A Feasibility Study of Risk Triggering Using Multimodal Trajectory Predictions in Mixed Traffic

Menglin Yang*[1], Sven Fröhlich[1], and Sebastian Pape[1]

[1]Research Associate, Chair of Traffic Process Automation, "Friedrich List" Faculty of Transport and Traffic Sciences, Technische Universität Dresden, Germany

Short summary

Ensuring safe interactions between vehicles and vulnerable road users (VRUs) is a critical challenge in mixed traffic environments. This study investigates the use of multimodal trajectory predictions to anticipate potential hazards involving VRUs and generate actionable risk warnings. A structured framework is proposed to extract safety indicators from predicted trajectories and evaluate multiple trigger strategies. Experimental results show that OR-logic aggregation achieves the highest overall performance by balancing hazard detection and false alarm control, while other strategies exhibit more conservative or imbalanced behavior. Furthermore, the choice of the number of aggregated trajectories (K) introduces an additional trade-off, where increasing K enhances risk coverage but may lead to reduced reliability due to more frequent warnings. These findings provide practical guidance for the design and deployment of risk warning systems.
Keywords: Proactive safety warning, Multimodal trajectory prediction, Risk-trigger strategies, Mixed traffic environments

1 Introduction

In mixed traffic environments, proactive safety warning systems are essential for supporting safe interactions between vehicles and vulnerable road users. Such systems require anticipation of potential future interactions to issue timely warnings, but future motion is inherently uncertain, and a single deterministic prediction is insufficient to capture potential risks.

Previous studies on proactive safety warning systems have largely relied on rule-based indicators such as time-to-collision (TTC) (Kiefer et al., 2005; Chen et al., 2013) or deterministic trajectory prediction models, such as constant velocity or kinematic models and Kalman filter-based approaches (Wu et al., 2014; Zhang et al., 2022), as commonly used in advanced driver assistance systems. While these approaches enable timely risk detection, they are limited in capturing the inherent uncertainty of future interactions. Probabilistic and multimodal prediction methods have been explored to represent diverse future behaviors and risk assessment (Kim & Kum, 2018; Xiong et al., 2018), yet their integration into systematic safety warning generation remains relatively underexplored. A key challenge is how to aggregate these predictions into a binary decision on whether to trigger a warning, particularly in the absence of explicit risk labels, as well as how prediction uncertainty affects the reliability of risk warnings.

To address this, we propose a framework that leverages multimodal trajectory predictions to generate actionable safety warnings. The framework consists of four hierarchical steps: obtaining multimodal predicted trajectories, extracting safety indicators from these trajectories, applying different trigger strategies to each predicted trajectory, and aggregating the resulting signals to produce a final risk alert. This structured pipeline provides a unified basis for evaluation and enables systematic handling of prediction uncertainty.

This study is conducted as part of the PROTEKT project, which focuses on the development of practical risk warning systems for mixed traffic scenarios. In this context, an effective proactive warning system must satisfy three key requirements: safety, acceptability, and timeliness (Wogalter et al., 1987; Bai et al., 2025; Weber et al., 2020). Safety ensures that hazardous situations are reliably detected; acceptability limits false alarms to maintain user trust; and timeliness allows meaningful intervention. These requirements motivate not only the design of the proposed framework but also the subsequent evaluation of how different trigger strategies balance hazard detection and false alarm control. Specifically, within the proposed framework, we examine several trigger

strategies—including individual signals (spatial interaction or behavioral response) and their logical combinations (AND- and OR-logic)—focusing on the trade-offs between safety (sensitivity to actual hazards) and reliability (minimizing false alarms), particularly in the absence of explicit risk labels. By comparing these strategies across multiple predicted trajectories, the study identifies how different approaches influence warning effectiveness and provides guidance for practical system deployment.
The paper is organized as follows: Section 2 presents the framework and methodology, Section 3 introduces the experiments and analyzes the results, and Section 4 discusses the results and concludes with directions for future work.

