Sustainable Cities and Society 101 (2024) 105168

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Unveiling pedestrian injury risk factors through integration of urban contexts using multimodal deep learning

Jeongyeop Baek, Lisa Lim *

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, South Korea

ARTICLE INFO

Keywords:
Pedestrian safety
Socio-spatial information
Street-level image
Urban planning and policy
Multimodal deep learning
Explainable AI

ABSTRACT

This study aimed to identify contributing risk factors for pedestrian injury by integrating socio-spatial and street-level contexts through multimodal deep learning to overcome the limitations of existing studies that only consider one type of data. To investigate how the two contexts assist in describing pedestrian injury risk, six multimodal deep learning models were established by varying the ratio integrating the two contexts. The developed model with the highest performance was interpreted by using two XAI methods: SHAP for socio-spatial context and Grad-CAM for street-level context. The results indicated that the street-level context mainly contributes to the pedestrian injury risk level, assisted by the socio-spatial context, which cannot be captured at the street-level. The three main contributing risk factors were identified through model interpretation: the fragmented sky view due to the locations of high-rise buildings, the placement of crosswalks in areas adjacent to public transits, and interregional sociodemographic disparities. This study provides insight into the use of integrating two different urban contexts to identify pedestrian injury risk factors, which are expected to support improvement strategies that enhance public health.

1. Introduction

Walking is a sustainable transportation mode that benefits the physical and mental health of pedestrians, and public health is of paramount importance in establishing sustainable cities and societies (Kelly et al., 2017). Despite the significant benefits of walking, there are risks associated with unintentional pedestrian injuries during walking, which can pose serious health threats to pedestrians (Oxley et al., 2018). Pedestrian injuries represent a global public health issue (WHO, 2018), and they pose a significant risk to both individuals and communities, particularly within urban settings (Hang et al., 2003). Despite pedestrian fatalities resulting from unintentional injuries accounting for a relatively small proportion (<15 %) of all fatalities, this percentage has consistently risen in recent years (J. Hu et al., 2023). This trend is especially concerning for elderly individuals, as unintentional injuries such as falls have emerged as a leading cause of death, with escalating mortality rates (Donald & Bulpitt, 1999; Fried et al., 2007).

Unintentional pedestrian injuries can lead to long-term disabilities, require extensive hospital treatments and frequent emergency department visits, and generate substantial direct medical costs (G. Hu et al., 2019). Consequently, these injuries exert considerable strain on healthcare systems and hinder overall economic development (Sun et al., 2021). To establish sustainable cities and achieve public health objectives, it is imperative to create streets that ensure safe walking conditions by identifying and refining urban risk factors that contribute to pedestrian injuries.

The influence of the urban built environment on pedestrian injury risk is currently primarily understood within two main contexts in previous studies, namely, socio-spatial and street-level characteristics. The socio-spatial context includes several risk factors, such as building and land characteristics (Bustos et al., 2021; Mentis & Papadopulos, 2021), road characteristics (Osama & Sayed, 2017; Wier et al., 2009), pedestrian facilities (Chan et al., 2021; Gkekas et al., 2020; Mooney et al., 2016), lighting conditions (J.-K. Kim et al., 2008), public transit (J. Kim et al., 2022; Lakhotia et al., 2020), and demographics (LaScala et al., 2000; Oxley et al., 2018). Meanwhile, the street-level context is mainly focused on perceptual risk factors, such as visual distraction perceived by pedestrians (Schwebel et al., 2012a; Tapiro et al., 2020; Tobis et al., 1985; H. Wang et al., 2022).

Since these two main contexts, both impact pedestrians' experiences and behaviors (Ferreira et al., 2022), there is a need to understand the relationship between pedestrian injury risk and the two main contexts

* Corresponding author.
E-mail address: lisalim@kaist.ac.kr (L. Lim).

https://doi.org/10.1016/j.scs.2023.105168
Received 15 August 2023; Received in revised form 4 November 2023; Accepted 28 December 2023
Available online 29 December 2023

together. However, existing studies have primarily focused on using a single source of socio-spatial or street-level information to identify its relationship with pedestrian injury risk (Suel et al., 2021). Therefore, the specific contributions of each urban context to pedestrian injury risk have not been thoroughly investigated (Ferreira et al., 2022). Thus, there is a need to clarify how the two urban contexts contribute to pedestrian injury risk individually within a comprehensive perspective through their integration.

To fill this gap, an advanced method that integrates diverse data sources and forms is needed. One of the data fusion methods, namely, multiview-based fusion, can be employed to jointly utilize data measured from different perspectives (Kumar & Daumé, 2011; Yi et al., 2018). Multiview-based methods have already been proven effective in various domains, such as urban crowd flow prediction (J. Zhang et al., 2017), urban air quality prediction (Zheng et al., 2015), and measuring environmental deprivation (Suel et al., 2021). Furthermore, multimodal deep learning (DL) methods that leverage different data forms have been shown to enable the joint utilization of structured and unstructured data, effectively representing complex urban big data (J. Liu et al., 2020; Srivastava et al., 2019).

Therefore, using a multimodal DL method, this study integrates socio-spatial and street-level contexts to predict pedestrian injury risk levels in a comprehensive way. This study aims to identify contributing risk factors in both socio-spatial and street-level contexts and to further explain how the identified risk factors are associated with pedestrian injury risk. The findings of the study are expected to support improvement strategies that enhance public health, which is a key component of sustainable cities.

2. Related works and contributions

In addressing the issue of pedestrian injuries, it is imperative to consider multiple scales of the related contexts, i.e., the demographic characteristics of the pedestrians (e.g., gender, age, and education level), the broader urban context (e.g., surrounding built environments and programs), and the street-level experiences (e.g., adjacent built environments, sensory perceptions and physical interactions).

The urban environment, comprising elements such as buildings, population, and geographical areas, contributes to the creation of this broad urban context. However, pedestrians tend to focus on specific details, such as street views and the characteristics of nearby buildings, which lead to their own sensory experiences (Kim & Mateo-Babiano, 2018). While the overall composition of the street can be influenced by the broader socio-spatial context, it is essential to emphasize the significance of considering what individuals perceive when walking (Cambra & Moura, 2020). Understanding how street-level experience complements the socio-spatial context is crucial for a comprehensive understanding of pedestrian injury risk factors.

This section describes previous studies that have illustrated the relationships between pedestrian injuries and various risk factors on two levels; the first is the social-spatial context of the urban environment, including demographic characteristics, and the second is street-level perceptual experiences of pedestrians, which are followed by the last section describing the need to integrate these two contexts.

2.1. Socio-spatial risk factors

In the context of socio-spatial information, urban risk factors are subdivided into six categories: building and land characteristics, road characteristics, pedestrian facilities, lighting conditions, public transit, and demographics.

Previous studies have consistently highlighted the impact of building and land characteristics on pedestrian injuries in urban environments (Bustos et al., 2021; Clifton et al., 2009; Harvey & Aultman-Hall, 2015; Mentis & Papadopulos, 2021; Wier et al., 2009; Zajac & Ivan, 2003). Several studies have focused on the impacts of the attributes of buildings, such as the number of buildings (Clifton et al., 2009), building height (Bustos et al., 2021; Clifton et al., 2009), density of buildings (Clifton et al., 2009; Harvey & Aultman-Hall, 2015; Zajac & Ivan, 2003), and building deterioration (Mentis & Papadopulos, 2021). On the other hand, the surroundings of buildings, such as land use patterns (Wier et al., 2009; Zajac & Ivan, 2003) and sidewalk conditions (Mentis & Papadopulos, 2021), have also been highlighted.

Previous studies have also identified road characteristics as significant factors influencing pedestrian injuries (Osama & Sayed, 2017; Wier et al., 2009). Road width has been found to affect pedestrian safety, with narrower roads being associated with higher rates of pedestrian injuries, potentially due to increased interactions among pedestrians, cyclists, and various traffic vehicles (Wier et al., 2009). Wier et al. (2009) found that the number of road segments and road length contribute to higher risks of traffic accidents in congested urban areas. Additionally, Osama and Sayed (2017) identified road types, such as arterial or collector roads, as one of the influential factors, indicating that different types of roads may contribute to variations in unintentional injury risk for pedestrians.

The design of pedestrian facilities, such as sidewalks and crosswalks, has an impact on pedestrian injury risk (Beitel et al., 2018; Chan et al., 2021; Clifton et al., 2009; Gkekas et al., 2020; Mooney et al., 2016; Wu et al., 2020). Sidewalk-level features play a crucial role in shaping the pedestrian-friendliness of urban roads (Clifton et al., 2007). The width of sidewalks and the presence of obstacles have also received significant attention (Beitel et al., 2018; Chan et al., 2021; Gkekas et al., 2020; Wu et al., 2020). Furthermore, Mooney et al. (2016) found that the presence of traffic islands and crosswalk infrastructures is associated with elevated numbers of pedestrian injuries in New York.

Previous studies have also highlighted the importance of illuminance levels on streets and the corresponding effects on pedestrian safety (J.-K. Kim et al., 2008; Suk & Walter, 2019). Suk and Walter (2019) reported that higher roadway illuminance levels are linked to reduced traffic accidents. Adequate illumination owing to street lighting is crucial for enhancing pedestrian safety, especially in areas characterized by darkness or low visibility due to a lack of street-lighting furniture (J.-K. Kim et al., 2008).

According to previous research, the location and accessibility of public transit facilities play a crucial role in influencing pedestrian safety (J. Kim et al., 2022; Lakhotia et al., 2020; Mooney et al., 2016; Osama & Sayed, 2017). For example, subway stations increase the population flow in nearby areas, leading to an increase in pedestrian injuries due to obstacles such as stairs and escalators (J. Kim et al., 2022; Osama & Sayed, 2017). Additionally, the number of bus stops in an urban area is directly related to the number of pedestrian injuries in that area (Lakhotia et al., 2020; Mooney et al., 2016).

Previous studies have also found that demographic factors, such as age, population density, and socioeconomic variables, significantly influence the pedestrian injury rate and severity (J. Kim & Lee, 2018; LaScala et al., 2000; Niebuhr et al., 2016; Oxley et al., 2018; Wier et al., 2009). Several studies have identified increasing pedestrian age as a notable factor associated with an increased probability of fatal pedestrian injury (J. Kim & Lee, 2018; Niebuhr et al., 2016; Oxley et al., 2018; Wier et al., 2009). Furthermore, regional population density such as the number of houses, resident population (Wier et al., 2009), population density, age composition, unemployment rate, gender, and education (LaScala et al., 2000), have been identified as risk factors for pedestrian injuries. Table 1 summarizes the reported socio-spatial risk factors influencing pedestrian injury in previous studies.

2.2. Street-level perceptual risk factors

The importance of visual perception for pedestrian safety has also been investigated for several decades through a variety of methods (Barton, 2006; Demetre et al., 1992; Lee et al., 1984; Morrongiello et al., 2015; Schwebel et al., 2012b; Van Schagen & Rothengatter, 1997).

