Transportation Research Part D 124 (2023) 103936

Contents lists available at ScienceDirect

Transportation Research Part D

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

Assessing the effects of the built environment and microclimate on cycling volume

Ting Zhou [a,*], Tao Feng [a,b], Astrid Kemperman [a]

[a] *Department of Built Environment, Eindhoven University of Technology, the Netherlands*
[b] *Urban and Data Science Lab, Graduate School of Advanced Science and Engineering, Hiroshima University, Japan*

ARTICLE INFO

Keywords:
Multi-scale urban built environment
Cycling volume
Gradient boosting decision tree

ABSTRACT

Cycling benefits human health and helps to mitigate environmental issues. However, limited evidence exists regarding how the built environment influences cycling volume under different time and weather conditions. In this paper, we employed a Gradient Boosting Decision tree method to analyze the non-linear and threshold effects of the multiscale built environment and microclimate on cycling volume. Results based on the multisource data of The Netherlands show that 27 out of the 28 variables have a non-linear and threshold effect on cycling volume. Temperature is found to be a dominant factor among all variables. At street level, slope is the most important factor, followed by the green view and sky view indexes. At neighborhood level, population density is the most important factor, followed by residential density, and the density of bus stops. These findings offer useful insights for planning a cycling-friendly urban built environment at different scales.

1. Introduction

Cycling contributes to human health and reduces the risk of many physical and mental problems, because it not only improves metabolic, mental, and cardiovascular health (Götschi et al., 2016; Haennel & Lemire, 2002), prevents falls, and reduces depression (Rissel & Watkins, 2014), but also encourages people to engage in vigorous physical activities more often and for longer periods (Kerr et al., 2016). As a sustainable transportation mode, cycling reduces private vehicle usage, thus mitigating traffic congestion and improving air quality (Li et al., 2011). Considering the health and environmental benefits of cycling, researchers and urban planners aim to understand the factors influencing cycling behavior to promote cycling levels. Previous studies have associated the urban built environment with different domains of cycling behavior. The majority of these studies focused on aspects such as bicycle mode choice (e.g., Rybarczyk et al., 2014), bicycle ridership (Duran-Rodas et al., 2019), cycling duration (e.g., Gao et al., 2018), and cycling frequency (Heesch et al., 2014). However, it is important to note that the effects of the built environment might vary across different locations and over time. Neglecting spatial and temporal variations may lead to reduced reliability in model output and inaccurate interpretation of the relationships between the relevant variables (Qian & Ukkusuri, 2015). In this regard, understanding the temporal fluctuation and spatial variation in cycling behavior can be facilitated by large cycling volume data that encompass sufficient spatial details and temporal coverage (Jestico et al., 2016). Furthermore, ample and accurate cycling volume data is beneficial for devising bicycle infrastructure investment strategies, e.g., prioritizing locations with high cycling volumes, and optimizing the allocation of

* Corresponding author.
E-mail address: t.zhou@tue.nl (T. Zhou).

https://doi.org/10.1016/j.trd.2023.103936
Received 15 June 2023; Received in revised form 3 September 2023; Accepted 4 October 2023
Available online 17 October 2023

limited funding for bicycle programs (Chen et al., 2017).

Realizing the importance of cycling volume in fostering active transport, several studies have investigated the relationships between urban built environment characteristics and cycling volume (Chen et al., 2017; Hochmair et al., 2019; Sun et al., 2017) and achieved interesting results. However, some research gaps still exist. Firstly, the majority of research focused on a single scale level of the urban built environment (e.g., street, or neighborhood level), and included a limited set of variables, making it difficult to provide a panoramic insight into the effects of the urban built environment. Secondly, most studies used traditional survey methods that are generally time-consuming, costly, and challenging to apply at different and larger environmental scales. Thirdly, previous studies have predominantly centered on aspects such as bike usage (Koohsari et al., 2020; Ji et al., 2023), cycling distance (Ji et al., 2022), and cycling frequency (Schmiedeskamp et al., 2016), while limited attention has been given to the exploration of cycling volume. Furthermore, studies typically ignored the effects of spatial and temporal dependency (e.g., location, time) and weather conditions (e.g., temperature, humidity). Neglecting these factors may result in incomplete and less convincing analyses, particularly given that cyclists might exhibit higher sensitivity than car drivers to both the built environment and the outdoor microclimate. As a consequence, the complicated effects of multi-scale urban built environment characteristics and microclimate conditions on cycling volume remain an under-explored issue.

To fill this research gap, this study aims to analyze the impacts of multi-scale urban built environment characteristics and microclimate on cycling volume and to identify threshold effects. This research contributes to the literature in two ways. Empirically, it fulfills the research gap in the built environment and cycling volume studies by identifying the non-linear and threshold effects of built environment characteristics at different scale levels on cycling volume. Besides, it enhances the accuracy of current evidence by assessing the relationships using a nonparametric GBDT model while accounting for variations across locations, time, and weather conditions. The results will provide information for urban planners and designers on how to enhance the urban built environment to promote cycling behavior. Moreover, it gives a more panoramic insight into the urban built environment assessment by involving urban built environment characteristics at multiple scales including street attributes and neighborhood characteristics.

The remainder of the paper is structured into five sections. Section 2 examines the literature concerning the factors associated with cycling volume and the modeling approaches used in this domain. This is followed by Section 3, which details the collected data and the methods employed to measure the multi-scale urban built environment characteristics and predict the relationships between these characteristics and cycling volume. Section 4 outlines the findings, and Section 5 provides conclusions and discusses implications for planning practices. Finally, the limitations and potential directions for future research are discussed.

Table 1
Factors associated with cycling volume included in the existing literature.

Variables		Relationship with cycling volume	Source
Neighborhood-level characteristics			
Infrastructure	Presence of metro stations	Positive	Faghih-Imani et al. (2014)
	Distance to the nearest bus stop	Not significant	Sun et al. (2017)
	Distance to city center	Negative	Dill et al. (2007)
		Not significant	Griswold et al. (2011); Sun et al. (2017)
	Distance to university campus	Negative	Griswold et al. (2011); Faghih-Imani et al. (2014)
Environment	Population density	Positive	McCahill et al. (2008)
	Residential density	Not significant	Sun et al. (2017); Chen et al. (2017)
	Household density	Not significant	Chen et al. (2017)
	Employment density	Positive	McCahill et al. (2008)
		Not significant	Sun et al. (2017); Chen et al. (2017)
	Density of bike lanes	Positive	Dill et al. (2003)
	Motor vehicle volume	Negative	Sun et al. (2017)
	Street connectivity	Positive	Sun et al. (2017)
	Land use mix	Positive	Griswold et al. (2011); Hankey et al. (2012); Chen et al. (2017)
Street-level characteristics			
Infrastructure	Road class	Not significant	Sun et al. (2017)
	Road length	Negative	Sun et al. (2017)
	Bike lane length	Positive	Chen et al. (2017)
	Presence of bike lanes	Positive	Reynolds et al. (2009); Buck. (2012); Hong et al. (2020)
	Presence of bike racks	Positive	Reynolds et al. (2009); Buck. (2012); Hong et al. (2020)
Environment	Presence of water bodies	Positive	Chen et al. (2017)
	Slope	Negative	Chen et al. (2017)
Non-urban built environment variables			
Temporal variables	Peak hours	Positive	Tin Tin et al. (2012)
	Weekends	Positive	Tin Tin et al. (2012)
	Non-winter seasons	Positive	Tin Tin et al. (2012); Chen et al. (2017)
Weather variables	Temperature	Positive	Miranda-Moreno et al. (2012); Schmiedeskamp et al. (2016)
	Precipitation	Negative	Miranda-Moreno et al. (2012); Schmiedeskamp et al. (2016)
	Humidity	Positive	Miranda-Moreno et al. (2012)

