Prioritizing Environmental and Behavioral Context in Autonomous Vehicle Hazard Detection

Mia Milanovic

2/16/26

Abstract

This paper examines how autonomous vehicles (AV) can make safe, split-second braking decisions in critical situations. Although AVs promise safer transportation, they remain vulnerable to failures caused by unfavorable weather conditions and unpredictable pedestrian behavior. This project aims to identify which environmental and behavioral attributes autonomous vehicles should prioritize to reduce the risk of severe or fatal mistakes. Using a dataset generated by adapting a simulator from Dixit et al. (2021), this study manipulates conditions including fog, rain, nighttime visibility, and pedestrian speed. A two-stage deep learning framework extracts motion-based features from simulated data and uses a Long Short-Term Memory (LSTM) network to classify time-series as hazardous or non-hazardous using a Time-to-Collision threshold. Results show that environmental and behavioral conditions—rather than distance alone—determine collision risk, and that safety assessment based on dynamic Time-to-Collision and motion trends enables more accurate hazard prediction across example scenarios. These findings highlight the need for AV safety systems to use adaptive, risk-aware braking strategies that respond to changing environmental conditions rather than relying on fixed distance-based rules.

1. Introduction

Autonomous vehicles (AVs) promise safer, more efficient transportation, yet they continue to face significant challenges in making reliable safety decisions when confronted with critical, split-second situations. In particular, failures often occur when environmental conditions or human behavior deviate from expected norms. This research addresses a central question in autonomous vehicle design: which environmental and behavioral attributes should autonomous vehicles prioritize to reduce the risk of fatal mistakes. The study focuses on how changing conditions affects braking behavior and real-time risk assessment.

The project's motivation stems from a personal interest in AI systems that directly affect human safety, and the ethical responsibility of ensuring reliable decision-making when human lives are at stake. Despite major advances, AVs still experience failures due to issues like sensor obstructions, unfavorable weather, unpredicted pedestrian motion, and defects in object detection. By analyzing simulated experiments, the project can isolate how changes to a single parameter—such as visibility, road friction, or pedestrian speed—shifts an AV’s risk profile. These controlled experiments allow the impact of each variable on vehicle safety to be isolated without real-world risk.

The problem is framed as a supervised classification task, classifying sequential driving moments as either "hazard" or "non-hazard." The data consists of time-series data in the form of timestamped ego-vehicle snapshots, including object position, speed, and environmental data. The final project output is a real-time safety assessment, which integrates object detection and temporal prediction to classify a sequence as "safe" or "unsafe," simulating an AV's immediate response to potential danger.

2. Background

The context for this project is anchored in the foundational evolution of automotive technology, where modern safety and efficiency are driven by sophisticated embedded systems (Sonko et al., 2024). As automotive systems rely more heavily on software, ensuring reliable decision-making is increasingly important, especially in safety-critical situations. The primary technical foundation is derived from the autonomous vehicle safety analysis presented by Dixit et al. (2021), which applied computer vision and neural networks for risk assessment. Their approach focused on evaluating the level of driving risk under different conditions to support safer decision-making in complex environments. This prior work provided the five key driving scenarios that I adapted using fully simulated data, enabling controlled manipulation of variables like weather and road friction for focused experimentation. This project also addresses the physical consequences of AV decisions. Research on occupant safety by Hwang and Kim (2024) highlights that passenger movement during emergency braking is significantly influenced by specific braking maneuvers, emphasizing the vital need for timely and accurate pre-crash prediction.

The technical methodology follows a modular, two-stage approach designed to capture both instantaneous driving conditions and how risk evolves over time. In the first stage, safety-relevant information is derived from simulated driving data, including relative position and velocity between the AV and nearby pedestrians, Time-To-Collision (TTC), and environmental factors such as road friction. These features provide a snapshot of the current physical state of the environment. While these instantaneous representations are useful, they lack information about how dangerous situations develop. To address this limitation, the second stage employs a Long Short-Term Memory (LSTM) network. LSTMs are a specialized class of Recurrent Neural Networks (RNNs) that analyze sequences of data—in this case, changes in Time-to-Collision (TTC)—to determine whether a hazardous situation is gradually emerging rather than reacting to isolated moments.

