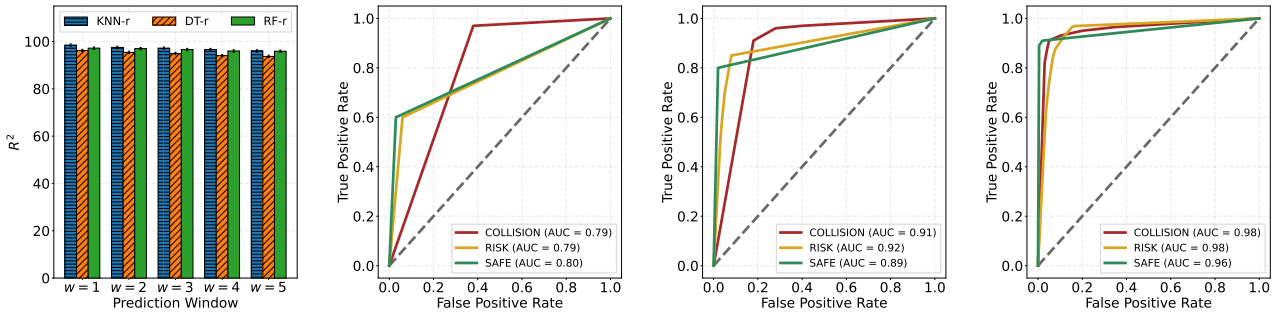
Input: Pre-trained input data
Output: Vehicle situation $situation$ for $i \in V$

```
/* Prediction windows                                          */
w ← [1, 2, 3, 4, 5]
foreach vehicle i ∈ V do
    Prediction()
    Communication()

Function Prediction():
    /* Regression Models                                       */
    foreach model r ∈ R that predicts vehicle behavior do
        Run the models with input data for each w
        Calculate performance metrics
    Select r ∈ R with the best metrics
    /* Classification Models                                   */
    foreach model c ∈ C that classifies vehicle situation do
        Run the models with input data for each w
        Calculate performance metrics
    Select c ∈ C with the best metrics
    situation ← Combine results from R and C for each w
    return situation for each w

Function Communication():
    if situation != SAFE then
        Send alert message to RSU
    else
        Send update message to RSU
```

D. Applying Machine Learning algorithms

After data collection, performed through real-time communication between vehicles and RSU, well-known ML algorithms for regression and classification were used to predict and calculate collision situations between vehicles that share their mobility information. Thus, the regression and classification versions of the following algorithms were considered: K-Nearest Neighbors (KNN), Random Forest (RF), and Decision Tree (DT). In this study, the regression versions of the algorithms are referred to as KNN-r, RF-r, and DT-r. Additionally, the classification versions of these algorithms are referred to as KNN-c, RF-c, and DT-c.

Algorithm 1 describes the risk prediction and communication system that is executed on each vehicle in the scenario. Initially, it receives mobility data from the vehicles. With this data, the system trains models to predict collision risks. It operates in two main stages: the first for prediction and the second for communication, called `Prediction` and `Communication`, respectively. First, vehicles continuously perform the prediction



(a) Regression results (b) DT-c (c) KNN-c (d) RF-c

Fig. 2. Simulation results considering different Machine Learning (ML) models and prediction time window.

process locally (Line 3). In the prediction stage, regression models are used to predict the vehicle's behavior in different observation windows (w) (Lines 6 and 8). After this prediction, classification models are applied to identify the situations the vehicle will encounter in each future w (Lines 10 and 12). The vehicle's situation is determined by combining these results, which can be *SAFE*, *RISK*, or *COLLISION* (Line 14). Finally, the vehicle sends a data message to the nearest RSU (Line 4), which can either be an alert message in the case of a *RISK* or *COLLISION* situation, or an update message if its situation is *SAFE* (Lines 17 and 20).

IV. Performance Evaluation

This section presents details of the performance evaluation carried out with the ML models in the context of vehicle accident detection and prevention.

A. Scenario description and Methodology

The simulation platform to evaluate the performance of the designed mechanism is composed of the Simulation of Urban Mobility (SUMO) 1.18.0, the network simulator OMNeT++ 6.0.0, and the vehicular networking framework Veins 5.2 [14], which implements the IEEE 802.11p protocol stack for V2X communication and signal attenuation. The ML algorithms were implemented in Python 3.13.4. We used a central sub-map of 114 km^2 from TAPASCologne trace[2], which reproduces vehicle traffic in the city of Cologne, Germany. We consider 2 hours of vehicular mobility and up to 700 vehicles. The simulation time was 800 seconds, with 100 initial seconds of warm-up. We ran the simulations 33 times to obtain a 95% confidence interval.

