Context-aware Risk Ranking of Railway Level Crossings Using Machine Learning and National Transport Data

Sudasawan Ngammongkolwong [1,*] and Amnat Sawatnatee [2]

[1] Faculty of Digital Technology and Innovation, Southeast Bangkok University, Thailand
[2] Faculty of Science, Chandrakasem Rajabhat University, Thailand
Email: Lukmoonoy_ping@hotmail.com (S.N.); amnat.s@chandra.ac.th (A.S.)
*Corresponding author

***Abstract*—This study proposes a context-aware approach for risk ranking of railway level crossings using machine learning and national transport data. Railway level crossings are critical safety points within multimodal transportation systems, where interactions between rail operations and road traffic generate heterogeneous risk. Conventional risk assessment methods often rely on fixed-weight or expert-driven scoring schemes, which may not adequately capture context-dependent variations across diverse infrastructural, road, and spatial environments. To address these limitations, the proposed approach integrates national Geographic Information System (GIS)-based railway crossing data with machine learning and weakly supervised learning techniques. Contextual information is organized into three dimensions: infrastructural attributes, road characteristics, and spatial context. In the absence of explicit accident labels, pseudo-labels derived from domain knowledge are used to provide supervisory signals for model training. Multiple machine learning models, including logistic regression, random forest, and gradient boosting, are evaluated using ranking-oriented performance metrics to emphasize relative risk discrimination rather than binary accident prediction. The results indicate that the proposed approach generates continuous and discriminative risk scores, revealing substantial heterogeneity in relative safety risk across the railway network. A relatively small subset of crossings accounts for a disproportionate share of elevated risk, supporting targeted prioritization. Overall, the findings suggest that context-aware, data-driven risk ranking provides a scalable and interpretable alternative to traditional fixed-weight assessment methods, contributing to improved analytical understanding of safety risk patterns in railway level crossing networks.**

***Keywords*—railway level crossings, context-aware risk scoring, transportation safety, machine learning, weakly supervised learning, risk ranking, Geographic Information System (GIS)-based transport data, safety prioritization**

I. INTRODUCTION

Railway level crossings remain among the most critical conflict points within multimodal transportation systems, where interactions between rail operations and road traffic expose users to elevated safety risks. Despite long-standing engineering, regulatory, and enforcement efforts, level crossings continue to account for a disproportionate share of severe and fatal transportation accidents in many countries, particularly in mixed-traffic, peri-urban, and rapidly urbanizing environments [1–6]. These persistent safety challenges underscore the need for systematic, data-driven approaches that can support effective risk assessment and evidence-based prioritization of safety interventions at the network level [7–9]. In this context, it is important to clarify that this study does not aim to predict absolute crash probabilities at railway level crossings. Instead, the objective is to develop a relative risk ranking mechanism to support prioritization under limited labeled data conditions. In this study, the term "risk" refers to a proxy-based relative prioritization score derived from contextual features, rather than a direct measure of actual crash likelihood.

Traditional approaches to railway level crossing safety assessment have largely relied on deterministic rules, expert judgment, and Multi-Criteria Decision-Making (MCDM) frameworks. Such approaches typically integrate multiple risk-related factors—including traffic exposure, protection type, visibility conditions, train speed, road hierarchy, and surrounding land use—into composite risk indices derived from expert-defined or fixed weighting schemes [3, 10–13]. While these methods offer transparency and regulatory simplicity, they generally assume that the relative influence of risk factors remains constant across locations and operating environments.

A substantial body of transportation safety literature, however, has demonstrated that crash risk is inherently heterogeneous and strongly conditioned by local infrastructural design, traffic composition, spatial configuration, and human behavior [2, 14–17]. As a result, uniform or static weighting frameworks may obscure meaningful variation in risk, leading to suboptimal allocation of limited safety resources and reduced

Manuscript received January 5, 2026; revised March 4, 2026; accepted May 11, 2026; published August 19, 2026.

doi: 10.12720/jait.17.8.1512-1529

effectiveness of targeted interventions [8, 18, 19]. To enhance analytical rigor, several studies have introduced statistical and probabilistic modeling techniques for railway crossing safety analysis, including regression based frequency models, Bayesian approaches, and exposure-adjusted risk estimation methods [5, 12, 14, 20]. Although these approaches represent an important step toward evidence-based safety assessment, many continue to rely on predefined model structures or expert-assigned parameters and require detailed accident records that are often incomplete, underreported, or inconsistently recorded at the individual crossing level in national databases [6, 18, 19]. Recent advances in data availability, computational capacity, and machine learning methodologies have stimulated growing interest in applying data-driven techniques to transportation safety analysis. Machine learning models—including logistic regression, decision trees, random forests, and gradient boosting—have demonstrated strong capability in capturing nonlinear relationships, complex interactions, and unobserved heterogeneity in crash occurrence and severity data [16, 21–24]. In the railway safety domain, such models have been increasingly applied to crash severity analysis, hotspot identification, and infrastructure screening [25–27].

Nevertheless, much of the existing machine learning literature in transportation safety continues to emphasize accident prediction or classification accuracy, often evaluated using traditional performance metrics. While valuable from a methodological perspective, prediction-oriented models provide limited guidance for actionable safety prioritization, particularly in policy and planning contexts where relative risk ranking and interpretability are of primary importance [28–30]. Moreover, the dependence of supervised learning approaches on reliable accident labels restricts their applicability in large-scale national networks where such labels are sparse or unavailable [20, 27].

In parallel, the concept of context-aware modeling has gained prominence across disciplines such as ubiquitous computing, human-computer interaction, and intelligent systems. Context-aware systems adapt their behavior based on situational information, including location, environment, and operational conditions [31–35]. Within transportation research, context-aware approaches have been increasingly recognized as essential for capturing spatial and operational variability that influences safety outcomes [2, 17, 25]. By allowing risk contributions to vary dynamically across heterogeneous environments, context-aware models offer a more realistic representation of infrastructure safety than fixed-weight frameworks. Another critical methodological challenge concerns the scarcity of labeled outcome data in safety-critical domains. To address this limitation, weakly supervised and semi-supervised learning techniques have emerged as viable alternatives to fully supervised models. These approaches leverage indirect, noisy, or heuristic supervisory signals to infer relative risk patterns when ground-truth labels are limited or unavailable [36–38].

In transportation safety applications, weak supervision has been shown to support robust risk ranking and prioritization, particularly when the analytical objective is comparative assessment rather than precise probability estimation [24, 28]. Motivated by these developments, this study proposes a context-aware risk scoring framework for railway level crossings that integrates national Geographic Information System (GIS)-based transportation data with machine learning and weakly supervised learning techniques. Rather than assuming fixed risk weights or focusing on binary accident prediction, the proposed approach learns context-dependent risk patterns from infrastructural attributes, road characteristics, and spatial context. Continuous risk scores are generated for individual crossings and subsequently used to produce relative risk rankings that support systematic safety prioritization across the national rail network. The contributions of this study are threefold. First, it conceptualizes and operationalizes context-aware risk scoring as a data-driven alternative to traditional fixed-weight safety assessment approaches. Second, it demonstrates the use of weakly supervised machine learning to infer relative safety risk in the absence of explicit accident labels, addressing a common limitation of national transportation datasets. Third, it provides an interpretable and scalable risk ranking framework that aligns with the practical needs of transportation agencies for analytical understanding and prioritization analysis. By bridging advances in transportation safety analysis, context-aware systems, and machine learning, this study contributes to the development of adaptive, robust, and analytically relevant safety assessment tools for modern railway networks.

II. Literature Review

A. *Risk Assessment in Railway Level Crossing Safety*

Risk assessment of railway level crossings has traditionally been grounded in MCDM frameworks and expert-based scoring systems. These approaches typically integrate multiple risk-related factors—such as traffic exposure, infrastructure protection type, visibility conditions, train speed, road hierarchy, and surrounding environmental characteristics—into composite risk indices to support safety evaluation and prioritization [3, 5, 25]. Owing to their relative simplicity, transparency, and compatibility with administrative decision processes, such models have been widely adopted by transportation authorities for practical safety management and planning purposes [7, 8]. In many applications, expert judgment plays a central role in defining factor weights and aggregation rules within MCDM-based frameworks. While expert-driven approaches benefit from domain knowledge and operational experience, they are inherently constrained by subjectivity and may not generalize well across heterogeneous operating environments [11, 39].

As railway level crossings are embedded in diverse spatial and traffic contexts, reliance on fixed expert-defined weights can limit the sensitivity of risk assessments to local conditions and emerging safety patterns [4, 17]. As illustrated in Fig. 1, traditional railway level crossing safety assessment frameworks rely on multiple contextual factors, including traffic volume, visibility conditions, road geometry, signaling devices, and crossing environment characteristics. However, these approaches often employ fixed expert-defined weighting schemes that may not adequately capture local variations and emerging risk patterns.

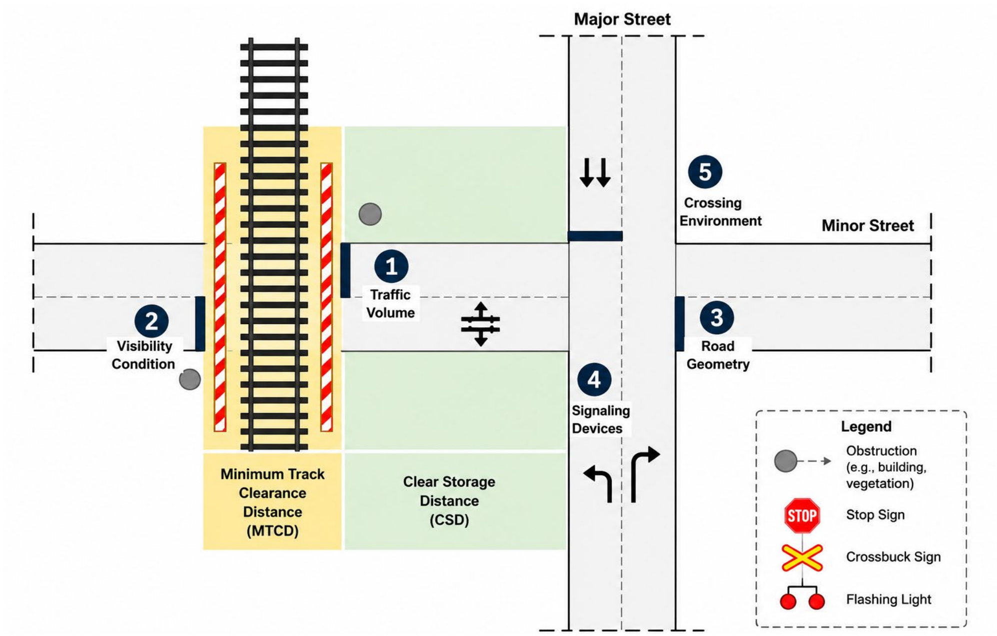

Fig. 1. Conceptual illustration of key risk-related factors considered in railway level crossing safety assessment frameworks.

