Electronic Research Archive

ERA, 34(5): 3050–3078.
DOI: 10.3934/era.2026138
Received: 28 December 2025
Revised: 20 February 2026
Accepted: 26 February 2026
Published: 08 April 2026

http://www.aimspress.com/journal/era

Research article

Comprehensive risk assessment on expressway open-toll Plaza with computational learning-based accident predictions and surrogate safety measure estimations

Hojae Kim[1], Juneyoung Park[2,*], Cheol Oh[2] and Chris Lee[3]

[1] Department of Transportation and Logistics Engineering, Hanyang University, Ansan 15588, South Korea
[2] Department of Transportation and Logistics Engineering & Smart City Engineering, Hanyang University, Ansan 15588, South Korea
[3] Department of Civil and Environmental Engineering, University of Windsor, ON N9B 3P4, Canada

* **Correspondence:** Email: juneyoung@hanyang.ac.kr; Tel: +821034994156; Fax: +82314368147.

Abstract: The rapid expansion of expressway open-toll plazas, where high-speed electronic toll collection lanes operate alongside slower manual and single-lane toll lanes, has introduced complex crash risks that are difficult to capture using traditional crash-frequency models alone. Sparse crash records, heterogeneous operating conditions, and evolving tolling configurations call for data-driven approaches that integrate simulation, surrogate safety metrics, and computational learning. This study aims to perform a comprehensive risk assessment of open-toll plaza configurations by combining surrogate safety measure (SSM) estimations with computational learning (CL)-based accident prediction models. First, a broad set of geometric and operational scenarios is generated by systematically varying lane layout, toll-lane composition, merging length, speed limit, traffic volume, and heavy-vehicle share. For each scenario, microscopic traffic simulations are conducted, and SSMs—including time to collision (TTC), post-encroachment time (PET), and conflict counts—are extracted to quantify instantaneous interaction risk at the vehicle level. These SSMs serve both as comparative safety indicators and as target outputs for subsequent CL models. Next, the CL model is trained to predict conflict frequencies from design and traffic features, while a complementary parametric count model is estimated for benchmarking. To enhance interpretability, eXplainable AI (XAI) attribution techniques are used to decompose the CL predictions into feature-level contributions, revealing nonlinear and interaction effects associated with high-risk operating regimes. The results

highlight the dominant influence of traffic volume and toll lane ratio on conflict occurrence. By integrating SSM-based simulation outputs with CL and XAI, the proposed approach provides quantitative and interpretable evidence that supports safer design and operation of expressway open-toll plazas within next-generation data-driven transportation systems.

Keywords: open-toll plaza; computational learning; count-aware machine learning; explainable artificial intelligence; micro-traffic simulation

Abbreviations: CI: Computational intelligence; CL: Computational learning; CV: Cross-validation; DT: Decision tree; ETC: Electronic toll collection; ETCR: ETC ratio; GLM: Generalized linear model; HSETC: High-speed ETC; HSETCR: HSETC ratio; HVR: Heavy vehicle ratio; LGBM: Light gradient boosting machine; LL: Lane layout; LTR: Lane type provisioning rule; ML: Merging length; NB: Negative binomial; NL: Number of mainline lanes; OT: Open-toll; PET: Post-encroachment time; R^2: Coefficient of determination; RF: Random forest; SPF: Safety performance function; SHAP: SHapley additive exPlanations; SSAM: Surrogate safety assessment model; SSM: Surrogate safety measure; TCS: Toll collection system; TCSR: TCS ratio; TTC: Time-to-collision; TV: Traffic volume; WMAPE: Weighted Mean Absolute Percentage Error; XAI: eXplainable artificial intelligence; XGB: eXtreme gradient boosting.

1. Introduction

In toll-plaza operations, OT concepts aim to reduce stopping and speed loss during toll transactions to preserve mainline flow. The approach has advanced with the development of electronic tolling technologies. ETC, a core enabling technology for OT, electronically charges tolls via communication between roadside infrastructure and in-vehicle units, thereby reducing stops at plazas and improving traffic flow [1–3]. With growing traffic demand and increasing needs for higher throughput, HSETC operations have been increasingly adopted, building on ETC while designing lane width, operating speed, and separation levels to be more mainline-like and to support near free-flow passage [4–6]. This evolution in technology and operations can be viewed as a response to toll-plaza bottlenecks, aiming to reduce congestion-related societal costs and to achieve more stable traffic flow.

However, HSETC is often implemented in mixed operations where it coexists with ETC or TCS at the same plaza, and the diverging/merging area may experience degraded safety due to pronounced speed heterogeneity across lanes and concentrated lane-choice and lane-changing interactions [4–6]. For instance, when high-speed streams aiming for near free-flow passage merge with streams that require stopping or low-speed passage, acceleration–deceleration patterns become inconsistent in the merge influence area, and abrupt lane changes and maneuvers may increase. Such conditions can elevate the likelihood of crashes such as rear-end and sideswipe events, and the associated risks may be amplified nonlinearly through interactions among design and operational factors, including lane composition, merging geometry, traffic demand, and vehicle mix. Accordingly, there is a need for an analysis framework that can quantitatively diagnose risk mechanisms in mixed OT environments and enable reproducible comparisons across operating conditions.

Traditionally, road safety evaluation has relied on crash-based SPF models, commonly using count-data formulations such as Poisson or negative binomial models to estimate crash frequencies [7].

However, because HSETC has not been deployed for long, and plaza configurations and operations may change rapidly, it is difficult to secure sufficient and stable crash records at the individual-plaza level. Moreover, risks in mixed operations emerge from the accumulation of microscopic interactions within a short influence area, where high-order interactions among lane configuration, merging length, demand level, and vehicle mix can be critical. Under these conditions, conventional SPF-based approaches may be limited in simultaneously addressing data sparsity, configuration changes, and nonlinear interaction structures. As a result, diagnosing risk mechanisms and comparing design–operation alternatives using crash data alone can be constrained in mixed OT environments under HSETC operations.

Accordingly, the objective of this study is to develop a comprehensive and interpretable risk assessment framework for mixed OT plazas by integrating simulation-based SSM estimation, count-aware CL modeling of conflict frequencies, and XAI interpretation. Using systematically designed simulation scenarios spanning diverse geometric and operational conditions, we first extract conflict-based surrogate outputs [8,9]. We then formulate conflict-frequency prediction as a supervised learning task and compare traditional SPF baselines (Poisson and negative binomial), a transparent nonlinear baseline (DT), and boosting-based CL models (XGB and LGBM) under both non-count-aware and count-aware objectives [7,10–12]. Model performance is evaluated via CV, using WMAPE and R^2 as primary metrics. Finally, beyond total conflict frequency, we examine generalization across conflict categories (e.g., rear-end-like, lane-change-like, and crossing conflict) to identify risk mechanisms that are reliably functionalized by scenario-level variables and those requiring additional behavioral or geometric descriptors, thereby providing actionable evidence for safer design and operation of OT plazas [13–20].

2. Literature review

2.1. Overseas cases of high-speed electronic toll collection implementation

HSETC-based high-speed electronic tolling operations have been progressively introduced and expanded across multiple countries in response to growing demands to reduce delay and increase throughput in toll collection segments. However, real-world deployment of OT systems must accommodate non-tag vehicles and legacy plaza assets and must operate within tolling and geometric constraints. Consequently, many agencies implement mixed operations in which HSETC coexists with ETC or TCS. As a result, mixed OT operations in which HSETC coexists with ETC or TCS are commonly observed in practice. The specific configuration and the pace of transition for such mixed OT operations can vary depending on national transition strategies and operational contexts, which suggests the need for generalized analyses targeting mixed OT environments.

In Korea, following the introduction of Hi-pass as a form of ETC at selected tollgates in 2000, ETC was expanded nationwide by 2007, and a wide range of benefits were reported, including reduced travel time, lower fuel consumption and operating costs, decreased emissions, mitigated congestion, and improved safety. According to the Public Data Portal (data.go.kr) dataset provided by the Korea Expressway Corporation, the ETC share of expressway traffic in 2023 was reported as 89.8%. As ETC has become the primary toll collection method and its effectiveness has been demonstrated, HSETC (high-speed electronic toll collection) upgrade projects have been pursued to increase throughput and reduce toll-plaza delay. Nevertheless, because it is necessary to simultaneously consider continuity of

existing toll-plaza operations and the processing of non-tag vehicles, mixed OT operations in which HSETC, ETC, and TCS coexist are implemented across many segments. Similar mixed OT operations are also found in the United States, where high-speed electronic tolling is being expanded, while on-site payment systems for non-tag users are maintained in parallel. For example, the Ohio Turnpike allows electronic tolling users to pass through via an open-road method, while providing non-users with the option to pay by cash or credit card at exits. This can be characterized as a representative form of mixed OT, in which high-speed electronic passage and on-site payment operations are implemented within the same system. In addition, many countries, including those in Europe, commonly operate plaza-based facilities where electronic tolling lanes and on-site payment lanes coexist. Although operating speeds and lane configurations for electronic tolling vary across countries, many systems already operate under mixed conditions that combine electronic tolling with on-site payment. In such settings, introducing HSETC will likely create a transitional mixed OT regime in which high-speed electronic passage lanes coexist with low-speed or stop-and-pay lanes.

In summary, while HSETC-based OT operations are expanding worldwide, mixed OT operations in which HSETC coexists with ETC or TCS remain prevalent in practice, and their implementation varies across countries and operating agencies. Accordingly, discussions confined to a single country or a specific operational case may be insufficient to fully capture the risk mechanisms of mixed OT. Therefore, a comprehensive safety analysis framework is needed to cover diverse combinations of lane configuration, geometry, demand, and vehicle composition while producing generalizable and transferable insights.

2.2. Applicability of simulation-based SSMs

Traditional roadway safety evaluation has largely evolved around SPF models based on crash data. For example, Lord and Mannering (2010) comprehensively summarized the methodological foundations of crash-frequency analysis, including the Poisson and negative binomial families, and demonstrated that crash-based approaches have served as a core tool for safety evaluation and policy/design decision-making [7]. However, because HSETC involves diverse operational and geometric configurations and remains in an early stage of deployment, it may be difficult to secure sufficient and stable crash observations at the individual toll-plaza level. In such cases, relying solely on crash-based approaches imposes constraints on diagnosing risk mechanisms and comparing/evaluating alternative design and operational strategies.

Under these conditions, an approach that generates vehicle trajectories via microscopic traffic simulation and evaluates risk by computing SSMs has been widely used in safety research. Gettman and Head (2003) presented procedures and applicability for computing SSMs from simulation outputs to compare safety across alternatives [8], and FHWA (2008) summarized that SSAM can extract conflicts using trajectory-based interaction information and support comparative safety evaluation [9]. Accordingly, conventional SSMs such as TTC and PET, as well as SSAM-based conflict frequencies, have been used as effective indicators to quantify vehicle interactions and compare relative risk across design/operational alternatives even when crash observations are limited.

