

Received 1 July 2025, accepted 21 July 2025, date of publication 24 July 2025, date of current version 6 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3592445

RESEARCH ARTICLE

Diagnostic Analysis of Crosswalk Safety Hazards in Pedestrian Environments: A SHAP-Enhanced Machine Learning Approach With Street-View Imagery

CARYL ANNE M. BARQUILLA^{ID} AND JEONGWOO LEE^{ID}

Department of Urban Design and Studies, Chung-Ang University, Seoul 06974, South Korea

Corresponding author: Jeongwoo Lee (jeongwoo@cau.ac.kr)

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (RS-2025-00523493). This research was also supported by the Chung-Ang University research grant in 2025.

ABSTRACT This study investigated the influence of built environment features on crosswalk safety in dense urban settings, with a focus on visual streetscape characteristics extracted from street-view imagery using both semantic and instance segmentation. It used data from 36,750 crosswalks in Seoul, South Korea, to rigorously evaluate multiple machine learning algorithms for predicting pedestrian crash risk. Among the models assessed, the Random forest (RF) demonstrated the highest precision, aligning with the objective of enhancing pedestrian safety through accurate risk identification. The RF model enhanced by SHapley Additive exPlanations (SHAP) achieved strong predictive performance (precision: 0.91), and SHAP analysis identified visual features, particularly sky openness ratio, building coverage, and sidewalk ratio, as influential factors affecting crash risk. A lower sky openness ratio combined with a higher building ratio was associated with increased crash likelihood, whereas greater sidewalk coverage and the presence of traffic control measures, including traffic lights and crosswalk time indicators, mitigated risk. Interaction effects further highlighted the complexity of urban safety, showing that combinations of streetscape and infrastructural elements can amplify or reduce hazards. These results highlight the importance of combining visual and structural data for thorough risk assessment and further the use of interpretable machine learning in urban safety research. The findings imply that to address particular combinations of built environment elements that increase the risk of crosswalk crashes, policy and planning initiatives should concentrate on context-sensitive interventions, particularly by placing bus stops strategically, maintaining tree canopies for visibility, clearing visual clutter, and improving pedestrian infrastructure.

INDEX TERMS Built environment, crosswalk safety, interpretable machine learning, pedestrian environment, street view imagery.

I. INTRODUCTION

Traffic collisions remain a pressing global public health concern, exacerbated by rapid urbanization and increasing population density in metropolitan areas. As more than half of the world's population now resides in urban settings, cities have become epicenters of traffic-related injuries and

fatalities [1]. In South Korea, accelerated industrialization and urban expansion have intensified this problem, with 76,548 pedestrian crashes recorded in 2022 alone [2].

In particular, a spatial concentration of such incidents was observed in Seoul: approximately 23% of these crashes occurred at crosswalks over the recent five-year period (Fig. 1) [2]. This paradox, where infrastructure intended to improve safety becomes a site of heightened risk, reflects a deeper challenge: pedestrian danger is shaped not only

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague^{ID}.

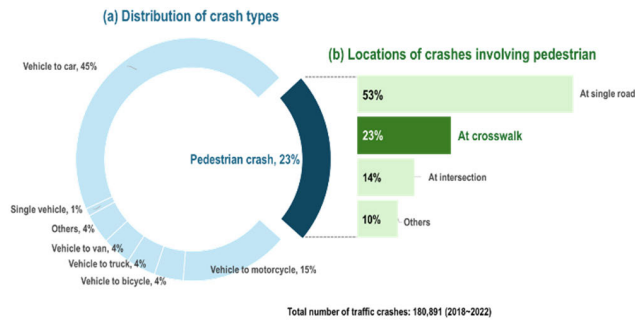


FIGURE 1. Traffic crashes in Seoul, South Korea, over a five-year period (2018–2022): (a) distribution of crash types; (b) locations of crashes involving pedestrians.

by the presence of infrastructure, but also by the quality and configuration of the surrounding environment. Factors such as poor visibility, discontinuous sidewalks, street clutter, and obstructive vegetation may turn even well-intentioned designs into hazardous zones. However, these fine-grained, street-level features are often missing from conventional datasets.

While prior studies have examined associations between pedestrian crashes and built environment factors such as land use, road network density, and traffic controls [3], [4], [5], [6], many rely on aggregated or remotely sensed data that fail to capture the nuanced, street-level conditions experienced by pedestrians [7]. As a result, there remains a significant research gap in capturing and modeling these micro-scale environmental factors, which play a crucial role in pedestrian safety, particularly at crosswalks. To address this gap, this study incorporates detailed SVI as a novel data source to represent pedestrian-level environmental features, such as building frontage, street furniture, crosswalk visibility, and vegetation coverage.

Furthermore, this study explores not only the individual effects of these visual features but also how interactions between visual and structural elements contribute to pedestrian crash risk at crosswalks. This study combines these visual cues with traditional built environment factors using machine learning (ML) models, which are enhanced by SHapley Additive exPlanations (SHAP), to offer a more comprehensive understanding of pedestrian safety. Use of SHAP enables identification of the features' contributions and the interactions among them, thus providing actionable insights for urban planning and policy interventions aimed at improving crosswalk safety.

In doing so, this study makes the following contributions:

- Integrates street-view imagery with conventional built environment data to capture human-scale features influencing pedestrian crash risk at crosswalks.
- Applies interpretable machine learning methods, using SHAP to quantify both individual and interaction effects of visual and structural variables.
- Conducts a spatially explicit analysis in Seoul, contributing new empirical insights from a dense, transit-oriented

urban context that remains underrepresented in current literature.

- Provides actionable recommendations for urban planners and policymakers, supporting data-driven design and safety interventions for crosswalk environments.

The remainder of this paper is organized as follows. Section II reviews the literature on pedestrian safety and built environment features and examines the methodologies used in prior studies. Section III comprehensively describes the study area and the methodology employed to analyze the crosswalk data, including the use of ML, deep learning (DL), and SHAP-based explainable ML. Section IV describes the results of predictive modeling and spatial analyses, and Section V discusses the implications of these findings and recommendations for improving crosswalk pedestrian safety. Finally, Section VI concludes the study and summarizes key insights and suggestions for future research.

II. LITERATURE REVIEW

A. BUILT ENVIRONMENT FEATURES AS DETERMINANTS OF PEDESTRIAN SAFETY

Urban pedestrian safety is shaped by a wide array of interrelated factors, among which the built environment is particularly influential. The spatial configuration of streets, land uses, and surrounding infrastructure not only guides pedestrian movement but also produces conditions that can either amplify or mitigate crash risk. In academic literature, built environment features are often grouped into two interrelated dimensions: (1) the *physical built environment*, referring to the tangible, spatial arrangement of urban form, and (2) the *functional built environment*, which captures how the environment operates in relation to pedestrian behavior, exposure, and connectivity [5], [8], [9], [10].

The physical form of the built environment, including street width, road geometry, presence of sidewalks, and intersection design, plays a role in determining pedestrian safety. These features influence pedestrian movement and exposure to risk in several ways. For example, wide roads or multiple lanes may increase vehicle speed and reduce visibility, heightening the likelihood of crashes [11], [12], [13]. Similarly, poorly designed intersections or areas without clear pedestrian crossings can create conflict points, where pedestrians and vehicles interact in hazardous conditions [9], [14]. The presence or absence of essential pedestrian infrastructure, namely sidewalks, crosswalks, traffic signals, and signage, further shapes pedestrian safety. Studies have shown that well-designed pedestrian infrastructure, including clear crosswalks and signals, reduces pedestrian crash risk by improving visibility and guiding pedestrian movement [15], [16].

Beyond the physical form of built environments, their functional dimensions, including density, connectivity, and land use patterns, play a critical role in shaping pedestrian exposure and crash risk. High population and employment densities, intricate street networks, and mixed-use zones often

concentrate pedestrian activity, thereby increasing the frequency of pedestrian-vehicle interactions [17], [18], [19]. Although areas with high population and employment densities or dense intersections may offer improved accessibility, they also generate more conflict points between pedestrians and vehicles [20]. Similarly, high street connectivity can enhance accessibility but also create more opportunities for pedestrian-vehicle interactions, raising the risk of crashes [20]. Mixed-use areas, often characterized by a combination of residential, commercial, and institutional buildings, are associated with higher pedestrian activity and, consequently, increased crash risk [21], [22]. These environments may improve accessibility but high pedestrian exposure is often not matched with adequate infrastructure, such as proper lighting, sidewalk width, and safe crossings [23], [24].

In summary, various built environment features have been consistently identified as key correlates of pedestrian crashes. While physical and functional built environment indicators provide a foundational understanding of pedestrian safety risks, integrating these with visual data derived from street-level imagery allows for a more nuanced and localized assessment of the streetscape environment.

B. SVI AS A DIAGNOSTIC TOOL FOR PEDESTRIAN SAFETY ANALYSIS

In recent years, SVI has emerged as a powerful diagnostic tool for analyzing visual aspects of the built environment that influence pedestrian safety. Unlike traditional methods, which often rely on subjective or survey-based assessments, SVI enables the objective extraction of visually observable streetscape elements, particularly sidewalks, street lighting, signage, pedestrian fences, trees, building coverage, sky openness, and street transparency. These features, directly observable from the street level, can influence pedestrian safety but are often underrepresented in conventional spatial datasets. SVI offers a systematic and scalable way to assess these features across large urban areas, making it a valuable tool in diverse settings [25], [26], [27].

Research has demonstrated a strong relationship between visual elements captured through SVI and pedestrian safety outcomes. For instance, sidewalks provide designated spaces for pedestrians, helping to reduce the risk of pedestrian-vehicle collisions [15], [16], [28]. Similarly, visual presence of streetlights and signages improves visibility and situational awareness for both drivers and pedestrians. Elements such as pedestrian fences and traffic signals provide a clear separation between pedestrian and vehicle spaces, enhancing safety at crosswalks and reducing the potential for accidents [29], [30].

While many built environment features directly enhance pedestrian safety, some—such as tree canopies—have more complex effects. Empirical studies indicate that the visual presence of tree canopies is associated with mixed outcomes regarding their influence on safety. Although tree canopies can create a sense of security for road users and alleviate

eye strain, thus enhancing overall traffic safety [30], [31], [32], poorly positioned canopies may obstruct sightlines and increase the likelihood of crashes [33], [34]. Additionally, SVI can capture features of vertical infrastructure, namely building height and size, that can obstruct sightlines, reduce visibility, and consequently lead to higher risks for pedestrians [4], [35], [36]. These findings emphasize the need for a careful urban design that considers visibility and safety.

An emerging metric derived from SVI is sky openness, which refers to the proportion of sky visible from the street level that can serve as a proxy for the perceived enclosure or openness of urban spaces. Lower levels of sky openness may restrict visibility and contribute to a sense of crowding or visual blockage, potentially affecting both pedestrian awareness and driver reaction times [26], [30].

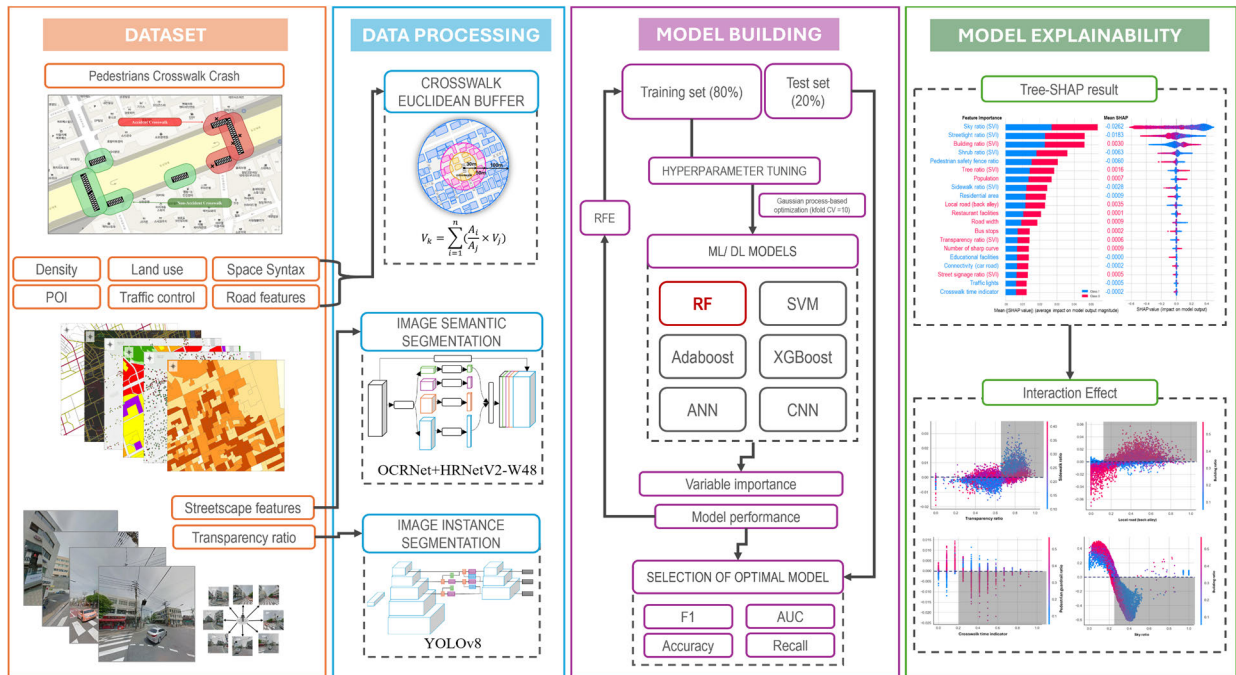
Street transparency, defined as the degree of visibility into building interiors from the street, such as through ground-level windows, is an underexamined visual element. Transparent facades enhance natural surveillance and foster the perception of safety by improving visibility for both pedestrians and drivers [37]. However, excessive transparency may contribute to “visual clutter,” distracting drivers and potentially causing cognitive overload, which may impair their ability to focus on critical road elements [38], [39]. This highlights the importance of balancing transparency in urban design to optimize both pedestrian safety and driver attention.

Collectively, existing studies showed that streetscape elements captured via SVI can influence pedestrian safety through factors such as visibility, distraction, and spatial enclosure. However, most of the studies isolated individual features and overlooked how these elements interact in complex urban environments. Few studies applied composite analyses that reflected the full range of streetscape conditions observable through SVI. This study addresses this gap by systematically quantifying multiple visual features and examining their combined influence on pedestrian safety outcomes. This approach enhances diagnostic safety modeling and supports evidence-based approaches to urban design.