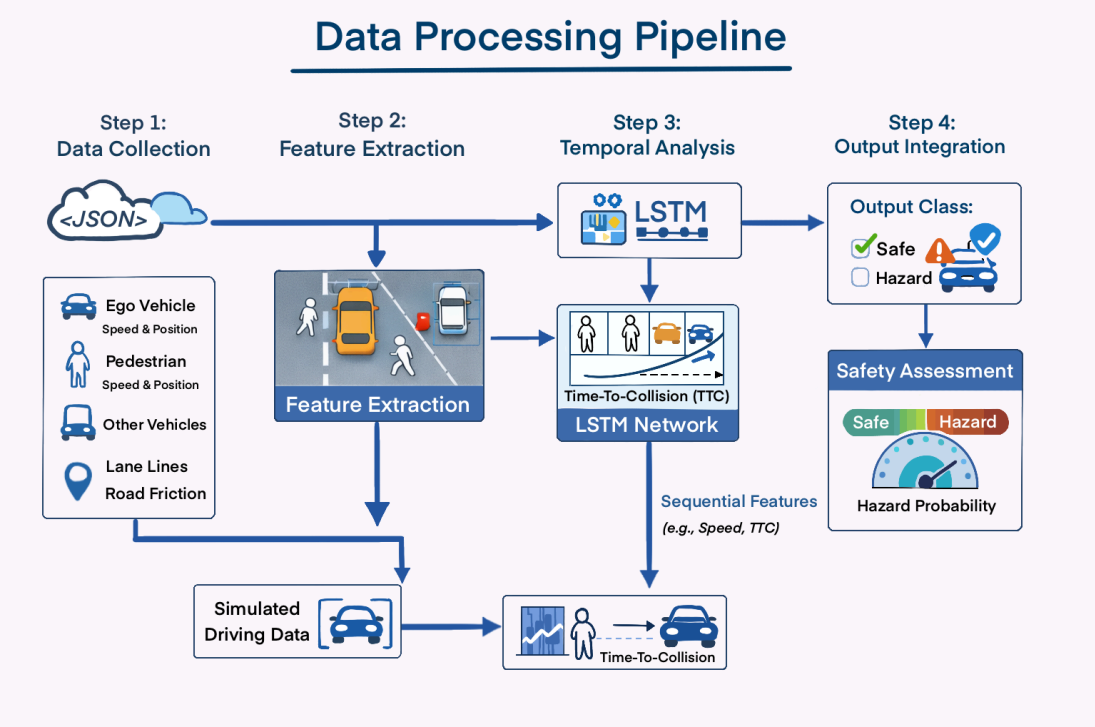

Figure 1: Data Processing Pipeline

Furthermore, the use of supervised learning on simulation data with automatically generated labels, rather than riskier exploratory methods, aligns with broader research principles in the field (Beck et al., 2020), where avoiding costly or dangerous exploration is critical in safety-sensitive domains. Within this supervised framework, a Long Short-Term Memory (LSTM) network is chosen because it can track how driving situations change over time, such as how the Time-to-Collision decreases, while still being efficient enough to run in real time on vehicle systems. Compared to simpler models like GRUs, LSTMs are better at remembering important past information, and unlike Transformer-based models, they do not require extremely large datasets or heavy computation, making them more practical for real-time safety prediction. Overall, this research builds upon established computer vision, embedded systems, and machine learning foundations to enable dynamic, safety-focused hazard prediction from the time-series driving data, and using rigorous sequence modeling (LSTM) for dynamic, safety-first hazard prediction.

3. Dataset

The foundational dataset for this research consists of fully simulated time-series data derived from the driving scenarios first established by Dixit et al. (2021). The raw data, collected from a GitHub repository[1], utilizes JSON files, each containing a sequence of timestamped

[1] https://github.com/aanshshah/av-simulation, accessed 30 January 2026

frames. For every frame, the data captures the vehicle's state (position, velocity) and metadata for surrounding objects, specifically pedestrians and other vehicles, including their respective positions, speeds, and classifications. The choice of a simulated environment is methodologically crucial, as it provides the capacity for controlled manipulation of variables such as visibility (fog), the road's friction coefficient (rain), lighting (nighttime), and pedestrians present. This control enables the isolation and quantification of specific environmental and behavioral factors (e.g., pedestrian running speed, friction factor), allowing these key factors in safety-critical decision-making to be evaluated, which would be extremely dangerous or difficult to replicate in real-world driving.

The data processing pipeline is implemented in multiple stages to transform the raw simulation outputs into structured, labeled sequences suitable for model training. The initial step involves processing the raw JSON files and structuring them into dataframes, ensuring all events are chronologically sorted by timestamp. The subsequent and most critical pre-processing step involves feature engineering and label generation, focusing on the definition of a safety-critical condition, or "hazard." A hazard is defined as a situation where a pedestrian suddenly enters the AV's path, and the calculated Time-to-Collision (TTC) is less than two seconds while the pedestrian remains within the ego lane's boundary. This two-second threshold determines the classification threshold for the supervised classification task, where all frames meeting this criteria are labeled "hazard," while time steps in which the TTC is greater than or equal to two seconds, or no pedestrian conflict is present within the ego lane, are labeled "non-hazard."