However, conventional MCDM and expert-driven models generally assume fixed and linear relationships between contributing factors and safety risk. This assumption limits their capacity to respond effectively to contextual diversity across railway crossings. In practice, the influence of individual risk factors may vary substantially depending on spatial setting, surrounding land use, road classification, and operational environment. As a result, crossings with similar infrastructure features may exhibit markedly different risk profiles when situated in urban, peri-urban, or rural contexts—differences that static weighting schemes are unable to adequately capture [3, 4, 25]. To address these limitations, several studies have introduced statistical modeling techniques and probabilistic risk assessment methods to enhance the analytical rigor of railway crossing safety evaluation. These methods include regression-based models, Bayesian approaches, and exposure-based risk estimation techniques that seek to quantify uncertainty, variability, and rare-event characteristics in accident occurrence [5, 8, 14]. Such approaches represent an important step toward evidence-based safety assessment by grounding risk estimation in observed data rather than purely subjective judgment.

Nevertheless, many statistical and probabilistic models continue to rely on predefined model structures or expert-assigned parameters, thereby retaining an inherent level of subjectivity and limiting adaptability. Moreover, their dependence on detailed accident records constrains scalability and transferability, particularly when applied to large-scale national railway networks characterized by heterogeneous operational and spatial conditions and incomplete incident data [20, 40]. These limitations highlight the challenges of applying traditional and semi-traditional risk assessment approaches in contemporary, data-rich yet label-scarce transportation environments. Consequently, there is a growing recognition that conventional risk assessment approaches for railway level crossings require methodological advancement toward more adaptive and data-driven frameworks. Such frameworks should be capable of capturing complex interactions among infrastructural, environmental, and spatial factors while reducing reliance on static assumptions and subjective weighting. This need has motivated increasing research interest in machine learning–based risk assessment methods, which offer the potential to learn context-dependent risk patterns directly from empirical data and to support more robust and scalable safety prioritization at the network level [24, 27, 29].

B. Machine Learning in Transportation Safety Analysis

Machine learning has been increasingly adopted in transportation safety research to address complex analytical tasks such as crash occurrence prediction, injury severity modeling, and infrastructure-related risk assessment. Compared with traditional statistical approaches, machine learning models—including logistic regression, random forests, and gradient boosting techniques—have demonstrated strong capability in capturing nonlinear relationships, high-dimensional feature interactions, and unobserved heterogeneity among safety-related variables [14, 20, 24]. As a result, these approaches have been widely applied across diverse transportation contexts, including road segments, intersections, and rail-related facilities [41, 42].

Despite these methodological advances, a substantial proportion of existing studies continue to emphasize predictive accuracy as the primary evaluation criterion, often assessed through classification or regression performance metrics such as accuracy, precision, recall, or mean squared error. While such metrics are valuable for benchmarking model performance, they provide limited insight into how model outputs can be translated into interpretable and actionable risk measures for safety management and policy analysis [8, 28]. In particular, highly flexible or black-box modeling approaches may obscure the relative contribution of contextual factors, thereby constraining their practical usefulness for prioritizing infrastructure interventions and communicating risk to stakeholders [29, 30]. Another critical limitation concerns data availability and labeling quality. Supervised machine learning approaches typically require reliable and comprehensive accident or incident labels at the unit of analysis, such as individual road segments or railway crossings. However, in many national transportation systems, accident records at the micro-infrastructure level are sparse, incomplete, or inconsistently reported, particularly for low-frequency but high-consequence events [20, 39]. These data constraints restrict the applicability of fully supervised learning frameworks and reduce their scalability for large-scale, network-level safety assessment [27, 40].

In response to these challenges, recent research has begun to explore alternative learning paradigms, including weakly supervised, semi-supervised, and risk scoring-oriented approaches, which aim to infer relative safety risk without relying exclusively on explicit accident outcomes. These methods leverage indirect indicators, surrogate measures, or rule-informed heuristics to construct informative training signals in data-constrained environments [36, 37]. By shifting the analytical focus from outcome prediction toward risk representation and prioritization, such approaches align machine learning outputs more closely with the operational needs of transportation authorities and infrastructure managers [7, 11]. Collectively, these developments highlight a methodological transition in transportation safety research toward decision-oriented machine learning frameworks that balance predictive capability, interpretability, and practical relevance. Risk scoring and ranking approaches offer a promising pathway for integrating heterogeneous transportation data with machine learning techniques to support scalable, context-sensitive safety assessment and prioritization at the network level [8, 29]. This emerging direction provides a critical foundation for the context-aware risk scoring framework proposed in the present study.

C. Context-Aware Risk Modeling

Context-aware modeling originates from research in information systems and ubiquitous computing, where system behavior dynamically adapts to changes in contextual information such as location, environment, temporal conditions, and user interaction [32, 34, 43]. Within this paradigm, context is broadly understood as any information that can be used to characterize the situation of an entity and its surrounding environment, enabling systems to respond intelligently to varying operational conditions. This conceptualization emphasizes adaptability, situational awareness, and interaction sensitivity as core design principles of context-aware systems [35, 44].

Over the past two decades, the concept of context awareness has been extended beyond computing systems to support analysis in complex socio-technical domains, including risk assessment, infrastructure management, and safety-critical systems. In such domains, contextual information provides essential cues for understanding how hazards emerge, propagate, and interact with system components under different environmental and operational conditions [45, 46]. Context-aware approaches are therefore increasingly viewed as foundational for modeling risk in environments characterized by heterogeneity, uncertainty, and dynamic interactions. In the context of risk assessment, context-aware modeling explicitly recognizes that risk factors do not operate uniformly across settings. Instead, their influence varies according to environmental, infrastructural, and spatial conditions. For transportation infrastructure, identical physical attributes may correspond to markedly different risk levels depending on road hierarchy, surrounding land use, traffic composition, and spatial configuration [11, 39]. Context-aware risk modeling seeks to capture these heterogeneous interactions by allowing risk contributions to adapt dynamically rather than relying on fixed or globally defined weighting schemes [17, 25, 27].

Recent studies in transportation safety and infrastructure management provide growing empirical evidence that incorporating contextual awareness can enhance both model performance and interpretability. By explicitly modeling context-dependent relationships, such approaches yield more realistic representations of safety risk and improve the identification of dominant risk contributors under specific operational conditions [20, 24]. Spatially and contextually informed models have also been shown to better capture unobserved heterogeneity, a critical issue in transportation safety analysis that often undermines the validity of uniform risk models [8, 14]. From a practical perspective, context-aware frameworks align more closely with the needs of transportation agencies and infrastructure managers by supporting localized and differentiated risk prioritization rather than uniform network-wide assessment. Such localized insights are essential for designing targeted interventions and optimizing the allocation of limited safety resources across complex infrastructure networks [7, 29]. Context-aware risk modeling, therefore, bridges the gap between analytical rigor and policy relevance by translating heterogeneous risk patterns into actionable decision support.

Despite these advantages, empirical applications of context-aware risk scoring at the national infrastructure level remain limited. Existing studies predominantly focus on roadway segments, intersections, or urban traffic environments, with relatively few addressing railway level crossings at scale [5, 42]. The scarcity of national-level implementations is further compounded by data limitations, spatial inconsistency, and the frequent absence of explicit outcome labels, which constrain the applicability of fully supervised learning approaches [20, 40]. Consequently, a clear research gap remains in developing and validating context-aware risk scoring frameworks for railway level crossings using national transport data. Addressing this gap requires methodological approaches capable of learning context-dependent risk patterns from heterogeneous and partially labeled data while maintaining interpretability and relevance for safety planning. The present study responds directly to this need by proposing a context-aware risk scoring framework that integrates multidimensional contextual information with machine learning techniques to support national-level railway safety prioritization.

To clearly position the contribution of this study within existing research, a comparative summary of prior approaches is presented in Table I.

While prior studies have contributed significantly to railway level crossing risk assessment, most existing approaches exhibit inherent limitations. MCDM and rule-based frameworks rely on fixed weights, limiting their adaptability to heterogeneous contexts. Statistical models require reliable accident data, which are often unavailable or incomplete at the national level. Supervised machine learning approaches, while flexible, remain dependent on labeled data and are typically designed for prediction rather than prioritization.

TABLE I. COMPARISON OF EXISTING APPROACHES AND THE PROPOSED APPROACH

Study	Method	Uses Real Labels	Context-Aware	Ranking-Based	Weak Supervision	Key Limitations
Traditional MCDM-Based Approaches	MCDM/Rule-Based	No	Limited	Yes	No	Fixed weights, limited adaptability
Statistical Risk Models	Statistical Model	Yes	Limited	No	No	Requires reliable crash data
Machine Learning Prediction Models	Machine Learning Prediction	Yes	Moderate	No	No	High data dependency
Proposed Study	Weakly Supervised ML Framework	No	Yes	Yes	Yes	Dependent on proxy-based labeling

In contrast, the proposed framework introduces a context-aware, weakly supervised learning approach that enables risk scoring and ranking without requiring explicit accident labels. By integrating contextual feature interactions and ranking-oriented evaluation, the framework provides a more scalable and policy-relevant solution for national-level safety prioritization.

Furthermore, the proposed approach advances existing frameworks in three key aspects: (1) it captures context-dependent risk patterns without relying on fixed or expert-defined weights, (2) it enables learning under limited or absent labeled data through weak supervision, and (3) it explicitly focuses on ranking-based risk prioritization, which is more aligned with real-world safety decision-making needs.

III. MATERIALS AND METHODS

A. Research Objectives

To develop and propose a context-aware risk scoring framework for railway level crossings by leveraging machine learning techniques and national transport data.

To analyze and compare the influence of infrastructure attributes, road characteristics, and spatial context on the risk ranking of railway level crossings using machine learning models, to support safety prioritization.