2. Literature review

2.1. Factors associated with cycling volume

The importance of the urban built environment influencing cycling volume has been well documented in previous studies. Neighborhood-level characteristics are the most frequently discussed in these studies. For instance, a higher level of population density, employment density (McCahill & Garrick, 2008), and land use mix (Chen et al., 2017; Griswold et al., 2011; Hankey et al., 2012), better street connectivity (Sun et al., 2017), a lower volume of motor vehicles (Sun et al., 2017) and more bike lanes (Dill & Carr, 2003) are shown positively associated with cycling volume. The presence of metro stations (Faghih-Imani et al., 2014), shorter distance to a university campus (Faghih-Imani et al., 2014; Griswold et al., 2011), and city center (Dill & Voros, 2007) also influence bicycle flows positively. However, in some other studies, residential density (Chen et al., 2017; Sun et al., 2017), employment density (Chen et al., 2017; Sun et al., 2017), household density (Chen et al., 2017), the distance to the city center (Griswold et al., 2011; Sun et al., 2017) and the distance to the nearest bus stop (Sun et al., 2017) are not significantly related to cycling volume.

In addition to these analyses, studies also examined the effects of street-level characteristics. Specifically, providing cycling infrastructure such as bike lanes and bike racks creates safety benefits and encourages cycling (Buck & Buehler, 2012; Hong et al., 2020; Reynolds et al., 2009). Besides, the presence of water bodies and bike lane length are positive predictors of cycling volume, whereas the presence of steep areas is a negative one (Chen et al., 2017). Sun et al. (2017) found that cycling volume is negatively associated with road length, but not road class.

Furthermore, some studies have stressed the effects of temporal characteristics and weather conditions on cycling volume. For instance, Tin Tin et al. (2012) pointed out that bicycle counts are much higher in peak hours, during non-winter seasons, and on weekends, and this view was supported by Chen et al. (2017). Moreover, high temperature and low precipitation prove to have a positive effect on cycling volume (Miranda-Moreno & Nosal, 2011; Schmiedeskamp & Zhao, 2016), however, the temperature has a negative effect when it is higher than 28 °C and humidity is greater than 60% (Miranda-Moreno & Nosal, 2011).

In sum, the existing studies have revealed several determinants of cycling volume, including neighborhood-level characteristics, street-level characteristics, and temporal and weather conditions. Table 1 summarizes the main factors that have been addressed in the existing literature. Nevertheless, some variables are omitted, providing a limited panoramic insight into the effects of the urban built environment. For example, some neighborhood-level characteristics (e.g., the density of outdoor leisure facilities) and street-level characteristics (e.g., the presence of streetlights, traffic lights, and greenness) have hardly been considered, while it can be assumed that they promote cycling. Thus, there is a need to incorporate more comprehensive urban built environment characteristics in explaining cycling volume.

2.2. Modeling approaches

A variety of regression-based models have been applied to identify the relationship between the urban built environment and cycling volume, such as the Poisson regression model (Jestico et al., 2016), negative binomial model (Noland et al., 2016), fixed-effects linear regression model (Hong et al., 2020). These models generally share two main drawbacks: the neglect of spatial dependency and the linear relationship assumption.

To incorporate the spatial effects, Lyu et al. (2020) proposed a multi-scale geographically weighted regression model to analyze the relationship between built environment characteristics and bicycle ridership. They found different influencing patterns between traditional downtown districts and newly built-up areas, particularly for variables like population density, road network, parking space, and points of interest (POI). Still, these findings are based on the assumption of linear relationships, ignoring the possibility of non-linear associations, and threshold effects. Threshold effects mean that certain variables have an influence or lack influence based on whether their values reach or surpass a particular threshold level. Thus, these linear-relationship-based methods might be unable to depict the complex relationships.

Nowadays, a growing number of studies have explored the nonlinearity and threshold effects that exist within the connection between the built environment and cycling behavior. Related studies use regression methods based on predefined nonlinear shapes (Boulange et al., 2017; Koohsari et al., 2020), instead of nonparametric assumptions. This means regression methods tend to fit the data with a predefined line. If this line underestimates the relationship between the built environment and cycling behavior, the results might be inaccurate.

Some machine learning models such as the gradient boosting decision trees do not assume that a built environment variable follows a predefined relationship with cycling behavior, thus might be able to demonstrate the true and accurate relationship. For instance, Duran-Rodas et al. (2019) are one of the few researchers who analyzed the non-linear relationships between built environment factors and bicycle ridership (in relation to arrivals and departures to and from a station) using a gradient boosting machine model. They validated the good prediction accuracy of this machine learning method to help understand the factors affecting bicycle ridership, and they found that the most influencing variables are city population, distance to the city center, and leisure-related establishments. However, it does not identify the threshold values of statistically significant variables, lacking a detailed description and exploration of the non-linear relationships.

Despite some studies exploring the non-linear association between the built environment and cycling behavior, the specific threshold values of the influential variables could be further explored. Besides, there is a scarcity of comprehensive and detailed investigation into the non-linear relationships between urban built environment characteristics and cycling volume.

3. Methodology

3.1. Data

Given that the purpose of this paper is to examine the relationships between multi-scale urban built environment characteristics and cycling volume, it is imperative to process data pertaining to both cycling volume and the built environment. Moreover, considering the documented impact of weather on cycling behavior (de Kruijf et al., 2021), the inclusion of weather data is also necessary. In this study, we use the cycling volume data extracted from NDW Dexter in The Netherlands (https://dexter.ndwcloud.nu/home). NDW counts the number of bikes passed through hundreds of locations across the whole country of The Netherlands. The cycling volume used in this study is aggregated on an hourly basis and covers 279 locations from May 13, 2022 to May 26, 2022 (excluding May 23, 2022, for which no data was collected). To be consistent, the weather data (temperature and humidity) was also downloaded during the same period from the website of The Royal Netherlands Meteorological Institute (https://www.knmi.nl/over-het-knmi/about).

To obtain the built environment data at both the street and neighborhood levels, we integrated multisource data from different open repositories, including OpenStreetMap, Google Street View images from Google Map, DEM (Digital Elevation Model) data from NASA Earthdata and POI (Point of Interest) data from OpenStreetMap. See Table 2 for an overview of the various data sources used in this study (see Table 3).