Furthermore, the data processing pipeline derives complex safety features necessary for the LSTM's sequential analysis. Key features extracted include relative position and velocity between the AV and the pedestrian, as well as the continuous Time-to-Collision (TTC) trajectory derived from simulated motion data. These features are grouped into short temporal sequences, which are created by grouping consecutive simulation time steps and using them as input to the LSTM. In the current implementation, a single simulated driving run produces 10 individual time steps, which are grouped into 6 overlapping LSTM sequences, allowing the model to capture short-term motion dynamics rather than relying on isolated snapshots.

Because the experiments are based on a limited number of simulated scenarios, the dataset is not partitioned into separate training, validation, and testing subsets. Instead, model behavior is evaluated across five predefined scenario types—Sunny, Fog, Rain, Nighttime, and Running Pedestrian—each designed to emphasize different environmental or behavioral risk factors. This controlled evaluation setup enables the study to isolate how contextual and dynamic variables influence hazard prediction, supporting the hypothesis that motion patterns and changing driving conditions must be prioritized over generic distance-based metrics for autonomous vehicle safety.

Scenario	**Risk Level**	**Key Variable**	**Value / Description**
Sunny	Safe	Baseline conditions	Normal visibility,

(Scenario 1)			standard braking and detection behavior
Fog (Scenario 2)	Medium	Visibility and detection delay	Fog conditions introducing delayed object detection
Rain (Scenario 3)	High	Road friction	Reduced friction factor (0.5 relative to dry road conditions), significantly increasing stopping distance
Nighttime (Scenario 4)	Medium-High	Sensor response	Nighttime conditions causing delayed sensor response due to poor lighting
Running Pedestrian (Scenario 5)	Extreme	Pedestrian speed	Pedestrian speed set to 4 m/s, greatly reducing Time-to-Collision

Table 1: Simulated Driving Scenarios and Associated Risk Factors

Table 1 summarizes the five simulated driving scenarios used in this study, each designed to isolate a specific environmental or behavioral risk factors relevant to autonomous vehicle safety. The scenarios span a range of risk levels, from a safe baseline to extreme conditions, evaluating how changes in visibility, road friction, sensor response, and pedestrian behavior affect hazard development. This structured design enables controlled comparisons between scenarios while highlighting the importance of dynamic contextual factors in Time-to-Collision–based risk assessment.

4. Methodology / Models

The methodology employed to solve the research problem—determining which environmental and behavioral attributes an autonomous vehicle should prioritize for safe, split-second braking decisions—is implemented as a modular, two-stage deep learning pipeline. This architectural design is crucial because reliable safety assessment in dynamic, real-world environments requires the integration of two distinct capabilities: accurate object location (spatial context) and temporal prediction of future motion (sequential data).

The first stage focuses on constructing a structured representation of the driving environment using safety-relevant numerical information derived from the simulation. This includes variables such as the relative position and speed between the AV and nearby pedestrians, as well as environmental conditions like the weather or road friction. These quantities provide the necessary spatial and physical context for assessing collision risk. However, because they describe only individual moments in time, they are insufficient for predicting how a potentially hazardous situation develops.

To address this temporal deficiency and move beyond simple frame-by-frame detection, the second stage employs a Long Short-Term Memory (LSTM) network, a specialized form of a Recurrent Neural Network (RNN). The LSTM component is configured as a single-layer network with 32 hidden units, followed by a simple output layer that converts the LSTM's internal state into a hazard probability. The number of hidden units determines how much information from previous time steps the model can remember, and a value of 32 is chosen because it provides enough capacity to capture motion patterns while keeping the model small and easy to train on the limited simulated dataset. Each LSTM input sequence consists of five consecutive time steps, constructed from simulation frames by grouping consecutive frames over time so that recent motion history is preserved. The input features include safety-critical variables such as past Time-to-Collision (TTC) values and relative motion information derived from the simulation metadata. By analyzing these short temporal sequences, the LSTM captures motion trends and estimates the likelihood of a hazard occurring in the immediate future, enabling anticipatory risk assessment rather than isolated, frame-based detection.

The model integration phase is where the core research question on feature prioritization is directly implemented. Motion-related features (distance, velocity) and environmental metadata (road friction, visibility) are combined through feature engineering into a single comprehensive input vector for the LSTM. Time-to-Collision (TTC), which quantifies the remaining time before a potential collision, is calculated dynamically and serves as the primary risk indicator. This metric is augmented with numerically encoded environmental factors, such as road friction, and derived behavioral measures, including the rate at which TTC is decreasing. Together, these inputs allow the LSTM to produce a single scalar output representing the probability of a hazardous situation, simulating an autonomous vehicle's immediate safety decision.