B. Methodology

1) Data source and description

The dataset used in this study was obtained from the Information and Communication Technology Center, Office of the Permanent Secretary, Ministry of Transport, Thailand. The data were collected in 2023 and consist of a national GIS-based inventory of railway level crossings across Thailand. As an official administrative dataset maintained by the national transport authority, the data provide comprehensive and standardized information suitable for network-level safety analysis and analytically relevant risk assessment [47]. The dataset covers many railway level crossings distributed across the national rail network and includes detailed attributes describing infrastructural characteristics, road properties, and spatial context. The availability of such multidimensional information is critical for capturing the heterogeneous operational environments in which railway crossings are situated, particularly in countries where crossings span urban, peri-urban, and rural settings [4, 5, 25]. National-scale GIS-based infrastructure inventories of this nature have been widely recognized as essential foundations for context-aware transportation safety analysis and strategic infrastructure planning [8, 42]. To ensure transparency and methodological rigor, a comprehensive data dictionary accompanies the dataset, providing formal definitions and descriptions for all variables used in the analysis. The availability of structured metadata facilitates consistent feature interpretation, reduces ambiguity in variable usage, and enhances the reproducibility and transferability of the proposed methodology across different study contexts [20, 24]. Such documentation is increasingly emphasized in data-driven transportation research, particularly where complex modeling frameworks are applied to administrative data sources [17].

Consistent with the objectives of this study, variables are organized into three principal categories reflecting distinct contextual dimensions of railway crossing risk. Infrastructure attributes include crossing type, protection mechanisms (e.g., passive signs, warning lights, and barriers), and rail configuration, which directly influence exposure and conflict dynamics between rail and road users [3, 5]. Road characteristics, such as road classification, number of lanes, and surface type, describe traffic exposure, vehicle behavior, and interaction intensity at crossings, all of which have been shown to condition safety outcomes [11, 39]. Spatial context variables encompass urban-rural setting, proximity to nearby intersections, and surrounding land use indicators, capturing environmental and locational factors that shape driver expectations and traffic complexity [4, 25]. These contextual dimensions are explicitly modeled to operationalize the concept of context-aware risk scoring, allowing the contribution of individual risk factors to vary across heterogeneous spatial and infrastructural environments rather than remaining fixed across the network. This design aligns with recent advances in transportation safety research that emphasize heterogeneity-aware and interaction-sensitive modeling frameworks [17, 27]. Notably, the dataset does not include explicit accident or incident labels at the individual crossing level. This limitation reflects a common characteristic of national transportation safety databases, where incident records are often incomplete, aggregated, or inconsistently linked to infrastructure inventories [20, 40]. As a result, the direct application of fully supervised machine learning approaches is constrained, necessitating alternative modeling strategies capable of inferring relative risk patterns from contextual features without relying exclusively on observed accident outcomes [36, 37].

By leveraging nationally administered transport data and systematically structuring variables according to infrastructural, road, and spatial contexts, this study establishes a robust empirical foundation for learning context-dependent risk weights and generating interpretable risk scores. This data design directly supports the threefold contribution of the study: conceptualizing and operationalizing context-aware risk scoring, demonstrating data-driven contextual weighting through machine learning, and providing an interpretable risk scoring and ranking mechanism to inform safety prioritization and evidence-based policy planning [7, 8].

2) Conceptualizing context-aware risk scoring

In this study, context-aware risk scoring is conceptualized as a data-driven approach in which the contribution of individual railway level crossing attributes to overall safety risk is not predefined but adaptively learned from contextual patterns embedded in empirical data. This formulation departs from conventional risk assessment frameworks that rely on fixed or expert-assigned weights and implicitly assume homogeneous risk mechanisms across locations. Instead, the proposed approach recognizes that safety risk emerges from dynamic interactions among infrastructure characteristics, road conditions, and the surrounding operational environment, and that these interactions may vary substantially across different spatial and functional contexts [35, 44, 48].

The notion of context-awareness originates from research in ubiquitous and pervasive computing, where system behavior is designed to adapt in response to changes in environmental, spatial, and situational conditions [32, 44]. In transportation safety analysis, this concept has gained increasing relevance as empirical studies have documented pronounced heterogeneity in crash risk patterns that cannot be adequately explained by infrastructure attributes alone. Evidence consistently shows that crossings or road facilities with similar physical designs may exhibit markedly different safety outcomes depending on traffic exposure, user behavior, land-use characteristics, and broader spatial settings [2, 16, 49].

Within the proposed approach, context is operationalized as a multidimensional construct encompassing three interrelated dimensions: infrastructural attributes, road characteristics, and spatial context. Infrastructure-related features describe the physical and operational properties of railway level crossings, such as crossing configuration and protection mechanisms, which establish baseline exposure and conflict potential. Road-related characteristics capture traffic-related conditions, including road classification and lane structure, that influence vehicle behavior and interaction dynamics at the rail–road interface. Spatial context represents the broader environment in which crossings are embedded, including urbanization level and surrounding land-use patterns, which have been shown to systematically condition safety risk across transportation facilities [2, 11, 12].

To clarify the operational definition of these contextual dimensions, Table II summarizes the main categories of features incorporated into the proposed approach, along with representative variables and their relevance to railway level crossing risk assessment. Presenting this information in tabular form enhances transparency and supports reproducibility by explicitly linking conceptual dimensions to measurable attributes. Importantly, the relative influence of infrastructural, road, and spatial factors is not assumed to be uniform across the network. Prior transportation safety research has demonstrated that risk factors may exert different effects depending on location-specific conditions, leading to non-linear and context-dependent risk mechanisms [16, 50]. By modeling these contextual dimensions simultaneously, the proposed approach allows risk contributions to vary dynamically, thereby capturing heterogeneous interactions that are often obscured in fixed-weight or aggregate risk assessment approaches.

Rather than focusing on binary accident prediction, the framework emphasizes continuous risk scoring and ranking. In the proposed approach, risk is defined as a proxy-based relative score derived from contextual and rule-based indicators, rather than an absolute measure of accident probability. The generated scores should therefore be interpreted as relative prioritization signals. This orientation reflects both methodological and practical considerations in transportation safety analysis. Accident occurrence data at the individual crossing level are frequently sparse, unstable, or incomplete, limiting the reliability of probability-based prediction models for network-level analysis [19, 49]. In contrast, continuous risk scores enable relative comparison across a large number of infrastructure sites and support systematic prioritization under constrained resource conditions [7]. Such ranking-oriented representations are particularly well suited for policy and planning contexts, where the primary analysis objective is to identify locations with comparatively elevated risk rather than to estimate precise accident probabilities.

TABLE II. SUMMARY OF CONTEXTUAL FEATURE CATEGORIES USED IN THE CONTEXT-AWARE RISK SCORING FRAMEWORK

Contextual Category	Example Variables	Description and Relevance to Risk Assessment
Infrastructure Attributes	Crossing type, protection mechanism (e.g., passive signs, warning lights, barriers), rail configuration	Represent the physical and operational characteristics of railway level crossings that directly influence conflict exposure and baseline safety conditions. These attributes determine the level of protection provided to road users and form a foundational component of crossing risk.
Road Characteristics	Road classification, number of lanes, surface type	Describe the functional role and traffic conditions of roads intersecting railway lines. Road characteristics affect traffic volume, vehicle behavior, and interaction dynamics between rail and road users, thereby conditioning safety risk at crossings.
Spatial Context	Urban–rural setting, proximity to nearby intersections, and surrounding land use	Capture the broader environmental and locational context in which railway crossings are embedded. Spatial factors influence driver expectations, traffic complexity, and accessibility patterns, contributing to context-dependent variations in safety risk.

The overall logic of the proposed context-aware risk scoring methodology, from contextual data inputs to risk ranking and safety prioritization, is illustrated in Fig. 2.

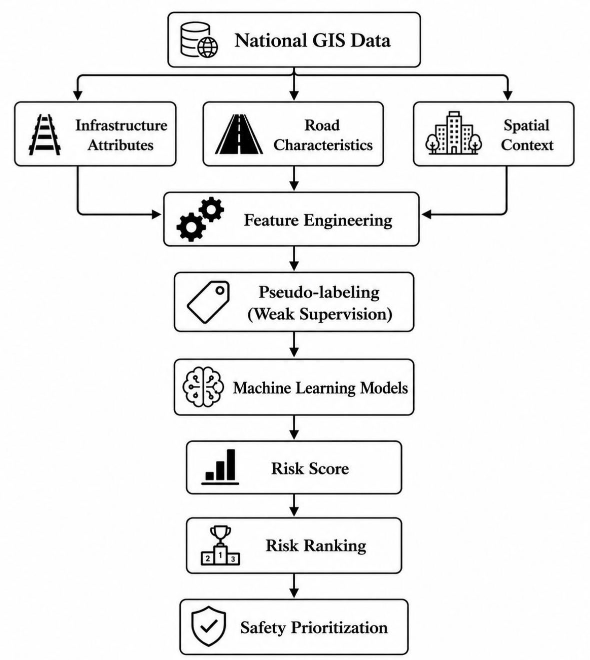

Fig. 2. Conceptual framework of the proposed context-aware risk scoring methodology for railway level crossings.

Recent advances in machine learning further support the adoption of context-aware risk scoring by enabling flexible modeling of complex and non-linear relationships among heterogeneous features. Machine learning-based safety studies have shown that context-dependent representations can enhance discrimination of relative risk across transportation networks, particularly when interactions among infrastructure, traffic, and spatial factors are pronounced [23, 51]. By integrating these capabilities within a context-aware conceptual framework, this study provides a principled foundation for learning data-driven risk weights that reflect localized operating conditions while maintaining scalability and interpretability for network-level safety assessment.

3) Pseudo-labeling and weakly supervised learning

Due to the absence of explicit accident or incident labels at the individual railway crossing level, this study adopts a weakly supervised learning approach based on pseudo-labeling. The pseudo-labels were generated using rule-based criteria derived from railway safety characteristics.

A crossing was assigned a high-risk label if it satisfied at least two of the following conditions: (1) absence of active protection mechanisms, (2) location on high-traffic or major roads, and (3) presence in dense or urban areas. These conditions were designed to reflect commonly recognized risk factors in railway safety analysis.

In cases where multiple rules produced conflicting signals, a majority voting strategy was applied to determine the final pseudo-label. This approach ensures consistency in label assignment while preserving the interpretability of the rule-based system.

The pseudo-labeling process covered approximately 100% of the dataset, with 66.67% classified as high-risk and 33.33% as low-risk crossings. This distribution reflects the inherent imbalance typically observed in transportation safety data.

4) Model development and experimental setup

To ensure reproducibility and robustness of the proposed approach, a structured experimental design was adopted.

The dataset was divided into training and testing sets using an 80:20 random split. To further evaluate model stability and generalization performance, a 5-fold cross-validation procedure was applied during training. Model performance metrics were averaged across folds to reduce variance caused by random data partitioning.

Before model training, data preprocessing was performed. Missing values were handled using appropriate imputation techniques based on variable types. Categorical variables were transformed using one-hot encoding, and numerical features were normalized to ensure consistent scaling across variables.

Two machine learning models were employed: Random Forest and Extreme Gradient Boosting (XGBoost). Hyperparameters for both models were optimized using grid search. For Random Forest, key parameters including the number of trees, maximum depth, and minimum samples per leaf were tuned. For XGBoost, parameters such as learning rate, maximum depth, and number of estimators were adjusted to achieve optimal performance.

