Accident Analysis and Prevention 197 (2024) 107455

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Can we trust our eyes? Interpreting the misperception of road safety from street view images and deep learning

Xujing Yu [a], Jun Ma [a,b,*], Yihong Tang [a], Tianren Yang [a,b], Feifeng Jiang [c]

[a] Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China
[b] Urban Systems Institute, The University of Hong Kong, Hong Kong, China
[c] Faculty of Architecture, The University of Hong Kong, Hong Kong, China

ARTICLE INFO

Keywords:
Road safety
Human perception
Street view images
Deep learning
Built environment

ABSTRACT

Road safety is a critical concern that impacts both human lives and urban development, drawing significant attention from city managers and researchers. The perception of road safety has gained increasing research interest due to its close connection with the behavior of road users. However, safety isn't always as it appears, and there is a scarcity of studies examining the association and mismatch between road traffic safety and road safety perceptions at the city scale, primarily due to the time-consuming nature of data acquisition. In this study, we applied an advanced deep learning model and street view images to predict and map human perception scores of road safety in Manhattan. We then explored the association and mismatch between these perception scores and traffic crash rates, while also interpreting the influence of the built environment on this disparity. The results showed that there was heterogeneity in the distribution of road safety perception scores. Furthermore, the study found a positive correlation between perception scores and crash rates, indicating that higher perception scores were associated with higher crash rates. In this study, we also concluded four perception patterns: "Safer than it looks", "Safe as it looks", "More dangerous than it looks", and "Dangerous as it looks". Wall view index, tree view index, building view index, distance to the nearest traffic signals, and street width were found to significantly influence these perception patterns. Notably, our findings underscored the crucial role of traffic lights in the "More dangerous than it looks" pattern. While traffic lights may enhance people's perception of safety, areas in close proximity to traffic lights were identified as potentially accident-prone regions.

1. Introduction

Road safety is a significant global issue (AlHamad et al., 2023). According to the World Health Organization's Global Status Report on Road Safety in 2018, approximately 1.35 million individuals die each year as a result of road traffic crashes, making it the primary cause of death for individuals aged 5–29 (World Health Organization, 2018). Governments and affiliated agencies have recognized the grave threat that road traffic crashes pose to human life and have dedicated substantial efforts to mitigating the risks associated with such incidents (Chen et al., 2022). The United Nations has identified the death rate resulting from road traffic injuries as a crucial indicator in the Sustainable Development Goals (UN Statistics, 2018).

Compared to actual or reported traffic crashes and safety issues, the perception of road safety refers to a subjective evaluation of the likelihood of road accidents and provides information about high-risk areas before actual accidents occur (Lund & Rundmo, 2009; Salonen, 2018). Road safety perception, also known as crash risk perception, holds significant importance in traffic safety research and management as it relates to the behavior of road users and aids in understanding how they identify risk factors (Diógenes and Lindau, 2010). Moreover, the perception of road safety plays a crucial role in proactive transportation safety planning. It enables the recognition of existing and potential road safety problems and facilitates the implementation of timely preventive measures, even in the absence of actual crashes (Lee et al., 2016; Schneider et al., 2004).

Despite the significance of road safety and human perception in this regard, there is a lack of research exploring the association between actual crash risk and the perception of road safety. Places that seem safe may not be. Particularly in cities with complex built environments, it remains unknown whether we can accurately perceive potential dangers on the road. In other words, our motivation for conducting this study

* Corresponding author.
E-mail address: junma@hku.hk (J. Ma).

https://doi.org/10.1016/j.aap.2023.107455
Received 27 September 2023; Received in revised form 20 December 2023; Accepted 31 December 2023
Available online 12 January 2024

stemmed from our curiosity about the consistency between perceived traffic safety risk and actual crash risk, and if inconsistency exists, the reasons behind such discrepancies. These can lead to the development of targeted interventions and educational campaigns for crash prevention.

To address these inquiries, this paper aims to investigate the correlation between human perception of road safety and actual crash rates, as well as analyze the impact of urban built environment factors on the misperception of road safety. The perception scores of road safety will be assessed using street view images and deep learning techniques. The remainder of the paper is structured as follows. Section 2 provides a comprehensive review of previous studies pertaining to road safety perception, its relationship with actual crash risks, the influence of built environment factors, and the utilization of computer vision and street view images in traffic research. Section 3 introduces the research area, data sources, and methods employed in this study. Section 4 presents the results of our research, while section 5 offers a detailed discussion of the results, highlighting our contributions and acknowledging the limitations of the study. Finally, Section 6 presents the concluding remarks of this research.

2. Literature review

2.1. Road safety perception and its discrepancy with actual traffic crashes

Several studies have focused on road users' perception of safety and its relationship with actual traffic crashes. Previous literature has indicated a positive correlation between perceived crash risk and actual transport risks (Elvik & Bjørnskau, 2005). Moreover, road safety perception and actual traffic crashes have a reciprocal influence. A higher subjective appraisal of traffic risk increases road users' inclination to adopt safer traffic behaviors, thereby potentially reducing the occurrence of traffic crashes (Boua et al., 2022; Deery, 1999). Conversely, high actual crash rates can heighten people's perceived crash risk on the road (Cho et al., 2009).

However, human perception does not always align with actual safety risks. Limited rationality or failures in rational decision-making can lead to inaccurate assessments of accident risks (Elvik, 2010). The perception of road safety is shaped by the information about potential hazards in the traffic environment and road users' ability to perceive potential traffic risks (Brown & Groeger, 1988). Factors such as urban infrastructure and individual personality characteristics can cause road users to over- or underestimate road safety conditions (DeJoy, 1989; von Stülpnagel & Lucas, 2020). However, there is a noticeable gap in research regarding this misperception.

The measurement of actual road safety differs from the measurement of road safety perception. Actual road safety is typically quantified using crash frequency and severity data, along with information on traffic volume, crashes, injuries, and fatalities (Schepers et al., 2014). Traffic crash data provides details such as crash location and time, which allow for the integration of other datasets, such as built environment data, enabling detailed spatial or spatiotemporal analyses like crash hotspot analysis (Bíl et al., 2019; Imprialou & Quddus, 2019). On the other hand, studies focusing on the perception of road safety primarily rely on questionnaires or interviews that include items related to individuals' perceptions of road traffic safety. These surveys can capture rich information, including demographic characteristics of the respondents. However, these studies often have a limited geographical scope due to their high costs, preventing large-scale analysis.

2.2. Built environment and traffic safety

The built environment plays a significant role in shaping human travel patterns and, consequently, influencing traffic safety (Ewing & Cervero, 2010; Ewing & Dumbaugh, 2009; Frank & Engelke, 2001). Extensive literature has examined the effects of various built environment factors on traffic crash frequency and injury, including land use (Mathew et al., 2022), road attribute (Han et al., 2018; Høye & Hesjevoll, 2020), points of interest (POI) (Jia et al., 2018), and transportation infrastructure (Ding et al., 2020).

The features of the built environment also have significant impacts on the human perception of road safety (Han et al., 2023). For example, a high number of crossings and a lack of traffic lights and signs can increase perceived road risk (Aceves-González et al., 2020; Lee et al., 2016), while proper street lighting and spacious road can contribute to a higher perceived road safety (Park & Garcia, 2020; Rankavat & Tiwari, 2016). Other built environment features, such as land use, have also been found to be associated with the perception of road safety. A higher fraction of open land use, for instance, may lead to a greater perception of comfort and safety for road users (Aziz et al., 2018).

In urban areas, various urban environments can significantly influence the state of road traffic safety and human perceptions of it. However, there is limited research studying how the built environment affects the relationship and discrepancy between road traffic safety and human perceptions of road safety.

2.3. Applying computer vision and street view images in traffic study

The availability of high-resolution street view images (SVIs) through services like Google Street View, combined with advancements in computer vision techniques, has facilitated the widespread use of SVIs for large-scale evaluation of urban environment attributes (Biljecki & Ito, 2021; Liu et al., 2017). SVIs, typically collected by dashcams mounted in vehicles driving through roads, are increasingly utilized in transportation and traffic studies (Biljecki & Ito, 2021; Sainju & Jiang, 2020).

In traffic research, SVIs serve two main purposes. Firstly, some studies employ pre-trained image segmentation and object detection models to extract environment attributes from SVIs. These attributes are then used as independent variables to investigate various traffic-related issues. For example, Hankey et al. (2021), Liu et al. (2019) and Cai et al. (2022) utilized SVIs to characterize the urban environment and predict bicycling and walking rates, urban traffic-solar energy, and traffic crash occurrences, respectively. Kwon & Cho (2020) extracted information from SVIs such as proportions of the sky, green space, road, sidewalk, as well as the number of pedestrians and vehicles. They explored the association of these attributes with perceived crash risk measured through a risk-perception survey. Secondly, another type of studies involves training deep learning or machine learning models directly from SVIs collected in the study areas for traffic scene recognition and classification. Examples include identifying traffic signs (Campbell et al., 2019), road strips (Sainju & Jiang, 2020), and road noise barriers (Zhang et al., 2022).

SVIs, when utilized in conjunction with computer vision techniques, offer a valuable data source characterized by high spatial resolution and extensive coverage. This combination empowers us to gain deeper insights into the spatial attributes of urban environments and address various traffic-related issues more effectively.

2.4. Research gaps

Undoubtedly, research on road safety and road safety perception has yielded significant achievements. However, there still exist some limitations and gaps in the current research. Firstly, while existing studies have examined the impacts of the built environment on road safety and perception separately, few have focused on the association and discrepancy between road safety and road safety perception. Places that seem safe may not be. Moreover, there is limited knowledge regarding the factors related to the built environment that contribute to this perception mismatch. It is important to note that a road appearing safer than it actually is may hinder the proper awareness of traffic collision risks, while a road appearing more dangerous than it is may imply a potential danger of traffic crashes. Interpreting this mismatch is crucial for proactive interventions and planning. The second gap pertains to the research coverage. Most previous research on human perception of road safety has been conducted on limited roads or in specific communities due to the substantial costs associated with data collection. This limitation restricts the spatial extent available for studying distribution patterns in cities. To address these limitations, this study will measure the perception of road safety using computer vision techniques and SVIs. SVIs, combined with computer vision techniques, offer a valuable data source characterized by high spatial resolution and extensive coverage. The primary objective of this study is to interpret the mismatch between road safety perception and traffic crash rates.

3. Data and methods

3.1. Study area

Manhattan, the most densely populated borough of New York City, USA, was selected as the study area. Traffic crashes pose a significant issue in New York, as evidenced by the number of people injured in traffic crashes increasing by 11 percent from 110,250 in 2013 to 122,713 in 2019 (Shaaban & Ibrahim, 2021). Given that Manhattan serves as the urban core of the New York metropolitan area, traffic safety in this region is of paramount concern. In addition, the rich open traffic and built environment dataset in Manhattan enabled us to conduct in-depth analyses. Our empirical analysis encompassed road segments within Manhattan. The road network data utilized in the study was obtained from the NYC Street Centerline (CSCL). Specifically, roads located in Manhattan were selected for analysis. To ensure data completeness, road segments without street view images and traffic data were eliminated from the dataset. As a result, a total of 6012 road segments were retained for the study. The research period focused on was 2013–2019, excluding the effects of the COVID-19 pandemic. Road segments were chosen as the analysis units because it was easier to link traffic data, such as average annual daily traffic, to specific roads. Additionally, road attributes and facilities were expected to play crucial roles as variables in this study. Therefore, the crash data and other variables were analyzed at the segment-level instead of the zonal or intersection level.