2 Methodology

Overall framework

This section presents the overall framework for leveraging multimodal trajectory predictions to generate safety risk triggers in mixed traffic scenarios. The approach follows a four-step hierarchical process, as illustrated in Fig. 1. First, multimodal predicted trajectories for all traffic participants are obtained. Next, relevant safety indicators are extracted from these trajectories. The third step applies the defined trigger strategies to each predicted trajectory, generating risk signals based on individual or combined proxy signals. Finally, the risk assessments across multiple predicted modes are aggregated using a Top-K approach to account for diverse possible futures, producing actionable triggers. This structured pipeline provides a unified basis for evaluating prediction-driven safety risk triggering.
Fig. 1 shows the whole framework of this study.

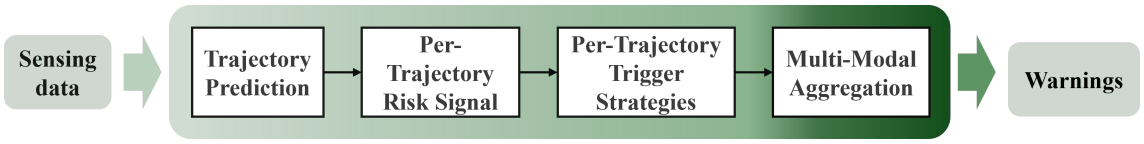


Figure 1: Example figure caption.

In this study, we utilized Multipath++ (Varadarajan et al., 2021) as the multimodal trajectory prediction model for all traffic participants over a 3-second horizon to provide adequate lead time for potential evasive actions. Each traffic participant is referred to as an agent. These predicted trajectories serve as the input for subsequent risk indicator extraction and trigger generation.

Per-Trajectory Risk Signals

Three safety indicators are first computed to assess potential conflicts between each pair of agents. These continuous indicators are then mapped to actionable risk signals by applying predefined thresholds, resulting in two complementary risk signals: the Interaction-based Collision Signal and the Behavioral Response Signal.

- **Interaction-based Collision Signal**: This signal captures the spatial proximity and temporal likelihood of a potential collision between the ego agent and other traffic participants. It is quantified using TTC (Hydén, 1987; Nadimi et al., 2020), defined as the predicted time remaining until a possible collision assuming constant velocities along the predicted trajectories.
- **Behavioral Response Signal**: This signal reflects the ego agent's kinematic responses indicative of risk perception. It is quantified using Maximum Deceleration (MaxDec), representing the largest negative acceleration, and Maximum Steering Rate (MaxSteer), representing the highest rate of change in heading angle, both over the prediction horizon. These indicators capture evasive maneuvers such as braking or abrupt steering.

Per-trjectory Trigger Strategies

Risk triggering is based on proxy signals derived from observable motion patterns, which provide measurable indicators of potential hazardous interactions. To investigate how these signals we

introduced above contribute to risk detection, four trigger strategies are defined, including individual signals and their logical combinations. This formulation enables a systematic comparison of different signal usage and integration schemes for risk triggering in the absence of explicit risk labels.

Let $I \in \{0, 1\}$ and $B \in \{0, 1\}$ denote the binary activation of the Interaction-based Collision Signal and the Behavioral Response Signal, respectively. These binary outcomes are determined by comparing continuous indicators against thresholds that vary according to the type of agents involved in a potential conflict (e.g., vehicle–pedestrian, vehicle–bicycle, or vehicle–vehicle), based on prior studies and empirical adjustment. Based on these two signals, four trigger strategies are defined:

$$\begin{aligned} \text{I-only:} \quad & T = I \\ \text{B-only:} \quad & T = B \\ \text{OR-logic:} \quad & T = I \vee B \\ \text{AND-logic:} \quad & T = I \wedge B \end{aligned} \tag{1}$$

where $T \in \{0, 1\}$ represents the final warning trigger.

Multimodal Trigger Aggregation

Multimodal trajectory prediction typically produces multiple plausible future motion hypotheses for each agent. However, safety warning systems require a single, actionable decision rather than mode-specific outputs.