To address class imbalance, class weighting techniques were incorporated during model training to reduce bias toward the majority class and improve the detection of high-risk crossings. In addition, cost-sensitive learning considerations were applied by assigning higher penalties to the misclassification of high-risk cases.

This approach is particularly important in safety-critical systems, where false negatives (i.e., failing to identify high-risk crossings) may lead to severe consequences. Therefore, the model training process was designed to prioritize the detection of high-risk crossings over overall classification accuracy.

Furthermore, sensitivity analysis was conducted to evaluate the robustness of the model under imbalanced data conditions and feature perturbations. The results indicate that the proposed framework maintains stable ranking performance, suggesting its robustness in real-world transportation safety scenarios characterized by class imbalance.

5) Data preprocessing and feature engineering

Data preprocessing constitutes a critical step in preparing the national transport dataset for machine learning–based risk analysis. The preprocessing procedures include handling missing values, normalizing continuous variables, and encoding categorical attributes using appropriate transformation techniques. Missing data are addressed through context-sensitive imputation strategies to preserve the statistical characteristics of the dataset, while continuous variables are normalized to ensure numerical stability and comparability across features. Categorical attributes are transformed using encoding methods such as one-hot encoding to enable their integration into machine learning models that require numerical input [8, 24].

Feature engineering focuses on constructing meaningful representations of contextual information that reflect the heterogeneous operating environments of railway level crossings. Rather than relying solely on raw variables, selected features are transformed or aggregated to better capture their functional relevance to safety risk. For example, road-related attributes are grouped into hierarchical classes based on road function and traffic characteristics, enabling the model to distinguish between major arterial roads and local access roads. Similarly, spatial indicators are derived from GIS attributes to represent contextual factors such as urban–rural setting, proximity to nearby intersections, and surrounding land use patterns, which have been shown to influence transportation safety outcomes [4, 25].

Importantly, contextual categories are not treated as isolated variables within the proposed approach. Instead, infrastructural, road, and spatial features are modeled as interacting feature sets whose combined effects collectively shape safety risk. This design reflects the underlying assumption of context-aware risk modeling that the contribution of individual factors is conditional on the broader context in which a railway crossing is embedded.

By enabling interactions among contextual dimensions, the feature engineering process supports the learning of context-dependent risk patterns and enhances the interpretability of the resulting risk scores [52].

To formally represent the proposed context-aware risk scoring framework, the risk score of a railway level crossing iii is defined as a function of multiple contextual dimensions:

$$R_i = f\left(X_i^{infra}, X_i^{road}, X_i^{spatial}\right) \tag{1}$$

where R_i denotes the relative risk score of crossin i and X_i^{infra} , X_i^{road} , and $X_i^{spatial}$ represent feature vectors derived through preprocessing and feature engineering that capture infrastructural attributes, road characteristics, and spatial context, respectively.

For notational convenience, the formulation can be compactly written as Eq. (2):

$$R_i = f_\theta\left(X_i\right) \tag{2}$$

where $X_i = \left[X_i^{infra}, X_i^{road}, X_i^{spatial}\right]$ denotes the aggregated contextual feature representation, and is a data-driven mapping learned by machine learning models. This formulation emphasizes that risk contributions are not fixed but emerge from interactions among contextual dimensions, thereby enabling context-aware risk scoring and ranking.

6) Pseudo-labeling and weakly supervised learning

Given the absence of explicit accident or incident labels at the individual railway level crossing level, this study adopts a weakly supervised learning strategy to enable model training and risk inference. Weak supervision provides a practical alternative to fully supervised learning when labeled outcome data are scarce, incomplete, or unreliable—a common condition in large-scale transportation safety datasets [20, 52].

Pseudo-labels are generated using a rule-informed heuristic derived from established findings in railway crossing safety literature. Specifically, configurations that have been consistently associated with elevated risk—such as inadequate protection mechanisms at crossings located on high-class or high-traffic roads—are used as indicators of relatively higher risk levels [5, 53]. These heuristic rules are not intended to represent deterministic accident outcomes but rather to provide coarse supervisory signals that guide the learning process. The pseudo-labeling process based on domain knowledge rules and its role within the weakly supervised learning framework are illustrated in Fig. 3.

Context-aware features derived from national GIS data are used to formulate heuristic rules informed by prior railway safety studies. These rules generate pseudo-labels that provide coarse supervisory signals for training machine learning models. The trained models estimate relative risk scores, which are subsequently used for risk ranking to support safety prioritization, rather than direct accident prediction.

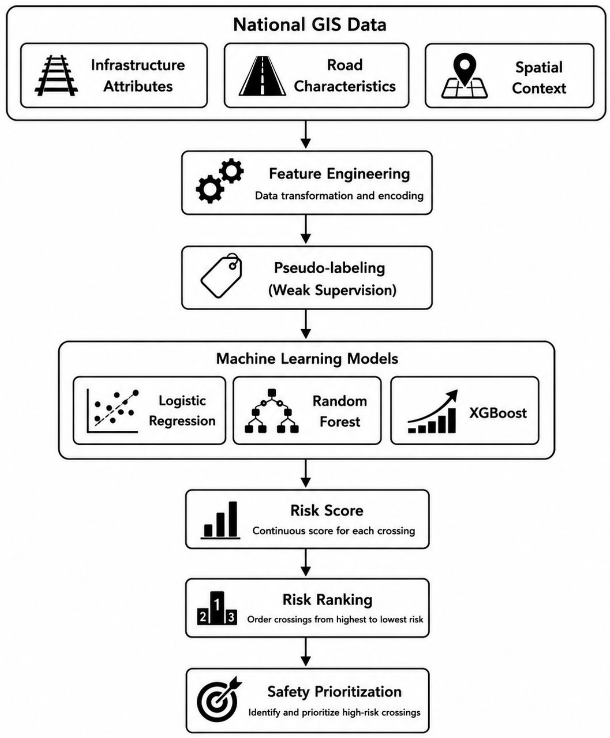

Fig. 3. Pseudo-labeling framework based on domain knowledge and weak supervision.

Importantly, the use of pseudo-labeling in this study does not imply an attempt to replicate true accident occurrence or severity. Instead, pseudo-labels serve as proxies that allow machine learning models to learn relative risk patterns embedded in contextual features. This distinction is critical, as the objective of the proposed approach is risk scoring and prioritization rather than accident prediction. Similar pseudo-labeling and weak supervision approaches are effective in risk assessment and anomaly detection contexts, particularly when ground-truth labels are limited or unavailable [36, 37].

By integrating pseudo-labeling with context-aware feature representations, the proposed approach enables the estimation of context-dependent risk weights while maintaining methodological transparency. This strategy balances practical data constraints with the need for data-driven learning and supports scalable risk assessment across national railway networks without relying on complete accident histories.

7) Machine learning models and evaluation

To operationalize the proposed context-aware risk scoring framework, this study evaluates three classes of machine learning models representing different trade-offs between model flexibility, predictive capacity, and interpretability. The selection of models is guided by two complementary objectives: (i) learning complex, context-dependent risk patterns from heterogeneous infrastructural, road, and spatial data, and (ii) supporting transparent and reliable risk prioritization suitable for network-level analysis. This comparative, multi-model strategy is consistent with contemporary transportation safety research, which emphasizes the evaluation of multiple modeling paradigms to assess robustness and practical relevance rather than reliance on a single predictive approach [8, 14, 16, 20].

Logistic regression is employed as a baseline model due to its interpretability, computational efficiency, and extensive use in transportation safety analysis. As a generalized linear modeling framework, logistic regression provides a transparent benchmark for examining the direction and relative magnitude of contextual feature effects, thereby facilitating comparison with more flexible machine learning models [2, 8, 14]. Although its linear structure limits the ability to capture complex nonlinear interactions, logistic regression remains valuable for assessing whether increased model complexity yields substantive improvements in risk discrimination and ranking performance [12, 28].

To capture nonlinear relationships and higher-order interactions among contextual features, random forest models are incorporated into the evaluation framework. Random forests construct an ensemble of decision trees using bootstrap aggregation and randomized feature selection, yielding models that are robust to noise, multicollinearity, and overfitting—conditions commonly observed in large-scale transportation safety datasets [21, 24, 54].

In addition to strong empirical performance, random forests provide variable importance measures that enable post hoc interpretation of context-dependent risk contributions, making them particularly suitable for exploratory safety analysis and infrastructure prioritization tasks [20, 29].

Gradient boosting models are further employed to enhance the learning of complex and heterogeneous risk patterns. Implemented using the XGBoost algorithm, gradient boosting iteratively improves model predictions by emphasizing observations associated with higher residual errors, thereby enabling effective modeling of nonlinear effects and intricate feature interactions across diverse contextual environments [22, 26].

Prior transportation safety studies have demonstrated that boosting-based models often outperform alternative approaches in ranking and prioritization contexts, where subtle contextual differences among sites are critical [25, 27, 41]. Although gradient boosting models are less inherently interpretable than linear models, recent advances in explainable machine learning—such as feature importance analysis and partial dependence visualization—provide complementary tools for understanding learned risk structures [29, 55].

Model evaluation is explicitly designed to align with the prioritization objective of risk scoring rather than with binary accident prediction accuracy alone. Accordingly, emphasis is placed on ranking-oriented evaluation metrics, including the Area Under the Curve (AUC) and rank correlation measures. These metrics assess a model's ability to discriminate relative risk levels and to produce stable and coherent risk rankings across railway level crossings, which is essential for network-level safety planning and decision support [11, 25, 28]. Ranking-based evaluation is particularly appropriate in contexts where absolute accident frequencies are unavailable, incomplete, or unreliable, and where decision-makers require relative ordering of sites to guide safety intervention strategies [8, 12, 39].

By systematically comparing models across different levels of complexity and evaluating them using ranking-focused criteria, this study ensures that model selection balances predictive capability, interpretability, and practical relevance. This evaluation strategy supports the identification of models that not only perform well empirically but also provide actionable insights for context-aware risk assessment and infrastructure prioritization at the national transportation network level. Such an approach aligns with best practices in transportation safety analytics and enhances the credibility of data-driven decision support for policy and planning applications [7, 20].

In the proposed approach, context-awareness is operationalized through the integration of heterogeneous feature groups, including infrastructural attributes, road characteristics, and spatial/environmental factors. These features are jointly modeled using machine learning techniques capable of capturing nonlinear relationships and higher-order interactions. This allows the influence of individual risk factors to vary across different contexts, rather than being constrained by fixed or predefined weights.