The metrics used for the evaluation were: *(1) R² (Coefficient of Determination*, for measuring the proportion of variance explained by the regression models; *(2) AUC (Area Under the ROC Curve)*, for assessing the discriminative ability of classifiers across multiple classes; and *(3) Prediction Time*, to measure the computational cost and real-time feasibility.

Figure 3 shows the spatial distribution of RSUs in the simulation scenario. Each RSU has a communication coverage area of 2600 meters, allowing it to reach a wide urban zone and communicate with multiple vehicles simultaneously.

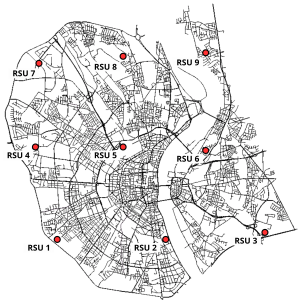


Fig. 3. Cologne sub-map and Road Side Unit (RSU) deployed in the city.

B. Simulation Results

Figure 2(a) presents the results for the coefficient of determination (R^2), expressed as a percentage, obtained by the regression models KNN-r, DT-r, and RF-r across different prediction windows ($w = \{1, 2, 3, 4, 5\}$). The R^2 value indicates the proportion of variance in the data explained by the model, with values closer to 100% reflecting higher predictive accuracy. The results show that for the $w = 1$ window, all models achieved excellent performance, with the KNN-r model reaching the highest R^2 value of 97.63%, followed by RF-r with 97.24%, and DT-r with 96.29%. As the value of w increases, a decreasing trend in the R^2 values is observed (an expected behavior, as longer-term predictions are generally more challenging to perform accurately). Despite this reduction, performance remains high across all windows. The KNN-r model remains the most effective across all prediction windows, still achieving an R^2 of 93.61% at $w = 5$. The RF-r model also shows consistent performance and outperforms the DT-r model, confirming the benefits of ensemble methods like Random Forest over individual decision trees. DT-r exhibits the lowest R^2 values across all windows but still maintains results above 91%, which is considered satisfactory. Overall, these results confirm the models' ability to accurately predict vehicular behavior, especially in short-term windows. Even with the expected performance drop as the prediction horizon increases,

[2]https://sumo.dlr.de/docs/Data/Scenarios.html

all models maintain R^2 values above 90%, highlighting the robustness of the proposed approaches.

Figures 2(b) to 2(d) present the ROC curves for the classification models DT-c, KNN-c, and RF-c, respectively, considering the prediction window $w = 5$ (most challenging scenario) and the three evaluated classes: *COLLISION*, *RISK*, and *SAFE*. The DT-c model exhibited a noticeable decline in performance, with AUC values dropping to 80% for the *SAFE* class and 79% for the *COLLISION* class, indicating higher sensitivity to increased prediction horizons and reduced generalization capability. In contrast, the KNN-c model maintained strong discriminative performance, with AUC values above 89% for all classes at $w = 5$, confirming its robustness even for longer-term predictions. The RF-c model delivered the best results overall, with AUC values above 96% across all classes, demonstrating high accuracy and strong resilience to longer prediction windows. The RF-c model shows the best predictive performance, followed by KNN-c, while DT-c performs worst in longer prediction scenarios.

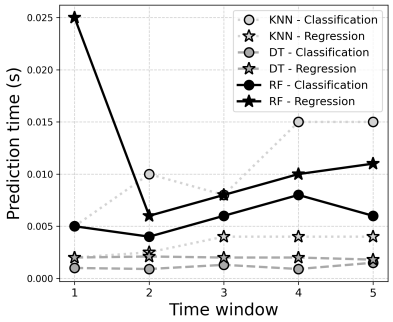


Fig. 4. Model prediction time results.