C. METHODOLOGICAL APPROACHES IN PEDESTRIAN SAFETY RESEARCH

Traditional approaches to traffic safety research, such as linear regression, Poisson models, and negative binomial regressions have been employed widely to examine the fundamental factors in crash occurrence [11], [40], [41], [42], [43], [44], [45], [46], [47], [48]. More advanced statistical methods, namely generalized additive models and Tobit or truncated regression models, have been employed to address non-normal distributions and handle censored or threshold-based data, such as those observed in high-severity accident analyses [49], [50], [51], [52], [53], [54], [55], [56], [57]. While these approaches have proven effective, they still face challenges in fully addressing the complex interdependencies inherent in traffic crash data.

Recent advancements in ML and DL techniques have gained recognition for their ability to manage large vol-



umes of heterogeneous data and uncover complex, nonlinear patterns. These methods excel at extracting intricate relationships from complex datasets to predict traffic crashes. Models including random forest (RF), support vector machines (SVMs), decision trees, gradient boosting, and artificial neural networks (ANNs) have consistently outperformed traditional statistical methods, particularly in accounting for the spatial and temporal complexities of traffic crashes [40], [58], [59], [60], [61], [62], [63]. For instance, Chang [40] highlighted the superior predictive performance of ANNs compared to the negative binomial regression model, further underscoring the transformative potential of ML and DL techniques. However, despite their predictive power, ML models often lack interpretability, which limits their utility in understanding the factors behind the predictions. This “black box” nature can be a barrier when translating model outputs into actionable policy recommendations.

To address this challenge, the integration of SHAP values has emerged as a promising solution to enhance model transparency. SHAP values, grounded in cooperative game theory, quantify the contribution of each feature to individual predictions, thereby helping to identify the most influential variables affecting crash risk. Recent studies have demonstrated SHAP’s effectiveness in diverse traffic safety contexts, such as predicting crash severity, identifying high-risk locations, and understanding the impact of environmental and infrastructural factors on road safety outcomes [64], [65], [66], [67]. For instance, Ma et al. [67] used a deep learning framework combined with SHAP to interpret the impact of detailed driving behaviors, such as sharp acceleration and braking, on crash risk, providing clear, interpretable

insights into the main contributors to traffic collisions. These methodological advances reinforce the rationale for employing SHAP in our crosswalk diagnostics, as they allow for transparent, actionable analysis of both individual and interacting environmental factors.

By quantifying the influence of each feature, SHAP values help identify critical variables driving crash risk, enabling more informed and actionable decision-making to develop policies for safe street environments [21], [68], [69].

D. RESEARCH GAPS AND NOVEL CONTRIBUTIONS

Pedestrian safety research has made substantial advances, yet critical gaps remain, particularly at the crosswalk level where pedestrian–vehicle interactions are most concentrated. Existing studies tended to focus on broader spatial units (e.g., intersections and neighborhoods) and relied on aerial or structured data, which, while valuable, often overlook visual attributes that influence crosswalk conditions and potential pedestrian risk. In addition, while many studies examined the effects of individual built environment features, few investigated how these features interact or influence pedestrian safety in nonlinear ways. Grasping these more complex relationships is essential for understanding how the built environment shapes pedestrian risk in real-world contexts.

This study addresses these gaps through three key contributions. First, it focuses on crosswalk-scale analysis, offering a micro-level perspective on pedestrian–vehicle conflict points that are often generalized in broader-scale studies. Second, it leverages SVI using segmentation techniques to capture detailed, pedestrian-level visual features that are often overlooked in conventional analyses. This includes visual open-

ness, building setbacks, and visibility barriers, all of which are critical for assessing crosswalk safety. Third, although ML models have been applied in many pedestrian safety studies, this study applies interpretable ML techniques to a novel combination of SVI and built environment variables, enabling the identification of nonlinear effects and interaction patterns that have not been explored in previous studies.

These contributions advance pedestrian safety research by offering a more nuanced, data-driven approach that incorporates both the physical and visual aspects of the built environment, providing valuable insights into urban planning and crosswalk safety improvements.

III. MATERIALS AND METHODS

The framework for this study (Fig. 2) consisted of three main components. First, the traffic crash data and built environment features of Seoul were analyzed. The occurrence of pedestrian crashes at crosswalks was treated as a binary variable (0 or 1) and used as the dependent variable in the model, with built environment features obtained from the geographic information system (GIS) and SVI serving as independent variables. This binary specification was chosen because it outperformed count-based models in predictive accuracy, precision, and better supports the practical goal of distinguishing hazardous from non-hazardous crosswalks for policy interventions. These datasets were used to train both the ML and DL models, with the optimal model selected based on its highest predictive accuracy. The chosen model was then utilized to identify the most prominent features associated with crosswalk pedestrian crashes, with SHAP values employed to explore nonlinear relationships and interaction effects among the variables. This approach allowed for a deeper understanding of how these factors impact crosswalk pedestrian safety.

A. STUDY AREA

Seoul, the capital and largest city of South Korea, spans approximately 605.21 km² and is home to 10 million residents. The city faces ongoing challenges in traffic safety due to its dense population and extensive transportation infrastructure [70]. In Seoul, nearly one-fourth of pedestrian crashes are concentrated at crosswalks where pedestrian-vehicle interactions are frequent and complex [2]. This study examined locations of crashes involving pedestrians over a five-year period, utilizing data from 36,750 identified crosswalks across the city. These crosswalks exhibit diverse characteristics, including variations in design and surrounding features that influence pedestrian safety. To analyze pedestrian safety, crosswalks were categorized based on their crash history, and various built environment features were examined across different buffer zones surrounding these crosswalks. Both visual and structural features were considered to capture the diversity of streetscapes across the study area. This variation prompted an examination of whether crosswalks' visual and structural characteristics are associated with the likelihood of pedestrian crashes.

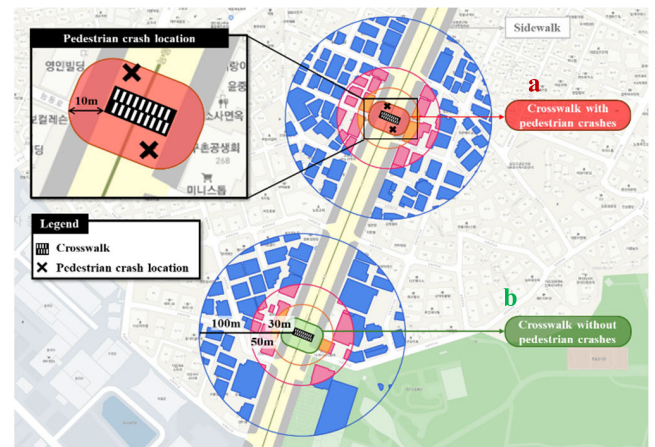


FIGURE 3. (a) Crosswalk with pedestrian crashes (red) and (b) crosswalk without pedestrian crashes (green) with areal interpolation applied on multiple Euclidean buffers to measure built environment features around crosswalks.

B. DATASET

1) TRAFFIC DATA

a: PEDESTRIAN CRASH DATA

This study used the Traffic Accident Analysis System (TAAS) to analyze pedestrian crashes at crosswalks and geocoded 17,418 crash locations collected from January 2018 to December 2022. Administered by the Korea Road Traffic Authority, TAAS is South Korea's official government database for traffic crashes, compiling detailed annual records from the Korean National Police Agency and insurance companies [71], [72], [73]. Furthermore, according to the Road Traffic Act, motorists are mandated to stop or refrain from parking within 10 m of crosswalks [74]. To classify crosswalks based on crash involvement, a 10-m buffer was created around the crosswalk polygon (Fig. 3a) to capture pedestrian crash occurrences just outside crosswalk boundaries. Crosswalks with one or more pedestrian crashes within the buffer were defined as crosswalks with pedestrian crashes, whereas those without them were classified as crosswalks without pedestrian crashes (Fig. 3b). This method identified 13,173 crosswalks with and 23,577 crosswalks without pedestrian crashes.

b: PEDESTRIAN VOLUME DATA

Pedestrian volume was used as the primary exposure variable for pedestrian crashes. This variable was derived from Nicebizmap [75], a business district analysis platform that provides street-level pedestrian traffic data across the Seoul metropolitan area. The measure served as a proxy for pedestrian volume, using the average pedestrian flow across each street segment. Pedestrian volume was collected for each segment, reflecting the intensity of pedestrian activity along specific streets in the vicinity of crosswalks.

To refine the analysis, an intersection method was applied in ArcGIS to combine pedestrian volume data with the buffer zones of each crosswalk environment. The mean value of pedestrian flow across the intersected street segments

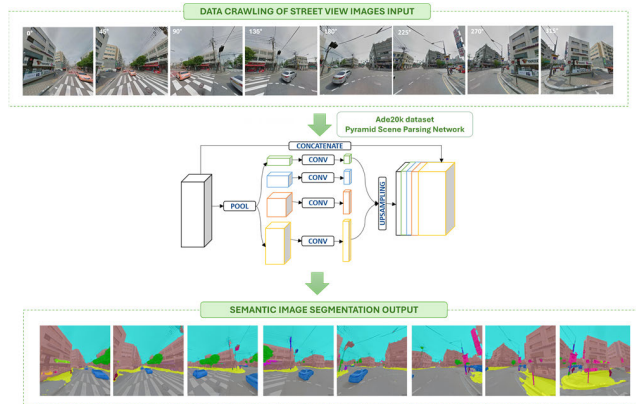


FIGURE 4. Semantic image segmentation using deep learning (OCRNet+HRNetV2-W48).

was then calculated to represent pedestrian volume at each crosswalk. This approach helped to assess the influence of pedestrian activity on crash likelihood.

2) VISUAL STREETSCAPE FEATURES

SVI was collected from the Kakao Map Road View [76] a leading mapping platform in South Korea that offers high-resolution, frequently updated panoramic images of urban environments. Kakao Map Road View was selected over more globally recognized alternatives such as Google Street View due to its superior spatial coverage and image currency within the Seoul metropolitan area [77], as well as its enhanced accessibility for Korean urban research applications. This platform's comprehensive and up-to-date imagery ensures accurate representation of local built environments, which is critical for the robust analysis of crosswalk safety [78].

For each crosswalk with pedestrian crashes, eight images were captured at 45-degree intervals—0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°—to create a complete 360-degree panoramic view, providing a comprehensive characterization of the environment at each crash location. For crosswalks without pedestrian crashes, the focus was on the centroid of the crosswalk, capturing eight images at the same intervals to ensure consistent coverage of the crosswalk area. Each image measured 640 × 640 pixels, facilitating uniform application of semantic and instance segmentation techniques for analyzing the streetscape features.

a: SEMANTIC IMAGE SEGMENTATION

For semantic image segmentation (Fig. 4), the OCRNet+HRNetV2-W48 model, trained on the ADE20K dataset [79], was employed due to its superior accuracy, achieving a mean intersection over union of 43.25 [80]. This model facilitated the precise detection and classification of various streetscape features, which include building coverage, pedestrian safety fences, pedestrian pathways, shrubs, sidewalks, sky openness, street signage, streetlights, trees, and vehicles (cars, trucks, and buses).

b: INSTANCE SEGMENTATION

Transparency is defined as the ratio of window area to façade area, in contrast with the wall interface, which can

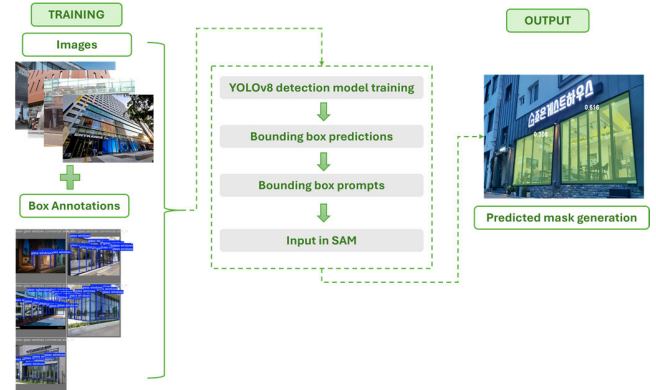


FIGURE 5. Instance segmentation by deep learning using the YOLOv8 model.

obstruct views. Conversely, street windows facilitate interaction between indoor activities and street life, enhancing the depth and vibrancy of street environments. Ground-floor transparency is crucial for creating high-quality pedestrian spaces [37]. To assess the transparency ratio, instance segmentation was performed using the YOLOv8 model, known for its accuracy, real-time processing, and precise object detection [82]. This model generated segmentation masks that accurately delineated window and façade areas in the SVIs.

The YOLOv8 model (Fig. 5) was trained using a comprehensive dataset of annotated building facades in Seoul. The training process included rigorous preprocessing—resizing, normalization, and augmentation—and was fine-tuned using the stochastic gradient descent optimizer and regularization techniques over 72 epochs to enhance performance. After training, the model calculated transparency ratios by extracting window and façade areas from the segmentation masks. The accuracy and reliability of these calculations were validated against ground-truth annotations using metrics such as intersection over union and F1-score [82].

3) BUILT ENVIRONMENT FEATURES

The built environment dataset used in this analysis was categorized into seven groups: density, land use, traffic control, and road and block features. Multiple Euclidean buffers of 30 m, 50 m, and 100 m were established around each crosswalk centroid to analyze and capture the variations in built environment features at different spatial scales. These methodologies are consistent with approaches employed in previous studies [83], [84]. Data sources were acquired from various government entities and organizations and extracted from GIS datasets. This approach incorporates an extensive examination of existing literature, creating a technique encompassing both broad and specific urban form elements.

a: DENSITY FEATURES

Population and business density were considered as density variables. Density metrics were calculated using “jipgyegu” units with areal-weighted interpolation (Eq. 1). For population density, the number of residents within each buffer zone

was assessed, and business density was determined based on the number of businesses located within the same area. The areal-weighted interpolation method (Fig. 3b) enabled the precise estimation of these densities by considering the spatial distribution of data points and ensuring accurate representation within each buffer distance using the following equation:

$$V_k = \sum_{i=1}^n \left(\frac{A_i}{A_j} \times V_j \right), \quad (1)$$

where A_i is the area of intersected feature i ,

A_j is the area of source feature j ,

V_j is the attribute value of source feature j ,

n is the number of intersected features in target buffer k , and

V_k is the sum of all estimated values for target buffer k .

b: LAND-USE FEATURES

To understand the spatial characteristics associated with pedestrian traffic crashes at crosswalks, this study employed two complementary methods to capture land-use features: (1) classification of land-use types based on geospatial data and (2) analysis of facility-level attributes using point of interest (POI) data.