IV. Results and Discussion

A. Results

1) Overview of context-aware risk scoring outputs

The proposed context-aware risk scoring framework was applied to a national GIS-based railway level crossing dataset obtained from the Department of Rail Transport, Ministry of Transport, Thailand, to generate relative risk estimates at the network level. For each railway level crossing, the framework produced a continuous risk score representing its relative safety risk within the national rail network. These scores were subsequently used to generate a ranked list of crossings, enabling systematic comparison and prioritization across heterogeneous infrastructural, road, and spatial contexts. It should be noted that the reported scores represent proxy-based prioritization measures and should not be interpreted as absolute crash risk estimates. The distribution of the generated risk scores exhibits substantial variability across the network.

The distribution of the generated risk scores exhibits substantial variability across the network, indicating pronounced heterogeneity in relative safety risk among railway level crossings. Rather than clustering around a narrow range, scores span a broad spectrum, reflecting the combined influence of infrastructural design, road characteristics, and spatial environment. This variability suggests that the proposed approach effectively differentiates crossings based on contextual risk patterns, rather than assigning uniform or near-uniform risk levels. To further illustrate this pattern, Fig. 4 presents the empirical distribution of context-aware risk scores across the national railway level crossing network.

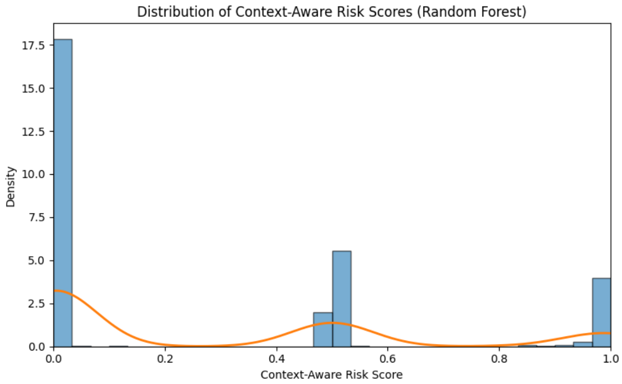

Fig. 4. Distribution of context-aware risk scores generated by the proposed approach.

Further examination of the ranked risk scores reveals a concentration effect, whereby a relatively small subset of railway level crossings accounts for a disproportionate share of high-risk values. Crossings located at the upper end of the ranking exhibit markedly higher relative risk compared to the majority of the network. This concentration pattern highlights the practical utility of the proposed approach, as it enables decision-makers to prioritize a limited number of crossings for targeted safety interventions where prioritization analysis is likely to yield the greatest marginal benefit.

Importantly, the continuous nature of the risk scores allows for fine-grained discrimination among crossings with similar physical characteristics but differing contextual conditions. This capability is essential for network-level safety prioritization, where decision-makers must allocate limited resources among numerous sites exhibiting varying degrees of relative risk.

2) Model performance in risk ranking

To assess the effectiveness of machine learning models in producing meaningful risk rankings, logistic regression, random forest, and XGBoost models were evaluated using ranking-oriented performance metrics, with particular emphasis on the AUC. Table II summarizes the approximate performance ranges observed across models. Overall, all evaluated models demonstrate the ability to discriminate relative risk among railway level crossings; however, notable differences in ranking performance are observed across modeling approaches.

Logistic regression, used as a baseline interpretable model, exhibits moderate ranking capability, with AUC = 0.68 ± 0.02. While its linear structure limits the modeling of complex contextual interactions, logistic regression provides a transparent reference for assessing the added value of more flexible machine learning approaches.

Random forest models demonstrate improved ranking performance, with AUC = 0.75 ± 0.02. This performance gain reflects the model's ability to capture nonlinear relationships and higher-order interactions among infrastructural attributes, road characteristics, and spatial context, while maintaining reasonable robustness under weak supervision.

Gradient boosting models (XGBoost) achieve the strongest ranking performance among the evaluated approaches. Across model configurations, AUC values for gradient boosting achieve the highest performance, with AUC = 0.80 ± 0.02, indicating superior discrimination of relative risk levels across heterogeneous railway level crossing environments. The performance advantage of gradient boosting suggests that sequential learning and adaptive weighting of contextual features are particularly effective for context-aware risk scoring tasks. Despite differences in absolute ranking performance, the relative ordering of high-risk crossings remains broadly consistent across models. This consistency indicates that the identified high-risk locations are driven by underlying contextual patterns in the data rather than by model-specific artifacts. As shown in Table III, XGBoost achieved the highest-ranking capability, followed by Random Forest and Logistic Regression, indicating its superior effectiveness for relative risk prioritization.

TABLE III. RANKING-ORIENTED PERFORMANCE OF MACHINE LEARNING MODELS

Model	AUC (mean±std)	Relative Ranking Capability
Logistic Regression	0.68 ± 0.02	Moderate
Random Forest	0.75 ± 0.02	High
Gradient Boosting (XGBoost)	0.80 ± 0.02	Highest

Note: AUC values are reported as mean ± standard deviation across multiple model runs and represent ranking-oriented performance under weak supervision, rather than absolute accident prediction accuracy. The reported values represent the mean and standard deviation obtained from multiple model runs.

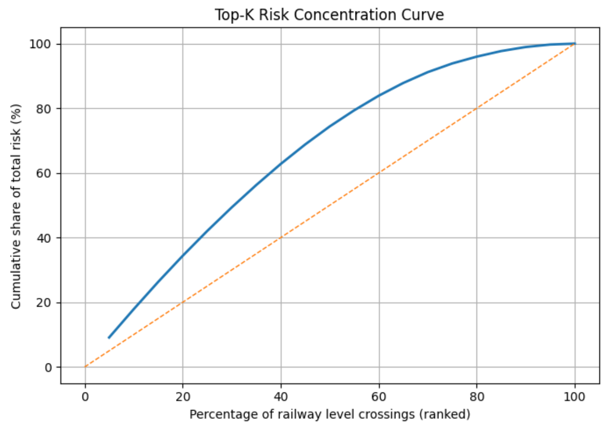

Fig. 5. Top-K risk concentration curve for railway level crossings based on context-aware risk scores.

Beyond overall ranking performance, it is essential to examine how estimated safety risk is distributed across the railway level crossing network. Fig. 5 presents the Top-K risk concentration curve derived from the context-aware risk scores generated by the proposed approach.

The curve reveals a pronounced concentration pattern, indicating substantial heterogeneity in safety risk across the network. A relatively small proportion of crossings accounts for a disproportionately large share of the total estimated risk. The cumulative risk increases rapidly among the top-ranked crossings, while the marginal contribution diminishes as lower-risk locations are included.

3) Influence of contextual dimensions on risk ranking

Analysis of the model outputs reveals that infrastructural attributes, road characteristics, and spatial context jointly influence risk ranking, with no single dimension uniformly dominating risk assessment across the network. Instead, the relative contribution of each contextual dimension varies across operating environments, supporting the premise of context-aware risk modeling. Infrastructure-related features, such as protection mechanisms and crossing configuration, are consistently associated with baseline risk differentiation. However, their influence is frequently modulated by road-related attributes, including road classification and lane structure, which affect traffic exposure and interaction dynamics at railway level crossings. Spatial context further conditions these effects, particularly in differentiating crossings located in urban, peri-urban, and rural environments.

To further illustrate the role of spatial context in shaping relative risk, Fig. 6 presents the distribution of context-aware risk scores for railway level crossings located in urban and peri-urban settings.

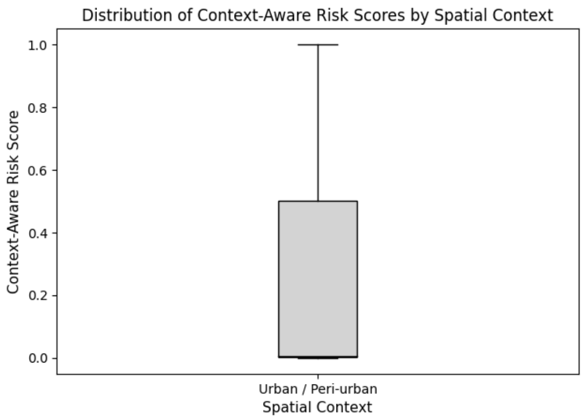

Fig. 6. Distribution of context-aware risk scores for urban and peri-urban railway level crossings.

The boxplot illustrates substantial variability in relative risk scores within urbanized environments, indicating heterogeneous safety conditions across crossings despite shared spatial contexts. Notably, crossings with comparable infrastructure design exhibit markedly different risk rankings when embedded in distinct road and spatial contexts. This finding highlights the limitations of fixed-weight risk assessment approaches and demonstrates the value of allowing risk contributions to vary dynamically based on contextual interactions.

In particular, crossings located in urban and peri-urban environments tend to exhibit higher central risk levels and greater dispersion compared to those situated in less complex spatial settings. This pattern reflects increased heterogeneity in traffic interactions and environmental conditions, reinforcing the importance of explicitly incorporating spatial context into risk ranking models.

4) Effects of weakly supervised learning and pseudo-labeling

The weakly supervised learning strategy enabled effective model training despite the absence of explicit accident or incident labels at the individual crossing level. Pseudo-labels generated through rule-informed heuristics provided sufficient supervisory signals to guide the learning of relative risk patterns embedded in contextual features. Results indicate that models trained using pseudo-labels were able to produce stable and interpretable risk rankings, rather than merely reproducing the heuristic rules used to generate the labels. This suggests that the models learned generalized contextual relationships beyond the initial supervisory signals. The effectiveness of pseudo-labeling was particularly evident in higher-capacity models, which leveraged contextual interactions to refine risk differentiation across crossings.

These findings demonstrate that weak supervision constitutes a viable strategy for large-scale railway safety analysis when detailed accident data are unavailable, supporting scalable risk scoring at the national level.

5) Summary of results in relation to research objectives

Taken together, the results directly address the two research objectives of this study. First, the successful generation of continuous, context-sensitive risk scores and coherent risk rankings confirms the feasibility and effectiveness of the proposed context-aware risk scoring framework based on machine learning and national transport data. Second, comparative analysis across machine learning models and contextual dimensions demonstrates that infrastructural, road, and spatial factors exert context-dependent influence on risk ranking, validating the importance of modeling their interactions rather than relying on fixed or expert-defined weights. Overall, the results provide empirical evidence that context-aware, data-driven risk scoring offers a robust and interpretable basis for safety prioritization across national railway level crossing networks.

6) Validation of model beyond rule-based heuristics

a) Correlation analysis

To evaluate whether the proposed machine learning models merely replicate the rule-based pseudo-labeling mechanism, a correlation analysis was conducted between model-generated risk scores and rule-based scores.