Figure 4 shows the prediction time of the models in both regression and classification scenarios, considering different values of $w = \{1, 2, 3, 4, 5\}$. This analysis enables the evaluation of the computational cost associated with each approach, highlighting performance differences among the tested models. The KNN-r and KNN-c models exhibit distinct behaviors. While the KNN-c model shows a significant increase in prediction time as w grows (suggesting a higher computational cost due to nearest neighbor searches in high-dimensional spaces) the KNN-r model remains more efficient, maintaining a stable prediction time across all values of w. The DT-r and DT-c models show smaller variations between regression and classification. The regression model maintains consistent performance, and the classification model exhibits only minor fluctuations, indicating that the hierarchical structure of decision trees can handle both tasks with minimal impact on inference time. In the case of RF, the RF-r model starts with a relatively high prediction time, which stabilizes as w increases. The RF-c model demonstrates more uniform performance across all w values, suggesting that the ensemble structure of RF handles classification efficiently and maintains predictable inference times. In general, the results indicate that regression models generally offer faster prediction times, while classification models tend to incur higher computational costs, particularly for algorithms like KNN.

V. Conclusion

This study demonstrated the effectiveness of Machine Learning (ML) models in predicting vehicle collisions, highlighting the potential of these approaches in enhancing road safety. The Random Forest (RF)-classifier model outperformed in all time windows analyzed, while the K-Nearest Neighbors (KNN)-regressor model showed superior results in the first three prediction steps, later surpassed by the RF-regressor in the last two. Future work will explore the possibility of in-vehicle edge training, as well as compare different distribution strategies using a federated learning model that exchanges only model weights. Additionally, Vehicle-to-Vehicle (V2V) communication will be investigated with mechanisms to ensure both data dissemination and privacy.

References

[1] M. M. Saad, M. A. Tariq, M. Ajmal, D. Kim, and G. Srivastava, "Federated multi-agent reinforcement learning for resource allocation in nr-v2x mode 2," *IEEE Internet of Things Journal*, 2025.

[2] A. Comi, O. Hriekova, and M. Nigro, "Exploring road safety in the era of micro-mobility: Evidence from rome," *Transportation research procedia*, vol. 78, pp. 55–62, 2024.

[3] J. B. da Costa, A. M. de Souza, R. I. Meneguette, E. Cerqueira, D. Rosário, C. Sommer, and L. Villas, "Mobility and deadline-aware task scheduling mechanism for vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, 2023.

[4] Y. Li, L. Li, and P. Fan, "Mobility-aware computation offloading and resource allocation for noma mec in vehicular networks," *IEEE Transactions on Vehicular Technology*, 2024.

[5] Z. Jin, T. Song, and W.-K. Jia, "An adaptive cooperative caching strategy for vehicular networks," *IEEE Transactions on Mobile Computing*, 2024.

[6] M. Hamidaoui *et al.*, "Survey of autonomous vehicles' collision avoidance algorithms," *Sensors (Basel, Switzerland)*, vol. 25, no. 2, p. 395, 2025.

[7] R. Parada, R. Corvillo, and P. Dini, "A dl-based estimation probability approach for vru collision avoidance," in *2023 IEEE Conference on Artificial Intelligence (CAI)*, IEEE, 2023, pp. 23–25.

[8] S. Alagarsamy, P. Nagaraj, B. Srikanth, C. V. Krishna, G. Bharath, and S. S. Kalyan, "A novel machine learning technique for predicting road accidents," in *2023 III International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, IEEE, 2023, pp. 1547–1551.

[9] S. Veluchamy, K. M. Mahesh, P. T. Sheeba, *et al.*, "Deepdrive: A braking decision making approach using optimized gan and deep cnn for advanced driver assistance systems," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106 111, 2023.

[10] B. Ribeiro, M. J. Nicolau, and A. Santos, "Using machine learning on v2x communications data for vru collision prediction," *Sensors*, vol. 23, no. 3, p. 1260, 2023.

[11] W. Farhat, O. Ben Rhaiem, H. Faiedh, and C. Souani, "A novel cooperative collision avoidance system for vehicular communication based on deep learning," *International Journal of Information Technology*, vol. 16, no. 3, pp. 1661–1675, 2024.

[12] S. A. Munaaf, B. Vasudevan, and S. Routray, "Regenerative braking solutions for electric vehicles: Advancing efficiency and energy recapture," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2024, pp. 331–337.

[13] K. Mattas, G. Albano, R. Donà, M. C. Galassi, R. Suarez-Bertoa, S. Vass, and B. Ciuffo, "Driver models for the definition of safety requirements of automated vehicles in international regulations. application to motorway driving conditions," *Accident Analysis & Prevention*, vol. 174, 2022.

[14] C. Sommer, R. German, and F. Dressler, "Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis," *IEEE Transactions on Mobile Computing (TMC)*, vol. 10, no. 1, Jan. 2011.