First, land-use types were classified using extensive geographic databases. The analysis focused on six categories commonly used in traffic research: commercial areas, forested areas, industrial areas, mixed-use areas, open spaces, and residential areas. They were measured within a defined buffer surrounding each crosswalk to reflect the general urban form and zoning context.

Second, POI data was used to represent the functional aspects of urban spaces near crosswalks. Seven types of POI facilities were selected based on the existing literature linking them to pedestrian crash rates: accommodation facilities, convenience stores, educational facilities, hospital facilities, hotel facilities, on-street parking lots, and restaurant facilities [85], [86], [87], [88], [89]. POI data were sourced from the VWorld Data Center of the Ministry of Land, Infrastructure, and Transport of South Korea, which provides high-quality, up-to-date national spatial information [90].

c: TRAFFIC CONTROL DEVICES AND TRANSIT INFRASTRUCTURE

To capture the effects of traffic control measures and transit infrastructure on pedestrian safety at crosswalks, this category includes variables related to signalization, signage, and nearby public transportation facilities. Specifically, traffic control features were evaluated through the number of (1) crosswalk time indicators, which provide pedestrians with timing information for safe crossing; (2) crosswalk warning signs, which alert drivers to the presence of crosswalks; (3) road surface direction signs, which direct vehicle movement; and (4) traffic lights, which control vehicle flow and affect pedestrian crossing intervals.

In addition, the presence of transit facilities can influence pedestrian volume and behavior around crosswalks. A higher concentration of public transit stops may increase pedestrian exposure to traffic risks, particularly when the infrastructure is not designed to accommodate increased pedestrian volume [91]. The number of bus stops and subway stations within the defined buffer zones around each crosswalk was assessed systematically to quantify their potential impact on pedestrian safety.

d: ROAD AND BLOCK FEATURES

To evaluate the impact of road features on pedestrian traffic crashes at crosswalks, a range of variables related to road design and spatial connectivity were employed, with several metrics derived from space syntax analysis. Specifically, angular choice and integration were calculated to understand potential movement paths and accessibility within the street network, providing insights into how both pedestrians and vehicles navigate urban spaces. Connectivity was evaluated separately for car roads and pedestrian pathways, highlighting the distinct impacts of these networks on pedestrian safety. Global integration was used to note the position of each crosswalk within the urban road network. All values were processed using AutoCAD for initial data refinement, the Space Syntax Toolkit for QGIS, and a network segmenter for network modeling. Furthermore, the distance from each crosswalk to the nearest expressway and main road was considered. These factors often influence traffic volume and subsequently the crash occurrence risk. The lengths of local roads (back-alley roads), expressways, and main roads within each buffer zone were calculated to understand the scale and density of these road types and their potential relationships with pedestrian crashes.

Road curvature was assessed using the Road Curvature Analyst tool in ArcGIS [92]. This tool identifies and characterizes horizontal curves in roadway layers, calculating curve radius and azimuth using a Bayes classifier and the Douglas–Peucker generalization process to reduce noise [93], [94], [95]. Curves with radii of less than 175 m, considered hazardous based on prior literature [96], were specifically counted within each crosswalk buffer zone to evaluate their potential contribution to pedestrian traffic crash risks. Additionally, the presence of four-way intersections and road widths within each buffer were examined, as these factors are crucial for understanding the dynamics of traffic at crosswalks and the overall configuration of roadways in the study areas.

Spatial morphology has been shown to affect pedestrian safety, with prior research highlighting that block length and area are critical factors influencing pedestrian crashes. Compact urban configurations with shorter block lengths can increase crash occurrence risk by influencing traffic flow patterns and pedestrian visibility [97]. To investigate these spatial characteristics, the block length (perimeter of blocks within designated buffer zones) and block area (total land area

TABLE 1. Hyperparameter tuning using gaussian process-based optimization.

Machine learning algorithms	Tuning hyperparameter	Crosswalk buffer		
		30 m	50 m	100 m
RF	n estimators	350	410	300
	Max depth	7	8	8
	Max features	auto	sqrt	log2
	Min sample leaf	4	2	3
	Min sample split	6	5	4
AdaBoost	n estimators	120	180	150
	Learning rate	0.9	0.7	0.5
XGBoost	n estimators	320	370	340
	Learning rate	0.2	0.1	0.15
SVM	Max depth	5	7	6
	C (regularization)	0.5	1	2
	Kernel	Linear	RBF	Polynomial
ANN	Gamma	0.7	0.5	0.9
	Number of layers	4	3	5
	Number of neurons	256	128	512
CNN	Activation function	tanh	relu	tanh
	Learning rate	0.005	0.01	0.001
	Batch size	64	32	128
	Number of filters	64	32	128
	Filter size	3×3	5×5	3×3
	Number of layers	4	3	5
	Activation function	relu	tanh	relu
	Learning rate	0.0005	0.001	0.0001

RF: random forest; SVM: support vector machine; AdaBoost: adaptive boosting; XGBoost: eXtreme gradient boosting; ANN: artificial neural network; CNN: convolutional neural network; RBF: radial basis function; tanh: hyperbolic tangent; relu: rectified linear unit.

of these blocks) were measured to evaluate their effects on traffic flow and pedestrian safety.

C. MODELLING

Before the model-building phase, min-max scaling (Eq. 2) was applied to normalize the preprocessed data. This scaling technique transforms each feature into a fixed range, typically $[0,1]$, based on the minimum and maximum values of the feature. The formula used for scaling is as follows:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

where $\min(x)$ and $\max(x)$ represent the minimum and maximum feature values, respectively. This normalization ensures that all features contribute equally to the model, preventing any single feature from disproportionately influencing the results owing to differences in scale [98].

After the scaling process, the data were split into training (80%) and testing (20%) sets. Various ML algorithms, including RF, adaptive boosting (AdaBoost), eXtreme gradient boosting (XGBoost), SVM, ANN, and convolutional neural network (CNN), were trained and analyzed using the Scikit-learn ML library in Python 3.8.4. Hyperparameter tuning was performed using the Gaussian process-based optimization with 10-fold cross-validation to enhance model performance (Table 1). This approach efficiently navigated the hyperparameter spaces by trading off exploration and exploitation [99]. The trained models were evaluated using the testing dataset, with performance metrics including classi-

fication accuracy, precision, recall, and F1-score; the optimal model was selected based on these metrics.

D. MODEL INTERPRETABILITY

Users commonly face the interpretability challenge in ML and DL. To address this problem, Bîl et al. [94] developed SHAP, a method rooted in game theory that provides a local explanation and optimal credit allocation for various ML models [100]. To enhance the computational efficiency of tree-based ensemble models, TreeSHAP (Eq. 3), which optimizes the calculation of Shapley values by leveraging tree model nodes, was introduced [101].

TreeSHAP accelerates computation and improves the analysis of variable dependencies and feature interactions. In this study, the RF model—a tree-based ensemble model—was chosen, utilizing TreeSHAP to visualize the explanations for the crash occurrence model. Table 2 summarizes the key variables utilized for this approach. The TreeSHAP equation used in the analysis is as follows:

$$\phi_i = \sum_{S \subseteq N-i} \frac{|S|! (M - |S| - 1)!}{M!} [f_x(S \cup i) - f_x(S)], \quad (3)$$

where ϕ_i represents the impact of feature i on the model output, based on its marginal effect [100]; f_x denotes the prediction function of the RF model classifier, and M represents the total number of features within the model. S refers to any subset of features excluding the i -th feature, with $|S|$ indicating the size of that subset. Applying the SHAP analysis to RF models provides a detailed examination of the factors influencing pedestrian crashes.

IV. RESULTS

A. MODEL COMPARISON

The performance of ML models was evaluated across three spatial scales—30 m, 50 m, and 100 m—using accuracy, precision, recall, and area under the curve of the receiver operating characteristic metrics. As shown in Table 3, RF consistently outperformed other models, achieving the highest performance across all metrics at the 50-m scale. XGBoost demonstrated competitive results, followed by CNN, whereas ensemble models such as AdaBoost and ANN showed moderate performance. SVM exhibited the lowest predictive capability. For comparison, logistic regression (LR) (Appendix D), noticeably lower performance than the machine learning models, further underscoring the advantages of advanced ML approaches in capturing complex spatial patterns.

Overall, the 50-m buffer produced the highest accuracy and stability, outperforming both the 30-m and 100-m buffers. RF's superior precision (0.907) indicates its effectiveness in minimizing false positives, making it particularly reliable for identifying crash-prone crosswalks. XGBoost, with a high recall (0.835), was effective at capturing crash-prone locations, though with a slightly lower precision. Receiver operating characteristic curve analysis (Fig. 6)

TABLE 2. Summary of key variables used in interpretable machine learning.

Features	Description (unit of analysis)	Crosswalk with pedestrian crashes				Crosswalk without pedestrian crashes			
		Mean	SD	Min	Max	Mean	SD	Min	Max
Exposure features									
Pedestrian volume	Mean pedestrian volume observed within a 50-m crosswalk buffer (foot traffic intensity)	816.9	767.3	0	7,436.7	704.3	651.2	0	7,397.0
Land-use features									
Residential area	Residential area within a 50-m crosswalk buffer (m²)	264.1	487	0	7,854	309.8	558.2	0	7,854
Commercial area	Number of commercial areas within a 50-m crosswalk buffer (m²)	140.9	294.4	0	7,556	129.8	302.9	0	7,556
Restaurant facilities	Number of restaurant facilities within a 50-m crosswalk buffer (m²)	4.2	5.7	0	72	2.9	4.8	0	133
Traffic control features									
Crosswalk time indicators	Number of crosswalk time indicators within a 50-m crosswalk buffer	1.1	2	0	14	1.4	2.2	0	14
Traffic lights	Number of traffic lights within a 50-m crosswalk buffer	3.9	4.5	0	30	4.8	4.8	0	29
Road features									
Connectivity (car road)	Mean car road connectivity within a 50-m crosswalk buffer measured by the Space Syntax tool (m²)	4.7	1.1	0	15.6	5	1.1	0	15.6
Local road (back alley)	Total length of the local road (back alley) within a 50-m crosswalk buffer (m)	75.4	80.2	0	501.1	68.6	81.1	0	602.9
Road width	Mean road width within a 50-m crosswalk buffer (m²)	15.9	9.2	0	100	15.2	9.5	0	79
Number of sharp curves	Number of sharp curves within a 50-m crosswalk buffer	0.2	0.6	0	10	0.2	0.6	0	10
Transit stops									
Bus stops	Number of bus stops within a 50-m crosswalk buffer	0.1	0.5	0	9	0.1	0.3	0	6
Streetscape features (SVI)									
Building ratio	Proportion of buildings derived from semantic segmentation (%)	24.6	12.4	0	76	22.2	10.2	0	66.1
Shrub ratio	Proportion of small vegetation derived from semantic segmentation (%)	0.4	0.7	0	14.4	0.9	1.2	0	14
Sidewalk ratio	Proportion of sidewalks derived from semantic segmentation (%)	4.6	3.3	0	42.9	5.9	4.3	0	43.8
Pedestrian safety fence ratio	Proportion of pedestrian safety fences in the area derived from semantic segmentation (%)	0.5	0.8	0	11	0.9	1.2	0	13.6
Sky openness ratio	Proportion of visible sky derived from semantic segmentation (%)	22.1	14.1	0	69.9	31.6	8.3	0	89.9
Street signage ratio	Proportion of street signage derived from semantic segmentation (%)	0.4	0.5	0	6.9	0.3	0.4	0	4.4
Streetlight ratio	Proportion of streetlights in the area derived from semantic segmentation (%)	0.1	0.1	0	2.4	0.2	0.2	0	2.8
Tree ratio	Proportion of tree coverage derived from semantic segmentation (%)	8.4	7.9	0	60.8	4.8	5.5	0	52.2
Transparency ratio	Proportion of building façade with transparent windows derived from instance segmentation (%)	19.7	13.1	0	56.0	18.6	20.4	0	68.1

SD: standard deviation; SVI: street-view imagery.

further confirmed RF's strong true-positive performance, followed closely by XGBoost and CNN, which also exhibited stable and consistent performances.

B. RESULTS OF INTERPRETABLE ML

Fig. 7 presents the SHAP analysis results for the RF model, elucidating the top-20 predictor variables ranked by their mean SHAP values. These values quantify the average impact of each feature on the model's predictions, with larger absolute mean SHAP values indicating greater influence.

The left panel ranks the variables by their importance, whereas the right panel shows violin plots of SHAP value distributions, which capture the variability of each feature's contribution to predicting pedestrian crash at crosswalks. The color gradient within the violin plots (red for high and blue for low feature values) illustrates the directionality of each predictor's influence.

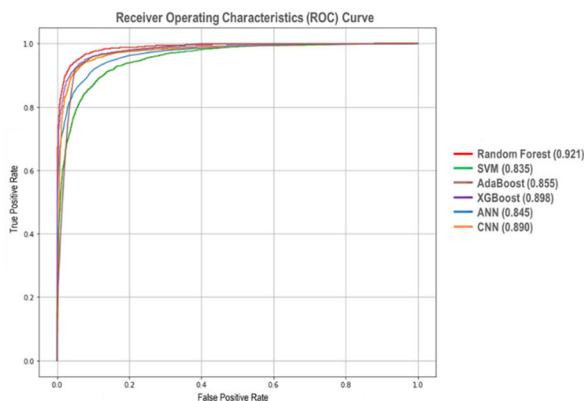
Several SVI variables have proven to be highly effective in predicting crash occurrences at crosswalks in Seoul. Particularly, sky openness ratio, streetlight ratio, shrub ratio, and sidewalk ratio have exhibited a negative association, indicating their substantial contribution to reducing the occurrence of crosswalk pedestrian crashes. In contrast, building ratio, tree ratio, and transparency ratio showed a positive influence, reflecting their association with increased pedestrian crash risks. Additionally, street signage ratio demonstrated a minimal positive influence. Collectively, these variables underscore the complex interplay of visual and structural elements in shaping crosswalk safety outcomes.

Pedestrian volume, representing direct exposure, emerged as a relevant predictor of crosswalk crashes, with higher pedestrian volume levels associated with increased crash likelihood.

TABLE 3. Comparison of accuracy, precision and recall with different model algorithms.