3.2. Data

3.2.1. Place Pulse 2.0 dataset

In this study, we utilized “Place Pulse 2.0,” an online platform developed by MIT, for collecting data on urban perception ratings. The platform presented participants with pairs of SVIs randomly selected from various cities worldwide. Participants were then asked to choose between the two images or indicate if they perceived them as equal in response to evaluative questions such as “which place looks safer?” and other questions like “beautiful,” “depressing,” “lively,” “wealthy,” and “boring” (Dubey et al., 2016; Zhang et al., 2018). The dataset, available for download online (https://centerforcollectivelearning.org/urbanperception), contains SVIs along with a corresponding document, recording the ID of each image, the study question, and the choice made by the volunteers. For this particular study, we focused solely on distilling the perception of safety from the question “which place looks safer?” Based on this, we would build an SVIs database specifically for road safety perception training.

3.2.2. Street view images of Manhattan

The researchers obtained road segment data from NYC Open Data. Additionally, we created sampling points along the centerlines of the road segments at 15-meter intervals, resulting in a total of 48,505 points. To retrieve the corresponding SVIs for each sample point within the road segments, we utilized ArcGIS to calculate the coordinates of the sampling points based on the road segment data. The location and time information of these points were then used as inputs to an API, which facilitated the retrieval of SVIs (Li et al., 2018; Li & Ratti, 2019). In this study, we only downloaded SVIs from 2013 to 2019 since our study period was in this time frame.

3.2.3. Reported traffic crash and traffic volume

To represent actual road safety of each road segment, the crash rate was used in this study, which was calculated and measured by the function below (Hou et al., 2020; Zeng et al., 2017):

$$CR_i = \frac{No\_crash_i}{AADT_i \times L_i \times \frac{365}{1000,000}} \tag{1}$$

in which CR_i is the average annual crash rate/frequency per 1 million vehicle kilometers traveled (VKT) on road segment i from 2013 to 2019; No_crash_i is the average annual number of traffic crashes that occurred on road segment i; $AADT_i$ is the average annual daily traffic on road segment i; and L_i is the length of segment i.

Crash data comes from NYC Open Data Portal, while AADT comes from New York State Department of Transportation. Crash data point was aggregated at the road segment level based on longitude and latitude information.

3.2.4. Built environment data

To investigate the impact of various built environment variables, road attribute, land use, POI and traffic-related facility were considered. Furthermore, visual indexes extracted from SVIs was also included, along with four categories of variables constituted the five aspects of built environment variables examined in this research. Table 1 shows the summary of the research variables.

We extracted and calculate visual indexes representing the percentage that the proportion of pixels occupied by specific visual elements in an SVI, to reflect eye-level urban design quality (Su et al., 2023). The panoptic segmentation technique was applied to realize the above extraction and calculation. Panoptic segmentation unifies the typically distinct tasks of semantic segmentation (assign a class label to each pixel) and instance segmentation (detect and segment each object instance) (Kirillov et al., 2019). It detects countable objects and uncountable regions concurrently, resulting in a comprehensive representation of images (Liu et al., 2023). The panoptic segmentation model we employed was Panoptic Feature Pyramid Network (FPN) and from the Detectron2, a popular AI library for computer vision tasks in the GitHub developed by the Facebook (Wu et al., 2019). The box average precision and mask average precision of the model can achieve 42.4 % and 38.5 % on the COCO dataset, respectively.

Road width and type were considered here in terms of road attributes. Road type was set as a dummy variable, to reflect the impact of highways and general roads on road safety.

Land use, another important variable in the built environment, was included in the analysis. The Primary Land Use Tax Lot Output (PLUTO) data from NYC Open Data was obtained, and the 11 original land use categories were transformed into 6 categories: residential land, commercial land, industrial land, transportation land, public land, and open space, shown in Table 2. The percentage of each land use type within a 500 m buffer of each road segment was calculated and recorded. To capture the level of land use mixture, the entropy of the six land use types was calculated (Ding et al., 2018).

POI data was also collected from NYC Open Data. To consider the nearby environments and the impact of distant surroundings on road safety, a buffer-based method was employed to calculate the density of various types of POIs within a 500 m buffer of each road segment (Luo et al., 2022; Wang et al., 2023). The study considered 13 types of POIs from Point of Interest (CommonPlace) dataset in New York, including residential, education facility, cultural facility, recreational facility, social services, transportation facility, commercial, government facility, religious institution, health services, public safety, water, and miscellaneous. The density of each POI type was recorded for analysis. We also

Table 1
Summary of the built environment variables.

Class	Description	Data	Source	Variable numbers
View indices	Proportion of sky, tree, road, pavement, building…… pixels to the whole image	Street View Images	Google Street View API	12
Road attribute	Road width	Road network polyline	New York open data	1
	Road types (Highway or not)			1
Land use	6 types of land use area in the buffer of each road segment	Land use polygon		6
	Mixed land use level			1
POI	Different POI numbersin the buffer of each road segment	POI point		13
	Road distance to the closest POI			13
	POI types			1
Traffic-related facility	Different numbersin the buffer of each road segment	Facility point	Geofabrik	5
	Road distance to the closest facility			5

Table 2
Land use category.

PLUTO classes	Reclassify
One &Two Family Buildings	Residential land
Multi-Family Walk-Up Buildings	
Multi-Family Elevator Buildings	
Mixed Residential & Commercial Buildings	Commercial land
Commercial & Office Buildings	
Industrial & Manufacturing	Industrial land
Transportation & Utility	Transportation land
Parking Facilities	
Public Facilities & Institutions	Public land
Open Space & Outdoor Recreation	Open space
Vacant Land	

generated distance features by calculated the distance from each road segment to the nearest POI of each type. Additionally, the study counted the number of POI types within the 500 m buffer of each road segment to reflect the diversity of facilities in the vicinity.

The study also considered important traffic facilities and characteristics, such as crossings, bus stops, stop signs, junctions, and traffic signals. Point data for these features were collected from Geofabrik, and a buffer-based method was also used to generate density and distance features for these variables. In total, the study prepared 58 independent variables related to the built environment. The data for all variables accessed within 2013–2019.

3.3. Methodology

In this paper, we proposed three questions to guide our thinking:

- Can we measure the human perception of road safety at a comparable spatial resolution to actual traffic crash rates, while achieving a high spatial coverage?
- How can we quantitatively measure the association/disparities between road safety perception and actual traffic crashes at the street level?
- If a discrepancy exists between perceived and actual road safety, what are the specific built environment factors that contribute to this misperception?

The methodology framework, as depicted in Fig. 1, employed a stepwise approach to address the research questions. Here is a breakdown of each step:

Step 1: SVIs from the Place Pulse 2.0 dataset were selected to build our dataset for training a deep learning model. The fitted model was then used to estimate road safety perception probability and score using SVIs from the study area.

Step 2: To quantitatively measure the association between road safety perception scores and actual crash rates, correlation analysis and regression analysis were conducted. Moreover, by comparing the road safety perception scores with the crash rates, patterns of road safety perception could be identified.

Step 3: The influences of the built environment on the patterns of

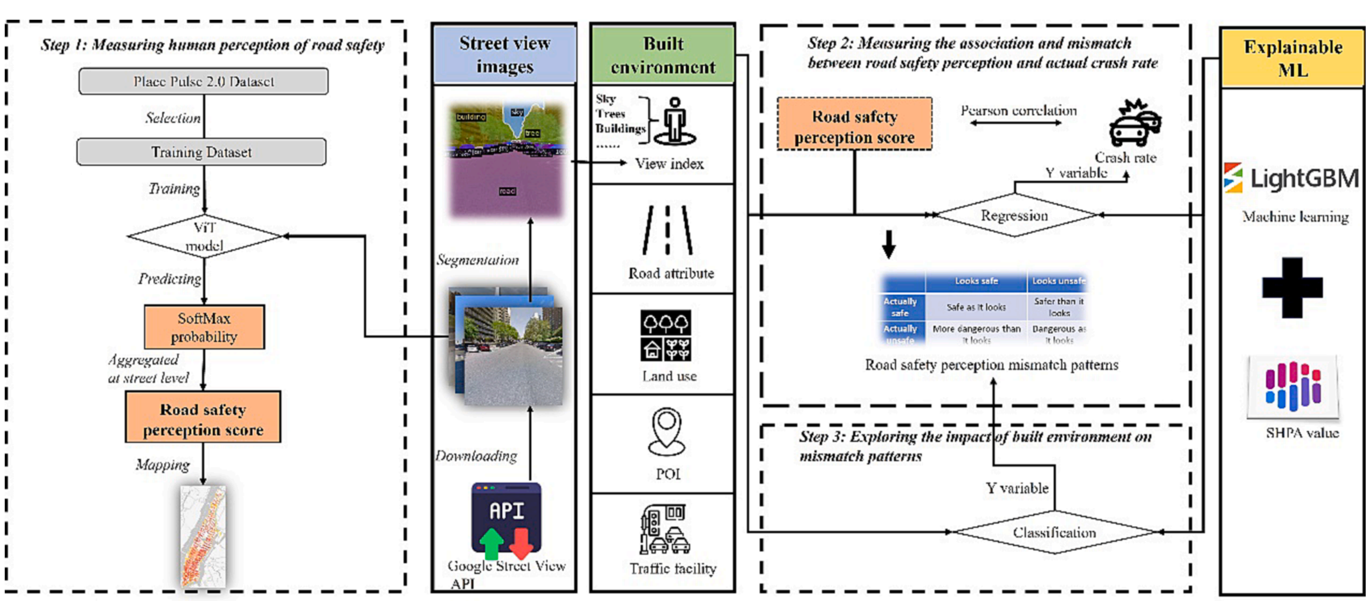

Fig. 1. Methodology framework.

perception mismatch were examined. In this step, road safety perception patterns were used as the dependent variable for classification. Explainable machine learning techniques were employed to explore the impacts of different built environment variables on the overall classification task as well as on specific classes.

SVIs were collected in the study area for two distinct purposes. One is to obtain the perception score, the other is to get the view index of various built environment objects, as an important aspect of built environment, along with other variables such as road attributes, land use, POI, and traffic facilities. Explainable machine learning method will be used in this study for regression and classification tasks.

3.3.1. Measuring road safety perception with a deep learning model

In this study, we aim to reconstruct road safety perception as an image classification problem using SVIs and deep learning techniques. The training of the deep learning model involves in several key elements, such as training data, labels, model architecture and so on. Finally, the trained model would be used for predicting perception scores in the study area.

3.3.1.1. Training dataset. As we mentioned, the Place Pulse 2.0 SVIs dataset was utilized for training our model. To ensure that the choice questions presented to volunteers specifically addressed “traffic safety” rather than other safety concepts, we retained only those choice records where both the left and right images depicted scenes related to traffic. Panoptic segmentation technique was used to recognize different typical traffic features, such as roads, pavements, cars, buss, and trunks. This technique allowed us to recognize these elements and calculate their proportion of pixels in the overall image. Here, images with less than 30 percentage of pixels related to main traffic elements (roads, pavements, cars, buses, and trunks) were eliminated from our image training process, shown in Fig. 2.

