Spatial heterogeneity in the nonlinear impact of built environment on commuting time of active users: A gradient boosting regression tree approach

DOI:
10.1155/2023/6217672

Document status and date:
Published: 01/01/2023

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 6217672, 15 pages
https://doi.org/10.1155/2023/6217672

Research Article

Spatial Heterogeneity in the Nonlinear Impact of Built Environment on Commuting Time of Active Users: A Gradient Boosting Regression Tree Approach

Jingxian Wu[1,2] **Guikong Tang**[2] **Huapeng Shen**[3] **and Soora Rasouli**[4]

[1]*Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Beijing, China*
[2]*Department of Transportation Engineering, University of Shanghai for Science and Technology, Jungong Road #516, Shanghai 200093, China*
[3]*Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, No. 100 Ping Le Yuan, Chaoyang District, Beijing 100124, China*
[4]*Urban Planning and Transportation Group, Eindhoven University of Technology, P.O. Box 513, Eindhoven 5600 MB, Netherlands*

Correspondence should be addressed to Huapeng Shen; shenhuapeng1997@163.com

Received 10 January 2022; Revised 24 January 2023; Accepted 2 February 2023; Published 13 February 2023

Academic Editor: Hongtai Yang

Many studies provided evidence regarding the influence of built environment (BE) on commuting time. However, few studies have considered the spatial heterogeneity of such impacts. Using data from Nanjing, China, this study employs two-step clustering and gradient boosted regression trees (GBRT) to segment the neighborhoods into different types and investigate the effects of BE characteristics on the commuting time of active users. The results show a strong effect of BE characteristics on commuting time, involving active modes. The importance of BE characteristics varies among neighborhood types. For active commuters in the internal region of Nanjing, commuting time is affected mostly by the land use mix at the work end. The lowest impact of BE in internal regions is associated with metro station density. For active commuters in external region of the city, the relative importance of intersection density at the home end is the largest (as high as 5.76%). Moreover, other significant differences are found in the associations between BE characteristics and active commuting time in the two regions.

1. Introduction

Active travel mode, referring to walking or cycling, is a viable alternative to driving in short-to-medium distance trips [1]. Regular active travel such as active commuting to work is thought to benefit both the environment and an individual's physical health [2]. To encourage active travel, several interventions such as bike-sharing programs and provisions of footpaths and cycle lanes have been implemented. However, such promotion does not bring a significant increase in the share of active commuting, and private cars are still the most widely used mode [3].

The low prevalence of active commuting can be attributed in part to urbanization's increasing average commute distance [4, 5]. Longer trip distances imply that active commuters will spend more time on the road and thus have less life satisfaction. Despite the fact that active mode commuting generates positive utilities such as a health-enhancing effect, commuters prefer to shorten it due to time budgets [6]. Active transportation, particularly cycling, can be used for longer distance trips if the time cost of commuting by active modes is appropriately reduced [4]. As a result, it is critical to investigate factors that influence the commuting time of active users.

Many studies on travel behavior have found that the built environment and sociodemographics are strongly associated with walking and cycling (e.g., see [7–14]). Individuals' active travel choices are influenced by built-environment characteristics such as walk-bike infrastructure [7], street pattern [8], route connectivity [9], street greenery [10], and population density [11]. Changes in the built environment, according to Handy et al. [12], influence travel mode choice primarily by altering travel time. Furthermore, Eldeeb et al. discovered that improving the built environment does not have a homogeneous impact on the likelihood of using active mode in different parts of the city [13]. However, rare studies have focused on the commuting time of active mode and examined whether the built environment has a spatially different impact on it.

Therefore, this study attempts to advance the literature by investigating the nonlinear associations between the built environment and travel time of active commuters, accounting for spatial heterogeneity. First, the two-step clustering method is used to differentiate the regions with various built-environment features. Next, gradient-boosted regression trees are generated for predicting the commuting time of active travelers in each region. Travelers' sociodemographics, trip characteristics, and built-environment characteristics are considered as conditional variables.

The remainder of this article is structured as follows. Section 2 presents a literature review on the association between built environment and active travel. Section 3 describes the data source used in this study and built-environment characteristics at the traffic analysis zone (TAZ) level, as well as commuters' personal profiles. Section 4 briefly describes the methodology used in this study, while Section 5 discusses the findings. Finally, we present the summary of the key findings and discuss their implications for planning practice.

2. Literature Review

2.1. Association between the Built Environment and Active Travel. The built environment is the physical setting designed to meet people's need to engage in activities. The built-environment characteristics that relate to residents' travel behavior are defined from "3Ds" into "5Ds," which are "density," "diversity," block "design," "destination," "accessibility," and "distance" to transit [15, 16]. Some scholars specifically focused on the impact of the "5Ds," as reflection of built-environment characteristics, on active travel (e.g., see [7–14, 17–19]).

Density, as a key component of built environment, has a paradoxical effect on active travel choice [20]. Density appeared to have a significant impact on choosing active travel modes in some studies, such as reference [21], but not in other studies (e.g., see [22]). Zhu et al. discovered that increasing population density increases residents' possibility of commuting in active mode [11]. Block design such as street crossing density, road density, and connected sidewalks have profound effects on active travel [8, 23, 24]. Accessibility indicators such as employment accessibility [18], destination accessibility [25], and transit accessibility [26] are found to be positively correlated with residents' use of active travel. Diversity that measures land use mix in a neighborhood/region is closely associated with the choice of active travel [13, 23, 27]. Furthermore, Raman and Roy found that mixing multiple land uses beyond a certain proportion can have an adverse effect [28]. For commuting, the mix of land use types that specify jobs and houses plays a major role. The appropriate job-to-housing ratio can be an indication of whether residents are likely to be employed in the neighboring area of their residence. This, in turn, influences commute distance and facilitates using active travel to go to work.

All the above-mentioned studies have confirmed the contribution of built environment to active travel primarily by using regression-based models such as multilevel regression [11], a logit-based model [8, 18, 24], and a structural equation model [23, 25]. However, these models are primarily based on a priori, often linear relation between the built environment and active mobility. A parallel stream of studies examines the association between built environment and travel behavior using machine learning methods. Ding et al. applied decision trees [29] to extract the nonlinear relationship between a built environment and commute mode choice. Tao et al. assessed the importance of BE components for the energy consumption of active users by applying gradient boosting decision trees [19]. They found the distance to the nearest park posed the greatest impact. Cheng et al. used a random forest to assess BE's impact on elderly active travel and found that population density had the highest contribution [30]. Liu et al. in their most recent study used the extreme gradient boosting approach to examine the association of built environment and active travel choice. They found that trip characteristics contributed more than the built environment [31].

2.2. Spatial Heterogeneity. Spatial heterogeneity refers to the varying impact of the same influential factors at different spatial scales or geographical locations. Several studies have investigated whether the relationship between the built environment and travel behavior varies across different types of neighborhoods [32, 33]. Srinivasan and Ferreira found the built environment around residential areas to pose different effects on travel mode choice compared to that around workplaces [34]. For instance, land use mix around households' residence has stronger effects on household travel mode choice and travel distances than that at job locations [35]. Using the geographically weighted regression (GWR) model, Tu et al. found significantly different impacts of built environment on travel mode choice [36]. Zhang et al. used a hierarchical linear model to explore the relationship between the neighborhood-built environment and trip distance [37]. They found significant spatial heterogeneity in the influence of the built environment on travel behavior. Neighborhoods located in different areas of the city sometimes share similar characteristics, but their effects can be different. Zhong et al. used a geographically weighted regression model to analyze the spatial heterogeneity of the effects of an urban built environment on road travel time and found that spatially varying relationships exist [38]. Ding et al. used spline in a mixed logit model and concluded that nonlinearity exists in the relation between built environment and commuting mode choice [39].