Table 1
Socio-spatial risk factors for unintentional pedestrian injury.

Category	Subcategory	Reference
Building and land characteristics	Building height	Bustos et al., 2021; Clifton et al., 2009
	Density of buildings	Clifton et al., 2009; Harvey & Aultman-Hall, 2015; Zajac & Ivan, 2003
	Building deterioration	Mentis & Papadopulos, 2021
	Land use	Zajac & Ivan, 2003
Road characteristics	Road density	Wier et al., 2009
	Road width	
	Road length	
	Road type	Osama & Sayed, 2017
Pedestrian facilities	Sidewalks	Beitel et al., 2018; Chan et al., 2021; Clifton et al., 2009; Gkekas et al., 2020; Wu et al., 2020
	Crosswalks	Mooney et al., 2016
Lighting conditions	Street lighting	Kim et al., 2008; Suk & Walter, 2019
Public transit	Subway station	Kim et al., 2022; Osama & Sayed, 2017
	Bus stops	Lakhotia et al., 2020; Mooney et al., 2016
Demographics	Number of houses	Wier et al., 2009
	Total population	Wier et al., 2009
	Age-related population	Kim & Lee, 2018; LaScala et al., 2000; Niebuhr et al., 2016; Oxley et al., 2018
	Unemployment rate	LaScala et al., 2000
	Gender	LaScala et al., 2000
	Education level	LaScala et al., 2000

When pedestrians encounter various components of the built environment, they are visually stimulated by the components, process their cognitive loads, and determine their own safety perception (Davis et al., 2019). Previous studies have emphasized the significant role of visual distraction as a crucial factor in unintentional pedestrian injury risk (Schwebel et al., 2012a; Tapiro et al., 2020; Tobis et al., 1985; H. Wang et al., 2022; Wu et al., 2020). H. Wang et al. (2022) found that pedestrians exhibit higher levels of oxyhemoglobin change in cortices related to visual processing and executive function when distracted by a higher cognitive load. Moreover, participants' responses to traffic lights are slower and result in higher activation in the prefrontal cortex when visually distracted.

Previous studies have indicated that the visual distraction of pedestrians can be caused by several objects observed at the street-level (Crundall et al., 2006; Decker et al., 2015; Oviedo-Trespalacios et al., 2019; Perez & Bertola, 2011). Crundall et al. (2006) demonstrated that street-level advertisements draw the visual attention of pedestrians, resulting in their longer fixations while walking. Additionally, Decker et al. (2015) found that the presence of numerous billboards at the street-level can cause visual distractions for pedestrians and drivers. Similarly, Oviedo-Trespalacios et al. (2019) indicated that the behavior of pedestrians and drivers is affected by the number of signs observed at the street-level, as well as their shape, content, and position. Moreover, Perez & Bertola (2011) highlighted that highly cluttered visual objects can increase risky behaviors, such as fixation for longer periods, due to visual distraction.

Visual distraction has a more pronounced impact on certain demographics than others (Schwebel et al., 2012a; Tapiro et al., 2020; Tobis et al., 1985; Wu et al., 2020). Wu et al. (2020) utilized real street images to accurately investigate perceived safety issues among older pedestrians. Similarly, Tobis et al. (1985) observed that older adults, especially among those with fall experiences, rely more on visual cues for posture maintenance. While visual distraction poses a critical risk for older adults, it also affects injury risk for children. Schwebel et al. (2012a) emphasized that the safety of children could be compromised if they become distracted while crossing streets. Additionally, Tapiro et al. (2020) demonstrated that a busy road environment poses significant hazards for child pedestrians, highlighting the impact of distractions on pedestrian crossing risk.

2.3. Integration of socio-spatial and street-level risk factors

Since urban data are collected from various sources, they pose a significant challenge in regard to data fusion due to their multisource and heterogeneous nature (J. Liu et al., 2020). Despite the complexities associated with urban data fusion, research on the integration of multimodal data is urgent and crucial owing to its strength in learning feature representation (J. Liu et al., 2020). In response to this need, several existing methods have been proposed and demonstrated to be effective in integrating urban data (Du et al., 2018; Shahrbabaki et al., 2018; Zhu et al., 2018). Furthermore, the multimodal approach is useful in many research domains, such as urban planning, urban environment, and urban public safety and security (Zheng et al., 2014).

A multimodal DL-based approach to urban data can also help in understanding the contributors to pedestrian injury risk. Previous studies have provided significant insights by identifying urban risk factors from socio-spatial and street-level perspectives. These two different urban contexts contain multimodal information captured from different viewpoints of the built environment (Suel et al., 2021).

Different contexts of data have both strengths and weaknesses (Suel et al., 2021). Socio-spatial information represents social and spatial data related to cities and their surrounding areas from a macroscale perspective. Such data include information regarding regional demographics, social structure, transportation, infrastructure, and environment. However, there are limitations in representing how these built environment components are perceived from the viewpoint of pedestrians. To address this issue, image sources such as Google Street View (GSV) are widely utilized to quantify street-level visual characteristics (Hanson et al., 2013; Li et al., 2022; Mooney et al., 2016; M. Wang & Vermeulen, 2021). Street-level images provide rich contextual information as perceived by pedestrians (Suel et al., 2021). Nevertheless, street-level images typically capture only sectional features from a point on the street, which limits capturing regional characteristics.

Considering that pedestrians experience two different contexts that coexist within the urban environment simultaneously, there is a need to integrate the two heterogeneous natures of urban data. Existing studies have primarily focused on using a single source of socio-spatial or street-level data to identify pedestrian injury risk, limiting the collaborative utilization of information from different data sources. Therefore, one of the main contributions of the current study is that it integrates socio-spatial characteristics and street-level perceptual pedestrian experience risk factors using multimodal DL to predict pedestrian injury risk.

3. Methods

This study aims to identify contributing risk factors in both socio-spatial and street-level contexts and to further explain how the identified risk factors are associated with pedestrian injury risk. To achieve this, this study explores how the two different contexts assist in describing pedestrian injury risk through variation of the integrating ratio. Fig. 1 illustrates the procedures of this study, and descriptions of the procedures are provided as follows.

1. Data collection: Randomly generated area points along street centerlines in Seoul were buffered by 50 m to extract three main categories of data: socio-spatial context, GSV images representing street-level context, and the frequency of pedestrian injuries as labels.
2. Data processing: The data were preprocessed depending on the data type. Socio-spatial characteristics were extracted, standardized, and saved in a tabular format; GSV images were segmented into 19 built environment components and resized; and the frequencies of pedestrian injuries were categorized into 3 risk levels.

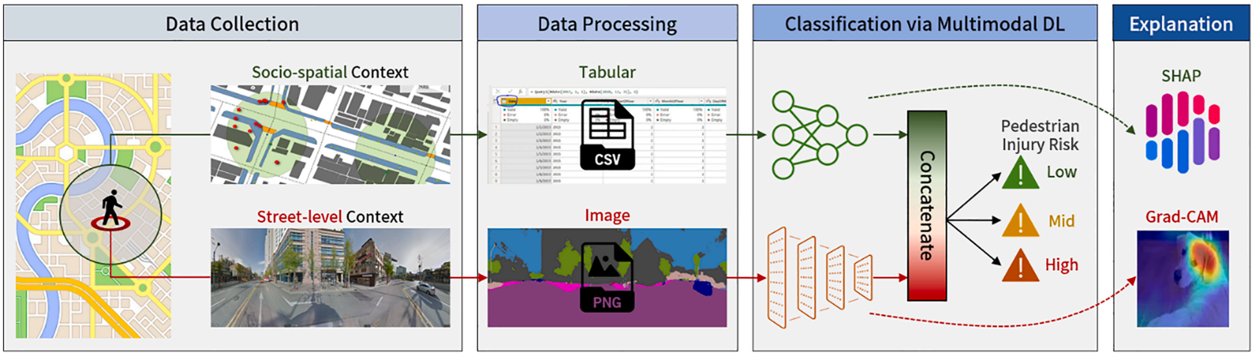

Fig. 1. Procedures used in this study.

3. Classification via multimodal DL: Pedestrian injury risk levels were classified using multimodal DL. Here, six different models with varying integration ratios of the two data types were employed and compared to investigate how the two different urban contexts assist in describing pedestrian injury risk.
4. Explanation: Appropriate XAI methods were employed to identify important risk factors for each data type. The SHapley Additive exPlanation (SHAP) approach was utilized for tabular data, and the gradient-weighted class activation mapping (Grad-CAM) method was applied for image data.

3.1. Study setting: city of Seoul in South Korea

This study investigated urban pedestrian injury risk factors in Seoul, South Korea as a study setting. Seoul accommodates more than 20 % of the South Korea's entire national population. Seoul has 10 million residents in an urban area covering approximately 605 km^2, resulting in a population density of 16,840 persons/km^2. Seoul's population continued to grow quickly until the early 1990s, mainly due to rural-to-urban migration (H. M. Kim & Han, 2012). Seoul's rapid population growth has led to both urbanization and a rapid increase in high-rise buildings, particularly apartments (Choi et al., 2019).

Seoul was selected as the study setting due to its dense and highly populated characteristics, which requires that much attention be paid to pedestrian safety as it is one of the major metropolitan areas in the world. This study is specifically limited to Seoul. Thus, while the findings of the study could provide insights for pedestrian safety in metropolitan contexts, they may not be applicable to other urban contexts due to the particular characteristics of Seoul. Therefore, future studies can expand related study settings to multiple cities with distinct socio-spatial characteristics and cultural backgrounds to better understand the role of urban contexts in pedestrian safety to design and enforce new requirements and policies.

3.2. Data

3.2.1. Pedestrian injury risk level as labels

Pedestrian injury risk was defined as the frequency of outdoor pedestrian injuries occurring within a 50-m radius of a specific point along a street centerline. The pedestrian injury data used in this study were obtained using the safe map application programming interface (API) of the public data portal provided by the Ministry of Public Administration and Security. The provided data contained pedestrian injuries reported to the emergency department from January 2018 to June 2021 in Seoul. The public data included four categories of pedestrian injuries recorded by paramedics: injuries, accident injuries, severe trauma, and other safety accidents. The outdoor pedestrian injury data were imported into the quantum geographical information system (QGIS), which is an open-source desktop geographic information system, through the web feature service (WFS) API in point vector format to geographically identify the injury locations.

To measure the risk of pedestrian injuries for each street segment in Seoul, 20,000 random points were generated along the street centerlines using QGIS. Subsequently, a buffer zone with a 50-m radius was created around each random point. Outdoor pedestrian injuries within each buffer zone were included by utilizing the QGIS superposition function, and the frequency of outdoor injuries within each buffer zone was recorded. The frequency data were saved in comma-separated value (CSV) format, along with the corresponding random point IDs.