3.2. Method

3.2.1. Measures of street-level and neighborhood-level characteristics

We used different methods to quantify the set of variables. As for street-level characteristics, firstly we identified the presence of sidewalks, bike lanes, streetlights, crosswalks, traffic lights, traffic signs, bike parking, potholes, water scenery, greenery index, and sky view index through semantic segmentation technique using Google Street View images. With Google's Street View Image Application Programming Interface (API), users can access Google Maps's extensive collection of street view images (Ki & Lee, 2021; Rundle et al., 2011). Several key parameters for the API need to be specified, including image size (640 × 640 pixels is the maximum image resolution for non-premium plan users), geographic location (geographic coordinates or addresses), the field of view (zoom level), the camera's angle relative to the Street View vehicle (default is 0), and heading (the camera's direction with 0°=north, 90°=east, 180°=south and 270°=west) (Nguyen et al., 2019). Specifically, according to the coordinate of sampling locations from cycling volume data, we downloaded Google Street View images in four headings (0°, 90°, 180°, and 270°) at each sampling location to capture a panoramic view of the environment, and image resolution is 640 × 640 pixels. These settings of the parameters were also used by Wang et al. (2019), Nguyen et al. (2019), and Keralis et al. (2020). The In-Place Activated BatchNorm model developed by Bulo et al. (2018) was adopted for segmentation due to its high accuracy (the mean intersection over union (mIoU)) of 53.42%. It was trained on Mapillary Vistas with WideResNet38 and DeepLab3. MIoU is the standard metric for segmentation, which calculates a ratio between the intersection and the union of two sets (Garcia-Garcia et al., 2017).

Moreover, compared to other datasets, Mapillary Vistas provides more cycling infrastructure-related categories such as bike lanes and bike parking. Specifically, variables including greenery and sky view were quantified by computing the ratios of the pixels classified as them over the total number of pixels in the image. In addition, sidewalks, bike lanes, streetlights, crosswalks, traffic lights, traffic signs, potholes, as well as water scenery were measured as binary variables (i.e. score 1 if an element is present in the image and 0 otherwise). Following Ito & Biljecki (2021), the type of road was collected by creating a 100 m buffer using ArcGIS 10.2 from 279 sampling locations where we downloaded cycling volume data, and converted categorical values to numerical ones by assigning different values to different types of roads, and selecting the type of roads that are with the longest road length, which is similar to prior works (Guo & He, 2020; Ito & Biljecki, 2021). Moreover, to determine the slope, we calculated this parameter using DE data and Spatial Analyst in ArcGIS 10.2.

As for neighborhood-level characteristics, firstly, we computed the distance to the nearest public park, metro station, and school using Network Analyst in ArcGIS 10.2. Chen et al. (2017) have demonstrated that the use of a 1-mile buffer to measure the built environment can lead to better results. Therefore, in our study, the density of outdoor leisure facilities (parks, dog parks, playgrounds, golf courses, picnic sites, etcetera), bus stops, residential density, and land use mix were aggregated and calculated based on a 1600-meter buffer. The land use mix was calculated using the following formula (equation (1)) which is adapted from Frank et al. (2005).

Table 2
Data source of measurement.

Type of data	Data source
Cycling volume data	https://dexter.ndwcloud.nu/home
Temperature data	https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens
Humidity data	https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens
OpenStreetMap data	https://www.openstreetmap.org/
Google Street View images	https://www.google.com/maps/
DEM	https://asterweb.jpl.nasa.gov/gdem.asp
POI data	https://www.arcgis.com/home/search.html?q = POI

Table 3
Description of variables.

Variables		Description	Mean / Percentage	SD
Urban built environment variables				
Street-level characteristics				
Infrastructure	Type of road	Type of roads with the longest road length within a 100 m buffer (service, track = 0.1; primary, primary link = 0.2; secondary, secondary link = 0.4; tertiary, tertiary link = 0.5; unclassified = 0.6; sidewalk = 0.8; bike lanes = 1; others = 0)	0.57	0.43
	Presence of sidewalks	Presence of sidewalk (yes:1; no:0)	Yes: 0.94; no: 0.06	N/A
	Presence of bike lanes	Presence of bike lanes (yes:1; no:0)	Yes: 0.57; no: 0.43	N/A
	Presence of streetlights	Presence of streetlight along the road (yes:1; no:0)	Yes: 0.81; no: 0.19	N/A
	Presence of crosswalks	Presence of a crosswalk (yes:1; no:0)	Yes: 0.22; no: 0.78	N/A
	Presence of traffic lights at intersections	Presence of traffic lights (yes:1; no:0)	Yes: 0.21; no: 0.79	N/A
	Presence of traffic signs	Presence of traffic signs (yes:1; no:0)	Yes: 0.72; no: 0.28	N/A
	Presence of bike parking	Presence of bike parking (yes:1; no:0)	Yes: 0.03; no: 0.97	N/A
Environment	Slope	Uphill ascent (m)	28.59	35.31
	Presence of potholes	Presence of potholes (yes:1; no:0)	Yes: 0.01; no: 0.99	N/A
	Presence of water scenery	Presence of water scenery (yes:1; no:0)	Yes: 0.54; no: 0.46	N/A
	Green view index	Proportion of green vegetation in Google Street View images	0.35	0.19
	Sky view index	Proportion of sky visibility in Google Street View images	0.29	0.15
Neighborhood-level characteristics				
Infrastructure	The density of outdoor leisure facilities	Number of outdoor leisure facilities within a 1600 m buffer	23.14	25.09
	The density of bus stops	Number of bus stops within a 1600 m buffer	33.25	25.52
	Distance to the nearest public park	Distance to the nearest public park/urban green space (m)	1087.32	1232.26
	Distance to the nearest metro station	Distance to the nearest metro station (m)	2477.55	2331.76
	Distance to the nearest school	Distance to the nearest school (m)	1128.39	897.82
Environment	Road density	Ratio of total road length and the urban area within a 1000 m buffer (km/km^2)	21.51	10.57
	Sidewalk density	Ratio of total sidewalk length and the urban area within a 1000 m buffer (km/km^2)	3.88	3.97
	Cycleway density	Ratio of total bike lanes length and the urban area within a 1000 m buffer (km/km^2)	2.99	1.76
	Population density	Ratio of the number of residents to the land area per km^2	2189.03	3440.40
	Residential density	Ratio of the number of residential dwellings to the land area within a 1600 m buffer	0.36	0.25
	Land use mix	Land use diversity index measured by equation (1)	0.34	0.16
Non-urban built environment variables				
Time variables	Peak hours	From 7:00 to 8:00 am and from 16:00 to 18:00 pm (yes:1; no:0)	NA	NA
	Weekends	Saturdays and Sundays (yes:1; no:0)	NA	NA
Weather variables	Temperature	Hourly temperature (degrees Celsius)	16.36	4.47
	Humidity	Hourly relative humidity (%)	86.00	17.27
Cycling volume	Cycling volume	Cycling volume (hourly aggregated at each location)	71.00	113.00

This measure is especially promising since it was closely related to the types of physical activities (Frank et al., 2005; Brown et al., 2009), and has been widely used by other researchers (Brown et al., 2009; Ewing et al., 2015).

$$y = -\frac{[x_0 * lnx_0 + x_1 * lnx_1 + x_2 * lnx_2]}{ln3} \tag{1}$$

where the y represents *land use mix index*, x_0, x_1, x_2 represent *residential share*, *commercial share*, and *office share*, respectively. *Shares* are computed based on the floor area of each use within a 1600 m buffer and *ln* refers to the natural logarithm. Additionally, we used the statistical population density of the smallest administrative area where each sampling point is located as population density. Besides, road density, sidewalk density, and cycleway density were measured as the ratio of total road/sidewalk/cycleway length and the urban area, and the unit is km/km^2, the same as the previous studies (Cervero et al., 2009; Cheng et al., 2020). Followed Cervero et al. (2009), considering the simplicity and accuracy of calculations, road density, sidewalk density, and cycleway density were aggregated and counted using a 1000-meter buffer. We provide a descriptive summary of the variables in Table 3.