3.3.1.2. Labels. Given that each choice record entails a comparison between two images, our objective was to transform the results of these paired comparisons into labeled single-samples, enabling the prediction of road safety perception in SVIs. Following the methodology outlined in the previous study (Zhang et al., 2018), we computed the Q-score using the pairwise comparison records. Specifically, for each perception, the win (W_i) and the loss (L_i) ratios of image i are defined as:

$$W_i = \frac{w_i}{w_i + l_i + e_i}; L_i = \frac{l_i}{w_i + l_i + e_i} \tag{2}$$

where w_i, l_i, e_i are number of times that image i wins, loses or equals to its paired image. Q-score ranging from 0 to 10 for each image i is defined as:

$$Q_i = \frac{10}{3}\left(W_i + \frac{1}{w_i}\sum_{j_1=1}^{w_i} W_{j_1} - \frac{1}{l_i}\sum_{j_2=1}^{l_i} L_{j_2} + 1\right) \tag{3}$$

where j_1 and j_2 denote images lose to or win image i in the comparison. Equation (3) simply corrects an image win ratio (W_i) by incorporating the average win ratio of the images it prevailed over and subtracting the loss ratio of the images that were chosen over image i. A prerequisite for generating reliable and consistent Q-scores is ensuring that each image’s perception score is derived from a minimum of 5 pairwise comparisons.

To mitigate the influence of human perception bias and avoid introducing noise, we adopted a strategy of selecting representative positive (safe) and negative (unsafe) samples from the entire dataset for the training task (Wei et al., 2022; Zhang et al., 2018). In this context, positive means Q_{high}, while negative means Q_{low}. The equation is as follows:

$$Q_{\mathrm{high}} = \overline{Q} + \sigma, Q_{\mathrm{low}} = \overline{Q} - \sigma \tag{4}$$

where $\overline{Q}$ is the average Q-score of all images concerning safety and σ denotes standard deviation. Therefore, we transformed the labels into binary category. After that, the number of SVIs with labels is 16,820 in the dataset, with 8,502 positives and 8,318 negatives.

Fig. 2. Images in the Place Pulse 2.0: (a) SVIs reflecting traffic scene; (b) SVIs not reflecting traffic scene and thus eliminated from image data training.

3.3.1.3. Model architecture. In order to measure perceived road safety, a deep learning model capable of recognizing and interpreting visual content with a deeper understanding of underlying traffic safety attributes was desired. For this purpose, the Vision Transformer (ViT) was selected as the base pre-trained model for transfer learning. ViT is a cutting-edge deep learning model architecture that utilizes the Transformer architecture, initially designed for natural language processing tasks, in the field of computer vision (Dosovitskiy et al., 2020). The model architecture, as depicted in Fig. 3, treats an image as a sequence of patches and employs self-attention mechanisms to capture global dependencies between these patches. This differs from traditional convolutional neural networks (CNNs) commonly used in computer vision tasks. For transfer learning, we adopted the pre-trained model and froze the parameters of the transformer encoder. We then fine-tuned a 3-layer MLP (Multi-Layer Perceptron) to adapt the model to our specific task.

3.3.1.4. Model evaluation. Categorical cross-entropy loss (L) and classification accuracy (Acc) are utilized to evaluate the model. They are calculated as follows:

$$L = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{5}$$

$$Acc = \frac{T}{N} \tag{6}$$

where $\hat{y}_i$ denotes predicted class probabilities; y_i represents ground-truth class probabilities; n represents the total number of classes; T represents the number of images that are correctly classified, and N represents the total number of images in the dataset. A lower L and a higher Acc indicate better model performance.

3.3.1.5. Model prediction. In the prediction process, the trained model was fed with SVIs from the study area. Similar to the training phase, SVIs that are not related to traffic scenes were removed before performing the inference process. This ensures that the predictions generated by the model are focused on traffic safety perception. We applied the softmax function to calculate the probability for predicting a certain class. The probabilities of each image classified into one class serve as indicators of human perception. By quantifying subjective human perception through the probability values, we could explicitly measure and quantify the perception of road safety in the study area. For SVIs labeled as negative, we took the negative sign of the probability values. This transformation reflects the unsafe perception associated with those images. On the other hand, for SVIs labeled as positive, we retained the probability values as the perceptual scores for each image.

3.3.2. Measuring the association and misperception with the explainable machine learning method

The advent of the big data era has led to the recognition of the superiority of machine learning methods over traditional statistical approaches, offering improved prediction performance (Ma et al., 2020a, Ma et al., 2020b). This has prompted the utilization of machine learning methods in various studies related to traffic crashes (Jiang & Ma, 2021; Santos et al., 2022). In this study, we employed the LightGBM machine learning method, which was introduced by Microsoft in 2017 (Ke et al., 2017). LightGBM has demonstrated high computational efficiency and prediction accuracy compared to other machine learning methods in previous research (Jiang et al., 2022; Wen et al., 2021). LightGBM was applied for both regression and classification tasks. Regression was employed to investigate the relationship between perception scores and crash rates, with crash rate as the dependent variable. In the classification model, the dependent variable consisted of different road safety mismatch types, including “Safer than it looks,” “Safe as it looks,” “Dangerous as it looks,” and “More dangerous than it looks” road segments.

While machine learning methods have shown promising results in modeling non-linear problems, they are often criticized for their lack of interpretability, often referred to as “black box” methods. To address this concern and quantify the impacts of built environment variables on different road safety misperception patterns, an explainable machine learning model is necessary. SHAP (SHapley Additive exPlanation) is a game-theoretic approach used to interpret the performance of machine learning models. SHAP employs an additive feature attribution method, where the output model is defined as a linear combination of input variables (Lundberg & Lee, 2017). SHAP can be integrated with various machine learning models, including LightGBM, using Python scripts.

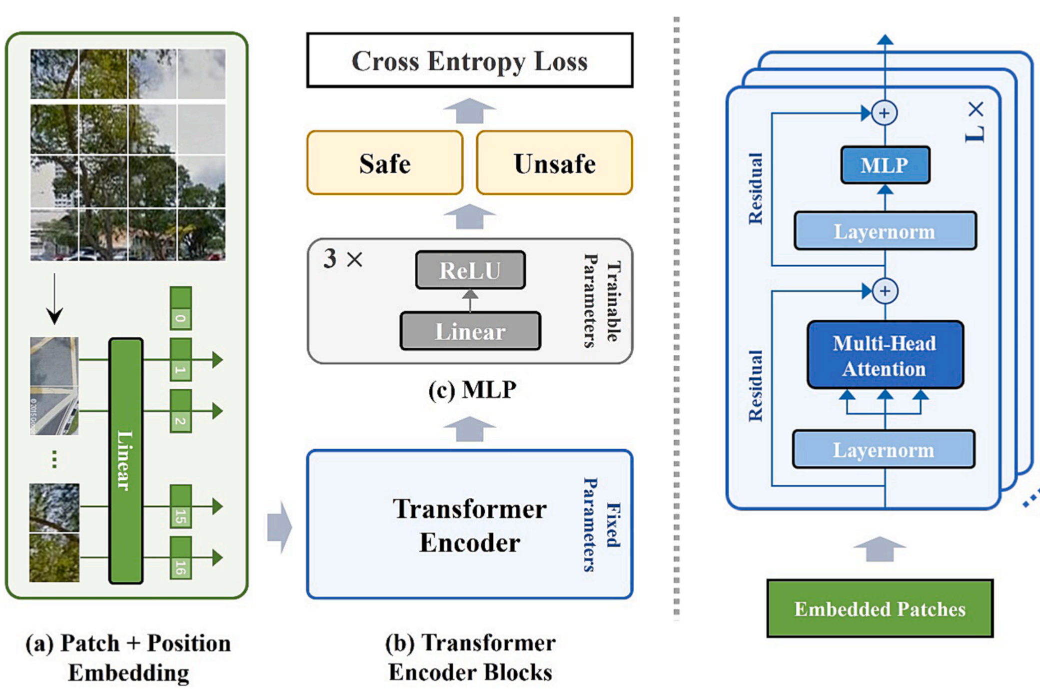

Fig. 3. Simplified ViT-based model architecture.

This integration enables the application of explainable machine learning techniques to analyze the relationships among built environment variables, crash rates, and road safety perception.

4. Results

4.1. ViT model training experiments and results

The hyperparameters used to train the ViT-based model are outlined in Table 3. To facilitate the training and validation of the model, the dataset was split. The test size of 0.3 indicates that 70 % of the data was allocated for training, while the remaining 30 % was reserved for validation. During training, the maximum number of epochs was set to 150, determining the number of complete passes the model makes over the training dataset. Early stopping was implemented to save training time and prevent overfitting. If the loss does not decrease for 40 consecutive epochs, the training is stopped early. The optimizer used was Adam, an optimization algorithm that adapts the learning rate during training. Additionally, we specified a batch size of 32, determining the number of samples processed by the model in each iteration. The learning rate was set to 1.00E-05, governing the magnitude of weight adjustments based on the loss gradient during training. A weight decay of 1.00E-02 was applied to reduce the complexity of the model and control overfitting.

To evaluate the performance of the ViT model, a famous CNN-based model, specifically Resnet50 (He et al., 2016), was chosen as the benchmark model to compare the training results. To assess the training and validation progress, we visualized the training loss and accuracy curves. Fig. 4 **(a)** depicted the results of the Resnet, while Fig. 4 **(b)** displayed the training loss and accuracy curves for the ViT-based model. By analyzing these curves, we can better understand how the model's performance evolves over the course of training.

Observing the loss curves, we noticed that in the initial epochs, the loss decreases significantly, indicating that the model is quickly learning to make better predictions. It is worth noting that the training loss decreases more rapidly than the testing loss, suggesting that the model is effectively learning from the training data but may not generalize as well to unseen data. As the training progresses, the rate of loss reduction slows down, eventually reaching a relatively stable level. This indicates that the model is gradually converging towards an optimal solution, refining its predictions. The accuracy curves exhibit a opposite trend to the loss curves. Initially, there is a notable increase in accuracy within the first few epochs, indicating that the model is learning to make more accurate predictions. However, as the training continues, the accuracy curves may exhibit fluctuations but generally stabilize at a certain level.

To compare the evaluation results of the ViT-based model and the benchmark Resnet-based model comprehensively, Table 4 shows the best accuracy and loss values achieved by both models. From the table, it can be observed that the ViT-based model outperforms the Resnet-based model in terms of accuracy (higher accuracy) and loss (lower loss). These results indicate that the ViT model better fits the dataset and demonstrates superior performance. As a result, the ViT model was selected for the prediction of SVIs in the study area.

Table 3
Hyper-parameters settings.

Hyper-parameters	Setting
Epoch	150
Early stop	Yes
Batch size	32
Test size	0.3
Learning rate	1.00E-05
Weight decay	1.00E-02
Optimizer	Adam

4.2. Measuring and mapping perception of road safety

Fig. 5 displays examples of perceived safe and unsafe SVIs after model inference. The top three images are predicted to be perceptually safe, and the bottom three images are classified perceptually unsafe. The figure also includes the results of panoptic segmentation, which involves the segmentation of elements such as trees, roads, and the sky in the images. Additionally, traffic objects like traffic signals and vehicles are identified and assigned confidence probabilities. This segmentation process enables the computation of various view indexes that represent the eye-level built environment.

To assess road safety perception at the road segment level, the perception scores of SVIs corresponding to a particular road segment were aggregated by calculating the average score. Table 5 presents the characteristics of road safety perception scores in the study area. When considering individual SVIs, approximately 70 % of the images were recognized as safe, while around 30 % of the images were recognized as unsafe. When examining road segments, nearly 75 % of the road segments were recognized as safe, while approximately 25 % of the road segments were recognized as unsafe.

Road safety perception scores can be visualized using ArcGIS Pro, as shown in Fig. 6 **(a)**. The map illustrates the spatial heterogeneity of human perception of road safety in Manhattan. The roads perceived as unsafe are predominantly located in the urban periphery. On the other hand, roads with a perceived lower traffic risk are distributed in the inner-city area, although there is a scattered distribution of low traffic safety perception. To gain further insights, a comparison can be made between the distribution pattern of perceived road safety and the actual road crash rate in Manhattan. Fig. 6 **(b)** displays the distribution of actual crash risk. Interestingly, the actual crash risk does not exhibit the same high concentration in the urban fringe areas as perceived road safety suggests. This discrepancy between perceived and actual traffic risk sparked our interest in studying the mismatch between the two.

4.3. Measuring the association between road safety perception scores and actual crash rates

To investigate the association between road safety perception scores and actual crash rates in Manhattan, we began by calculating the correlation coefficient between the two variables. Before performing the correlation analysis, a logarithmic transformation was applied to the crash rate variable. The study unit remained as each road segment. The Pearson correlation coefficient between the perceptions of road safety and crash rates was determined to be 0.503, which is significant at the 0.01 level. This indicates a strong positive correlation between perceived road safety scores and actual crash rates. In simpler terms, roads that are perceived as safer tend to have higher crash rates.

To further explore the relationship between perception scores and crash rates, a regression analysis was performed. The crash rate after logarithmic transformation was used as the dependent variable, while the perception score and other built environment features served as independent variables. The machine learning model LightGBM was employed for the regression analysis, and its performance was compared with traditional ordinary least squares regression (OLS) and other machine learning methods, including ridge regression, random forest (RF), and gradient boosting decision tree (GBDT). Cross-validation (CV) with 5-fold CV was implemented to prevent overfitting. The regression metrics used were mean squared error (MSE) and R2. A higher R2 indicates that the model explains a larger proportion of the variance in the data, while a lower MSE indicates that the model has smaller errors and is more accurate in predicting the target variable. Table 6 presents the results of the regression analysis. Two main findings can be observed. Firstly, the LightGBM model outperformed other models, as evidenced by its highest R2 and smallest MSE. Secondly, when comparing the five pairs of models, the models incorporating the traffic safety perception score as an independent variable demonstrated higher R2 and smaller

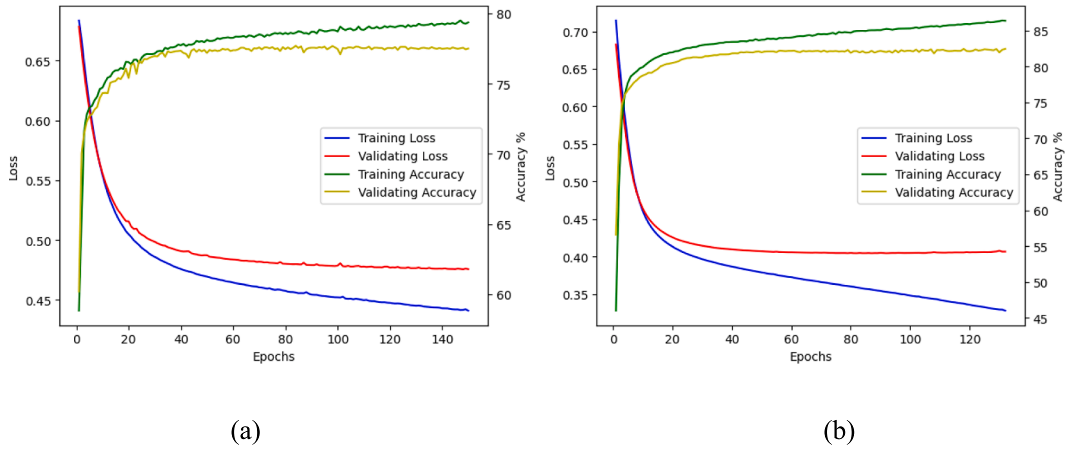

Fig. 4. Accuracy curve and loss curve of two models: (a) Resnet-based model; (b) ViT-based model.

Table 4
Best results of two deep learning models.

Model	Accuracy	Loss	Early stop at epoch
Resnet-based	77.626 %	0.476	149
ViT-based	82.442 %	0.404	132

MSE compared to the models without the traffic safety perception score. This suggests that by including the perception score of traffic safety in the models, the fit with the data is improved.

To assess the importance of the road safety perception score in explaining the crash rate and determine its positive or negative impact, SHAP values were employed following the regression analysis. Fig. 7 illustrates the global importance factor of the input variables, which is estimated as the average of absolute SHAP values per feature across the data. The input variables are ranked based on their importance for the crash rate, with higher mean SHAP values indicating greater

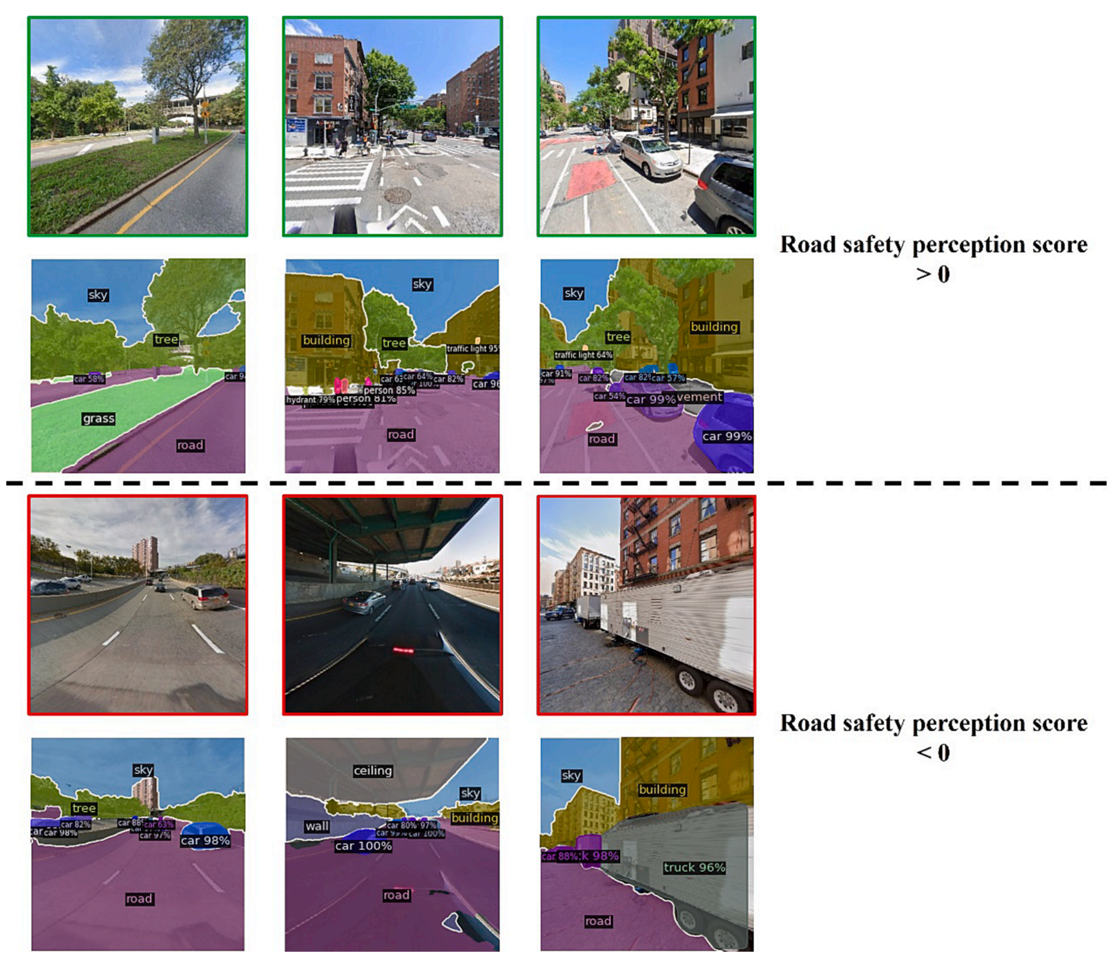

Fig. 5. Perceived safe/unsafe image samples in the study area and the corresponding results of image panoptic segmentation.

Table 5
Descriptive statistics of SVIs.

	Count	Safe	Unsafe	Mean	Std	Max	Min	Median
Perception score of each SVI	436,308	69.704 %	30.296 %	0.312	0.699	0.995	−0.999	0.684
Perception score of each segment*	6,012	74.900 %	25.100 %	0.269	0.529	0.949	−0.980	0.494

Notes: *perception score of each segment was obtained by calculating the average scores of all SVIs on the road segment.

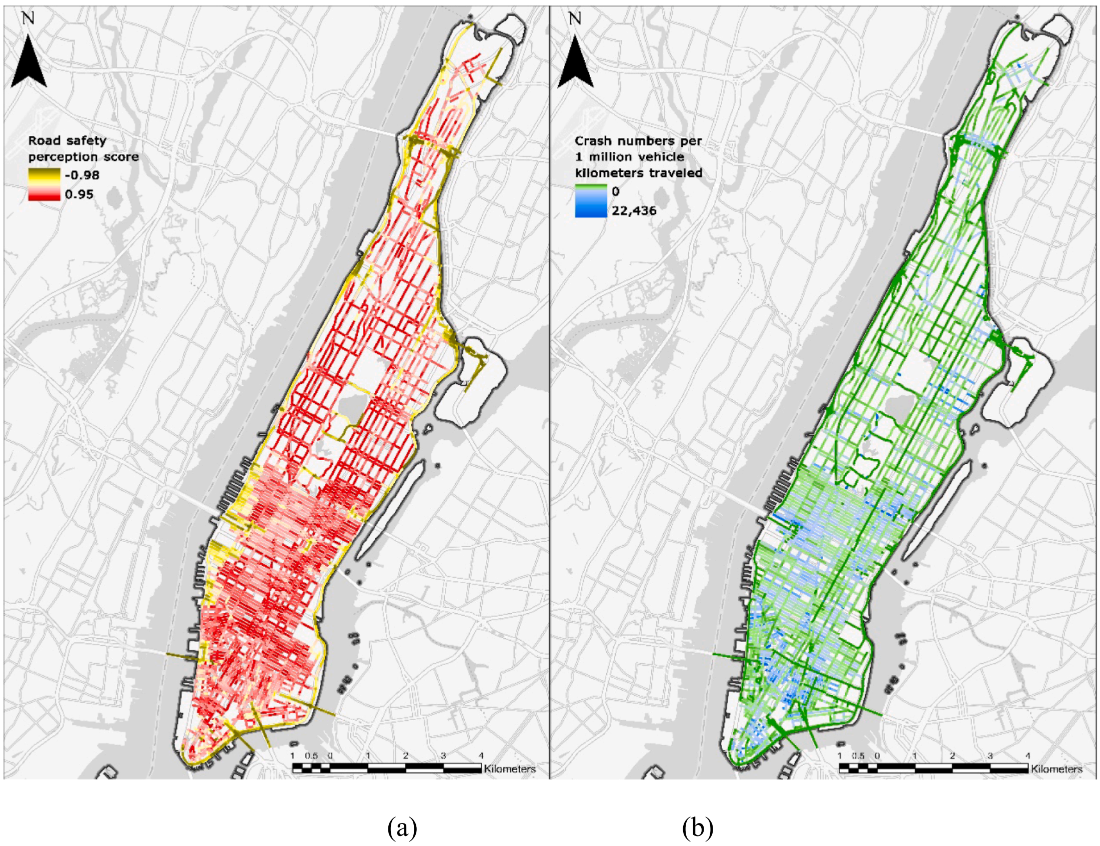

Fig. 6. Distribution of road safety perception scores and crash rates.