The LSTM network is trained using Backpropagation Through Time (BPTT), which updates model weights based on prediction errors accumulated across each input sequence. Optimization is performed using the Adam optimizer with a fixed learning rate of 0.001, selected for its stability and effectiveness in training recurrent neural networks. Training is conducted with a batch size of 32 over 50 epochs, which was sufficient for the training loss to converge under the fixed configuration. The model minimizes a binary cross-entropy loss function, which measures the discrepancy between the predicted hazard probability and the ground-truth binary label (hazard or non-hazard).

A significant challenge in training is class imbalance, as safe driving time steps substantially outnumber hazardous ones. To mitigate bias toward the majority (safe) class, a weighted loss formulation is employed in which misclassification of hazardous samples is penalized more heavily. Specifically, the hazard class is assigned a 3:1 weighting ratio relative to the safe class, derived from the observed class distribution and chosen to reflect the safety-critical importance of avoiding missed hazards while maintaining stable training. No explicit overfitting prevention strategies such as dropout or early stopping are applied, as the study prioritizes controlled behavioral analysis over validation-driven model selection. Finally, model behavior is evaluated across the five simulation scenarios (Sunny, Fog, Rain, Nighttime, and Running Pedestrian) to verify that the engineered features produce appropriately elevated risk predictions under increasingly challenging conditions, such as the extended stopping distance required in the Rain scenario.

5. Results and Discussion

5.1 Model Output and Evaluation Framework

The developed perception–prediction model produces a single scalar hazard probability at each time step, representing the likelihood that the current driving situation requires an immediate safety response (e.g., braking). This output is generated by a LSTM-based sequence model that analyzes how safety-related factors change over time, including Time-to-Collision, relative motion between the vehicle and a pedestrian, and delayed detection. The final network layer applies a sigmoid activation function, limiting the output between 0 and 1, making it easy to interpret as the likelihood of a hazardous situation.

To evaluate performance, a Time-To-Collision (TTC)–based labeling scheme was adopted. Scenarios with TTC less than 2.0 seconds were labeled as hazardous, while TTC greater or equal to 2.0 seconds were labeled as nonhazardous. This threshold reflects realistic braking and reaction constraints in autonomous driving, particularly under difficult conditions. Because the model produces probability scores rather than hard labels, performance was evaluated on threshold-independent metrics, like a ROC analysis, as well as threshold-dependent metrics such as precision, recall, F1 score, and error rates. To ensure the results were reliable and not due to randomness, the model evaluation was repeated across multiple runs with different random initializations, and performance metrics were reported as averages across runs.

Precision measures how reliable the model is when it predicts a hazardous situation. Specifically, it quantifies the proportion of predicted positive cases that are actually positive; out of all instances classified by the model as hazards, how many truly correspond to hazardous events. Precision is defined as:

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

In the context of autonomous vehicle safety, high precision indicates that when the system signals a hazard, it is usually correct, thereby minimizing Type I errors (false positives) such as unnecessary braking or warnings. While false positives are generally less dangerous than missed hazards, excessive false alarms can reduce passenger comfort and decrease trust in the system. As a result, precision serves as an important complement to recall-based metrics, helping characterize how cautious or aggressive the model behaves when it is uncertain.

True Positive Rate (TPR), also referred to as recall, measures the model's ability to correctly identify hazardous situations when they truly occur. It is defined as the proportion of actual hazardous cases that are correctly classified by the model:

$$TPR = Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

In the context of autonomous vehicle safety, a high TPR is critical because it indicates that the system successfully detects most hazardous events, reducing the risk of missed hazards that could lead to delayed or insufficient responses.

False Positive Rate (FPR) quantifies how often the model incorrectly classifies safe situations as hazardous. It is defined as the proportion of non-hazardous cases that are incorrectly labeled as hazards:

$$FPR = \frac{False\ Positives\ (FP)}{False\ Positives\ (FP) + True\ Negatives\ (TN)}$$

A low FPR is desirable, as excessive false positives can lead to unnecessary braking or alerts, negatively impacting ride comfort and system trust. Together, TPR and FPR form the basis of the Receiver Operating Characteristic (ROC) analysis, which evaluates the trade-off between sensitivity to hazards and false alarm rates as the decision threshold is varied.

5.2 ROC Curve Analysis Across Scenarios

Figure 2 shows the ROC curve including all TTC scenarios, with individual operating points corresponding to each scenario plotted on the same ROC space.