In summary, existing studies have empirically done a lot on the measurement of BE on active travel behavior or spatial heterogeneity of BE on travel behavior. However, none have investigated how the built environment influences the commuting time of active modes and comprehensively considered the potential impact of spatial heterogeneity. Thus, this study contributes to the existing body of the literature by exploring nonlinear relation (without any a priori assumptions) between built environment and active commuting time taking into account spatial heterogeneity. It is important to note that the duration of active mobility for commuting purpose has never been the subject of examination within the topic of nonlinear relationship between built environment and active mobility.

3. Data

3.1. Study Area and Data Sources. The data used in this study originated from Nanjing, China. Nanjing is a mega-city and the provincial political and economic center of Jiangsu. It is located in the eastern region of China, downstream of the Yangtze River. It is an important gateway city for the central and western regions' development, which is fueled by radiation from the Yangtze River Delta. Nanjing is divided into 11 administrative regions, covering a total area of 6,587 km^2 and a built-up area of 868 km^2. The resident population was 9.42 million in 2021, with an urban population of 8.19 million and an urbanization rate of 86.9%. In 2021, the city's gross regional product reached 163,532 billion yuan. This study concentrates on the most urbanized regions, including Gulou, Qinhuai, Xuanwu, Jianye, Yuhuatai, Qixia, Jiangning, Pukou, and Luhe (two regions, Gaochun and Lishui, were newly designated as regions during the urbanization and were not involved in this survey). Seven of them are on the southern side of the Yangtze River, while two are on the northern side. For transportation census and management, the Nanjing transportation planning agency divides the entire study area of Nanjing into 766 traffic analysis zones (TAZ) based on land use and administration boundaries. Figure 1 depicts the study area of Nanjing, China.

In this study, four data sources are used: Nanjing Household Travel Survey data from 2016; Nanjing urban GIS data; points of interest (POIs) from Baidu map; and an open street map (OSM). The Nanjing Household Travel Survey is an annual survey conducted by the Nanjing transportation planning agency. It is carried out through household interviews in order to learn about the daily mobility patterns of urban residents. In 2016, the survey employed the stratified random sampling technique to guarantee that the sample size was proportional to the population size. In total, 8,387 people from 3,015 households were invited to participate in the survey. The survey collected individual's sociodemographic information (e.g., household income, car ownership, gender, and age) and their travel diaries (e.g., trip origin, destination, purpose, departure time, and travel mode) on a given day. Based on the provided trip purpose and travel mode information, 1,937 commuters used active modes, such as walking, bicycling, and e-cycling. In China, riding an e-bike does not imply much higher speed than ordinary bikes due to the infrastructure limitation. Thus, this study included it as one of the active travel modes, from home to work in the morning were chosen for this study.

The built-environment characteristics are measured at the TAZ level using the software of ArcGIS. Data sources, including Baidu map POIs, open street maps, and urban land use GIS data, are used. Many studies have focused on the built-environment characteristics surrounding residences, but others, such as Sun et al. [22] and Ding et al. [29] have emphasized the importance of trip destination characteristics. Thus, this study measures the TAZ characteristics for both the home and workplace ends. Among all these data sources, POIs provide geographic information about specific points and are used to calculate transit-related indicators such as intersection density, bus stop density, and metro station density within the TAZ. The open street map is used to calculate the density of roads in each TAZ. The urban land use GIS data are used to calculate three indicators of land use: land use mix and the ratios of residences and working places to the area of TAZ. The land use mix is determined by an entropy index of nine land use types around commuters' residences and work places. The nine different types of land uses include residence, industrial use, public administration, commercial services, green space and plazas, construction, transportation, public facilities, and warehousing. The indicator is calculated as follows:

$$\text{LandMix} = \frac{-1}{\ln n} \sum p_i \ln\ p_i, \tag{1}$$

where p_i is proportion of the type i land use, and n is the number of land use types.

3.2. Data Description

3.2.1. Built Environment at the TAZ Level. Table 1 shows the definitions and descriptive statistics of eight built-environment variables obtained at the TAZ level. All 766 TAZs have the BE characteristics in five dimensions: density, design, distance to transit, destination accessibility, and diversity. Housing density corresponds to the density of residential development. Road density and intersection density that describe the characteristics of a street network represent street design. Bus stop density and metro station density measure the accessibility of bus and subway services and are related to distance to transit. Job density indicates destination accessibility; land use mix represents diversity; and distance to the CBD reflects regional location.

3.2.2. Statistics of Active Commuters. We matched BE features for each active commute trip based on both home-end and work-end TAZs in the trip records. Table 2 depicts the sociodemographics, trip characteristics, and BE characteristics of 1,937 commuters by active modes. Males account for 43.9% of the sample, which is slightly lower than females. 56.4% have a bachelor degree or higher the majority are aged between 30 and 49 and 52.7% own a driving license. The average household has 0.63 cars, while the average number of children aged six years old or below is 0.12. 53.5% of the

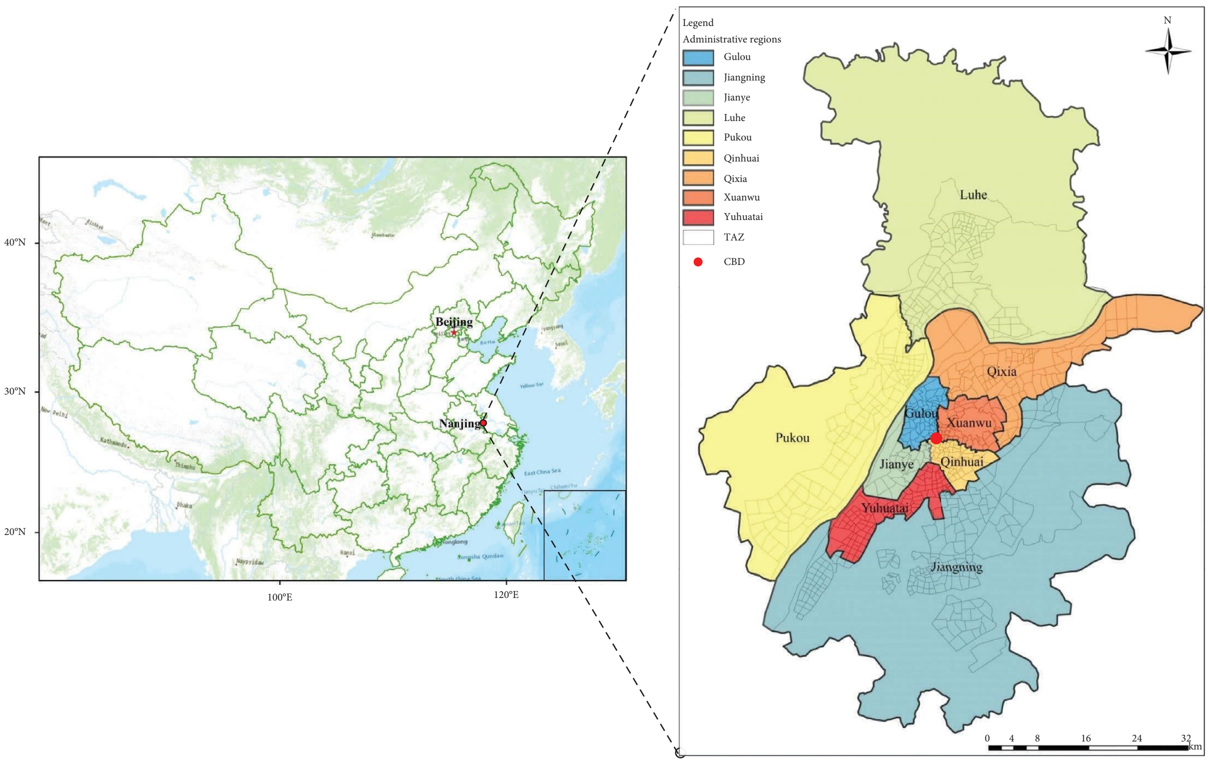


Figure 1: Case study area in Nanjing.