		Crosswalk Buffer		
Model	Metrics	30m	50m	100m
RF	Recall	0.824	0.835	0.830
	Accuracy	0.910	0.934	0.912
	Precision	0.885	0.907	0.899
AdaBoost	Recall	0.775	0.785	0.770
	Accuracy	0.860	0.870	0.846
	Precision	0.832	0.837	0.825
XGBoost	Recall	0.825	0.835	0.835
	Accuracy	0.890	0.909	0.900
	Precision	0.885	0.891	0.888
SVM	Recall	0.732	0.745	0.740
	Accuracy	0.780	0.820	0.813
	Precision	0.815	0.829	0.825
ANN	Recall	0.776	0.780	0.773
	Accuracy	0.840	0.865	0.850
	Precision	0.830	0.817	0.820
CNN	Recall	0.815	0.813	0.816
	Accuracy	0.898	0.900	0.890
	Precision	0.898	0.904	0.895

RF: random forest; SVM: support vector machine; AdaBoost: adaptive boosting; XGBoost: eXtreme gradient boosting ANN: artificial neural network; CNN: convolutional neural network

**FIGURE 6.** Receiver operating characteristics curve of different models for crosswalk traffic crash with 50-m crosswalk buffer. SVM: support vector machine; AdaBoost: adaptive boosting; XGBoost: eXtreme gradient boosting ANN: artificial neural network; CNN: convolutional neural network.

Several land-use features also exhibited notable effects: Restaurant facilities and commercial areas showed positive SHAP values, indicating the higher incidence of crashes in their crosswalk areas, which are characterized by heightened activity.

In contrast, residential areas demonstrated a negative SHAP value, reflecting their distinct association with reduced occurrence of pedestrian incidents at crosswalks.

Road features and transit stop variables revealed varying impacts on crosswalk safety. Features such as length of local roads, road width, and number of bus stops primarily exhib-

ited positive associations, whereas connectivity of car roads demonstrated a minimal negative influence on pedestrian crash occurrence. Lastly, traffic control features consistently showed a mitigating effect on crosswalk crash risks. Variables including the number of traffic lights and crosswalk time indicators exhibited negative SHAP values, highlighting their role in regulating pedestrian-vehicle interactions and enhancing safety at crosswalks.

The violin plots in Fig. 7 also reveal patterns of variability of key features. For example, the building ratio showed a wide distribution of SHAP values, indicating its varying impact on predictions across the dataset. In contrast, features such as traffic lights and crosswalk time indicators displayed narrower distributions, suggesting that their contributions are more uniform. These results provide a comprehensive data-driven ranking of predictors influencing the model's outputs. The integration of feature importance, directionality, and variability offers valuable insights into the factors affecting the risk of pedestrian crashes at crosswalks, setting the stage for further interpretation and discussion.

1) NONLINEAR RELATIONSHIPS BETWEEN CRASH OCCURRENCE AND BUILT ENVIRONMENT FEATURES

Scatter plots with locally weighted scatterplot smoothing (Appendix VI) illustrate the nonlinear relationships between built environment features and crosswalk pedestrian crashes. For instance, sky and streetlight ratios exhibited decreasing trends in pedestrian crash occurrences, with the rate of decrease tapering off at both high and low levels. Similarly, the shrubs ratio showed an initial decrease in pedestrian crashes, stabilizing as the ratio increased. Sidewalk coverage revealed a consistent decreasing trend, with stronger effects observed at lower coverage levels. Pedestrian safety fences displayed a similar pattern, with crash occurrences decreasing more prominently at lower ratios before leveling off at higher densities. In contrast, tree and building ratios showed an increasing trend in pedestrian crashes, with higher densities corresponding to greater risks.

Traffic signs showed a stable or slightly increasing trend, leveling off beyond certain ranges. SHAP values for traffic control measures and road configuration variables (Appendix B) also indicated nonlinear patterns. The number of traffic lights and crosswalk time indicators demonstrated decreasing trends in crash occurrences, with the latter stabilizing beyond certain thresholds. Features, namely length of local roads and road width exhibited increasing trends, particularly for roads with intermediate width. The number of sharp curves and connectivity of car roads displayed marginal increases in crash occurrences, with connectivity showing stabilization or slight decreases at higher levels. Moreover, POIs displayed varying trends in relation to crosswalk pedestrian crashes (Appendix C). Residential density and number of educational facilities were associated with decreasing crash occurrences, whereas restaurants and bus stops showed increasing trends.

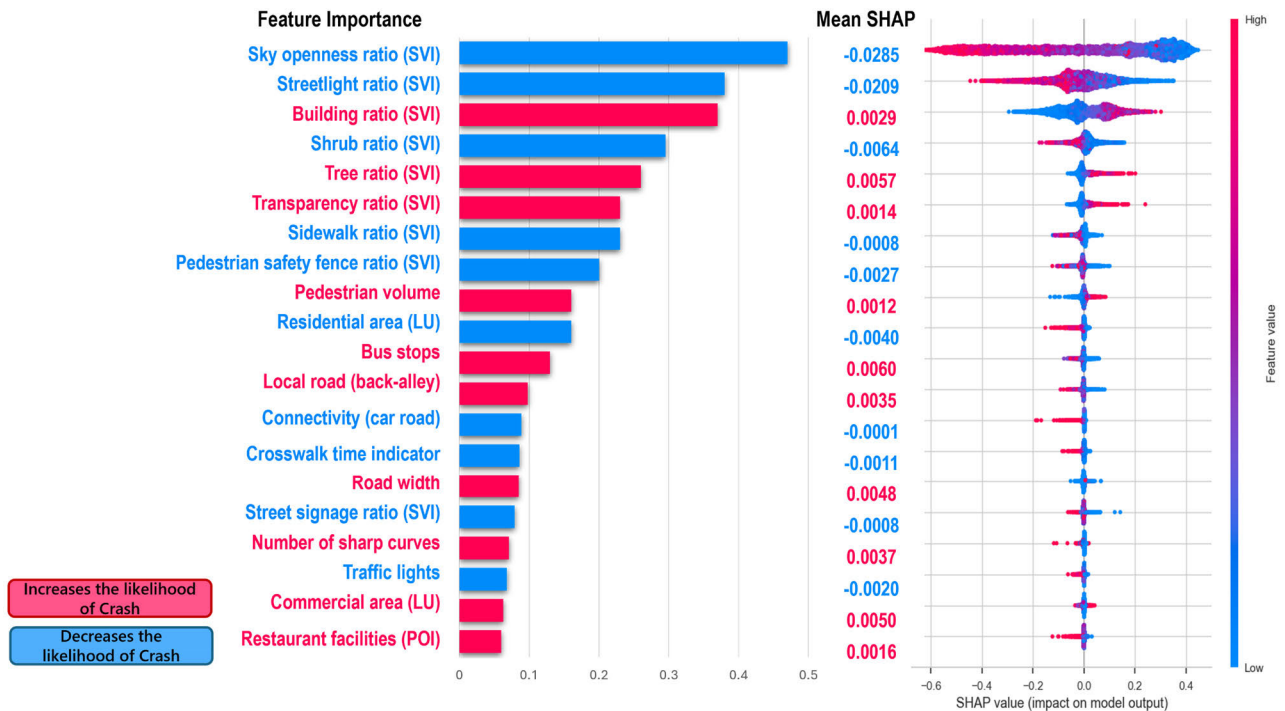


FIGURE 7. Variable importance of the top-20 predictor variables for random forest and their corresponding mean SHapley Additive exPlanations (SHAP) values. The violin plot shows SHAP value distribution, with colors representing the values of predictor variables (red: high value; blue: low value). SVI: street-view imagery; LU: land use; POI: point of interest.

2) INTERACTION EFFECTS OF BUILT ENVIRONMENT FEATURES ON CROSSWALK PEDESTRIAN CRASH

The interaction analysis (Fig. 8) revealed complex relationships between built environment characteristics and the likelihood of crosswalk pedestrian crashes. Crosswalks with a lower sky openness ratio combined with a higher building ratio were associated with an elevated likelihood of crash occurrence (Fig. 8a), whereas those with a higher sky openness ratio and lower building ratio exhibited a reduced likelihood. Similarly, environments with a higher tree ratio and lower sky openness ratio were linked to a lower likelihood of crash occurrence (Fig. 8b).

Interactions involving pedestrian infrastructure indicated that crosswalks with more time indicators and a higher ratio of pedestrian safety fences were associated with a lower likelihood of crosswalk pedestrian crashes (Fig. 8c). In contrast, crosswalks near back alleys with high building ratios were identified as higher-risk locations (Fig. 8d).

Proximity to public transportation and road design also influenced the likelihood of crosswalk pedestrian crashes. Crosswalks located near multiple bus stops and wider roadways exhibited a higher probability of crash occurrences (Fig. 8e), as did those near multiple bus stops with a lower pedestrian safety fence ratio (Fig. 8f). Additionally, crosswalks with high transparency ratios and narrow sidewalks were associated with an increased likelihood of pedestrian crashes (Fig. 8g). These findings underscore the complex interplay between built environment variables, offering a

detailed lens through which pedestrian safety at crosswalks can be understood.

V. DISCUSSION

A. METHODOLOGICAL CONTRIBUTIONS

This study advances research on urban safety by integrating interpretable ML with SVI, utilizing both semantic and instance segmentation. While previous research primarily relied on semantic segmentation to classify general streetscape elements such as roads, sidewalks, and vegetation [30], [32], [102], this study fills a critical gap by applying instance segmentation to quantify building façade transparency. Specifically, the study measures the ratio of window area to total façade area, offering a novel indicator of pedestrian visibility conditions in the built environment. The RF model, optimized at a 50-meter spatial resolution, outperformed other ML and DL models in terms of predictive accuracy. Its integration with SHAP analysis provided a transparent and interpretable framework for identifying nonlinear relationships and interaction effects between built environment features and pedestrian crash risks. This methodological approach enhances spatial granularity and contextual sensitivity, offering a replicable model for future research on pedestrian safety in urban environments.

B. KEY PREDICTORS AND INTERACTION EFFECTS

SHAP analysis revealed that streetscape features extracted from street view imagery, such as sky openness, building

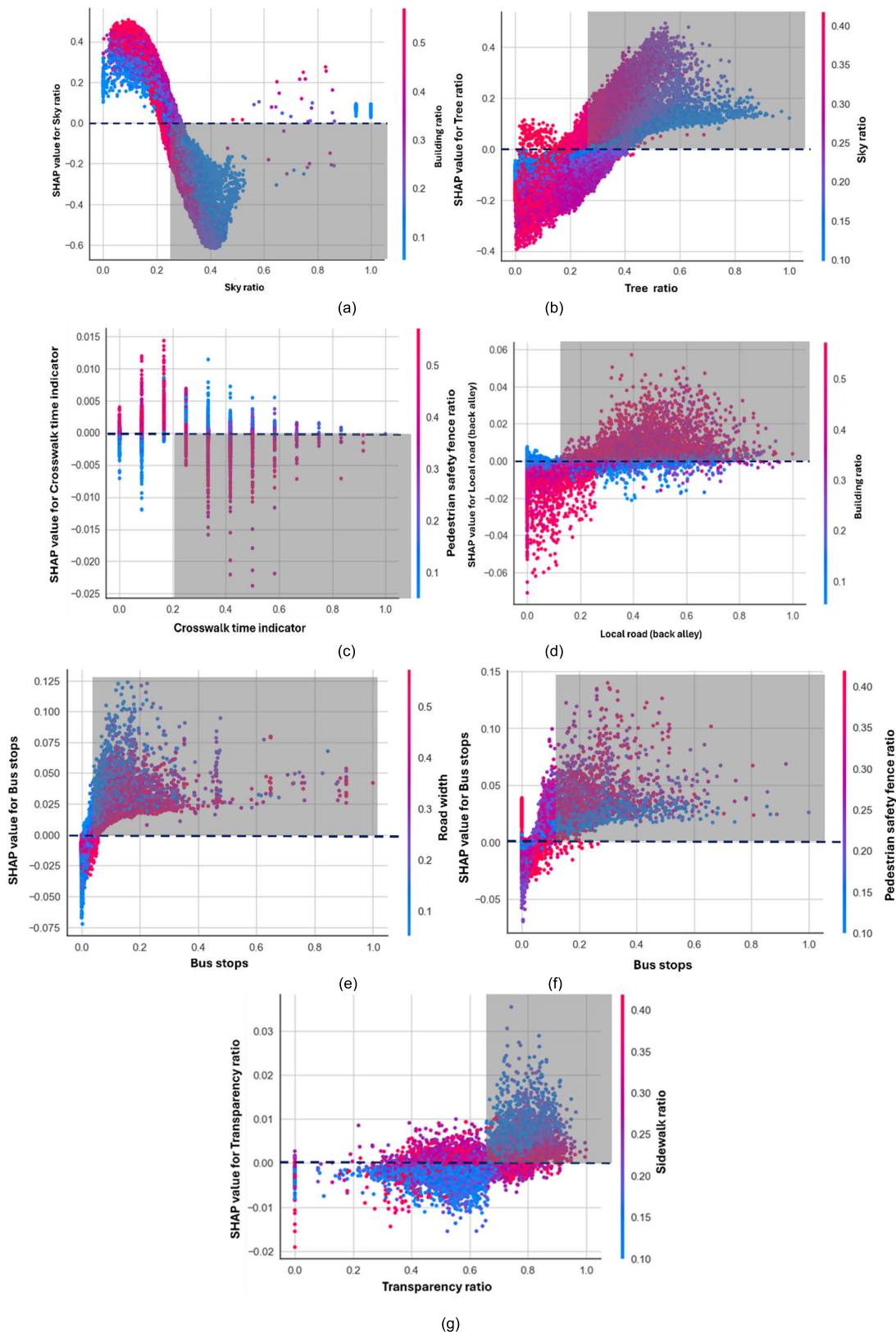


FIGURE 8. Interaction effects between built environment features and the occurrence of pedestrian crashes at crosswalks (a~g). SHAP: SHapley Additive exPlanations.

coverage, and sidewalk ratios, were the most influential predictors of crosswalk crashes. Greenery exhib-

ited context-dependent impacts: while shrubs tended to improve safety by preserving clear sightlines, dense tree