The results of extracting pedestrian injury occurrences from a randomly generated 20,000 points showed a significant decrease in the number of points as the occurrence increased (Fig. 2a). This indicates that there is an inequality in pedestrian injury occurrences throughout Seoul. This inequality in the labels, also known as the class imbalance issue, can lead to performance degradation and ambiguous interpretations in classification results (Japkowicz & Stephen, 2002). Therefore, an attempt was made to balance the class distribution at each risk level to address the class imbalance issue. The frequencies of pedestrian injuries were categorized into 3 different levels. The risk levels were defined as follows: 'Low' if the number of pedestrian injuries was 0, 'Mid' if it was 1 or 2, and 'High' if it was 3 or more. Fig. 2b illustrates the frequency of the samples according to the corresponding pedestrian injury risk levels.

3.2.2. Socio-spatial data

In this study, 26 variables were established as socio-spatial risk factors in a tabular input format. Table 2 shows an overview of the socio-spatial variables utilized in this study. All spatial information was imported into QGIS, and the information located within a 50-m buffer of random points along the streets was extracted. The extracted information was recorded with respect to the IDs of the randomly generated points, along with the corresponding pedestrian injury risk levels.

The building and land characteristics included variables such as the number of buildings, building area, average building height, average building age, and land use. The number of buildings represents the number of buildings in each buffer zone, and the building area refers to the sum of the area of all buildings within each buffer zone. Additionally, the average building age was calculated based on the elapsed years from the use approval date of buildings to 2023. Furthermore, the calculated average building height and the recorded land use, such as restricted development area, commercial area, green area, residential area, and industrial area, were included.

The road characteristics consisted of variables such as the number of road segments, average road width, average road length, and road type. A road segment refers to a division of the road when the direction changes, and the number of road segments represents the count of segments within each buffer zone. The calculated average width and length of each segment and the road type (e.g., highway, general road, or provincial road) were also included in the analysis. In cases in which

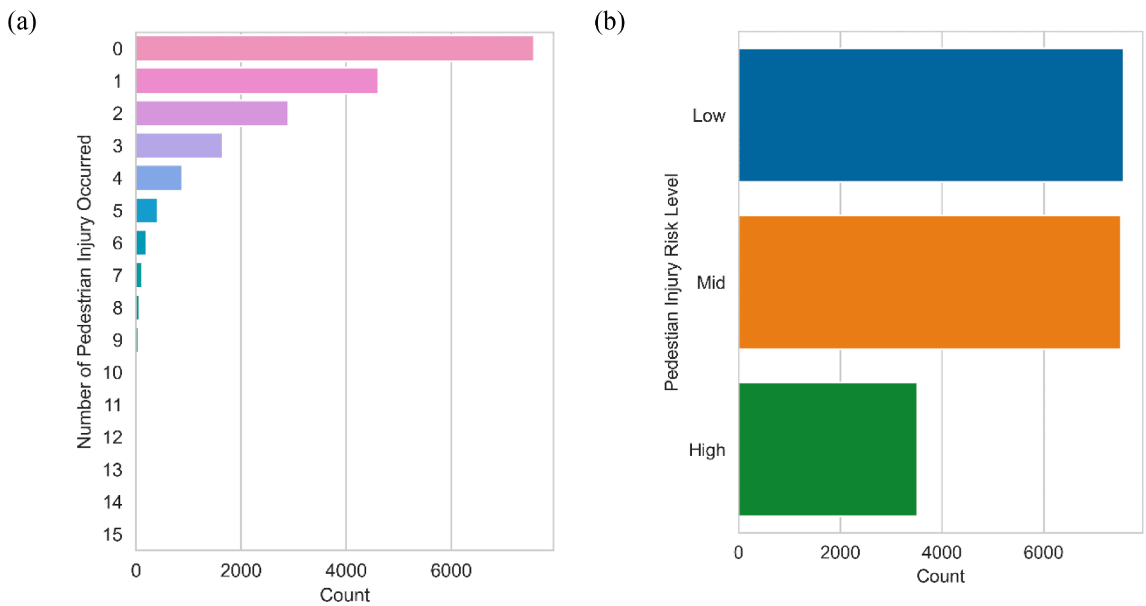

Fig. 2. Frequency of samples based on the (a) number of pedestrian injuries and (b) pedestrian injury risk levels.

Table 2
Overview of the socio-spatial variables utilized in this study.

Category	Variable	Source	Data Type
Building and land characteristics	Number of buildings	National Spatial Information Portal	Numerical
	Building area		
	Average building height		
	Average building age		
	Land use		Categorical
Road characteristics	Number of road segments	National Spatial Information Portal	Numerical
	Average road width		
	Road type		Categorical
Pedestrian facilities	Area of sidewalk	Ministry of the Interior and Safety	Numerical
	Area of crosswalk		
Lighting conditions	Number of streetlamps	Ministry of the Interior and Safety	Numerical
Public transit	Distance to subway station	Seoul Open Data Square	Numerical
	Number of bus stops	Seoul Transport Operation and Information Service	
Demographics	Number of houses	National Spatial Information Portal	Numerical
	Population age 10		
	Population age 20		
	Population age 30		
	Population age 40		
	Population age 50		
	Population age 60		
	Unemployment rate	Seoul Open Data Square	
	Number of male residents		
	Number of female residents		
	Education level		

there were multiple road types, the one with the highest frequency was selected.

Pedestrian facilities included variables such as the area of the sidewalk and crosswalk. The superstition function in QGIS was used to extract each spatial information within each buffer zone. The areas of the sidewalk and crosswalk within each buffer zone were calculated and recorded.

Lighting conditions included the number of streetlamps, and they were extracted using spatial information. By employing the superstition function in QGIS, the number of streetlamps within each buffer zone was extracted and recorded.

Public transit included variables such as the distance to the nearest subway stations and the number of bus stops. Similarly, the distance to the nearest subway station from each randomly generated point and the number of bus stops within each buffer zone were recorded.

Demographics included variables such as the number of houses, total population, the percentage of population by age group, number of male and female residents, unemployment rate, and education level. Demographic information for South Korea was provided by the National Spatial Information Portal in a 100-m grid format. By using the join function in QGIS, the demographic information for the closest grid to each randomly generated point was extracted.

However, multicollinearity caused by a high level of correlation among variables can potentially degrade the performance of DL models and create confusion in quantifying the impact of variables on the dependent variable (Basu & Maji, 2022). Therefore, variables with an absolute Pearson correlation coefficient greater than 0.80 were removed through correlation analysis. Among the 26 variables used in this study, the following high-level correlations were observed: 0.82 for total population and number of houses and 0.93 for average road length and average road width. Consequently, the following two variables were not considered in this study: total population and average road length.

Furthermore, numerous random points were excluded from these study samples because, no buildings were observed within a 50-m buffer of these points. In instances where buildings were not observed, variables such as building area, average building height, average building age, and land use could not be extracted. As a result, a total of 2034 random points out of 20,000 were excluded from the scope of this study.

3.2.3. Street-level images

When walking, individuals receive various visual stimuli from the street, which influence their behaviors and perceptions (Li et al., 2022). To quantify the visual stimuli experienced by pedestrians, GSV images were utilized in this study. Images depicting the components of the built environment were employed instead of using raw panoramic views. The reason for applying segmented images is that they provide the location of each element in the built environment and its local connectivity with other elements at the pedestrian level.

To segment GSV images into 19 built environment components, a pretrained model comprising DeepLabv3+ and MobileNetV2 was employed. This architecture was introduced by Zhao et al. (2017) and demonstrated commendable performance in the Pascal Visual Object Classification Challenge and for the Cityscapes test dataset (P. Liu et al., 2019; M. Wang & Vermeulen, 2021). The model relies on semantic segmentation using the Cityscapes dataset (Cordts et al., 2016). By leveraging this approach, pixel-level semantic information for 19 physical components, including vegetation, roads, sidewalks, terrain, buildings, walls, fences, sky, cars, trucks, buses, trains, motorcycles, bicycles, poles, traffic lights, traffic signs, people, and riders, were extracted. Panoramic views are interpreted by assigning color groups to each segmented element, achieving a mean intersection over union value of 72.1 % (Zhao et al., 2017).

The 20,000 randomly generated points were assigned unique IDs, which included coordinates of latitude and longitude as attributes. Using the coordinates, panoramic views were obtained from the Street View Static API of the Google Maps platform. A panoramic view was generated by horizontally merging three images with a resolution of 640 × 640 ($W \times H$) and a field of view of 120°.

However, GSV panoramic views are not always available and are often acquired with an insufficient quality, such as with positioning errors or occlusions (Srivastava et al., 2019). Since this issue existed in this study as well, points that fell outside the GSV scene or were associated with low quality were visually identified and excluded from the sample of this study. As a result, a total of 1396 random points were excluded from the study. Fig. 3 represents an overview of socio-spatial and street-level data extraction.

3.3. Integration of socio-spatial and street-level contexts

The main contribution of the current study is integrating socio-spatial and street-level contexts to identify pedestrian injury risk factors using multimodal DL. While the socio-spatial context included 24 variables after excluding 2 highly correlated variables, the street-level context contained GSV images segmented into 19 built environment components and their local positions perceived by pedestrians. A total of 16,570 random points were considered after data filtering and socio-spatial variables were standardized to have a mean of 0 and a variance of 1 before being input into the multimodal DL model.

Furthermore, as an attempt to investigate the contribution ratio of the socio-spatial and street-level contexts in describing pedestrian injury risk levels, six multimodal DL models were established by varying the integration ratio of the two contexts (e.g., 0:100, 20:80, 40:60). Fig. 4 illustrates the architecture of the multimodal DL model used in this study, which integrates socio-spatial and street-level contexts in varying ratios.

The 24 variables representing socio-spatial characteristics were passed through three layers of a fully connected network with 256, 512, and 1024 neurons, respectively. Consequently, 1024 features were extracted from the last layer of socio-spatial-related networks.

In parallel, a densely connected convolutional network (DenseNet) (Huang et al., 2017) was employed to extract features from the street-level context. DenseNet has been widely used for image classification tasks (Li et al., 2022), and several studies have evaluated and predicted human perception along streets by employing this network (Dubey et al., 2016; Min et al., 2019; Verma et al., 2020; F. Zhang et al., 2018). A segmented GSV image was resized to a resolution of 332 × 112 ($W \times H$) and input into pretrained DenseNet-201, and 57,600 features were extracted from the last layer of street-level related networks.

A total of 1024 socio-spatial features and 57,600 street-level features were integrated depending on the designated ratio of six different models. The number of neurons in the concatenation layer was fixed at 2560 for all models but varied in terms of the ratio of data types considered. For example, in Model 1, only the socio-spatial context was utilized, expanding 22 features to 2560, while in Model 6, only the street-level context was considered, resulting in the contraction of 57,600 features to 2560.