Table 6
Regression result of crash rates.

Model	Whether corporates perception score (YES/NO)	R2	MSE
OLS	NO	40.506 %	2.831
	YES	40.654 %	2.824
Ridge	NO	40.281 %	2.841
	YES	40.466 %	2.831
RF	NO	58.822 %	1.925
	YES	59.108 %	1.923
GBDT	NO	53.350 %	2.215
	YES	53.915 %	2.187
LightGBM	NO	63.049 %	1.752
	YES	63.332 %	1.736

importance. From the figure, it is evident that the road safety perception score holds the second position in terms of importance, just behind the variable “distance to the nearest traffic signal” (distance_poi_traffic_traffic_signals). This finding confirms that the road safety perception has a critical impact on actual crash rates. Additionally, the figure reveals that a relatively high road safety perception score (indicated by the red color) contributes to a positive impact on the crash rate, meaning it increases the crash rate. Conversely, a relatively low score (indicated by the blue color) contributes to a negative impact on the crash rate, reducing it. This analysis highlights a significant mismatch between perceived road safety and actual road safety in Manhattan.

4.4. *Exploring the mismatch between road safety perception and actual crash risk*

To explore the misperception between perceived and actual road safety, we categorized all road segments into four perception classes/patterns based on their characteristics. The categorization criteria are as follows:

Class 0 “Safer than it looks” road segments - Road safety perception score is less than zero, but the crash rate is less than the top 25 % of all segments.

Class 1 “Safe as it looks” road segments - Road safety perception score is more than zero, and the crash rate is less than the top 25 % of all segments.

Class 2 “Dangerous as it looks” road segments - Road safety perception score is less than zero, and the crash rate is larger than 75 % of all segments.

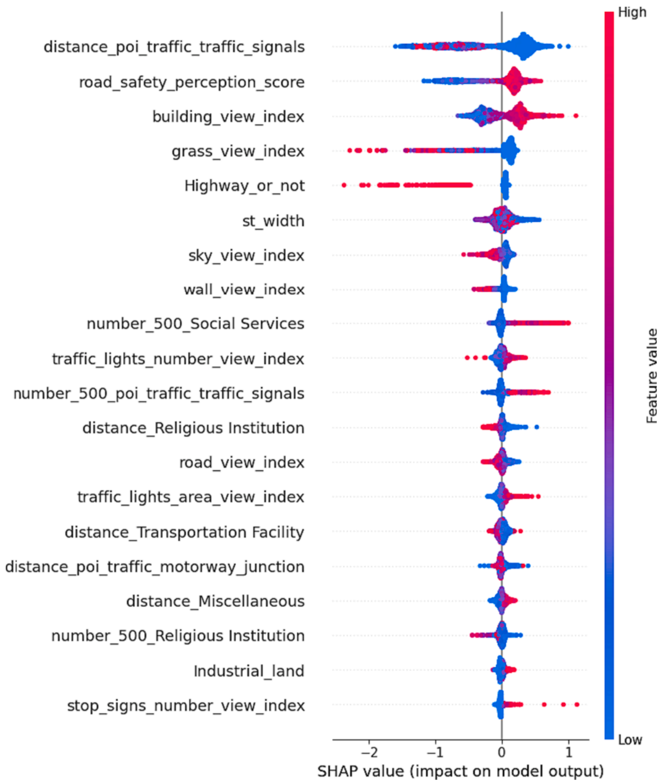

Fig. 7. SHAP value of global importance factors on explaining crash rates.

Class 3 "More dangerous than it looks" road segments - Road safety perception score is larger than zero, and the crash rate is larger than 75 % of all segments.

To differentiate between safe and unsafe road segments, the upper quartile of crash rates was used. Additionally, the distribution of road safety perception scores greater than zero and less than zero was close to 75:25.

Table 7 presents the number of road segments in each safety misperception pattern based on the classification. The majority of segments fall into the "Safe as it looks" class, indicating that the perceived and actual road safety align relatively well for these segments. There are only 200 segments in the "Dangerous as it looks" class. The classes with more pronounced safety perception mismatches are "Safer than it looks" and "More dangerous than it looks," encompassing 1,309 and 1,303 segments, respectively.

Mapping the road safety misperception can provide visual insights into the spatial distribution of these patterns. Fig. 8 illustrates the spatial distribution of the four safety perception classes. We can see that "Safer than it looks" primarily located in the urban fringe of Manhattan. "More dangerous than it looks" segments are concentrated in southern part of Manhattan. "Safe as it looks" roads occupy most of the space in the city. "Dangerous as it looks" segments are found in a few concentrated areas in the central region of Manhattan.

Table 7
Road segments number of different safety perception patterns.

	Looks safe	Looks unsafe	Total
Actually safe	Safe as it looks(3,200)	**Safer than it looks (1,309)**	4,509
Actually unsafe	**More dangerous than it looks(1,303)**	Dangerous as it looks (200)	1,503
Total	4,503	1,509	6,012

4.5. Impact of built environment in road safety perception patterns

To investigate the contribution of different built environment characteristics to the various mismatch patterns, a classification task was performed using the LightGBM model. The label for this machine learning method represents the four classes of road safety perception patterns, while the features consist of 58 built environment variables. Table 8 presents the classification results obtained from the LightGBM model. The accuracy of the model is approximately 0.8, indicating a reasonably high level of correct predictions. The weighted-average precision, recall, and F1-score are 0.79, 0.80, and 0.78, respectively. These metrics demonstrate the overall performance and effectiveness of the model in classifying the different perception patterns based on the built environment characteristics.

SHAP value was then used for model interpretation. Fig. 9 displays the rank of the top 20 global importance factors, indicating the variables that contribute the most to the model's predictions. The top 10 variables, in order of importance, are: wall view index, tree view index, building view index, distance to the nearest traffic signals, street width, land use mix, distance to the nearest road junction, proportion of transportation land area, pavement view index, and distance to the nearest religious institution.

Furthermore, summary plots were generated to illustrate the range and distribution of the impacts of built environment variables on the prediction of each road safety perception mismatch type. Fig. 10 highlights the influential variables for each perception type, which may differ slightly from the top global importance factors. It is evident that the wall view index, tree view index, and building view index are three significant factors influencing road safety perception. A lower wall view index, higher tree view index, and higher building view index are more likely to be categorized as "Looks safe" classes ("Safe as it looks" and "More dangerous than it looks"), while a higher wall view index, lower tree view index, and lower building view index are more likely to be categorized as "Looks dangerous" classes ("Safer than it looks" and "Dangerous as it looks").

On the other hand, street width and distance to the nearest traffic light emerge as critical factors influencing actual road safety. A wider street and being farther away from traffic lights are more likely to be categorized as "Actual safe" classes ("Safer than it looks" and "Safe as it looks"), whereas a narrower street and closer proximity to traffic lights are more likely to be categorized as "Actual dangerous" classes ("Dangerous as it looks" and "More dangerous than it looks").

Based on Fig. 8 and the description provided, it appears that the peripheral areas of Manhattan have fewer views of trees and buildings but more views of walls. These areas are farther away from dense traffic signal locations, where the accident rate tends to be higher due to heavy vehicle and pedestrian traffic. As a result, many road segments in these peripheral areas may be "Safer than it looks." While the central region of Manhattan, which comprises the city center, generally has wider roads with more treescapes. These areas are labeled as "Safe as it looks" because the visual perception aligns with the actual safety level. In specific concentrated areas within the central region of Manhattan, labeled as "Dangerous as it looks," there is visibility of many walls and pavements, a high degree of mixed land use, and proximity to road junctions. These factors contribute to a higher risk of accidents in these areas, making them visually perceived as dangerous as well. The southern part of Manhattan is characterized by concentrated segments labeled as "More dangerous than it looks." These areas have a significant number of signals and buildings visible, but the close proximity to the signals combined with the narrowness of the roads makes them more susceptible to accidents, thereby being "More dangerous than it looks."

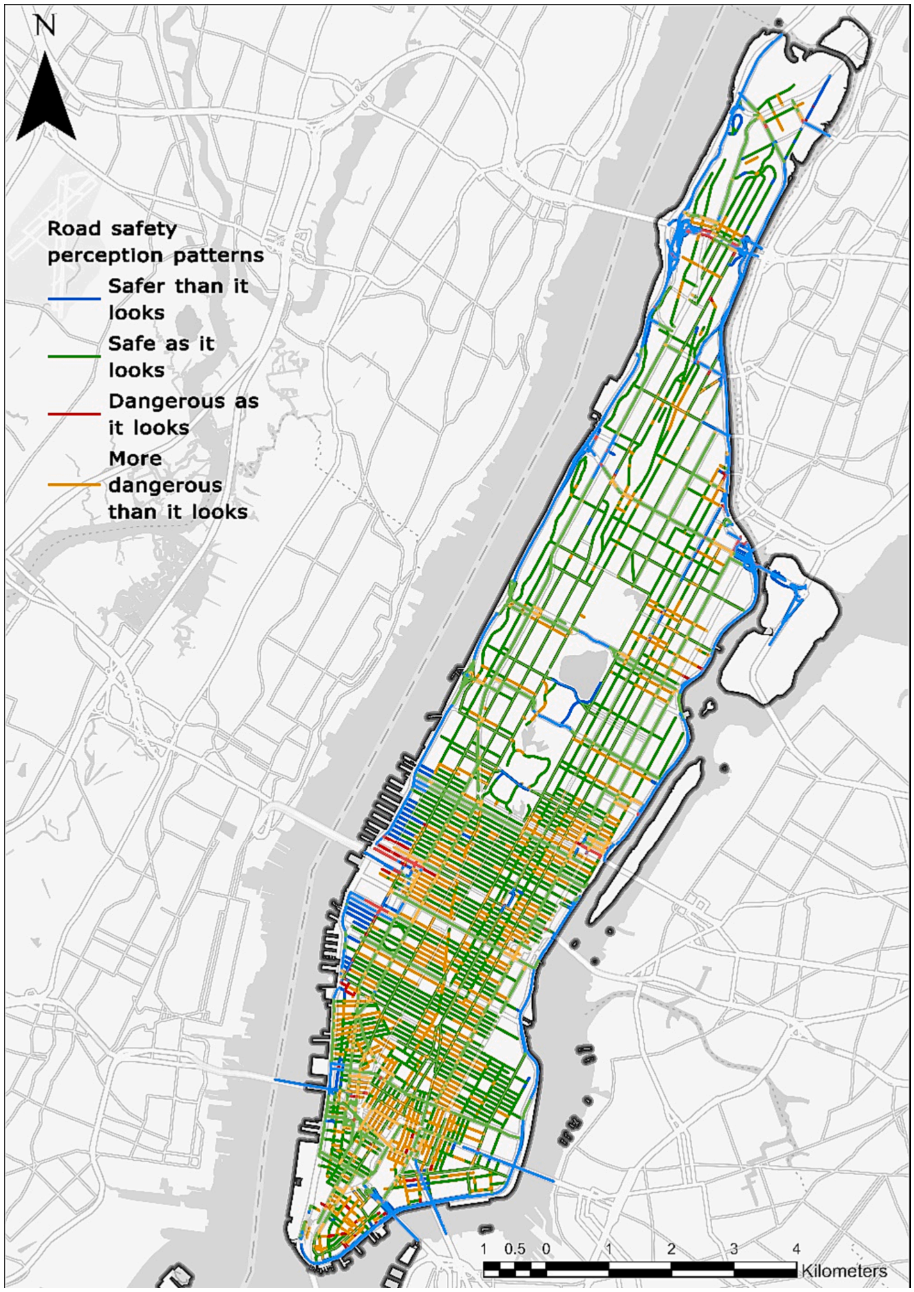

Fig. 8. Distribution of different road safety perception patterns.