Table 1: Built-environment characteristics for 766 TAZs.

Names	Variable descriptions	Means (S.D.)
Road density	Total road length/TAZ area (km/km^2)	6.85 (4.57)
Intersection density	Intersections/TAZ area (count/km^2)	4.39 (11.55)
Bus stop density	Bus stops/TAZ area (count/km^2)	4.85 (4.95)
Metro station density	Metro stations/TAZ area (count/km^2)	0.23 (0.61)
House density	Residential area/TAZ area	0.19 (0.20)
Job density	Industrial, public administration, and commercial area/TAZ area	0.24 (0.21)
Land use mix	An entropy index of nine types of land use	0.41 (0.19)
Distance to CBD	Euclidean distance from TAZ centroids to CBD* (km)	16.40 (9.66)

*CBD refers to the center of the TAZ in which Xinjiekou business district is located.

respondents have an annual household income over 100,000 CNY. The majority (81.5%) leaves for work between 7:00 am and 8:30 am. The sample's average commute time is 21.52 minutes, and the average trip distance to work is 4.98 km.

4. Methodology

To examine the impact of built-environment characteristics on active commuting time, this study first divides Nanjing's 766 zones into different types using a two-step clustering method. Then, in each region, gradient-boosted regression trees are constructed to investigate the determinants and relative importance of the influential factors on active commuting time. The following sections elaborate the specifics of the analysis.

4.1. Two-Step Clustering Method. The two-step clustering method has been extensively used in the transportation field due to its flexibility and capability in data processing [40, 41]. It has an advantage over other clustering techniques in that it can handle both continuous and discrete variables simultaneously. In addition, it can determine the optimal number of clusters automatically and its clustering accuracy is unaffected by the size of data [42, 43].

The clustering consists of two procedures. First, it clusters the 766 zones into groups according to their similarity in BE characteristics. Then, the merging algorithm is used to gradually combine these groups until only one group is left. The optimal clustering number is determined using Bayesian information criterion (BIC). Interested readers can refer to Chiu et al. [44].

Table 2: Sample description of active commuters ($N = 1937$).

Names	Variable descriptions	Means (S.D./percent)
Sociodemographics		
Gender	Gender: 1 = male; 0 = female	0 = 56.1%, 1 = 43.9%
Education	Hold a bachelor degree or above: 1 = yes; 0 = no	0 = 43.6%, 1 = 56.4%
Age	Respondent's age: 1 = 20–29 years old; 2 = 30–39 years old; 3 = 40–49 years old; 4 = 50 or more years old	1 = 17.6%, 2 = 27.1%, 3 = 36.1%, 4 = 19.2%
License	Hold a driving license: 1 = yes; 0 = no	0 = 47.3%, 1 = 52.7%
Cars	Number of cars owned by a household (count)	0.63 (0.58)
Child	Number of children at 6 years old or younger (count)	0.12 (0.34)
Income	Household income per year: 1 = over 100,000 CNY; 0 = other	0 = 46.7%, 1 = 53.3%
Trip attributes		
Departure time	The commute trip occurs in morning peak hours from 7:00 am to 8:30 am: 1 = yes; 0 = no	0 = 18.5%, 1 = 81.5%
Trip distance	Euclidean distance from residential TAZ centroid to workplace TAZ centroid (km)	4.98 (6.38)
Commuting time	Commuting time spent on road (min)	21.52 (11.98)
Built environment at home end		
Road density	Total road length per TAZ area (km/km^2)	9.78 (3.54)
Intersection density	Intersections/TAZ area (count/km^2)	10.12 (16.99)
Bus stop density	Bus stops/TAZ area (count/km^2)	8.32 (5.39)
Metro station density	Metro stations/TAZ area (count/km^2)	0.53 (0.92)
House density	Residential area/TAZ area	0.31 (0.19)
Job density	Industrial, public administration, and commercial area/TAZ area	0.30 (0.21)
Land use mix	An entropy index of nine types of land use	0.52 (0.16)
Distance to CBD	Euclidean distance from TAZ centroids to CBD* (km)	9.39 (8.38)
Built environment at work end		
Road density	Total road length per TAZ area (km/km^2)	9.77 (3.91)
Intersection density	Intersections/TAZ area (count/km^2)	11.48 (19.05)
Bus stop density	Bus stops/TAZ area (count/km^2)	8.48 (5.47)
Metro station density	Metro stations/TAZ area (count/km^2)	0.46 (0.86)
House density	Residential area/TAZ area	0.31 (0.19)
Job density	Industrial, public administration, and commercial area/TAZ area	0.30 (0.20)
Land use mix	An entropy index of nine types of land use	0.52 (0.16)
Distance to CBD	Euclidean distance from TAZ centroids to CBD* (km)	9.35 (8.59)

*CBD refers to the centroid of the TAZ in which Xinjiekou business district is located.

4.2. Gradient-Boosting Regression Trees. Gradient-boosted regression trees (GBRT) are an ensemble model that combines gradient-boosting and regression trees [45, 46]. It has myriad merits over the traditional linear regression methods and has been often used in transportation research [29, 37, 45]. First, GBRT is more effective at data prediction and interpretation than general linear regressions or even just a single tree due to its tree-based ensemble feature. Second, it accommodates data with missing values and avoids multicollinearity of explainable variables. Third, it can calculate the relative importance of each variable without making assumptions about the variables' relationships. Fourth, it is adaptable to both continuous and categorical types and is applicable to small data sets. Furthermore, it avoids the overfitting issue that frequently arises as the number of tree nodes rises by using gradient boosting.

The GBDT model combines multiple regression trees sequentially with each new tree adding up to correct the errors of the previous ones. Given the training data $\{(y_i, \mathbf{x}_i)\}_1^N$, the specific learning steps are as follows:

(1) Initialize the base model $F_0(\mathbf{x})$ to be a constant:

$$F_0(\mathbf{x}) = \operatorname{argmin}_\gamma \sum_{i=1}^{N} L(y_i, \gamma), \tag{2}$$

where y_i is the observed value, γ is the predicted value, and N is the number of observation. Squared error is chosen as the loss function for the regression.

(2) For $m = 1$ to M (M is the times of iterations or optimal number of trees), compute the residual which is mathematically calculated by the negative derivation of loss function with respect to the previous model outcome:

$$r_{mi} = -\left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})}, i = \{1, \ldots, N\}, \tag{3}$$

where r_{mi} is negative gradient, and $F(x_i)$ is the previous model.

(3) Fit a regression tree to the residuals r_{mi} and minimize the loss function:

$$h_m(\mathbf{x}) = \sum_{j=1}^{J} \gamma_{mj} I\left(\mathbf{x} \in R_{mj}\right),$$
$$\gamma_{mj} = \operatorname{argmin}_{\gamma} \sum_{\mathbf{x}_i \in R_{mj}} L\left(y_i, F_{m-1}(\mathbf{x}_i) + \sum_{j=1}^{J} \gamma I\left(\mathbf{x} \in R_{mj}\right)\right), \tag{4}$$

where $h_m(\mathbf{x})$ is the m th regression tree, J is the tree depth, referring to the number of terminal nodes, R_{mj} is the disjoint region partitioned by the terminal nodes of m th tree, γ_{mj} is the optimal coefficient for R_{mj}, and $I(\mathbf{x} \in R_{mj})$ equals to 1 when $(\mathbf{x} \in R_{mj})$, or 0 otherwise.