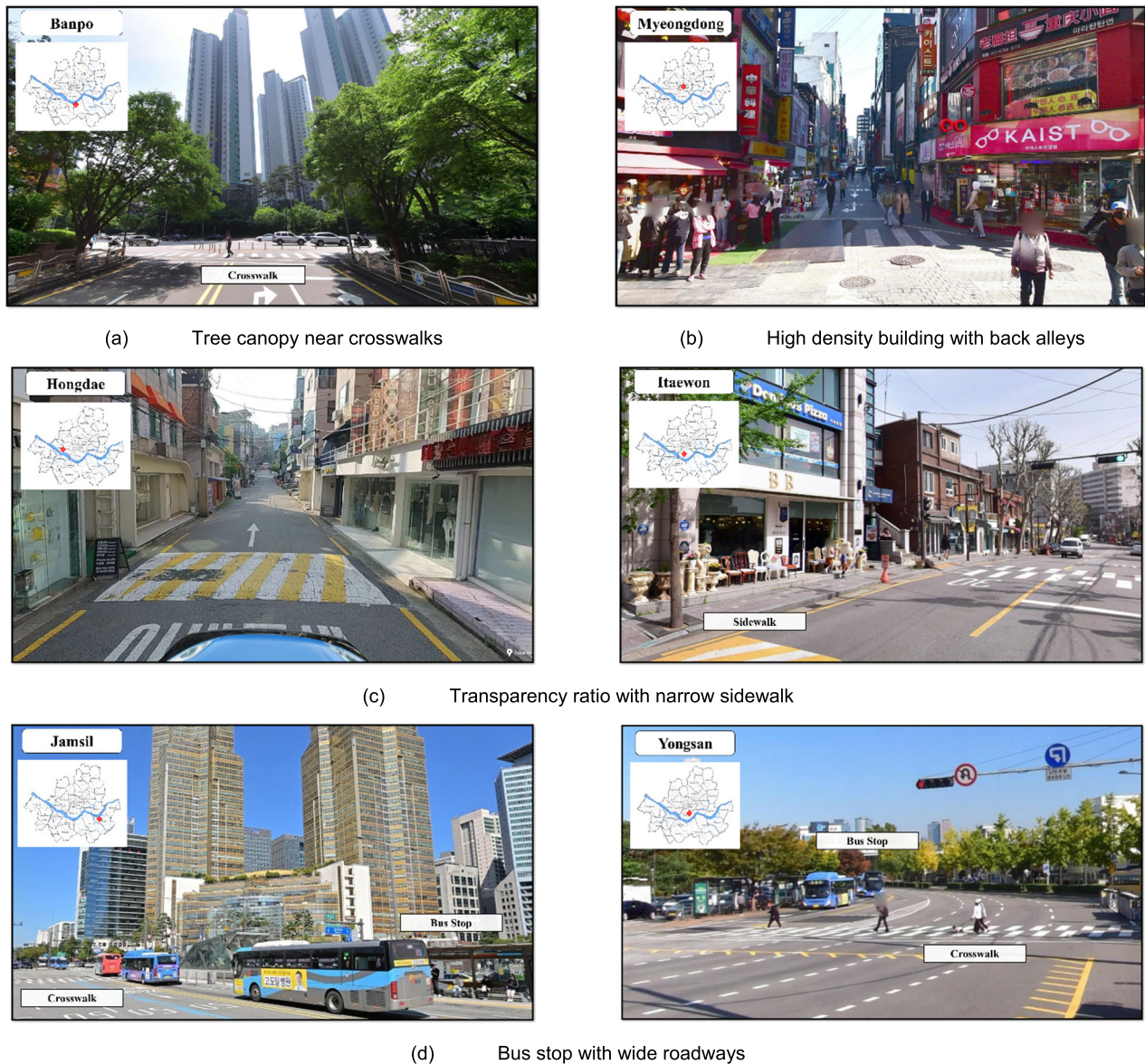


FIGURE 9. Street view imagery of crosswalks across Seoul, highlighting the interaction of key features identified in the study's findings.

canopies near crosswalks, as observed in Seoul's Banpo and Seocho districts, increased collision risks by obstructing visibility (Fig. 9a) [30], [103]. This phenomenon has also been documented in other urban contexts, such as Yokohama, Japan [35]. In addition, pedestrian volume emerged as a critical factor, underscoring the need for targeted interventions in high-traffic zones such as Myeongdong and Sinsa-dong, where visibility obstructions and unpredictable pedestrian movements are more prevalent (Fig. 9b) [104].

Furthermore, the analysis identified specific combinations of built environment features that amplified safety risks. In commercial districts, the presence of high-rise buildings adjacent to back alleys restricted visibility, a situation exacerbated by pedestrian congestion and obstructive shop

displays. Similarly, crosswalks located near transparent building facades in areas such as Apgujeong and Hongdae were associated with increased risks due to visual distractions and constrained pedestrian spaces (Fig. 9c) [105]. Bus stops situated on arterial roads, for example in Jamsil and Yongsan, created additional challenges by allowing stationary buses to obstruct sightlines and prompt unsafe pedestrian behaviors (Fig. 9d) [106], [107], [108]. These patterns mirror issues documented in other high-traffic urban environments. Collectively, these findings highlight the importance of analyzing interactions between built environment features, rather than focusing solely on individual variables, in developing effective, context-sensitive strategies and mitigating pedestrian crash risks at crosswalks.

C. IMPLICATIONS FOR ENHANCING CROSSWALK SAFETY

This study's results advocate data-driven, site-specific interventions to enhance pedestrian safety at crosswalks. One important implication is the need for strategic greenery management, such as trimming tree canopies near crosswalks to prevent visibility obstructions while maintaining shrubs to preserve clear sightlines. In addition, upgrading pedestrian infrastructure by widening sidewalks, installing refuge islands, and deploying countdown timers in high-volume zones can address safety concerns in areas with intense pedestrian activity [9], [109], [110].

Design adjustments should be both enhanced and complemented to address the unique challenges of high-risk areas. For instance, repositioning bus stops to minimize conflicts with pedestrian crossings; limiting obstructive shop displays, especially those combined with transparent facades in commercial areas; and managing visual clutter can help mitigate visual and spatial hazards. By integrating advanced computational analysis with practical urban design, this study provides planners with actionable strategies to prioritize pedestrian safety while sustaining the vibrancy and accessibility of urban environments.

This study, conducted in a dense, transit-oriented metropolitan setting of Seoul, employs a methodological approach that integrates interpretable ML with semantic and instance segmentation of SVI, which is broadly applicable to other cities. These techniques allow for scalable, fine-grained analysis of built environment features in diverse contexts. However, we acknowledge that the specific risk patterns identified here may not directly generalize to cities with lower density, different mobility patterns, or car-dominated transport systems. Urban form influences nature and prevalence of pedestrian safety risks: for instance, wide suburban arterials may present different hazard configurations than narrow commercial corridors in Seoul. Therefore, while our core analytical approach can be adapted across geographies, applying it in new settings would require contextual calibration and possibly new variables to reflect local infrastructure, behavioral norms, and policy environments. Future research should test this approach in varied urban systems to assess the robustness of key predictors and support context-sensitive pedestrian safety interventions.

VI. CONCLUSION

This study investigated pedestrian crash risks at crosswalks by integrating interpretable ML techniques with advanced semantic and instance segmentation to better capture the visual and spatial characteristics of the built environment. SHAP analysis revealed that streetscape features such as sky openness, building coverage, and sidewalk conditions were among the most influential predictors of crash risk. Notably, the innovative use of instance segmentation to quantify building façade transparency provided a novel proxy for pedestrian visibility, complementing traditional semantic features. Key predictors identified through RF modeling

included tree coverage, bus stop density, land-use mix, curved street segments, and façade transparency. These features often interact synergistically, with their combined influence on crash risk exceeding the sum of individual effects, underscoring the complexity of pedestrian safety dynamics.

While the model demonstrated strong predictive accuracy at a fine spatial resolution, several limitations must be acknowledged. The reliance on data from a single urban context constrains the generalizability of findings to cities with different built environments, transport systems, or policy frameworks. The study primarily focused on quantifiable environmental features and did not incorporate behavioral or perceptual factors, such as pedestrian awareness, driver attentiveness, or social conditions.

In addition to spatial constraints, temporal considerations must be addressed. The five-year dataset (2018–2022) does not account for seasonal fluctuations in pedestrian volume or crash patterns. It also excludes the effects of recent infrastructure upgrades and overlaps with the COVID-19 pandemic, during which travel behavior and mobility patterns changed dramatically. These dynamics may have affected pedestrian exposure in ways not captured by the model. Future research incorporating real-time, seasonal, and post-pandemic data would help address these gaps.

Another limitation lies in the binary modeling approach adopted in this study. Although binary classification models often outperform count-based models due to their reduced complexity, they may overlook the nuanced variation in crash frequencies. Subsequent studies should consider alternative specifications, such as count-based, hierarchical, or hybrid models, to better capture outcome variability while preserving predictive accuracy. Further research should also aim to validate and extend these findings across diverse urban contexts with varying morphological and policy conditions to enhance external validity. Incorporating real-time and longitudinal data could enable dynamic modeling that better reflects changing urban environments and pedestrian-vehicle interactions. Furthermore, integrating qualitative insights and hybrid modeling approaches that combine ML with traditional statistical techniques may improve both the predictive power and interpretability of pedestrian safety assessments. Such advances would support the development of more holistic and actionable interventions.

In conclusion, this study contributes to urban safety research by demonstrating how detailed visual data and interpretable ML methods can uncover complex built environment factors influencing pedestrian crash risk. These insights have practical implications for urban planning, highlighting the importance of strategic transit stop placement, thoughtful vegetation management, and enhanced sidewalk and crosswalk design to promote pedestrian safety. By advancing data-driven approaches, this research lays the groundwork for more effective, targeted interventions to reduce pedestrian crashes in urban settings.

APPENDIX A

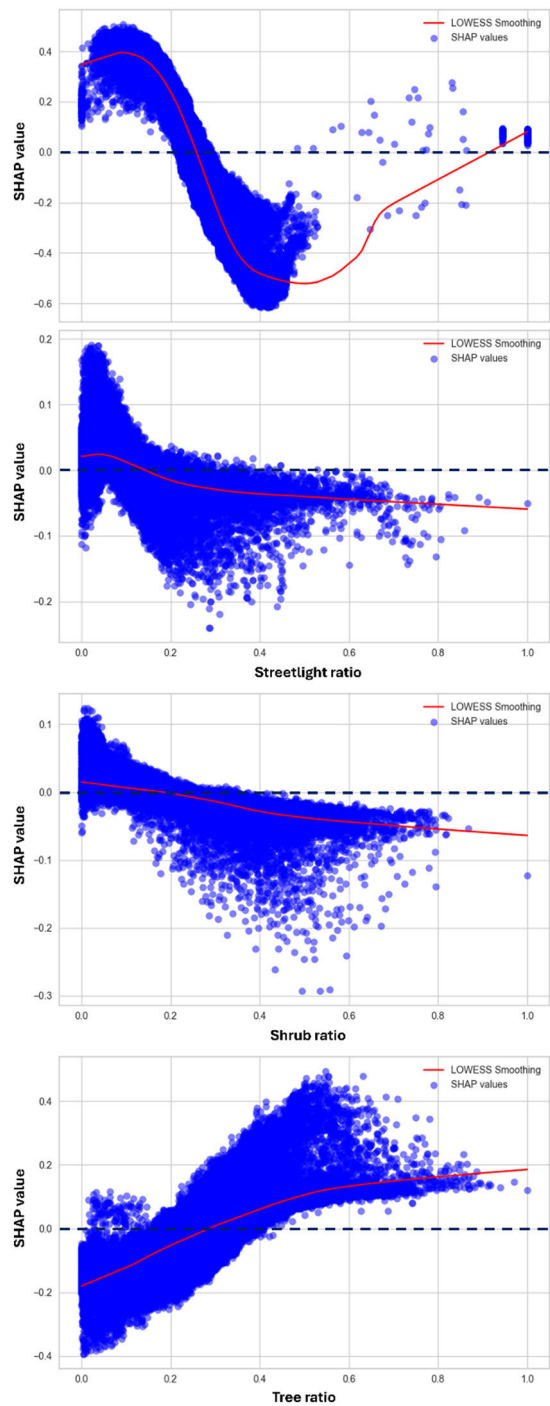


FIGURE 10. Nonlinear relationship between pedestrian crosswalk crashes and streetscape features.

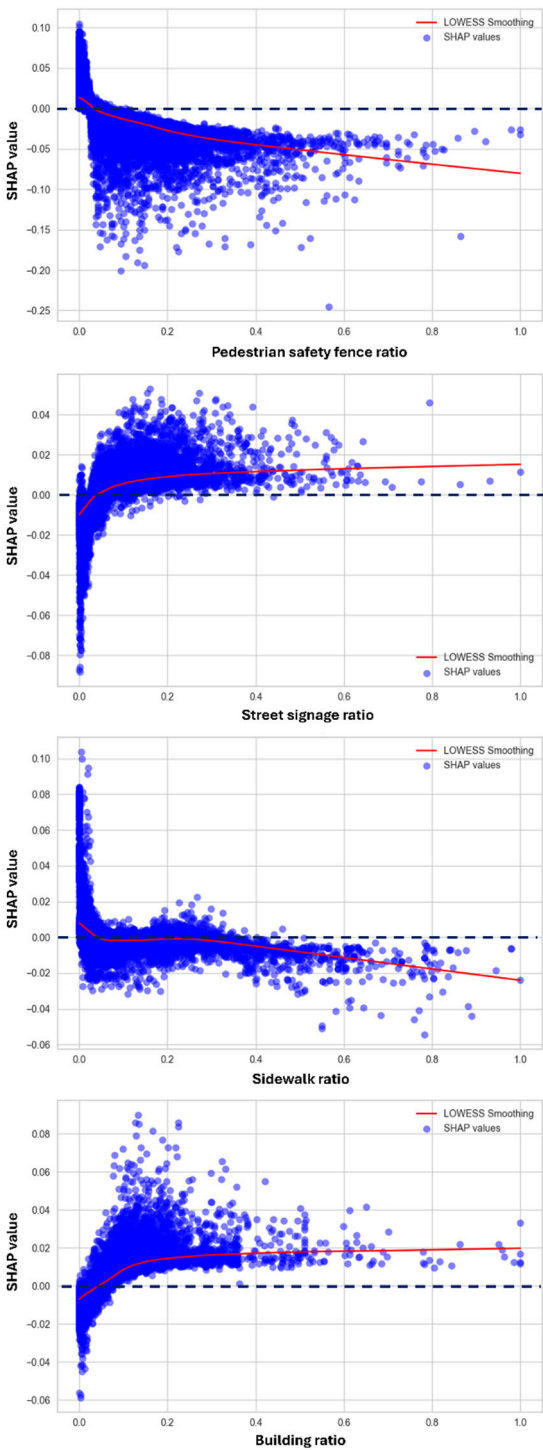


FIGURE 10. (Continued.) Nonlinear relationship between pedestrian crosswalk crashes and streetscape features.