Table 8
The result of LightGBM model.

Metrics	Precision	Recall	F1-score	Accuracy
Weighted-average	0.79	0.80	0.78	\
Accuracy				0.80

5. Discussions

5.1. Mismatch between road safety perception and actual road safety

The presence of a mismatch between perceived road safety and actual road safety, as observed in our results, is a notable finding. Interestingly, similar phenomena have been identified in previous studies, and researchers have offered plausible explanations to shed light on this discrepancy.

One explanation is referred to as "risk avoidance." Cho et al. (2009), in their study on pedestrian and bicyclist safety in Montgomery County, MD, near Washington, DC, found a mismatch between perceived and actual safety. They proposed that road users may actively avoid areas with a high-perceived risk based on their perception of safety and the number of trips taken by the participants. In other words, individuals tend to steer clear of road segments they perceive as unsafe, resulting in lower crash rates in those areas.

Another explanation is termed "caution enhancement." Rather than avoiding dangerous roads altogether, individuals may exhibit heightened caution and adopt safer behaviors when navigating road segments perceived as having high traffic risks (Dinh et al., 2020; Edwards, 1999). This cautious approach may involve reducing driving speed, avoiding traffic violations, and maintaining a higher level of attentiveness. As a result, the crash rates in these segments may decrease due to the enhanced caution exercised by road users (Boua et al., 2022; Schneider

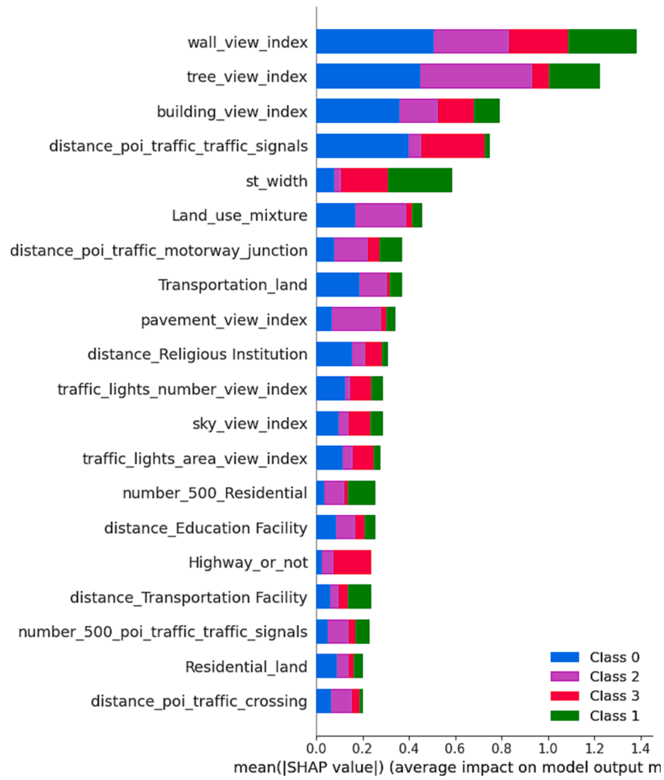

Fig. 9. SHAP value of global importance factors on explaining road safety patterns.

et al., 2004).

Both explanations provide plausible insights into understanding the existence of the mismatch between perceived and actual road safety. These explanations suggest that individuals' perceptions of safety influence their behavior on the road, contributing to the observed differences between perceived and actual road safety outcomes.

5.2. *"More dangerous than it looks" road segments: Traffic lights may deceive your eyes*

In order to gain further insights into the road sections and their associated built environment variables, an exploration of the SHAP values can be conducted. Fig. 10 **(d)** specifically highlights the factors contributing to the "More dangerous than it looks" pattern. According to the figure, the distance to the nearest traffic signal emerges as the most influential factor responsible for the "More dangerous than it looks" pattern. An increase in the distance to the nearest traffic signal corresponds to smaller SHAP values for this pattern and its associated probability. Additionally, the traffic lights number view index (Perceived number of traffic lights in the SVI) and area view index (Perceived proportion of traffic light pixels in the SVI) also play significant roles in explaining this bias. Higher values of the traffic lights number view index and area view index increase the susceptibility to the "More dangerous than it looks" pattern.

Fig. 11 displays SHAP dependence plots, which provide a quantitative analysis of how individual variables influence the model's output. In the plots, the horizontal axis represents the values of the independent variable being investigated. The vertical axis corresponds to the SHAP value, which indicates the contribution of the specific independent variable to the model's output. Fig. 11 **(a)** shows that when there is a traffic light on the road (distance value close to 0 in the horizontal axis), there is a high probability (SHAP value greater than 0) of the roadway falling into the "More dangerous than it looks" category. Conversely,

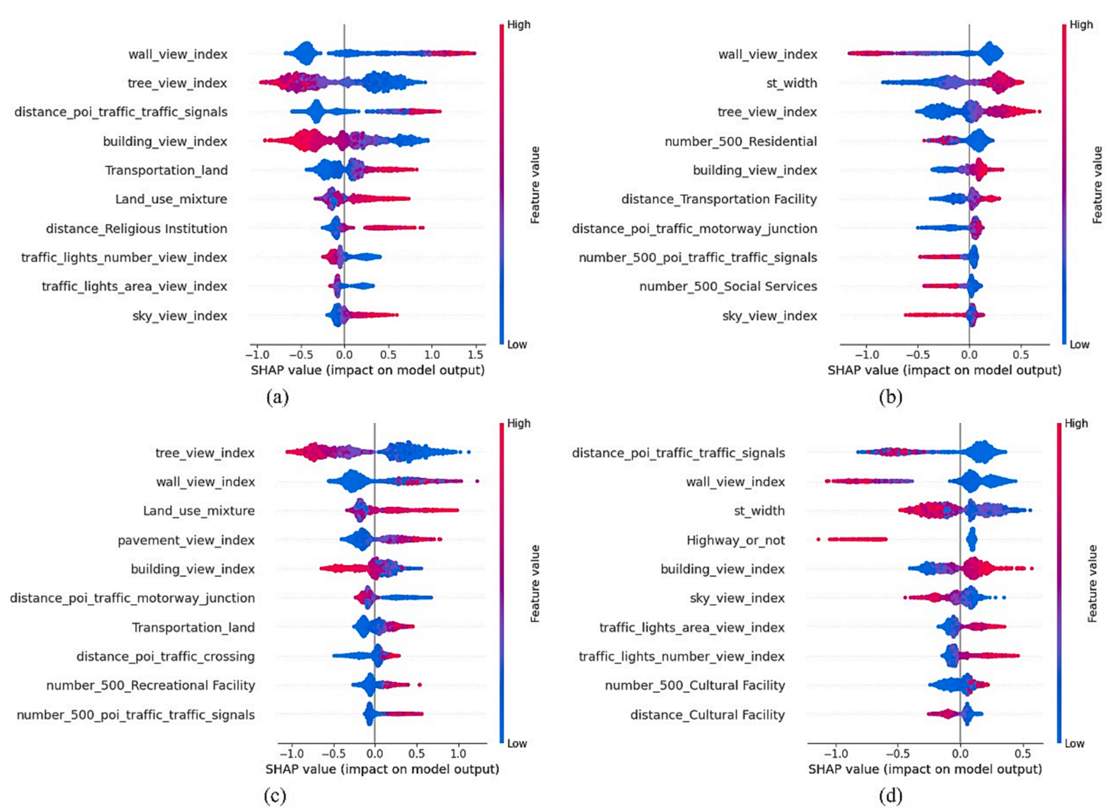

Fig. 10. SHAP summary plots on explaining various road safety perception patterns: (a) "Safer than it looks"; (b) "Safe as it looks"; (c) "Dangerous as it looks"; (d) "More dangerous than it looks".

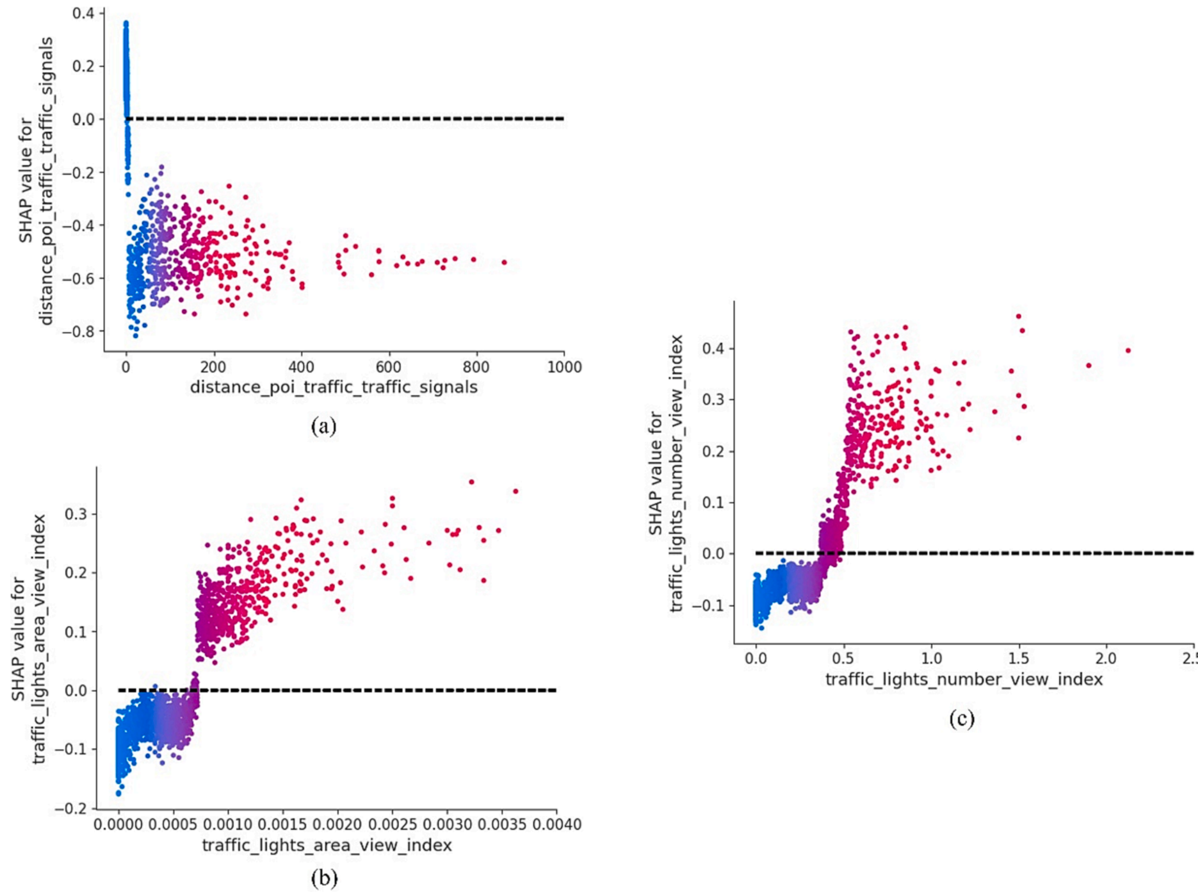

Fig. 11. SHAP dependence plots for traffic light elements in "More dangerous than it looks" roads: (a) distance to the nearest traffic signal; (b) traffic lights area view index; (c) traffic lights number view index.

when the road is farther away from traffic lights, it is almost never categorized as such. Fig. 11 **(b)** and **(c)** indicate that streets with a traffic light area view index below approximately 0.0007 or a number view index below 0.4 are unlikely to be categorized as "More dangerous than it looks." However, once these threshold values are exceeded, the probability of the roadway being classified as "More dangerous than it looks" increases greatly.