SSM-based evaluation has also been applied to toll-plaza safety assessment, where merging/diverging and lane-changing interactions are concentrated. Xing et al. (2020) analyzed temporal variability in conflict occurrence in toll-plaza approach/influence areas and showed that changes in operational conditions, such as demand levels, can be associated with conflict patterns [13].

Wang and Stamatiadis (2013) established a framework for deriving surrogate safety measures from microsimulation-based conflict analysis [21]. Essa and Sayed (2015) further reported that simulated conflicts were broadly consistent with field-measured conflict patterns, supporting the use of simulation-based SSMs for proactive safety evaluation under varying geometric and operational scenarios [22]. In addition, Fei et al. (2025) evaluated safety performance in a toll-plaza diverging area based on an improved simulation platform that considers weak-constraint lane behaviors and showed that conflict-based outcomes can vary across conditions [10]. Collectively, these studies support that variations in lane configuration, geometry, demand, and toll collection composition can materially affect conflicts in toll-plaza environments, and suggest that SSM/conflict-based analysis can be empirically used to diagnose risk formation mechanisms and compare alternatives in mixed OT environments.

Nevertheless, because simulation-based SSM results can vary depending on scenario combinations, it is important to construct scenario datasets that systematically cover diverse design and operational conditions and to accumulate and compare the resulting outputs in a reproducible manner. Lares et al. (2025) discussed the need to generalize condition-specific risk structures through data-driven learning in complex traffic safety problems and proposed research directions for organizing scenario-based outputs into decision-relevant forms [11]. Therefore, for mixed OT environments, an analysis framework is required that leverages simulation-derived SSM/conflict outcomes while ensuring comparability and reproducibility across a wide range of geometric and operational combinations.

2.3. Computational learning–based crash and conflict prediction studies

Simulation-based SSM/conflict analysis is effective for diagnosing risk mechanisms and comparing alternatives; however, an additional modeling framework is required to cover a broader range of design and operational conditions and to organize the results into prediction and interpretation forms that can be used for decision-making. Recent advances in CI and CL provide a foundation for learning risk outcomes from scenario datasets and developing generalizable prediction functions. In particular, tree-based and boosting-based models such as DT, XGB, and LGBM can effectively capture complex nonlinearities and high-order interactions, and many studies in the traffic safety domain have reported applications of these models for risk prediction and for analyzing contributing factors [13–20]. For example, Ma et al. (2023) reported an approach that uses tree/boosting-based model predictions to quantify the contributions of key risk factors in a freeway safety-related prediction problem [19].

Meanwhile, because safety evaluation results are directly linked to design and operational decision-making, it is important to secure not only predictive performance but also model explainability and trustworthiness. XAI has been widely used as a method to decompose complex predictions into consistent feature-level contributions and to interpret both variable effects and interactions. Lundberg et al. (2020) summarized the SHAP interpretation framework by presenting an approach that extends explanations of tree-based models from local insights to global understanding [16], and Parsa et al. (2020) demonstrated the applicability of combining XGBoost and SHAP to explain key contributing factors in safety-related predictions [17]. In addition, Al-Mahamid et al. (2025) reported an approach that integrates tree-based predictive models with explainability techniques to interpret variable effects and condition-specific changes in risk [18]. Therefore, a framework combining CL and XAI can organize outputs derived from simulation-based SSM/conflict analysis into reproducible

prediction and interpretation forms across diverse design and operational spaces, thereby contributing to diagnosing risk mechanisms and supporting decision-making for mixed OT operations.

2.4. Implications

Based on the literature review, this study investigated diverse implementation configurations of HSETC-based mixed OT and examined (i) the suitability of conflict-based analysis under conditions where crash data are limited, as well as (ii) generalizable predictive modeling and an interpretation framework built upon such conflict-based outcomes. Building on these discussions, we defined the safety risk of mixed OT in terms of conflict frequency rather than crash frequency, systematically designed simulation scenarios, and then compared and validated traditional SPFs (Poisson and negative binomial) against CL models (DT, XGB, and LGBM), while interpreting modeling results using SHAP. Overall, this study implies that, in mixed OT systems where diverse configurations are likely to coexist, shifting conflict-based evaluation from retrospective ex post comparisons to a generalizable prediction-and-interpretation framework enables reproducible risk diagnosis and actionable evidence for design and operational improvements even under data scarcity and complex nonlinear and interaction effects.

3. Materials and methods

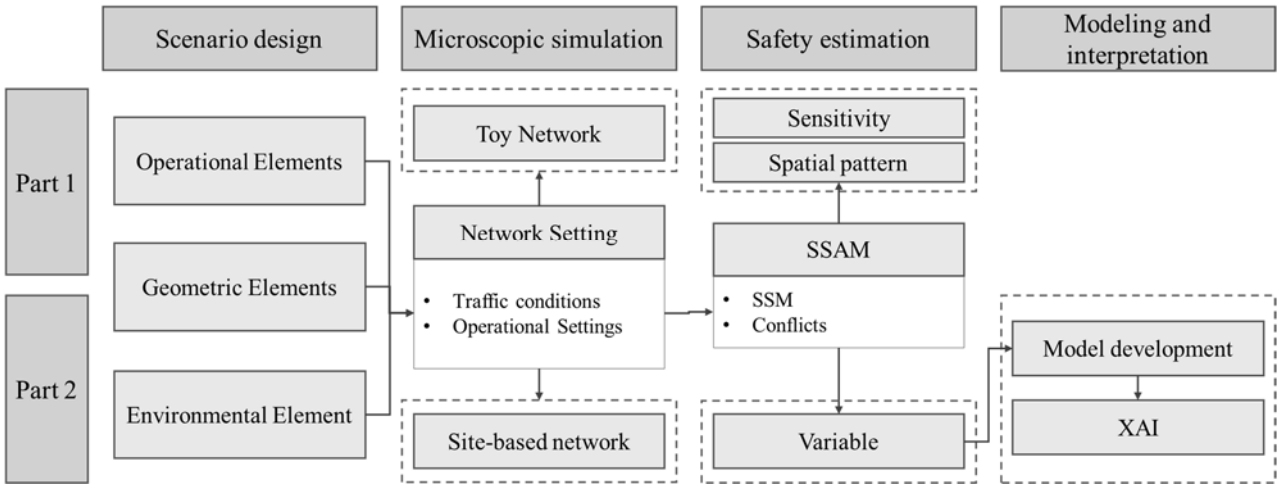

Figure 1. Overall research framework.

This study applies a four-step analytical framework (Figure 1) to comprehensively evaluate traffic safety risk in OT environments. The framework consists of 1) scenario design, 2) microscopic traffic simulation, 3) SSAM-based estimation of SSMs and extraction of traffic conflicts, followed by Part-1 diagnostic analyses focusing on design-factor sensitivity and spatial risk patterns, and 4) Part-2 modeling, which involves CL-based prediction of conflict frequency and XAI-based interpretation. To address the distinct objectives of Parts 1 and 2, two separate scenario sets were constructed. Each scenario set was simulated using a microscopic traffic simulation model to generate vehicle trajectory data. Based on these trajectories, TTC, PET, and total and type-specific conflict frequencies were extracted using SSAM. In Part 1, the derived SSMs and conflict outcomes were directly analyzed to

quantitatively identify the effects of design and operational factors on safety performance and to detect spatial patterns of risk concentration within the merge influence area. Part 2 formulates conflict frequency prediction as a supervised learning problem, using scenario-level geometric and operational variables as inputs and conflict frequency as the target variable. In this stage, traditional SPF baselines based on Poisson and NB models are compared with tree-based and boosting-based CL models under both non-count-aware and count-aware objective functions. Model performance is evaluated through CV, with WMAPE and R^2 adopted as the primary evaluation metrics to select the optimal model. Finally, beyond total conflict frequency, the generalization of prediction performance and explanation structures is examined across rear-end-type and lane-change-type conflicts.

3.1. Scenario design

The scenario design was structured to support both diagnostic inference on risk formation mechanisms in OT environments and predictive modeling of conflict frequencies under diverse OT conditions. Accordingly, two complementary scenario sets were prepared. Part 1 is diagnostic-oriented, aiming to (i) quantify the sensitivity of conflicts to key design and operational factors and (ii) characterize spatial risk patterns within the merge influence area. In contrast, Part 2 is modeling-oriented, aiming to establish predictive relationships between scenario-level design–operation alternatives and conflict frequencies, thereby enabling generalizable risk estimation beyond the simulated cases. Across both parts, OT is consistently conceptualized as a mixed toll environment in which HSETC lanes operate alongside conventional ETC and TCS lanes.

3.1.1. Rationale for scenario setting

The selection of scenario elements was guided by two complementary considerations: relevance to safety risk formation in OT environments and consistency with the distinct analytical objectives of Parts 1 and 2.

First, the scenario elements were chosen to represent operational and geometric factors that can be realistically adjusted in the planning and operation of OT plazas [23–26]. These elements directly influence the core mechanisms through which safety risks emerge in OT environments, including speed differentials between lane groups, the intensity of lane-changing demand, and the feasibility of merging maneuvers within constrained distances. In particular, the coexistence of HSETC lanes with ETC and TCS lanes inherently produces heterogeneous traffic streams, and the interaction of these streams is strongly shaped by lane layout, speed limit, traffic demand, lane-type composition, merging length, number of mainline lanes, and heavy vehicle ratio. These elements jointly determine how vehicles distribute across lane groups upstream, how speed adjustments occur near the tolling area, and where conflicts are likely to concentrate within the merge influence area.

Second, the scenario elements were selected to support two analytically distinct but methodologically connected purposes. Part 1 is diagnostic-oriented and focuses on identifying how and where safety risks form, emphasizing interpretable contrasts and spatial patterns of conflicts. Accordingly, Part 1 employs a subset of scenario elements and analyzes them directly without further variable transformation. In contrast, Part 2 is modeling-oriented and aims to establish predictive relationships between design–operation alternatives and conflict frequencies. To achieve this, Part 2 expands the combinatorial space of scenario elements and subsequently transforms selected elements

into quantitative modeling variables suitable for supervised learning. This dual structure ensures that the scenario design simultaneously supports mechanistic interpretation and predictive generalization, while maintaining a consistent conceptualization of OT environments across both parts.

3.1.2. Definition and interpretation of scenario variables

This study distinguishes scenario elements from modeling variables to separate design levers used to construct OT scenarios from the numerical inputs used for prediction models. The scenario elements are the controlled geometric and operational factors specified during scenario design, whereas the modeling variables are transformed or encoded quantities derived from those elements for statistical analysis and CL-based prediction.