(4) Update the model:

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \sum_{j=1}^{J} \gamma_{mj} I\left(\mathbf{x} \in R_{mj}\right). \tag{5}$$

To prevent overfitting in the training procedure, hyperparameters including optimal number of trees M, learning rate ν, and tree depth J should be estimated by using test data or cross-validation. The model is replaced by

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \nu \sum_{j=1}^{J} c_{mj} 1\left(\mathbf{x} \in R_{mj}\right), \tag{6}$$

where ν is the learning rate that scales the contribution of each tree. It has the value range from 0 to 1. Smaller values of learning rate give rise to larger M value and results in minor test error.

The optimal values of these parameters are determined by performing the 5-fold cross-validation. Root mean squared error (RMSE) is chosen as the performance measurement. Parameters that result in the lowest cross-validation error are preferred in the final model. For the 5 test datasets in cross-validation, $\text{RSME}_5^{\text{test}}$ is calculated as follows:

$$\text{RSME}_5^{\text{test}} = \frac{1}{5} \sum_{f=1}^{5} \sqrt{\frac{1}{N_f'} \sum_{t=1}^{N_f'} \left(y_t - \hat{y}_t\right)^2}, \tag{7}$$

where N_f' is the data number in test set f.

Meanwhile, the learned regression trees $\{T_m\}_1^M$ provide interpretative results that show the relative influence of an explanatory variable x_κ as follows [45]:

$$\hat{I}_\kappa^2 = \frac{1}{M} \sum_{m=1}^{M} \hat{I}_\kappa^2(T_m),$$
$$\hat{I}_\kappa^2(T_m) = \sum_{t=1}^{J-1} \hat{\tau}_t^2 I\left[v_t = \kappa\right], \tag{8}$$

where J is the number of terminal nodes, J-1 is the number of the nonterminal nodes, v_t is the feature associated with the node t, $\hat{\tau}_t^2$ is the improvement in squared error after the splitting node t, and $I[v_t = \kappa]$ equals to 1 when $v_t = x_\kappa$, or 0 otherwise.

5. Results

5.1. Identification of the Neighborhood Types. As shown in Table 3, there are 766 TAZs in Nanjing with varying built-environment characteristics. The two-step clustering method is used to cluster these TAZs with more homogeneous spatial features. To eliminate the influence of collinearity on clustering results, the Pearson correlation coefficient is used to test the association between pairs of BE variables, and the variance inflation factors (VIF) are calculated to measure the degree of collinearity. Except for bus stop density, all BE characteristics have coefficients less than 0.6 (0.7–1.0 indicates strongly correlated) and all VIFs calculated are less than 3. Both indicate that all BE variables are suitable for clustering. In a stepwise approach, the ratio change in BIC and ratio of distance measures for a variety of clusters are identified. A model with two clusters appears to be optimal, with a silhouette coefficient value of 0.5.

Table 3 shows the centroids for the two clustered groups as well as the significance of their differences in each BE characteristic. All 766 TAZs are divided into two groups: Cluster-1 with 263 TAZs and Cluster-2 with 503 TAZs. The spatial heterogeneity of TAZs has been interpreted using the centroids for each group. TAZs in Cluster-1 are featured by a higher road density (10.64 km/km^2), more intersections (11.04 count/km^2), more access to metro stations (0.61 count/km^2), higher ratio of residential land (0.40), more job opportunities (0.33), higher land use mix (0.58), and closer proximity to CBD (8.11 km). TAZs in Cluster-2 have a lower road density (4.86 km/km^2), fewer intersections (0.92 count/km^2), less developed metro service (0.04 count/km^2), less residential land use (0.08), lower job coverage (0.19), a lower land use mix (0.33), and are located far away from the CBD (20.74 km). Given the spatial difference between the centroids, we named cluster-1 as the internal region and cluster-2 as the external region. The Mann–Whitney U test method is used to compare the differences in BE characteristics between the two groups. The result demonstrates their spatial differences. The two types of TAZs in Nanjing are shown in Figure 2.

5.2. Results of GBRT. Using the GBM package in RStudio, GBRT models for the commute time of active commuters living in each region are estimated. The relative importance of influential factors is calculated for both identified regions. The relative importance is measured by comparing the error reduction of one variable in commute time compared to other variables. All variables included have a total importance that adds up to 100%. Prior to modeling, hyper-parameters including learning rate, optimal number of iterations (or the number of trees), and tree depth must be tuned. Ridgeway recommended setting the learning rate for

Table 3: Centroids for the TAZ clustering results.

Attributes	Cluster-1 (263 TAZs)	Cluster-2 (503 TAZs)	Mann–Whitney U	Sig.
Road density	10.64	4.86	15990.00	<0.001
Intersection density	11.04	0.92	27695.00	<0.001
Metro station density	0.61	0.04	41884.00	<0.001
House density	0.40	0.08	9588.50	<0.001
Job density	0.33	0.19	34587.00	<0.001
Land use mix	0.58	0.33	16321.00	<0.001
Distance to CBD	8.11	20.74	13653.00	<0.001

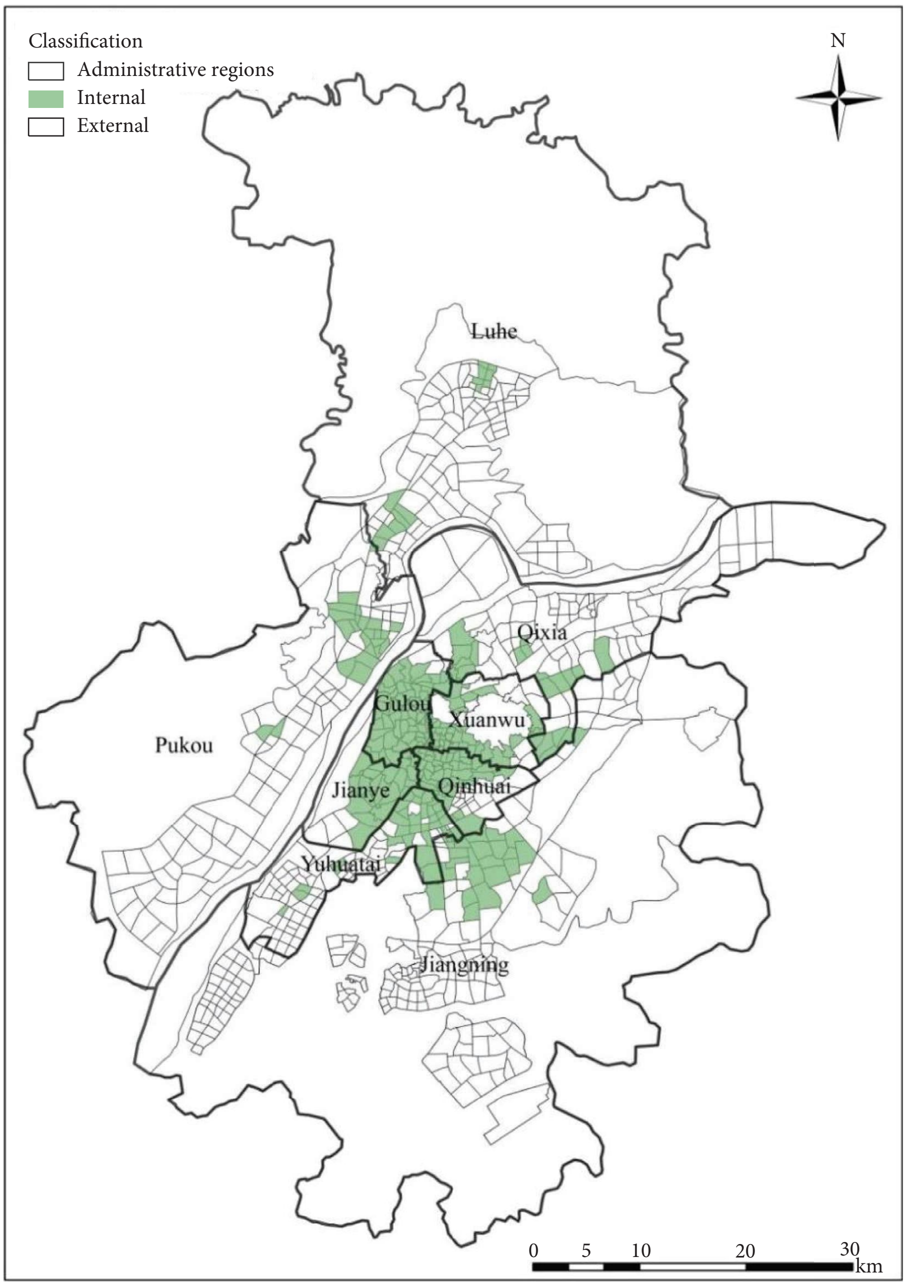


Figure 2: Spatial distributions of TAZs in clustered neighborhood types.