3.2.2. Relationship between the urban built environment and cycling volume

This research employed a machine learning model—gradient boosting decision trees. The model was first developed to predict data in computer science (Friedman, 2001), and has lately been utilized in urban research to explore the relationship between built environment characteristics and driving distance (Ding et al., 2018). The gradient boosting decision trees model offers various advantages over regression models. Firstly, it can illustrate the non-linear relationship, as shown in several existing studies (Ding et al., 2018; Yin et al., 2020). Secondly, through stagewise learning of the data, it can modify weights based on predictors, hence improving prediction accuracy. (Wu et al., 2020). Thirdly, not like traditional regression models, the gradient boosting decision trees model considers interactions among predictors, thus addressing the multi-collinearity issue Elith et al., 2008). Moreover, it can accommodate predictors with missing data (Ding et al., 2018).

Gradient boosting decision trees algorithm is an iterative approach that builds a series of decision trees for prediction. The model works on the principle that many weak learners (i.e. trees) jointly create a more accurate predictor. Initially, a single decision tree is trained using the training data (***x***,***y***) (***x*** is a vector of explanatory variables, including the multi-level built environment characteristics, f(x) is an approximation function of the response variable, the cycling volume y). The residuals (the difference between the predicted and actual values) are calculated using a loss function L(y, f(x)). Then, the second decision tree is trained to predict the residuals from the first tree. This process of creating a new tree to predict the residuals is repeated multiple times. Each new tree is trained on the residuals from the previous trees. The predictions from all the trees are then added together to create the final model prediction.

The parameter learning of the gradient boosting decision trees model is described as follows:

(1) Setup the model to minimize the square difference in the loss function L(y, f(x)):

$$L(y,f(x)) = (y - f(x))^2 \tag{2}$$

$$f_0(x) = argmin_\gamma \sum_{i=1}^{N} L(y_i, \gamma) \tag{3}$$

(2) Perform a total of M iterations for $m = 1, 2, \ldots, M$, and for each iteration, four tasks need to be completed.

(a) For $i = 1, 2, \ldots, N$, compute the negative target, which is the estimated value of the residual by using derivation α:

$$\gamma_{im} = -\left[\frac{\alpha L(y_i, f(x_i))}{\alpha f(x_i)}\right]_{f=f_{m-1}} \tag{4}$$

(b) Fit the regression tree to the target γ_{im} (the estimated value of the residual that was calculated in equation (4)) giving terminal regions R_{jm}.

j (the size of each of the constituent trees) = 1, 2, …, J.

(c) Compute a gradient descent step size:

$$\gamma_{jm=argmin_\gamma} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + \gamma) \tag{5}$$

(d) Update the model:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} I(x \in R_{jm}), I = 1 if x \in R_{jm} \tag{6}$$

(3) Output of the final model when the loss function is minimized after M iterations:

$$f(x) = f_M(x) \tag{7}$$

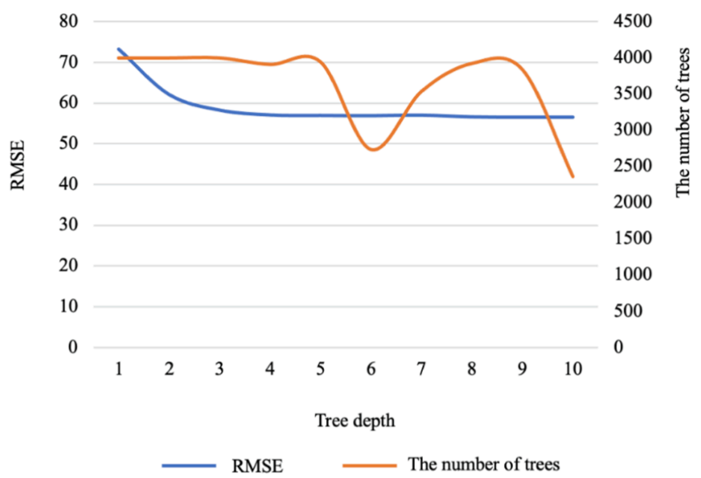

Fig. 1. Tree depth versus RMSE and the number of trees.

The learning rate, number of trees, and tree depth are three important parameters of the gradient boosting decision trees model. The learning rate is also called shrinkage value. Usually, the smaller the shrinkage, the less the loss, and the better the result of the model. The number of trees represents how many iterations will be conducted. More iterations could potentially enhance model performance, but can also overfit a model. Typically, smaller shrinkage requires more iterations, thus having a longer computation time. The tree depth represents the complexity of trees, also known as the number of splits that are used to fit each decision tree. In the next section, the process to find the best possible parameters will be described.

4. Results

In this study, we used the "gbm" package in R to estimate the model. Generally, the setting ranges for the number of trees and the learning rate are 3000 to 10000, and 0.1 to 0.001, respectively (Ridgeway, 2017). Firstly, we set the learning rate at 0.01 and chose the tree depth of 5. But even if we increased the number of trees to 10000, the model cannot converge. Subsequently, we set the learning rate at 0.1 and the tree depth at 5. For the number of trees, we changed it from 1000 to 5000 with an interval of 1000 and estimated 5 models with a number of trees at 1000, 2000, 3000, 4000, and 5000, respectively. We used root mean square error (RMSE) to assess model performance. The smaller the RMSE is, the better the performance of the model. When we set the number of trees at 4000, the best model performed was achieved with the smallest RMSE at 57.109. Finally, we set the learning rate at 0.1 and the number of trees at 4000. To find the optimal value of the tree depth, we changed it from 1 to 10 and estimated 10 models. Since 5-fold cross-validation consistently performs best for most datasets (Ridgeway, 2017), we applied 5-fold cross-validation to identify the optimal number of trees for each of the models. As shown in Fig. 1, RMSE drops quickly at first and becomes relatively flat when tree depth is 4. Since models with larger tree depth are computationally intensive in terms of model efficacy, we chose 4 as the tree depth. Consequently, we kept the learning rate at 0.1, set the number of trees at 4000, and chose a tree depth of 4. After 3910 iterations, the model produced the best result (RMSE is 57.081, the squared error loss is 3258.241).

