V2V-Enhanced Collision Risk Prediction Beyond the Sensing Horizon Using Prefix-Based Temporal Modeling

Daryel I. Leon Cachott*, Chadi Assi*, Maurice J. Khabbaz†, Floriano De Rango‡,
*Concordia University, Montreal, Canada †American University of Beirut, Beirut, Lebanon
‡University of Calabria, Rende, Italy

***Abstract*—This paper presents a dynamic collision risk prediction framework based on a variable-length, prefix-based temporal modeling strategy that enables continuous, frame-by-frame risk estimation under partial observation histories. Unlike conventional fixed-length approaches, the proposed method progressively refines predictions as new observations become available, allowing early risk estimation from limited temporal context. The framework fuses monocular vision and 2D LiDAR to extract spatial and kinematic features, which are processed by a recurrent temporal model to infer a probabilistic collision risk at each time step. To overcome the line-of-sight limitation of ego-centric sensing, a Vehicle-to-Vehicle (V2V) cooperative extension is introduced, where neighboring vehicles broadcast raw kinematic observations that are incorporated as synthetic feature inputs into the temporal model, enabling risk estimation for occluded threats beyond the sensing horizon. The system is evaluated in a Digital Twin simulation environment across safe driving, near-collision, collision, and occluded intersection scenarios. Results show that the ego-centric model achieves stable predictions within $0.7s$, with a precision of 0.94, recall of 0.90, and collision warnings up to $1.75s$ in advance. The cooperative extension further improves early warning in occluded scenarios, recovering $1.35s$ of warning time lost (WTL) and reaching a collision warning up to approximately $2.3s$ in advance, while maintaining consistent probability evolution during transitions between cooperative and onboard sensing.**

***Index Terms*—Collision risk prediction, autonomous vehicles, LSTM, variable-length sequences, partial observability, cooperative perception, V2V communication, Digital Twin simulation.**

I. Introduction

A. Preliminaries:

Advanced Driver Assistance Systems (ADAS) comprise a broad set of onboard technologies designed to improve driving safety by integrating sensing, control, and actuation capabilities that continuously monitor the surrounding environment and assist drivers in critical situations [1]. Evolving from early functions such as anti-lock braking to modern real-time perception and decision-making systems, ADAS aim at reducing collision risk, enhancing situational awareness, and progressively automating driving tasks [2]. Within this ecosystem, Collision Avoidance Systems (CAS) play a central role by identifying hazardous scenarios and issuing timely warnings or interventions based on information obtained from multimodal sensors, including radar, LiDAR, and vision systems [3], [4]. Despite notable reductions in accident rates, CAS performance remains constrained by sensing range, perception uncertainty, and decision latency, particularly under complex and adverse driving conditions [5], [6].

LiDAR and vision-based sensing have emerged as dominant technologies among contemporary perception modalities due to their complementary strengths [7], [8]. LiDAR offers accurate geometric representation of the environment and robustness to lighting variations, while camera-based systems provide rich semantic information essential for object detection and classification. However, LiDAR sensors incur high cost, generate large data volumes, and impose significant processing overhead, limiting their widespread deployment [9]. Conversely, monocular camera systems are cost-effective and information-rich but suffer performance degradation under poor illumination and adverse weather conditions [10], [11]. Consequently, hybrid multi-modal perception strategies that fuse LiDAR-derived geometric features with vision-based semantic cues have gained attention as a means to enhance robustness and reliability in real-world ADAS.

Despite these advances, improvements in sensing capabilities alone do not address a fundamental constraint inherent to all onboard perception systems: their ego-centric line-of-sight-limited nature [12]. In urban environments characterized by intersections and dense infrastructure, occluded threats represent a significant collision risk that onboard sensors alone cannot address. Vehicle-to-Vehicle (V2V) communication has emerged as a complementary paradigm to overcome this limitation, enabling vehicles to share kinematic information and extend their effective sensing horizon beyond physical occlusions. However, existing cooperative approaches predominantly rely on sharing high-level decisions, limiting their integration with onboard temporal prediction pipelines [13].

Beyond spatial awareness, safe navigation additionally requires the ability to anticipate how traffic situations evolve over time. Effective collision avoidance therefore depends not only on accurate perception, but also on predictive models that capture the temporal dynamics of interacting agents. In this context, Digital Twin–based simulation platforms have been widely adopted to generate diverse, controllable driving scenarios for training and validating perception and prediction models. However, maintaining consistency between simulated

environments and real-world deployments remains a critical challenge, as discrepancies can impair generalization in dynamic and multi-agent settings [14]. To address this limitation, data-driven temporal learning approaches, particularly Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures, have demonstrated strong potential for modeling sequential dependencies in vehicular trajectories and sensor streams. When combined with multi-modal perception and Digital Twin environments, LSTM-based frameworks enable robust and responsive collision risk prediction, supporting proactive decision-making in safety-critical autonomous driving applications.

B. Motivation and Problem Statement:

Beyond these limitations, a fundamental challenge lies in how collision prediction is formulated. A large body of prior work treats perception and prediction as disjoint modules, placing primary emphasis on object detection rather than continuous collision risk estimation [15]. As highlighted in [16], this separation limits the ability of such systems to capture the temporal evolution of Vehicle-to-Vehicle (V2V) interactions, which is essential for anticipating collisions before they become imminent. Consequently, many existing approaches fail to model collision risk as a time-evolving quantity, instead relying on isolated or late-stage predictions.

Furthermore, many existing methods assume simplified or quasi-static driving scenarios, reducing their robustness to diverse traffic behaviors and highly dynamic environments [17]. As a result, early-warning mechanisms are often triggered only at advanced stages of risk escalation, providing insufficient time for effective evasive maneuvers. This limitation is further exacerbated by the inability of such systems to reason under partial and evolving observations, a condition inherent to real-world driving.

Most importantly, existing temporal models typically rely on fixed-length observation windows, requiring a predefined amount of data before performing inference [18], [19]. This design introduces an inherent delay in prediction and limits the applicability of these methods in real-time settings, where early-stage observations are critical for timely collision anticipation. Moreover, prior approaches often fail to capture the continuous evolution of collision risk, treating prediction as a frame-level or static decision rather than a progressively refined estimate [20].

Beyond temporal modeling limitations, existing ego-centric frameworks are fundamentally constrained by the physical sensing horizon of the ego vehicle [21]. In occluded intersection scenarios, the interval between first detection and collision may be insufficient for effective evasive action, regardless of the quality of the temporal model. This limitation cannot be resolved through improved onboard sensing alone and requires cooperative information sharing among vehicles. Yet existing V2V-based approaches typically exchange processed outputs or rely on cloud infrastructure, rather than integrating raw kinematic broadcasts directly into an onboard temporal prediction model.

These limitations highlight the absence of a unified framework capable of performing continuous, context-aware collision risk estimation under partial temporal observations and beyond the ego vehicle's physical sensing horizon. To address this gap, collision prediction must be reformulated as a temporal inference problem in which risk is dynamically updated as new observations become available, rather than inferred from fixed-length temporal windows or limited to ego-centric perception. Motivated by this perspective, this work proposes a variable-length, prefix-based collision prediction framework that enables frame-by-frame estimation of imminent collision risk from incomplete observation histories. The proposed system integrates multi-modal sensory information from monocular vision and 2D LiDAR to jointly capture geometric relationships and inter-vehicle kinematic dynamics, which are sequentially processed to produce a continuously updated probabilistic collision risk at each time step. In addition, a cooperative extension is introduced in which raw kinematic information broadcast by neighboring vehicles — rather than derived risk scores or high-level decisions — is directly incorporated as synthetic feature inputs into the temporal model, enabling collision risk estimation for occluded threats without relying on external infrastructure or cloud-assisted processing.

C. Novel Contributions:

This paper addresses the problem of early and reliable collision prediction for autonomous vehicles operating in dynamic traffic environments. The main contributions of this work are:

- A variable-length, prefix-based temporal modeling strategy is proposed that formulates collision prediction as a continuous temporal inference problem, enabling frame-by-frame risk estimation under partial observation histories. Unlike fixed-length approaches, the method dynamically refines risk estimates as new observations arrive, enabling early anticipation of hazardous scenarios.
- A unified perception–prediction pipeline is developed that fuses monocular vision and 2D LiDAR features with synthetic feature vectors derived from V2V kinematic broadcasts into a single sequential temporal model. This design enables continuous collision risk estimation for both visible and occluded threats.
- The framework is evaluated in a Digital Twin simulation environment across diverse scenarios including occluded intersections, demonstrating progressive convergence of risk estimates and reliable early prediction under limited observation horizons. Cooperative perception is shown to improve Time-To-Accident awareness beyond the ego vehicle's physical sensing horizon.

D. Paper Organization:

The remainder of this paper is organized as follows. Section II reviews related work on collision detection, prediction, and avoidance in automotive systems. Section III presents the proposed system model, including the multimodal perception pipeline, ego-centric and cooperative feature extraction, and

the prefix-based temporal collision risk prediction framework. Section IV evaluates the performance of the proposed approach under diverse driving scenarios, analyzing prediction accuracy and temporal responsiveness. Finally, Section V concludes the paper and outlines directions for future research.

II. Literature Survey

Early CAS have explored monocular camera–based collision prediction systems. In [22], a vision-based method was proposed to estimate Time-to-Contact (TTC) using object size and image position cues, thereby avoiding explicit 3D scene understanding. The approach was validated on real vehicles and controlled test tracks, demonstrating the feasibility of camera-only collision estimation. However, monocular perception alone remains sensitive to illumination changes, occlusions, and depth ambiguity, limiting robustness in complex environments. Critically, occlusion in this context extends beyond sensor sensitivity: in urban intersection scenarios, a threatening vehicle may be geometrically invisible [23]–[25] to all onboard sensors regardless of their quality, representing a fundamental limitation of ego-centric perception architectures.

More recently, deep learning has been extensively adopted for object detection and motion analysis in autonomous driving. The You Only Look Once (YOLO) family of algorithms has demonstrated strong performance in visual perception tasks. More recent variants have reported high mean average precision in automotive applications, making them attractive for onboard perception pipelines, [26]–[28]. Nevertheless, these approaches are primarily designed for instantaneous object detection and do not inherently capture temporal risk evolution or collision likelihood over time.

Several studies have investigated sensor fusion and inter-vehicle association techniques to enhance collision prediction accuracy. In [29], point-matching algorithms were employed to estimate relative pose offsets between measurements obtained from different vehicles. Other works have focused on LiDAR-centric collision avoidance strategies (*e.g.*, [30]), emphasizing the trade-off between computational complexity and prediction accuracy, while monocular vision–based warning systems have highlighted challenges related to depth estimation and uncertainty, [31]. Although integrating visual detection with temporal modeling using LSTM networks has been explored (*e.g.*, [32]), most existing approaches rely on fixed-length temporal windows and training data derived from standard driving scenarios, which constrains adaptability under diverse and unpredictable traffic behaviors.

Cooperative collision avoidance systems have been proposed to overcome the sensing limitations of individual vehicles by sharing information among neighboring agents [12], [13]. However, existing approaches predominantly exchange processed outputs such as detected object lists, risk scores, or trajectory predictions, implicitly assuming that each vehicle has already perceived the relevant threat locally, an assumption that fails precisely in occluded scenarios where cooperation is most needed [33]–[35]. To the best of the authors' knowledge, no existing approach directly incorporates raw V2V kinematic broadcasts as synthetic feature inputs into an onboard recurrent temporal prediction model, leaving the integration of cooperative sensing and sequential risk inference as an open problem.

To support realistic evaluation of such frameworks, Digital Twin–based simulation environments have been increasingly adopted to enable safe and repeatable evaluation of perception and prediction algorithms under diverse and controllable conditions [36], [37]. However, existing deployments typically rely on Vehicle-to-Cloud communication for computation, introducing latency that reduces suitability for fully onboard real-time prediction. In this work, the Digital Twin environment serves purely as a high-fidelity evaluation platform, with all sensing, cooperative kinematic exchange, and temporal inference performed onboard without cloud dependency.

In contrast to existing approaches that rely on high-cost 3D LiDAR configurations, cloud-assisted computation, fixed-horizon temporal models, or processed-output sharing in cooperative systems, this work presents a unified collision risk prediction framework that combines ego-centric multimodal sensing with direct integration of V2V kinematic broadcasts into a single recurrent temporal model. This design addresses both the temporal limitation of fixed-length prediction and the spatial limitation of ego-centric sensing, enabling early collision risk estimation for visible and occluded threats within a single onboard inference pipeline.

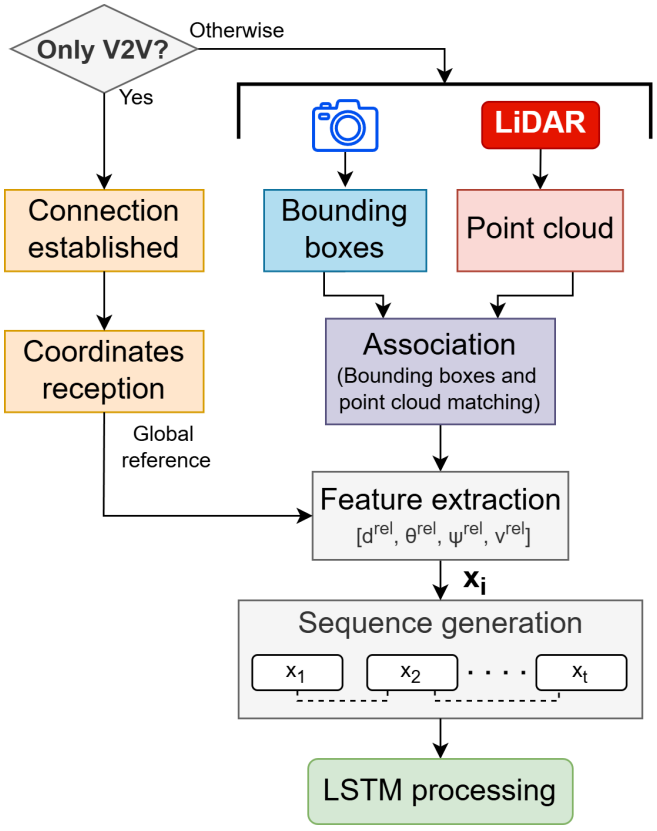


Fig. 1: Pipeline integrating ego-centric multimodal sensing and V2V cooperative kinematic broadcasts into a unified LSTM-based temporal inference framework.

III. System Model

Fig. 1 illustrates the processing pipeline of the proposed collision prediction framework. A monocular camera and a 2D LiDAR operate in parallel to capture visual and spatial information from the surrounding environment. Visual detections obtained from the camera are associated with LiDAR point clusters through a data association process, enabling the identification of corresponding vehicles across sensing modalities. For each matched object, geometric and kinematic features are extracted and organized into temporal sequences. When the target vehicle is occluded and unavailable to onboard sensors, kinematic data broadcast by neighboring vehicles via V2V communication are processed as synthetic observations and incorporated into the same sequential representation. In both cases, the resulting feature sequences are processed by the temporal model, which outputs a frame-by-frame collision probability.

A. Scenario Description:

To evaluate the proposed collision prediction framework, a realistic traffic scenario is implemented within the Interactive Quanser QLabs simulation environment, which provides a Digital Twin of the Quanser QCar [38] vehicle (Fig. 2) platform, a 1/10 scale research vehicle equipped with an ego-facing RGB monocular camera and a 2D RPLiDAR sensor. The camera provides frontal semantic perception, while the LiDAR delivers 360° planar range measurements.

Fig. 2: Platform used for experiments.

In addition to onboard sensing, the proposed framework assumes the availability of short-range Vehicle-to-Vehicle (V2V) communication for cooperative perception. Each vehicle periodically broadcasts its kinematic state, including position, velocity, and orientation, using standard vehicular communication protocols such as IEEE 802.11p (DSRC) or Cellular-V2X (C-V2X) [39], [40].

To abstract the communication process, V2V interaction is modeled as a low-latency kinematic broadcast channel. The transmitted information is assumed to be received reliably within the temporal resolution of the perception pipeline. Communication delays, packet losses, and synchronization errors are not explicitly modeled, as the focus of this work lies in the integration of cooperative information into the temporal prediction framework. Under this abstraction, V2V data are treated as an additional sensing modality that extends the observable state space beyond the ego vehicle's line-of-sight constraints.

Each simulation instance generates stochastic interactions between two autonomous vehicles. Scenario diversity is achieved through a combination of uniformly and Gaussian-distributed random variables. Initial vehicle positions and orientations are sampled from uniform distributions to ensure broad spatial coverage across the test area. Driving behavior is subsequently randomized using Gaussian-distributed parameters to emulate natural variations in human-like acceleration and speed control. Specifically, the maximum vehicle speed follows $v_{\max} \sim \mathcal{N}(3.75, 1.318^2)$ m/s, the acceleration is modeled as $a_\mu \sim \mathcal{N}(0.25, 0.05^2)$ m/s^2, and the per-frame acceleration variability follows $\sigma_a \sim \mathcal{N}(0.08, 0.02^2)$ m/s^2. These parameters govern each vehicle's instantaneous acceleration according to

$$a_i(t) = \mathcal{N}(a_\mu, \sigma_a^2) \times \left(1 - \frac{v_i(t)}{v_{\max}}\right). \tag{1}$$

B. Camera-Based Object Detection and Position Estimation:

Vehicle perception is performed using the monocular camera in conjunction with a YOLO-based object detection model [41]. This module identifies bounding-boxes of surrounding traffic participants. The resulting visual detections provide semantic context that supports subsequent data association and multi-modal feature extraction. The relative position of detected vehicle is estimated using image-based geometric cues.

1) Distance Estimation: The relative longitudinal distance, d, between the ego vehicle and a detected vehicle is approximated as a function of the bounding box height h (in *pixels*):

$$d_i^{ego} = f_y \frac{H_{\text{real}}}{h}, \tag{2}$$

where f_y denotes the focal length along the vertical image axis and H_{real} represents the real-world vehicle height, which is 0.182 m according to the QCar manufacturer.

2) Angle Estimation: The relative bearing angle, θ, of a detected vehicle with respect to the ego vehicle is estimated based on the horizontal position of the bounding box center in the image plane:

$$\theta_i^{ego} = \arctan\left(\frac{c_{bx} - c_{ix}}{f_x}\right), \tag{3}$$

where c_{bx} denotes the horizontal coordinate of the bounding box center, c_{ix} the horizontal center of the image, and f_x is the focal length along the horizontal axis. All the camera intrinsic parameters were provided by the manufacturer.

C. Data Association:

For each vehicle detected in the camera image, the estimated relative distance and bearing obtained from bounding box geometry are used to guide the association with 2D LiDAR measurements. The association process matches camera-detected vehicles with their corresponding LiDAR point clusters by exploiting spatial proximity and orientation consistency. Specifically, candidate LiDAR clusters are evaluated based on their relative distance and angular alignment with

respect to the ego vehicle. Once association is established, the LiDAR-derived distance and relative angle are selected as the final geometric features.

Once the LiDAR points are determined during the association, the relative speed feature and the relative motion direction is estimated as follows. Let $\Delta\tilde{x}_i^m$ and $\Delta\tilde{y}_i^m$ denote the differences between two consecutive frames LiDAR points:

$$v_i^m = \frac{\sqrt{(\Delta\tilde{x}_i^m)^2 + (\Delta\tilde{y}_i^m)^2}}{\Delta t} \qquad \psi_i^m = \mathrm{atan2}(\Delta\tilde{y}_i^m,\ \Delta\tilde{x}_i^m) \tag{4}$$

with Δt measured from timestamp differences between consecutive frames.

D. Cooperative Feature Extraction

Once communication is established, upon reception, the ego vehicle applies a coordinate transformation to convert the received global position into its own ego-centric body frame, accounting for the rear-axle to body-center offset and the ego vehicle heading. The forward and lateral body-frame components of the inter-vehicle displacement vector are computed as:

$$b_x = \cos(\psi_{ego})(x_B - x_{ego}) + \sin(\psi_{ego})(y_B - y_{ego}) \tag{5}$$

$$b_y = -\sin(\psi_{ego})(x_B - x_{ego}) + \cos(\psi_{ego})(y_B - y_{ego}) \tag{6}$$

where (x_B, y_B) is the broadcast position of target vehicle, (x_{ego}, y_{ego}) is the ego body center in the world frame, and ψ_{ego} is the ego heading. These components are then used to compute the cooperative distance and bearing:

$$d_i^{coop} = \sqrt{b_x^2 + b_y^2} \qquad \theta_i^{coop} = atan2(-b_y, b_x) \tag{7}$$

The target speed v_i^{coop} and heading ψ_i^{coop} are estimated from the frame-to-frame displacement of the broadcast position using angle unwrapping to prevent discontinuities:

$$v_i^{coop} = \frac{\sqrt{\Delta b_x^2 + \Delta b_y^2}}{\Delta t} \qquad \psi_i^{coop} = atan2(\Delta b_x, \Delta b_y) \tag{8}$$

where $\Delta b_x = b_x^{(i)} - b_x^{(i-1)}$, $\Delta b_y = b_y^{(i)} - b_y^{(i-1)}$, and Δt is the inter-frame time interval.

E. Collision Risk Prediction:

Collision prediction is formulated as a temporal inference problem over partial observation histories. Rather than relying on fixed-length temporal windows, the proposed framework estimates collision risk at each time step using the sequence of observations available up to that point.

For each detected and associated target vehicle, a feature vector is extracted at frame i from the onboard sensing modalities as

$$x_i^{\text{ego}} = [d_i^{\text{ego}},\ \theta_i^{\text{ego}},\ \psi_i^{\text{ego}},\ v_i^{\text{ego}}], \tag{9}$$

In the cooperative setting, kinematic information derived from V2V broadcasts is organized into an auxiliary feature vector:

$$x_i^{\text{coop}} = [d_i^{\text{coop}},\ \theta_i^{\text{coop}},\ \psi_i^{\text{coop}},\ v_i^{\text{coop}}], \tag{10}$$

These cooperative features are treated as synthetic observations of vehicles that are occluded from the ego vehicle's direct sensing field.

To enable unified temporal modeling, both ego-centric and cooperative observations are expressed in a common representation. Accordingly, each target vehicle is described by a generalized feature vector:

$$x_i = [d_i^{\text{rel}},\ \theta_i^{\text{rel}},\ \psi_i^{\text{rel}},\ v_i^{\text{rel}}], \tag{11}$$

where the superscript "rel" denotes features expressed relative to the ego vehicle, regardless of whether they originate from onboard sensing or V2V-based cooperative perception. The operating mode is determined at each frame by the availability of onboard sensor observations: if the target vehicle is successfully detected by the camera and associated with a LiDAR cluster, ego-centric features are used exclusively; if the target is occluded and no valid onboard detection is available, cooperative features derived from V2V kinematic broadcasts are substituted into the unified feature representation, extending risk estimation beyond the physical sensing horizon.

At time step t, the available temporal context is represented by the observation prefix

$$X_{1:t} = [x_1, x_2, \ldots, x_t], \qquad 2 < t \leq T, \tag{12}$$

where T denotes the maximum sequence length.

The proposed modeling strategy infers collision risk from variable-length prefixes rather than fixed-horizon temporal windows. This allows the model to generate predictions from incomplete temporal information and to refine these predictions as additional observations become available. Accordingly, at each frame t, the temporal model produces a collision risk estimate

$$\hat{y}_t = f(X_{1:t}), \tag{13}$$

where $\hat{y}_t$ is a time-evolving probability that should converge as t increases. This prefix-based formulation enables progressive estimation of collision risk as the temporal context increases.