To aggregate risk information across multiple predicted trajectories into a unified risk trigger, we adopt the Top-K aggregation strategy, which triggers a risk warning if any of the top-K most probable trajectories satisfies the risk condition. In this study, our prediction algorithm generates six trajectories per agent. Accordingly, we evaluate the Top-K logic with $K = 1, 2, 3, 4, 5, 6$ to systematically analyze how including more predicted modes affects trigger activation and detection performance.

To further analyze the inter-modal consistency of the model's risk assessment, we additionally examine the k-modal risk support distribution, which characterizes how many predicted modes independently indicate a potential risk. This analysis provides insight into the consistency of risk predictions across multiple modes and reveals how sensitive the trigger results are to prediction uncertainty and mode diversity, enabling a systematic trade-off analysis between recall and false alarms.

3 Experiments and analysis

Experimental Setup

Dataset and Preprocessing: The experiments are conducted on the Waymo Open Motion Dataset (Waymo LLC, Google Brain, 2021), which provides trajectory and state information for traffic participants in mixed traffic scenarios, including vehicles and vulnerable road users.

Since only processed trajectory data are available, a light data quality control procedure is applied before analysis. Unrealistic interactions caused by perception noise (e.g., abnormally small inter-object distances are filtered using geometric plausibility checks and TTC, while predicted trajectories exhibiting abnormal kinematic behavior, such as near-zero motion or excessively high velocities, are identified as low-quality and their contribution to safety indicator evaluation is suppressed. These steps ensure reliable risk assessment without altering the overall data distribution, as low-quality trajectories are retained but their influence is suppressed to avoid biasing the resulting risk-triggering outcomes.

Case Illustration: Fig. 2 illustrates a representative traffic scenario involving a potential conflict between a left-turning cyclist and a straight-moving vehicle, highlighting the interactions between agents and the key points considered for risk assessment. To be specific, The six predicted trajectories of the ego agent (the cyclist) for 30 frames (3 s) are depicted in a warm orange gradient, along with the surrounding agents. Along each predicted trajectory which generates a risk trigger (solid lines), the point at which the predicted minimum TTC occurs is marked, as well as the locations corresponding to the predicted maximum deceleration and steering rate. In addition, the interacting agent (the vehicle) associated with the minimum TTC is highlighted, together with its position at the corresponding time step.

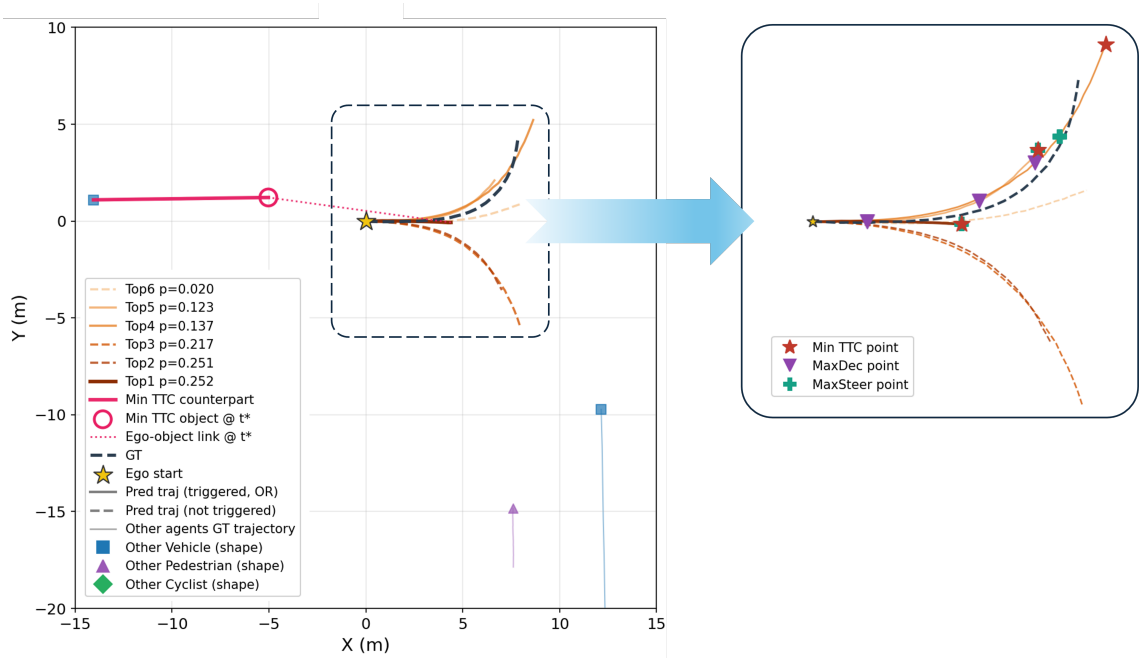


Figure 2: A representative traffic scenario

Signal Thresholding: Thresholds for converting safety indicators into actionable risk signals are selected based on prior studies and empirical observations (Lubbe & Rosén, 2014; Lenard et al., 2018; Sun et al., 2022; Nadimi et al., 2020), and are adapted to different interaction types. Detailed parameter settings are provided in the Appendix.

Evaluation Metrics: Since explicit annotations of safety-critical events are unavailable, the absolute correctness of risk detection cannot be established. Therefore, the evaluation focuses on the relative effectiveness of different trigger strategies under a unified definition of risk, by comparing the consistency between risk triggers derived from predicted trajectories and those obtained from ground-truth trajectories.

To systematically assess each strategy, standard classification metrics are adopted, including recall (R), precision (P), and F1 score (Powers, 2011):

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2(P \times R)}{P + R} \tag{4}$$

where TP and FP refer to the examples correctly predicted as positives and negatives, respectively, and FN refers to the positive examples incorrectly predicted as negatives.

These metrics provide a quantitative interpretation of the operational objective of the warning system: maximizing recall to ensure safety, maintaining adequate precision to preserve user acceptance, and capturing the trade-off between these two aspects with the F1 score as a balanced measure of overall performance. Using these measures, the effectiveness of the four risk trigger strategies can be compared systematically.

Trigger Strategy Evaluation

Fig. 3 shows the Precision-Recall curves under four strategies, and Fig. 4 presents the $F1$ score under different triggering strategies.

Among the four trigger strategies, it is observed that the **OR-logic** strategy achieves the highest recall and precision, indicating that its warnings largely correspond to actual risks and few hazardous events go undetected. In contrast, the **AND-logic** strategy exhibits relatively low recall and precision, issuing fewer warnings than there are actual risks; this conservatism reduces false alarms but increases the likelihood of missed alerts. The **I-only** strategy demonstrates high recall

but moderate precision, reflecting strong sensitivity to potential conflicts while producing a moderate number of false positives. Conversely, the **B-only** strategy achieves high precision at the cost of substantially lower recall, suggesting that behavioral cues provide reliable evidence of risk but may fail to capture many hazardous situations. Overall, these results highlight the trade-offs among sensitivity, specificity, and coverage across strategies.

As shown in Fig. 4, the OR-logic strategy achieves the highest F1-score, confirming its superior ability to balance safety and reliability. In contrast, the AND-logic strategy yields the lowest F1-score, indicating that its conservative nature leads to insufficient detection of hazardous events despite reduced false alarms. The I-only and B-only strategies attain moderate F1-scores, reflecting their complementary but imbalanced characteristics—one being more sensitive and the other more specific.