As summarized in Table 1, seven scenario elements were considered: LL, SL, TV, LTR, ML, NL, and HVR. LL represents the left-to-right lane arrangement of toll-lane types within the toll-plaza influence area, reflecting how HSETC, ETC, and TCS lanes are spatially allocated and how mixed streams are forced to sort and merge. In this study, LL has three discrete layouts, HSETC–TCS–ETC–TCS, HSETC–TCS–ETC, and HSETC–TCS, which were selected to reflect realistic mixed toll operations and to induce distinct lane-changing and merging demands. SL denotes the speed-limit setting applied to the HSETC operating environment within the influence area and is set to 60, 80, and 100 km/h to represent alternative speed-management regimes. TV describes demand intensity; in Part 1, it is set as peak vs. non-peak to support interpretable contrasts, while in Part 2, it is configured by LOS A–D at the scenario-setting stage and then represented by observed traffic volume (continuous) in modeling to preserve demand granularity and improve learnability of count outcomes. LTR is a lane-type composition rule indicating how ETC and TCS lane counts are provided relative to HSETC, with two levels: (i) equivalent to the HSETC lane count and (ii) 50% of the HSETC lane count for both ETC and TCS. ML denotes the available merging length downstream of the tolling zone and is set to 280, 380, and 480 m to evaluate the constraint effect of merging distance. NL represents the number of mainline lanes (3 vs. 5) and captures differences in upstream/downstream capacity and lane-sorting space. HVR denotes the heavy-vehicle ratio (10%, 25%, and 40%) to reflect realistic composition shifts that can amplify interaction risks in mixed operations.

The analytical use of these elements differs by part (Tables 2 and 3). Part 1 is diagnostic-oriented; therefore, the elements are used directly without variable transformation. Part 2 is modeling-oriented; therefore, selected elements are transformed into modeling variables to create a compact, learnable, and interpretable feature set for conflict-frequency prediction. In Part 2, LL is encoded as a three-category attribute using a one-hot representation. SL and HVR are retained as numeric inputs (km/h and %, respectively). TV is entered as the observed traffic volume to represent continuous demand effects on conflict frequency. Because LTR and NL describe lane provisioning and the scale relationship between the plaza and the mainline, they are not used as raw categorical labels; instead, they are transformed into ratio-type modeling variables that explicitly quantify lane-type composition and relative lane scale. Specifically, lane-type composition is represented by HSETCR, ETCR, and TCSR, defined as the proportions of HSETC, ETC, and TCS lanes among all toll-plaza lanes, such that HSETCR + ETCR + TCSR = 1. The scale relationship between toll lanes and the mainline is represented by TLR, defined as the ratio of the total number of toll-plaza lanes to the NL. This element-to-variable formulation preserves the design intent of LTR and NL while providing numerically stable predictors aligned with discrete count modeling.

Table 1. Scenario elements.

Element	Description
LL	Lane layout (left-to-right ordering of HSETC/ETC/TCS lanes)
SL	Speed limit of HSETC (km/h)
TV	Demand intensity
LTR	Lane-type provisioning rule (ETC/TCS relative to HSETC)
ML	Merging length (m)
NL	Number of mainline lanes
HVR	Heavy vehicle ratio (%)

3.1.3. Scenario set

Two complementary scenario sets were prepared to align with the distinct objectives of Parts 1 and 2 (Tables 2 and 3).

Part 1 quantifies how conflicts respond to key design and operating factors and pinpoints spatial hotspots within the merge influence area. In mixed OT operations, conflicts tend to cluster non-uniformly, particularly where streams with different speeds converge and where limited merging distance restricts feasible maneuvers. Accordingly, the Part 1 scenario set prioritizes interpretable, section-based diagnostics over exhaustive combinatorial coverage. Accordingly, Part 1 uses five elements directly: LL (3 levels), SL (3 levels), LTR (2 levels), ML (3 levels), and TV (2 levels), as summarized in Table 2. This structure enables controlled contrasts to examine whether conflict concentration shifts toward the tolling zone or downstream segments under changes in demand (TV), speed regime (SL), lane-type provisioning (LTR), lane arrangement (LL), and merging distance (ML), thereby yielding mechanistic insights beyond scenario-level totals.

Table 2. Part 1 diagnostic scenario set.

Element	Standard scenario	Comparison scenario 1	Comparison scenario 2	Total
LL	HSETC-TCS-ETC-TCS	HSETC-TCS-ETC	HSETC-TCS	3
SL	80 km/h	60 km/h	100 km/h	3
LTR	Equivalent to HSETC	50% of HSETC lane count	-	2
ML	380 m	280 m	480 m	3
TV	Peak hour	Non-Peak hour	-	2

Part 2 aims to develop and compare predictive models of conflict frequency across a broad design–operation space and to test whether performance and explanation structures generalize from total conflicts to type-specific conflicts. For this purpose, Part 2 expands the scenario space using LL (3 levels), SL (3 levels), LTR (2 levels), NL (2 levels), HVR (3 levels), and TV configured by LOS A–D (4 levels), yielding 432 scenarios in a full-factorial structure ($3 \times 3 \times 2 \times 2 \times 3 \times 4$). The modeling inputs are defined as the transformed variables summarized in Table 3 (LL one-hot; SL, HVR numeric; TV as observed traffic volume; LTR and NL transformed into HSETCR/ETCR/TCSR and TLR). The prediction targets are scenario-level conflict frequencies extracted from SSAM, including total conflicts and type-specific conflicts, which supports a direct comparison between non-count-aware

objectives and count-aware objectives aligned with discrete frequency outcomes.

By design, the two scenario sets are complementary: Part 1 prioritizes interpretability and spatial diagnosis (why and where conflicts form), while Part 2 prioritizes predictive generalization across unseen combinations (how conflicts change under new design–operation alternatives). This dual structure enables an integrated OT safety assessment pipeline that links diagnostic evidence to model-based decision support.

Table 3. Part 2 modeling variables derived from scenario elements.

Elements	Variable	Description
LL	LL_0, LL_1, LL_2	One-hot encoding
SL	SL	Numeric
TV	TV_vol	Numeric
HVR	HVR	Numeric
LTR	HSETCR, ETCR, TCSR	Ratios (HSETCR + ETCR + TCSR = 1)
NL	TLR	(# toll plaza lanes)/NL

3.2. Microscopic simulation environment

This study conducted a microscopic traffic simulation using PTV VISSIM, with car-following behavior specified based on the built-in Wiedemann 99 model [27]. To realistically reproduce lane-choice and merging/diverging interactions that emerge when HSETC, ETC, and TCS lanes coexist in mixed OT environments, we considered it essential to reflect plaza usage shares and traffic-flow patterns. Accordingly, VISSIM's vehicle route and desired speed decision functions were used to implement lane group–specific routes and speed-control schemes in accordance with scenario conditions. The SeoulTG influence area was selected as the base network, and validation was performed by comparing the simulation outputs with the as-built traffic configuration of the SeoulTG influence area, confirming that key traffic-flow patterns under mixed operations were reproduced consistently.

Subsequently, while maintaining the SeoulTG-based base network as a common foundation, the simulation environment was adjusted to match the distinct scenario settings of Parts 1 and 2. Part 1 compares the sensitivity of conflicts to changes in design factors and diagnoses spatial risk patterns. Part 2 applies CI/CL models to functionalize conflict frequencies based on an expanded set of design–operation combinations. In both parts, the simulation environment was constructed on the base network so that the essential OT mechanisms (i.e., speed heterogeneity, lane changing, and constrained merging interactions arising from the coexistence of HSETC, ETC, and TCS lanes) were reproduced consistently, and high-resolution vehicle trajectory data were generated to enable consistent SSAM-based conflict extraction.

3.2.1. Simulation environment

The Part 1 simulation environment is a simplified, generic OT toy network constructed to enable interpretable comparisons across scenario factors (Section 3.1). The network includes (i) an upstream approach segment where lane-choice and speed adaptation begin, (ii) a toll-lane operation zone representing mixed HSETC–ETC–TCS streams, and (iii) a downstream merge influence area where streams converge into the mainline. The downstream extent was selected to capture post-toll turbulence,

including forced merges, residual lane changes, and speed recovery. The toy network intentionally minimizes site-specific geometry and external disturbances so that changes in conflict outcomes can be attributed primarily to scenario-level factors in Part 1 (i.e., LL, SL, LTR, ML, and TV). This environment is particularly suitable for Part 1 because it supports section-based spatial aggregation: conflicts can be mapped and compared across predefined longitudinal sections to identify risk concentration zones and systematic shifts in conflict locations driven by alternative OT layouts and operating regimes.

The Part 2 simulation environment is based on the SeoulTG corridor and is intended to reflect realistic geometric and operational complexity around a representative OT facility. The network covers the upstream–downstream influence area required to capture OT-related interactions, including approach dynamics, lane allocation and sorting, toll-lane operations, and downstream merging and recovery. Unlike the toy network, the SeoulTG-based environment preserves key site-anchored characteristics (e.g., merging topology and operational constraints) that affect conflict formation under realistic traffic compositions and demand levels. This environment supports the construction of a comprehensive modeling dataset (432 scenarios) by systematically varying the Part 2 scenario elements (LL, SL, demand condition, LTR, NL, and HVR) and by operationalizing modeling variables derived from these elements (e.g., HSETCR, ETCR, TCSR, and TLR), with TV represented by observed traffic volume to preserve demand granularity for learning. Because the objective of Part 2 is to learn predictive mappings from scenario-level descriptors to discrete conflict frequencies, the simulation outputs are aggregated at the scenario level after SSAM processing, producing the target variables used in the CI/CL modeling pipeline [28,29,31].

3.2.2. Traffic demand, operations, and behavioral settings

In both environments, demand is controlled using scenario-specific inputs. In Part 2, demand scenarios are configured by LOS categories (A–D), while TV is operationalized as observed traffic volume for modeling to retain continuous demand information across scenarios. Speed management is implemented through SL settings, which interact with toll-lane composition to generate heterogeneous speed profiles across lane groups. Vehicle composition and baseline car-following and lane-changing behavior are held consistent within each part to ensure that differences in conflicts reflect scenario factor changes rather than ad hoc behavioral recalibration.

To reduce stochastic noise intrinsic to microscopic simulation, scenarios are executed with multiple random seeds and results are aggregated before modeling (Part 2) or before spatial diagnostics (Part 1), depending on the analytical objective. This procedure improves the stability of estimated conflict frequencies and the robustness of spatial pattern comparisons.

3.2.3. Trajectory output preparation for SSAM

For both Parts 1 and 2, high-resolution vehicle trajectory data are exported as the direct input to SSAM. Trajectory sampling resolution is set to capture rapid longitudinal and lateral maneuvers that are typical in OT merging environments. By ensuring consistent trajectory quality, the subsequent SSAM-based conflict extraction is not confounded by output resolution differences, enabling reliable comparison across scenarios and between the two analytical parts [27,33].

3.3. Surrogate safety estimation and conflict extraction

Surrogate safety measures were estimated using SSAM, which identifies potential crash events from high-resolution vehicle trajectory data by evaluating spatiotemporal proximity and interaction severity. SSAM is particularly suitable for OT environments because it enables safety assessment under conditions where crash data are sparse, heterogeneous, or unavailable, while still capturing vehicle-level interaction dynamics such as abrupt deceleration, forced merging, and lateral intrusion that are characteristic of mixed HSETC–ETC–TCS operations. In this study, vehicle trajectories generated from microscopic simulations were processed using SSAM to compute TTC- and PET-based interaction metrics and to extract discrete conflict events. To ensure consistency with the established surrogate safety literature and comparability across scenarios, default SSAM parameter settings recommended in prior validation studies were adopted [8,9,21,22,33]. In this study, the TTC threshold was set to 1.5 s, and the PET threshold was set to 5.0 s.

Each detected interaction was classified into one of three conflict types based on relative vehicle trajectories and motion vectors. Rear-end conflicts represent longitudinal following interactions associated with critical TTC values. Lane-change conflicts represent lateral intrusion events occurring during active lane-changing maneuvers. Crossing conflicts represent interactions with intersecting trajectories that involve both longitudinal and lateral components. Events that could not be consistently assigned to one of these categories were labeled as unclassified; however, such cases were negligible and therefore excluded from subsequent analysis. For each scenario, conflict frequencies were aggregated to produce four target measures: total conflicts, rear-end conflicts, lane-change conflicts, and crossing conflicts. However, within freeway toll-plaza segments, most vehicles travel in the same direction, and interactions primarily arise during diverging, lane-changing, and merging processes. Under this operational mechanism, the conditions required for crossing conflicts defined by interactions with intersecting trajectories are inherently limited, and the detected frequency of such events was extremely low in our simulations. Accordingly, we regarded crossing conflicts as rare events that are unlikely to represent the dominant risk mechanisms in the study setting and focused the subsequent analyses on rear-end and lane-change conflicts. These aggregated measures play distinct analytical roles in Parts 1 and 2, consistent with the dual diagnostic–predictive structure of the study.

In Part 1, SSAM outputs were used primarily for diagnostic analysis. Conflict locations were mapped onto predefined longitudinal sections of the OT toy network described in Section 3.2, enabling spatial examination of conflict concentration and migration within the merge influence area. This section-based aggregation facilitates the identification of high-risk zones and enables systematic comparisons of how changes in LL, SL, TV, LTR, and ML alter both the location and the mechanism of conflict emergence. In addition to spatial diagnostics, aggregated conflict counts and summary SSM statistics were used to assess the sensitivity of safety outcomes to individual scenario elements. Accordingly, the emphasis of Part 1 is on understanding why and where conflicts occur under alternative OT design and operating conditions, rather than on predictive generalization.

In Part 2, SSAM-derived conflict frequencies were treated as response variables in a supervised modeling framework. Across the 432 scenarios in the SeoulTG-based simulation environment (Section 3.2.1), scenario-level variables are constructed from LL, SL, TV, LTR, NL, and HVR and transformed as described in Section 3.1. These variables are used as model inputs, and section-aggregated conflict counts are used as the outputs. This transformation converts surrogate safety assessment into a conflict-frequency modeling problem, enabling the construction of predictive models

that estimate expected conflict frequencies under previously unobserved OT configurations. Unlike Part 1, spatial information is not explicitly retained in Part 2; instead, conflicts are aggregated over the full OT influence area, consistent with the objective of deriving compact, scenario-level risk functions suitable for planning- and design-oriented evaluation. Separate response variables are defined for total, rear-end, lane-change, and crossing conflicts to support conflict type–specific modeling and generalization analysis in the subsequent CI/CL framework.

3.4. Modeling approach for conflict frequency functionalization

3.4.1. Problem definition and learning targets

Part 2 develops predictive models that map scenario-level design and operational descriptors to SSAM-derived conflict frequencies aggregated over the OT influence area [28–32]. Let x_i denote the descriptor vector for scenario i and y_i denote the corresponding conflict count. The input vector is defined as

$$\mathbf{x}_i = [LL_i, SL_i, TV_i, TLR_i, HSETCR_i, ETCR_i, TCSR_i, HVR_i] \tag{1}$$

and the prediction target is

$$\hat{y}_i = f(\mathbf{x}_i) \tag{2}$$

Here, y_i is defined as the scenario-level aggregated conflict frequency obtained from SSAM outputs, and the primary target is total conflicts, with additional targets for rear-end, lane-change, and crossing conflicts [28–33]. Because SSAM conflicts are nonnegative, discrete, and typically right-skewed, we evaluate both conventional parametric count models and computational learning models under objectives that are either generic (squared error) or count-consistent (Poisson) [31,34].

3.4.2. Candidate models and rationale

To functionalize conflict frequencies under heterogeneous OT scenarios, this study considers a set of candidate models that progressively relax the assumptions of traditional safety performance functions while maintaining interpretability and predictive stability.

As a benchmark, traditional parametric SPFs based on Poisson regression and negative binomial (NB) regression are included. These models have been widely used in traffic safety analysis to model discrete crash or conflict counts and provide a statistically well-established reference framework. The Poisson model assumes equality of mean and variance, whereas the NB model introduces an overdispersion parameter to accommodate extra-Poisson variability. Although these models offer clear parameter interpretation, their restrictive functional forms limit their ability to capture nonlinear effects and higher-order interactions among OT design and operational variables [7].

To relax linearity assumptions while preserving a transparent structure, a DT model is included as a nonlinear but interpretable baseline. DTs recursively partition the feature space and can naturally represent threshold effects in variables such as lane layout, demand level, and electronic-lane composition. However, single-tree models are known to be unstable and may suffer from limited predictive accuracy in complex, high-dimensional settings. In this study, the DT serves primarily as a reference point between parametric SPFs and more flexible ensemble models.

Gradient boosting models, including XGBoost and LightGBM, are adopted to capture nonlinear relationships and complex interactions across OT scenarios. In their standard formulations, these models minimize squared-error-based loss functions, treating conflict frequencies as continuous outcomes. Such non-count-aware objectives are commonly used in traffic-related machine learning studies due to their computational efficiency and strong predictive performance. Including these formulations allows direct comparison with much of the existing CI/CL literature and provides a baseline for assessing the added value of distribution-aware learning [35–37].

In addition to standard boosting formulations, this study explicitly considers count-aware boosting models using Poisson-based objectives. Conflict frequencies are nonnegative integer outcomes whose variability tends to increase with their mean, a property that is not explicitly reflected in squared-error loss functions. Poisson objectives directly align the learning process with the assumed data-generating mechanism of count variables, potentially improving calibration and stability across low- and high-conflict regimes. By evaluating both non-count-aware and count-aware objectives within the same boosting frameworks, this study systematically examines whether distribution-consistent learning offers advantages for functionalizing OT conflict frequencies.

Overall, the candidate model set is designed to span a spectrum from statistically established SPFs to flexible CI/CL models. This structure enables (i) benchmarking against conventional safety analysis approaches, (ii) assessing the benefits of nonlinear and interaction-aware learning, and (iii) examining the role of count-consistent objective functions in improving the robustness of conflict-frequency prediction under diverse OT operating conditions.

3.4.3. Cross-validation and evaluation metrics

Because the response variables are nonnegative conflict counts aggregated by scenario, data exhibit typical count characteristics (discreteness, right-skewness, and mean-variance dependence). To evaluate candidate models fairly under these characteristics, we adopted a unified validation protocol and multiple complementary performance measures.

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote the dataset with $N = 432$ scenarios. We performed K-fold cross-validation with $K = 5$. The index set $\{1, \ldots, N\}$ is partitioned into K disjoint folds $\mathcal{I}_1, \ldots, \mathcal{I}_K$. For fold k, the training and test sets are

$$\mathcal{D}_{\text{train}}^{(k)} = \{(\mathbf{x}_i, y_i): i \notin \mathcal{I}_k\}, \mathcal{D}_{\text{test}}^{(k)} = \{(\mathbf{x}_i, y_i): i \in \mathcal{I}_k\} \tag{3}$$

A model $f^{(k)}(\cdot)$ is fit on $\mathcal{D}_{\text{train}}^{(k)}$ and evaluated on $\mathcal{D}_{\text{test}}^{(k)}$. Metrics are computed for each fold and summarized as mean ± standard deviation across folds:

$$\bar{M} = \frac{1}{K}\sum_{k=1}^{K} M^{(k)}, s_M = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\left(M^{(k)} - \bar{M}\right)^2} \tag{4}$$

To quantify predictive error in a scale-consistent manner, we used WMAPE as the primary metric. For fold k, with test set $\mathcal{I}_k$, WMAPE is defined as

$$\text{WMAPE}^{(k)}(\%) = 100 \times \frac{\sum_{i\in\mathcal{I}_k}|y_i - \hat{y}_i|}{\sum_{i\in\mathcal{I}_k} y_i} \tag{5}$$

where $\hat{y}_i = f^{(k)}(\mathbf{x}_i)$. Compared with MAPE, WMAPE avoids instability caused by small denominators because errors are normalized by the total observed count in the test fold rather than by each individual y_i. This property is particularly important in OT scenarios where conflict magnitudes can vary substantially across demand and lane-composition regimes.

To assess how well each model explains scenario-level variance, we used R^2 as a complementary metric. For fold k,

$$R^{2\,(k)} = 1 - \frac{\sum_{i\in\mathcal{I}_k}(y_i-\hat{y}_i)^2}{\sum_{i\in\mathcal{I}_k}(y_i-\bar{y}_k)^2}, \bar{y}_k = \frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k} y_i \tag{6}$$

While R^2 is widely interpretable, it is sensitive to outliers and the scale of counts; therefore, it is used alongside WMAPE to balance proportional accuracy and variance explanation.

The optimal CL model for total conflicts was selected based on the minimum mean WMAPE across folds. R^2 was used as a secondary check to ensure that the chosen model also maintains high explanatory power and does not achieve lower WMAPE by over-smoothing large-count regimes. This selection strategy is consistent with the intended application of the model as a scenario-level risk functionalization tool for OT planning and design evaluation.

3.4.4. Conflict type–specific modeling for transferability and mechanism insight

After identifying the best-performing model for total conflicts, the same candidate modeling logic was applied to conflict subtypes (rear-end, lane-change, and crossing) using the identical feature set. Each subtype was modeled independently to evaluate (i) predictive transferability across conflict mechanisms and (ii) whether the same OT scenario descriptors yield consistent explanatory power across interaction types.

This design directly supports mechanism-level interpretation. For example, strong performance for rear-end and lane-change conflicts suggests that demand and lane-composition descriptors adequately capture longitudinal and lateral interaction pressures, whereas degraded performance for crossing conflicts may indicate that additional geometric constraints or maneuver-specific cues are needed beyond scenario-level descriptors.

3.4.5. XAI interpretation using SHAP

To provide a transparent, SPF-like interpretation of CL predictions, SHAP was applied to the selected boosting model(s). For each scenario, SHAP decomposes the predicted conflict frequency into additive feature contributions, enabling both global and local interpretation. Global importance was summarized by mean absolute SHAP values, while feature-effect patterns were examined across the range of TV and lane-composition descriptors to identify nonlinear responses and interaction regimes relevant to OT operations. This XAI framework ensures that high-capacity CI models remain auditable and policy-relevant, supporting evidence-based decisions on OT lane composition and operational strategies.

4. Results

4.1. Part 1: Design-factor sensitivity and spatial pattern diagnosis

4.1.1. Aggregated sensitivity results across design and operating factors

Table 4 summarizes changes in average TTC and conflict frequencies across scenarios. Safety improvement is identified when TTC increases or conflicts decrease relative to the standard scenario.

For LL, under non-peak demand, the standard case reported an average TTC of 0.314 s. The HSETC–TCS–ETC layout showed a lower TTC of 0.254 s, whereas the HSETC–TCS layout showed a higher TTC of 0.341 s. Despite these differences in TTC, both alternative layouts reduced total conflict frequency. Under peak demand, the standard TTC was 0.407 s, and TTC values under the two alternative layouts were 0.389 s and 0.396 s, respectively, while total conflicts decreased from 50.567 in the standard case to 42.633 and 48.967 in the alternative layouts. Rear-end and lane-change conflicts also decreased in most scenarios, but under peak demand lane change, conflicts increased from 17.022 to 17.978 for one alternative layout. These findings suggest that LL can redistribute speed profiles and lane-choice patterns in a way that affects TTC and conflicts differently under OT operations.

For SL, non-peak TTC in the standard case was 0.358 s. When SL was reduced to 60 km/h, TTC decreased to 0.245 s, and total conflicts increased to 5.333. When SL was increased to 100 km/h, TTC was 0.305 s, and total conflicts decreased to 3.917. Under peak demand, the standard TTC was 0.401 s, becoming 0.370 s at 60 km/h and 0.421 s at 100 km/h. Total conflicts under peak demand increased to 53.300 at 60 km/h and decreased to 43.244 at 100 km/h, compared to 45.622 in the standard scenario.

Rear-end and lane-change conflicts followed a similar pattern, increasing at 60 km/h and decreasing at 100 km/h. Overall, SL changes influence safety by modifying speed differentials and interaction pressure between heterogeneous tolling streams. A notable implication is that safety variation is driven more strongly by speed heterogeneity between the mainline and the toll-plaza influence area than by speed heterogeneity within the toll-lane groups alone. This is because the speed-limit setting reshapes the deceleration and re-acceleration transition as vehicles enter and exit the influence area, which in turn changes longitudinal following pressure and the forcing of lane-changing maneuvers. As a result, SL acts as an operational lever that governs conflict generation primarily through mainline-to-influence-area speed discontinuities and transition-induced interaction intensity.

For LTR, non-peak TTC increased from 0.267 s in the standard case to 0.339 s when the number of TCS and ETC lanes was reduced, while total conflicts decreased from 4.810 to 4.233. Under peak demand, TTC changed marginally from 0.396 to 0.398 s, and total conflicts decreased from 49.267 to 45.511. Rear-end and lane-change conflicts also decreased consistently under both non-peak and peak demand, indicating that toll-lane composition adjustments broadly alleviate interaction pressure. This pattern indicates that adjusting toll-lane composition can stabilize interactions and reduce conflicts, with TTC improvements being more pronounced under non-peak conditions.

For ML, under non-peak demand, the standard TTC was 0.296 s, and total conflicts were 3.578. Shortening ML to 280 m reduced TTC to 0.240 s and increased total conflicts to 6.089, whereas extending ML to 480 m increased TTC to 0.372 s. However, under non-peak demand, total conflicts increased slightly from 3.578 to 3.898, and rear-end and lane-change conflicts also increased, suggesting that the safety benefit of extending merging length may not manifest uniformly across demand regimes. Under peak demand, the standard TTC was 0.392 s; shortening ML to 280 m yielded

a TTC of 0.386 s with total conflicts of 56.578, while extending ML to 480 m yielded a TTC of 0.415 s with total conflicts of 37.411. Under peak demand, rear-end and lane-change conflicts also decreased, indicating that providing additional merging space contributes more directly to suppressing conflicts when demand is high. This result suggests that the effect of merging length extension can be nonlinear because maneuver feasibility and speed adaptation change with demand level.

Table 4. Summary of SSM changes compared to the standard scenario.

Elements	TV	SSM	Standard scenario	Comparison scenario 1	Comparison scenario 2
LL	Non-peak hour	TTC	0.314	0.254	0.341*
		Total conflict	5.556	4.032*	3.978*
		Rear-end conflict	3.069	1.877*	1.667*
		Lane-change conflict	2.486	2.155*	2.311*
	Peak hour	TTC	0.407	0.389	0.396
		Total conflict	50.567	42.633*	48.967*
		Rear-end conflict	33.544	27.389*	30.967*
		Lane-change conflict	17.022	15.233*	17.978
SL	Non-peak hour	TTC	0.358	0.245	0.305
		Total conflict	4.315	5.333	3.917*
		Rear-end conflict	2.221	2.433	1.958*
		Lane-change conflict	2.094	2.900	1.958*
	Peak hour	TTC	0.401	0.370	0.421*
		Total conflict	45.622	53.30	43.244*
		Rear-end conflict	29.878	32.944	29.078*
		Lane-change conflict	15.733	20.344	14.156*
LTR	Non-peak hour	TTC	0.267	0.339*	-
		Total conflict	4.81	4.233*	-
		Rear-end conflict	2.409	2.000*	-
		Lane-change conflict	2.401	2.233*	-
	Peak hour	TTC	0.396	0.398*	-
		Total conflict	49.267	45.511*	-
		Rear-end conflict	32.081	29.185*	-
		Lane-change conflict	17.178	16.311*	-
ML	Non-peak hour	TTC	0.296	0.240	0.372*
		Total conflict	3.578	6.089	3.898
		Rear-end conflict	1.422	3.678	1.513
		Lane-change conflict	2.156	2.411	2.385
	Peak hour	TTC	0.392	0.386	0.415*
		Total conflict	48.178	56.578	37.411*
		Rear-end conflict	31.1	38.122	22.678*
		Lane-change conflict	17.067	18.433	14.733*

Note: *Scenario associated with safety deterioration.

Overall, LL and ML show demand-dependent changes in TTC and conflicts, while SL at 100 km/h and LTR under reduced TCS and ETC lane provision exhibit relatively consistent safety-improving tendencies. Because the same design or operational change can yield different responses across demand regimes and conflict types, interpretation should jointly consider rear-end and lane-change conflicts in addition to average TTC and total conflicts.

4.1.2. Spatial risk patterns and hotspot shifts

The spatial diagnostics in Figure 2 complement the aggregated results by illustrating where conflicts concentrate and how hotspots shift across factor groups. In the baseline OT configuration, conflicts are primarily concentrated around the upstream route-selection region and the downstream acceleration and merge influence area. Rear-end conflicts show stronger clustering, while lane-change conflicts are more spatially dispersed, reflecting their different interaction mechanisms under OT operations.

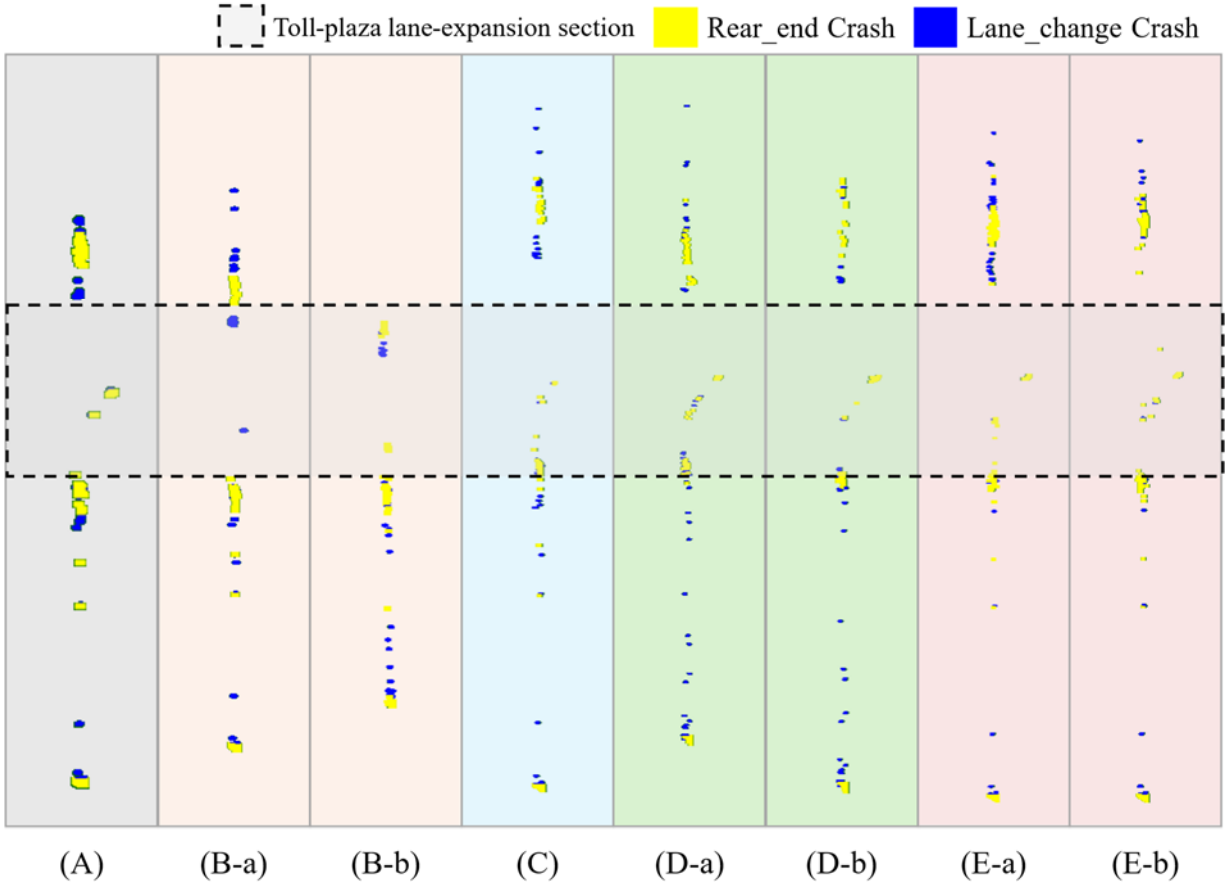

Figure 2. Conflict points distribution maps for different scenarios. A: Standard; B: LL scenario; C: LTR scenario; D: ML scenario; E: SL scenario; a: comparison scenario 1; b: comparison scenario 2.

When LL is modified, hotspots tend to shift closer to the toll plaza and the merging area, indicating that the arrangement of HSETC, ETC, and TCS lanes changes where intensive interactions

occur. For LTR changes, the overall hotspot regions remain similar, but conflict points spread more widely within the merging area, suggesting broader lane-changing pressure distribution when toll-lane composition is altered.

The most pronounced spatial deterioration is observed when ML is shortened. In this case, conflict points become denser inside and immediately around the toll plaza, and the conflict footprint expands both upstream and downstream, implying that constrained merging space amplifies and spatially spreads interaction risk. When ML is extended, hotspot intensity becomes lower, and the pattern appears closer to the baseline, consistent with reduced conflict frequencies in Table 4. Speed limit changes on HSETC lanes also broaden the conflict footprint around critical merging points, suggesting that operational policies can reshape not only the magnitude but also the spatial distribution of risk.

4.2. Part 2 results: Model performance and interpretation

4.2.1. Model development results

This subsection reports the predictive performance of competing CI and CL models for modeling conflict frequencies under OT operating conditions. Model comparison focused on predicting total conflicts using 5-fold CV and two evaluation metrics, WMAPE and R^2, to jointly capture scale-normalized error and goodness-of-fit. Specifically, the full dataset was partitioned into five equally sized folds; in each iteration, four folds (80%) were used for training, and the remaining one fold (20%) was used for testing/validation. This procedure was repeated five times so that each scenario appeared exactly once in the test fold, and performance was compared based on the mean of fold-wise metrics. In addition to WMAPE and R^2, we also report MAE as a supplementary error measure. However, model selection and primary comparison are based on WMAPE and R^2 because WMAPE provides scale-normalized error suitable for cross-scenario count magnitudes, while R^2 captures explanatory power. The candidate set covered two conventional SPF baselines (Poisson GLM and NB), a tree-based benchmark (DT), and gradient-boosting models with two learning objectives, namely a non-count-aware objective RMSE and a count-aware objective Poisson loss.

Table 5 summarizes the CV results for total conflicts. Overall, the data-driven tree/boosting models substantially outperformed the conventional SPF baselines. Both Poisson GLM and NB2 showed markedly lower explanatory power and larger relative error, indicating a limited suitability of purely parametric SPF forms for capturing the nonlinear and interactive mechanisms embedded in the OT scenario space. In contrast, DT and boosting models achieved consistently higher R^2 and lower WMAPE, demonstrating that nonlinear learners better align with the conflict-generation structure induced by heterogeneous operational mixes and merging patterns.

Within boosting family, the relative benefit of adopting a count-aware objective depended on the algorithm. For XGB, the Poisson objective produced lower WMAPE than the RMSE objective while maintaining comparable R^2, supporting the rationale that explicitly reflecting count data characteristics can improve predictive stability. For LGBM, the Poisson objective did not outperform the RMSE objective in this dataset, suggesting that the gain from count-aware learning is not guaranteed and should be empirically validated rather than assumed. Based on the total-conflict task, XGB with the Poisson objective was selected as the primary CL model for subsequent type-specific evaluation and XAI analyses.

To examine whether the learned modeling structure transfers across conflict types, we fixed the

selected model family (XGB Poisson) and evaluated performance for rear-end and lane-change conflicts using the same CV protocol (Table 6). The results show that rear-end and lane-change conflicts were predicted with accuracy comparable to total, indicating that the current scenario-level feature set supports stable modeling for both longitudinal and lane-change mechanisms under OT operations.

Table 5. CV performance for total conflict functionalization.

Model	WMAPE (%)	MAE	R^2
NB	52.62	1013.38	0.456
Poisson GLM	52.62	1013.38	0.456
DT	12.08	230.82	0.955
XGB	13.14	250.19	0.953
LGBM	13.60	259.31	0.954
XGB (Poisson)	11.96	229.29	0.954
LGBM (Poisson)	15.74	302.84	0.940

Table 6. Type-specific generalization using XGB (Poisson).

Target	WMAPE	R^2
Total	13.90	0.937
Rear-end	14.38	0.934
Lane-change	12.49	0.949

Although mean WMAPE values remain in the low-teens, this level of error is considered acceptable for scenario-level conflict frequency prediction, given the strong right-skewness and high dispersion of SSAM-derived counts across heterogeneous OT operating regimes. In particular, the consistently high R^2 indicates that the selected nonlinear learners explain most of the across-scenario variability, while WMAPE remains stable across folds.

4.2.2. SHAP interpretation results

This section explains why the selected count-aware boosting model produces the predicted conflict frequencies and whether the learned explanatory structure is transferable across conflict types. We relied on SHAP, which decomposes each prediction into an additive sum of feature contributions relative to a baseline (expected) prediction. For each target (total, rear-end, and lane-change), we reported four complementary indices (Tables 7–9): (i) mean absolute SHAP value, mean(|SHAP|), (ii) mean signed SHAP value, mean (SHAP), (iii) the proportion of samples with positive SHAP values, pos_ratio, and (iv) the correlation between feature values and SHAP values, corr(x, SHAP). These indices jointly characterize global importance, directionality, sign heterogeneity, and monotonic tendency under nonlinear interactions. Figure 3 visualizes the SHAP results using global importance (bar) and beeswarm summaries for each target.

Mean(|SHAP|) quantifies global importance by averaging the magnitude of SHAP contributions across samples. A larger value indicates that the variable frequently shifts predictions away from the baseline by a larger amount. This metric is comparable across variables within the same target, and it

provides a stable ranking even when the direction of effects varies by operating regime. Mean(SHAP) captures the average direction of contribution. A positive mean(SHAP) suggests that, on average, the variable increases predicted conflict frequency relative to the baseline, whereas a negative mean(SHAP) indicates an average decrease. However, because tree ensembles learn nonlinear and interaction-driven relationships, mean(SHAP) should be interpreted together with pos_ratio and corr(x, SHAP). A negative mean(SHAP) does not necessarily mean the variable reduces risk in general; it may reflect that negative contributions are large in magnitude for some scenarios. pos_ratio measures sign heterogeneity by indicating how often the variable contributes positively across samples. Values near 0.5 imply frequent sign switching, whereas values closer to 1 (or 0) indicate that contributions are predominantly positive (or negative). In highly nonlinear settings, a variable can show high importance while also having pos_ratio near 0.5, signaling strong regime dependence. corr(x, SHAP) provides a diagnostic of monotonic tendency between a variable and its contribution. A strong positive correlation implies that larger feature values tend to be associated with larger positive contributions (or less negative contributions) on average, whereas a negative correlation implies the opposite. Importantly, corr(x, SHAP) is not a proof of monotonicity; it is a summary diagnostic that can still be consistent with conditional reversals due to interactions.

Table 7. SHAP results: total conflict.

Variable	Mean(|SHAP|)	Mean(SHAP)	Pos_ratio	Corr(x, SHAP)
TV	1.412	-0.178	0.690	0.832
TLR	0.330	-0.068	0.670	0.282
LL_2	0.299	0.082	0.375	
HVR	0.243	-0.036	0.365	0.754
ETCR	0.223	0.043	0.550	-0.650
SL	0.087	0.003	0.470	-0.732
TCSR	0.046	-0.008	0.390	-0.686
HSETCR	0.035	0.006	0.530	0.901
LL_0	0.033	-0.006	0.435	
LL_1	0.027	0.002	0.370	

Table 8. SHAP results: rear-end conflict.

Variable	Mean(|SHAP|)	Mean(SHAP)	Pos_ratio	Corr(x, SHAP)
TV	1.420	-0.177	0.690	0.834
TLR	0.340	-0.068	0.670	0.286
LL_2	0.302	0.083	0.375	
HVR	0.251	-0.038	0.365	0.766
ETCR	0.232	0.045	0.555	-0.659
SL	0.085	0.004	0.465	-0.740
TCSR	0.047	-0.007	0.405	-0.690
HSETCR	0.037	0.007	0.535	0.910
LL_0	0.032	-0.006	0.445	
LL_1	0.028	0.002	0.345	

Table 9. SHAP results: lane-change conflict.

Variable	Mean(|SHAP|)	Mean(SHAP)	Pos_ratio	Corr(x, SHAP)
TV	1.320	-0.184	0.715	0.797
TLR	0.256	-0.075	0.665	0.303
ETCR	0.223	0.040	0.550	-0.640
LL_2	0.217	0.057	0.375	
HVR	0.191	-0.023	0.415	0.587
SL	0.078	-0.001	0.630	-0.704
TCSR	0.046	-0.011	0.435	-0.651
LL_0	0.032	-0.005	0.405	
HSETCR	0.022	0.003	0.560	0.824
LL_1	0.018	0.001	0.405	

LL was treated as a nominal scenario descriptor and encoded using one-hot indicators (LL_0, LL_1, LL_2). For a one-hot encoded category, the SHAP value of LL_k should be interpreted as the marginal contribution of being in category k relative to the model baseline, conditional on other variables for that scenario. Because only one category is active per observation, comparing the contributions across LL_0, LL_1, and LL_2 provides insight into how different LL regimes shift predicted conflict frequencies. Two practical notes are critical. First, one-hot SHAP values do not represent a continuous slope, but rather a category-specific shift. Second, because SHAP is computed relative to the model's global baseline (expected value), a positive SHAP for LL_2 does not automatically imply LL_2 is riskier than LL_0 unless the comparison is made explicitly (e.g., by comparing the distribution of LL_k SHAP values across categories or by using category-wise average predictions).

Across total, rear-end, and lane-change, the SHAP rankings and diagnostics were highly consistent (Tables 7–9; Figure 3). TV was the dominant driver under all three targets, indicating that TV is the most influential scenario-level exposure factor in the OT scenario space. While mean(SHAP) for TV was negative, pos_ratio remained high, and corr(x, SHAP) was strongly positive, implying a regime-dependent effect: TV frequently increases predicted conflicts, and larger TV tends to be associated with larger contributions, yet a subset of scenarios receives large negative contributions that dominate mean(SHAP). Such behavior is consistent with tree ensembles capturing interactions and saturation effects, where the marginal impact of TV depends on the joint configuration of LL and lane-composition variables.

TLR also appeared consistently as a major predictor across the three targets. Similar to TV, TLR exhibited a negative mean(SHAP) but a positive corr(x, SHAP), suggesting that increases in TLR can elevate conflict predictions in many regimes while producing negative contributions in others. HVR showed a comparable pattern with a positive corr(x, SHAP), indicating that higher HVR tends to be associated with larger contributions, while the sign of its contribution remains regime-dependent. Among the composition variables, ETCR was consistently influential across targets and exhibited a mixed directional signature, with a positive mean(SHAP) but a negative corr(x, SHAP), indicating strong interaction dependence with the rest of the feature set. SL and TCSR exhibited smaller global importance but contributed as secondary modulators whose contributions vary across operating regimes. Overall, the similarity of SHAP rankings and diagnostics across total, rear-end, and lane-change indicates that these outcomes share a transferable, scenario-level explanatory structure under OT operations.

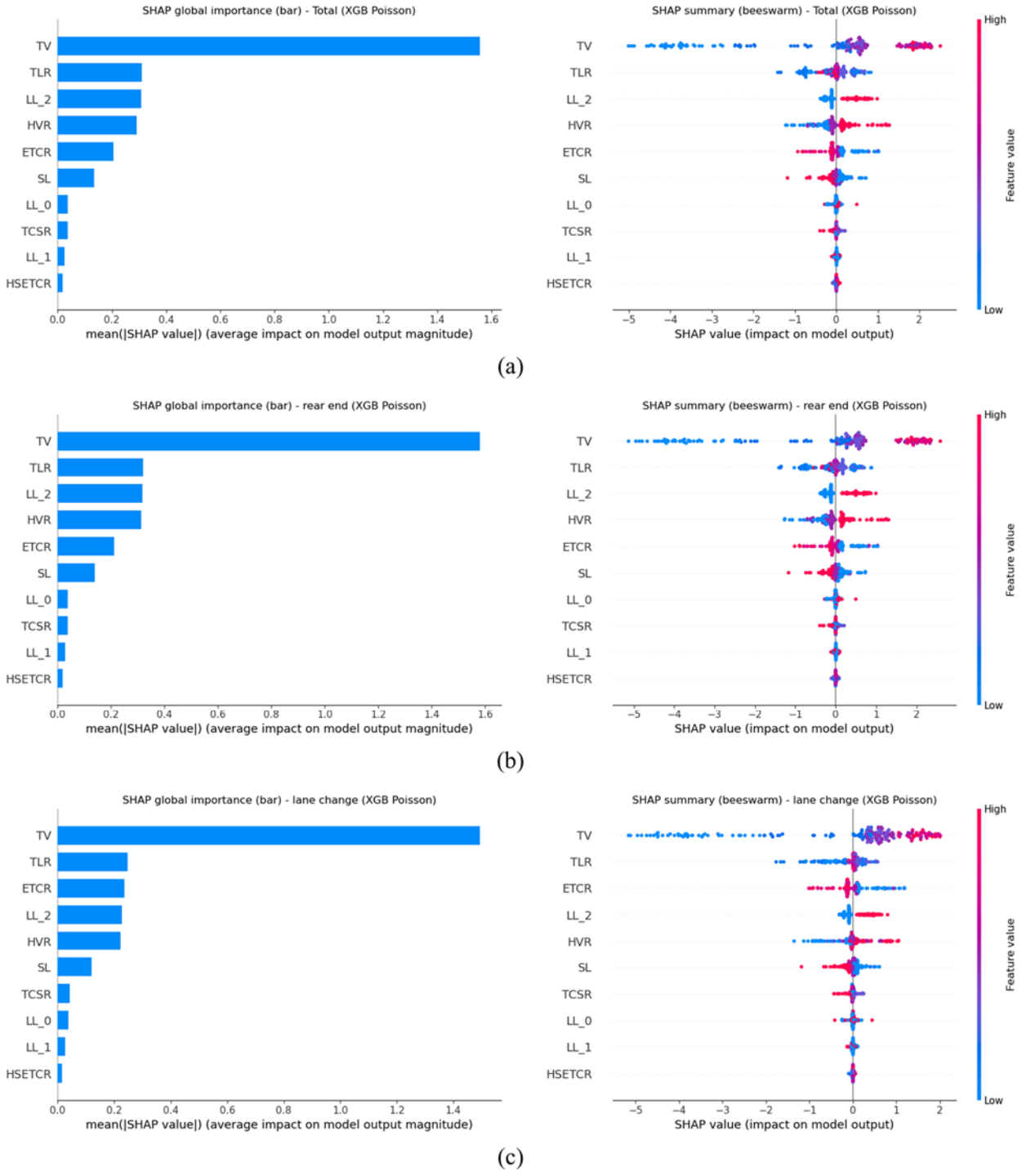

Figure 3. SHAP interpretation results for conflict frequency models. a: Total conflicts; b: rear-end conflicts; c: lane-change conflicts.

5. Discussion

Part 1 complements Part 2 by diagnosing how OT configurations shape conflict-generation mechanisms in both magnitude and spatial concentration. The scenario-based sensitivity comparisons and the spatial distribution patterns consistently show that conflicts are not uniformly produced across

the plaza area; instead, they concentrate in merging and speed-transition zones where mixed toll streams interact most intensely. This finding is important because it indicates that the safety consequences of OT design and operations are driven by localized interaction regimes created by LL, speed differentials, and merging opportunities, rather than by a single global factor. At the same time, Part 1 highlights a practical limitation of purely scenario-by-scenario SSAM assessment: although it yields intuitive, configuration-specific diagnostics, it does not directly provide a general prediction function for screening untested combinations. These observations motivate Part 2, where scenario-level conflict counts are modeled using CI/CL and interpreted using XAI to support broader OT design and operational decision-making.

The results consistently indicate that data-driven tree/boosting models provide substantially better explanatory and predictive utility for scenario-level conflict-frequency modeling in OT scenarios than conventional SPF baselines. In particular, the Poisson GLM and NB2 models show markedly weaker performance, suggesting that the count process generated under heterogeneous OT operations cannot be sufficiently represented by a simple parametric mean-variance structure alone. By contrast, DT and gradient boosting models capture nonlinearities and higher-order interactions embedded in the scenario descriptors, leading to clear improvements in WMAPE and R^2. This supports the use of CI/CL as a practical complement to simulation-based SSAM evaluation when a transferable prediction function is needed for screening OT alternatives.

Within the boosting family, the comparison between count-aware and non-count-aware objectives provides a key methodological implication. XGB with the Poisson objective achieves clearly improved accuracy relative to its squared-error formulation, indicating that a count-consistent objective can improve stability and fidelity when learning discrete event frequencies derived from SSAM. This is meaningful because the targets are nonnegative counts, and the modeling goal is not merely interpolation but robust prediction across diverse OT configurations spanning different LL regimes, SL settings, demand levels (TV), and lane-allocation descriptors. Therefore, the Poisson objective is justified as a data-consistent choice whose advantage is verified empirically in the results.

The XAI results provide actionable insights beyond global feature rankings by clarifying directionality and regime dependence. Across total, rear-end, and lane-change targets, TV is consistently the dominant driver, reinforcing exposure dominance in scenario-level conflict generation. TLR also appears as a major contributor, indicating that the relative scale of the toll-lane system to the mainline meaningfully conditions interaction pressure. In addition, LL category indicators persist among top contributors, confirming that LL regimes systematically shape the interaction structure in OT. For LTR-related composition, ETCR, HSETCR, and TCSR exhibit mixed directional diagnostics, which is consistent with strong interactions among composition, demand, and operating settings. Taken together, these patterns explain why linear SPF specifications may be inadequate and show how CI/CL coupled with SHAP can deliver SPF-like interpretability while preserving nonlinear learning capacity for OT safety evaluation.

6. Conclusions

This study proposed an integrated framework to evaluate safety in OT expressway plazas by combining microscopic simulation–based SSAM outputs, count-consistent CI/CL models, and XAI. To address sparse crash data and complex, nonlinear interaction mechanisms induced by mixed toll-lane operations, SSAM-derived conflict frequency was used as the proxy safety outcome.

Two complementary analyses were conducted. Part 1 provided diagnostic evidence that conflict risk is strongly scenario-dependent and spatially concentrated around interaction-intensive zones (e.g., merging and speed-transition areas), supporting the need for design- and operation-sensitive assessment rather than aggregate interpretations. Part 2 modeled scenario-level conflict frequencies using scenario descriptors and compared conventional count-based SPF baselines with tree and boosting-based CI/CL models. The results showed that DT and boosting models substantially outperformed Poisson GLM and NB2 in predictive accuracy, highlighting the importance of capturing nonlinearities and higher-order interactions in the OT scenario space. Among the candidate learning models, XGB with the Poisson objective provided the most reliable basis for modeling count outcomes, enabling stable prediction while respecting the distributional nature of SSAM-derived counts. Finally, SHAP-based XAI decomposed predictions into feature-level contributions, providing transparent interpretation while preserving nonlinear learning capacity. Across targets, TV consistently emerged as the dominant contributor, and TLR and LL category indicators also played persistent conditioning roles. The LTR-related composition descriptors (ETCR, HSETCR, and TCSR) exhibited interaction-driven, regime-dependent contribution patterns, offering practical insights for scenario screening and risk-informed OT planning. Overall, the proposed framework enables (i) scenario-driven diagnostics, (ii) generalizable prediction for untested OT alternatives, and (iii) feature-level explanations for decision support.

Mixed open-tolling operations are expanding worldwide, and expressway toll plazas where HSETC, ETC, and TCS lanes coexist can structurally induce safety degradation because speed heterogeneity and intensive lane-changing and merging interactions overlap within a constrained area [34–37]. Recognizing the practical constraint of crash data often being sparse and heterogeneous in such environments, this study provides an operationally relevant evaluation framework by integrating microscopic simulation with SSAM-based conflict analysis and by offering a scenario-level, generalizable predictive model with XAI-based interpretation. In particular, the diagnostic results in Part 1 show that risk concentrates in specific zones rather than being uniformly distributed, which can be directly used to prioritize design interventions. The functionalization results in Part 2 further improve decision efficiency by estimating expected conflict levels and identifying key contributing factors even for untested combinations of design and operating conditions, thereby supporting rapid scenario screening and comparative evaluation. In addition, the proposed framework can serve as a foundation for risk-informed operational strategies because it can provide supporting evidence when assessing additional countermeasures such as variable speed limits, strengthened lane guidance, and adjustments to lane-group operating schemes.