These results suggest that traffic lights have the effect of making roads look safer, but road segments in close proximity to traffic lights may actually be more dangerous. An increased number of traffic lights is associated with a reduced perception of traffic danger, while a lack of traffic lights may improve perceived safety risk (Aceves-González et al., 2020; Rothman et al., 2015). However, traffic lights are typically located in areas with high pedestrian volumes, which can increase the risk of collisions (Gu & Peng, 2021). In this study, the results further indicate that the safety perception associated with traffic lights may be misleading.

5.3. Practical implications

The findings of this study have practical implications for road safety and can inform policy-making and safety planning efforts. The analysis suggests that increasing the view of road trees can contribute to a perception of safer roads. Urban street trees act as a visual barrier, separating the roads from other infrastructures, which can create a sense of familiarity and security for road users (Harvey et al., 2015; Naderi et al., 2008). Furthermore, road greening can help alleviate eye strain for road users, thereby improving the overall safety of the traffic environment (Chiang et al., 2022). Therefore, efforts to improve the urban treescape can be considered as a potential strategy to enhance road safety perception. Widening the width of the road is another potential intervention that can be considered. Increasing the space available for road users can enhance safety by allowing more room for transportation and reducing congestion. In addition, in areas characterized by the "More dangerous than it looks" pattern, certain built environment characteristics, such as traffic lights, may impede a proper awareness of crash risk. Therefore, proactive inventions involved with education and engineering should be taken (Schneider et al., 2004).

5.4. Contributions and limitations

This research makes several valuable contributions. Firstly, it utilizes deep learning methods and SVIs to measure and map human perception of road safety. Unlike traditional studies that are often limited to small urban areas due to the high cost of surveys, this research covers a large area of Manhattan at the street level. By conducting a city-scale road safety perception measurement, the study focuses on identifying road safety perception patterns and interpreting the influence of the built environment on the disparities between human perception and actual crash rates in Manhattan.

However, it is important to acknowledge the limitations of this study. Firstly, the measure of road safety perception relied on SVIs and computer vision techniques. Although these methods provide valuable insights, visual information extracted from images may not capture the entirety of human perceptions comprehensively. Secondly, while the Place Pulse 2.0 dataset has been widely used for human perception measurements in various case study areas, it should be noted that the images and the choices made by volunteers in the dataset were collected from different countries. This introduces the possibility of perception biases due to variations in urban settings and cultural factors. And the use of a pre-trained model for obtaining the view index could benefit from being trained and evaluated in conjunction with the SVIs dataset, such as Place Pulse 2.0. It could potentially enhance the accuracy of panoptic segmentation results in the urban context and improve the validity of the data used for assessing road safety perceptions. Thirdly, this study relies solely on crash rates as a representation of the actual road safety situation. It does not consider other important aspects of road traffic safety, such as fatal rates and injury rates, which provide insights into the severity of traffic collisions. It may limit the comprehensive assessment of road safety and may introduce potential biases. In addition, this study didn't consider the differences among different road users and explore the misperception of pedestrians, cyclists and different motorists separately. This limitation is mainly due to data acquisition constraints, such as the lack of data on pedestrian volume. It may also introduce bias regarding the impact of variables, as the built environment may indirectly affect the number of traffic crashes by influencing the exposure of pedestrians, while pedestrian volume was not covered in this study. The use of street view images with deep learning for road safety perception measurements hinders the acquisition of rich demographic information about volunteers and real-time street information, as would be obtained through field surveys.

In future work, it would be beneficial to address these limitations. Efforts can be made to measure human perception of road safety in a more localized and context-specific manner, considering the specific characteristics of the study area. To validate the accuracy and reliability of this measurement, future work should include surveys that directly assess people's perception of road safety based on their daily experiences and collect demographic data and real-time street data, such as weather conditions, as control variables. Additionally, incorporating a broader range of road safety indicators, such as injury and fatality rates, would provide a more comprehensive understanding of the road safety situation.

6. Conclusions

In this study, the ViT deep learning model and SVIs were employed to measure human perception of road safety in Manhattan. The analysis revealed that roads perceived as unsafe were primarily located in the urban periphery, indicating the heterogeneity of the distribution. The association between the perception of road safety and crash rates was further investigated. It was observed that perception score of road safety had a positive impact on crash rates. This could be attributed to the fact that road users tend to exercise greater caution in roads perceived as unsafe or may even avoid these areas altogether. Conversely, in roads perceived as safe, people may exhibit reduced caution, leading to potentially higher crash rates. Four distinct perception patterns were identified: "Safer than it looks," "Safe as it looks," "More dangerous than it looks," and "Dangerous as it looks." The perception patterns were influenced by several significant factors, including the wall view index, tree view index, building view index, distance to the nearest traffic signals, and street width. Importantly, the study highlighted the pivotal role of traffic lights in the "More dangerous than it looks" pattern. While traffic lights may enhance people's perception of safety, areas in close proximity to traffic lights were identified as potentially accident-prone regions. This finding underscores the complex relationship between perception and actual safety, where the presence of traffic lights may create a false sense of security while posing inherent risks to road users.

Overall, the study provides valuable insights into the misperception of road safety and the factors influencing them. By understanding these patterns and the associated factors, policymakers and urban planners can make informed decisions to improve road safety in areas prone to misperceptions.

CRediT authorship contribution statement

Xujing Yu: Conceptualization, Methodology, Software, Writing – original draft. **Jun Ma:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition. **Yihong Tang:** Writing – review & editing, Methodology. **Tianren Yang:** Writing – review & editing. **Feifeng Jiang:** Writing – review & editing.

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

Data will be made available on request.

Acknowledgments

This study was jointly supported by the General Research Fund (No. 17200422) from the Hong Kong Research Grant Council, and the Young Scientists Fund (No. 42201092) from the National Natural Science Foundation of China. The authors would also like to acknowledge the support from the University of Hong Kong Seed Fund for PI Research – Basic Research (No. 202111159236).

References

Aceves-González, C., Ekambaram, K., Rey-Galindo, J., & Rizo-Corona, L. (2020). The role of perceived pedestrian safety on designing safer built environments. Traffic Injury Prevention, 21(sup1), S84–S89. https://doi.org/10.1080/15389588.2020.1812062.

AlHamad, S., Almallah, M., Naser, M.N., Alhajyaseen, W.K.M., de Roos, M.P., 2023. Examining the role of road safety audits worldwide: exploring road safety expert's opinions. Int. J. Inj. Contr. Saf. Promot. 30 (1), 106–115. https://doi.org/10.1080/17457300.2022.2114090.

Aziz, H.M.A., Nagle, N.N., Morton, A.M., Hilliard, M.R., White, D.A., Stewart, R.N., 2018. Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data. Transportation 45 (5), 1207–1229. https://doi.org/10.1007/s11116-017-9760-8.

Bíl, M., Andrášik, R., Sedoník, J., 2019. A detailed spatiotemporal analysis of traffic crash hotspots. Appl. Geogr. 107, 82–90. https://doi.org/10.1016/j.apgeog.2019.04.008.

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: a review. Landsc. Urban Plan. 215, 104217 https://doi.org/10.1016/j.landurbplan.2021.104217.

Boua, M., Kouabenan, D.R., Belhaj, A., 2022. Road safety behaviors: role of control beliefs and risk perception. Transport. Res. F: Traffic Psychol. Behav. 91, 45–57. https://doi.org/10.1016/j.trf.2022.09.021.

Brown, I.D., Groeger, J.A., 1988. Risk perception and decision taking during the transition between novice and experienced driver status. Ergonomics 31 (4), 585–597.

Cai, Q., Abdel-Aty, M., Zheng, O., Wu, Y., 2022. Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety. Transportation Research Part c: Emerging Technologies 135, 103541. https://doi.org/10.1016/j.trc.2021.103541.

Campbell, A., Both, A., Sun, Q.C., 2019. Detecting and mapping traffic signs from google street view images using deep learning and GIS. Comput. Environ. Urban Syst. 77, 101350.

Chen, F., Shi, L., Li, Y., Wang, Q., Sun, H., Tang, X., Zu, J., Sun, Z., 2022. Measuring road safety achievement based on EWM-GRA-SVD: a decision-making support system for APEC countries. Knowl.-Based Syst. 252, 109373 https://doi.org/10.1016/j.knosys.2022.109373.

Chiang, Y.-C., Ke, R.-A., Li, D., Weng, P.-Y., 2022. Greening and safety: the influence of road greenness on driver's attention and emergency reaction time. Environ. Behav. 54 (9–10), 1195–1226.

Cho, G., Rodríguez, D.A., Khattak, A.J., 2009. The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. Accid. Anal. Prev. 41 (4), 692–702. https://doi.org/10.1016/j.aap.2009.03.008.

Deery, H.A., 1999. Hazard and risk perception among young novice drivers. J. Saf. Res. 30 (4), 225–236. https://doi.org/10.1016/S0022-4375(99)00018-3.

DeJoy, D.M., 1989. The optimism bias and traffic accident risk perception. Accid. Anal. Prev. 21 (4), 333–340.

Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach. Accid. Anal. Prev. 112, 116–126. https://doi.org/10.1016/j.aap.2017.12.026.

Ding, H., Sze, N.N., Li, H., Guo, Y., 2020. Roles of infrastructure and land use in bicycle crash exposure and frequency: a case study using Greater London bike sharing data. Accid. Anal. Prev. 144, 105652.