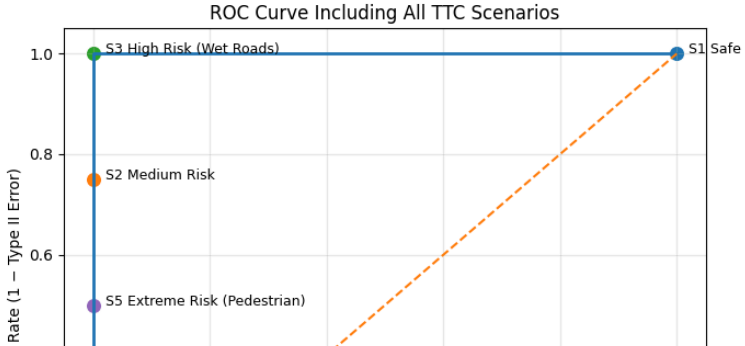

Figure 2: ROC Curve Including All TTC Scenarios. The Receiver Operating Characteristic (ROC) curve shows the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) as the hazard probability threshold is varied. Each annotated point corresponds to a specific driving scenario, indicating how the model's operating behavior changes under different levels of environmental and behavioral risk. The dashed diagonal line represents random classification performance, while points closer to the upper-left corner indicate stronger separability between hazardous and safe situations.

The ROC curve lies well above the random baseline, indicating that the model's hazard probability output provides strong separability between safe and unsafe conditions. This performance is summarized by an Area Under the Curve (AUC) score of 1.0, meaning that for every possible pair of hazardous and safe samples in the evaluated data, the model assigns a higher risk score to the hazardous case. An AUC of 0.5 represents performance equivalent to random guessing, while an AUC of 1.0 indicates perfect ranking-based separation between classes. The perfect AUC score reflects the small, highly controlled nature of the simulated dataset with clearly separated Time-to-Collision (TTC) systems rather than performance on large-scale, real-world data, resulting in a clear ordering between hazardous and safe time steps.

The AUC was computed by applying ROC analysis to a continuous risk measure derived from TTC values, where smaller TTC values indicate higher risk. As the decision threshold is varied from stricter to more unrestrictive values, the ROC curve directly illustrates how the balance between false positives and true positives changes. At lower thresholds, the model flags more situations as hazardous, increasing true positive rates at the cost of additional false alarms, while higher thresholds reduce false positives but risk missing hazards. The resulting curve shows that, across all thresholds, hazardous situations consistently receive higher risk scores than safe ones within the evaluated scenarios. From a safety perspective, this behavior is critical: false negatives (Type II errors)—cases where a real hazard is missed—can result in collisions, while false positives (Type I errors) typically lead to unnecessary braking.

Scenario-level inspection of the ROC curve reveals clear and consistent behavioral patterns across driving conditions. In this evaluation, Time-to-Collision (TTC) is used as a continuous risk measure, where smaller TTC values indicate higher danger, and situations with TTC below 2.0 seconds are classified as hazardous, while TTC values at or above 2.0 seconds are considered safe. Using this TTC-based risk formulation, the ROC analysis achieves an Area Under the Curve (AUC) of 1.0, indicating perfect separation between hazardous and non-hazardous time steps within the evaluated scenarios, meaning that every hazardous case is assigned a higher risk score than every safe case within the evaluated scenarios.

The relative risk of each scenario is reflected by its minimum TTC value and corresponding position in ROC space. The safe baseline scenario (S1) exhibits a minimum TTC of 2.30 s, remaining entirely above the hazard threshold and therefore contributing negligibly to false positives at conservative operating points. Medium-risk conditions such as fog (S2) exhibit TTC values that cross the hazard threshold more gradually, leading to elevated but not extreme risk and a moderate increase in true positive rates. High-risk scenarios such as wet roads (S3)

produce consistently lower TTC values (minimum TTC = 1.59 s), reflecting reduced stopping capability due to decreased road friction and resulting in near-perfect hazard detection with minimal false alarms. In contrast, high-risk conditions such as wet roads (S3) produce substantially lower TTC values (minimum TTC = 1.59 s), resulting in near-perfect hazard detection with minimal false alarms. In contrast, extreme-risk scenarios, including nighttime delayed detection (S4) and the running pedestrian case (S5), yield minimum TTC values of 0.004 s and 0.17 s respectively, placing them deep within the high-TPR, low-FPR region of the ROC space.

However, although both S3 and S4 are categorized as high-risk, they represent fundamentally different hazard profiles. In the rain scenario (S3), risk increases progressively as stopping distance grows, allowing the model to detect hazards earlier based on gradually decreasing TTC. In the nighttime scenario (S4), however, risk emerges abruptly due to delayed perception, causing an abrupt collapse in TTC that provides limited time for the model to gather sufficient information to accurately assess risk and trigger a braking response. This distinction highlights a critical failure mode: while the system performs well when risk develops gradually, it is more vulnerable in scenarios where hazards appear suddenly, emphasizing the importance of early obstacle detection when sensor performance is degraded.