practice between 0.01 and 0.001 [46]. The smaller learning rate is thought to improve model performance. We set the learning rate as 0.001 in accordance with Tao et al. [19]. In order to find the best GBRT, we initially developed the model with the depth of the tree ranging from 1 to 49 in increments of 1. The optimal parameters are then determined using the RMSE value of five-fold cross-validation, which varies as tree depth increases. Figures 3 and 4 visualize the RMSE values versus tree depth and the optimal number of iterations for the internal and external regions, respectively. The RMSE in the internal region decreases with increasing tree depth until it reaches 28. In the external region, however, this indicator becomes stable at a depth of 19. As a result, 28 were set as the tree depth for the model in the internal region and 19 for the model in the external region. The difference between internal and external regions with respect to the number of iterations is even larger. According to the results, the commuting time model in the internal region iterated 3,140 times before convergence, while the model in the external region iterated 2,728 times. Both models fit well, with pseudo-R^2 values of 0.637 and 0.585 in the internal and external regions, respectively. These values are greater than those of traditional linear regressions, which are 0.207 and 0.237. For comparison, we estimated a general GBRT for all active commuters and found that the model has the lower pseudo-R^2 (0.509). This indicates that incorporating spatial heterogeneity in creating the GBRT improves the model fit.

5.2.1. Relative Importance of Influential Factors. Relative importance is commonly used in machine learning to measure how much a factor influences a dependent variable. All variables in this study have a relative importance that sums up to 100%. The greater the relative importance of the factor, the greater it contributes. Table 4 is the calculated relative importance of each influential factor in determining active commuting time in internal and external regions. The result demonstrates that built-environment characteristics have a higher collective importance than social demographics. This is consistent with the findings of some earlier studies [30, 31]. The importance of built-environment characteristics at both commute trip ends is 63.29% and 54.92%, respectively, for internal and external regions. The roughly 8% gap could be due to the more spatially constrained nature of active commute trips in the internal region. In both regions, built-environment features at the work-end pose higher importance than those at home end, which is consistent with the finding of Ding et al. [29]. Similarly, we find differences in the collective importance of sociodemographics in both the regions. They are 5.49% more important in the external region than in the internal region. Active commuters in the external region have more flexibility in determining their commuting time than those in the internal region.

In the internal region, road network density at both trip ends contributes significantly to active commuting time, accounting for 5.24% (ranking 3rd) and 5.23% (ranking 4th), respectively. The effectiveness of active commuting is closely related to the connectivity of the street network, particularly the routes for cycling and walking. This is consistent with the findings of Cao [47]. The intersections density at the work end (4.41%) is shown to have an impact on the trip time of active commuters. Land use mix, job density, and house density at the work end have higher rankings than those at the home end. This can be explained by the high aggregation of morning commutes at the destination over the origin. Similarly, bus stop density at the work end is as high as 4.32%, greater than that at the home end. This confirms the roles of transit accessibility on travel behavior [48], as well as the fact that the resultant trip time is more influenced by the BE feature at the trip end. Metro station density is the least important BE factor, and its importance at both ends is less than 2.00%. This could be explained by the least variation in metro services in the internal region. Geographical locations of home and work ends that are presented by distance to CBD pose the contributions, 3.60% and 3.84%, respectively.

In the external region, intersection density at the home end is the most influential BE factor, with a relative importance of 5.76% while its importance at the work end is 4.47%. Land use characteristics at the work end, including job density, house density, and land use mix, have higher contributions in the external region, ranking third, fourth, and ninth, respectively. In contrast to that in the internal region, the density of metro stations at work ends in the external region has a greater impact, accounting for 4.49% of the total. This may be due to the proximity of metro services to the workplace. The bus stop density at both ends contributes around 3.5%, which is comparable to the internal region. The remaining BE variables had only a minor influence. For active commuters in the external region, the distance from the work end to the CBD (4.22%) contributes more to their trip duration than the home end (3.08%) does.

5.3. Spatial Heterogeneity in BE Impact. To describe the spatial heterogeneity of BE impact, a more thorough comparison of derived BE importance as well as BE associations with active commuting time was made.

5.3.1. BE Importance. Figure 5 shows the comparison of BE importance to active commuting time in two regions. In the internal region, nearly all BE variables have a relative importance more than 3%, with the exception of metro station density at both ends and intersection density at the home end. In external region, all BE variables at the work end and three out of eight BE variables (intersection density, bus stop density, and distance to CBD) at the home end have relative importance over 3%. The most significant disparity is in the roles of street network-related factors, metro station density, and land use-related factors.

The road network density at both ends contributes 1.78%~2.69% more in the internal region than it does in the external region. The intersection density at the home end in the internal region, on the other hand, contributes half as much as it does in the external region. Although metro station density at the home end contributes the least in both regions, its importance in the internal region is six times that of the external region. At the workplace, metro station density is three times as important in the external region as it is in the internal region. When land use variables related to BE variables at home end, such as land use mix, job density, and house density are

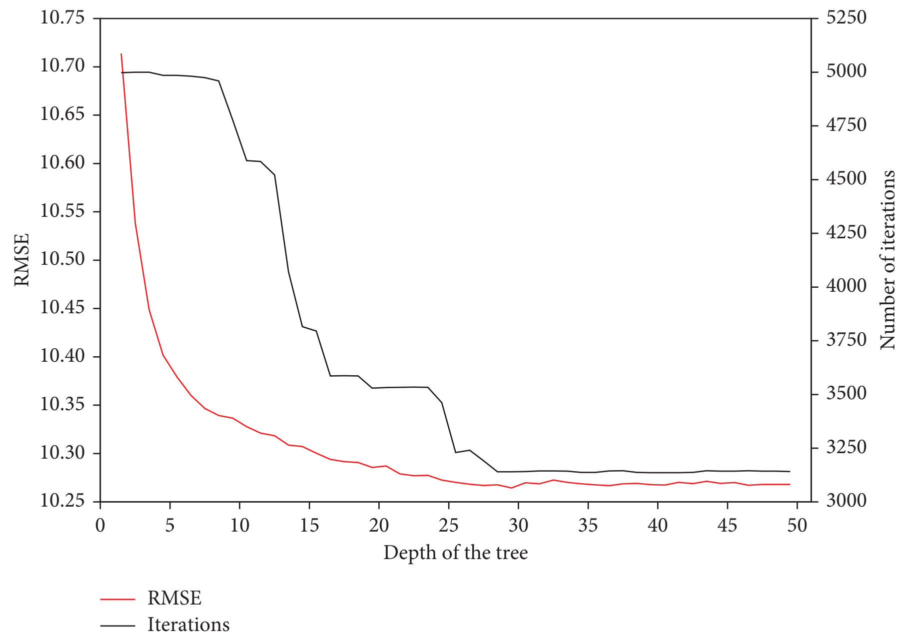


Figure 3: Result of RMSE in the internal region.