Dinh, D.D., Vũ, N.H., McIlroy, R.C., Plant, K.A., Stanton, N.A., 2020. Effect of attitudes towards traffic safety and risk perceptions on pedestrian behaviours in Vietnam. IATSS Research 44 (3), 238–247.
Diógenes, M. C., Lindau, L. A., & de Transportes–LASTRAN, L. de S. (2010). Risk perception of pedestrians at midblock crossings in Brazil. Proceedings of the 12th World Conference on Transport Research (WCTR), 1–19.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., & Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv Preprint arXiv: 2010.11929.
Dubey, A., Naik, N., Parikh, D., Raskar, R., Hidalgo, C.A., 2016. Deep Learning the City: Quantifying Urban Perception at a Global Scale. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Vol. 9905. Springer International Publishing, pp. 196–212. https://doi.org/10.1007/978-3-319-46448-0_12.
Edwards, J.B., 1999. Speed adjustment of motorway commuter traffic to inclement weather. Transport. Res. F: Traffic Psychol. Behav. 2 (1), 1–14.
Elvik, R., 2010. Why some road safety problems are more difficult to solve than others. Accid. Anal. Prev. 42 (4), 1089–1096.
Elvik, R., Bjørnskau, T., 2005. How accurately does the public perceive differences in transport risks?: An exploratory analysis of scales representing perceived risk. Accid. Anal. Prev. 37 (6), 1005–1011. https://doi.org/10.1016/j.aap.2005.05.003.
Ewing, R., Cervero, R., 2010. Travel and the built environment. J. Am. Plann. Assoc. 76 (3), 265–294. https://doi.org/10.1080/01944361003766766.
Ewing, R., Dumbaugh, E., 2009. The built environment and traffic safety: a review of empirical evidence. J. Plan. Lit. 23 (4), 347–367. https://doi.org/10.1177/0885412209335553.
Frank, L.D., Engelke, P.O., 2001. The built environment and human activity patterns: exploring the impacts of urban form on public health. J. Plan. Lit. 16 (2), 202–218.
Gu, Z., Peng, B., 2021. Investigation into the built environment impacts on pedestrian crash frequencies during morning, noon/afternoon, night, and during peak hours: a case study in Miami County, Florida. Journal of Transportation Safety & Security 13 (9), 915–935.
Han, C., Huang, H., Lee, J., Wang, J., 2018. Investigating varying effect of road-level factors on crash frequency across regions: a Bayesian hierarchical random parameter modeling approach. Analytic Methods in Accident Research 20, 81–91.
Han, C., Xu, G., Pervez, A., Gao, F., Huang, H., Pei, X., Zhang, Y., 2023. Modeling traveler's speed-route joint choice behavior with heterogeneous safety concern. Analytic Methods in Accident Research 37, 100253. https://doi.org/10.1016/j.amar.2022.100253.
Hankey, S., Zhang, W., Le, H.T., Hystad, P., James, P., 2021. Predicting bicycling and walking traffic using street view imagery and destination data. Transp. Res. Part D: Transp. Environ. 90, 102651.
Harvey, C., Aultman-Hall, L., Hurley, S.E., Troy, A., 2015. Effects of skeletal streetscape design on perceived safety. Landsc. Urban Plan. 142, 18–28.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, 770–778. https://doi.org/10.1109/CVPR.2016.90.
Hou, Q., Huo, X., Leng, J., 2020. A correlated random parameters tobit model to analyze the safety effects and temporal instability of factors affecting crash rates. Accid. Anal. Prev. 134, 105326 https://doi.org/10.1016/j.aap.2019.105326.
Høye, A.K., Hesjevoll, I.S., 2020. Traffic volume and crashes and how crash and road characteristics affect their relationship – a meta-analysis. Accid. Anal. Prev. 145, 105668 https://doi.org/10.1016/j.aap.2020.105668.
Imprialou, M., Quddus, M., 2019. Crash data quality for road safety research: current state and future directions. Accid. Anal. Prev. 130, 84–90.
Jia, R., Khadka, A., Kim, I., 2018. Traffic crash analysis with point-of-interest spatial clustering. Accid. Anal. Prev. 121, 223–230.
Jiang, F., Ma, J., 2021. A comprehensive study of macro factors related to traffic fatality rates by XGBoost-based model and GIS techniques. Accid. Anal. Prev. 163, 106431 https://doi.org/10.1016/j.aap.2021.106431.
Jiang, F., Ma, J., Li, Z., 2022. Pedestrian volume prediction with high spatiotemporal granularity in urban areas by the enhanced learning model. Sustain. Cities Soc. 79, 103653 https://doi.org/10.1016/j.scs.2021.103653.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Proces. Syst. 30.
Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 9404–9413.
Kwon, J.-H., Cho, G.-H., 2020. An examination of the intersection environment associated with perceived crash risk among school-aged children: using street-level imagery and computer vision. Accid. Anal. Prev. 146, 105716 https://doi.org/10.1016/j.aap.2020.105716.
Lee, G., Park, Y., Kim, J., Cho, G.-H., 2016. Association between intersection characteristics and perceived crash risk among school-aged children. Accid. Anal. Prev. 97, 111–121. https://doi.org/10.1016/j.aap.2016.09.001.
Li, X., Ratti, C., 2019. Mapping the spatio-temporal distribution of solar radiation within street canyons of Boston using Google Street View panoramas and building height model. Landsc. Urban Plan. 191, 103387 https://doi.org/10.1016/j.landurbplan.2018.07.011.
Li, X., Ratti, C., Seiferling, I., 2018. Quantifying the shade provision of street trees in urban landscape: a case study in Boston, USA, using Google Street View. Landsc. Urban Plan. 169, 81–91. https://doi.org/10.1016/j.landurbplan.2017.08.011.
Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., Mamouei, M., 2023. An interpretable machine learning framework for measuring urban perceptions from panoramic street view images. iScience 26 (3), 106132. https://doi.org/10.1016/j.isci.2023.106132.
Liu, L., Silva, E.A., Wu, C., Wang, H., 2017. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. Computers, Environment and Urban Systems 65, 113–125. https://doi.org/10.1016/j.compenvurbsys.2017.06.003.
Liu, Z., Yang, A., Gao, M., Jiang, H., Kang, Y., Zhang, F., Fei, T., 2019. Towards feasibility of photovoltaic road for urban traffic-solar energy estimation using street view image. J. Clean. Prod. 228, 303–318.
Lund, I.O., Rundmo, T., 2009. Cross-cultural comparisons of traffic safety, risk perception, attitudes and behaviour. Saf. Sci. 47 (4), 547–553. https://doi.org/10.1016/j.ssci.2008.07.008.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst. 30.
Luo, Y., Liu, Y., Xing, L., Wang, N., Rao, L., 2022. Road safety evaluation framework for accessing park green space using active travel. Front. Environ. Sci. 10, 864966.
Ma, J., Cheng, J.C.P., Jiang, F., Chen, W., Zhang, J., 2020a. Analyzing driving factors of land values in urban scale based on big data and non-linear machine learning techniques. Land Use Policy 94, 104537. https://doi.org/10.1016/j.landusepol.2020.104537.
Ma, J., Ding, Y., Cheng, J.C.P., Jiang, F., Tan, Y., Gan, V.J.L., Wan, Z., 2020b. Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. J. Clean. Prod. 244, 118955 https://doi.org/10.1016/j.jclepro.2019.118955.
Mathew, S., Pulugurtha, S.S., Duvvuri, S., 2022. Exploring the effect of road network, demographic, and land use characteristics on teen crash frequency using geographically weighted negative binomial regression. Accid. Anal. Prev. 168, 106615 https://doi.org/10.1016/j.aap.2022.106615.
Naderi, J.R., Kweon, B.S., Maghelal, P., 2008. The street tree effect and driver safety. ITE Journal on the Web 78 (2), 69–73.
Park, Y., Garcia, M., 2020. Pedestrian safety perception and urban street settings. Int. J. Sustain. Transp. 14 (11), 860–871. https://doi.org/10.1080/15568318.2019.1641577.
Rankavat, S., Tiwari, G., 2016. Pedestrians risk perception of traffic crash and built environment features – Delhi, India. Saf. Sci. 87, 1–7. https://doi.org/10.1016/j.ssci.2016.03.009.
Rothman, L., Buliung, R., To, T., Macarthur, C., Macpherson, A., Howard, A., 2015. Associations between parents′ perception of traffic danger, the built environment and walking to school. J. Transp. Health 2 (3), 327–335. https://doi.org/10.1016/j.jth.2015.05.004.
Sainju, A.M., Jiang, Z., 2020. Mapping road safety features from streetview imagery: a deep learning approach. ACM/IMS Transactions on Data Science 1 (3), 1–20. https://doi.org/10.1145/3362069.
Salonen, A.O., 2018. Passenger's subjective traffic safety, in-vehicle security and emergency management in the driverless shuttle bus in Finland. Transp. Policy 61, 106–110. https://doi.org/10.1016/j.tranpol.2017.10.011.
Santos, K., Dias, J.P., Amado, C., 2022. A literature review of machine learning algorithms for crash injury severity prediction. J. Saf. Res. 80, 254–269. https://doi.org/10.1016/j.jsr.2021.12.007.
Schepers, P., Hagenzieker, M., Methorst, R., van Wee, B., Wegman, F., 2014. A conceptual framework for road safety and mobility applied to cycling safety. Accid. Anal. Prev. 62, 331–340. https://doi.org/10.1016/j.aap.2013.03.032.
Schneider, R.J., Ryznar, R.M., Khattak, A.J., 2004. An accident waiting to happen: a spatial approach to proactive pedestrian planning. Accid. Anal. Prev. 36 (2), 193–211. https://doi.org/10.1016/S0001-4575(02)00149-5.
Shaaban, K., Ibrahim, M., 2021. Analysis and identification of contributing factors of traffic crashes in New York City. Transp. Res. Procedia 55, 1696–1703. https://doi.org/10.1016/j.trpro.2021.07.161.
Su, N., Li, W., Qiu, W., 2023. Measuring the associations between eye-level urban design quality and on-street crime density around New York subway entrances. Habitat Int. 131, 102728 https://doi.org/10.1016/j.habitatint.2022.102728.
UN Statistics. (2018). Indicator 3.6.1—E-Handbook on SDG Indicators—UN Statistics Wiki. https://unstats.un.org/wiki/display/SDGeHandbook/Indicator+3.6.1.
von Stülpnagel, R., Lucas, J., 2020. Crash risk and subjective risk perception during urban cycling: evidence for congruent and incongruent sources. Accid. Anal. Prev. 142, 105584 https://doi.org/10.1016/j.aap.2020.105584.
Wang, J., Ma, S., Jiao, P., Ji, L., Sun, X., Lu, H., 2023. Analyzing the risk factors of traffic accident severity using a combination of random forest and association rules. Appl. Sci. 13 (14), 8559.
Wei, J., Yue, W., Li, M., Gao, J., 2022. Mapping human perception of urban landscape from street-view images: a deep-learning approach. Int. J. Appl. Earth Obs. Geoinf. 112, 102886 https://doi.org/10.1016/j.jag.2022.102886.
Wen, X., Xie, Y., Wu, L., Jiang, L., 2021. Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP. Accident Analysis & Prevention 159, 106261. https://doi.org/10.1016/j.aap.2021.106261.
World Health Organization, 2018. Global Status Report on Road Safety 2018. World Health Organization. https://www.who.int/publications/i/item/9789241565684.
Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. https://github.com/facebookresearch/detectron2.

Zeng, Q., Wen, H., Huang, H., Pei, X., Wong, S.C., 2017. A multivariate random-parameters Tobit model for analyzing highway crash rates by injury severity. Accid. Anal. Prev. 99, 184–191. https://doi.org/10.1016/j.aap.2016.11.018.

Zhang, K., Qian, Z., Yang, Y., Chen, M., Zhong, T., Zhu, R., Lv, G., Yan, J., 2022. Using street view images to identify road noise barriers with ensemble classification model and geospatial analysis. Sustain. Cities Soc. 78, 103598.

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. Landsc. Urban Plan. 180, 148–160. https://doi.org/10.1016/j.landurbplan.2018.08.020.