APPENDIX B

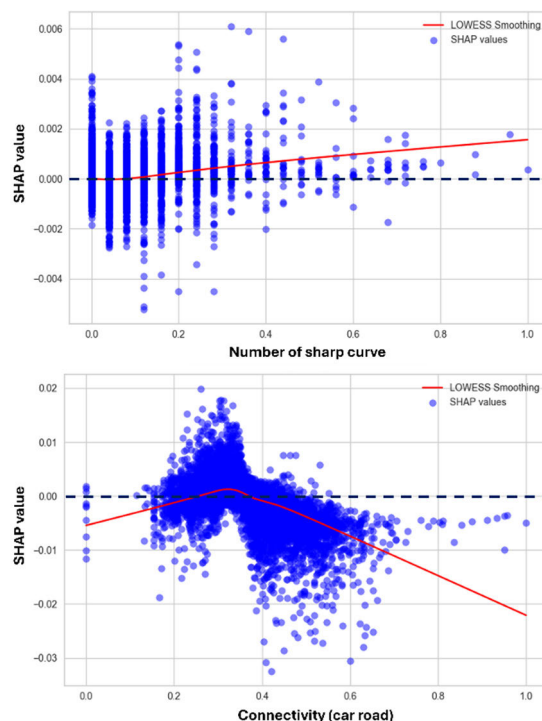
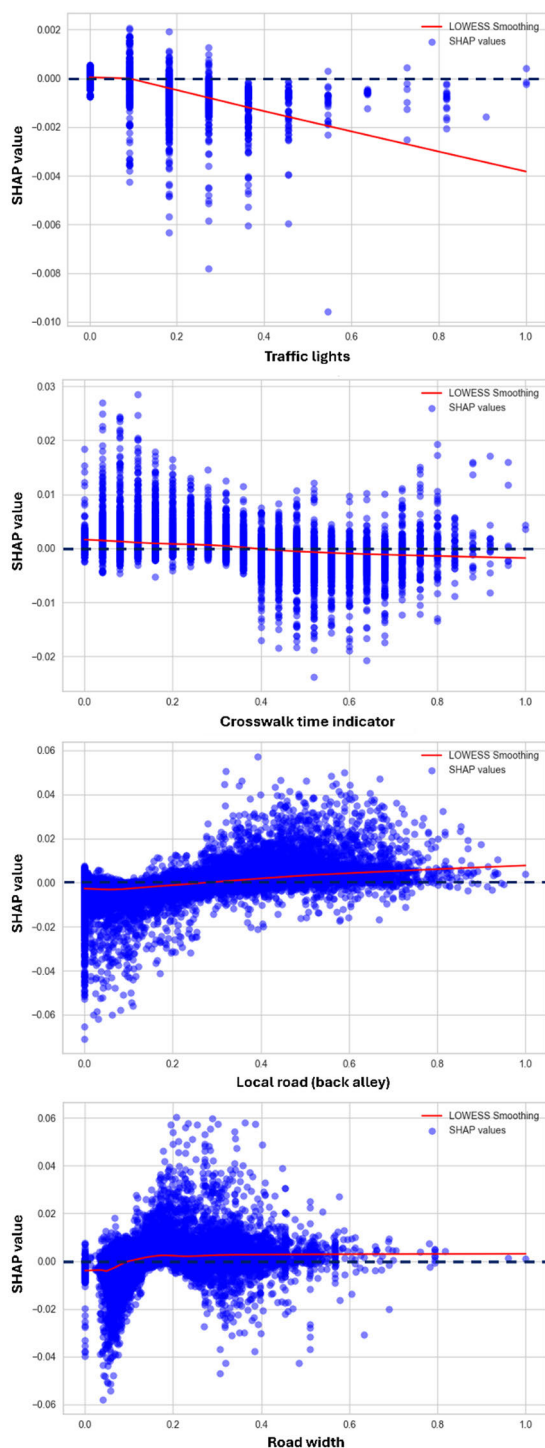


FIGURE 11. (Continued.) Nonlinear relationship between pedestrian crosswalk crashes and traffic control and road features.

FIGURE 11. Nonlinear relationship between pedestrian crosswalk crashes and traffic control and road features.

APPENDIX C

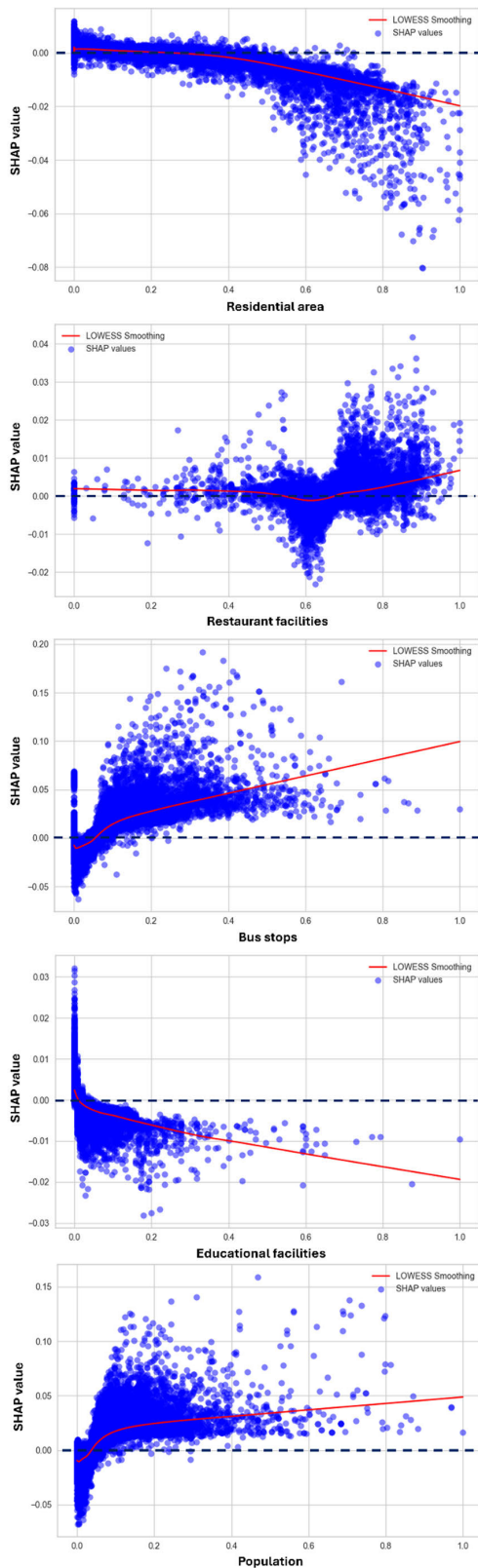


FIGURE 12. Nonlinear relationship between pedestrian crosswalk crashes and land use and density features.

APPENDIX D

TABLE 4. Table comparison of traditional statistic model (LR) and machine learning models (RF, AdaBoost, XGBoost, SVM, ANN, CNN) evaluation result.

Model	Recall	Accuracy	Precision
LR	0.635	0.700	0.714
RF	0.835	0.934	0.907
AdaBoost	0.785	0.870	0.837
XGBoost	0.835	0.909	0.891
SVM	0.745	0.820	0.829
ANN	0.780	0.865	0.817
CNN	0.813	0.900	0.904

ACKNOWLEDGMENT

This article is a modified version of a section of Barquilla’s dissertation, Exploring Street-Level Environmental Influences on Pedestrian Crashes Using Street View Images and Machine Learning. The authors would like to thank Dr. Myung-Jin Jun, an Emeritus Professor at Chung-Ang University, for his insightful comments on a previous version of this work.

REFERENCES

- [1] *Global Status Report on Road Safety*, World Health Organization, Geneva, Switzerland, 2023, vol. 15, no. 4.
- [2] *Traffic Accident Analysis System (TAAS)*, (KoROAD) Road Traffic Authority, Seoul, South Korea, 2023.
- [3] C. Xin, R. Guo, Z. Wang, Q. Lu, and P.-S. Lin, “The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized ordered probability model with heterogeneity in means and variances,” *Analytic Methods Accident Res.*, vol. 16, pp. 117–132, Dec. 2017, doi: [10.1016/j.amar.2017.10.001](https://doi.org/10.1016/j.amar.2017.10.001).
- [4] G. Lee, Y. Park, J. Kim, and G.-H. Cho, “Association between intersection characteristics and perceived crash risk among school-aged children,” *Accident Anal. Prevention*, vol. 97, pp. 111–121, Dec. 2016, doi: [10.1016/j.aap.2016.09.001](https://doi.org/10.1016/j.aap.2016.09.001).
- [5] S. Ukkusuri, L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez, “The role of built environment on pedestrian crash frequency,” *Saf. Sci.*, vol. 50, no. 4, pp. 1141–1151, Apr. 2012, doi: [10.1016/j.ssci.2011.09.012](https://doi.org/10.1016/j.ssci.2011.09.012).
- [6] S. S. Zajac and J. N. Ivan, “Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut,” *Accident Anal. Prevention*, vol. 35, no. 3, pp. 369–379, May 2003, doi: [10.1016/s0001-4575\(02\)00013-1](https://doi.org/10.1016/s0001-4575(02)00013-1).
- [7] Z. An, B. Xie, and Q. Liu, “No street is an island: Street network morphologies and traffic safety,” *Transp. Policy*, vol. 141, pp. 167–181, Sep. 2023, doi: [10.1016/j.tranpol.2023.07.023](https://doi.org/10.1016/j.tranpol.2023.07.023).
- [8] E. Dumbaugh and W. Li, “Designing for the safety of pedestrians, cyclists, and motorists in urban environments,” *J. Amer. Planning Assoc.*, vol. 77, no. 1, pp. 69–88, Dec. 2010, doi: [10.1080/01944363.2011.536101](https://doi.org/10.1080/01944363.2011.536101).
- [9] P. Stoker, A. Garfinkel-Castro, M. Khayesi, W. Odero, M. N. Mwangi, M. Peden, and R. Ewing, “Pedestrian safety and the built environment: A review of the risk factors,” *J. Planning Literature*, vol. 30, no. 4, pp. 377–392, Nov. 2015, doi: [10.1177/0885412215595438](https://doi.org/10.1177/0885412215595438).
- [10] R. Ewing and E. Dumbaugh, “The built environment and traffic safety: A review of empirical evidence,” *J. Planning Literature*, vol. 23, no. 4, pp. 347–367, May 2009, doi: [10.1177/0885412209335553](https://doi.org/10.1177/0885412209335553).
- [11] M. A. Abdel-Aty and A. E. Radwan, “Modeling traffic accident occurrence and involvement,” *Accident Anal. Prevention*, vol. 32, no. 5, pp. 633–642, Sep. 2000, doi: [10.1016/s0001-4575\(99\)00094-9](https://doi.org/10.1016/s0001-4575(99)00094-9).