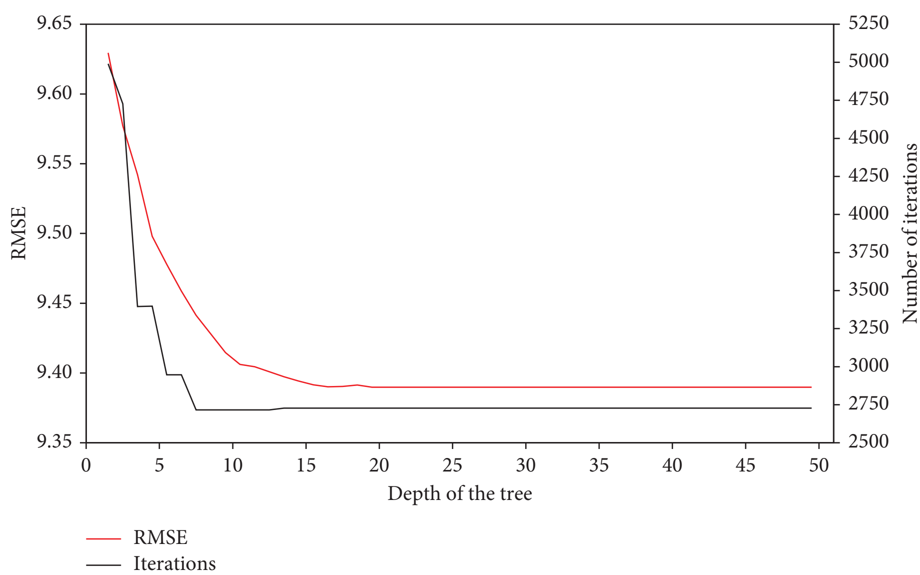


Figure 4: Result of RMSE in the external region.

compared, their roles range between 3.92%~4.28% in the internal region and 1.98%~2.21% in the external region. Notice that at the work end land use mix still holds a more important role in internal region than in external region (5.49% versus 3.47%). These varying effects from region to region are closely related to land use characteristics. The diverse and well-developed land use pattern within the internal region implies greater job options for active commuters. As a result, these factors in the internal region have great effects. TAZs in the external region, on the other hand, are generally less developed in large blocks with homogeneous land use. Distance to the CBD and bus stop density have similar roles in both regions.

5.3.2. Nonlinear Associations between BE and Active Commuting Time. Partially dependent curves are used to present the nonlinear associations between BE characteristics and active commuting time. In GBRT models, partial dependence curves are commonly used to visualize the marginal effects of independent variables on the dependent variable. Figure 6 shows the relationships between BE at home (columns 1 and 2) and work ends (columns 3 and 4) and active commuting time in internal (columns 1 and 3) and external (columns 2 and 4) regions.

Figures 6(a) and 6(b) show nonlinear associations between street network-related characteristics and active commute times. The active commuting time for the internal

Table 4: The relative importance of influential factors in both regions.

Variables	Internal region Rank	Internal region Relative importance (%)	Internal region Sum (%)	External region Rank	External region Relative importance (%)	External region Sum (%)
Built environment at home end			28.97			21.65
Road density	3	5.24		16	2.55	
Intersection density	16	2.68		2	5.76	
Bus stop density	15	3.43		8	3.79	
Metro station density	17	1.70		25	0.26	
House density	11	3.92		18	2.21	
Job density	9	4.28		19	2.02	
Land use mix	10	4.12		20	1.98	
Distance to CBD	13	3.60		12	3.08	
Built environment at work end			34.32			33.27
Road density	4	5.23		10	3.45	
Intersection density	7	4.41		6	4.47	
Bus stop density	8	4.32		11	3.43	
Metro station density	18	1.49		5	4.49	
House density	6	4.69		4	4.69	
Job density	5	4.85		3	5.05	
Land use mix	2	5.49		9	3.47	
Distance to CBD	12	3.84		7	4.22	
Trip attributes			29.68			32.54
Departure time	14	3.50		14	2.72	
Trip distance	1	26.18		1	29.82	
Sociodemographics			7.05			12.54
Gender	19	1.47		17	2.26	
Education	23	0.90		23	1.18	
Age	20	1.40		15	2.72	
License	22	1.08		22	1.25	
Cars	21	1.09		21	1.85	
Child	25	0.33		24	0.30	
Income	24	0.78		13	2.98	
Total			100			100

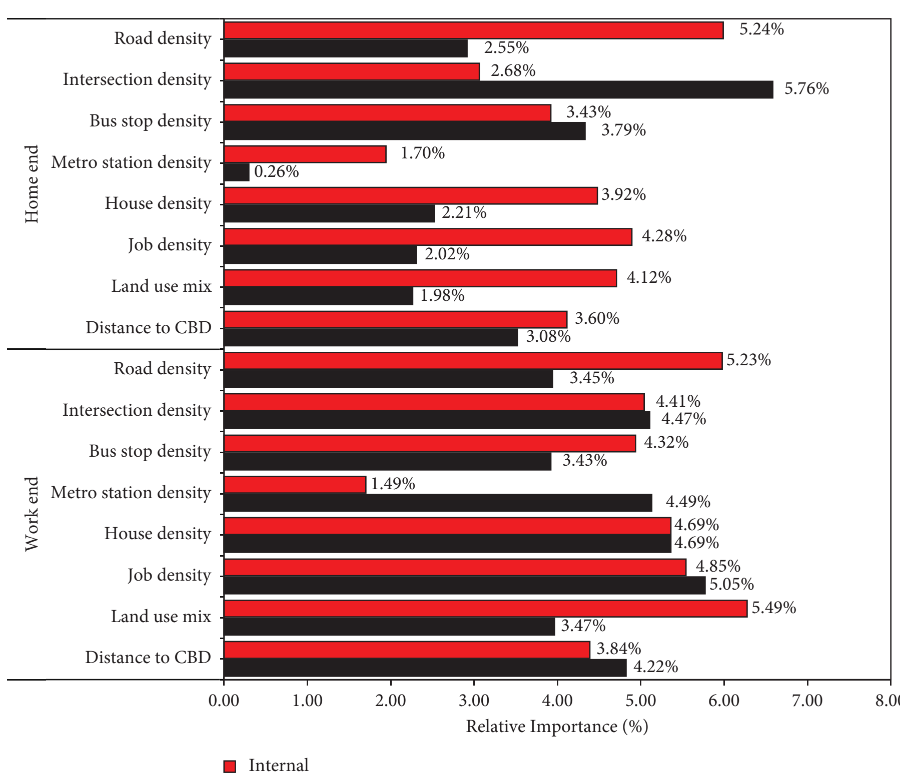


Figure 5: Comparison of relative importance in internal and external regions.