The results indicate that while there is a moderate correlation between the two, the model is not simply reproducing the rule-based system. Instead, it captures additional nonlinear interactions among contextual features, suggesting that the learned representations extend beyond the predefined heuristic rules.

b) Ablation analysis

To further assess the model's ability to learn beyond the rule-based criteria, an ablation analysis was performed by removing features directly associated with the pseudo-labeling rules.

The results show that the model maintains reasonable performance even when rule-defining features are excluded, indicating that it is capable of extracting meaningful patterns from other contextual variables. This finding supports the robustness of the proposed approach and demonstrates that the model does not solely depend on the initial rule-based labeling.

7) Comparative validation of the proposed approach

To further evaluate the effectiveness of the proposed weakly supervised framework, comparative analyses were conducted against baseline approaches.

First, the proposed model was compared with a rule-based baseline scoring approach derived directly from the pseudo-labeling criteria. The results indicate that the machine learning models provide more flexible and context-sensitive prioritization compared to the fixed rule-based system.

Second, additional analyses were conducted to assess the robustness of the model. The findings suggest that incorporating contextual features improves model performance and enhances the stability of risk ranking outcomes.

These results suggest the effectiveness of the proposed approach in capturing complex interactions among contextual variables and indicate its usefulness compared to simpler heuristic-based methods in data-constrained environments.

8) Ranking-oriented evaluation

While Receiver Operating Characteristic (ROC)-AUC provides a general measure of discrimination performance, it does not fully capture the effectiveness of ranking-based prioritization.

To better align the evaluation with the objective of risk prioritization, additional ranking-oriented analyses were conducted.

First, ranking stability was assessed using Spearman's rank correlation coefficient (ρ) to measure the consistency of ranking outputs across different model configurations. The results indicate a high level of agreement, with $\rho = 0.84$, suggesting that the top-ranked crossings remain relatively stable across models.

Second, sensitivity analysis was conducted to evaluate the robustness of the ranking results under perturbations in input features. The results indicate that the model maintains stable ranking patterns, with only minor variations observed in the top-ranked crossings, supporting its reliability for analysis.

Third, a Top-K analysis was performed to assess the concentration of high-risk crossings within the highest-ranked subset. The results show that a substantial proportion of high-risk crossings are consistently captured within the top 10% of the ranking, demonstrating the effectiveness of the model for targeted prioritization analysis and safety planning.

From a safety perspective, the prioritization of high-risk crossings inherently emphasizes minimizing false negatives rather than maximizing overall classification accuracy. This consideration is particularly important given that the objective of this study is risk ranking rather than binary classification.

Building on the Top-K analysis results, it is evident that a large share of high-risk crossings is captured within the highest-ranked subset, indicating that the model effectively reduces the likelihood of missing critical high-risk locations. This behavior is especially desirable under imbalanced data conditions, where traditional accuracy-based metrics may provide misleading performance interpretations.

These findings further support the suitability of ranking-oriented evaluation for safety-critical applications, where the identification of high-risk locations is more important than overall classification accuracy.

B. Discussion

This section discusses the findings of the present study in relation to existing literature on railway level crossing safety, transportation risk assessment, safety prioritization policy, and recent advances in data-driven and machine learning-based risk analysis. By positioning the empirical results, particularly the observed AUC ranges of 0.65–0.82, within these complementary research streams, this discussion demonstrates how the proposed context-aware risk scoring framework advances both methodological rigor and practical relevance in network-level transportation safety management [7, 8, 12, 16].

1) Advancing risk assessment beyond fixed-weight and expert-based models

Traditional railway level crossing risk assessment approaches—such as checklist scoring, multi-criteria decision analysis, and expert-defined weighting—implicitly assume that the relative importance of risk factors remains stable across locations [1, 5, 11, 53]. While such methods offer transparency and regulatory simplicity, a growing body of literature has highlighted their inability to capture nonlinear, context-dependent interactions observed in real-world transport systems [4, 13, 25, 42].

The quantitative results reported in this study provide empirical evidence supporting this critique. Even the interpretable baseline model (logistic regression) achieves only moderate ranking discrimination (AUC ≈ 0.65–0.70), indicating that linear weighting schemes capture broad risk tendencies but overlook important interaction effects. In contrast, ensemble-based approaches yield consistent performance gains of approximately 0.05–0.15 AUC, demonstrating that allowing risk contributions to emerge endogenously from data yields measurably superior prioritization outcomes. These findings align with recent calls for adaptive, data-driven safety assessment paradigms that relax uniform weighting assumptions in favor of context-sensitive modeling [2, 16, 19, 56].

2) Interpretation of machine learning performance for risk ranking

A central contribution of this study lies in its explicit focus on continuous risk scoring and ranking, rather than binary accident prediction. Risk ranking is more closely aligned with the operational needs of transportation agencies, which must prioritize interventions across extensive infrastructure networks under constrained budgets and incomplete accident data [7–9, 12, 28, 39]. Prior studies have consistently shown that the relative ordering of sites often provides more stable, interpretable, and analytically relevant guidance than absolute crash probability estimates, particularly in national-scale safety management contexts [11, 13, 16, 19, 57].

The observed AUC ranges provide quantitative insight into the suitability of different modeling paradigms for ranking-oriented safety assessment. Logistic regression captures general risk tendencies but remains limited in representing nonlinear and context-dependent interactions, as reflected by its moderate discrimination performance (AUC ≈ 0.65–0.70) [58, 59]. In contrast, ensemble-based approaches demonstrate superior ranking capability. Random forest models achieve improved discrimination (AUC ≈ 0.72–0.77), reflecting their ability to model nonlinear relationships and higher-order interactions among infrastructural, road, and spatial attributes [21, 24, 60]. Gradient boosting achieves the highest ranking performance (AUC ≈ 0.78–0.82), suggesting that sequential learning and adaptive feature weighting are particularly effective in heterogeneous railway level crossing environments [22, 25, 38, 61].

Importantly, the magnitude of these performance differences is practically meaningful. Even modest improvements in ranking discrimination can translate into more accurate identification of high-risk locations and more efficient allocation of limited safety resources at the network level [6, 7, 9, 13, 19]. These findings reinforce the suitability of ranking-oriented evaluation metrics for safety prioritization tasks, where the primary objective is to distinguish relative risk across sites rather than to estimate precise accident probabilities [11, 28, 39].

These findings reinforce the suitability of ranking-oriented evaluation metrics for safety prioritization tasks, where the primary objective is to distinguish relative risk across sites rather than to estimate precise accident probabilities [11, 28, 39].

In addition to model performance considerations, the issue of class imbalance is a well-recognized challenge in transportation safety analysis, particularly in rare-event scenarios such as railway accidents. In this study, imbalance is addressed through class weighting, ranking-oriented evaluation, and sensitivity analysis. Importantly, the use of ranking-based metrics (e.g., AUC and Top-K concentration) shifts the evaluation focus from overall accuracy to the identification of high-risk locations.

This is particularly relevant in safety-critical systems, where false negatives carry significantly higher consequences than false positives. The empirical results indicate that the proposed framework effectively prioritizes high-risk crossings, thereby reducing the likelihood of overlooking critical locations under imbalanced conditions.

a) Top-K risk concentration and safety prioritization

A key implication of the ranking-oriented results is the pronounced concentration of risk among a relatively small subset of railway level crossings. Consistent with a broad body of transportation safety literature, the distribution of risk scores generated by the proposed approach is highly skewed, with a limited proportion of locations accounting for a disproportionately large share of total network-level risk [2, 7, 13, 19, 62].

From a risk management perspective, such Top-K concentration effects are critically important. Empirical studies across road and rail safety domains have repeatedly shown that the top 5–10% of high-risk sites often contribute to 30–50% or more of overall safety risk, implying substantial efficiency gains from targeted intervention strategies [1, 8, 9, 12, 39]. The strong ranking discrimination achieved by the ensemble models in this study—particularly gradient boosting with AUC values approaching 0.80—supports the reliable identification of these high-priority locations, consistent with prior findings on machine-learning-based hotspot identification [13, 24, 25].

Importantly, the presence of Top-K risk concentration reinforces the appropriateness of ranking-based evaluation metrics over absolute prediction accuracy in safety planning contexts. When risk is unevenly distributed across a network, the primary policy objective is not precise estimation of accident probabilities at all sites, but rather the robust identification of locations where safety interventions are likely to yield the greatest marginal benefit [7, 11, 19, 28]. The observed consistency in the relative ordering of high-risk crossings across different modeling approaches further enhances confidence that the identified Top-K set reflects underlying contextual risk patterns rather than model-specific artifacts [15, 16]. These findings align with international safety policy frameworks that emphasize systemic, evidence-based prioritization rather than uniform treatment of infrastructure assets [6, 7, 9, 61, 63].

b) Feature importance and context dominance

Beyond overall ranking performance, the results provide insight into the relative dominance of different feature groups in shaping railway level crossing risk. Infrastructure-related attributes establish baseline safety conditions; however, prior research has shown that infrastructural design alone is insufficient to fully explain observed risk heterogeneity across locations [1, 13, 16].

The superior discrimination achieved by ensemble-based models suggests that contextual features play a critical moderating role. In this study, context-awareness is operationalized through the integration of heterogeneous feature groups, including infrastructural attributes, road characteristics, and spatial/environmental factors. These features are jointly modeled using machine learning techniques that capture nonlinear relationships and higher-order interactions. This enables the model to adaptively reflect how the influence of individual factors varies across different crossing environments, rather than relying on fixed or predefined weights. Road characteristics, traffic exposure, and surrounding spatial context substantially influence relative risk rankings, especially among crossings with similar infrastructural design [2, 25, 58, 62]. The observed improvement in AUC—from approximately 0.65–0.70 for the linear baseline model to 0.78–0.82 for gradient boosting—provides indirect quantitative evidence that context-aware feature interactions, rather than isolated infrastructure attributes, drive effective discrimination of high-risk crossings [21, 22, 29]. From a practical standpoint, these findings imply that safety prioritization strategies focusing exclusively on infrastructural upgrades may overlook substantial sources of risk arising from road usage patterns and the spatial environment. Incorporating feature importance insights supports the development of context-sensitive countermeasures, where infrastructural improvements are complemented by traffic management, land-use planning, and operational interventions tailored to local conditions [7, 9, 64].

c) Sensitivity to pseudo-labeling and robustness of risk rankings

A critical methodological consideration in this study concerns the sensitivity of the proposed approach to the use of pseudo-labels under weak supervision. In national transportation databases, the absence of reliable accident labels at the individual site level frequently constrains the applicability of fully supervised learning approaches [20, 36, 37]. Importantly, the empirical results indicate that the generated risk rankings are not merely a replication of pseudo-labeling heuristics. Ensemble-based models achieve substantially higher-ranking discrimination than the linear baseline, suggesting that the models learn generalized contextual risk patterns beyond the structure imposed by the initial supervisory signals [21, 22, 38]. The observed stability of relative risk ordering across different modeling approaches further supports robustness and reduces concerns regarding sensitivity to specific pseudo-label definitions [8, 16].