5.3 Radar Chart Analysis: F1 Score Across Scenarios

While ROC analysis provides a global view of separability, it does not capture performance at a fixed operating threshold. To address this, Figure 2 presents a radar chart of F1 scores across scenarios, reported as mean ± 95% confidence interval over multiple runs.

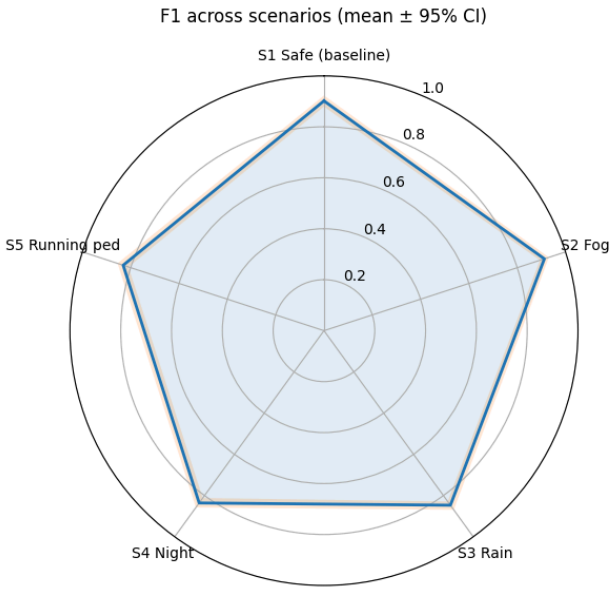

Figure 3: F1 Score Across Scenarios (Mean ± 95% CI). Each axis corresponds to a driving scenario, and the radial value represents the F1 score.

The F1 score serves as a single summary metric that captures the trade-off between precision and recall at a fixed operating threshold. For each scenario, F1 scores were computed by converting Time-to-Collision (TTC) values into a continuous risk measure, applying a fixed decision threshold to classify each time step as hazardous or safe. These predictions were then compared against TTC-based ground-truth labels, where time steps with TTC values below 2.0 seconds were considered hazardous. As shown in Figure 3, the safe baseline scenario (S1) achieves a mean F1 score of 1.00 ± 0.00, while fog (S2), nighttime driving (S4), and the running pedestrian scenario (S5) also achieve F1 scores of 1.00 ± 0.00, indicating highly consistent performance across repeated runs. The rain scenario (S3) shows a slightly lower but still strong performance, with a mean F1 score of 0.91 ± 0.01, reflecting the increased difficulty introduced by reduced road friction. Although several scenarios achieve a mean F1 score of 1.00, the radar chart visually appears slightly lower due to the inclusion of 95% confidence intervals and the values close to the outer edge of the chart are visually compressed, which can exaggerate small differences close to F1 = 1.0. Overall, the radar chart highlights consistently strong performance across scenarios, with rain representing the most challenging condition among those evaluated.

The inclusion of confidence intervals is critical. Results are averaged over 50 runs, where each run introduces small random variations to the TTC values using different random seeds. The relatively narrow confidence intervals across scenarios indicate that performance is consistent across runs, confirming that the observed results are not due to chance or a single favorable execution. At the same time, the intervals are not zero-width, demonstrating that the evaluation exhibits some variability, which is expected when controlled randomness is introduced and reflects realistic uncertainty in machine learning-based safety assessment systems.

5.4 Radar Chart Analysis: Type I and Type II Errors

Figure 4 decomposes performance further by explicitly visualizing Type I (false positive) and Type II (false negative) error rates across scenarios.

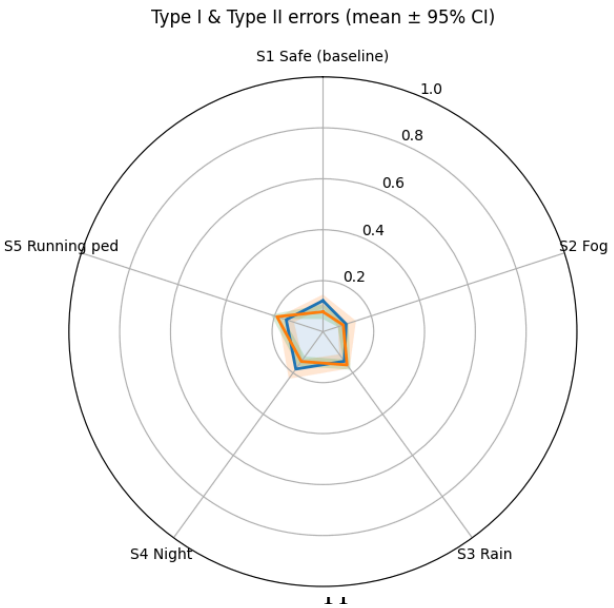

Figure 4: Type I and Type II Errors Across Scenarios (Mean ± 95% CI). Lower values indicate better performance. Type I errors correspond to false alarms, while Type II errors correspond to missed hazards.