- [12] Y. Ma, F. Wang, S. Chen, G. Xing, Z. Xie, and F. Wang, "A dynamic method to predict driving risk on sharp curves using multi-source data," *Accident Anal. Prevention*, vol. 191, Oct. 2023, Art. no. 107228, doi: [10.1016/j.aap.2023.107228](https://doi.org/10.1016/j.aap.2023.107228).
- [13] Z. Sawalha and T. Sayed, "Evaluating safety on urban arterial roadways," *J. Transp. Eng.*, vol. 2, no. 127, pp. 151–158, 2001.
- [14] A. Sheykhsfard, F. Haghighi, E. Papadimitriou, and P. Van Gelder, "Analysis of the occurrence and severity of vehicle-pedestrian conflicts in marked and unmarked crosswalks through naturalistic driving study," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 76, pp. 178–192, Jan. 2021, doi: [10.1016/j.trf.2020.11.008](https://doi.org/10.1016/j.trf.2020.11.008).
- [15] O. F. Hamim and S. V. Ukkusuri, "Towards safer streets: A framework for unveiling pedestrians' perceived road safety using street view imagery," *Accident Anal. Prevention*, vol. 195, Feb. 2024, Art. no. 107400, doi: [10.1016/j.aap.2023.107400](https://doi.org/10.1016/j.aap.2023.107400).
- [16] X. F. Ji and X. Qiao, "Nonlinear influence of built environment on pedestrian traffic accident severity," *J. Transp. Syst. Eng. Inf. Technol.*, vol. 23, no. 1, pp. 314–323, 2023, doi: [10.16097/j.cnki.1009-6744.2023.01.033](https://doi.org/10.16097/j.cnki.1009-6744.2023.01.033).
- [17] K. Kim, I. M. Brunner, and E. Y. Yamashita, "Influence of land use, population, employment, and economic activity on accidents," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1953, no. 1, pp. 56–64, Jan. 2006, doi: [10.1177/0361198106195300107](https://doi.org/10.1177/0361198106195300107).
- [18] Z. Wang, L. Hu, F. Wang, M. Lin, and N. Wu, "Assessing the impact of different population density scenarios on two-wheeler accident characteristics at intersections," *Sustainability*, vol. 16, no. 5, p. 1737, Feb. 2024, doi: [10.3390/su16051737](https://doi.org/10.3390/su16051737).
- [19] R. Yan, L. Hu, J. Li, and N. Lin, "Accident severity analysis of traffic accident hot spot areas in Changsha city considering built environment," *Sustainability*, vol. 16, no. 7, p. 3054, Apr. 2024, doi: [10.3390/su16073054](https://doi.org/10.3390/su16073054).
- [20] R. Ewing, D. Choi, F. Kiani, J. Kim, and S. Sabouri. (Sep. 2020). *Street Network Connectivity, Traffic Congestion, and Traffic Safety*. Utah Dep. Transp. Res. Innov. Div. [Online]. Available: www.udot.utah.gov/go/research%0Ahttps://rosap.ntl.bts.gov/view/dot/54922
- [21] J. Wang, L. Ji, S. Ma, X. Sun, and M. Wang, "Analysis of factors influencing the severity of Vehicle-to-Vehicle accidents considering the built environment: An interpretable machine learning model," *Sustainability*, vol. 15, no. 17, p. 12904, Aug. 2023, doi: [10.3390/su151712904](https://doi.org/10.3390/su151712904).
- [22] M. Wier, J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia, "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning," *Accident Anal. Prevention*, vol. 41, no. 1, pp. 137–145, Jan. 2009, doi: [10.1016/j.aap.2008.10.001](https://doi.org/10.1016/j.aap.2008.10.001).
- [23] V. Gitelman, D. Balasha, R. Carmel, L. Hendel, and F. Pesahov, "Characterization of pedestrian accidents and an examination of infrastructure measures to improve pedestrian safety in Israel," *Accident Anal. Prevention*, vol. 44, no. 1, pp. 63–73, Jan. 2012, doi: [10.1016/j.aap.2010.11.017](https://doi.org/10.1016/j.aap.2010.11.017).
- [24] M. Zhu, H. Li, N. N. Sze, and G. Ren, "Exploring the impacts of street layout on the frequency of pedestrian crashes: A micro-level study," *J. Saf. Res.*, vol. 81, pp. 91–100, Jun. 2022, doi: [10.1016/j.jsr.2022.01.009](https://doi.org/10.1016/j.jsr.2022.01.009).
- [25] X. Ye, S. Li, W. Gong, X. Li, X. Li, B. Dadashova, W. Li, J. Du, and D. Wu, "Street view imagery in traffic crash and road safety analysis: A review," *Appl. Spatial Anal. Policy*, vol. 18, no. 2, Jun. 2025, doi: [10.1007/s12061-025-09653-7](https://doi.org/10.1007/s12061-025-09653-7).
- [26] S. Hu, H. Xing, W. Luo, L. Wu, Y. Xu, W. Huang, W. Liu, and T. Li, "Uncovering the association between traffic crashes and street-level built-environment features using street view images," *Int. J. Geographical Inf. Sci.*, vol. 37, no. 11, pp. 2367–2391, Nov. 2023, doi: [10.1080/13658816.2023.2254362](https://doi.org/10.1080/13658816.2023.2254362).
- [27] T. Tanprasert, C. Siripanpornchana, N. Surasvadi, and S. Thajchayapong, "Recognizing traffic black spots from street view images using environment-aware image processing and neural network," *IEEE Access*, vol. 8, pp. 121469–121478, 2020, doi: [10.1109/ACCESS.2020.3006493](https://doi.org/10.1109/ACCESS.2020.3006493).
- [28] E. J. Kim, J. Won, and J. Kim, "Is Seoul walkable? Assessing a walkability score and examining its relationship with pedestrian satisfaction in Seoul, Korea," *Sustainability*, vol. 11, no. 24, p. 6915, Dec. 2019, doi: [10.3390/su11246915](https://doi.org/10.3390/su11246915).
- [29] P. D. Isola, J. N. Bogert, K. M. Chapple, S. Israr, T. L. Gillespie, and J. A. Weinberg, "Google street view assessment of environmental safety features at the scene of pedestrian automobile injury," *J. Trauma Acute Care Surgery*, vol. 87, no. 1, pp. 82–86, 2019, doi: [10.1097/ta.0000000000002338](https://doi.org/10.1097/ta.0000000000002338).
- [30] H. Yue, "Investigating the influence of streetscape environmental characteristics on pedestrian crashes at intersections using street view images and explainable machine learning," *Accident Anal. Prevention*, vol. 205, Sep. 2024, Art. no. 107693, doi: [10.1016/j.aap.2024.107693](https://doi.org/10.1016/j.aap.2024.107693).
- [31] J. R. Naderi, "Landscape design in clear zone: Effect of landscape variables on pedestrian health and driver safety," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1851, no. 1, pp. 119–130, Jan. 2003, doi: [10.3141/1851-12](https://doi.org/10.3141/1851-12).
- [32] X. Yu, J. Ma, Y. Tang, T. Yang, and F. Jiang, "Can we trust our eyes? Interpreting the misperception of road safety from street view images and deep learning," *Accident Anal. Prevention*, vol. 197, Mar. 2024, Art. no. 107455, doi: [10.1016/j.aap.2023.107455](https://doi.org/10.1016/j.aap.2023.107455).
- [33] W. E. Marshall, Y. Golombek, N. Coppola, and B. Janson. (2019). *The Unresolved Relationship Between Street Trees and Road Safety*. [Online]. Available: http://safety.fhwa.dot.gov/local_rural/training/fhwas07018/vegetationfv1108.pdf
- [34] W. E. Marshall P. E., N. Coppola, and Y. Golombek, "Urban clear zones, street trees, and road safety," *Res. Transp. Bus. Manage.*, vol. 29, pp. 136–143, Dec. 2018, doi: [10.1016/j.rtbm.2018.09.003](https://doi.org/10.1016/j.rtbm.2018.09.003).
- [35] M. Mori, S. Horino, S. Kitajima, M. Ueyama, T. Ebara, and T. Itani, "Ergonomics solution for crossing collisions based on field assessment of visual environment at urban intersections in Japan," *Appl. Ergonom.*, vol. 39, no. 6, pp. 697–709, Nov. 2008, doi: [10.1016/j.apergo.2007.12.007](https://doi.org/10.1016/j.apergo.2007.12.007).
- [36] D. Jones and M. K. Jha, "The effect of urban form on traffic accident incidence," *Recent Adv. Comput. Eng. Appl.*, pp. 212–222, Jan. 2010.
- [37] J. T. White, "Measuring urban design: Metrics for livable places," *J. Urban Design*, vol. 20, no. 2, pp. 291–292, Mar. 2015.
- [38] M. L. Cunningham, M. A. Regan, N. Elazar, and D. McTiernan, "A new and novel method for assessing visual clutter in the driving environment," in *Proc. Australas. Road Saf. Conf.*, 2017.
- [39] R. Rosenholtz, Y. Li, and L. Nakano, "Measuring visual clutter," *J. Vis.*, vol. 7, no. 2, pp. 1–22, Aug. 2007, doi: [10.1167/7.2.17](https://doi.org/10.1167/7.2.17).
- [40] L.-Y. Chang, "Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network," *Saf. Sci.*, vol. 43, no. 8, pp. 541–557, Oct. 2005, doi: [10.1016/j.ssci.2005.04.004](https://doi.org/10.1016/j.ssci.2005.04.004).
- [41] S. C. Joshua and N. J. Garber, "Estimating truck accident rate and involvements using linear and Poisson regression models," *Transp. Planning Technol.*, vol. 15, no. 1, pp. 41–58, Jun. 1990, doi: [10.1080/03081069008717439](https://doi.org/10.1080/03081069008717439).
- [42] S. Miaou, P. S. Hu, T. Wright, A. K. Rathi, and S. C. Davis, "Relationship between truck accidents and highway geometric design: A Poisson regression approach," *Transp. Res. Rec.*, vol. 6366, no. 1376, p. 37831, 1992.
- [43] E. Coruh, A. Bilgic, and A. Tortum, "Accident analysis with aggregated data: The random parameters negative binomial panel count data model," *Analytic Methods Accident Res.*, vol. 7, pp. 37–49, Jul. 2015, doi: [10.1016/j.amar.2015.07.001](https://doi.org/10.1016/j.amar.2015.07.001).
- [44] S. Heydari, L. Miranda-Moreno, and A. J. Hickford, "On the causal effect of proximity to school on pedestrian safety at signalized intersections: A heterogeneous endogenous econometric model," *Analytic Methods Accident Res.*, vol. 26, Jun. 2020, Art. no. 100115, doi: [10.1016/j.amar.2020.100115](https://doi.org/10.1016/j.amar.2020.100115).
- [45] H.-C. Park, B.-J. Park, and P. Y. Park, "A multiple membership multilevel negative binomial model for intersection crash analysis," *Analytic Methods Accident Res.*, vol. 35, Sep. 2022, Art. no. 100228, doi: [10.1016/j.amar.2022.100228](https://doi.org/10.1016/j.amar.2022.100228).
- [46] M. A. Khasawneh, A. A. Al-Omari, and B. Ganam, "Forecasting traffic accidents in Jordan using regression techniques," *Jordan J. Civil Eng.*, vol. 12, no. 4, pp. 570–579, 2018.
- [47] A. Mahmoudabadi, "Comparison of weighted and simple linear regression and artificial neural network models in freeway accidents prediction," in *Proc. 2nd Int. Conf. Comput. Netw. Technol.*, 2010, pp. 392–396, doi: [10.1109/iccnt.2010.73](https://doi.org/10.1109/iccnt.2010.73).
- [48] D. C. Eckley and K. M. Curtin, "Evaluating the spatiotemporal clustering of traffic incidents," *Comput., Environ. Urban Syst.*, vol. 37, pp. 70–81, Jan. 2013, doi: [10.1016/j.compenvurbsys.2012.06.004](https://doi.org/10.1016/j.compenvurbsys.2012.06.004).
- [49] M. B. Islam and S. Hernandez, "Fatality rates for crashes involving heavy vehicles on highways: A random parameter tobit regression approach," *J. Transp. Saf. Secur.*, vol. 8, no. 3, pp. 247–265, Jul. 2016, doi: [10.1080/19439962.2015.1027071](https://doi.org/10.1080/19439962.2015.1027071).

- [50] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Anal. Prevention*, vol. 45, pp. 628–633, Mar. 2012, doi: [10.1016/j.aap.2011.09.015](https://doi.org/10.1016/j.aap.2011.09.015).
- [51] A. Khoda Bakhshi and M. M. Ahmed, "Real-time crash prediction for a long low-traffic volume corridor using corrected-impurity importance and semi-parametric generalized additive model," *J. Transp. Saf. Secur.*, vol. 14, no. 7, pp. 1165–1200, Jul. 2022, doi: [10.1080/19439962.2021.1898069](https://doi.org/10.1080/19439962.2021.1898069).
- [52] Y. Xie and Y. Zhang, "Crash frequency analysis with generalized additive models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2061, no. 1, pp. 39–45, Jan. 2008, doi: [10.3141/2061-05](https://doi.org/10.3141/2061-05).
- [53] Y. Zhang, Y. Xie, and L. Li, "Crash frequency analysis of different types of urban roadway segments using generalized additive model," *J. Saf. Res.*, vol. 43, no. 2, pp. 107–114, Apr. 2012, doi: [10.1016/j.jsr.2012.01.003](https://doi.org/10.1016/j.jsr.2012.01.003).
- [54] S. C. Barry and T. J. O'Neill, "The analysis of group truncated binary data with random effects: Injury severity in motor vehicle accidents," *Biometrics*, vol. 56, no. 2, pp. 443–450, Jun. 2000, doi: [10.1111/j.0006-341x.2000.00443.x](https://doi.org/10.1111/j.0006-341x.2000.00443.x).
- [55] T. J. O'Neill and S. C. Barry, "Truncated Logistic Regression," *Int. Biometric Soc. Stable*, vol. 51, no. 2, pp. 533–541, 1995. [Online]. Available: <https://www.jstor.org/stable/2532941>
- [56] Y. Qi, R. Srinivasan, H. Teng, and R. Baker, "Analysis of the frequency and severity of rear-end crashes in work zones," *Traffic Injury Prevention*, vol. 14, no. 1, pp. 61–72, Jan. 2013, doi: [10.1080/15389588.2012.675109](https://doi.org/10.1080/15389588.2012.675109).
- [57] X. Xu, S. C. Wong, and K. Choi, "A two-stage bivariate logistic-tobit model for the safety analysis of signalized intersections," *Analytic Methods Accident Res.*, vols. 3–4, pp. 1–10, Oct. 2014, doi: [10.1016/j.amar.2014.08.001](https://doi.org/10.1016/j.amar.2014.08.001).
- [58] R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, "Comparison of machine learning algorithms for predicting traffic accident severity," in *Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT)*, Jordan, Apr. 2019, pp. 272–276, doi: [10.1109/JEEIT.2019.8717393](https://doi.org/10.1109/JEEIT.2019.8717393).
- [59] C. Liu, Y. Wang, W. Li, L. Tao, S. Hu, and M. Hao, "An urban built environment analysis approach for street view images based on graph convolutional neural networks," *Appl. Sci.*, vol. 14, no. 5, p. 2108, Mar. 2024, doi: [10.3390/app14052108](https://doi.org/10.3390/app14052108).
- [60] S. Krishnaveni and M. Hemalatha, "A perspective analysis of traffic accident using data mining techniques," *Int. J. Comput. Appl.*, vol. 23, no. 7, pp. 40–48, Jun. 2011, doi: [10.5120/2896-3788](https://doi.org/10.5120/2896-3788).
- [61] A. Iranitalab and A. Khatkhat, "Comparison of four statistical and machine learning methods for crash severity prediction," *Accident Anal. Prevention*, vol. 108, pp. 27–36, Nov. 2017, doi: [10.1016/j.aap.2017.08.008](https://doi.org/10.1016/j.aap.2017.08.008).
- [62] F. R. Moghaddam, S. H. Afandizadeh, and M. Ziyadi, "Prediction of accident severity using artificial neural networks," *Int. J. Civ. Eng.*, vol. 9, no. 1, pp. 41–48, 2011.
- [63] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang, and Q. Shangguan, "Dynamic prediction of traffic incident duration on urban expressways: A deep learning approach based on LSTM and MLP," *J. Intell. Connected Vehicles*, vol. 4, no. 2, pp. 80–91, 2021, doi: [10.1108/JICV-03-2021-0004](https://doi.org/10.1108/JICV-03-2021-0004).
- [64] J. Li, X. Wang, X. Yang, Q. Zhang, and H. Pan, "Analyzing free-way safety influencing factors using the CatBoost model and interpretable machine-learning framework, SHAP," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2678, no. 7, pp. 563–574, Jul. 2024, doi: [10.1177/03611981231208903](https://doi.org/10.1177/03611981231208903).
- [65] S. Dong, A. Khatkhat, I. Ullah, J. Zhou, and A. Hussain, "Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley additive exPlanations," *Int. J. Environ. Res. Public Health*, vol. 19, no. 5, p. 2925, Mar. 2022, doi: [10.3390/ijerph19052925](https://doi.org/10.3390/ijerph19052925).
- [66] A. Rafe and P. A. Singleton, "Exploring the determinants of pedestrian crash severity using an AutoML approach," in *Proc. Int. Conf. Transp. Develop.*, Jun. 2024, pp. 442–455, doi: [10.1061/9780784485514.039](https://doi.org/10.1061/9780784485514.039).
- [67] X. Ma, Z. Huo, J. Lu, and Y. D. Wong, "Deep forest with SHAPley additive explanations on detailed risky driving behavior data for freeway crash risk prediction," *Eng. Appl. Artif. Intell.*, vol. 141, Feb. 2025, Art. no. 109787, doi: [10.1016/j.engappai.2024.109787](https://doi.org/10.1016/j.engappai.2024.109787).
- [68] I. Chang, H. Park, E. Hong, J. Lee, and N. Kwon, "Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP," *Accident Anal. Prevention*, vol. 166, Mar. 2022, Art. no. 106545, doi: [10.1016/j.aap.2021.106545](https://doi.org/10.1016/j.aap.2021.106545).
- [69] Y. Ye, C. Zhong, and E. Suel, "Unpacking the perceived cycling safety of road environment using street view imagery and cycle accident data," *Accident Anal. Prevention*, vol. 205, Sep. 2024, Art. no. 107677, doi: [10.1016/j.aap.2024.107677](https://doi.org/10.1016/j.aap.2024.107677).
- [70] M. W. Adler and R. Ahrend, "Traffic safety in Korea: Understanding the vulnerability of elderly pedestrians," OECD Reg. Dev., OECD Publishing Paris, Working Paper 2017/03, Sep. 2017. [Online]. Available: <http://dx.doi.org/10.1787/22fca93d-en>
- [71] K. Kim, S. Yang, and J. Lee, "A study on the Feng Shui characteristics of 'pedestrian accident Hotspots' in Seoul (focusing on traffic accident analysis system data)," *J. Seoul Stud.*, vol. 95, pp. 33–82, May 2024, doi: [10.17647/jss.2024.05.95.33](https://doi.org/10.17647/jss.2024.05.95.33).
- [72] I. Ashraf, S. Hur, M. Shafiq, and Y. Park, "Catastrophic factors involved in road accidents: Underlying causes and descriptive analysis," *PLoS ONE*, vol. 14, no. 10, Oct. 2019, Art. no. e0223473, doi: [10.1371/journal.pone.0223473](https://doi.org/10.1371/journal.pone.0223473).
- [73] J. Lee and T.-H.-T. Gim, "A spatial econometrics perspective on the characteristics of urban traffic accidents: Focusing on elderly drivers' accidents in Seoul, South Korea," *Int. J. Injury Control Saf. Promotion*, vol. 27, no. 4, pp. 520–527, Oct. 2020, doi: [10.1080/17457300.2020.1817945](https://doi.org/10.1080/17457300.2020.1817945).
- [74] Korean Legislation Research Institute. (2016). *Road Traffic Act (Act, no. 13458, Article 32)*. Accessed: Dec. 2023. [Online]. Available: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=35814&lang=ENG
- [75] L. NICE Genie Data Co. (2025). *NICEbizmap*. [Online]. Available: <https://m.nicebizmap.co.kr/>
- [76] Kakao. (2022). *Kakao Maps*. Accessed: Oct. 2023. [Online]. Available: <https://map.kakao>
- [77] S. Kim, J. Choi, and Y. Kang, "Measurement and visualization of perceived fear of crime by street level using street view images and deep learning technology," *J. Korean Cartographic Assoc.*, vol. 24, no. 1, pp. 71–84, Apr. 2024, doi: [10.16879/jkca.2024.24.1.071](https://doi.org/10.16879/jkca.2024.24.1.071).
- [78] T. Lee. (2021). *How to Determine the Most Suitable Map API_Web Development Targeted for Korean Market*. [Online]. Available: https://www.theseus.fi/bitstream/handle/10024/496518/HowtodeterminethemostsuitablemapAPI_webdevelopmenttargetedforKoreanmarket_TaeheeLee.pdf?sequence=2
- [79] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ADE20K dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5122–5130, doi: [10.1109/CVPR.2017.544](https://doi.org/10.1109/CVPR.2017.544).
- [80] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 173–190, doi: [10.1007/978-3-030-58539-6_11](https://doi.org/10.1007/978-3-030-58539-6_11).
- [81] J. Gehl, "Life between buildings," 6th ed., Island Press, Washington, DC, USA, Tech. Rep., 2011.
- [82] G. Jocher, A. Charurasia, and J. Qiu. (Dec. 2023). *Ultralytics YOLOv8, Version 8.0.0*. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [83] N. H. Rahman, R. Rainis, S. H. Noor, and S. M. S. Mohamad, "The buffering analysis to identify common geographical factors within the vicinity of severe injury related to motor vehicle crash in Malaysia," *World J. Emergency Med.*, vol. 7, no. 4, p. 278, 2016, doi: [10.5847/wjem.j.1920-8642.2016.04.007](https://doi.org/10.5847/wjem.j.1920-8642.2016.04.007).
- [84] S. S. Pulugurtha and P. R. Kusam, "Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2291, no. 1, pp. 53–60, Jan. 2012, doi: [10.3141/2291-07](https://doi.org/10.3141/2291-07).
- [85] R. Jia, A. Khadka, and I. Kim, "Traffic crash analysis with point-of-interest spatial clustering," *Accident Anal. Prevention*, vol. 121, pp. 223–230, Dec. 2018, doi: [10.1016/j.aap.2018.09.018](https://doi.org/10.1016/j.aap.2018.09.018).
- [86] J. Edquist, C. M. Rudin-Brown, and M. G. Lenné, "The effects of on-street parking and road environment visual complexity on travel speed and reaction time," *Accident Anal. Prevention*, vol. 45, pp. 759–765, Mar. 2012, doi: [10.1016/j.aap.2011.10.001](https://doi.org/10.1016/j.aap.2011.10.001).
- [87] S. Biswas, S. Chandra, and I. Ghosh, "Effects of on-street parking in urban context: A critical review," *Transp. Developing Economies*, vol. 3, no. 1, pp. 1–14, Apr. 2017, doi: [10.1007/s40890-017-0040-2](https://doi.org/10.1007/s40890-017-0040-2).
- [88] T. Lian, B. P. Y. Loo, and Z. Fan, "Advances in estimating pedestrian measures through artificial intelligence: From data sources, computer vision, video analytics to the prediction of crash frequency," *Comput., Environ. Urban Syst.*, vol. 107, Jan. 2024, Art. no. 102057, doi: [10.1016/j.compenvurbsys.2023.102057](https://doi.org/10.1016/j.compenvurbsys.2023.102057).

- [89] L. Brühwiler, C. Fu, H. Huang, L. Longhi, and R. Weibel, "Predicting individuals' car accident risk by trajectory, driving events, and geographical context," *Comput., Environ. Urban Syst.*, vol. 93, Apr. 2022, Art. no. 101760, doi: [10.1016/j.compenvurbysys.2022.101760](https://doi.org/10.1016/j.compenvurbysys.2022.101760).
- [90] Vworld. (2023). *VWorld Data Center*. Accessed: Oct. 2023. [Online]. Available: <http://map.vworld.kr/map/maps.do>
- [91] C. M. Craig, N. L. Morris, R. Van Houten, and D. Mayou, "Pedestrian safety and driver yielding near public transit stops," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2673, no. 1, pp. 514–523, Jan. 2019, doi: [10.1177/0361198118822313](https://doi.org/10.1177/0361198118822313).
- [92] M. Bíl, R. Andrášik, J. Sedoník, and V. Cícha, "ROCA—An ArcGIS toolbox for road alignment identification and horizontal curve radii computation," *PLoS ONE*, vol. 13, no. 12, Dec. 2018, Art. no. e0208407, doi: [10.1371/journal.pone.0208407](https://doi.org/10.1371/journal.pone.0208407).
- [93] C. Ariën, K. Brijs, G. Vanroelen, W. Ceulemans, E. M. M. Jongen, S. Daniels, T. Brijs, and G. Wets, "The effect of pavement markings on driving behaviour in curves: A simulator study," *Ergonomics*, vol. 60, no. 5, pp. 701–713, May 2017, doi: [10.1080/00140139.2016.1200749](https://doi.org/10.1080/00140139.2016.1200749).
- [94] M. Bíl, R. Andrášik, and J. Sedoník, "Which curves are dangerous? A network-wide analysis of traffic crash and infrastructure data," *Transp. Res. A, Policy Pract.*, vol. 120, pp. 252–260, Feb. 2019, doi: [10.1016/j.tra.2019.01.001](https://doi.org/10.1016/j.tra.2019.01.001).
- [95] H. H. Awan, A. Pirdavani, A. Houben, S. Westhof, M. Adnan, and T. Brijs, "Impact of perceptual countermeasures on driving behavior at curves using driving simulator," *Traffic Injury Prevention*, vol. 20, no. 1, pp. 93–99, Jan. 2019, doi: [10.1080/15389588.2018.1532568](https://doi.org/10.1080/15389588.2018.1532568).
- [96] M. H. Islam, L. Teik Hua, H. Hamid, and A. Azarkerdar, "Relationship of accident rates and road geometric design," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 357, no. 1, Nov. 2019, Art. no. 012040, doi: [10.1088/1755-1315/357/1/012040](https://doi.org/10.1088/1755-1315/357/1/012040).
- [97] R. Ewing and R. Cervero, "Travel and the built environment," *J. Amer. Planning Assoc.*, vol. 76, no. 3, pp. 265–294, Jun. 2010, doi: [10.1080/01944361003766766](https://doi.org/10.1080/01944361003766766).
- [98] K. Koc and A. P. Gurgun, "Scenario-based automated data preprocessing to predict severity of construction accidents," *Autom. Construct.*, vol. 140, Aug. 2022, Art. no. 104351, doi: [10.1016/j.autcon.2022.104351](https://doi.org/10.1016/j.autcon.2022.104351).
- [99] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 2951–2959.
- [100] L. S. Shapley, "Stochastic games," *Proc. Nat. Acad. Sci. USA*, vol. 39, no. 10, pp. 1095–1100, Oct. 1953, doi: [10.1073/pnas.39.10.1095](https://doi.org/10.1073/pnas.39.10.1095).
- [101] S. M. Lundberg and S.-I. Lee, "Consistent feature attribution for tree ensembles," 2017, *arXiv:1706.06060*.
- [102] J. Stiles, Y. Li, and H. J. Miller, "How does street space influence crash frequency? An analysis using segmented street view imagery," *Environ. Planning B, Urban Anal. City Sci.*, vol. 49, no. 9, pp. 2467–2483, Nov. 2022, doi: [10.1177/23998083221090962](https://doi.org/10.1177/23998083221090962).
- [103] M. Zhu, N. N. Sze, and S. Newnam, "Effect of urban street trees on pedestrian safety: A micro-level pedestrian casualty model using multi-variate Bayesian spatial approach," *Accident Anal. Prevention*, vol. 176, Oct. 2022, Art. no. 106818, doi: [10.1016/j.aap.2022.106818](https://doi.org/10.1016/j.aap.2022.106818).
- [104] H. Tapiro, T. Oron-Gilad, and Y. Parmet, "Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior," *J. Saf. Res.*, vol. 72, pp. 101–109, Feb. 2020, doi: [10.1016/j.jsr.2019.12.003](https://doi.org/10.1016/j.jsr.2019.12.003).
- [105] A. Sołowczuk, "Efficient improvement of the visibility of pedestrians on junctions in Tempo-30 zones," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 603, no. 2, 2019, Art. no. 022042, doi: [10.1088/1757-899x/603/2/022042](https://doi.org/10.1088/1757-899x/603/2/022042).
- [106] S. Bendak, A. M. Alnaqbi, M. Y. Alzarooni, S. M. Aljanaahi, and S. J. Alsuwaidi, "Factors affecting pedestrian behaviors at signalized crosswalks: An empirical study," *J. Saf. Res.*, vol. 76, pp. 269–275, Feb. 2021, doi: [10.1016/j.jsr.2020.12.019](https://doi.org/10.1016/j.jsr.2020.12.019).
- [107] D. A. Quistberg, E. J. Howard, B. E. Ebel, A. V. Moudon, B. E. Saelens, P. M. Hurvitz, J. E. Curtin, and F. P. Rivara, "Multilevel models for evaluating the risk of pedestrian-motor vehicle collisions at intersections and mid-blocks," *Accident Anal. Prevention*, vol. 84, pp. 99–111, Nov. 2015, doi: [10.1016/j.aap.2015.08.013](https://doi.org/10.1016/j.aap.2015.08.013).
- [108] S. Rossetti, "In field assessment of safety, security, comfort and accessibility of bus stops: A planning perspective," *Eur. Transport/Trasporti Europei*, vol. 80, no. 6, pp. 1–17, Dec. 2020, doi: [10.48295/et.2020.80.8](https://doi.org/10.48295/et.2020.80.8).
- [109] Y. Hahm, H. Yoon, and Y. Choi, "The effect of built environments on the walking and shopping behaviors of pedestrians; a study with GPS experiment in sinchon retail district in Seoul, South Korea," *Cities*, vol. 89, pp. 1–13, Jun. 2019, doi: [10.1016/j.cities.2019.01.020](https://doi.org/10.1016/j.cities.2019.01.020).
- [110] B. Pessaro, M. Catalá, Z. Wang, and M. Spicer. (Oct. 2017). *Impact of Transit Stop Location on Pedestrian Safety*. [Online]. Available: [file:///C:/Users/N01458748/Downloads/dot_34939_DS1\(1\).pdf](file:///C:/Users/N01458748/Downloads/dot_34939_DS1(1).pdf)



CARYL ANNE M. BARQUILLA received the B.S. degree in forestry from the University of the Philippines and the M.S. degree in urban development from Chung-Ang University, Seoul, South Korea, where she is currently pursuing the Ph.D. degree in urban engineering with the Department of Civil Engineering, majoring in urban design and planning. Before pursuing her graduate studies, she worked as a Forester under the Department of Environment and Natural Resources, Philippines. She is a licensed Forester in Philippines. She has published in high-impact journals, such as *Sustainable Cities and Society* and *Land* and presented her work at various international conferences. Her research focuses on the intersection of urban environments and public health, drawing on expertise in machine learning, geographic information systems, greenspace planning, pedestrian safety, and urban air quality prediction. She is a member of Korea Planning Association and has received multiple Best Paper Awards, including the Young Urban Planners Network Award at the International Conference of Asia-Pacific Planning Societies. She has also been involved in national research projects funded by the National Research Foundation of Korea related to green infrastructure and environmental health modeling.



JEONGWOO LEE received the Ph.D. degree in urban planning from the Sol Price School of Public Policy, University of Southern California, in 2013. She is currently an Associate Professor with the Department of Urban Design and Studies, Chung-Ang University, Seoul, South Korea. She teaches site planning and urban design theory at the undergraduate level and spatial quantitative research methods and public health and city planning at the graduate level at Chung-Ang University. She has led and contributed to numerous nationally funded research projects supported by the National Research Foundation of Korea, Korea Land and Housing Corporation, and Gyeonggi Housing and Urban Development Corporation. Her work centers on developing healthy city frameworks, walkability indicators, age-friendly infrastructure, and spatial modeling of environmental health impacts in urban areas, with an emphasis on evidence-based planning tools, policy applications, and the integration of urban design with public health outcomes. Her research focuses on smart urban design, public health and urban planning, green infrastructure, integrated urban models using big data, and pedestrian environments. She serves as a reviewer for various international and domestic journals and is a board member of several academic organizations, including the Urban Design Institute of Korea, Korea Planning Association, Academy of Urban Health, Korean Regional Development Association, and Korea Landscape Council. She also contributes to public-sector advisory committees under Seoul Metropolitan Government and national agencies.

...