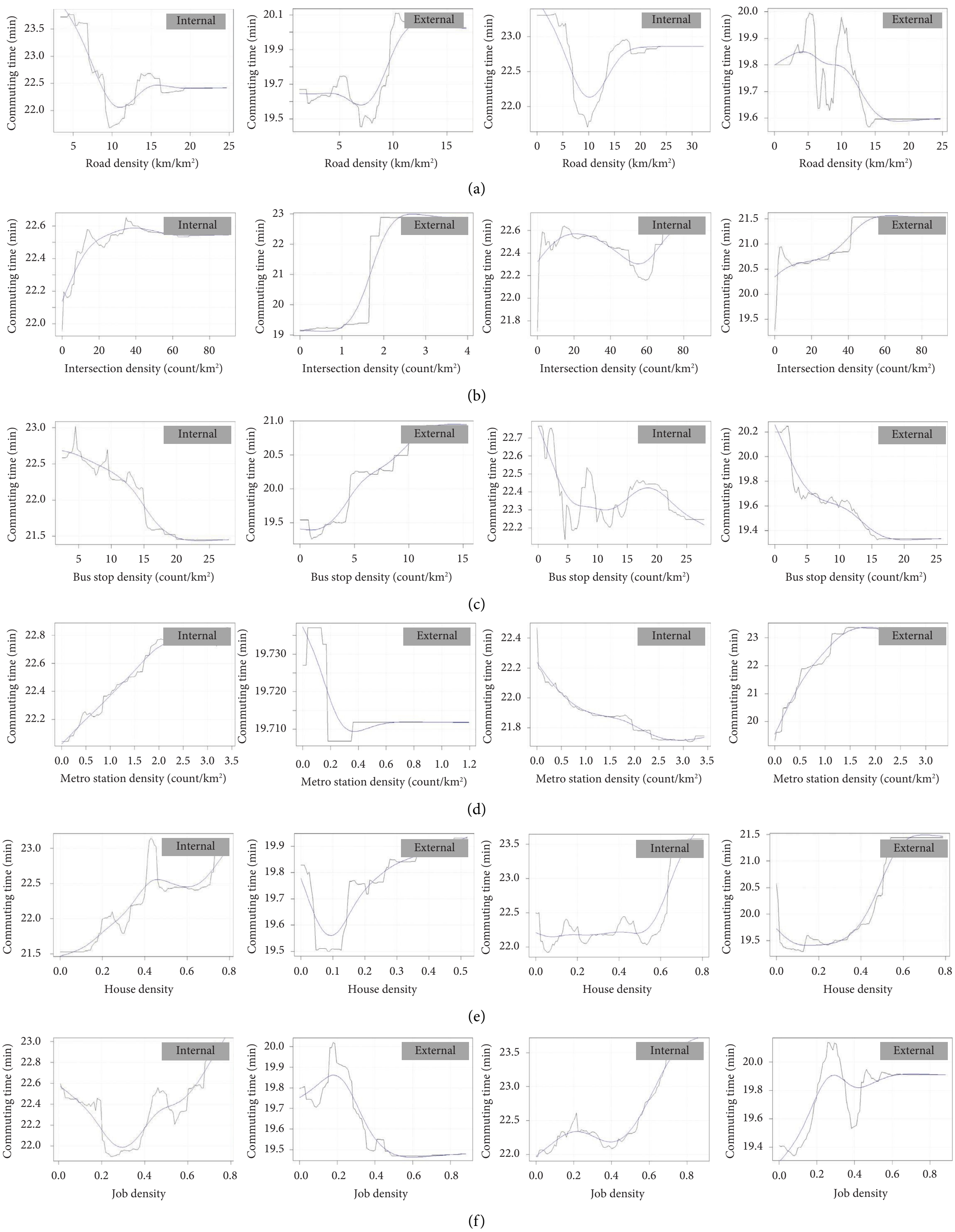
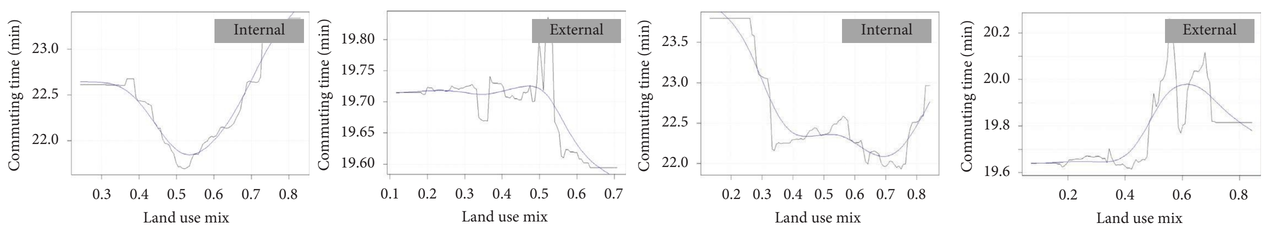
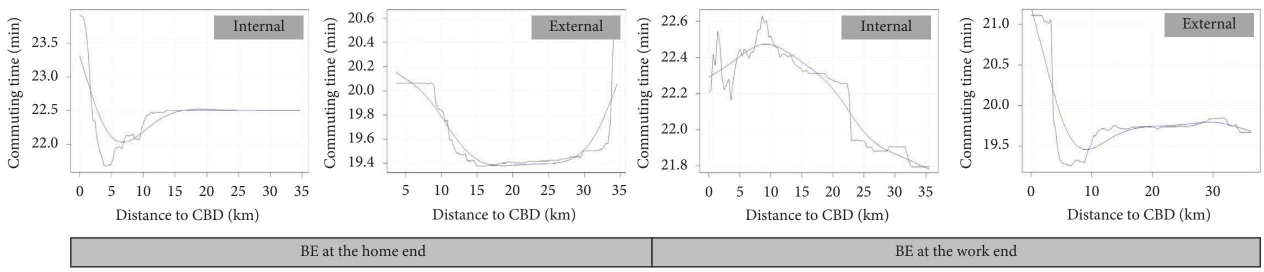


Figure 6: Continued.



(g)



(h)

Figure 6: Nonlinear impact of BE characteristics. (a) Road density (km/km^2). (b) Intersection density (count/km^2). (c) Bus stop density (count/km^2). (d) Metro station density (count/km^2). (e) House density. (f) Job density. (g) Land use mix. (h) Distance to CBD (km).

region decreases rapidly as the home-end road density increases from 0 to 11 km/km^2 and bottoms at 22 minutes. It then gradually increases. Active commuting in the external region decreases until the home-end road density reaches 7 km/km^2, and then it goes up sharply. The curves of work-end road density in both regions also have a nonlinear feature. Within 10 km/km^2, the active commuting time in the internal region decreases sharply in a nearly linear pattern. After that, it climbs to 22.8 min and then remains stable. With increasing work-end road density in the external region, active commuting time fluctuates in an approximate inverted U shape; when the road density at the work end reaches 5 km/km^2 in the external region, the increasing trend in commuting time stops. As shown in columns 1 and 2 of Figure 6, intersection density at home end is positively associated with active commuting time, when it is in the range of 0~40 per km^2 in internal region and 0~3 per km^2 in external region. Higher intersection density is often associated with longer stopping times, which in turn increases active commute times. Both findings reinforce the ambiguous impact of street network design on active mobility [49]. Improved street network connectivity may demonstrate that there are more alternative shortcuts to reach destinations, reducing travel time further, but it may also increase commute time due to more intersections.

Figures 6(c) and 6(d) compare the effects of transit-related variables in both regions. The density of bus stops at the home end has the opposite effects. Active commuting time increases when home-end bus stops increase in the external region but decreases when they increase in the internal region. Both regions see a similar trend in the impact of work-end bus stop density, namely, a reduction in active commuting time with an increase in bus stop density. A minor difference between the two curves is in the range of 10~18 stops/km^2 in the internal region where a fluctuation exists. The impact of metro station density at both ends exhibits the contrast patterns. The extension of home-end metro stations increases commuting time in the internal region while decreasing them in the external region. The extension of work-end metro stations reduces active commuting time in the internal region while increasing it in the external region.

Figures 6(e) and 6(g) illustrate the effects of land use-related variables, including house density, job density, and land use mix. In general, active commuting time increases with increasing house density at the home end, although there are some fluctuations in the curve of the external region within the ratio of 0.1. Because the data points in these intervals are too sparse to interpret, the fluctuations can be ignored. In both regions, house density at the work end has a positive relationship with active commuting time. Work-end house density, as presented by the ratio of residential land at the work end, is better kept within 0.5 for the internal region and within 0.3 for the external region. The general trend can be explained by the greater sense of safety while walking/cycling in areas with higher house density, which leads people to be willing to walk/cycle longer to their work. Job density at the home end has a U-shaped relationship with active commuting time in internal region. The threshold of 0.3 indicates the ideal ratio of work-related land. However, in the external region, the association is generally negative. Commuting time stops decreasing when the ratio of work-related lands reaches 0.5. Active commuting time increases in both regions as job density increases at the workplace.