Several limitations should be noted, as this study relies on conflicts as a surrogate safety outcome. Conflicts are valuable for capturing interaction risk related to potential crash occurrence, but their alignment with observed crashes is context-dependent and may vary by network characteristics, traffic composition, driver behavior, and operational policies. Therefore, the predictions in this study should be interpreted not as direct estimates of crash risk but as evidence for relative comparison and screening of interaction risk across alternatives, which remains consistent with the intended use of the framework for design and operational decision support. To strengthen external validity, once sufficient field data with properly matched operating and geometric conditions are accumulated, it is necessary to further verify the framework's consistency by comparing simulation-derived conflict frequencies with observed crash records and trajectory-based surrogate measures [38]. Future work should also strengthen robustness by conducting sensitivity analyses of key simulation parameters, including

driver aggressiveness, and by testing how stable the findings remain under different behavioral assumptions beyond the current geometry- and operation-focused settings. In addition, further research is needed to improve predictive and generalization performance through model refinement, including the application of additional CL methodologies and systematic hyperparameter optimization. Finally, because this study is based on scenario-level aggregation, it does not sufficiently decompose time-varying risk [39]. Future studies should aggregate conflicts at finer temporal resolutions, such as 15-min intervals, and incorporate traffic flow time-series dynamics to derive more nuanced and refined risk patterns [40].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2022-00142565).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. Weng, R. Wang, M. Wang, J. Rong, Fuel consumption and vehicle emission models for evaluating environmental impacts of the ETC system, *Sustainability*, **7** (2015), 8935–8950. https://doi.org/10.3390/su7078934
2. A. Eilbert, A. Mittelman, A. M. Chouinard, M. Glaze, C. Ho, *Estimating Emission Benefits of Electronic Open-Road Tolling Conversion Projects*, *John A. Volpe National Transportation Systems Center*, Report No. 1174898, 2022. Available from: https://rosap.ntl.bts.gov/view/dot/64625.
3. N. Alemazkoor, M. Burris, Examining potential travel time savings benefits due to toll rates that vary by lane, *J. Transp. Technol*., **4** (2014), 267–286. https://doi.org/10.4236/jtts.2014.43024
4. M. Abuzwidah, M. Abdel-Aty, Crash risk analysis of different designs of toll plazas, *Saf. Sci.*, **107** (2018), 77–84. https://doi.org/10.1016/j.ssci.2018.02.024
5. M. Abuzwidah, M. Abdel-Aty, Safety assessment of the conversion of toll plazas to all-electronic toll collection system, *Accid. Anal. Prev.*, **80** (2015), 153–161. https://doi.org/10.1016/j.aap.2015.03.039
6. M. Saad, M. Abdel-Aty, J. Lee, Analysis of driving behavior at expressway toll plazas, *Transp. Res. Part F Traffic Psychol. Behav*., **61** (2019), 163–177. https://doi.org/10.1016/j.trf.2017.12.008
7. D. Lord, F. Mannering, The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, *Transp. Res. Part A Policy Pract*., **44** (2010), 291–305. https://doi.org/10.1016/j.tra.2010.02.001

8. D. Gettman, L. Head, Surrogate safety measures from traffic simulation models, *Transp. Res. Rec.*, **1840** (2003), 104–115. https://doi.org/10.3141/1840-12
9. D. Gettman, L. Pu, T. Sayed, S. G. Shelby, *Surrogate Safety Assessment Model and Validation: Final Report*, *Turner-Fairbank Highway Res. Cent.*, Report No. FHWA-HRT-08-051, 2008. Available from: https://rosap.ntl.bts.gov/view/dot/39210.
10. Y. Fei, K. Long, L. Xing, X. Pei, X. Li, L. Yao, Safety performance analysis of toll plaza diverging area based on an improved simulation platform for weak-constraint driving behaviors, *Accid. Anal. Prev.*, **220** (2025), 108177. https://doi.org/10.1016/j.aap.2025.108177
11. O. Lares, H. Zhen, J. J. Yang, Feature group tabular transformer: A novel approach to traffic crash modeling and causality analysis, *Appl. Comput. Intell.*, **5** (2025), 29–56. https://doi.org/10.3934/aci.2025003
12. J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Stat*., **29** (2001), 1189–1232. https://doi.org/10.1214/aos/1013203451
13. L. Xing, J. He, M. Abdel-Aty, Y. Wu, J. Yuan, Time-varying analysis of traffic conflicts at the upstream approach of toll plaza, *Accid. Anal. Prev.*, **141** (2020), 105539. https://doi.org/10.1016/j.aap.2020.105539
14. L. Xing, M. Abdel-Aty, Q. Cai, Comparison of different models for evaluating vehicle collision risks at upstream diverging area of toll plaza, *Accid. Anal. Prev.*, **135** (2019), 105343. https://doi.org/10.1016/j.aap.2019.105343
15. W. Xiang, C. Wang, X. Li, Q. Xue, X. Liu, Optimizing guidance signage system to improve drivers' lane-changing behavior at the expressway toll plaza, *Transp. Res. Part F Traffic Psychol. Behav*., **90** (2022), 382–396. https://doi.org/10.1016/j.trf.2022.09.008
16. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, et al., From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell*., **2** (2020), 56–67. https://doi.org/10.1038/s42256-019-0138-9
17. A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, A. Mohammadian, Toward safer highways: Application of XGBoost and SHAP for real-time accident detection and feature analysis, *Accid. Anal. Prev*., **136** (2020), 105405. https://doi.org/10.1016/j.aap.2019.105405
18. H. Al-Mahamid, D. Al-Nabulsi, A. Torok, Developing safety performance functions incorporating pavement roughness using Poisson regression and machine learning models on Jordan's Desert Highway, *Transp. Res. Interdiscip. Perspect*., **34** (2025), 101659. https://doi.org/10.1016/j.trip.2025.101659
19. Y. Ma, J. Zhang, J. Lu, S. Chen, G. Xing, R. Feng, Prediction and analysis of likelihood of freeway crash occurrence considering risky driving behavior, *Accid. Anal. Prev*., **192** (2023), 107244. https://doi.org/10.1016/j.aap.2023.107244
20. W. Wang, Y. Yang, X. Yang, V. V. Gayah, Y. Wang, J. Tang, et al., A negative binomial Lindley approach considering spatiotemporal effects for modeling traffic crash frequency with excess zeros, *Accid. Anal. Prev*., **207** (2024), 107741. https://doi.org/10.1016/j.aap.2024.107741
21. C. Wang, N. Stamatiadis, Surrogate safety measure for simulation-based conflict study, *Transp. Res. Rec.*, **2386** (2013), 72–80. https://doi.org/10.3141/2386-09
22. M. Essa, T. Sayed, Simulated traffic conflicts: Do they accurately represent field-measured conflicts, *Transp. Res. Rec.*, **2514** (2015), 48–57. https://doi.org/10.3141/2514-06

23. *Korea Expressway Corporation*, *Design Guidelines for Toll Plaza Areas Considering On-Site Toll Collection Lanes*, 2019. Available from: https://www.codil.or.kr/viewDtlCostSave.do?gubun=costsave&pMetaCode=EDGCODA00606.
24. *Transportation Research Board, Highway Capacity Manual*, 2000. Available from: https://onlinepubs.trb.org/onlinepubs/trnews/rpo/rpo.trn129.pdf.
25. S. Kim, The toll plaza optimization problem: Design, operations, and strategies, *Transp. Res. Part E Logist. Transp. Rev.*, **45** (2009), 125–137. https://doi.org/10.1016/j.tre.2008.03.004
26. H. T. Abdelwahab, Traffic micro-simulation model for design and operational analysis of barrier toll stations, *Ain Shams Eng. J.*, **8** (2017), 507–513. https://doi.org/10.1016/j.asej.2016.05.010
27. H. U. Ahmed, Y. Huang, P. Lu, A review of car-following models and modeling tools for human and autonomous-ready driving behaviors in micro-simulation, *Smart Cities*, **4** (2021), 314–335. https://doi.org/10.3390/smartcities4010019
28. A. Dijkstra, P. Marchesini, F. Bijleveld, V. Kars, H. Drolenga, M. Van Maarseveen, Do calculated conflicts in microsimulation model predict number of crashes, *Transp. Res. Rec.*, **2147** (2010), 105–112. https://doi.org/10.3141/2147-13
29. C. Caliendo, M. Guida, Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts, *J. Transp. Eng.*, **138** (2012), 1453–1467. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000473
30. L. Xing, D. Zou, Y. Fei, K. Long, J. Wang, Safety evaluation of toll plaza diverging area considering different vehicles' toll collection types, *Appl. Sci.*, **13** (2023), 9005. https://doi.org/10.3390/app13159005
31. K. El-Basyouny, T. Sayed, Safety performance functions using traffic conflicts, *Saf. Sci.*, **51** (2013), 160–164. https://doi.org/10.1016/j.ssci.2012.04.015
32. L. Xing, J. He, M. Abdel-Aty, Q. Cai, Y. Li, O. Zheng, Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data, *Accid. Anal. Prev.*, **125** (2019), 174–187. https://doi.org/10.1016/j.aap.2019.01.034
33. F. Huang, P. Liu, H. Yu, W. Wang, Identifying if VISSIM simulation model and SSAM provide reasonable estimates for field measured traffic conflicts at signalized intersections, *Accid. Anal. Prev*., **50** (2013), 1014–1024. https://doi.org/10.1016/j.aap.2012.08.018
34. P. Song, N. N. Sze, O. Zheng, M. Abdel-Aty, Addressing unobserved heterogeneity at road user level for the analysis of conflict risk at tunnel toll plaza: A correlated grouped random parameters logit approach with heterogeneity in means, *Anal. Methods Accid. Res.*, **36** (2022), 100243. https://doi.org/10.1016/j.amar.2022.100243
35. M. Abuzwidah, M. Abdel-Aty, M. Ahmed, Safety evaluation of hybrid main-line toll plazas, *Transp. Res. Rec.*, **2435** (2014), 53–60. https://doi.org/10.3141/2435-07
36. F. Zahedieh, C. Lee, Impacts of a toll information sign and toll lane configuration on queue length and collision risk at a toll plaza with a high percentage of heavy vehicles, *Vehicles*, **6** (2024), 1249–1267. https://doi.org/10.3390/vehicles6030059
37. Y. He, J. Xia, Comprehensive evaluation on traffic safety of mixed traffic flow in a freeway merging area using cloud model, *Symmetry*, **17** (2025), 855. https://doi.org/10.3390/sym17060855
38. S. Son, K. Kwon, J. Park, M. Abdel-Aty, Applying an improved calibration method in the safety evaluation framework for the open tolling system, *Transportmetrica A Transport Sci.*, **21** (2025), 2270759. https://doi.org/10.1080/23249935.2023.2270759

39. N. Park, J. Park, C. Lee, Conditional generative adversarial network-based roadway crash risk prediction considering heterogeneity with dynamic data, *J. Safety Res.*, **92** (2025), 217–229. https://doi.org/10.1016/j.jsr.2024.12.001
40. J. Park, J. Park, C. Oh, J. Jeong, S. Lee, Risk driving indicator-based safety performance estimation by various aggregation level using hard braking event data, *Sustainability*, **18** (2026), 1914. https://doi.org/10.3390/su18041914

AIMS Press