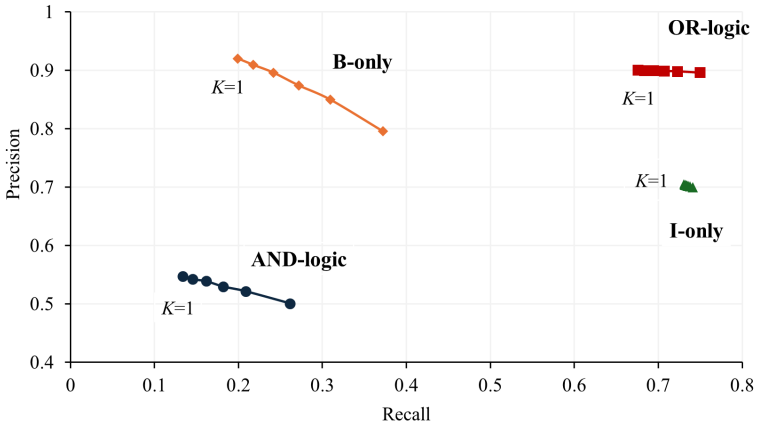


Figure 3: Precision-Recall curves under four strategies

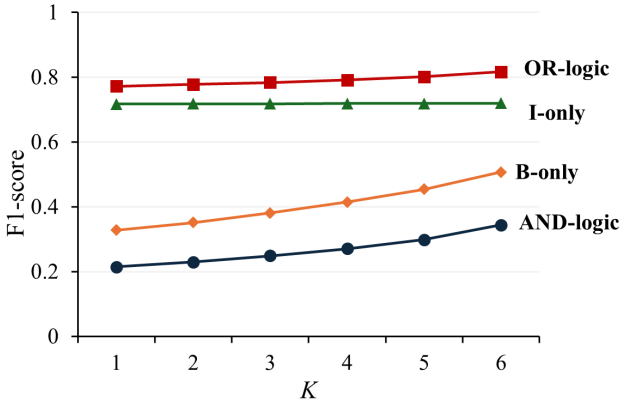


Figure 4: Comparison of F1 Scores across Top-K Trajectory Aggregation Strategies

Top-K Aggregation Analysis

To further investigate how risk information is aggregated across multiple predicted trajectories, the impact of the Top-K aggregation strategy on system performance is analyzed. As presented in Fig. 4, as K increases, both recall and $F1$-score consistently improve, indicating that considering more predicted trajectories increases the likelihood of capturing potential hazardous events. In contrast, precision gradually decreases with increasing K, as a larger set of trajectories also introduces more false positives, leading to more frequent but less reliable warnings.

Moreover, the OR-logic and I-only strategies exhibit relatively stable precision and $F1$-score across different values of K, suggesting that these strategies are less sensitive to the choice of K and maintain consistent performance. This behavior indicates that their trigger mechanisms are inherently more robust to variations in the number of considered trajectories.

Table 1 summarizes the frequency of risk-triggering trajectories across probability-ranked predicted

modes for the AND- and OR-logic strategies. Here, frequency is calculated as the number of triggered trajectories at each rank divided by the total number of triggering cases under that strategy.
For both strategies, lower-probability predicted trajectories tend to generate risk triggers more frequently. This indicates that abnormal or extreme predicted modes, although less likely, are more prone to satisfy risk conditions, underscoring the importance of considering multiple plausible futures in proactive warning systems. Notably, OR-logic exhibits consistently high triggering frequencies, reflecting its more aggressive detection behavior, while AND-logic shows a gradual increase in frequency with decreasing trajectory probability, highlighting its conservative nature.

Table 1: Frequency of risk-triggering trajectories by probability-ranked predicted modes for AND- and OR-logic strategies

Rank (by probability)	AND-trigger Frequency (%)	OR-trigger Frequency (%)
1	46.72	89.79
2	47.10	89.90
3	48.02	90.21
4	50.90	90.65
5	54.32	91.41
6	65.39	92.69

Application-Oriented Discussion

Table 2 provides a qualitative summary of these characteristics for quick reference.

Table 2: Qualitative comparison of different strategies

Strategy	Safety (R)	Reliability (P)	Overall ($F1$)	Observation
I-only	Relatively high	Relatively low	Medium	Sensitive to spatial conflicts
B-only	Relatively low	Relatively high	Medium	Conservative behavior-based detection
OR-logic	High	Low	Best	Maximizes coverage but increases false alarms
AND-logic	Low	High	Worst	Reduces false alarms but misses some risks

In practical deployment, safety is the foremost concern for a proactive warning system. Among the evaluated strategies, the OR-logic approach demonstrates the most consistent coverage of potential hazards, while the AND-logic and single-signal strategies exhibit trade-offs between sensitivity (safety) and specificity (reliability). Analysis of Top-K aggregation shows that, while increasing K generally improves recall, it can also elevate false alarm rates; notably, the OR-logic and I-only strategies exhibit relatively stable performance across different K values. These findings highlight that, in practical deployment, both the choice of trigger strategy and the selection of an appropriate K should be guided by application requirements to ensure reliable hazard detection while mitigating unnecessary warnings.

4 Conclusions

This study presents a framework for leveraging multimodal trajectory predictions to generate early safety risk warnings in mixed traffic scenarios. A structured four-step pipeline was developed, including trajectory prediction, per-trajectory risk evaluation, application of multiple trigger strategies, and aggregation across predicted modes. Two complementary risk signals—spatial interaction and behavioral response—were employed, and four trigger strategies, including single-signal and logical combinations, were systematically compared.

Experimental results demonstrate the distinct characteristics and trade-offs of different strategies. The OR-logic approach consistently achieves the highest recall and F1-score, providing robust coverage of potential hazards, while the AND-logic and single-signal strategies exhibit complementary strengths and limitations. Top-K aggregation analysis further highlights that performance trends vary with the number of considered predicted modes, indicating that both strategy selection and parameter choice play a critical role in practical deployment.
Future work will focus on incorporating trajectory datasets with explicit risk labels. Such data will enable systematic calibration of thresholds and aggregation parameters to reduce false alarms while maintaining high hazard detection, thereby further enhancing the operational effectiveness of the proposed warning system.

Acknowledgements

The work was carried out as part of the PROTEKT project. This project PROTEKT (Grant Nr.: 100757572) is co-funded by the European Union and co-financed from tax revenues on the basis of the budget adopted by the Saxon State Parliament.

Appendix

Tables A1 and A2 present the threshold settings for risk signals used in this study. These values are determined based on prior studies on traffic safety and interaction behavior (Lubbe & Rosén, 2014; Lenard et al., 2018; Sun et al., 2022; Nadimi et al., 2020), and further refined through observations and analysis of the collected trajectory data.

Table A1: Behavioral response signal thresholds by agent type

Type	MaxDec threshold (m/s^2)	MaxSteer threshold (deg/s)
Vehicle	3.000	15.0
Pedestrian	1.800	25.0
Cyclist	2.200	20.0

Table A2: Interaction-based collision signal thresholds by agent type

Ego → Other	TTC threshold (s)
Vehicle → Vehicle	2.0
Vehicle → Pedestrian	2.5
Vehicle → Cyclist	2.2
Pedestrian → Vehicle	2.5
Pedestrian → Pedestrian	1.8
Pedestrian → Cyclist	2.2
Cyclist → Vehicle	2.2
Cyclist → Pedestrian	2.2
Cyclist → Cyclist	2.0

References

Bai, Y., Meng, W., Gong, Y., Li, Y., & Gong, L. (2025). Advanced real-time highway risk early warning system integrating comprehensive weighting techniques. *International Journal of Intelligent Transportation Systems Research*, *23*(2), 1145–1162. Retrieved from `https://doi.org/10.1007/s13177-025-00505-9` doi: 10.1007/s13177-025-00505-9

Chen, Y.-L., Shen, K.-Y., & Wang, S.-C. (2013). Forward collision warning system considering both time-to-collision and safety braking distance. In *2013 ieee 8th conference on industrial electronics and applications (iciea)* (p. 972-977). doi: 10.1109/ICIEA.2013.6566508

Hydén, C. (1987). *The development of a method for traffic safety evaluation: The swedish traffic conflicts technique*. Ch. Hydén. Retrieved from https://books.google.de/books?id=XGRfNAAACAAJ

Kiefer, R. J., LeBlanc, D. J., & Flannagan, C. A. (2005). Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments. *Accident Analysis Prevention*, *37*(2), 295-303. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001457504000946 doi: https://doi.org/10.1016/j.aap.2004.09.003