For land use mix, its association with active commuting time varies by region. For active commuters in the internal region, the best home-end land use mix is around 0.5. However, in the external region, a land use mix of 0.5 or higher is preferable for active commuters. The work-end land use mix, in contrast to the home end, has a winding, decreasing association with active commuting time in the internal region. The active commuting time reaches its lowest point (of 22 minutes) when the land use mix is 0.7. In external region, the work-end land use mix has an inverted U-shaped relationship with active commuting time. It suggests that increasing the land use mix in the external region at the work end is only recommended to a certain extent in order to promote active commuting.

Figure 6(h) shows the relationship between distance to the CBD and active commute time. The residence distance to the CBD has a U-shaped relationship with active commute time in both regions. For active commuters in the internal region, with the distance of their residence to the CBD around 5 km, active commuting time reaches its lowest value (of 21.5 minutes); however, for commuters in the external region, when the distance is in the range of 15~20 km, the time keeps short. At the workplace, its average distance to the CBD is 8.8 km for the internal region and 16.9 km for the external region. We focus more on curve intervals of less than 10 km for internal region and over 15 km for external region. Active commuting time increases with the increasing distance of the workplace from the CBD in both intervals. The rise in active commuting time as a consequence of increasing distance from home to CBD is monotonic.

6. Conclusions and Discussion

With the evidence from Nanjing, China, this paper investigates the spatial heterogeneity in the BE impact on active commuting time. It uses the two-step clustering method to cluster 766 TAZs according to their BE features. Gradient-boosted regression trees are then constructed for each distinguished cluster to examine the heterogeneity in the importance of BE for active commuting time. It has been concluded that built-environment characteristics have more importance than sociodemographics because they contribute 63.29% to active commuting time in the internal region and 54.92% in the external region. This confirms the conclusions of Cheng et al. that despite of the minor impact of single BE factor, their total impacts were larger than that of sociodemographics [30]. The spatial heterogeneity of BE's role to active commuting time is further proved by comparing the nonlinear impact of BE on active commuting time in two regions. Such heterogeneity is so pronounced that in many cases the impact of BE on active commuting time is opposite in two regions.

The implications of this study for practice are two-fold. First, if the goal of urban planners and practitioners is to encourage the use of active modes, they need to prioritize different policies in each region depending on their importance in reaching the ultimate goal. This is especially important when there is a limited budget that can be spared for improving one or a few built-environment characteristics. Second, it is very important for policymakers to be aware of the adverse effects of certain policies in some regions if those policies are developed on the basis of average effects.

The current work comes with some limitations, which leave space for future improvements. First, if the data allows, walkability measures such as sidewalk presence and width should be included as part of conditional variables. Second, the temporal impact of BE on active commuting time deserves attention if new data sources become available. Finally, while the study's findings are applicable to Nanjing and may provide basic references for cities in similar contexts to Nanjing, the BE impact varies from city to city [50]. More case studies are recommended in the future.

Data Availability

The household travel survey data used to support the findings of this study are not available because of data privacy and protection.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Jingxian Wu and Huapeng Shen were in charge of conceptualization, methodology, formal analysis, and initial draft preparation. Jingxian Wu, Guikong Tang, and Soora Rasouli were responsible for review and editing. Jingxian Wu was in charge of funding acquisition, investigation, and supervision. Guikong Tang was in charge of data visualization. Soora Rasouli was responsible for language checks and supervision.

Acknowledgments

This research work was supported by the Scientific Research Fund of the Institute of Engineering Mechanics, China Earthquake Administration (grant no. 2020D23), National Natural Science Foundation of China (grant no. 52122215), and the Soft Science Key Project of Shanghai (grant no. 21692105100).

References

[1] S. Cook, L. Stevenson, R. Aldred, M. Kendall, and T. Cohen, "More than walking and cycling: what is 'active travel'?" *Transport Policy*, vol. 126, pp. 151–161, 2022.

[2] R. Buehler, T. Kuhnimhof, A. Bauman, and C. Eisenmann, "Active travel as stable source of physical activity for one third of German adults: evidence from longitudinal data," *Transportation Research Part A: Policy and Practice*, vol. 123, pp. 105–118, 2019.

[3] R. Buehler, J. Pucher, and A. Bauman, "Physical activity from walking and cycling for daily travel in the United States, 2001–2017: demographic, socioeconomic, and geographic variation," *Journal of Transport & Health*, vol. 16, Article ID 100811, 2020.

[4] A. Banerjee, M. Łukawska, A. F. Jensen, and S. Haustein, "Facilitating bicycle commuting beyond short distances: insights from existing literature," *Transport Reviews*, vol. 42, no. 4, pp. 526–550, 2022.

[5] R. Goel, A. Goodman, R. Aldred et al., "Cycling behaviour in 17 countries across 6 continents: levels of cycling, who cycles, for what purpose, and how far?" *Transport Reviews*, vol. 42, no. 1, pp. 58–81, 2022.

[6] A. Páez and K. Whalen, "Enjoyment of commute: a comparison of different transportation modes," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 7, pp. 537–549, 2010.
[7] H. M. A. Aziz, N. N. Nagle, A. M. Morton, M. R. Hilliard, D. A. White, and R. N. Stewart, "Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data," *Transportation*, vol. 45, no. 5, pp. 1207–1229, 2018.
[8] J. Feng, "The built environment and active travel: evidence from Nanjing, China," *International Journal of Environmental Research and Public Health*, vol. 13, no. 3, p. 301, 2016.
[9] R. Cervero, O. L. Sarmiento, E. Jacoby, L. F. Gomez, and A. Neiman, "Influences of built environments on walking and cycling: lessons from Bogotá," *International journal of sustainable transportation*, vol. 3, no. 4, pp. 203–226, 2009.
[10] J. Wu, B. Wang, N. Ta, K. Zhou, and Y. Chai, "Does street greenery always promote active travel? Evidence from Beijing," *Urban Forestry and Urban Greening*, vol. 56, Article ID 126886, 2020.
[11] Z. Zhu, Z. Li, Y. Liu, H. Chen, and J. Zeng, "The impact of urban characteristics and residents' income on commuting in China," *Transportation Research Part D: Transport and Environment*, vol. 57, pp. 474–483, 2017.
[12] S. L. Handy, M. G. Boarnet, R. Ewing, and R. E. Killingsworth, "How the built environment affects physical activity: views from urban planning," *American Journal of Preventive Medicine*, vol. 23, no. 2, pp. 64–73, 2002.
[13] G. Eldeeb, M. Mohamed, and A. Páez, "Built for active travel? Investigating the contextual effects of the built environment on transportation mode choice," *Journal of Transport Geography*, vol. 96, Article ID 103158, 2021.
[14] Y. Wang, C. K. Chau, W. Y. Ng, and T. M. Leung, "A review on the effects of physical built environment attributes on enhancing walking and cycling activity levels within residential neighborhoods," *Cities*, vol. 50, pp. 1–15, 2016.
[15] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: density, diversity, and design," *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp. 199–219, 1997.
[16] R. Ewing and R. Cervero, "Travel and the built environment," *Journal of the American Planning Association*, vol. 76, no. 3, pp. 265–294, 2010.
[17] L. Cheng, X. Chen, S. Yang, Z. Cao, J. De Vos, and F. Witlox, "Active travel for active ageing in China: the role of built environment," *Journal of Transport Geography*, vol. 76, pp. 142–152, 2019.
[18] L. Yang, S. Griffin, K. T. Khaw, N. Wareham, and J. