Travel Behaviour and Society 36 (2024) 100814

Contents lists available at ScienceDirect

Travel Behaviour and Society

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

Exploring non-linear effects of environmental factors on the volume of pedestrians of different ages using street view images and computer vision technology

Dongwei Liu [a], Yi Lu [a,*], Linchuan Yang [b]

[a] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong, Hong Kong Special Administrative Region
[b] Department of Urban and Rural Planning, School of Architecture, Southwest Jiaotong University, Chengdu, China

ARTICLE INFO

Keywords:
Pedestrian volume
Aging friendly
Street view imagery
Computer vision
Walking behavior
Gradient Boosting Decision Tree

ABSTRACT

Creating pedestrian-friendly neighborhoods and encouraging walking activities not only improves urban liveliness but also delivers health and environmental benefits. Previous research has largely focused on individual walking behavior, which often exhibits stronger associations with personal traits than with built environment characteristics. Pedestrian volume, a significant indicator of urban vitality and collective walking behavior, may have a stronger relationship with environmental characteristics. Moreover, prior studies often hypothesize a linear relationship between built environmental determinants and walking behavior. The exploration of potential non-linear influences may help policy makers and urban planners by identifying minimum, maximum, and optimal values of built environmental variables conducive for walking. Furthermore, studies focusing on the collective walking of older pedestrians remain scarce in the context of aging societies.

In this study, we utilize street view imagery and advanced computer vision algorithms to estimate citywide pedestrian volumes and their corresponding age classifications (older adults vs. all). Additionally, we assessed both micro and macro built environmental factors. The possible non-linear impact is examined using Gradient Boosting Decision Tree (GBDT). Our findings reveal disparities in the influence of environmental determinants on the volume of older pedestrians versus that of all pedestrians. Also, the significance of environmental elements exhibits variations across different spatial resolutions. Further, both eye-level vegetation and building level show an inverted U-shaped influence on the pedestrian count.

1. Introduction

The investigation into walking activity forms a crucial component of urban research (Adkins et al., 2017; Forsyth et al., 2009; Handy et al., 2005; He et al., 2022; Saelens & Handy, 2008; Wang et al., 2016; Wu et al., 2023). Firstly, collective pedestrian activity, e.g., pedestrian volume, often stimulates an area's socioeconomic vitality and social interaction (Chen et al., 2020; Im & Choi, 2019; Kim, 2018; Li et al., 2022). Secondly, having adequate walking time for an individual halts the trends of sedentary lifestyle and private automobile dependence, manifesting considerable health (Aune et al., 2016; Boyle et al., 2012; Manson et al., 1999; Smith et al., 2007; Wu et al., 2013) and environmental benefits (Maioli et al., 2019; Pellicer-Chenoll et al., 2021).

To enhance these positive effects, urban planners and researchers have endeavored to discern the diverse built environmental factors shaping walking behavior and pedestrian demand for decades. Significant built environmental correlates include macroscale built environmental factors, e.g., 'five Ds': density, diversity, design, accessibility to destination, and distance to public transit (Adkins et al., 2017; Carr et al., 2010; Im & Choi, 2019; Jacobs, 1961; Lynch, 1984) and microscale elements, e.g., features of buildings (Boarnet et al., 2011), green view (Lu, 2019; Yang et al., 2021), sky view (Yin & Wang, 2016), sidewalk (Nagata et al., 2020). Some exploratory qualitative studies also classified these variables into facilitator and barrier through interviews and focus group studies (Ahlport et al., 2008; Galea et al., 2008; Lockett et al., 2005; Močnik et al., 2022). The facilitator represents the variables making neighborhoods attractive for walking or improve accessibility, which includes the availability of restaurants, retail, public services,

* Corresponding author at: Department of Architecture and Civil Engineering, City University of Hong Kong, 5408, 5/F, Mong Man-wai Building, 83 Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong, Hong Kong Special Administrative Region.
E-mail addresses: dongweliu3-c@my.cityu.edu.hk (D. Liu), yilu24@cityu.edu.hk (Y. Lu), yanglc0125@swjtu.edu.cn (L. Yang).

https://doi.org/10.1016/j.tbs.2024.100814
Received 29 November 2023; Received in revised form 21 March 2024; Accepted 30 April 2024
Available online 6 May 2024

schools, sidewalks, bus stop density, and vegetation. On the contrary, the barrier represents the variables undermining the accessibility or sense of safety, comfort, and convenience including distance to the business center, distance to railway stations, and traffic volume (traffic-related variables). Besides, there are variables that exhibit dual impacts, either positive or negative, or whose effects on pedestrian remain unverified. Pedestrians perceive and evaluate these objective factors from multiple subjective dimensions, e.g., safety, comfort, aesthetics (Arellana et al., 2020; De Vos et al., 2023), and then make travel decisions about walking, e.g., whether to walk or not, when and where to walk. It is worth noting the same objective elements can be perceived differently by respondents with different sociodemographic elements, including age, gender, income, and education attainment, as well as personal health, lifestyle choices, and cultural norms (Delclòs-Alió et al., 2022; Frank et al., 2008; Guzman et al., 2020; Hearst et al., 2013; Huang et al., 2022; Yang et al., 2023).

1.1. Impacts of built environmental factors on travel behavior

Existing research on the built environmental determinants of travel behavior bifurcates into two major categories. The first employs spatial units as the focal point of the investigation, deploying measures such as walkability indexes or walking scores. These composite built environment metrics primarily concentrate on the features of pedestrian environments, e.g., the walkable infrastructure and opportunities for pedestrian activity (Adkins et al., 2017; Carr et al., 2010; Zhao et al., 2021). However, they generally omit individual's socioeconomic and personal variables, such as age, which can significantly influence perceived walkability and therefore impact individual walking behavior (De Vos et al., 2023; Lättman et al., 2018; Tiznado-Aitken et al., 2020).

The second adopts individual person as study subjects, emphasizing their individual walking propensities. These personal attributes can either amplify or discourage pedestrian's perception of different aspects of the built environment (Feng, 2017; Tcymbal et al., 2020; Wang et al., 2020; Yang et al., 2022), and thereby influence their travel behaviors.

1.1.1. Specific characteristics of older pedestrians

As a special population subgroups, older pedestrians are often involved with mobility limitations and are more sensitive to certain environmental factors, e.g., road surfaces and lighting, than other pedestrians (Feng, 2017; Ghani et al., 2018). Compared with younger adults, older adults tend to be more sedentary, make fewer and shorter trips, and have smaller activity spaces (Harvey et al., 2015; Sjögren et al., 2014). Older pedestrians also exhibit heightened sensitivity to uneven walkways, which prompts a more cautious walking strategy in older adults compared to younger adults (Ippersiel et al., 2021). Older people with walking difficulty tend to report poor lighting, distances, and terrain as barriers in the outdoor environment more often than those without walking difficulty (Rantakokko et al., 2012).

Therefore, there is a need to study how environmental factors can impact on different age groups, especially older pedestrians, who may be more sensitive to walking barriers.

1.1.2. Difference between collective walking behaviors and individual walking behaviors

As we mentioned earlier, most previous studies focused on the links between built environmental characteristics and individual walking behavior, such as speed, duration, frequency, and gaits of walking. Besides built environmental determinants, a multitude of individual factors, inclusive of health status and motivation, can cause deviation in individual walking behaviors. Instead, collective walking patterns, including pedestrian volume, are less impervious to these influences. Researchers found that built environment features can partially explain two-thirds of the variance in pedestrian volume, and scarcely one-third of the variability in individual walking duration (Jiang et al., 2021). Further, pedestrian volume is a significant indicator of urban vitality that cannot be replaced by aggregating the individual walking behavior of urban residents in a community (Cambra & Moura, 2020; Lee et al., 2017; Middleton, 2018). Longer average walking duration of residents in a community is not equivalent to higher pedestrian volume. The gap may be attributed to the low population density in this area or to the fact that walking occurs outside the community (Jiang et al., 2021). Similarly, the associations between built environment and pedestrian volume may differ among different population subgroups. For example, older pedestrians are more likely to concentrate in residential areas while less appear in suburban and mountainous areas compared with their counterparts (Liu et al., 2023a). However, evidence on the heterogeneity in the built environment-pedestrian volume link among different population subgroups remains scarce, partly due to methodological limitation. It remains challenging to collect sociodemographic information for all or most pedestrians in an urban area.

1.2. Collecting individual information for collective walking behavior

It's crucial to recognize that pedestrian demand embodies a synthesis of the activities of both local residents and non-local visitors, and excluding the latter could potentially produce incomplete or biased results. However, collecting sociodemographic data from visitors is indeed challenging due to their transient nature – they may not be easily reachable for one-time or even longitudinal surveys. Traditional methods, such as spot-based pedestrian counts, are notably costly and time-intensive (Im and Choi, 2019; Kang, 2018; Li et al., 2024). To enhance the accuracy and efficiency of pedestrian volume measurements, researchers have employed alternative methods. The widespread use of mobile Internet and location-based services (LBS) via smartphones has enabled the accumulation of extensive human mobility records, providing researchers with useful information on human mobility distribution (Chen et al., 2022c; Liu et al., 2015). Several studies have verified the high correlation between mobile phone location-based records and the spatiotemporal characteristics of human activities (Dunkel, 2015; Gariazzo et al., 2016). Nevertheless, the efficacy and precision of mobile signal data analysis are innately tied to the density and arrangement of signal towers or base stations (Caceres et al., 2012). Further, mobile signal data may be inaccurate to identify a pedestrian, and it remains imprecise to discern user activities—whether an individual is standing on streets, or simply staying indoors next to a street (Du et al., 2017). Furthermore, researchers have utilized individual-level portable-GPS-based sensing data to measure mobility of human and their exposure to various environment (Roberts & Helbich, 2021; Vich et al., 2019). However, this approach requires participants to be equipped with GPS-based devices, making it only suitable for a limited sample size. Besides, geolocated data from social media platforms such as Instagram, Twitter, and Weibo have been utilized to measure human mobility (Cesare et al., 2019; Hawelka et al., 2014; Roberts et al., 2019). These data not only provide insights into the volume of people in a location but also the associated emotions and activities. However, Social media data might not represent the entire population as it is usually created by a subset of the population, which can introduce demographic biases (Morstatter & Liu, 2017). Nightlight data extracted from satellite images has been extensively explored to depict density of human activities. However, this dataset tends to primarily capture human activity only at night and cannot measure pedestrian volume directly. Furthermore, some researchers have utilized origin–destination (OD) data from large-scale travel survey to capture aggregate travel patterns, including commuter flows, and subsequently calculated environmental exposure (Hankey et al., 2017; Widener et al., 2013). This approach can be complex, and the process may introduce errors or oversimplify the environmental context in which individuals travel (Cheng et al., 2020). Certain studies have fused various data sources to estimate human volume. For instance, a environmental behavior model integrating both pilot observation counts and built environmental factors is employed to predict pedestrian counts for entire cities (Cambra & Moura, 2020). Additionally, nighttime light images (VIIRS), geolocated Weibo data, and mobile signal data are combined to measure pedestrian volume in another study (Song et al., 2018). While these methods are innovative and can be valuable in specific scenarios, they tend to be overly complex, leading to potential overfitting of the training data. Moreover, they rely on multisource data and their resolution depends on the lowest quality data source available.

As an emerging and prevalent data source, Street View Images (SVIs) provide a novel data source for obtaining not only environmental elements (Biljecki & Ito, 2021) but also pedestrian volume (Chen et al., 2020; Lian et al., 2024; Liu et al., 2023b; Yin et al., 2015). These images serve as an invaluable resource, allowing researchers to gather geographically precise data on pedestrian volumes. Street view images allow a comprehensive, visual capture of pedestrian activity in specific locations, enabling the observation of intricate details that would otherwise go unnoticed, such as pedestrian demographics including age, and activities (Liu et al., 2023b). Besides, SVI enables the capture of detailed visual data about the real conditions of streets and the various factors that might influence walking behaviors, including the size of buildings, view of the sky, sidewalk width, road condition, green view index, and the presence of street furniture such as traffic lights and traffic signs (Cordts et al., 2016).

1.3. Needs for exploring non-linear relationships

It is worth noting that most studies assumed a predefined linear link between pedestrian behavior and its influencing components, which may provide misleading evidence to guide urban planning and policy making (L. Chen et al., 2022b; Chen et al., 2022a; Lu, 2019). This intricate dynamic of pedestrian interaction with the built environment might foster a non-linear association rather than a linear association (Cheng et al., 2020; Liu et al., 2021; Yang et al., 2021; Yang et al., 2022). Certain built environmental variables have been discerned to possess optimal values so that they may influence pedestrian behavior in a inverted U-shaped manner (Christiansen et al., 2016; Yang et al., 2021; Yang et al., 2022). The specificities of these optimal values exhibit potential differences across different regions and among people with varying characteristics (Christiansen et al., 2016; Lu et al., 2019).

In recent years, research has applied machine learning methodologies to delve into this non-linear relationship between walking behavior and environmental elements (Liu et al., 2021; Tao et al., 2020; Xiao et al., 2021; Yang et al., 2021). However, most investigations primarily focus on the individual walking behavior, such as the duration, distance, and frequency of walking. Nonetheless, the evidence of the link between built environmental factors and pedestrian volume has remained predominantly unexplored. There are considerable spatial discrepancies between pedestrian volume and individual behavior, and the factors that drive individuals to engage in walking may not necessarily lead to high levels of pedestrian activity in specific spaces and vice versa (Jiang et al., 2021). Therefore, it necessitates an investigation into the non-linear correlation that exists between environmental factors and collective pedestrian activities within certain streets or locations.

1.4. Current study

In this study, SVI was utilized as a data resource to obtain the total and older pedestrian counts at each SVI sampling location. We measured environmental factors with both macroscale geographic data from multi-sources and microscale elements extracted from SVIs. These factors were then applied across various spatial resolutions (at the SVI-collecting point and in the buffer of 100 m, 200 m, 500 m, and 1-kilometer) as predictive variables. Subsequently, we explored the non-linear relationship between these environmental variables and both the total and older pedestrian volume. The novelty and contributions of our research are threefold. First, we employ SVI to obtain the pedestrian volume of both all people and older adults. Second, we pioneer the investigation of the non-linear relationship between environmental factors and collective walking behavior. Third, we have examined the variation in the effects of environmental factors across different spatial scales on pedestrian volume.

2. Data and methodology

2.1. Pedestrian volume

In this study, we collected data on pedestrian activity in the form of total and older pedestrian counts based on the Street View Images (SVIs) of the whole Hong Kong area (Fig. 2), which were downloaded from Google Maps. Existing studies have proposed and validated that SVIs, when synergistically combined with computer vision technology, present a cost-efficient data source for capturing pedestrian volume at the street level (Chen et al., 2020; Yin et al., 2015). The SVIs covered 31,972 street segments in Hong Kong including nearly all streets in urban areas and most roads in suburban. To minimize any overlapping or missing road sections, SVIs were downloaded at intervals of 50 m along each street. Subsequently, a total of 70,021 sampling points were selected, encompassing 199 Tertiary Planning Units (TPUs).

Among these sampling points, 11,467 were found to have pedestrians, referred to as SVI sample points (SSP). For each SSP, a Street View panorama image was obtained. Following the approach proposed by Liu et al. (2023a), we recognized pedestrians and classified them into two groups: older pedestrians and non-older pedestrians (Fig. 1). Consequently, we obtained the count of pedestrians and oldre pedestrians for

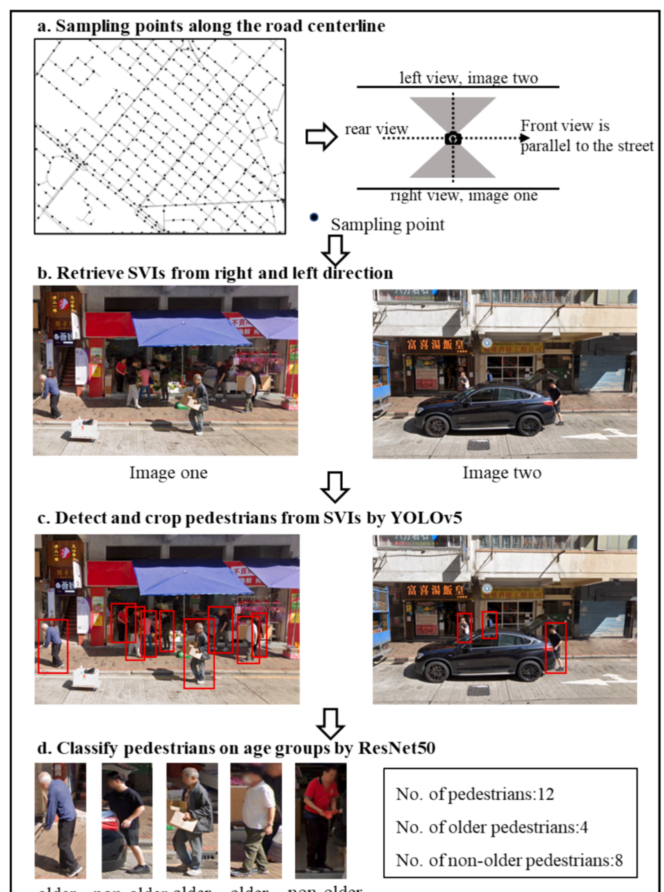

Fig. 1. Workflow of the detection of pedestrians and older pedestrians. (a) Sampling points along the road centerline; (b) Retrieve SVIs from right and left direction; (c) Detect and crop pedestrians from SVIs by YOLOv5; (d) Classify pedestrians on age groups by ResNet50.

each SSP. Overall, a total of 35,353 pedestrians were detected, comprising 7,375 older pedestrians and 27,978 non-older pedestrians. Most of detected pedestrians concentrated in the streets within Kowloon and north Hong Kong Island (Fig. 2).

2.2. Microscale environmental factors

In addition to pedestrian volume, this study also assessed microscale environmental factors through Street View Images (SVIs). Contrasting with GIS-based urban big data, the microscale streetscapes extracted from Street View Images (SVIs) offer a more congruent representation of the pedestrian's perceptual experience. Numerous investigations have harnessed SVIs to evaluate street-level built environments and urban phenomena with computer vision technology (Biljecki and Ito, 2021; Chen et al., 2022b; Li et al., 2015; Liu et al., 2023c; Lu, 2019; Yang et al., 2021).

This study employed the Pyramidal Scene Parsing Network (PSPNet) (Zhao et al., 2017), a highly efficacious machine learning algorithm tailored for semantic segmentation, to extract microscale environmental variables. We selected some microscale built-environment features based on the previous findings and the data availability from the image segmentation models. The selected features should be supported by previous empirical findings (Ewing & Handy, 2009; Nagata et al., 2020). They should also be measurable from SVIs by the image segmentation

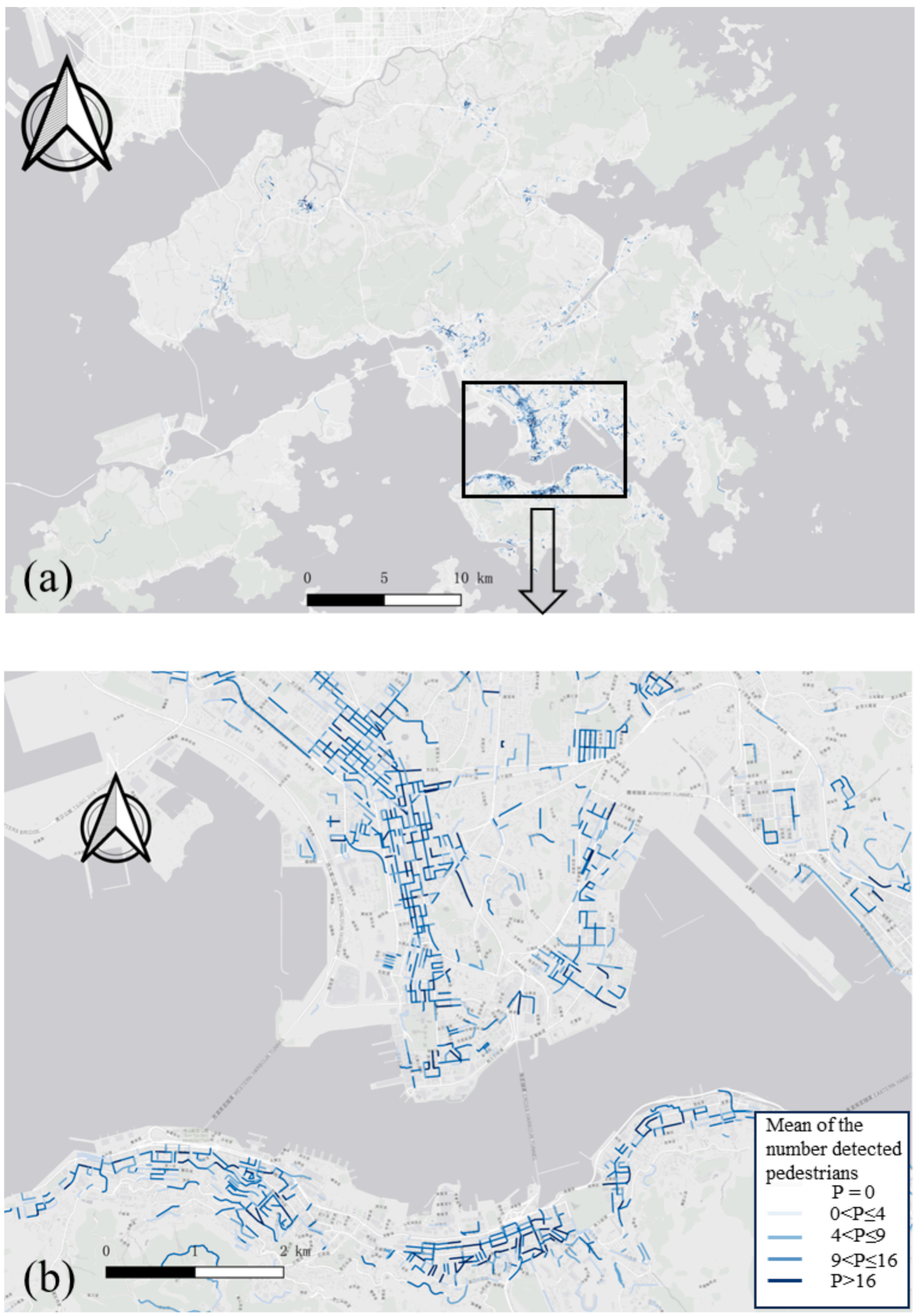

Fig. 2. Study area and the distribution of detected pedestrians. (a) Spatial distribution of detected pedestrians; (b) Spatial distribution of detected pedestrians in the black basket of (a) (Kowloon and north Hong Kong Island).

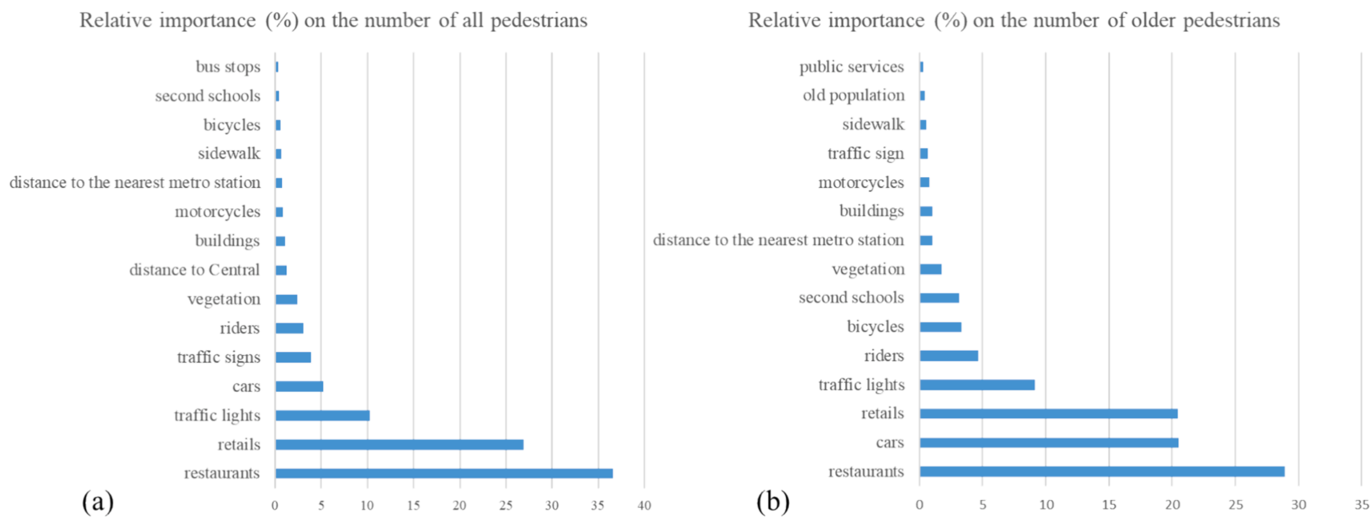

Fig. 3. The RI of the 15-highest predictor variables on the number of (a) all pedestrians and (b) older pedestrians.

Table 1
Description and spatial resolution range of predictor variables.

Category	Variable	Description	Buffer-relevant	Mean (SD) in 500 m buffer	Spatial resolution
Street-level built environment factors	Vegetation	Pixel proportion of vegetation in SVI	Yes	0.18 (0.16)	Single point, 100 m, 200 m, 500 m, 1 km
	Building	Pixel proportion of building in SVI	Yes	0.38 (0.15)	
	Sidewalk	Pixel proportion of sidewalk in SVI	Yes	0.01 (0.01)	
	Road	Pixel proportion of road in SVI	Yes	0.28 (0.06)	
	Traffic light	Pixel proportion of traffic light in SVI	Yes	0.00 (0.00)	
	Traffic sign	Pixel proportion of traffic sign in SVI	Yes	0.00 (0.00)	
	Wall	Pixel proportion of wall in SVI	Yes	0.00 (0.01)	
	Pole	Pixel proportion of pole in SVI	Yes	0.00 (0.00)	
	Fence	Pixel proportion of fence in SVI	Yes	0.02 (0.01)	
	Sky	Pixel proportion of sky in SVI	Yes	0.07 (0.06)	
Traffic volume related factors	Car	Pixel proportion of car in SVI	Yes	0.03 (0.03)	
	Truck	Pixel proportion of truck in SVI	Yes	0.00 (0.00)	
	Rider	Pixel proportion of rider in SVI	Yes	0.00 (0.00)	
	Bicycle	Pixel proportion of bicycle in SVI	Yes	0.00 (0.00)	
Socioeconomic factors	Population density	Number of residents per 100 sqm	Yes	3.14 (3.25)	500 m, 1 km
	population density for older residents	Number of residents (Age $\geq$ 60) per 100 square meters	Yes	0.63 (0.61)	500 m, 1 km
	Population density	Number of residents (Age < 60) per 100 square meters	Yes	2.51 (2.67)	500 m, 1 km
	Median household income	The median house hold income （HKD/month） of residents within the study area	Yes	33,294 (28373)	500 m, 1 km
Neighborhood built environment factors	Building floor density	Average building floor density	Yes	1.75 (2.27)	500 m, 1 km
	Job density	Number of jobs per 100 square meters	Yes	1.39 (2.05)	500 m, 1 km
	Land use mix	Entropy for the diversity of land uses within the buffer area. $Landusemix = \Sigma_i(p_i lnp_i)/lnN$, where pi is the ratio of the ith land use, and N is the total types of land uses. Five categories of land uses (residential, business, industrial, open space, and commercial) are considered (N = 5).	Yes	0.59 (0.28)	500 m, 1 km
	Density of road intersection	Number of road intersections per square kilometer	Yes	62.87 (30.50)	500 m, 1 km
	Normalized Difference Vegetation Index (NDVI)	The difference between near-infrared (which vegetation strongly reflects) and red light (which vegetation absorbs)	Yes	0.38 (0.14)	500 m, 1 km
	Restaurant	Count of restaurant POIs	Yes	54.14 (33.68)	Single point, 100 m, 200 m, 500 m, 1 km
	Retail	Count of retail POIs	Yes	37.25 (23.67)	
	School	Count of school POIs (per kilometer)	Yes	7.39 (4.23)	
	Elementary school	Count of elementary school POIs	Yes	1.71 (0.87)	
	Secondary school	Count of secondary school POIs	Yes	1.55 (0.85)	
	Park	Count of park POIs	Yes	0.42 (0.12)	
	Public service	Count of public service POIs	Yes	3.04 (1.71)	
	Distance to Central	Distance to Central (kilometer)	No	13.3 (7.93)	
	Density of bus stop	Count of bus stops per square kilometer	Yes	16.47 (12.14)	100 m, 200 m, 500 m, 1 km
	Distance to the nearest rail transit station	Distance to the nearest rail transit station (kilometer)	No	0.92 (1.17)	

models based on by two existing annotated image datasets, i.e., Cityscapes (Cordts et al., 2016) and ADE20K (Zhou et al., 2017). With these two criteria, we select microscale built-environment features including greenery, open sky, buildings, roads, sidewalks, traffic lights, traffic signs, cars, bicycles, and trucks. We test the pretrained model on our test dataset which contains 2000 SVIs. This test dataset was randomly selected from our targeted dataset and labelled following the rules of ADE20K. The model achieves a pixel accuracy of 81.31 %.

We exclude the feature of "person" from SVIs as it is directly associated with the pedestrian count in our study. For each category, we calculated the proportion of pixels corresponding to the specific object in relation to the total pixel count (except the pixel of "person") in the image.

2.3. Macroscale environmental factors

The selection of macroscale environmental variables was based on the 5Ds framework (Ewing & Cervero, 2010; Handy et al., 2005; Kang, 2018). These variables include density, diversity, design, destination accessibility, and distance to transit. Density contains the population density; diversity was measured by the entropy of land use diversity; design was evaluated by the density of street intersections, Normalized Difference Vegetation Index (NDVI); destination accessibility was measured by the count of POIs of different destinations, the distance to Central (the central business center of Hong Kong); Distance to transit was measured by the count of bus stops and the distance to the nearest rail transit station.

Population density data was sourced from Census and Statistics Department of Hong Kong, while all other macroscale environmental factors were calculated using GIS data obtained from Land Department of Hong Kong in 2021. The operational definition of each variable is described in Table 1.

2.4. Independent variables

The variables are grouped into two categories: buffer relevant and buffer irrelevant variables. Buffer-relevant variables refer to those that can vary across different buffers, e.g., population density and road intersection density varying across different buffers (100-meter, 200-meter, 500-meter, and 1000-meter) for the same sampling points. On the other hand, buffer irrelevant variables refer to these are constant across different buffers including the distance to Central (the central business district in Hong Kong) and the distance to the nearest rail transit stations. A comprehensive list of all the variables is summarized in Table 1.

To analyze, we first divided all predictor variables into four categories according to their scales and whether are built-environment variables (Fig. 4). Among them, macroscale variables are grouped into neighborhood-built environmental variables and sociodemographic variables. Neighborhood-built environmental variable includes building density, land use mix, the density of road intersections, the distance to the nearest rail transit stations, the distance to Central, and the count of surrounding restaurants, retails, bus stops, schools, primary schools, secondary schools, and public services. Sociodemographic variables encapsulate all non-built-environmental macroscale factors including the density of all population, old population, new-born children, the density of jobs, and the median household income. Microscale variables are divided into two categories: street-level built environment variables and traffic-related variables. Street-level built environment variable comprises all static environmental elements detected in SVIs including buildings, vegetation, sky, sidewalks, traffic lights, and traffic signs. Traffic-related variable encompasses all non-built-environmental microscale factors including cars, riders, bicycles, and trucks.

Subsequently, variables were classified into three distinct categories following a walk-along study conducted by (Močnik et al., 2022): facilitator, barrier, and other (Fig. 5). Facilitators represent those who make neighborhoods attractive for walking or improve accessibility, which includes restaurants, retail, public services, schools, sidewalks, bus stop density, and vegetation. On the contrary, barriers portray those who have the potential to undermine the accessibility or sense of safety, comfort, and convenience including distance to Central, distance to rail transit stations, and traffic volume (traffic-volume-related variables). Other contains all other variables that exhibit dual impacts, either positive or negative, or whose effects on pedestrian conduct remain unverified.

2.5. Gradient boosting decision trees

In this study, we deployed a machine learning technique, namely Gradient Boosting Decision Trees (GBDT) (Ke et al., 2017), to assess the non-linear link between environmental factors and pedestrian demand. Addressing the constraints of standard linear assumptions between dependent and independent variables, GBDT accommodates intricate non-linear associations, which is often the case with various factors from multiple sources in urban studies such as population density, land use diversity, and access to facilities.

As a non-parametric technique, GBDT makes no assumption about independence among observations. It subdivides data and explanatory variables based on the most informative features, independent of their spatial relationships. Therefore, it has no problem with spatial autocorrelation. Integrating the straightforwardness of decision trees with flexibility, GBDT selectively employs a subset of instances and predictor variables for one individual tree. This effectively reduces variance, demonstrating robustness in the face of anomalies and statistical

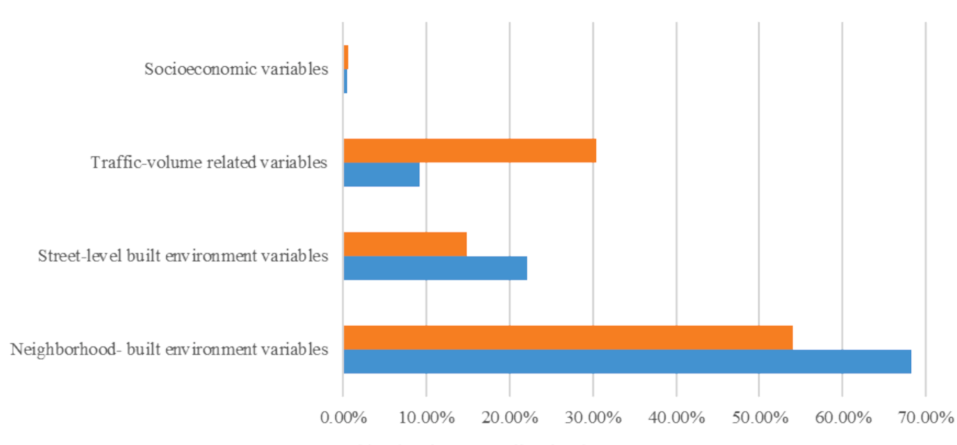

Fig. 4. Comparative visualization of the RI of four variable categories pertaining to both the number of all pedestrians and older pedestrians.

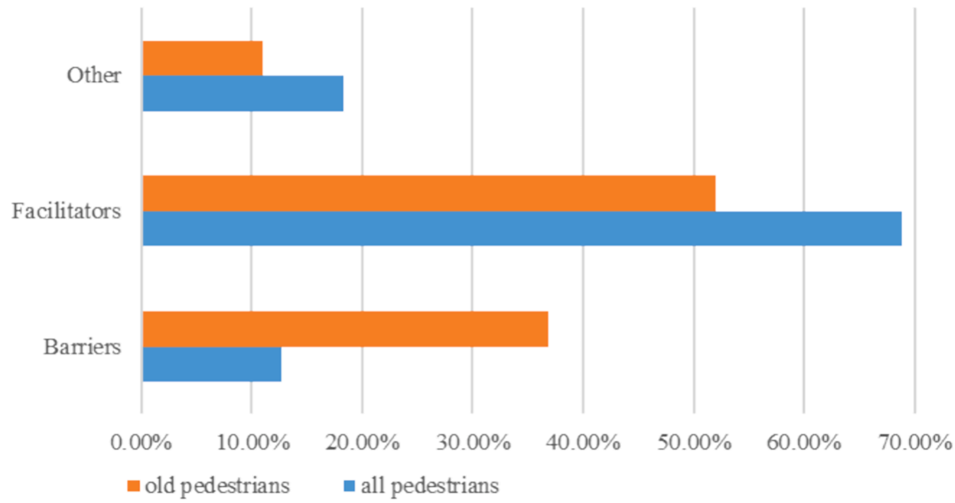

Fig. 5. Comparative visualization of the RI of facilitators, barriers, and others pertaining to both the number of all pedestrians and older pedestrians.

outliers. This subsequently enhances stability and precision. Notably, its obviation of overfitting and favorable calibration for unseen data, as a consequence of the law of large numbers, underscores its reliability and adaptability.

Contrasting with traditional linear regression methodologies, GBDT offers no t-statistics, p-values, or indicators of statistical significance (Ke et al., 2017). Its key output is relative importance (RI), representing the significance of each independent variable in determining the outcome variable (Ke et al., 2017). RI of a specific variable records the mean decrease in model fit (R^2) when this variable is permuted across all observations. An additional output of GBDT modeling is partial dependence plots (PDPs), elucidating the relationship between the dependent and independent variables contingent on the levels of independent variables. When applied to our studies, RI and PDPs can be instrumental in understanding the specific factors that influence pedestrian volume. In this study, we compared changes in the RI ranks across models. The changes of RI rankings of the same factor across models reveal change of relative contribution of this factor. Following previous studies, we can identify which environmental factors have more or less influence on pedestrian volume across different buffers by analyzing the changes in the RI ranks (Fisher et al., 2019; Hu et al., 2023; Liu et al., 2023b). PDPs illustrate the relationship between the dependent variable and independent variables, considering the varying levels of the independent variables. PDPs provide a visual representation of how changes in independent variables impact the predicted outcome, offering valuable insights into the complex interactions within the model. This comprehensive approach allows for a nuanced understanding of the relationship between environmental factors and pedestrian volume.

To utilize the random forest algorithm, it is necessary to define or optimize three particular parameters. These parameters contain the maximum depth of a tree, the number of features (referred to as splitting variables) allocated to each tree, and the overall count of trees comprising the forest. We employed a widely used technique known as grid search (Claesen & De Moor, 2015) to identify the optimal combination of these three parameters.

Initially, we define the possible intervals for the three parameters: maximum tree depth (scaling from 1 to 20), the number of features per tree (ranging from 2 to 20), and the number of trees (varying from 1000 to 30,000 in increments of 1000). Subsequently, the data is partitioned into training and testing datasets adhering to a ratio of 7:3. The GBDT model is then trained on the former dataset, with its performance evaluated on the latter utilizing out-of-bag error as the metric for model performance (L. Cheng et al., 2020b). The model exhibits optimal performance when the maximum tree depth is 12, the feature number for each tree is also 12, and the count of trees is 21000. Consequently, additional analyses are executed with this optimal model configuration.

3. Results

3.1. Relative importance (RI)

Initially, we implemented the GBDT utilizing predictor variables across all spatial resolutions for both the number of all pedestrians and older pedestrians. In each model, we calculated the RI of each object by the number or the average pixel proportion of it across all buffer zones of the sampling points. For instance, the RI of restaurants is computed as the cumulative number of the restaurants situated within the 100-meter, 200-meter, 500-meter, and 1-kilometer buffers of each sampling point. The RI of vegetation is computed as the average pixel proportion of vegetation in the SVI, and all SVIs sampled within the 100-meter, 200-meter, 500-meter, and 1-kilometer buffers of each sampling point.

We visualized the RI of the 15-highest predictor variables on the number of both all pedestrians and older pedestrians, respectively (Fig. 3). The count of restaurants is the most influential for both all pedestrians (36.6 %) and older pedestrians (28.9 %). Besides the count of restaurants, the count of retail and second schools, the pixel proportion of vegetation, traffic lights, traffic signs, cars, riders, buildings, bicycles, motorcycles, sidewalks, and the distance to the nearest rail transit station also belong to the 15-highest predictor variables of both response variables. The distance to Central and the count of bus stops are only the 15-highest predictor variable of all pedestrians, while the number of older residents (≥65 years old) and the count of public services are only in the list of older pedestrians.

Neighborhood- built environment variable has the highest RI for both the number of all pedestrians and older pedestrians, while socioeconomic variable is the least important for both. Compared to the number of all pedestrians, that of older pedestrians are more influenced by traffic-volume related variable.

The facilitator group is the most significant for both the number of all pedestrians and older pedestrians. Nonetheless, compared with the number of all pedestrians, that of older pedestrians exhibits heightened sensitivity to the barriers while being less susceptible to the facilitators.

3.2. Model fit of model across spatial resolutions

Subsequently, we employed GBDT with respective predictors for each spatial resolution: 100-meter, 200-meter, 500-meter, 1-kilometer, and in the image of the sampling point. Table 2 presents the model fit (R^2) of the models at varied spatial resolutions. Generally speaking, models of smaller buffer size exhibited better model fit in terms of R^2

Table 2
Model fit and the highest five variables of models across the different buffer sizes.

	100 *m*		200 *m*		500 *m*		1000 *m*		In the image	
	All	Elderly	All	Elderly	All	Elderly	All	Elderly	All	Elderly
Model fit (R^2)	0.258	0.234	0.168	0.172	0.101	0.112	0.051	0.052	0.295	0.275
The five variables with the highest relative importance										
1	restaurant	restaurant	restaurant	car	restaurant	second school	vegetation	public services	traffic light	traffic light
2	retail	retail	car	restaurant	traffic light	traffic light	restaurant	vegetation	vegetation	rider
3	traffic light	car	traffic sign	traffic sign	retail	retail	retail	building	building	motorcycle
4	car	bicycle	traffic light	building	distance to Central	restaurant	car	density of road intersection	traffic sign	vegetation
5	vegetation	vegetation	building	vegetation	car	old population	bus stops	Second school	distance to Central	sidewalk

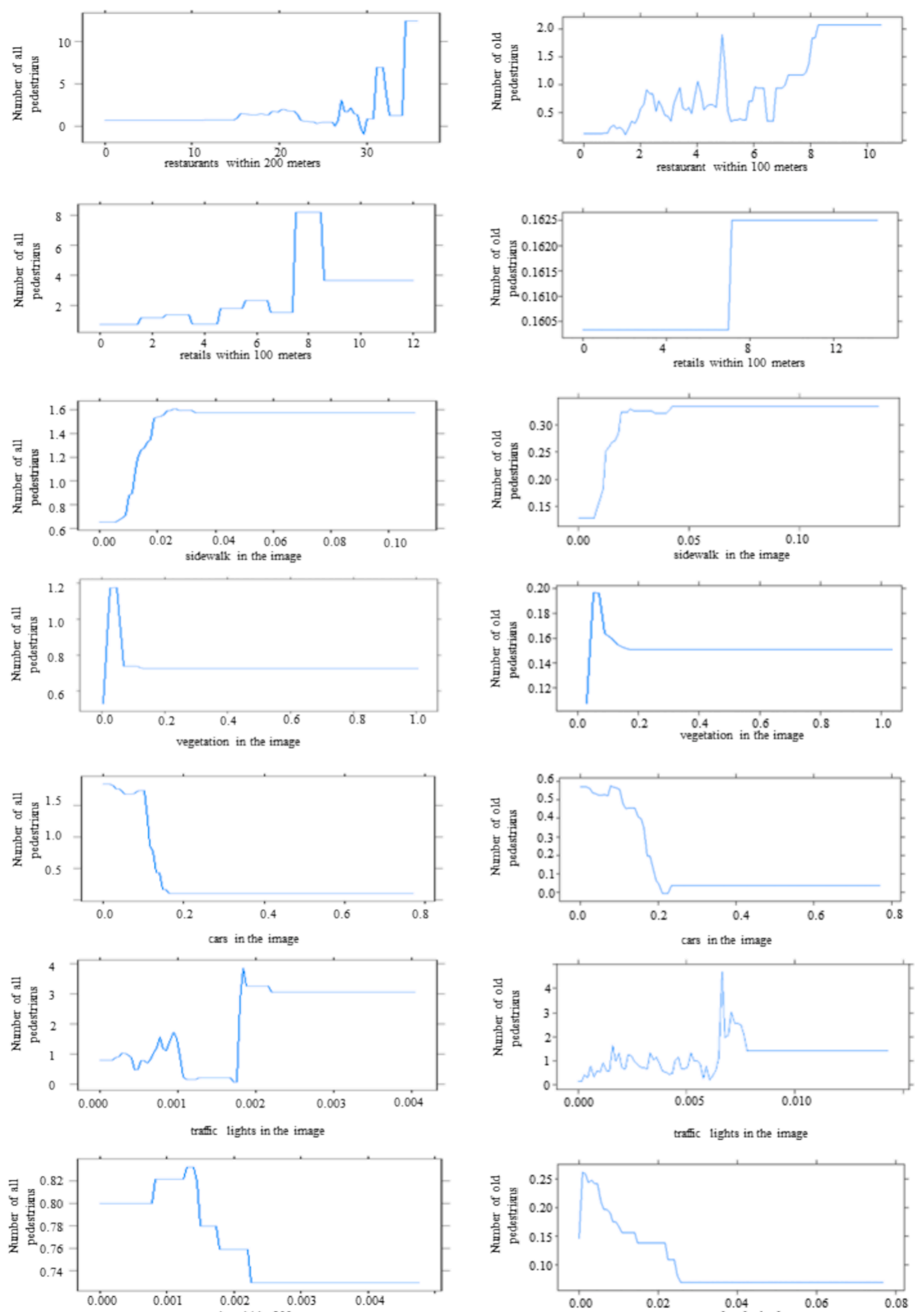

Fig. 6. PDPs of the 7 predictor variables with the highest RI on both the volume of all pedestrians (left) and the volume of older pedestrians (right).

value. The model using variables in the image attained the highest R^2 value for both the number of all pedestrians (0.295) and the older pedestrians (0.275), while models with the 1000 m buffer yield the smallest R^2 value for both outcomes.

The rank of the RI of variables in models with different spatial resolutions is distinct from one another. Variables such as restaurants, retail, vegetation, traffic lights, traffic signs, and cars constitute the most frequent occurrences within the list. Certain variables, including traffic signs and traffic lights, hold greater relative significance within models of smaller spatial resolution. Conversely, variables such as vegetation are of higher RI in models exhibiting larger spatial resolution.

3.3. Partial dependence plots (PDPs)

Besides, GBDT also produces Partial Dependence Plots (PDP) for each predictive variable. In this study, PDPs were demonstrated for the top seven items that have the highest RI on the both total volume of pedestrians and the volume of older pedestrians (Fig. 6) correspondingly. For each such influential item, the variable within the specific spatial resolution that exhibits the highest RI has been chosen for analysis.

The PDPs of these variables can be divided into three types: positive, negative, and winding. The first category includes predictor variables that maintain a broad positive relationship with the independent variables except for some slight fluctuations. For the count of all pedestrians, this list contains the number of restaurants within 200 m, the number of retails within 100 m, and the pixel proportion of traffic lights, traffic lights, and sidewalks in the image. For the count of older pedestrians, this list includes the number of restaurants and retails within 100 m, the number of second schools within 1000 m, the number of public services within 500 m, the pixel proportion of traffic signs and sidewalks, and the population of old residents.

The second category comprises predictor variables displaying a persistent negative correlation with the dependent variable except for some slight fluctuations. For the number of all pedestrians, this comprises the pixel proportion of cars and riders in the image, the pixel proportion of motorcycles within 200 m, the pixel proportion of bicycles within 100 m, the distance to Central, and the distance to the nearest rail transit station. For the number of older pedestrians, it includes the pixel proportion of cars and motorcycles in the image.

The third category incorporates predictor variables that manifest a varied correlation with the response variable, to such an extent that drawing clear, definitive trends becomes impracticable. For instance, the Partial Dependence Plots (PDPs) representing the pixel proportion of buildings in the image exhibit a flat trend up to 0.5. Subsequently, there is an increase until 0.6, after which it decreases until 0.8, eventually flattening out. The discrepancy observed between the PDPs for all pedestrians and older pedestrians lies in the end value of the y-axis. For all pedestrians, the conclusion value is lower than the initial value, whereas for older pedestrians, the terminating value is higher than the onset value. In the PDPs of the pixel proportion of vegetation in the image, the counts of both all and older pedestrians demonstrate parallel trends. Within both PDPs, the number of pedestrians rapidly ascends to a peak, maintains this peak briefly, and then decreases by approximately 50 percent of the increment, followed by a sustained plateau until the end. For the count of all pedestrians, the count reaches its apex when the pixel proportion of vegetation is roughly 0.04, then descends to the plateau phase when the pixel proportion of vegetation is around 0.08. In contrast, for the count of older pedestrians, the count achieves its maximum when the pixel proportion of vegetation is approximately 0.06 and diminishes to the level phase when the pixel proportion of vegetation is at about 0.018.

4. Discussion

4.1. Discrepancy in the impact of built environmental variables on all and older pedestrians

Numerous studies have identified that individual physical and cognitive abilities can moderate the impact of environmental variables on individual walking behavior (De Vos et al., 2023). These abilities often correlate with age, particularly among older adults (Lui & Wong, 2021; Wang & Lee, 2010). Utilizing the GBDT, this study seeks to assess how the RI and PDPs of specific environmental variables may differ between older pedestrians and all pedestrians.

Some features exhibit evidently different RI between all pedestrians and older pedestrians. Among the 15 variables with the highest RI rankings, bicycles, public services, and the old resident population are unique for older pedestrians. The association between the count of older residents and older pedestrians is unsurprising. Public services encompass healthcare facilities, social services, and public cultural amenities. Old individuals use these services, especially healthcare facilities, frequently. Hong Kong government poured resources into public services in recent decades to cope with the aging population (Guida et al., 2022). Therefore, there is a strong link between the public services and older pedestrians. Aside from bicycles, other traffic-related variables like trucks, cars, and motorcycles also demonstrated higher RI rankings for older pedestrians than for all pedestrians. High traffic volume can deter elder individuals from walking, largely due to the heightened risk perception (Won et al., 2016). Conversely, other pedestrians, who may not be as sensitive to these risks, are less impacted by these traffic-related factors (Wang et al., 2019). Furthermore, the same features manifest different “true causally relevant” areas (Kwan, 2012) for all pedestrians and older pedestrians. For instance, while the number of restaurants exerts influence on both models, it exhibits the highest RI rankings within a 200-meter buffer for all pedestrians and a 100-meter buffer for older pedestrians. It is conceivable that mobility limitations reduce the appeal of more distant restaurants for older pedestrians.

The variables were classified using two distinct methodologies. The first method grouped variables by their nature and scales: neighborhood-level (or macroscale) built environment, street-level (or microscale) built environment, traffic-related, and sociodemographic variables. Our results suggest that compared to all pedestrians, older pedestrians exhibit stronger associations with traffic-related elements while showing weaker associations with neighborhood-level built environment variables. This observation aligns with prior research indicating that older pedestrians are predominantly influenced by microscale built environments (Adkins et al., 2017; Forsyth et al., 2009). Conversely, younger or more physically adept pedestrians are largely impervious to these traffic-related factors but are more affected by macroscale factors. There are two possible reasons. First, older adults exhibit heightened sensitivity to microscale environmental and traffic factors due largely to frequently decreased physical and cognitive abilities (Adkins et al., 2017). Age-related changes in vision, hearing, strength, balance, and cognitive capacity might increase the risk of falls and traffic-related accidents (Adkins et al., 2017). Second, the limited mobility of older people curtails their walking range, limiting their access to remote destinations and infrastructure (Adkins et al., 2017; Wei et al., 2023). Therefore, the impact of microscale factors overrides that of macroscale factors.

The second method categorized variables by their impacts on the walking behavior of pedestrians according to previous qualitative studies: facilitator, barrier, and others. The results show that the number of older pedestrians has stronger associations with barriers but weaker with facilitators compared with the count of all pedestrians. An assumption asserts that older and disadvantaged people may exhibit lower walking propensity due to potential barriers to walking, such as apprehensions for crime or traffic safety, which diminish the benefits of supportive built environment determinants, thereby deterring walking behavior (Adkins et al., 2017). For example, an older adult may be reluctant to walk because of concerns about criminal or traffic safety, even living in a place with well-connected street networks accessing varying destinations.

4.2. Varying effects of environment factors across different spatial resolutions

Previous research examining the non-linear implications of environmental factors on walking behaviors mainly employed dependent variables at a single uniform spatial resolution, such as a 500-meter radius. This study, however, employed predictors across a spectrum of spatial scales, ranging from site-specific (as in the images) to buffers of 100 m, 200 m, 500 m, and 1 km. Our findings indicate that environmental factors exert different influences with varying buffers.

The changes of RI rankings of a variable across different buffer sizes reveal decrease or increase of contribution of a factor across different buffer sizes, and help us identify which factor has a significant impact on pedestrian volume across different buffers (Kwan, 2012). For example, if a factor has high RI rankings across all buffer sizes, we can infer the contribution of this factor is high and stable. In other words, a feature with downward RI rankings across buffer size may own to weak intrinsic correlations or inherent noise (Hu et al., 2023).

For variables representing destinations such as restaurants and retail, a clear pattern emerged wherein their RI rank dropped as buffer distance increased. This downward trend in RI rank manifested more rapidly for the number of older pedestrians compared to that of all pedestrians. To illustrate, the number of restaurants ranks first within 100 m, 200 m, and 500 m, ranking second within the 1000-meter radius for all pedestrians. In contrast, the rank of restaurant dropped more precipitously for older pedestrians, commencing at the highest rank within a 100-meter radius, then shifting to a second rank at 200 m, dropping to the fourth at 500 m, and finally, descending outside the top five within a 1000-meter radius.

Contrarily, variables such as vegetation exhibit an inverse correlation, with a rising rank in RI with larger buffer sizes, namely 500 m and 1000 m, compared to those ascertained onsite, 100 m, and 200 m. Despite not being a destination in itself, vegetation has been documented to enhance walkability and engender walking activity. It seems that the benefits provided by vegetation might not solely be directly attributable to vegetation visual exposure itself but could also be indirectly the ecological services like improving air quality and biodiversity, which is associated with a broader spatial range of impacts.

4.3. Partial dependence plots (PDPs) and non-linear effects

Conventional linear regression analyses have yielded substantial empirical data delineating both the negative and positive implications of environmental elements on travel patterns. However, these associations may not invariably exhibit linearity. Certain predictive variables may demonstrate symmetrical and inverted U-shaped interactions with outcome variables, manifesting effective ranges and breakpoints in non-linear ramifications across predictors and outputs.

Past research illustrates that vegetation can promote walking and physical activities. The utilization of Partial Dependence Plots (PDPs) elucidates similar patterns for both the number of all pedestrians and older pedestrians. Firstly, it is observed that the outcome rapidly elevates towards a peak, which can be regarded as the maximum benefit of vegetation. Then, there is a decline to stable effect, which is approximately half of that of the peak. Some corroborative literature also ascertains that at a 0.24 Green View Index in SVIs, the benefits of vegetation on walking is optimal (Yang et al., 2021).

The pixel proportion of buildings in the image displays an inverted U-pattern correlation with the number of all pedestrians and older pedestrians. This pixel proportion of buildings has an association with the building density, potentially serving as an approximation for urban density. Traditional hypotheses maintained that pedestrian activity was invariably positively influenced by urban density because that increased density implied an abundant availability of adjacent amenities, thereby fostering pedestrian activities. However, studies from high-density cities (L. Cheng et al., 2020a; Lu et al., 2019) have pointed out that once urban density surpasses a specific threshold, it may result in negative effects on walking activity. The rationale behind this phenomenon is that overcrowding and heightened risks of injury in extremely dense areas may discourage walking activity (L. Cheng et al., 2020a). Hence, within a reasonable range, urban density may effectively promote walking behavior. Urban density outside the range, being either too low or too high, may hinder walking.

4.4. Planning and policy implication

The intricate relationship between pedestrian dynamics and urban planning features discovered in this research should inform a series of urban and transportation policies. Specific actions include modifying urban infrastructure to remove physical obstacles like uneven pathways and incorporating features such as benches and improved lighting. Additionally, traffic policies need to be reassessed, potentially resulting in the implementation of reduced speed limits and longer crossing times in areas heavily used by older individuals. Urban planning should carefully consider the balance between residential density and green spaces, guided by the demonstrated inverted U-shaped relationship to pedestrian traffic levels, ensuring that these elements are coordinated to promote secure and inviting pedestrian pathways. Furthermore, the diverse impacts of environmental factors at different scales require customized interventions, such as targeted tree plantings and the strategic placement of public services, to directly address the unique needs of specific groups of users. Overall, these recommendations advocate for an urban approach that prioritizes pedestrian-centered design, particularly customizing environments to facilitate the movement and safety of aging populations.

4.5. Limitations and future studies

First, this study only extracted one attribute of pedestrians overlooking the significance of other attributes such as gender and disability. Future research endeavors should allocate more attention to these additional attributes. Second, the generalization of our study's findings to other places is uncertain due to Hong Kong's unique ultra-dense, mixed land-use urban fabric. Further empirical studies conducted in diverse settings are imperative to validate or challenge our results. Third, the study only illustrated the association between environmental factors and pedestrian volume, it necessitates longitudinal studies to establish causality linkage among variables. Fourth, it is essential to acknowledge that pedestrian volume itself serves as an environmental factor influencing pedestrian behavior. Future investigations should aim to comprehend the bidirectional relationship involved therein.

5. Conclusion

This research delved into the non-linear relationship between environmental factors and pedestrian volume for both all pedestrians and older pedestrians, which are estimated from Street View Imagery (SVI) at citywide sampling locations. It incorporates a broad range of environmental factors from multiple data sources and at different buffer sizes to predict pedestrian volumes. We find that environmental variables can impact different age groups at varying spatial extents, notably highlighting the distinctive influences of traffic-volume-related variables on older pedestrians. To create walking-friendly cities, the results of the study suggested that urban planners and designers should consider the diverse needs of pedestrians and the effective ranges of environmental features beyond one-size-fits-all, which may exhibit non-linear effects with walking behavior.

Funding

The work described in this paper was fully supported by the Research Grants Council of the Hong Kong SAR (Project No. CityU11207520).

CRediT authorship contribution statement

Dongwei Liu: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft. **Yi Lu:** Project administration, Supervision, Writing – review & editing. **Linchuan Yang:** Supervision, 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.

References

Adkins, A., Makarewicz, C., Scanze, M., Ingram, M., Luhr, G., 2017. Contextualizing walkability: do relationships between built environments and walking vary by socioeconomic context? J. Am. Plann. Assoc. 83 (3), 296–314.
Ahlport, K.N., Linnan, L., Vaughn, A., Evenson, K.R., Ward, D.S., 2008. Barriers to and facilitators of walking and bicycling to school: formative results from the non-motorized travel study. Health Educ. Behav. 35 (2), 221–244.
Arellana, J., Saltarín, M., Larrañaga, A.M., Alvarez, V., Henao, C.A., 2020. Urban walkability considering pedestrians' perceptions of the built environment: a 10-year review and a case study in a medium-sized city in Latin America. Transp. Rev. 40 (2), 183–203. https://doi.org/10.1080/01441647.2019.1703842.
Aune, D., Sen, A., Henriksen, T., Saugstad, O.D., Tonstad, S., 2016. Physical activity and the risk of gestational diabetes mellitus: a systematic review and dose–response meta-analysis of epidemiological studies. Eur. J. Epidemiol. 31 (10), 967–997.
Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: a review. Landsc. Urban Plan. 215, 104217.
Boarnet, M.G., Forsyth, A., Day, K., Oakes, J.M., 2011. The street level built environment and physical activity and walking: results of a predictive validity study for the Irvine Minnesota Inventory. Environ. Behav. 43 (6), 735–775.
Boyle, T., Keegel, T., Bull, F., Heyworth, J., Fritschi, L., 2012. Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. J. Natl Cancer Inst. 104 (20), 1548–1561.
Caceres, N., Romero, L.M., Benitez, F.G., del Castillo, J.M., 2012. Traffic flow estimation models using cellular phone data. IEEE Trans. Intell. Transp. Syst. 13 (3), 1430–1441.
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. J. Transp. Health 16, 100797.
Carr, L.J., Dunsiger, S.I., Marcus, B.H., 2010. Walk score™ as a global estimate of neighborhood walkability. Am. J. Prev. Med. 39 (5), 460–463.
Cesare, N., Nguyen, Q.C., Grant, C., Nsoesie, E.O., 2019. Social media captures demographic and regional physical activity. BMJ Open Sport Exerc. Med. 5 (1), e000567.
Chen, S., Bao, Z., Lou, V., 2022a. Assessing the impact of the built environment on healthy aging: A gender-oriented Hong Kong study. Environ. Impact Assess. Rev. 95, 106812.
Chen, L., Lu, Y., Sheng, Q., Ye, Y., Wang, R., Liu, Y., 2020. Estimating pedestrian volume using Street View images: a large-scale validation test. Comput. Environ. Urban Syst. 81, 101481.
Chen, L., Lu, Y., Ye, Y., Xiao, Y., Yang, L., 2022b. Examining the association between the built environment and pedestrian volume using street view images. Cities 103734.
Chen, L., Zhao, L., Xiao, Y., Lu, Y., 2022c. Investigating the spatiotemporal pattern between the built environment and urban vibrancy using big data in Shenzhen, China. Comput. Environ. Urban Syst. 95, 101827.
Cheng, L., De Vos, J., Zhao, P., Yang, M., Witlox, F., 2020. Examining non-linear built environment effects on elderly's walking: a random forest approach. Transp. Res. Part D: Transp. Environ. 88, 102552.
Cheng, Z., Jian, S., Rashidi, T.H., Maghrebi, M., Waller, S.T., 2020b. Integrating household travel survey and social media data to improve the quality of OD matrix: a comparative case study. IEEE Trans. Intell. Transp. Syst. 21 (6), 2628–2636. https://doi.org/10.1109/TITS.2019.2958673.
Christiansen, L.B., Cerin, E., Badland, H., Kerr, J., Davey, R., Troelsen, J., Van Dyck, D., Mitáš, J., Schofield, G., Sugiyama, T., 2016. International comparisons of the associations between objective measures of the built environment and transport-related walking and cycling: IPEN adult study. J. Transp. Health 3 (4), 467–478.
Claesen, M., & De Moor, B. (2015). Hyperparameter search in machine learning. arXiv Preprint arXiv:1502.02127.
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.
De Vos, J., Lättman, K., Van der Vlugt, A.-L., Welsch, J., Otsuka, N., 2023. Determinants and effects of perceived walkability: a literature review, conceptual model and research agenda. Transp. Rev. 43 (2), 303–324.
Delclòs-Alió, X., Rodríguez, D.A., Medina, C., Miranda, J.J., Avila-Palencia, I., Targa, F., Moran, M.R., Sarmiento, O.L., Quistberg, D.A., 2022. Walking for transportation in large Latin American cities: walking-only trips and total walking events and their sociodemographic correlates. Transp. Rev. 42 (3), 296–317.
Du, Y., Yue, J., Ji, Y., Sun, L., 2017. Exploration of optimal Wi-Fi probes layout and estimation model of real-time pedestrian volume detection. Int. J. Distrib. Sens. Netw. 13 (11), 1550147717741857.
Dunkel, A., 2015. Visualizing the perceived environment using crowdsourced photo geodata. Landsc. Urban Plan. 142, 173–186.
Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. J. Am. Plann. Assoc. 76 (3), 265–294.
Ewing, R., Handy, S., 2009. Measuring the unmeasurable: urban design qualities related to walkability. J. Urban Des. 14 (1), 65–84.
Feng, J., 2017. The influence of built environment on travel behavior of the elderly in urban China. Transp. Res. Part D: Transp. Environ. 52, 619–633.
Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously.
Forsyth, A., Oakes, J.M., Lee, B., Schmitz, K.H., 2009. The built environment, walking, and physical activity: is the environment more important to some people than others? Transp. Res. Part D: Transp. Environ. 14 (1), 42–49.
Frank, L.D., Kerr, J., Sallis, J.F., Miles, R., Chapman, J., 2008. A hierarchy of sociodemographic and environmental correlates of walking and obesity. Prev. Med. 47 (2), 172–178.
Galea, M.N., Bray, S.R., Ginis, K.A.M., 2008. Barriers and facilitators for walking in individuals with intermittent claudication. J. Aging Phys. Act. 16 (1), 69–84.
Gariazzo, C., Pelliccioni, A., Bolignano, A., 2016. A dynamic urban air pollution population exposure assessment study using model and population density data derived by mobile phone traffic. Atmos. Environ. 131, 289–300.
Ghani, F., Rachele, J.N., Loh, V.H., Washington, S., Turrell, G., 2018. Do differences in built environments explain age differences in transport walking across neighbourhoods? J. Transp. Health 9, 83–95.
Guida, C., Carpentieri, G., Masoumi, H., 2022. Measuring spatial accessibility to urban services for older adults: An application to healthcare facilities in Milan. Eur. Transp. Res. Rev. 14 (1), 23.
Guzman, L.A., Peña, J., Carrasco, J.A., 2020. Assessing the role of the built environment and sociodemographic characteristics on walking travel distances in Bogotá. J. Transp. Geogr. 88, 102844.
Handy, S., Cao, X., Mokhtarian, P., 2005. Correlation or causality between the built environment and travel behavior? Evidence from Northern California. Transp. Res. Part D: Transp. Environ. 10 (6), 427–444. https://doi.org/10.1016/j.trd.2005.05.002.
Hankey, S., Lindsey, G., Marshall, J.D., 2017. Population-level exposure to particulate air pollution during active travel: planning for low-exposure, health-promoting cities. Environ. Health Perspect. 125 (4), 527–534.
Harvey, J.A., Chastin, S.F., Skelton, D.A., 2015. How sedentary are older people? A systematic review of the amount of sedentary behavior. J. Aging Phys. Act. 23 (3).
Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. 41 (3), 260–271.
He, D., Lu, Y., Xie, B., Helbich, M., 2022. How greenway exposure reduces body weight: a natural experiment in China. Landsc. Urban Plan. 226, 104502.
Hearst, M.O., Sirard, J.R., Forsyth, A., Parker, E.D., Klein, E.G., Green, C.G., Lytle, L.A., 2013. The relationship of area-level sociodemographic characteristics, household composition and individual-level socioeconomic status on walking behavior among adults. Transp. Res. A Policy Pract. 50, 149–157.
Hu, S., Xiong, C., Chen, P., Schonfeld, P., 2023. Examining nonlinearity in population inflow estimation using big data: an empirical comparison of explainable machine learning models. Transp. Res. A Policy Pract. 174, 103743 https://doi.org/10.1016/j.tra.2023.103743.
Huang, G., Li, D., Yu, L., Yang, D., Wang, Y., 2022. Factors affecting sustainability of smart city services in China: From the perspective of citizens' sense of gain. Habitat International 128, 102645.
Im, H.N., Choi, C.G., 2019. The hidden side of the entropy-based land-use mix index: clarifying the relationship between pedestrian volume and land-use mix. Urban Stud. 56 (9), 1865–1881.
Ippersiel, P., Robbins, S.M., Dixon, P.C., 2021. Lower-limb coordination and variability during gait: the effects of age and walking surface. Gait Posture 85, 251–257.
Jacobs, J. (1961). The uses of city neighborhoods.
Jiang, Y., Chen, L., Grekousis, G., Xiao, Y., Ye, Y., Lu, Y., 2021. Spatial disparity of individual and collective walking behaviors: a new theoretical framework. Transp. Res. Part D: Transp. Environ. 101, 103096.
Kang, C.-D., 2018. The S+ 5Ds: spatial access to pedestrian environments and walking in Seoul, Korea. Cities 77, 130–141.
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. Advances in Neural Information Processing Systems, 30.
Kim, Y.-L., 2018. Seoul's Wi-Fi hotspots: Wi-Fi access points as an indicator of urban vitality. Comput. Environ. Urban Syst. 72, 13–24.
Kwan, M.-P., 2012. The uncertain geographic context problem. Ann. Assoc. Am. Geogr. 102 (5), 958–968.

Lättman, K., Olsson, L.E., Friman, M., 2018. A new approach to accessibility–examining perceived accessibility in contrast to objectively measured accessibility in daily travel. Res. Transp. Econ. 69, 501–511.
Lee, S., Sung, H., Woo, A., 2017. The spatial variations of relationship between built environment and pedestrian volume: focused on the 2009 Seoul Pedestrian flow survey in Korea. Journal of Asian Architecture and Building Engineering 16 (1), 147–154.
Li, Z., Lu, Y., Zhuang, Y., Yang, L., 2024. Influencing factors of spatial vitality in underground space around railway stations: A case study in Shanghai. Tunn. Undergr. Space Technol. 147, 105730.
Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. Urban For. Urban Green. 14 (3), 675–685.
Li, X., Li, Y., Jia, T., Zhou, L., Hijazi, I.H., 2022. The six dimensions of built environment on urban vitality: fusion evidence from multi-source data. Cities 121, 103482.
Lian, T., Loo, B.P., Fan, Z., 2024. Advances in estimating pedestrian measures through artificial intelligence: from data sources, computer vision, video analytics to the prediction of crash frequency. Comput. Environ. Urban Syst. 107, 102057.
Liu, X., Chen, X., Tian, M., De Vos, J., 2023b. Effects of buffer size on associations between the built environment and metro ridership: a machine learning-based sensitive analysis. Journal of Transport Geography 113, 103730. https://doi.org/10.1016/j.jtrangeo.2023.103730.
Liu, D., Jiang, Y., Wang, R., Lu, Y., 2023c. Establishing a citywide street tree inventory with street view images and computer vision techniques. Comput. Environ. Urban Syst. 100, 101924.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social sensing: a new approach to understanding our socioeconomic environments. Ann. Assoc. Am. Geogr. 105 (3), 512–530.
Liu, J., Wang, B., Xiao, L., 2021. Non-linear associations between built environment and active travel for working and shopping: an extreme gradient boosting approach. J. Transp. Geogr. 92, 103034.
Liu, D., Wang, R., Grekousis, G., Liu, Y., Lu, Y., 2023a. Detecting older pedestrians and aging-friendly walkability using computer vision technology and street view imagery. Comput. Environ. Urban Syst. 105, 102027.
Lockett, D., Willis, A., Edwards, N., 2005. Through seniors' eyes: an exploratory qualitative study to identify environmental barriers to and facilitators of walking. Canadian Journal of Nursing Research Archive 48–65.
Lu, Y., 2019. Using Google Street View to investigate the association between street greenery and physical activity. Landsc. Urban Plan. 191, 103435.
Lu, Y., Sun, G., Gou, Z., Liu, Y., Zhang, X., 2019. A dose–response effect between built environment characteristics and transport walking for youths. J. Transp. Health 14, 100616.
Lui, C.K., Wong, T.W., 2021. Associations among perceived walkability of neighborhood environment, walking time, and functional mobility by older adults: An exploratory investigation. Ageing Int. 1–14.
Lynch, K., 1984. Good city form. MIT press.
Maioli, H.C., de Carvalho, R.C., de Medeiros, D.D., 2019. SERVBIKE: riding customer satisfaction of bicycle sharing service. Sustain. Cities Soc. 50, 101680.
Manson, J.E., Hu, F.B., Rich-Edwards, J.W., Colditz, G.A., Stampfer, M.J., Willett, W.C., Speizer, F.E., Hennekens, C.H., 1999. A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. N. Engl. J. Med. 341 (9), 650–658.
Middleton, J., 2018. The socialities of everyday urban walking and the 'right to the city'. Urban Stud. 55 (2), 296–315.
Močnik, Š., Moogoor, A., Yuen, B., 2022. Exploring facilitators and barriers of older adults' outdoor mobility: a walk-along study in Singapore. J. Transp. Health 26, 101386.
Morstatter, F., Liu, H., 2017. Discovering, assessing, and mitigating data bias in social media. Online Social Networks and Media 1, 1–13. https://doi.org/10.1016/j.osnem.2017.01.001.
Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., Inoue, S., 2020. Objective scoring of streetscape walkability related to leisure walking: statistical modeling approach with semantic segmentation of Google Street View images. Health Place 66, 102428.
Pellicer-Chenoll, M., Pans, M., Seifert, R., López-Cañada, E., García-Massó, X., Devís-Devís, J., González, L.-M., 2021. Gender differences in bicycle sharing system usage in the city of Valencia. Sustain. Cities Soc. 65, 102556.
Rantakokko, M., Iwarsson, S., Mänty, M., Leinonen, R., Rantanen, T., 2012. Perceived barriers in the outdoor environment and development of walking difficulties in older people. Age Ageing 41 (1), 118–121.
Roberts, H., Helbich, M., 2021. Multiple environmental exposures along daily mobility paths and depressive symptoms: a smartphone-based tracking study. Environ. Int. 156, 106635.
Roberts, H., Sadler, J., Chapman, L., 2019. The value of Twitter data for determining the emotional responses of people to urban green spaces: a case study and critical evaluation. Urban Stud. 56 (4), 818–835.
Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. Med. Sci. Sports Exerc. 40 (7 Suppl), S550.
Sjögren, P., Fisher, R., Kallings, L., Svenson, U., Roos, G., Hellénius, M.-L., 2014. Stand up for health—avoiding sedentary behaviour might lengthen your telomeres: secondary outcomes from a physical activity RCT in older people. Br. J. Sports Med.
Smith, T.C., Wingard, D.L., Smith, B., Kritz-Silverstein, D., Barrett-Connor, E., 2007. Walking decreased risk of cardiovascular disease mortality in older adults with diabetes. J. Clin. Epidemiol. 60 (3), 309–317.
Song, Y., Huang, B., Cai, J., Chen, B., 2018. Dynamic assessments of population exposure to urban greenspace using multi-source big data. Sci. Total Environ. 634, 1315–1325.
Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., Ramaswami, A., 2020. Exploring the nonlinear relationship between the built environment and active travel in the Twin Cities. J. Plan. Educ. Res. 0739456X20915765.
Tcymbal, A., Demetriou, Y., Kelso, A., Wolbring, L., Wunsch, K., Wäsche, H., Woll, A., Reimers, A.K., 2020. Effects of the built environment on physical activity: a systematic review of longitudinal studies taking sex/gender into account. Environ. Health Prev. Med. 25 (1), 1–25.
Tiznado-Aitken, I., Lucas, K., Muñoz, J.C., Hurtubia, R., 2020. Understanding accessibility through public transport users' experiences: a mixed methods approach. J. Transp. Geogr. 88, 102857.
Vich, G., Marquet, O., Miralles-Guasch, C., 2019. Green exposure of walking routes and residential areas using smartphone tracking data and GIS in a Mediterranean city. Urban For. Urban Green. 40, 275–285.
Wang, Y., Chau, C.K., Ng, W.Y., Leung, T.M., 2016. A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods. Cities 50, 1–15.
Wang, Z., Lee, C., 2010. Site and neighborhood environments for walking among older adults. Health Place 16 (6), 1268–1279.
Wang, X., Shao, C., Yin, C., Guan, L., 2020. Built environment, life events and commuting mode shift: focus on gender differences. Transp. Res. Part D: Transp. Environ. 88, 102598.
Wang, R., Yuan, Y., Liu, Y., Zhang, J., Liu, P., Lu, Y., Yao, Y., 2019. Using street view data and machine learning to assess how perception of neighborhood safety influences urban residents' mental health. Health Place 59, 102186.
Wei, D., Lu, Y., Wu, X., Ho, H.C., Wu, W., Song, J., Wang, Y., 2023. Greenspace exposure may increase life expectancy of elderly adults, especially for those with low socioeconomic status. Health & Place 84, 103142.
Widener, M.J., Farber, S., Neutens, T., Horner, M.W., 2013. Using urban commuting data to calculate a spatiotemporal accessibility measure for food environment studies. Health Place 21, 1–9.
Won, J., Lee, C., Forjuoh, S.N., Ory, M.G., 2016. Neighborhood safety factors associated with older adults' health-related outcomes: a systematic literature review. Soc Sci Med 165, 177–186.
Wu, X., Lu, Y., Jiang, B., 2023. Built environment factors moderate pandemic fatigue in social distance during the COVID-19 pandemic: a nationwide longitudinal study in the United States. Landsc. Urban Plan. 233, 104690.
Wu, Y., Zhang, D., Kang, S., 2013. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. Breast Cancer Res. Treat. 137 (3), 869–882.
Xiao, L., Lo, S., Liu, J., Zhou, J., Li, Q., 2021. Nonlinear and synergistic effects of TOD on urban vibrancy: applying local explanations for gradient boosting decision tree. Sustain. Cities Soc. 72, 103063.
Yang, L., Ao, Y., Ke, J., Lu, Y., Liang, Y., 2021. To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. J. Transp. Geogr. 94, 103099.
Yang, Y., Sasaki, K., Cheng, L., Liu, X., 2022. Gender differences in active travel among older adults: non-linear built environment insights. Transp. Res. Part D: Transp. Environ. 110, 103405.
Yang, H., Wen, J., Lu, Y., Peng, Q., 2023. A quasi-experimental study on the impact of park accessibility on the mental health of undergraduate students. Urban For. Urban Green. 86, 127979.
Yin, L., Cheng, Q., Wang, Z., Shao, Z., 2015. 'Big data'for pedestrian volume: exploring the use of Google Street View images for pedestrian counts. Appl. Geogr. 63, 337–345.
Yin, L., Wang, Z., 2016. Measuring visual enclosure for street walkability: using machine learning algorithms and Google Street View imagery. Appl. Geogr. 76, 147–153.
Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2881–2890.
Zhao, J., Sun, G., Webster, C., 2021. Walkability scoring: why and how does a three-dimensional pedestrian network matter? Environment and Planning b: Urban Analytics and City Science 48 (8), 2418–2435.
Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). Scene parsing through ade20k dataset. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 633–641.