Kim, J., & Kum, D. (2018). Collision risk assessment algorithm via lane-based probabilistic motion prediction of surrounding vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *19*(9), 2965-2976. doi: 10.1109/TITS.2017.2768318

Lenard, J., Welsh, R., & Danton, R. (2018). Time-to-collision analysis of pedestrian and pedal-cycle accidents for the development of autonomous emergency braking systems. *Accident Analysis Prevention*, *115*, 128-136. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001457518300964 doi: https://doi.org/10.1016/j.aap.2018.02.028

Lubbe, N., & Rosén, E. (2014). Pedestrian crossing situations: Quantification of comfort boundaries to guide intervention timing. *Accident Analysis Prevention*, *71*, 261-266. Retrieved from https://www.sciencedirect.com/science/article/pii/S0001457514001717 doi: https://doi.org/10.1016/j.aap.2014.05.029

Nadimi, N., Ragland, D. R., & Amiri, A. M. (2020). An evaluation of time-to-collision as a surrogate safety measure and a proposal of a new method for its application in safety analysis. *Transportation Letters*, *12*(7), 491–500. Retrieved from https://doi.org/10.1080/19427867.2019.1650430 doi: 10.1080/19427867.2019.1650430

Powers, D. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63.

Sun, S., Zhang, Z., Zhang, Z., Deng, P., Tian, K., & Wei, C. (2022). How do human-driven vehicles avoid pedestrians in interactive environments? a naturalistic driving study. *Sensors*, *22*(20). Retrieved from https://www.mdpi.com/1424-8220/22/20/7860 doi: 10.3390/s22207860

Varadarajan, B., Hefny, A., Srivastava, A., Refaat, K. S., Nayakanti, N., Cornman, A., ... Sapp, B. (2021, nov). MultiPath++: Efficient Information Fusion and Trajectory Aggregation for Behavior Prediction. *2022 International Conference on Robotics and Automation (ICRA)*, 7814–7821. Retrieved from https://ieeexplore.ieee.org/document/9812107/http://arxiv.org/abs/2111.14973 doi: 10.1109/ICRA46639.2022.9812107

Waymo open motion dataset. (2021). https://waymo.com/open/. (Accessed: 2024-12-12)

Weber, B., Dangelmaier, M., Diederichs, F., & Spath, D. (2020). User rating and acceptance of attention-adaptive driver safety systems. *European Transport Research Review*, *12*(1), 26. Retrieved from https://doi.org/10.1186/s12544-020-00414-w doi: 10.1186/s12544-020-00414-w

Wogalter, M. S., Godfrey, S. S., Fontenelle, G. A., Desaulniers, D. R., Rothstein, P. R., & Laughery, K. R. (1987). Effectiveness of warnings. *Human Factors*, *29*(5), 599-612. Retrieved from https://doi.org/10.1177/001872088702900509 doi: 10.1177/001872088702900509

Wu, C., Peng, L., Huang, Z., Zhong, M., & Chu, D. (2014). A method of vehicle motion prediction and collision risk assessment with a simulated vehicular cyber physical system. *Transportation Research Part C: Emerging Technologies*, *47*, 179-191. Retrieved from https://www.sciencedirect.com/science/article/pii/S0968090X14002046 (Special Issue: Emerging Technologies Special Issue of ICTIS 2013 – Guest Editors: Liping Fu and Ming Zhong and Special Issue: Visualization Visual Analytics in Transportation – Guest Editors: Patricia S. Hu and Michael L. Pack) doi: https://doi.org/10.1016/j.trc.2014.07.002

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. doi: 10.1109/TITS.2017.2699191

Zhang, L., Yuan, K., Chu, H., Huang, Y., Ding, H., Yuan, J., & Chen, H. (2022). Pedestrian collision risk assessment based on state estimation and motion prediction. *IEEE Transactions on Vehicular Technology*, *71*(1), 98-111. doi: 10.1109/TVT.2021.3127008