4.1. The relative importance of explanatory variables

The relative importance of the explanatory variables in predicting cycling volume is shown in Table 4. The urban built environment characteristics play a more important role than time and weather variables in cycling volume, reaching a 63.18% contribution. Specifically, neighborhood-level characteristics exert a much more evident effect than street-level characteristics, with contribution rates of 52.92% and 10.26%, respectively. Among all the urban built environment characteristics, population density is the most important factor influencing cycling volume, with a contribution of 8.13%. This is followed by residential density (6.52%), the density of bus stops (6.10%), and cycleway density (5.81%). Road density, distance to the nearest park, distance to outdoor leisure activities,

Table 4
Relative importance of explanatory variables.

Variables		Relative importance (%)	Rank
Urban built environment total 63.18			
Street-level characteristics	Slope	3.59	12
Total 10.26	Green view index	2.10	15
	Sky view index	1.92	17
	The presence of bike lanes	1.11	18
	The presence of a crosswalk	0.30	20
	The presence of streetlight	0.28	21
	The type of road	0.27	22
	The presence of traffic signs	0.26	23
	The presence of a bike rack	0.21	24
	The presence of traffic lights	0.13	25
	The presence of water	0.08	26
	The presence of a sidewalk	0.01	27
	The presence of potholes	0.00	28
Neighborhood-level characteristics	Population density	8.13	4
Total 52.92	Residential density	6.52	5
	The density of bus stops	6.10	6
	Cycleway density	5.81	7
	Road density	5.45	8
	Distance to the nearest park	5.30	9
	The density of outdoor leisure activities	4.78	10
	Sidewalk density	3.64	11
	Land use mix	2.90	13
	Distance to the nearest metro station	2.32	14
	Distance to the nearest school	1.97	16
Non-urban built environment total 36.82			
Time variables	Peak hours	8.55	3
Total 9.47	Weekends	0.92	19
Weather variables	Temperature	18.48	1
Total 27.35	Humidity	8.87	2

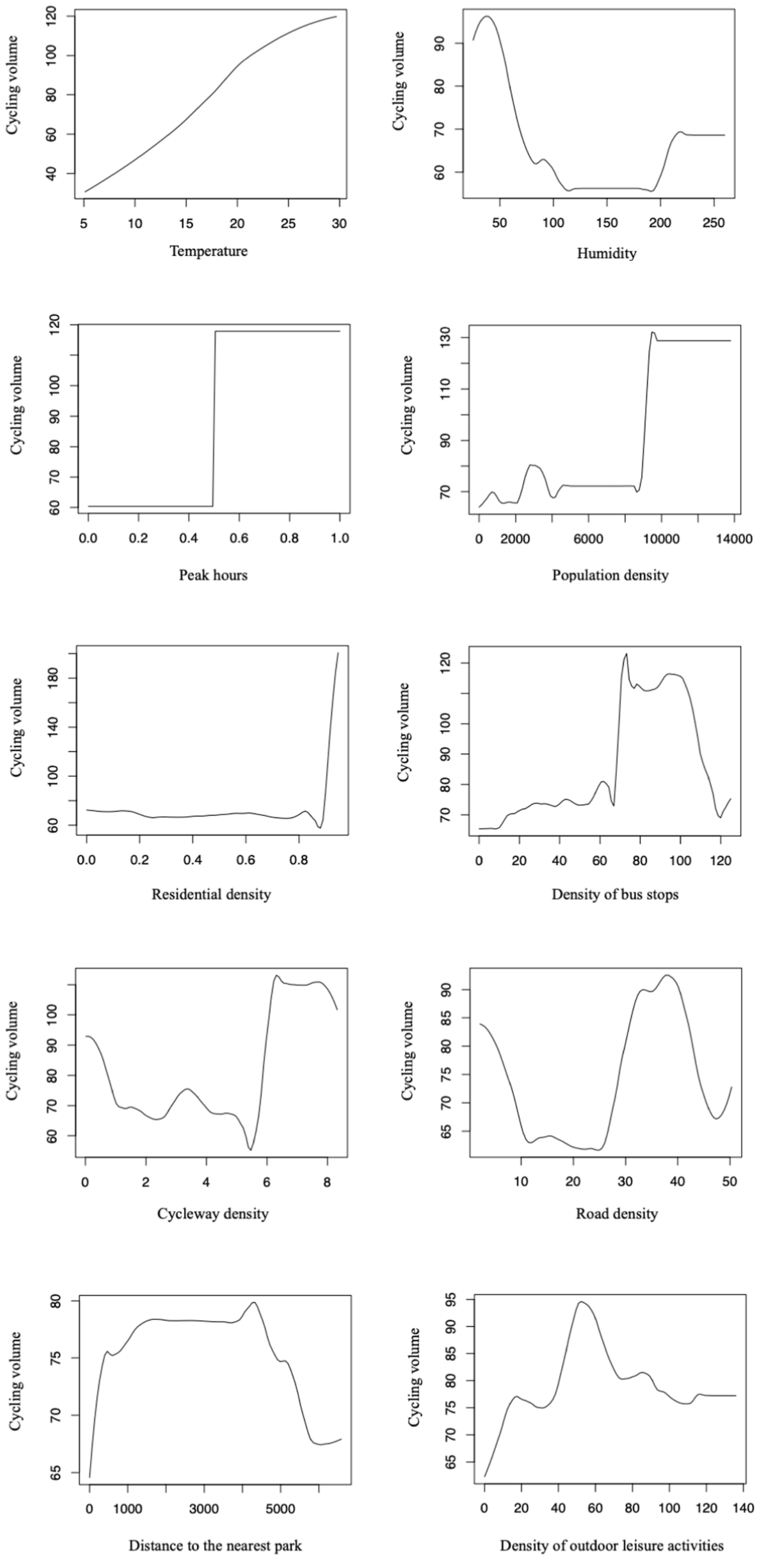

Fig. 2. Non-linear effects of key explanatory variables on cycling volume.

and sidewalk density also have some predictive power, with effects of more than 4%. In contrast, the type of road, the presence of crosswalks/streetlights/ traffic signs/ traffic lights/ water/sidewalks/ potholes have little effect on cycling volume, and the contribution is all < 1%.

Temperature is found to be a dominant factor in predicting cycling volume among all variables, accounting for 18.48% of the total contribution, which is followed by humidity, with a contribution rate of 8.87%. Time variables also expert an evident effect on cycling volume, especially peak hours showing a strong effect explaining cycling volume with an 8.55% contribution.

4.2. Non-linear effects of the key explanatory variables

The partial dependence plots provide a relatively accurate analysis of the complicated relationship between explanatory variables and cycling volume as shown in Figure A in the appendix. Fig. 2 shows the effects of the key explanatory variables (>4%).

Specifically, temperature is generally positively correlated with cycling volume. When it ranges between 5 and 20 °C, cycling volume increases rapidly with the increase in temperature. When it rises from 20 to 30 °C, the growth in cycling volume becomes slower; however, the drop in cycling volume with the increase in temperature has not been observed. The result was likely due to the Netherlands' moderate spring climate and the lack of extremely hot days in our dataset. Specifically, the mean temperature in May is just around 13 °C from 1991 to 2020 according to the Climate Change Knowledge Portal (https://climateknowledgeportal.worldbank.org/country/netherlands/climate-data-historical). Besides, in our dataset, we only have 156 out of 87,000 data observations where the temperature is higher than 29 °C. None of them exceeds 30 °C.