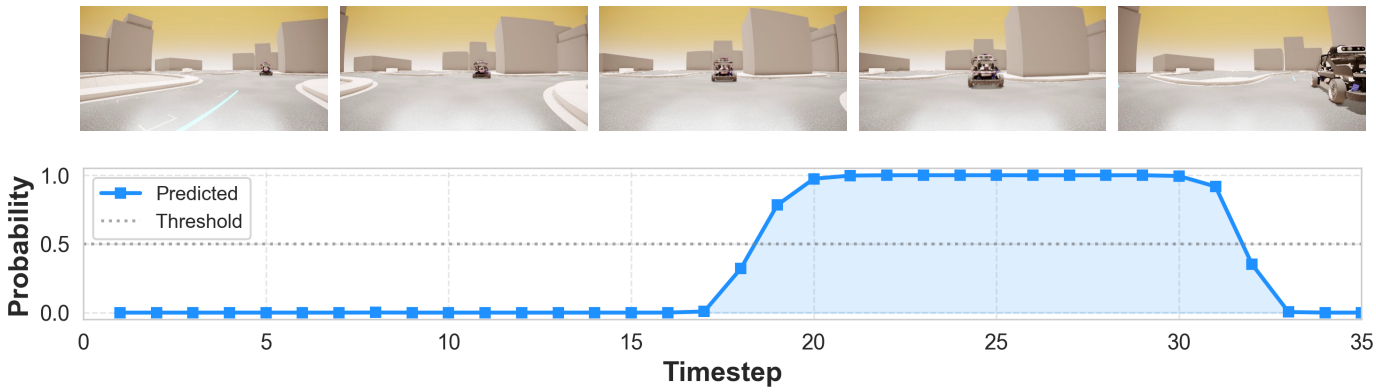
During training, the model is exposed to variable-length prefixes $X_{1:t}$, with $t \in [2, T]$, enabling it to learn risk estimation under different levels of temporal context.

To quantify prediction earliness, the Time-To-Accident (TTA) metric is defined as

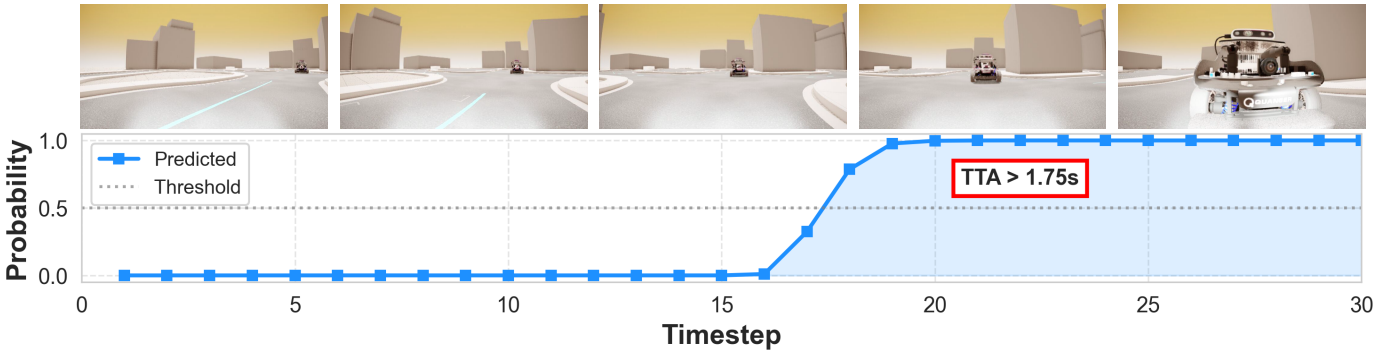
$$\mathrm{TTA} = \frac{f_{\text{collision}} - f_{\text{first}}}{\text{Frames per second (FPS)}}, \tag{15}$$

where f_{first} is the first frame at which the predicted collision probability exceeds a predefined threshold, and $f_{\text{collision}}$ is the frame at which the collision occurs.

A fixed probability threshold of 0.5 is used to convert predicted collision likelihood into a binary risk decision, following standard practice in the literature [42] and providing a balanced trade-off between sensitivity and false alarms.



(a) Realistic simulated near-miss scenario.



(b) Realistic simulated collision scenario.

Fig. 3: Qualitative analysis of collision risk prediction. (a) Near-miss scenario where the predicted risk increases during a potentially dangerous interaction but decreases after the ego vehicle changes direction, avoiding the collision. (b) Collision scenario where the predicted risk increases several frames before impact as inter-vehicle distance decreases.

IV. Simulation Framework and System Performance Evaluation

For model training and evaluation, a dataset comprising 975 randomly generated driving scenarios is constructed, with a collision to non-collision ratio of approximately 40/60. Each scenario consists of 15 consecutive frames sampled at a frequency of 7.4 Hz. By considering variable-length temporal subsequences, the effective dataset size increases to 13.650 labeled samples. All sequences are annotated using the simulator's ground-truth collision flag. To ensure robust performance evaluation, a 5-fold stratified cross-validation procedure is employed, preserving the class distribution in each fold.

A. Quantitative Analysis:

The performance of the proposed methodology is evaluated against three baseline temporal models, Temporal Convolutional Networks (TCN), GRU, and LSTM, is presented in Fig 5. Overall, the proposed approach achieves strong and consistent performance. The LSTM-based implementation attains the highest performance, with precision close to 0.94, recall around 0.90, and an F1-score of 0.92, indicating a well-balanced trade-off between false positives and false negatives. GRU achieves comparable precision (≈ 0.94) but lower recall (≈ 0.82), resulting in a reduced F1-score (≈ 0.87). The TCN model also maintains relatively high precision (≈ 0.93), but its recall decreases further (≈ 0.75), leading to the lowest F1-score (≈ 0.83) among the considered approaches.