After the concatenation of the two contexts, the concatenated features were passed through two fully connected layers with 512 and 128 neurons and then through the classification layer. In the classification

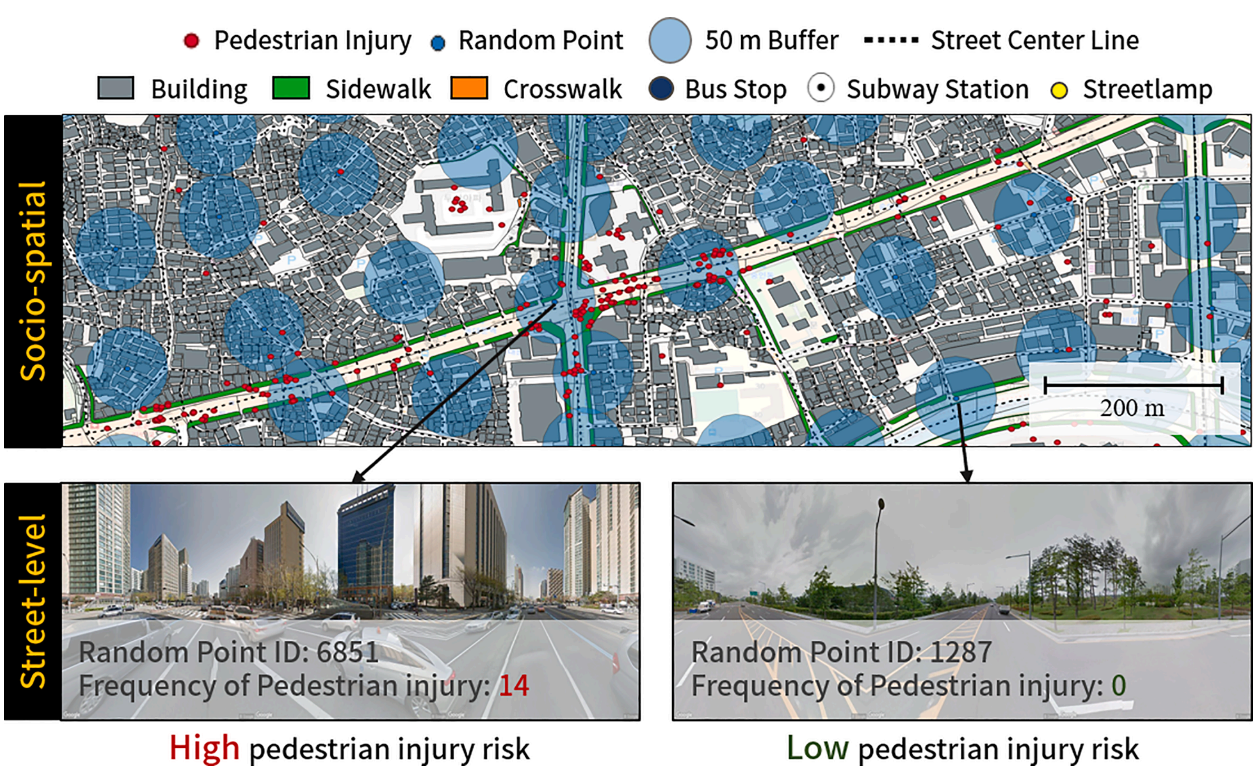

Fig. 3. Overview of socio-spatial and street-level data extraction.

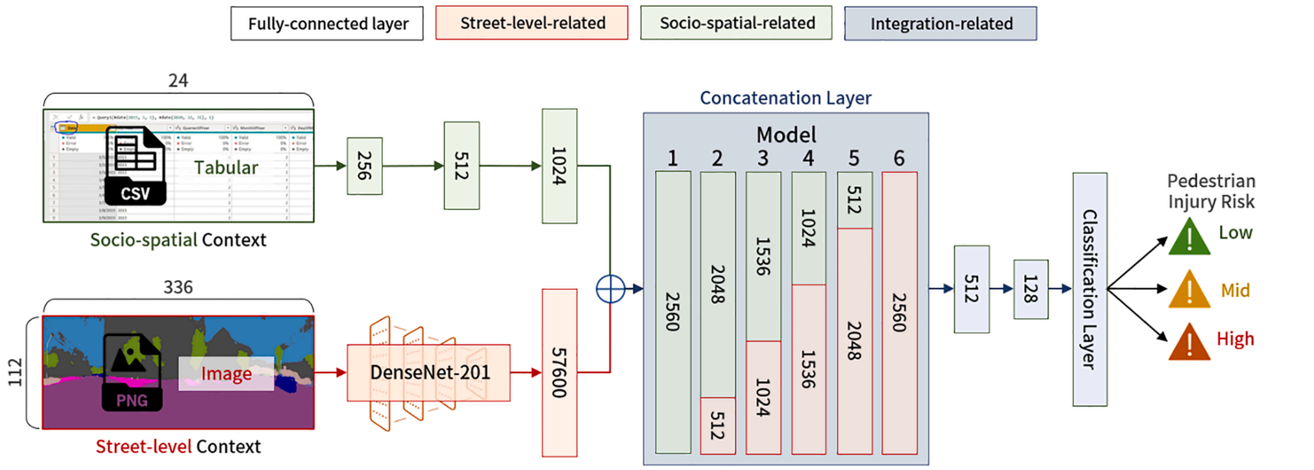

Fig. 4. The architecture of the multimodal DL model integrates socio-spatial and street-level contexts.

layer, the model predicted the possibilities of three levels of pedestrian injury risk.

To compare the performances of all six models that use different integration ratios, the hyperparameters of the multimodal DL model were fixed as follows: a rectified linear unit (ReLU) as an activation function; 0.0005 as the learning rate; 50 as the number of epochs; 128 as the batch size to overcome class imbalance; and cross-entropy loss as a loss function. Furthermore, the trained models were evaluated using performance metrics, such as accuracy, precision, recall, and the F1-score.

3.4. Interpretation of the multimodal DL model

3.4.1. Socio-spatial interpretation: SHAP

The contribution and effect of socio-spatial risk factors were calculated through SHAP scores. SHAP, which was introduced by Lundberg and Lee (2017), is a model-agnostic technique in which an explanatory model is constructed using training data and the trained model. SHAP values are used to assign fair levels of importance among features (Z. Zhang et al., 2022), with a player representing either a single feature value or a set of feature values.

SHAP has been utilized in urban settings to investigate contributing pedestrian safety risk factors in previous studies, such as those focused on residential and road segment density for pedestrian casualties (Chang et al., 2022), and older pedestrian characteristics for traffic crash severity (Guo et al., 2021). A SHAP-based interpretation is used in this study to explore which socio-spatial context of a built environment has a significant influence on the pedestrian injury risk level based on the trained multimodal DL model. This interpretation approach aids in exploring strategies for improvements in the built environment and mitigating pedestrian injury risk.

3.4.2. Street-level interpretation: Grad-CAM

A DL tool called Grad-CAM was employed to assess the street-level context that affects the pedestrian injury risk. Grad-CAM is based on a conventional approach that combines gradient information with feature mapping for gradient weighting (Fu et al., 2020). The resulting gradient-weighted class activation map aids in locating regions within the input samples that are crucial for class discrimination (Selvaraju et al., 2017). These regions are highlighted using feature maps from the final convolutional layer, which retain spatial information from captured visual patterns (Li et al., 2022).

Grad-CAM visualizations help interpret the annotated objects relevant to the DL model's predictions when they are applied to GSV images segmented into 19 components (Sangers et al., 2022). Several previous studies have already proven the effectiveness of using this approach in urban settings to investigate which built environment components affect pedestrians' perceptions, such as street infrastructure for walkability (Li et al., 2022), and observed cars for safety (Sangers et al., 2022). This study also employs this technique to understand which components perceived by pedestrians are related to pedestrian injury risk. This approach can provide valuable insight into a strategy for the placement of built environment components to reduce pedestrian injury risk.

4. Results

4.1. Data descriptions according to pedestrian injury risk levels

4.1.1. Socio-spatial context and pedestrian injury risk level

The characteristics of the 24 socio-spatial variables depending on the pedestrian injury risk level are described in Fig. 5. White circles indicate the average of the variables collected from the whole study samples. As a result, several socio-spatial variables reported an increase or decrease in the mean values as the pedestrian injury risk level increased.

Among the demographic variables, the following variables reported changes according to the risk levels: the number of houses, the population of those in their 60 s, the unemployment rate, and the education level. Higher risk levels were found to be associated with a higher number of houses, a higher unemployment rate, and a greater population of those in their 60 s, and a lower education level.

All of the building-related variables showed changes depending on the injury risk levels as well. Higher risk levels were found to be associated with a higher number of buildings, a higher average building height, a larger building area, and a lower average building age. However, while the land use variable did not show a distinct difference between mid and high levels of pedestrian injury risk, it did show a different pattern in low-risk levels.

The characteristics of the road network also showed different patterns according to the risk levels. A higher number of road segments was reported at higher risk levels. Despite the categorical attributes of the road type, the variable showed different patterns depending on the risk levels.

Pedestrian facilities, such as areas of sidewalks and crosswalks, showed a similar pattern. Although they did not show a linear trend as the risk level increased, they showed distinctively higher values in the high-risk level.

Furthermore, public transit, such as the distance to the nearest subway station and the number of bus stops, showed changes according to the risk levels. At higher risk levels, the distance to the nearest subway stations was shorter, and more bus stops were observed.

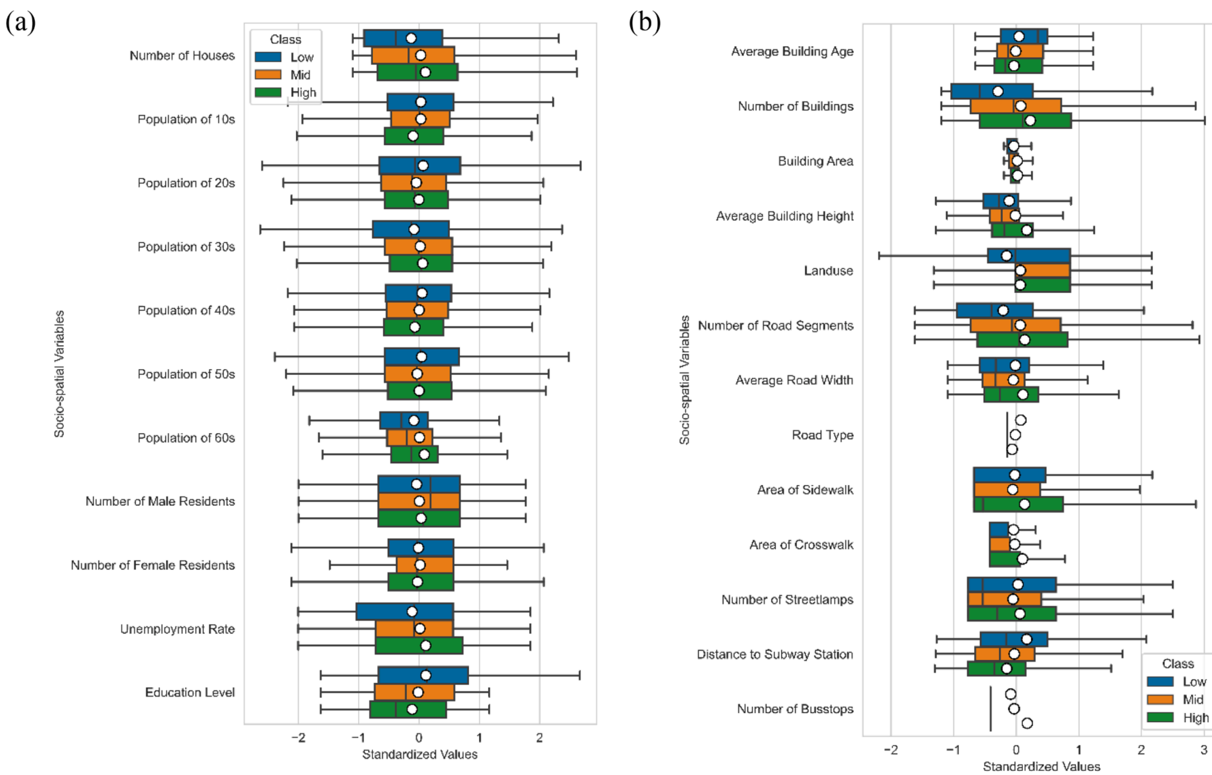

Fig. 5. Descriptions of the 24 socio-spatial context variables according to the pedestrian injury risk levels.

4.1.2. Street-level context and pedestrian injury risk level

The street-level characteristics according to the pedestrian injury risk level are illustrated in Fig. 6. For the street-level context, GSV images were obtained from random points and classified based on the risk levels. Consequently, distinct visual characteristics in the collected GSV images were observed depending on the risk level.

When observing representative GSV images, distinct visual characteristics of the built environment components were observed according to the three different pedestrian injury risk levels. At the low-risk level, there were few components aligned along the horizon, allowing for a more expansive view of the sky. On the other hand, as the risk level increased, a higher number of components, such as buildings, trees, and mobility, were observed. Therefore, the proportion of visible sky was found to be decreased at higher levels since it was blocked by

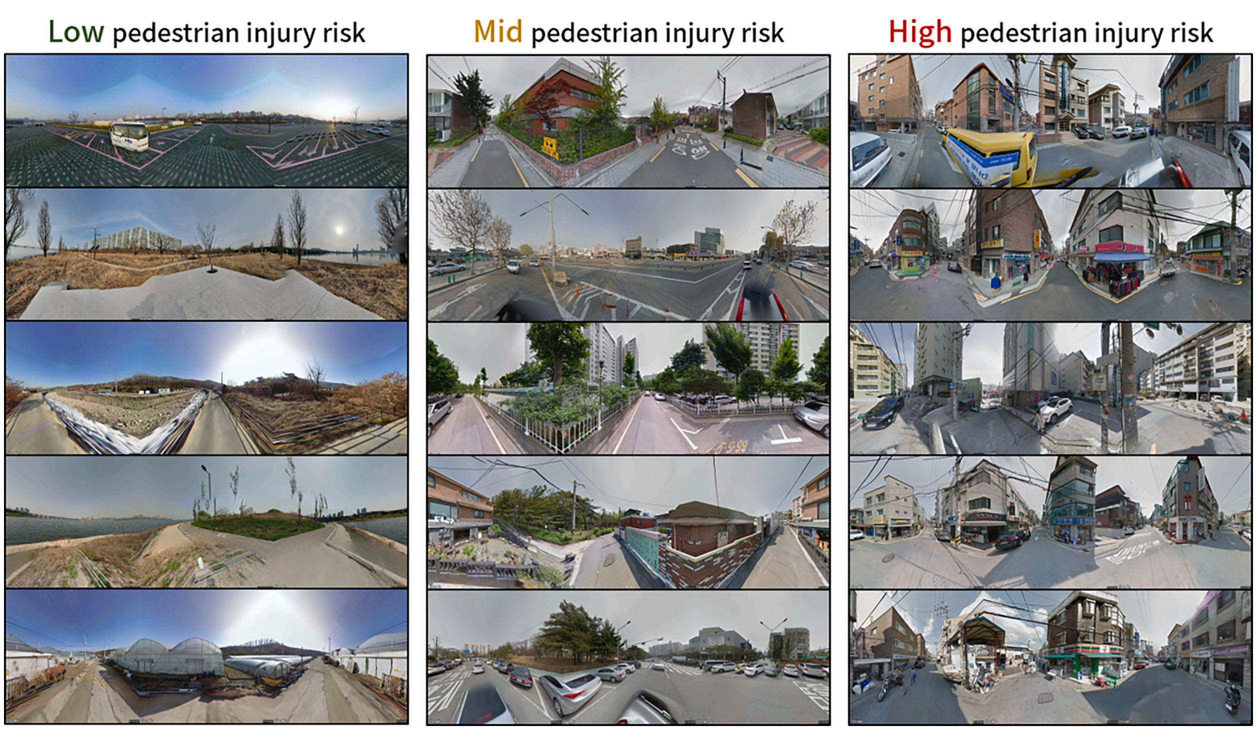

Fig. 6. Representative GSV images of the three pedestrian injury risk levels.

the other components. The visually perceived characteristics of the built environment components can be described through the segmented GSV images and their corresponding pedestrian injury risk levels.

4.2. Development of the multimodal DL model integrating socio-spatial and street-level contexts

To investigate how the socio-spatial and street-level contexts assist in describing pedestrian injury risk levels, six multimodal DL models were established by adjusting the integration ratio of the two contexts. Table 3 summarizes the performances of six multimodal DL models based on the integration ratios of the two urban contexts.

While Model 1 (using only socio-spatial context variables) displayed a relatively low accuracy of 0.62, Model 6 (using only street-level context variables) achieved a relatively higher accuracy of 0.86. Furthermore, among the integrated models, Model 4 (using a 40:60 ratio of socio-spatial and street-level contexts) reported the highest accuracy of 0.93. Model 4 also achieved the highest precision, recall, and F1-score outcomes, indicating the model's ability to minimize the errors in classifying from low to high level, or vice versa.

These results help highlight the contributions of socio-spatial and street-level contexts to pedestrian injury risk levels. For instance, looking at only socio-spatial context variables (Model 1) might not fully explain the pedestrian injury risk factors as much as looking at street-level context variables (Model 6). Furthermore, the combination of both socio-spatial and street-level contexts (Model 4) helps improve the understanding of pedestrian injury risk factors compared to looking at only one contextual context. Furthermore, the accuracy difference between Model 1 and Model 4 was 0.31, while the difference between Model 6 and Model 4 was 0.07. Such results indicate that the inclusion of street-level context is critical for understanding pedestrian injury risk factors, while the inclusion of socio-spatial context slightly improves the understanding of pedestrian injury risk factors.

4.3. Identification of risk factors for pedestrian injury

4.3.1. Socio-spatial risk factors

The contribution and effect of the socio-spatial variables were further analyzed by SHAP values obtained from Model 4, which reported the highest performance among the models. Fig. 7 shows the effect of the socio-spatial variables that contribute to classifying high-risk levels of pedestrian injury. Fig. 7a represents the overall importance and effect of the variables, and Fig. 7b presents the directions based on the feature SHAP values.

The top 10 variables with the highest contribution in the socio-spatial context were as follows: crosswalk area, education level, average building height, number of bus stops, average road width, number of buildings, population of those in their 60 s, unemployment rate, distance to nearest subway station, and number of road segments.

The variable that showed the highest contribution to high-risk levels was the area of crosswalks. It was found that a large crosswalk area can lead to a high risk of pedestrian injuries. This result is aligned with a previous study that highlighted that the presence of traffic islands and crosswalk infrastructures is associated with elevated numbers of pedestrian injuries (Mooney et al., 2016).

Furthermore, a high number of buildings and high average building heights contribute to a high level of pedestrian injury risk, which is a critical issue in Seoul due to the rapidly increasing number of high-rise buildings (H. M. Kim & Han, 2012). Additionally, a high number of bus stops and short distances to subway stations contribute to a higher risk level of pedestrian injury. Generally, public transit systems promote increased population flows (J. Kim et al., 2022), which can be a contributing factor to injury risk. Given Seoul's development of public transportation to accommodate its population density, these findings highlight the importance of the configuration of the built environment to mitigate injury risk.

Additionally, a high number of road segments and wide road width are also contributing factors to pedestrian injuries. The number of road segments is linked to both public transit and building characteristics. Roads are positioned between buildings, and they are especially prevalent at intersections with public transit. Moreover, wide road widths along complex roads were found to further increase the risk of pedestrian injury.

The following socio-spatial demographics are also high-risk factors: a low education level, a large population of those in their 60 s, and a high unemployment rate. These results are consistent with previous studies that have highlighted the importance of the characteristics of residents (LaScala et al., 2000; Niebuhr et al., 2016; Oxley et al., 2018). Such findings indicate that socio-spatial demographics also need to be considered in improvements of urban environments to alleviate pedestrian injury risk.

4.3.2. Street-level risk factors

Moreover, the contribution of built environment components represented in the street-level images was analyzed by image features extracted from the last layer of DenseNet-201 of Model 4. Fig. 8 shows the operation procedure used for Grad-CAM, which is as follows: 1. GSV images were segmented into 19 built environments and input to the multimodal DL model; 2. A gradient-weighted class activation map representing Grad-CAM importance (ranging from 0 to 1) for each pixel was calculated using Model 4; and 3. the activation map was overlaid on the segmented GSV images.

For further analysis, a comparative investigation between built environment components and Grad-CAM importance was conducted. First, the proportion of the pixel count for each observed component to whole image pixels (a resolution of 336 × 112) from all samples used in the study was calculated, and the components observed in all pixels were matched with their Grad-CAM importance. Fig. 9a shows the proportion occupied by each component within a single image for all samples, and Fig. 9b illustrates the distribution of Grad-CAM importance for each component in all samples.

As illustrated in Fig. 9a, buildings, roads, vegetation, sky, and cars were most frequently observed in GSV images on average. The proportions of buildings, roads, vegetation, and sky also varied based on the pedestrian injury risk level. More buildings but less sky, vegetation, and roads were observed at higher injury risk levels.

However, interestingly, the Grad-CAM importance values of the four main visual contents varied in their order of importance (Fig. 9b). The model exhibited a greater focus on sky and vegetation, rather than buildings and roads, which were more frequently observed in GSV

Table 3
Multimodal DL model performance based on the integration ratios of the two urban contexts.

Model	Description	Integration ratio (%)		Performance evaluation matrix			
		Socio-spatial	Street-level	Accuracy	Precision	Recall	F1-score
1	Socio-spatial only	100	0	0.62	0.65	0.58	0.60
2	Integration of socio-spatial and street-level	80	20	0.80	0.83	0.79	0.77
3		60	40	0.85	0.82	0.82	0.82
4		**40**	**60**	**0.93**	**0.93**	**0.94**	**0.93**
5		20	80	0.89	0.89	0.87	0.87
6	Street-level only	0	100	0.86	0.87	0.86	0.86

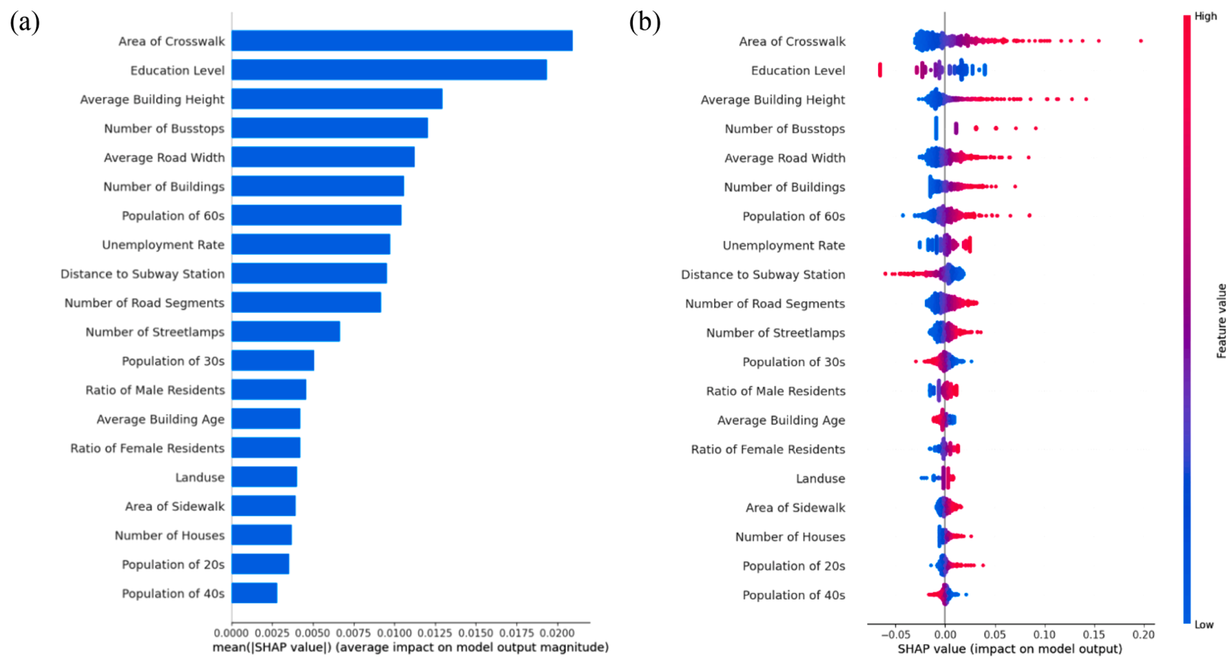

Fig. 7. SHAP results from socio-spatial variables describing (a) feature importance and (b) directions based on feature values.

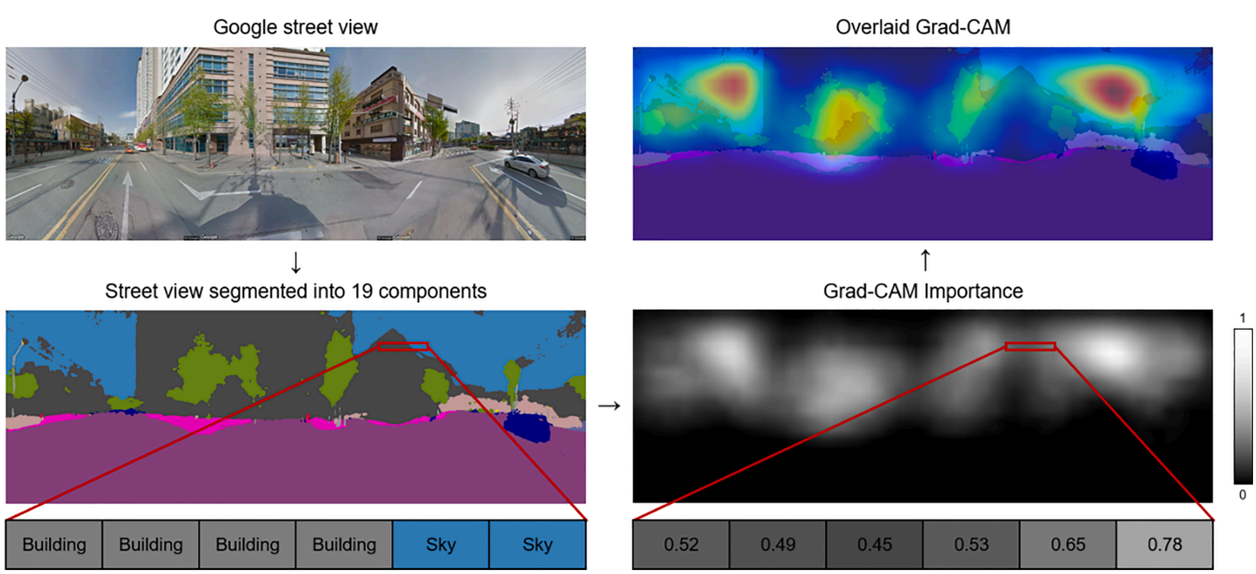

Fig. 8. Procedures of the Grad-CAM operation.

images. Furthermore, the Grad-CAM importance of trucks and buses followed that of the sky, while these items were relatively infrequently detected in GSV images. This indicates that the presence of trucks and buses can increase pedestrian injury risk by obstructing the street-level view of pedestrians by their presence, even though they were rarely detected.

5. Discussion

5.1. Role of multimodality from socio-spatial and street-level contexts

This study employed multimodal DL to integrate socio-spatial and street-level contexts to identify pedestrian injury risk factors. Furthermore, six multimodal DL models were established by adjusting the integration ratio of the two contexts to investigate how the socio-spatial and street-level contexts assist in describing pedestrian injury risk levels.

As a result of the performance evaluation of the six multimodal DL models, the model with an integration ratio of 40:60 for socio-spatial and street-level contexts achieved the highest accuracy of 0.93. It was found that the street-level context plays a critical role in the classification of pedestrian injury risk level, while the socio-spatial context assists in understanding the pedestrian injury risk level.

The developed multimodal DL model learned different aspects from two urban contexts. In the context of socio-spatial context, the identified risk factors were as follows: the area of crosswalks as pedestrian facilities; the number and heights of buildings as building characteristics; the width and number of road segments as road characteristics; and the education level, the population of those in their 60 s, and the unemployment rate as demographical characteristics. On the other hand, the identified risk factors for street-level context were as follows: sky, trucks, buses, and vegetation. Even though the characteristics of buildings and roads were identified as high-risk factors in the socio-spatial context,

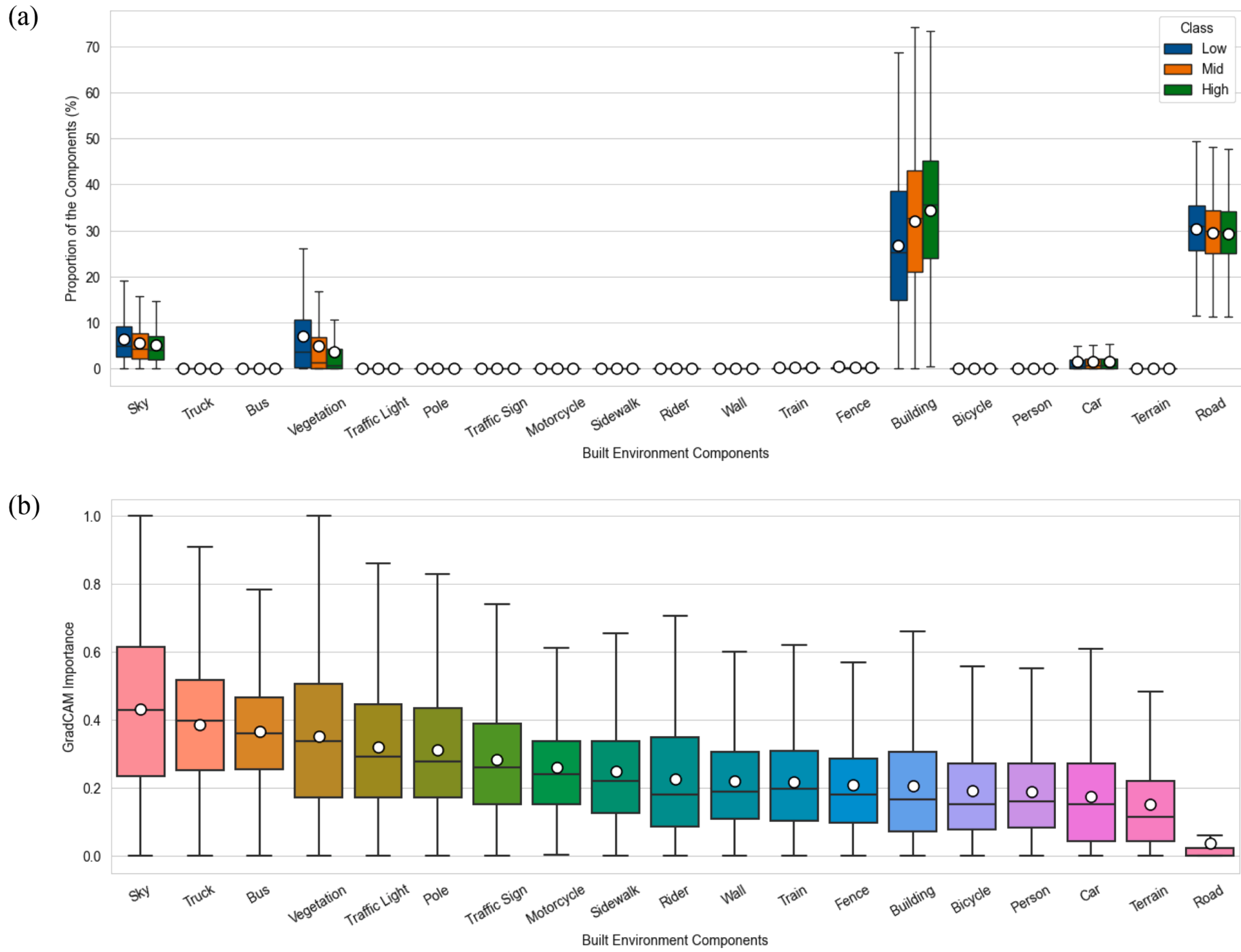

Fig. 9. Distribution of each component's (a) occupying proportion within a segmented GSV image (b) Grad-CAM importance.

their importance was found to be relatively low in the street-level context.

Typically, urban design policy improvement strategies are derived by focusing on the socio-spatial context. However, according to the results of this study, the street-level perceptual context is the main contributor to pedestrian injuries, followed by the socio-spatial context. The examination of these two urban contexts through multimodal deep learning provided different perspectives, enhancing the performance of models predicting pedestrian injury risk levels. Therefore, the development of urban design policy improvement strategies to mitigate pedestrian injury risk should consider both contexts.

5.2. Contributing risk factors to pedestrian injuries

The contributing risk factors for the two contexts were identified with SHAP and Grad-CAM processes. By combining the SHAP results in the socio-spatial context with the Grad-CAM results in the street-level context, three main contributing risk factors to pedestrian injuries can be summarized, namely, the fragmented sky view due to numerous tall buildings resulting from multiple road segments, the placement of crosswalks in areas adjacent to public transit sites, and interregional sociodemographic disparities.

The SHAP and Grad-CAM results emphasized the potential visual distraction caused by fragmented sky views divided by the locations of tall buildings. The SHAP results identified the number and height of buildings as risk factors, while the Grad-CAM results reported the view of the sky as the most important factor (while the view of buildings was less important). These findings are aligned with those of previous studies, which have suggested that complex variations in the silhouettes of multiple buildings decrease the level of perceived safety from the street-level visual perspective (Harvey & Aultman-Hall, 2015; Lindal & Hartig, 2013).

Especially in Seoul, there has been a rapid increase in high-rise buildings due to urbanization to accommodate rapid population growth (Choi et al., 2019). Therefore, urban policy improvements regarding silhouette variation could focus on measures such as maintaining a similar level of building heights and related elements in adjacent districts.

Another identified risk factor is the placement of crosswalks in areas near public transit. The SHAP results highlighted that the crosswalk area is the highest contributing risk factor in the socio-spatial context. They also indicated that numerous bus stops and short distances to subway stations are related to a high risk of pedestrian injuries. Additionally, the Grad-CAM results highlighted that the model captured the presence of buses and trucks as important determinants.

Several previous studies have revealed that the presence of public transit can increase population flows, potentially making adjacent areas more susceptible to pedestrian injuries (J. Kim et al., 2022; Mooney et al., 2016; Osama & Sayed, 2017). Such areas that are adjacent to public transit require a more refined crosswalk placement since pedestrians' behavior is more likely to be affected by visual cognitive load from movements of the components (H. Wang et al., 2022). Considering that 64.4 % of Seoul's land area is within a one-kilometer distance to a subway station (Jang, 2008), urban policy for pedestrian safety could be further improved through the proper placement of crosswalks adjacent to public transit.

The findings identified that interregional sociodemographic disparities need to be proactively considered in urban environment improvement strategies. The SHAP results revealed that a low education level, a high unemployment rate, and a higher population of those in their 60 s contribute to pedestrian injuries.

Previous research has shown that sensory decline in older individuals may lead to a reduced level of awareness of risky situations compared to that of younger individuals (Ayers et al., 2014; Gillespie et al., 2003; Oxley et al., 2018; Tobis et al., 1985). Additionally, the study results are consistent with a previous study that identified the unemployment rate and education level as risk factors for pedestrian injuries (LaScala et al., 2000). Overpopulation in Seoul has created a significant divide between rich and poor, resulting in discrepancies in income, education levels, road networks, and aging demographics (H. M. Kim & Han, 2012). Therefore, improvement strategies for pedestrian safety can be implemented proactively by considering interregional sociodemographic disparities.

6. Conclusions

While walking offers significant benefits, it also exposes one to unintentional pedestrian injuries that can pose serious health threats (Oxley et al., 2018). To create sustainable cities and meet public health objectives, it is crucial to ensure safe walking conditions by identifying and improving urban risk factors contributing to pedestrian injuries.

Previous studies have primarily explored the influence of the urban built environment on pedestrian injury risk within two key contexts, namely, socio-spatial and street-level contexts. However, these studies have often focused on using a single source of socio-spatial or street-level information to establish its relationship with pedestrian injury risk (Suel et al., 2021). Thus, there is a need to clarify how both urban contexts individually contribute to pedestrian injury risk from a comprehensive perspective through their integration.

To bridge this gap, the current study integrated socio-spatial and street-level contexts to predict pedestrian injury risk levels using a multimodal DL approach. This study aimed to identify risk factors in both contexts and explain how these factors relate to pedestrian injuries. Additionally, this study also explored how the two contexts assist in describing pedestrian injury risk by varying the integration ratio. The results indicated that an integration ratio of 40:60 for socio-spatial and street-level contexts, respectively, describes the urban contexts that most affect pedestrian injury risk levels.

Moreover, the multimodal DL model highlighted different aspects from both urban contexts. By combining SHAP results from the socio-spatial context with Grad-CAM results from the street-level context, three main contributing risk factors were identified: sky division due to numerous tall buildings resulting from multiple road segments, the placement of crosswalks in areas adjacent to public transit, and interregional sociodemographic disparities.

These findings align with those of previous studies, which have suggested that complex variations in the shapes of multiple buildings decrease the perceived level of safety from a street-level visual perspective (Harvey & Aultman-Hall, 2015; Lindal & Hartig, 2013). They are also consistent with previous research highlighting that transit areas may pose increased injury risks due to specific infrastructure components such as stairs, escalators, and transitions from sidewalks to roadways (J. Kim et al., 2022; Lakhotia et al., 2020; Mooney et al., 2016). Additionally, the importance of interregional sociodemographic disparities, such as education level, unemployment rate, and the size of the population of those in their 60 s, have also been emphasized in previous studies (Ayers et al., 2014; Gillespie et al., 2003; LaScala et al., 2000; Oxley et al., 2018; Tobis et al., 1985).

However, this study has a limitation in that it did not incorporate specifications regarding the physical condition of the individuals who had experienced unintentional injuries. Previous studies have indicated that an individual's physical condition can contribute to their level of pedestrian injury risk (Angal & Jagtap, 2016; Berg et al., 1992; Evans et al., 2001; Gillespie et al., 2003; Kallin et al., 2002). Therefore, the joint relationship between urban contexts and the physical condition of individuals who have experienced pedestrian injuries can be further investigated in future studies. Additionally, this study limited its study setting to Seoul, South Korea. While the findings of this study provide insights for densely populated metropolitan areas similar to Seoul, further investigation considering cultural and local contexts needs to be applied in other cities and countries.

By integrating socio-spatial and street-level perspectives, this study enhances the understanding of the complex interplay between urban contexts and pedestrian safety. This multiview and multimodal approach that integrates diverse urban contexts can inform more targeted and effective urban planning and policies. An understanding of the complex nature of urban big data through a multimodal approach can help create cities that are not only sustainable but also safer for pedestrians.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2022R1F1A1075136).

This work is financially supported by Korea Ministry of Land, Infrastructure and Transport (MOLIT) as 「Innovative Talent Education Program for Smart City」.

References

Angal, Y., & Jagtap, A. (2016). *Fall detection system for older adults. In 2016 IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT)* (pp. 262–266). IEEE. December.

Ayers, E. I., Tow, A. C., Holtzer, R., & Verghese, J. (2014). Walking while talking and falls in aging. *Gerontology, 60*(2), 108–113.

Barton, B. K. (2006). Integrating selective attention into developmental pedestrian safety research. *Canadian Psychology/Psychologie Canadienne, 47*(3), 203.

Basu, I., & Maji, S. (2022). Multicollinearity correction and combined feature effect in Shapley values. In *Australasian Joint Conference on Artificial Intelligence* (pp. 79–90).

Beitel, D., Stipancic, J., Manaugh, K., & Miranda-Moreno, L. (2018). Assessing safety of shared space using cyclist-pedestrian interactions and automated video conflict analysis. *Transportation Research Part D: Transport and Environment, 65*, 710–724.

Berg, K. O., Maki, B. E., Williams, J. I., Holliday, P. J., & Wood-Dauphinee, S. L. (1992). Clinical and laboratory measures of postural balance in an elderly population. *Archives of Physical Medicine and Rehabilitation, 73*(11), 1073–1080.

Bustos, C., Rhoads, D., Solé-Ribalta, A., Masip, D., Arenas, A., Lapedriza, A., & Borge-Holthoefer, J. (2021). Explainable, automated urban interventions to improve pedestrian and vehicle safety. *Transportation Research Part C: Emerging Technologies, 125*, Article 103018.

Cambra, P., & Moura, F. (2020). How does walkability change relate to walking behavior change? Effects of a street improvement in pedestrian volumes and walking experience. *Journal of Transport & Health, 16*, Article 100797.

Chan, E. T. H., Schwanen, T., & Banister, D. (2021). Towards a multiple-scenario approach for walkability assessment: An empirical application in Shenzhen, China. *Sustainable Cities and Society, 71*, Article 102949.

Chang, I., Park, H., Hong, E., Lee, J., & Kwon, N. (2022). Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP. *Accident Analysis & Prevention, 166*, Article 106545.

Choi, C. G., Lee, S., Kim, H., & Seong, E. Y. (2019). Critical junctures and path dependence in urban planning and housing policy: A review of greenbelts and New Towns in Korea's Seoul metropolitan area. *Land use policy, 80*, 195–204.

Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built

environment? *Transportation Research Part D: Transport and Environment, 14*(6), 425–436.
Clifton, K. J., Smith, A. D. L., & Rodriguez, D. (2007). The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning, 80*(1–2), 95–110.
Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3213–3223).
Crundall, D., Van Loon, E., & Underwood, G. (2006). Attraction and distraction of attention with roadside advertisements. *Accident Analysis & Prevention, 38*(4), 671–677.
Davis, S. J., Pugliese, B. J., & Barton, B. K. (2019). The intersection of pedestrian safety and multimodal perception. *Transportation Research Part F: Traffic Psychology and Behaviour, 67*, 205–216.
Decker, J. S., Stannard, S. J., McManus, B., Wittig, S. M., Sisiopiku, V. P., & Stavrinos, D. (2015). The impact of billboards on driver visual behavior: A systematic literature review. *Traffic injury prevention, 16*(3), 234–239.
Demetre, J. D., Lee, D. N., Pitcairn, T. K., Grieve, R., Thomson, J. A., & Ampofo-Boateng, K. (1992). Errors in young children's decisions about traffic gaps: Experiments with roadside simulations. *British Journal of Psychology, 83*(2), 189–202.
Donald, I. P., & Bulpitt, C. J. (1999). The prognosis of falls in elderly people living at home. *Age and Ageing, 28*(2), 121–125.
Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep learning the city: Quantifying urban perception at a global scale. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14* (pp. 196–212).
Du, S., Li, T., Gong, X., & Horng, S.-J. (2018). A hybrid method for traffic flow forecasting using multimodal deep learning. *ArXiv Preprint ArXiv:1803.02099*.
Evans, D., Hodgkinson, B., Lambert, L., & Wood, J. (2001). Falls risk factors in the hospital setting: a systematic review. *International Journal of Nursing Practice, 7*(1), 38–45.
Ferreira, M. C., Costa, P. D., Abrantes, D., Hora, J., Felício, S., Coimbra, M., & Dias, T. G. (2022). Identifying the determinants and understanding their effect on the perception of safety, security, and comfort by pedestrians and cyclists: A systematic review. *Transportation Research Part F: Traffic Psychology and Behaviour, 91*, 136–163.
Fried, L. F., Biggs, M. L., Shlipak, M. G., Seliger, S., Kestenbaum, B., Stehman-Breen, C., Sarnak, M., Siscovick, D., Harris, T., & Cauley, J. (2007). Association of kidney function with incident hip fracture in older adults. *Journal of the American Society of Nephrology, 18*(1), 282–286.
Fu, R., Hu, Q., Dong, X., Guo, Y., Gao, Y., & Li, B. (2020). Axiom-based grad-cam: Towards accurate visualization and explanation of cnns. *ArXiv Preprint ArXiv: 2008.02312*.
Gillespie, L. D., Gillespie, W. J., Robertson, M. C., Lamb, S. E., Cumming, R. G., & Rowe, B. H. (2003). *Interventions for preventing falls in elderly people (Cochrane review). Cochrane library*. Chichester: John Wiley and Sons.
Gkekas, F., Bigazzi, A., & Gill, G. (2020). Perceived safety and experienced incidents between pedestrians and cyclists in a high-volume non-motorized shared space. *Transportation Research Interdisciplinary Perspectives, 4*, Article 100094.
Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., & Wang, W. (2021). Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability, 13*(2), 926.
Hang, H. M., Ekman, R., Bach, T. T., Byass, P., & Svanström, L. (2003). Community-based assessment of unintentional injuries: A pilot study in rural Vietnam. *Scandinavian Journal of Public Health, 31*(62_suppl), 38–44.
Hanson, C. S., Noland, R. B., & Brown, C. (2013). The severity of pedestrian crashes: An analysis using google street view imagery. *Journal of Transport Geography, 33*, 42–53.
Harvey, C., & Aultman-Hall, L (2015). Urban streetscape design and crash severity. *Transportation Research Record, 2500*(1), 1–8.
Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4700–4708).
Hu, G., Ivers, R. Q., & Xiang, H. (2019). Advancing injury prevention in China. *Injury prevention* (pp. 1–2). BMJ Publishing Group Ltd. Vol. 25, Issue 1.
Hu, J., Flannagan, C., Ganesan, S., Bowman, P., Sun, W., Farooq, I., Kalra, A., & Rupp, J. (2023). Understanding the new trends in pedestrian injury distribution and mechanism through data linkage and modeling. *Accident Analysis & Prevention, 188*, Article 107095. https://doi.org/10.1016/j.aap.2023.107095
Jang, K. S. (2008). *Calculation for capital gains caused by high-density development around subway stations*. Seoul, Korea: Korea Appraisal Board.
Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429–449.
Kallin, K., Lundin-Olsson, L., Jensen, J., Nyberg, L., & Gustafson, Y. (2002). Predisposing and precipitating factors for falls among older people in residential care. *Public health, 116*(5), 263–271.
Kelly, P., Murphy, M., & Mutrie, N. (2017). The health benefits of walking. *Walking* (pp. 61–79). Emerald Publishing Limited. Vol. 9.
Kim, H. M., & Han, S. S. (2012). Seoul. *Cities (London, England), 29*(2), 142–154.
Kim, H. M., & Mateo-Babiano, I. (2018). Pedestrian crossing environments in an emerging Chinese city: Vehicle encountering, seamless walking, and sensory perception perspectives. *Sustainability, 10*(7), 2200.
Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention, 40*(5), 1695–1702.
Kim, J., & Lee, E. (2018). Understanding review expertise of developers: A reviewer recommendation approach based on latent Dirichlet allocation. *Symmetry, 10*(4). https://doi.org/10.3390/SYM10040114
Kim, J., Lee, J., Chung, S., & Jang, K. (2022). Effects of built environment in subway stations on pedestrian injuries. *Journal of Transport & Health, 26*, Article 101389.
Kumar, A., & Daumé, H. (2011). A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th International Conference on Machine Learning (ICML-11)* (pp. 393–400).
Lakhotia, S., Lassarre, S., Rao, K. R., & Tiwari, G. (2020). Pedestrian accessibility and safety around bus stops in Delhi. *IATSS Research, 44*(1), 55–66.
LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis. *Accident Analysis & Prevention, 32*(5), 651–658.
Lee, D. N., Young, D. S., & McLaughlin, C. M. (1984). A roadside simulation of road crossing for children. *Ergonomics, 27*(12), 1271–1281.
Lindal, P. J., & Hartig, T. (2013). Architectural variation, building height, and the restorative quality of urban residential streetscapes. *Journal of Environmental Psychology, 33*, 26–36.
Liu, J., Li, T., Xie, P., Du, S., Teng, F., & Yang, X. (2020). Urban big data fusion based on deep learning: An overview. *Information Fusion, 53*, 123–133.
Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., & Zhang, Y. (2019). Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. *Remote Sensing, 11*(7), 830.
Li, Y., Yabuki, N., & Fukuda, T. (2022). Measuring visual walkability perception using panoramic street view images, virtual reality, and deep learning. *Sustainable Cities and Society, 86*, Article 104140.
Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.
Mentis, A.-F. A., & Papadopulos, J. S. (2021). Near-collapse buildings and unsafe sidewalks as neglected urban & public health issue: A qualitative study. *Urban Science, 5*(2), 47.
Min, W., Mei, S., Liu, L., Wang, Y., & Jiang, S. (2019). Multi-task deep relative attribute learning for visual urban perception. *IEEE Transactions on Image Processing, 29*, 657–669.
Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D. M., Teitler, J. O., Sheehan, D. M., Jack, D. W., & Rundle, A. G. (2016). Use of google street view to assess environmental contributions to pedestrian injury. *American Journal of Public Health, 106*(3), 462–469.
Morrongiello, B. A., Corbett, M., Switzer, J., & Hall, T. (2015). Using a virtual environment to study pedestrian behaviors: How does time pressure affect children's and adults' street crossing behaviors? *Journal of Pediatric Psychology, 40*(7), 697–703.
Niebuhr, T., Junge, M., & Rosen, E. (2016). Pedestrian injury risk and the effect of age. *Accident Analysis & Prevention, 86*, 121–128.
Osama, A., & Sayed, T. (2017). Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety. *Canadian Journal of Civil Engineering, 44*(12), 1036–1044.
Oviedo-Trespalacios, O., Truelove, V., Watson, B., & Hinton, J. A. (2019). The impact of road advertising signs on driver behaviour and implications for road safety: A critical systematic review. *Transportation research part A: Policy and practice, 122*, 85–98.
Oxley, J., O'Hern, S., Burtt, D., & Rossiter, B. (2018). Falling while walking: A hidden contributor to pedestrian injury. *Accident Analysis & Prevention, 114*, 77–82.
Perez, W., & Bertola, M. A. (2011). The effect of visual clutter on driver eye glance behavior. In *Driving Assessment Conference (Vol. 6, No. 2011)*. University of Iowa.
Sangers, R., van Gemert, J., & van Cranenburgh, S. (2022). Explainability of deep learning models for urban space perception. *ArXiv Preprint ArXiv:2208.13555*.
Schwebel, D. C., Davis, A. L., & O'Neal, E. E (2012a). Child pedestrian injury: A review of behavioral risks and preventive strategies. *American Journal of Lifestyle Medicine, 6* (4), 292–302.
Schwebel, D. C., Stavrinos, D., Byington, K. W., Davis, T., O'Neal, E. E., & De Jong, D (2012b). Distraction and pedestrian safety: How talking on the phone, texting, and listening to music impact crossing the street. *Accident Analysis & Prevention, 45*, 266–271.
Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626).
Shahrbabaki, M. R., Safavi, A. A., Papageorgiou, M., & Papamichail, I. (2018). A data fusion approach for real-time traffic state estimation in urban signalized links. *Transportation Research Part C: Emerging Technologies, 92*, 525–548.
Srivastava, S., Vargas-Munoz, J. E., & Tuia, D. (2019). Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. *Remote Sensing of Environment, 228*, 129–143.
Suel, E., Bhatt, S., Brauer, M., Flaxman, S., & Ezzati, M. (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment, 257*, Article 112339.
Suk, J. Y., & Walter, R. J. (2019). New nighttime roadway lighting documentation applied to public safety at night: A case study in San Antonio, Texas. *Sustainable Cities and Society, 46*, Article 101459.
Sun, J., Yu, Q., Niyazbek, M., & Chu, F. (2021). Analysis of connection behaviour of communication network flow based on semantic understanding. *Microprocessors and Microsystems, 80*.
Tapiro, H., Oron-Gilad, T., & Parmet, Y. (2020). Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior. *Journal of Safety Research, 72*, 101–109.

Tobis, J. S., Reinsch, S., Swanson, J. M., Byrd, M., & Scharf, T. (1985). Visual perception dominance of fallers among community-dwelling older adults. *Journal of the American Geriatrics Society, 33*(5), 330–333.

Van Schagen, I., & Rothengatter, T. (1997). Classroom instruction versus roadside training in traffic safety education. *Journal of Applied Developmental Psychology, 18* (2), 283–292.

Verma, D., Jana, A., & Ramamritham, K. (2020). Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. *Building and Environment, 186*, Article 107340.

Wang, H., Li, D., Wang, Q., Schwebel, D. C., Miao, L., & Shen, Y. (2022). How distraction affects pedestrian response: Evidence from behavior patterns and cortex oxyhemoglobin changes. *Transportation Research Part F: Traffic Psychology and Behaviour, 91*, 414–430.

Wang, M., & Vermeulen, F. (2021). Life between buildings from a street view image: What do big data analytics reveal about neighbourhood organisational vitality? *Urban Studies, 58*(15), 3118–3139.

WHO, V. (2018). *Global status report on road safety 2018*. World Health Organization.

Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention, 41*(1), 137–145.

Wu, H., Chen, Y., Zhang, Z., & Jiao, J. (2020). The impact of street characteristics on older pedestrians' perceived safety in Shanghai, China. *Journal of Transport and Land Use, 13*(1), 469–490.

Yi, X., Zhang, J., Wang, Z., Li, T., & Zheng, Y. (2018). Deep distributed fusion network for air quality prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 965–973).

Zajac, S. S., & Ivan, J. N. (2003). Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention, 35*(3), 369–379.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning, 180*, 148–160.

Zhang, J., Zheng, Y., & Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. In *, 31. Proceedings of the AAAI Conference on Artificial Intelligence*.

Zhang, Z., Damiani, E., Hamadi, H.Al, Yeun, C.Y., & Taher, F. (2022). Explainable artificial intelligence to detect image spam using convolutional neural network. *ArXiv Preprint ArXiv:2209.03166*.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2881–2890).

Zheng, Y., Capra, L., Wolfson, O., & Yang, H. (2014). Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*(3), 1–55.

Zheng, Y., Yi, X., Li, M., Li, R., Shan, Z., Chang, E., & Li, T. (2015). Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2267–2276).

Zhu, L., Guo, F., Polak, J. W., & Krishnan, R. (2018). Urban link travel time estimation using traffic states-based data fusion. *IET Intelligent Transport Systems, 12*(7), 651–663.