Humidity, population density, and cycleway density seem to have similar non-linear relationships with cycling volume. That is, from 50% to 210%, from 3,000 to 9,000 persons per km2, and from 0 to 6.2 km/km^2, respectively, humidity, population density and cycleway density have a U-shaped relationship with cycling volume, after which the impact becomes negligible. However, for humidity, cycling volume reaches the highest level on the left side of the U-shape curve, whereas, for population density and cycleway density, cycling volume reaches the peak on the right side. Specifically, the humidity of 50% is considered relatively dry weather in the Netherlands, considering that the mean humidity from 13th May 2022 to 26th May 2022 (except 23rd May 2022 when no data was collected) is 86%. This means people in the Netherlands are more likely to cycle in relatively dryer weather. Additionally, population density is a major contributor to cycling volume, which is similar to previous studies (Lyu et al., 2020). The first assumption is that densely populated areas are usually armed with more facilities and services such as supermarkets, pharmacies, restaurants, and offices are convenient for people to cycle for different purposes like commuting, recreation, and shopping. Another assumption is that in densely populated areas, even if people have the same probability of choosing to cycle, the cycling volume would be higher because of the large population base. Moreover, denser cycleways might encourage people to cycle.

Besides, Fig. 2 shows people cycle more during peak hours. This is probably because cycling is the most vital commuting mode in the Netherlands (Statistical Netherlands, 2018). Residential density has no obvious influence on cycling volume before 0.9. But after that, cycling volume climbs manifestly and reaches the peak at approximately 200 with the increase in residential density. Similar results were published by Beenackers et al. (2012), who found that residential density was positively related to cycling. This is because the large size of residential communities is usually armed with neighborhood supermarkets, restaurants, shops, and other facilities. Shorter distances between home and potential destinations might be the reason for people choosing to cycle.

The density of bus stops and the road density have similar patterns with cycling volume. For the density of bus stops, cycling volume increases slightly when the density is within the interval of 0 and 68 bus stops within a 1600 m buffer. But when it rises from 68 to 75, cycling volume experiences a sharp increase, which is followed by a fluctuation at a relatively high level ranging from 110 to 125. After that, there is a dramatic drop with the increase in the density of bus stops. This result validates the first assumption, as described in Section 4.2, that areas with denser bus stops usually have more facilities and would be more convenient for people to cycle. However, ultra-dense bus stops might have the opposite effect. Since most bus stops are probably located in the city center, where it is too busy to cycle. Instead, taking a bus or walking might be a better option for people to travel for different purposes like commuting, recreation, socializing, and shopping. With regards to the road density, which is measured as the ratio of total road length and the urban area, when it is less than 32 km/km^2, the road density follows a U-shaped relationship with cycling volume. Then with the increase in road density, cycling volume starts to fluctuate until it reaches a peak when the road density is around 38 km/km^2. After that cycling volume drops dramatically and starts to fluctuate again. The results show that the areas equipped with higher road density are more attractive for people to cycle. This is consistent with the findings by Cervero et al. (2009) who found that, in a setting with relatively high road densities, Bogota residents are nearly twice (0 min or more per weekday) as likely to cycle for utilitarian purposes. Furthermore, our findings indicate that the presence of ultra-dense road networks could hinder large cycling volume. This phenomenon might be attributed to the fact that ultra-dense road networks are usually situated in city centers where cycling might not be the primary mode of travel for individuals.

Additionally, the density of outdoor leisure activities (parks, dog parks, playgrounds, golf courses, picnic sites, etcetera) is positively associated with cycling volume when it ranges from 0 to 58. The impact becomes relatively negative after the cycling volume increases to the highest level at about 94. This is reasonable, since dense large parks, greenspace, golf courses, and other outdoor leisure activities are might in suburban areas, far away from where many people live. Instead of cycling, many people might prefer to drive to these outdoor leisure places with their families, friends, and pets. As for the distance to the nearest park, the curve of the relationship with cycling volume starts with a reverse U shape, and the turning point is around 4300 m. Areas that are too far away from the nearest park might be in remote places, which is inconvenient for people to cycle for their activities.

5. Conclusion and discussion

Using multi-source data such as OpenStreetMap data, Google Street View images, and POI data of 279 locations in the Netherlands and the gradient boosting decision trees model, we identified the non-linear and threshold effects of urban built environment characteristics on cycling volume. Moreover, we clarified the associations between time variables, weather conditions, and cycling volume. To the best of our knowledge, this is the first study that uses the gradient boosting decision trees model to systematically analyze the association between the built environment and cycling volume, offering additional insights into the non-linear effects and giving policy implications to promote active transport.

First, the urban built environment makes a substantial contribution to predicting cycling volume, reaching 63.18%. Each built environment variable's relative importance and threshold effects further imply the hierarchy of intervention priorities. In particular, population density is the dominant factor, effectively promoting cycling volume within a certain range of 8500 to 9000 persons per km^2. This greatly challenges the assumptions adopted in the previous study by Sun et al. (2017). Probably this is because of the selection of study areas. People in the Netherlands are more likely to cycle in relatively densely populated areas, whereas people in Glasgow cycle regardless of population density. Secondly, residential density, the density of bus stops, and cycleway density are also salient predictors for cycling volume. The effective ranges are all relatively high values, from 0.9 to 0.92, 70 to 75 bus stops within a 1600 buffer, and 5.8 to 6 km/km^2, respectively. Low and excessive values may not affect cycling volume. Thirdly, we found that road density, distance to the nearest park, and the density of outdoor leisure activities also have relatively slight threshold effects on cycling volume. The threshold values are 28 and 42 km/km^2, 0 and 200 m, and 0 and 65 outdoor leisure activities within a 1600 m buffer, respectively. Noticeably, the turning points for the aforementioned three variables are also relatively high values. Thus, policies to increase the road density, distance to the nearest park, and the density of outdoor leisure activities have to be implemented within different effective ranges. Among the variables included in this study, temperature is the most dominant predictor for cycling volume, reaching 18.48%. However, when the temperature is above 29 °C, it may not affect cycling volume. Humidity is the second dominant predictor for cycling volume, but it should be noted that the effective values range from 20% to 70%. Consequently, people in the Netherlands prefer to cycle in more moderate and dryer weather.

The added values of this study include at least the following perspectives. Firstly, the proposed model increases the accuracy of the current evidence by exploring the complicated relationship between urban built environment characteristics and cycling volume. Secondly, the current study gives a more detailed and panoramic picture of assessing urban built environment by considering multi-level urban built environment characteristics and using multi-source of data. Thirdly, we demonstrate the possibility of applying the gradient boosting decision trees model and offering a new framework to model the association between the built environment and cycling behavior. Fourthly, the results inform policy initiatives in terms of increasing cycling by planning and designing a more cycling-friendly urban built environment. For example, urban planners should consider the importance of population density, residential density, the density of bus stops, cycleway density, road density and distance to the nearest park, and the density of outdoor leisure activities. To offer policy implications to promote cycling volume, recommended values for these variables in the Netherlands are approximately 9000 persons per km^2, 0.92, 75 bus stops within a 1600 buffer, 6 km/km^2, 42 km/km^2, 200 m, 65 outdoor leisure activities within a 1600 m buffer, respectively.