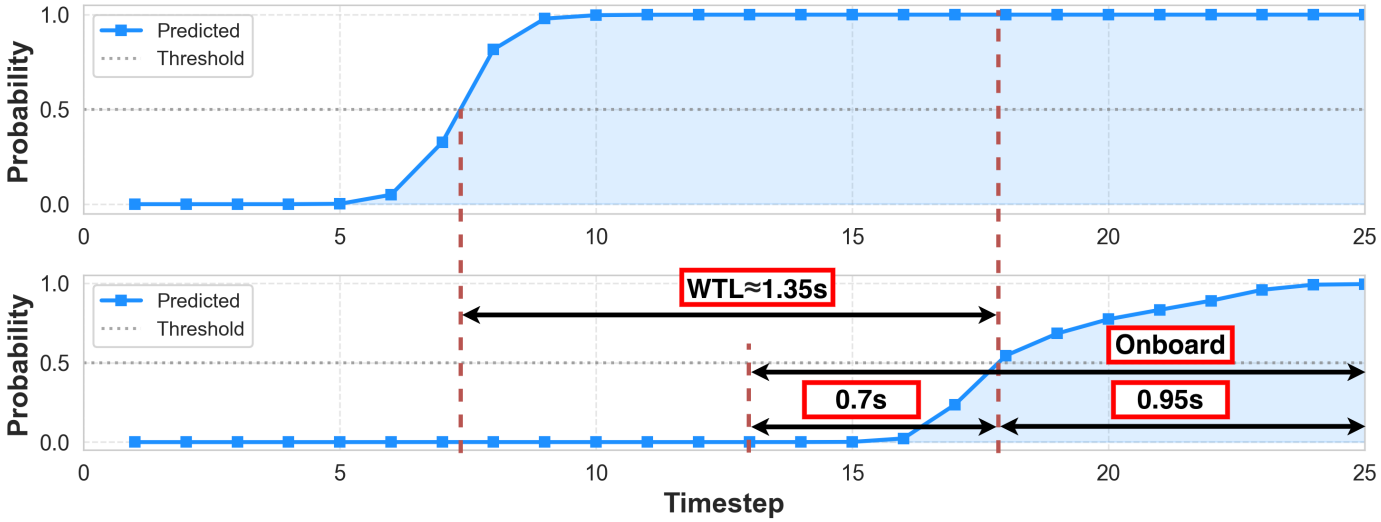
TABLE I: K-Fold Validation Results

Metric	F1	F2	F3	F4	F5	95% C.I.
Precision	0.97	0.94	0.95	0.91	0.98	[0.937–0.956]
Recall	0.88	0.85	0.89	0.83	0.90	[0.861–0.884]
F_1-Score	0.92	0.91	0.93	0.89	0.93	[0.913–0.927]
AUC	0.94	0.93	0.94	0.91	0.94	[0.929–0.940]

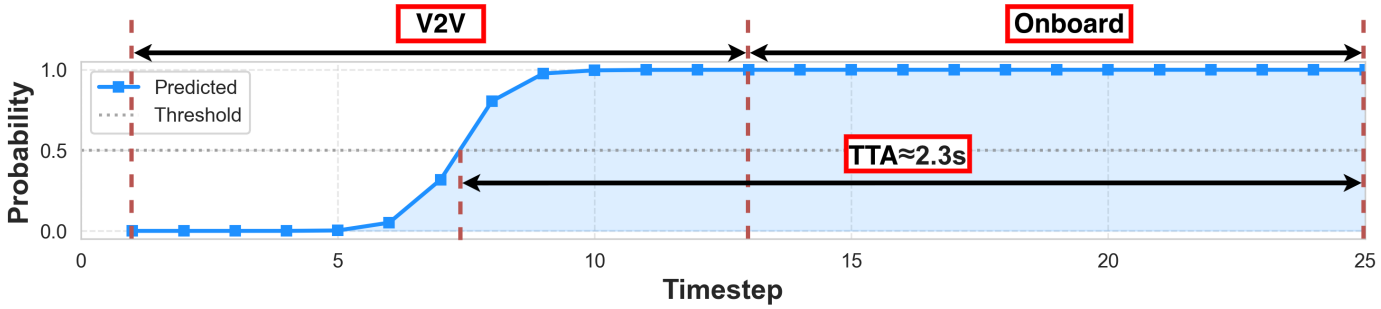
Table I summarizes the mean performance metrics along with their corresponding 95% confidence intervals across all folds, highlighting the stability and reliability of the proposed framework. The evaluation is conducted using the LSTM-based implementation, which achieved the best overall performance among the considered temporal models.

B. Qualitative Scenario Analysis:

Figure 3a presents representative qualitative results for non-collision and Figure 3b for collision scenarios. In the non-collision case, the predicted collision probability stabilizes near zero during safe driving phases. When proximity increases, the probability rises accordingly and subsequently



(a) Ego-centric prediction: non-occluded (top) vs. occluded (bottom) scenario



(b) Cooperative prediction for the occluded scenario

Fig. 4: Predicted collision probability trajectories comparing ego-centric and cooperative sensing under non-occluded and occluded intersection scenario.

decreases as the ego vehicle alters its trajectory, demonstrating adaptive response to dynamic interactions.

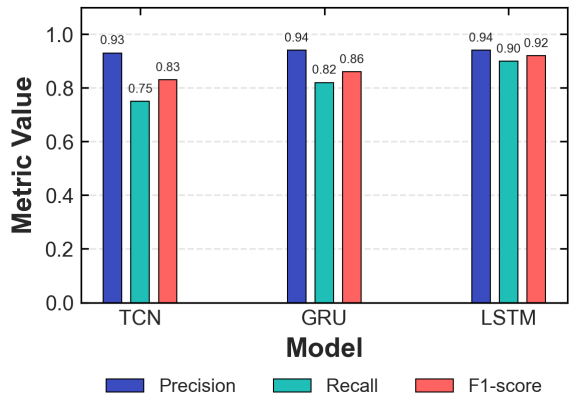


Fig. 5: Performance comparison of different baselines models.

In contrast, the collision scenario shows a rapid and monotonic increase in predicted collision probability after approximately frame 16, converging toward unity as the ego vehicle continues its approach without evasive action. The rapid convergence and subsequent stability of the prediction illustrate the robustness of the temporal modeling framework once sufficient observations are accumulated.

C. Cooperative Evaluation:

To assess the benefit of cooperative sensing under occlusion, an additional evaluation is conducted using a dedicated occluded intersection scenario, (Figure 6). In this configuration, a building structure positioned at the intersection corner prevents the ego vehicle from observing the approaching target vehicle through onboard sensors until both vehicles enter the intersection area. This represents a worst-case condition for ego-centric collision prediction, where the target vehicle is geometrically invisible regardless of sensing quality.

Figure 4a compares ego-centric predictions for the non-occluded and occluded versions of the same collision scenario. In the non-occluded case (top), the predicted collision probability crosses the 0.5 threshold at approximately frame 7, providing an early and stable warning. In the occluded case (bottom), the target vehicle remains invisible to all onboard sensors until frame 13, and after only $0.7s$ the probability rises sharply but with substantially reduced reaction time. The delay between the two threshold crossings amounts to $1.35s$

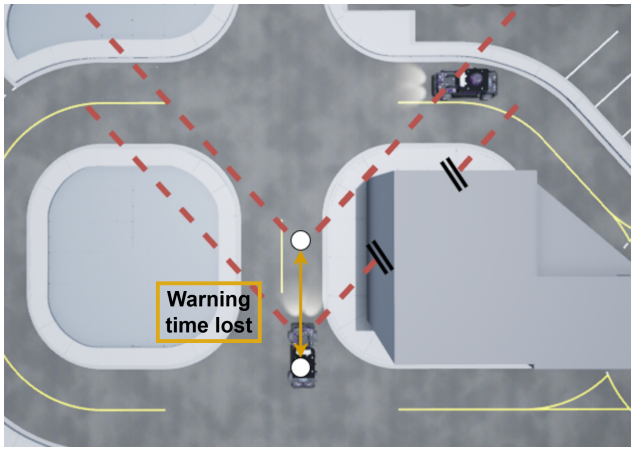


Fig. 6: Intersection scenario with occlusion.

at the operating frequency, quantifying the warning time lost due to occlusion under ego-centric sensing alone.

Figure 4b shows the cooperative prediction for the same occluded scenario. V2V kinematic broadcasts from the target vehicle are incorporated as synthetic feature inputs from the earliest frames, enabling the model to estimate collision risk during the full occlusion phase. As a result, the probability threshold is crossed at approximately frame 7, recovering the early warning timing observed in the non-occluded reference case. Upon visual detection of the target vehicle at frame 13, the system seamlessly transitions from V2V-derived to onboard sensor features with no discontinuity in the predicted probability trajectory. These results confirm that the cooperative extension recovers the $1.35s$ of warning time lost (WTL) under occlusion, extending the temporal model's effective sensing horizon beyond the ego vehicle's physical line-of-sight limitations and achieving a TTA of $\approx 2.3s$.

D. Model Robustness to Label Flipping:

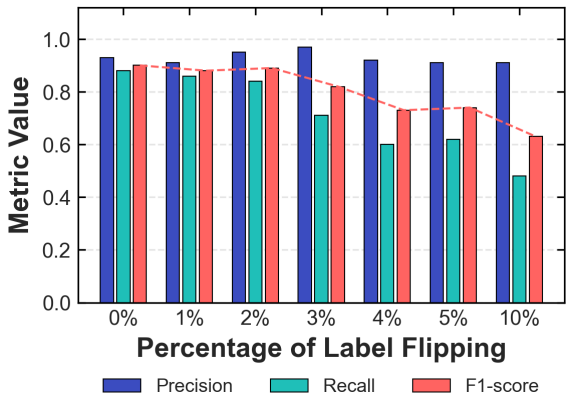


Fig. 7: Label flipping robustness evaluation.

To evaluate robustness under imperfect supervision, an additional experiment is conducted using a synthetically contaminated dataset. Label flipping is applied by deliberately mislabeling a percentage of collision samples as non-collision instances. Figure 7 illustrates the impact of label noise. Increasing the noise ratio directly correlates with a degradation in performance.

V. Conclusions

The experimental results demonstrate that the proposed framework achieves a well-balanced trade-off between precision and recall, confirming its effectiveness in distinguishing collision from non-collision scenarios. The high AUC value of approximately 0.94 further indicates strong discriminative capability. Importantly, these results validate that reliable collision risk prediction can be achieved from partial observation prefixes.

From a temporal perspective, the predicted collision risk exhibits an initial uncertainty phase, followed by transient fluctuations, and progressively converges toward a stable confidence level as additional observations are accumulated. This behavior reflects the continuous refinement of risk estimation as temporal context increases. Notably, the model is capable of anticipating collision events more than $1.75s$ in advance, while reaching a stable prediction state within approximately $0.7s$ or 5 frames at the current frequency. These findings highlight the effectiveness of the proposed prefix-based temporal modeling strategy for early collision risk prediction.

The cooperative extension further demonstrates that integrating V2V kinematic broadcasts as synthetic feature inputs into the temporal model successfully extends collision risk estimation beyond the ego vehicle's physical sensing horizon. In the occluded intersection scenario, ego-centric sensing alone delayed the collision-risk threshold crossing, producing a warning time loss of $1.35s$ compared with the non-occluded case. By incorporating the approaching vehicle's position and heading broadcasts during the full occlusion phase, the cooperative framework recovered this lost warning time and achieved a TTA of approximately $2.3s$. This result confirms that V2V information can compensate for temporary visual occlusions and extend the effective prediction horizon of the temporal model. Moreover, the continuous transition from V2V-derived features to onboard sensor-based features after visual detection shows that both observation sources can be integrated within a unified feature representation without modifying the temporal model architecture.

The robustness analysis further shows that performance degrades when label contamination exceeds 2%, indicating sensitivity to annotation noise. This observation underscores the importance of developing learning mechanisms that are inherently robust to imperfect supervision. As future work, noise-resilient temporal learning approaches may be explored to mitigate the impact of corrupted labels.

References

[1] R. Bishop, "Intelligent Vehicle Systems: Yesterday, Today, and Tomorrow," IEEE VTS News, 2000.
[2] A. Eskandarian, "Handbook of Intelligent Vehicles," Springer Science & Business Media, 2012.
[3] S. D. Pendleton et al., "Perception, Planning, Control, and Coordination for Autonomous Vehicles," Machines, 2017.

[4] Insurance Institute for Highway Safety (IIHS), "Real-world benefits of crash avoidance technologies," 2021.

[5] National Highway Traffic Safety Administration (NHTSA), "Critical reasons for crashes investigated in the National Motor Vehicle Crash Causation Survey," 2015.

[6] S. E. Shladover, "Connected and Automated Vehicle Systems: Introduction and Overview," Journal of Intelligent Transportation Systems, 2018.

[7] V. Kilic et al., "LiDAR and Camera Fusion for Road Object Detection," IEEE Access, 2021.

[8] M. Tsakmakopoulou et al., "Review of Sensor Fusion Methods for Autonomous Driving," Sensors, 2022.

[9] M. H. Rahman et al., "Challenges and Opportunities in LiDAR Data Processing," IEEE Reviews in Biomedical Engineering, 2020.

[10] A. K. Shetty, "Monocular Visual Odometry for Autonomous Vehicles," International Journal of Vehicle Autonomous Systems, 2021.

[11] M. Dreissig et al., "Deep Learning-Based Object Detection in Adverse Weather," IEEE Transactions on Intelligent Transportation Systems, 2020.

[12] Shi, J., Zhao, J., Zhuo, L., Wang, X., Zhan, X., & Liu, H. (2025). V2V cooperative perception with adaptive communication loss for autonomous driving. IEEE Transactions on Intelligent Transportation Systems.