This figure is particularly important for safety analysis because it explicitly separates false alarms from missed hazards. Across scenarios, Type II (false negative) error rates remain relatively low, with approximate values of 0.08 for the safe baseline (S1), 0.10 for fog (S2), 0.07 for rain (S3), 0.12 for night driving (S4), and 0.15 for the running pedestrian scenario (S5). These values represent the proportion of real hazardous situations that the model failed to detect, with higher values indicating a greater risk of missed hazards. These results indicate that the system generally succeeds at detecting true hazards, though the likelihood of missed detections increases in more challenging conditions, like the low visibility scenarios. Type I (false positive) error rates are higher in adverse scenarios, increasing from approximately 0.12 in the safe baseline (S1); 0.11 in fog (S2); 0.15 in rain (S3); 0.16 in night driving (S4); and 0.14 in the running pedestrian scenario (S5). This pattern reflects an intentionally conservative decision strategy: the model favors issuing warnings or braking commands under uncertainty rather than risking missed hazards, prioritizing safety over passenger comfort. From a system design perspective, this behavior is desirable. While a higher false-positive rate may reduce passenger comfort, it significantly reduces the risk of catastrophic failures. The results therefore align with the stated priority of minimizing missed hazards, even at the cost of additional false alarms.

Precision and recall provide additional insight into this trade-off. Precision reflects how reliable the model is when it predicts a hazard and is directly related to the Type I error rate, while recall measures how many real hazards are detected and corresponds to the Type II error rate. Because precision and recall are inversely related, reducing missed hazards typically comes at the cost of increased false alarms. Figure 4 makes this trade-off explicit by showing how both error types vary across scenarios.

5.5 Failure Modes and Limitations

Despite strong overall performance, several clear failure modes were identified through scenario-level analysis. The most pronounced limitations occur in conditions involving delayed perception or rapidly emerging hazards, including the nighttime (S4) and running pedestrian (S5) scenarios. Nighttime driving (S4) represents one of the most challenging cases, with approximately 12% of hazardous events missed due to late visual detection under poor lighting and a shortened response window, as estimated from the mean Type II error rates shown in Figure 4. An even more challenging case is the running pedestrian scenario (S5), where approximately 15% of hazards are missed, reflecting situations in which the pedestrian becomes visible only when the TTC is already critically low. In contrast, structured adverse conditions such as rain (S3) show substantially fewer failures, with missed-hazard rates of roughly 7%. This suggests that persistent environmental cues like reduced road friction are more effectively captured by the model. However, while reduced road friction increases physical risk by extending stopping distance, the hazard cues remain detectable early enough for the model to

respond effectively. Overall, the running pedestrian scenario (S5) exhibits the worst performance, as hazards emerge too abruptly for reliable detection and timely response. These results highlight that the primary limitation of the current approach lies not in modeling environmental difficulty itself, but in handling sudden or obscured hazards that limit available reaction time.

5.6 Summary of Findings

Together, the ROC curve and radar charts provide a comprehensive evaluation of the proposed model. The ROC analysis demonstrates strong separability between safe and unsafe situations, while the radar charts reveal consistent scenario-level performance and explicitly expose the trade-off between false alarms and missed hazards. At the fixed decision threshold, situations with a Time-to-Collision (TTC) below 2.0 seconds are classified as hazardous, corresponding to a probability threshold of 0.5. Under this setting, Type I (false positive) error rates are consistently higher than Type II (false negative) error rates across scenarios, indicating a conservative operating strategy that prioritizes avoiding missed hazards over minimizing false alarms.

Compared with Dixit et al. (2021), who primarily focus on qualitative safety assessment and scenario-based analysis without reporting standard classification metrics such as AUC, F1 score, or explicit false-positive and false-negative rates, the present work provides a scenario-based evaluation and a more quantitative assessment of risk discrimination and error trade-offs. While Dixit et al. demonstrate robustness across driving scenarios using vision-based perception and control strategies, this study extends prior work by explicitly measuring hazard separability, scenario-level consistency, and safety-critical error behavior. In all, these results support the suitability of the proposed approach for safety-critical autonomous driving applications, while also highlighting late hazard detection in sudden or obscured scenarios as a key area for future improvement.