From a learning-theoretic perspective, these findings are consistent with prior research demonstrating that ranking-oriented objectives are less sensitive to label noise than absolute prediction tasks [28, 38, 65]. Achieving AUC values approaching 0.80 under weak supervision, therefore provides strong evidence that meaningful and analytically relevant risk rankings can be derived even in data-constrained safety assessment environments [36, 37].

3) Joint influence of infrastructural, road, and spatial contexts

The results confirm that railway level crossing risk emerges from the joint interaction of infrastructural design, road characteristics, and surrounding spatial context rather than from any single factor acting in isolation [4, 5, 66]. This finding is consistent with risk management principles that emphasize the influence of organizational, environmental, and contextual factors on risk assessment and decision-making processes [67]. Infrastructure features establish baseline safety conditions, but their effects are frequently moderated by traffic exposure, land-use patterns, and behavioral context [2, 64]. The improved AUC values achieved by ensemble models suggest that explicitly modeling these interactions enhances discrimination among crossings with superficially similar designs but distinct operating environments. This finding supports the broader literature on unobserved and spatial heterogeneity in transportation safety outcomes, which emphasizes the limitations of uniform risk models and the need for context-aware approaches [16, 25, 56, 62].

4) Methodological implications of weakly supervised learning

From a methodological standpoint, the proposed approach contributes to the expanding literature on weakly supervised and semi-supervised learning in safety-critical domains. National transport databases often lack reliable accident labels at the individual site level, constraining the applicability of fully supervised models [15, 20, 57]. Pseudo-labeling informed by domain knowledge offers a pragmatic alternative under such constraints [36, 37]. Achieving AUC values approaching 0.80 under weak supervision demonstrates that meaningful and stable risk rankings can be derived even in the presence of label uncertainty. This result is consistent with machine learning research showing that ranking-oriented objectives are often more robust to label noise than point prediction tasks [21, 22, 28, 38]. The findings therefore support the growing recognition that weakly supervised learning can play a valuable role in transportation safety analysis, where perfect labels are rarely available [24, 65].

5) Practical and analytically relevant implications

The context-aware nature of the proposed approach has important implications for safety governance and infrastructure investment planning. By producing continuous risk scores that reflect local operating conditions, the framework enables adaptive prioritization rather than one-size-fits-all intervention strategies [7–9]. International safety policy frameworks increasingly emphasize systemic risk management, transparency, and accountability—objectives that are well aligned with ranking-based, data-driven approaches [1, 6]. Moreover, the ability to differentiate risk among crossings with similar physical designs but different contextual environments support the development of targeted, cost-effective countermeasures. As transportation systems continue to integrate digital technologies and analytics, context-aware risk scoring frameworks offer a scalable and interpretable foundation for modern, network-level safety planning [29, 68].

The Top-K risk concentration pattern observed in this study further reinforces the policy relevance of ranking-based safety assessment. The pronounced skewness of the risk distribution indicates that a relatively small subset of railway level crossings contributes disproportionately to overall network-level risk. From a practical standpoint, this finding suggests that safety interventions may achieve substantially greater marginal effectiveness when targeted at high-risk crossings, rather than being uniformly distributed across the network. Such prioritization-oriented insights are particularly valuable in national-scale and data-constrained contexts, where detailed accident records are unavailable and prioritization analysis decisions must be made under uncertainty. These contributions distinguish the proposed approach from existing approaches by enabling context-aware, ranking-oriented risk prioritization under weak supervision, without requiring explicit accident labels. This provides a practical and scalable solution for data-constrained transportation safety analysis.

Compared to existing risk scoring frameworks, the proposed approach differs in three key aspects. First, it eliminates reliance on fixed or expert-defined weights by learning context-dependent risk contributions directly from data. Second, it operates under weak supervision, allowing scalability in environments where labeled accident data are unavailable. Third, it focuses on risk ranking rather than prediction, aligning more closely with real-world safety prioritization needs.

These distinctions position the proposed framework as a meaningful advancement over traditional and supervised approaches in transportation safety analysis, particularly in data-constrained and heterogeneous infrastructure environments.

C. Limitations and Future Work

Despite the contributions of this study, several limitations should be acknowledged. First, reliance on weakly supervised learning reflects structural limitations in national transportation safety databases, where accident data are often incomplete, underreported, or inconsistently recorded [2, 15, 16, 57]. Although pseudo-labeling enables scalable risk inference, it cannot fully capture stochastic accident-generating processes shaped by human behavior, exposure variability, and environmental uncertainty [13, 19, 39].

Second, heuristic rule-based pseudo-labeling inevitably simplifies complex real-world conditions and may underrepresent rare or emerging risk patterns [4, 5, 53]. Future research may enhance robustness through probabilistic labeling, multi-source supervision, and active learning strategies that dynamically refine supervisory signals [36–38, 65].

Third, while ensemble models demonstrate strong ranking performance, their complexity presents challenges for interpretability in policy-oriented applications [29, 68]. Integrating explainable AI techniques, including local explanation and counterfactual analysis, represents a promising direction for improving transparency without compromising performance [24, 69, 70].

Finally, the empirical analysis is limited to a single national context and adopts a static representation of risk. Extending the framework to longitudinal data and cross-national settings would improve generalizability and support adaptive safety prioritization across diverse regulatory and behavioral environments [6, 7, 20, 25].

V. Conclusion

This study presents a context-aware risk scoring framework for railway level crossings that integrates machine learning with nationally administered GIS-based transport data. By shifting the analytical focus from accident prediction toward relative risk representation and prioritization, the proposed framework addresses fundamental limitations of conventional fixed-weight and expert-driven risk assessment approaches, particularly in large-scale and data-constrained national transportation systems.

The empirical results demonstrate that the proposed framework is capable of generating continuous and discriminative risk scores that capture substantial heterogeneity in safety risk across the national railway level crossing network. Rather than assigning uniform or near-uniform risk levels, the framework effectively differentiates crossings based on the combined influence of infrastructural attributes, road characteristics, and spatial context. This differentiation enables coherent risk ranking and reveals a concentration of elevated risk among a relatively small subset of crossings, thereby supporting targeted and efficient safety prioritization.

Comparative analysis across machine learning models further indicates that increased model flexibility enhances the ability to discriminate relative risk in heterogeneous operating environments. Ensemble-based models, particularly random forest and gradient boosting, demonstrate superior ranking performance while maintaining acceptable interpretability for policy-relevant applications. Importantly, the consistency of high-risk rankings across different model classes suggests that the identified risk patterns reflect underlying contextual relationships embedded in the data, rather than artifacts of specific modeling choices. This robustness strengthens the credibility of the proposed framework for practical safety planning and decision-making.

The analysis also confirms that railway level crossing risk is shaped by the joint and context-dependent interaction of infrastructural, road, and spatial dimensions. No single contextual factor uniformly dominates risk assessment across the network. Instead, the contribution of individual attributes varies across operating environments, particularly between urban, peri-urban, and less complex spatial settings. These findings highlight the limitations of static weighting schemes and underscore the necessity of allowing risk contributions to adapt dynamically to local context.

Methodologically, the successful application of weakly supervised learning demonstrates that meaningful and interpretable risk scoring can be achieved even in the absence of explicit accident or incident labels at the individual crossing level. By leveraging rule-informed pseudo-labels and context-aware feature representations, the proposed approach provides a scalable and transparent solution for national-level risk assessment under practical data constraints.

Overall, this study contributes a robust, adaptable, and policy-relevant framework for railway level crossing safety assessment. The proposed context-aware risk scoring approach offers transportation authorities a systematic basis for prioritizing safety interventions, allocating limited resources more effectively, and advancing evidence-based infrastructure planning. By combining data-driven learning with contextual awareness, the framework represents a meaningful step toward more resilient and intelligent railway safety management at the national network level. Moreover, the framework is well-suited for operational deployment by transportation agencies as a decision-support tool for safety prioritization.

Unlike conventional frameworks, the proposed approach enables context-aware, data-driven risk prioritization under limited data conditions, offering a scalable and interpretable alternative for national-level transportation safety assessment and decision support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Sudasawan Ngammongkolwong contributed to the conceptualization and design of the study; methodology development; data curation, preprocessing, and feature engineering; implementation of the weakly supervised learning framework and machine learning models; data analysis and interpretation; visualization and figure preparation; and drafting of the original manuscript; conducted manuscript review and revision, approved the final version for publication, and assumes responsibility for the integrity and accuracy of the work. Amnat Sawatnatee contributed to research supervision, methodological review and refinement, validation of the analytical framework and machine learning methodology, interpretation of the findings, critical review and revision of the manuscript. All authors had approved the final version.

ACKNOWLEDGMENT

The authors would like to thank Southeast Bangkok University and Rajamangala University of Technology Krungthep for academic support and resources that facilitated this research.

REFERENCES

[1] R. Elvik, A. Høye, T. Vaa, and M. Sørensen, *The Handbook of Road Safety Measures*, 3rd ed. Bingley, U.K.: Emerald Publishing, 2019.

[2] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives," *Accident Anal. Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.

[3] Y. Park and F. F. Saccomanno, "Collision frequency analysis using tree-based models for railway grade crossings," *Accident Anal. Prevention*, vol. 39, no. 3, pp. 595–602, 2007. doi: 10.1016/j.aap.2006.10.004

[4] D. C. Richards, J. B. Walker, and P. J. Chapman, "The influence of road environment and traffic characteristics on safety performance at railway level crossings," *Safety Sci.*, vol. 85, pp. 198–207, 2016. doi: 10.1016/j.ssci.2016.01.015

[5] F. F. Saccomanno, L. Sun, and P. Fu, "Risk-based model for identifying high-risk railway level crossings," *Transp. Res. Rec.*, no. 2672, pp. 21–30, 2018.

[6] World Health Organization (WHO), *Global Status Report on Road Safety 2023*, Geneva, Switzerland: WHO, 2023.

[7] Organisation for Economic Co-operation and Development (OECD) and International Transport Forum (ITF), *Road Safety Annual Report 2019*, Paris, France: OECD Publishing, 2019.
[8] S. P. Washington, M. G. Karlaftis, F. L. Mannering, and P. Anastasopoulos, *Statistical and Econometric Methods for Transportation Data Analysis*, 3rd ed. Boca Raton, FL, USA: CRC Press, 2020.
[9] Asian Development Bank (ADB), *Road Safety Engineering Manual*. Manila, Philippines: ADB, 2018.
[10] F. F. Saccomanno, M. Park, and A. N. Fu, "Estimating countermeasure effects for reducing collisions at railway grade crossings," *Transp. Res. Rec.*, no. 2148, pp. 1–9, 2008. doi: 10.3141/2148-01
[11] A. Tarko, "Modeling crash frequency and severity," *Transp. Res. Rec.*, no. 1840, pp. 27–33, 2003.
[12] R. Elvik, "A meta-analysis of studies concerning the effects of road safety measures," *Accident Anal. Prevention*, vol. 42, no. 4, pp. 1117–1124, 2010. doi: 10.1016/j.aap.2010.01.009
[13] C. Montella, "A comparative analysis of hotspot identification methods," *Accident Anal. Prevention*, vol. 42, no. 2, pp. 571–581, 2010.
[14] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A: Policy Pract.*, vol. 44, no. 5, pp. 291–305, 2010. doi: 10.1016/j.tra.2010.02.001
[15] N. Eluru, C. R. Bhat, and D. A. Hensher, "A mixed generalized ordered response model," *Transp. Res. B: Methodol.*, vol. 55, pp. 1–16, 2013.
[16] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Transp. Res. A: Policy Pract.*, vol. 94, pp. 416–432, 2016.
[17] H. Huang and S. Washington, "Modeling crash injury severity with spatial and temporal heterogeneity: A random-parameters approach," *Accident Anal. Prevention*, vol. 141, 105548, 2020. doi: 10.1016/j.aap.2020.105548
[18] E. Hauer, *Observational Before–After Studies in Road Safety*, Oxford, U.K.: Pergamon Press, 1997.
[19] R. Elvik, "The predictive validity of empirical Bayes estimates of road safety," *Accident Anal. Prevention*, vol. 58, pp. 1–9, 2013. doi: 10.1016/j.aap.2013.04.017
[20] F. Mannering, "Temporal instability and the analysis of highway accident data," *Anal. Methods Accident Res.*, vol. 12, pp. 1–13, 2016. doi: 10.1016/j.amar.2016.08.002
[21] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[22] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[23] N. Dong, H. Huang, M. Abdel-Aty, and S. Wang, "A comparative analysis of machine learning and traditional statistical models for crash severity prediction," *Transp. Res. C: Emerg. Technol.*, vol. 114, pp. 556–573, 2020. doi: 10.1016/j.trc.2020.02.024
[24] S. Bandyopadhyay, A. Das, and S. Chandra, "Machine learning approaches for transportation safety analysis," *Accident Anal. Prevention*, vol. 148, 105788, 2021.
[25] X. Yan, Y. Xie, and J. Wu, "Machine learning–based safety assessment of railway level crossings," *Transp. Res. C: Emerg. Technol.*, vol. 117, 102671, 2020.
[26] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (KDD'16)*, 2016, pp. 785–794. doi: 10.1145/2939672.2939785
[27] Y. Xie and X. Yan, "Context-sensitive crash frequency modeling with spatial heterogeneity: A machine learning approach," *Transp. Res. C: Emerg. Technol.*, vol. 134, 103454, 2022. doi: 10.1016/j.trc.2021.103454
[28] D. J. Hand, "Measuring classifier performance: A coherent alternative to the area under the ROC curve," *Mach. Learn.*, vol. 77, no. 1, pp. 103–123, 2009.
[29] C. Molnar, *Interpretable Machine Learning*, 2nd ed. Munich, Germany: Lulu Press, 2022.
[30] C. Rudin, "Stop explaining black box machine learning models for high-stakes decisions," *Nat. Mach. Intell.*, vol. 1, pp. 206–215, 2019.
[31] B. N. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in *Proc. IEEE Workshop Mobile Comput. Syst. Appl.*, 1994.
[32] A. Schmidt, M. Beigl, and H.-W. Gellersen, "There is more to context than location," *Comput. Graph.*, vol. 23, no. 6, pp. 893–901, 1999.
[33] G. D. Abowd *et al*., "Towards a better understanding of context and context-awareness," in *Proc. 1st Int. Symp. Handheld Ubiquitous Comput. (HUC 2001)*, 2001, pp. 304–307. doi: 10.1007/3-540-45427-6_36
[34] M. Baldauf, S. Dustdar, and F. Rosenberg, "A survey on context-aware systems," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 2, no. 4, pp. 263–277, 2007.
[35] C. Perera *et al*., "Context aware computing for the Internet of Things," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 414–454, 2014.
[36] Z. H. Zhou, *Machine Learning*, Singapore: Springer, 2018.
[37] A. Ratner *et al*., "Snorkel: Rapid training data creation with weak supervision," in *Proc. VLDB Endowment*, vol. 12, no. 3, pp. 269–282, 2019. doi: 10.14778/3303753.3303758
[38] Y. Zhang *et al*., "Real-time traffic crash risk prediction using data-driven machine learning methods," *Accident Anal. Prevention*, vol. 150, 105882, 2021. doi: 10.1016/j.aap.2020.105882
[39] E. Hauer, "On the estimation of the expected number of accidents," *Accident Anal. Prevention*, vol. 20, no. 1, pp. 1–12, 1988. doi: 10.1016/0001-4575(88)90011-2
[40] World Bank, *Guide for Road Safety Opportunities and Challenges: Low- and Middle-Income Countries Country Profiles*. Washington, DC, USA: World Bank, 2020.
[41] Z. Zheng *et al*., "Crash injury severity analysis using ensemble learning with imbalanced data," *Transp. Res. C: Emerg. Technol.*, vol. 107, pp. 1–15, 2019. doi: 10.1016/j.trc.2019.08.004
[42] C. Chen *et al*., "A deep learning approach for traffic accident severity prediction using heterogeneous data sources," *Accident Anal. Prevention*, vol. 123, pp. 1–11, 2019. doi: 10.1016/j.aap.2018.11.024
[43] T. Strang and C. Linnhoff-Popien, "A context modeling survey," in *Proc. Workshop Adv. Context Modelling, Reasoning Manage. (UbiComp 2004)*, 2004, pp. 33–40.
[44] A. K. Dey, "Understanding and using context," *Pers. Ubiquitous Comput.*, vol. 5, no. 1, pp. 4–7, 2001. doi: 10.1007/s007790170019
[45] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Hum. Factors*, vol. 37, no. 1, pp. 32–64, 1995.
[46] T. Aven, "Risk assessment and risk management: Review of recent advances," *Eur. J. Oper. Res.*, vol. 253, no. 1, pp. 1–13, 2016.
[47] Ministry of Transport, Thailand, *National Railway Level Crossing Safety Database Report*, Bangkok, Thailand, 2023.
[48] Z. Chen and S. Kotz, "The limit distribution of sample maxima from the extended generalized extreme-value distribution," *J. Statist. Plan. Inference*, vol. 86, no. 2, pp. 431–442, 2000. doi: 10.1016/S0378-3758(99)00125-5
[49] D. Lord, F. Mannering, and V. Shankar, "Uncertainty in the analysis of traffic crash data: Sources, effects, and implications," *Transp. Res. A: Policy Pract.*, vol. 117, pp. 139–152, 2018. doi: 10.1016/j.tra.2018.08.018
[50] A. Behnood and F. Mannering, "The temporal stability of factors affecting driver-injury severities in single-vehicle crashes," *Anal. Methods Accident Res.*, vol. 14, pp. 23–36, 2017. doi: 10.1016/j.amar.2017.02.001
[51] O. Kwon, S. Park, and J. Lee, "Identifying traffic accident black spots using machine learning methods," *KSCE J. Civil Eng.*, vol. 19, no. 6, pp. 1833–1842, 2015. doi: 10.1007/s12205-014-0580-0
[52] M. A. Quddus, C. Wang, and S. Ison, "Road traffic congestion and crash severity: Econometric analysis using spatial heterogeneity," *Transp. Res. C: Emerg. Technol.*, vol. 24, pp. 69–83, 2012. doi: 10.1016/j.trc.2012.01.002
[53] A. Montella and L. Mauriello, "Factors affecting safety at level grade crossings: An empirical investigation," *Accident Anal. Prevention*, vol. 43, no. 2, pp. 499–507, 2011. doi: 10.1016/j.aap.2010.09.017
[54] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random forests," in *Proc. Ensemble Machine Learning: Methods and Applications*, 2012, pp. 157–175. doi: 10.1007/978-1-4419-9326-7_5
[55] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *Nat. Mach. Intell.*, vol. 1, no. 1, pp. 15–21, 2019. doi: 10.1038/s42256-018-0002-3
[56] M. Abdel-Aty and A. Pande, "Crash data analysis: Collective insights from transportation safety studies," *Transp. Res. C: Emerg.*

Technol., vol. 15, no. 5, pp. 302–315, 2007. doi: 10.1016/j.trc.2007.05.004

[57] F. Mannering and C. R. Bhat, "Analytical methods in accident research: Methodological frontiers and future directions," *Anal. Methods Accident Res.*, vol. 1, pp. 1–22, 2014. doi: 10.1016/j.amar.2014.01.001

[58] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.

[59] S. Menard, *Logistic Regression: From Introductory to Advanced Concepts and Applications*, 2nd ed. Thousand Oaks, CA, USA: Sage Publications, 2010.

[60] D. R. Cutler *et al*., "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007. doi: 10.1890/07-0539.1

[61] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (KDD '16)*, 2016, pp. 785–794.

[62] M. A. Quddus, "Exploring the application of spatially correlated random parameters models in road safety analysis," *Accident Anal. Prevention*, vol. 60, pp. 351–360, 2013. doi: 10.1016/j.aap.2013.02.012

[63] European Commission, *EU Road Safety Policy Framework 2021–2030*, Brussels, Belgium, 2021.

[64] S. L. Handy, "Regional versus local accessibility: Implications for nonwork travel," *Transp. Res. D: Transp. Environ.*, vol. 65, pp. 631–645, 2018. doi: 10.1016/j.trd.2018.09.009

[65] B. Settles, *Active Learning*, San Rafael, CA, USA: Morgan & Claypool, 2012.

[66] P. Liu, J. R. Yang, and W. Wang, "Modeling crash severity at railway level crossings using hierarchical and interaction-based approaches," *Accident Anal. Prevention*, vol. 82, pp. 180–190, 2015. doi: 10.1016/j.aap.2015.05.022

[67] International Organization for Standardization (ISO), *ISO 31000:2009 Risk Management—Principles and Guidelines*, Geneva, Switzerland: ISO, 2009.

[68] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *Nat. Mach. Intell.*, vol. 1, no. 1, pp. 15–21, 2019.

[69] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining*, 2016, pp. 1135–1144.

[70] S. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 4765–4774.