Some limitations of this study are noteworthy. First, whether the complicated relationships between the built environment characteristics and cycling volume are comparable to other countries is unclear, since this study just focuses on one single country, the Netherlands. It would be of interest to apply a similar approach to other countries and compare the findings. Secondly, the impact of the built environment characteristics on cycling volume might vary by different ages/gender/income levels of people. Due to data limitations, socio-demographic characteristics are not included in this research. Thus, future studies are needed when detailed data becomes available. Thirdly, due to the efficacy of the model and the lack of data on cycling volume, only two weeks data are included in this research, which may create biases in weather conditions since only one season is considered. Fourthly, due to the lack of complete hourly precipitation data for these two weeks, precipitation is not included in this research. Nonetheless, future studies to assess the non-linear and threshold effects of the built environment on cycling volume are needed to support planning and designing a built environment that supports active transportation.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trd.2023.103936.

References

Beenackers, M.A., Foster, S., Kamphuis, C.B.M., Titze, S., Divitini, M., Knuiman, M., Van Lenthe, F.J., Giles-Corti, B., 2012. Taking up cycling after residential relocation: Built environment factors. Am. J. Prev. Med. 42 (6), 610–615. https://doi.org/10.1016/j.amepre.2012.02.021.

Boulange, C., Gunn, L., Giles-Corti, B., Mavoa, S., Pettit, C., Badland, H., 2017. Examining associations between urban design attributes and transport mode choice for walking, cycling, public transport and private motor vehicle trips. J. Transp. Health 6, 155–166. https://doi.org/10.1016/j.jth.2017.07.007.

Brown, B.B., Yamada, I., Smith, K.R., Zick, C.D., Kowaleski-Jones, L., Fan, J.X., 2009. Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity. Health Place 15 (4), 1130–1141. https://doi.org/10.1016/j.healthplace.2009.06.008.

Buck, D., Buehler, R., 2012. Bike lanes and other determinants of capital bikeshare trips. In: 91st Transportation Research Board Annual Meeting, pp. 703–706 http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Bike+Lanes+and+Other+Determinants+of+Capital+Bikeshare+Trips#0.

Bulo, S.R., Porzi, L., Kontschieder, P., 2018. In-place activated batchnorm for memory-optimized training of dnns. In: In ProceedIngs of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5639–5647.

Cervero, R., Sarmiento, O.L., Jacoby, E., Gomez, L.F., Neiman, A., 2009. Influences of built environments on walking and cycling: Lessons from Bogotá. Int. J. Sustain. Transp. 3 (4), 203–226. https://doi.org/10.1080/15568310802178314.

Chen, P., Zhou, J., Sun, F., 2017. Built environment determinants of bicycle volume: A longitudinal analysis. J. Transp. Land Use 10 (1), 655–674. https://doi.org/10.5198/jtlu.2017.892.

Cheng, L., Yang, J., Chen, X., Cao, M., Zhou, H., Sun, Y., 2020. How could the station-based bike sharing system and the free-floating bike sharing system be coordinated? In. J. Transp. Geogr. 89 https://doi.org/10.1016/j.jtrangeo.2020.102896.

de Kruijf, J., van der Waerden, P., Feng, T., Böcker, L., van Lierop, D., Ettema, D., Dijst, M., 2021. Integrated weather effects on e-cycling in daily commuting: A longitudinal evaluation of weather effects on e-cycling in the Netherlands. Transp. Res. A Policy Pract. 148, 305–315. https://doi.org/10.1016/j.tra.2021.04.003.

Dill, J., Carr, T., 2003. Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. Transp. Res. Rec. 1828, 116–123. https://doi.org/10.3141/1828-14.

Dill, J., Voros, K., 2007. Factors affecting bicycling demand: Initial survey findings from the Portland, Oregon, region. Transp. Res. Rec. 2031, 9–17. https://doi.org/10.3141/2031-02.

Ding, C., Cao, X. (Jason), & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transp. Res. Part A: Policy and Practice*, 110(August 2017), 107–117. https://doi.org/10.1016/j.tra.2018.02.009.

Duran-Rodas, D., Chaniotakis, E., Antoniou, C., 2019. Built environment factors affecting bike sharing ridership: data-driven approach for multiple cities. TransportationResearch Record 2673 (12), 55–68. https://doi.org/10.1177/0361198119849908.

Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J. Anim. Ecol. 77 (4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Greene, W., 2015. Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Stud. 52 (13), 2330–2348. https://doi.org/10.1177/0042098014560991.

Faghih-Imani, A., Eluru, N., & El-Geneidy, A. M. (2014). For citation please use: Imani, A., Eluru, N., El-Geneidy, A., Rabbat, M. & Haq, U. (accepted). How does land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. *J. Transport Geography*, February, 1–20.

Frank, L.D., Schmid, T.L., Sallis, J.F., Chapman, J., Saelens, B.E., 2005. Linking objectively measured physical activity with objectively measured urban form: findings from SMARTRAQ. Am. J. Prev. Med. 28 (2), 117–125. https://doi.org/10.1016/j.amepre.2004.11.001.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Gao, J., Kamphuis, C.B.M., Dijst, M., Helbich, M., 2018. The role of the natural and built environment in cycling duration in the Netherlands. Int. J. Behav. Nutr. Phys. Act. 15 (1), 1–16. https://doi.org/10.1186/s12966-018-0715-z.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). *A Review on Deep Learning Techniques Applied to Semantic Segmentation*. 1–23. http://arxiv.org/abs/1704.06857.

Götschi, T., Garrard, J., Giles-Corti, B., 2016. Cycling as a part of daily life: a review of health perspectives. Transp. Rev. 36 (1), 45–71. https://doi.org/10.1080/01441647.2015.1057877.

Griswold, J. B., Medury, A., & Schneider, R. J. (2011). *UC berkeley research reports title pilot models for estimating bicycle intersection volumes permalink https://escholarship.org/uc/item/380855q6 Publication Date*. http://www.safetrec.berkeley.edu.

Guo, Y., He, S.Y., 2020. Built environment effects on the integration of dockless bike-sharing and the metro. Transp. Res. Part D: Transp. Environ. 83 (April) https://doi.org/10.1016/j.trd.2020.102335.

Haennel, R.G., Lemire, F., 2002. Physical activity to prevent. 48, 65–71.

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis. MN. *Landscape and Urban Planning* 107 (3), 307–316. https://doi.org/10.1016/j.landurbplan.2012.06.005.

Heesch, K.C., Giles-Corti, B., Turrell, G., 2014. Cycling for transport and recreation: Associations with socio-economic position, environmental perceptions, and psychological disposition. Prev. Med. 63, 29–35. https://doi.org/10.1016/j.ypmed.2014.03.003.

Hochmair, H. H., Bardin, E., Ahmouda, A., 2019. Estimating bicycle trip volume for Miami-Dade county from Strava tracking data. J. Transport Geography, 75 (November 2018), 58–69. https://doi.org/10.1016/j.jtrangeo.2019.01.013.

Hong, J., Philip McArthur, D., Stewart, J.L., 2020. Can providing safe cycling infrastructure encourage people to cycle more when it rains? The use of crowdsourced cycling data (Strava). Transp. Res. Part A: Policy and Practice, 133(August 2019), 109–121. https://doi.org/10.1016/j.tra.2020.01.008.

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transp. Res. Part C: Emerging Technol. 132 (May), 103371 https://doi.org/10.1016/j.trc.2021.103371.

Jestico, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. J. Transp. Geogr. 52, 90–97. https://doi.org/10.1016/j.jtrangeo.2016.03.006.

Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E., Sun, Z., 2022. Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: a non-linear and interaction effect analysis. J. Transp. Geogr. 103, 103414 https://doi.org/10.1016/j.jtrangeo.2022.103414.

Ji, S., Heinen, E., Wang, Y., 2023. Non-linear effects of street patterns and land use on the bike-share usage. Transp. Res. Part D: Transport Environ. 116, 103630 https://doi.org/10.1016/j.trd.2023.103630.

Keralis, J.M., Javanmardi, M., Khanna, S., Dwivedi, P., Huang, D., Tasdizen, T., Nguyen, Q.C., 2020. Health and the built environment in United States cities: Measuring associations using Google Street View-derived indicators of the built environment. BMC Public Health 20 (1), 1–10. https://doi.org/10.1186/s12889-020-8300-1.

Kerr, J., Emond, J.A., Badland, H., Reis, R., Sarmiento, O., Carlson, J., Sallis, J.F., Cerin, E., Cain, K., Conway, T., Schofield, G., Macfarlane, D.J., Christiansen, L.B., Van Dyck, D., Davey, R., Aguinaga-Ontoso, I., Salvo, D., Sugiyama, T., Mitáš, J., Natarajan, L., 2016. Perceived neighborhood environmental attributes associated with walking and cycling for transport among adult residents of 17 cities in 12 countries: The IPEN study. Environ. Health Perspect. 124 (3), 290–298. https://doi.org/10.1289/ehp.1409466.

Ki, D., Lee, S., 2021. Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning. Landsc. Urban Plan. 205 (June 2020), 103920 https://doi.org/10.1016/j.landurbplan.2020.103920.

Koohsari, M.J., Cole, R., Oka, K., Shibata, A., Yasunaga, A., Hanibuchi, T., Sugiyama, T., 2020. Associations of built environment attributes with bicycle use for transport. Environ. Planning B: Urban Anal. City Sci. 47 (9), 1745–1757. https://doi.org/10.1177/2399808319845006.

Li, S., Liu, Y., Zhang, J., 2011. Lose some, save some: Obesity, automobile demand, and gasoline consumption. J. Environ. Econ. Manag. 61 (1), 52–66. https://doi.org/10.1016/j.jeem.2010.10.001.

Lyu, C., Liu, Y., Liu, Z., Wu, X., Yang, X., 2020. Exploring multi-scale spatial relationship between built environment and public bicycle ridership: A case study in Nanjing. J. Transp. Land Use 13 (1), 447–467. https://doi.org/10.5198/jtlu.2020.1568.

McCahill, C., Garrick, N.W., 2008. The applicability of space syntax to bicycle facility planning. Transp. Res. Rec. 2074, 46–51. https://doi.org/10.3141/2074-06.

Miranda-Moreno, L., Nosal, T., 2011. Weather or not to cycle: Temporal trends and impact of weather on cycling in an urban environment. Transp. Res. Rec. 2247, 42–52. https://doi.org/10.3141/2247-06.

Nguyen, Q.C., Khanna, S., Dwivedi, P., Huang, D., Huang, Y., Tasdizen, T., Brunisholz, K.D., Li, F., Gorman, W., Nguyen, T.T., Jiang, C., 2019. Using Google Street View to examine associations between built environment characteristics and U.S. health outcomes. Prev. Med. Rep. 14 (April), 100859 https://doi.org/10.1016/j.pmedr.2019.100859.

Noland, R.B., Smart, M.J., Guo, Z., 2016. Bikeshare trip generation in New York City. Transp. Res. A Policy Pract. 94, 164–181. https://doi.org/10.1016/j.tra.2016.08.030.

Qian, X., Ukkusuri, S.V., 2015. Spatial variation of the urban taxi ridership using GPS data. Appl. Geogr. 59, 31–42. https://doi.org/10.1016/j.apgeog.2015.02.011.

Reynolds, C.C.O., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: A review of the literature. Environ. Health: A Global Access Science Source 8 (1). https://doi.org/10.1186/1476-069X-8-47.

Rissel, C., Watkins, G., 2014. Impact on cycling behavior and weight loss of a national cycling skills program (AustCycle) in Australia 2010–2013. J. Transp. Health 1 (2), 134–140. https://doi.org/10.1016/j.jth.2014.01.002.

Rundle, A.G., Bader, M.D.M., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using google street view to audit neighborhood environments. Am. J. Prev. Med. 40 (1), 94–100. https://doi.org/10.1016/j.amepre.2010.09.034.

Rybarczyk, G., Wu, C., 2014. Examining the impact of urban morphology on bicycle mode choice. Environ. Plann. B. Plann. Des. 41 (2), 272–288. https://doi.org/10.1068/b3.

Schmiedeskamp, P., Zhao, W., 2016. Estimating daily bicycle counts in Seattle, Washington, from seasonal and weather factors. Transp. Res. Rec. 2593 (2593), 94–102. https://doi.org/10.3141/2593-12.

Sun, Y., Du, Y., Wang, Y., Zhuang, L., 2017. Examining associations of environmental characteristics with recreational cycling behaviour by street-level strava data. Int. J. Environ. Res. Public Health 14 (6), 644. https://doi.org/10.3390/IJERPH14060644.

Tin Tin, S., Woodward, A., Robinson, E., Ameratunga, S., 2012. Temporal, seasonal and weather effects on cycle volume: An ecological study. Environ. Health: A Global Access Sci. Source 11 (1), 1–9. https://doi.org/10.1186/1476-069X-11-12.

Wang, R., Liu, Y., Lu, Y., Yuan, Y., Zhang, J., Liu, P., Yao, Y., 2019. The linkage between the perception of neighbourhood and physical activity in Guangzhou, China: Using street view imagery with deep learning techniques. Int. J. Health Geogr. 18 (1), 1–11. https://doi.org/10.1186/s12942-019-0182-z.

Wu, X., (Jason) Cao, X., & Ding, C., 2020. Exploring rider satisfaction with arterial BRT: An application of impact asymmetry analysis. Travel Behav. Soc. 19 (August 2019), 82–89. https://doi.org/10.1016/j.tbs.2019.12.006.

Yin, C., Cao, J., Sun, B., 2020. Examining non-linear associations between population density and waist-hip ratio: An application of gradient boosting decision trees. Cities 107 (August), 102899. https://doi.org/10.1016/j.cities.2020.102899.