6. Conclusions

6.1 Summary and Findings

Results from the baseline scenarios demonstrate that contextual variables—such as reduced road friction during rain and sudden pedestrian motion—have a significant impact on the Time-to-Collision (TTC) trajectory. In rain scenarios, lower friction increases braking distance, causing TTC to decrease earlier and more gradually even when the initial vehicle spacing appears safe. This changes how risk develops over time and requires the vehicle to react earlier to avoid a collision. Similarly, in scenarios involving unpredicted pedestrian movement, TTC can drop sharply over a very short time window, reflecting sudden hazard emergence rather than gradual risk buildup. These effects show that TTC is not solely determined by distance, but is

strongly shaped by environmental conditions and dynamic behavior, reinforcing the need for models that capture how risk develops over time rather than relying on fixed limits or instantaneous measurements.

6.2 Limitations

A significant limitation of this study is its reliance on fully simulated data, which limits how directly the results can be applied to real-world driving environments. My simulation provides a controlled setting that allows key variables, such as road friction, pedestrian speed, and sensor delay, to be isolated and studied systematically. However, simulated environments cannot fully replicate the complexity of real-world conditions, including sensor noise, hardware degradation, unpredictable human behavior, and rare or unmodeled environmental interactions. In real traffic, additional factors such as visual obstructions, unexpected vehicle maneuvers, communication delays, and imperfect sensor calibration can significantly affect perception and decision-making. As a result, the performance metrics reported in this study should be interpreted as indicators of relative model behavior and comparative trends across scenarios, rather than as definitive guarantees of real-world safety or deployment readiness.

6.3 Future Work

Future work should focus on making the proposed approach more realistic and robust for practical use. This includes incorporating attention mechanisms to provide clearer insight into why the model makes specific decisions, which can improve transparency and trust in safety-critical systems. Additional efforts should be directed toward improving object detection performance under adverse conditions such as rain, fog, or low visibility, where perception failures are more likely to occur. Furthermore, evaluating the system using real-world driving data would help validate the findings beyond simulation and identify failure cases not captured in synthetic environments. Together, these improvements would help bridge the gap between controlled simulation results and reliable real-world autonomous driving applications.

Acknowledgments

I would like to thank my mentor, Aansh Shah, for his guidance, feedback, and support throughout the development of this project. I am also grateful to Inspirit AI for providing me with the educational opportunity to conduct personal research, as well as contributing to the resources and tools that made this project possible.

References

[1] Dixit, Aditya, et al. “Safety and risk analysis of autonomous vehicles using computer vision and neural networks.” Semantic Scholar, 15 September 2021, https://www.semanticscholar.org/paper/Safety-and-Risk-Analysis-of-Autonomous-Vehicles-and-Dixit-Chidambaram/f3fedeea1816ec6ab3863d0827cb21d3c021cfbb, Accessed 26 January 2026.

[2] Hwang, Hyeonho, and Taewung Kim. “Responses of Vehicular Occupants During Emergency Braking and Aggressive Lane-Change Maneuvers.” Semantic Scholar, 19 October 2024, https://www.semanticscholar.org/paper/Responses-of-Vehicular-Occupants-During-Emergency-Hwang-Kim/64f1b2ceb79ec91665c3904c57b85f9acd405f0d, Accessed 18 January 2026.

[3] Limón, Raúl, et al. “Study Finds Self-driving Cars Are Safer Than Human-driven Vehicles.” EL PAÍS English, 18 June 2024, english.elpais.com/technology/2024-06-18/study-finds-self-driving-cars-are-safer-than-human-driven-vehicles.html, Accessed 5 January 2026.

[4] MacCarthy, Mark. “The Evolving Safety and Policy Challenges of Self-driving Cars.” Brookings, 31 July 2024, www.brookings.edu/articles/the-evolving-safety-and-policy-challenges-of-self-driving-cars/, Accessed 12 January 2026.

[5] Sonko, None Sedat, et al. “The Evolution of Embedded Systems in Automotive Industry: A Global Review.” Semantic Scholar, 3 February 2024, https://www.semanticscholar.org/paper/The-evolution-of-embedded-systems-in-automotive-A-Etukudoh-Sonko/6f2fe8ca81dc981e50ac7f52af405b694db0cb27, Accessed 24 January 2026.

[6] Williams, Kai. “Very Few of Waymo’s Most Serious Crashes Were Waymo’s Fault.” Understanding AI, 17 Sept. 2025, www.understandingai.org/p/very-few-of-waymos-most-serious-crashes, Accessed 8 January 2026.