[13] Li, L., Cheng, Y., Sun, C., & Zhang, W. (2024, June). ICOP: Image-based cooperative perception for end-to-end autonomous driving. In 2024 IEEE Intelligent Vehicles Symposium (IV) (pp. 2367-2374). IEEE.

[14] C. Wu et al., "Bridging the Sim2Real Gap in Autonomous Driving," IEEE Transactions on Neural Networks and Learning Systems, 2022.

[15] L. Dal'Col, M. Oliveira, and V. Santos, "Joint Perception and Prediction for Autonomous Driving: A Survey," arXiv preprint arXiv:2412.14088, 2024. [Online]. Available: https://arxiv.org/abs/2412.14088

[16] Y. Zhang, Y. Zou, S. Selpi, and others, "Spatiotemporal Interaction Pattern Recognition and Risk Evolution Analysis During Lane Changes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 6, pp. 6663–6673, 2023. [Online]. Available: http://dx.doi.org/10.1109/TITS.2022.3233809

[17] C. A.-O. Chitraranjan, V. Vipulananthan, and T. Sritharan, "Vision-Based Collision Warning Systems with Deep Learning: A Systematic Review. LID - 10.3390/jimaging11020064 [doi] LID - 64," (in eng), no. 2313-433X (Electronic).

[18] S. Mozaffari, M. Alizadeh, and B. Moshiri, "Trajectory Prediction with Observations of Variable-Length for Motion Planning in Highway Merging Scenarios," 2024.

[19] M. Fan, J. Wang, and Y. Liu, "Temporal-Adaptive Progressive Distillation for Observation-Adaptive Trajectory Forecasting," arXiv preprint arXiv:2603.06231, 2026.

[20] T. Zhao, H. Chen, and Y. Zhang, "Accident Anticipation via Temporal Occurrence Prediction," 2025.

[21] Gao, W., Zhou, L., & Luo, X. (2026, February). CampusSyn: A Real World Complex Environment Dataset for Vehicle-to-Vehicle Collaborative Perception. In 2026 14th International Conference on Intelligent Control and Information Processing (ICICIP) (pp. 197-204). IEEE.

[22] E. Dagan et al., "Forward Collision Warning with a Single Camera," IEEE Intelligent Vehicles Symposium, 2004.

[23] Moller, K., Schwarzmeier, L., & Betz, J. (2025, June). From shadows to safety: Occlusion tracking and risk mitigation for urban autonomous driving. In 2025 IEEE Intelligent Vehicles Symposium (IV) (pp. 1883-1890). IEEE.

[24] Zhang, X., Li, Y., Wang, J., Qin, X., Shen, Y., Fan, Z., & Tan, X. (2025). InScope: A new real-world 3D infrastructure-side collaborative perception dataset for open traffic scenarios. Information Fusion, 103951.

[25] Yu, M. Y., Vasudevan, R., & Johnson-Roberson, M. (2020, May). Risk assessment and planning with bidirectional reachability for autonomous driving. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 5363-5369). IEEE.

[26] M. Gasparovic et al., "YOLOv5 for Real-Time Object Detection," 2022.

[27] M. A. Widyadara et al., "Evaluation of YOLOv8 for Vehicle Detection," 2023.

[28] P. Egger et al., "Deep Learning in Automotive Computer Vision," 2021.

[29] A. Rauch, S. Maier, F. Klanner, and K. Dietmayer, "Inter-vehicle object association for cooperative perception systems," in 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), 2013: IEEE, pp. 893-898.

[30] E. Candela, Y. Feng, D. Mead, Y. Demiris, and P. Angeloudis, "Fast collision prediction for autonomous vehicles using a stochastic dynamics model," in 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), 2021: IEEE, pp. 211-216.

[31] A. M. Ibrahim, R. M. Hassan, A. E. Tawfiles, T. Ismail, and M. S. Darweesh, "Real-time collision warning system based on computer vision using mono camera," in 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020: IEEE, pp. 60-64.

[32] P. Goudarzi and B. Hassanzadeh, "Collision risk in autonomous vehicles: classification, challenges, and open research areas," Vehicles, vol. 6, no. 1, pp. 157-190, 2024.

[33] Yan, S., Wu, Y., Liu, Z., & Xie, C. (2026). Research on Cooperative Vehicle–Infrastructure Perception Integrating Enhanced Point-Cloud Features and Spatial Attention. World Electric Vehicle Journal, 17(4), 164.

[34] Hussain, M., Ali, N., & Hong, J. E. (2022). Vision beyond the field-of-view: A collaborative perception system to improve safety of intelligent cyber-physical systems. Sensors, 22(17), 6610.

[35] Zhang, Q., Kong, H., Gu, P., Meng, Y., & Liu, T. (2025). Spatio-temporal risk prediction-based longitudinal motion control for autonomous vehicles in blind zone environments. Transactions of the Institute of Measurement and Control, 01423312251376714.

[36] Z. Wang et al., "A digital twin paradigm: Vehicle-to-cloud based advanced driver assistance systems," in 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020: IEEE, pp. 1-6.

[37] K. M. Alam and A. El Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," IEEE access, vol. 5, pp. 2050-2062, 2017.

[38] Quanser, "QCar User Manual and Platform Specifications," 2022.

[39] Guo, S., Aloi, D. N., Li, J., & Zhao, H. (2025). Physical layer evaluation on IEEE 802.11 p with different configurations in NLOS scenarios for V2V communications. IEEE Access, 13, 44428-44444.

[40] Zadobrischi, E., & Havriliuc, Ş. (2024). Enhancing scalability of C-V2X and DSRC vehicular communication protocols with LoRa 2.4 GHz in the scenario of urban traffic systems. Electronics, 13(14), 2845.

[41] Panchal, A., Sheth, N., & Prakash, C. (2025, December). Monocular Vehicle Detection and Distance Estimation Using YOLO. In 2025 IEEE International Conference on Smart Power, Energy, Renewables, and Transportation (SPERT) (pp. 1-5). IEEE.

[42] F. Mahmood, D. Jeong, and J. Ryu, "A new approach to traffic accident anticipation with geometric features for better generalizability," IEEE Access, vol. 11, pp. 29263-29274, 2023.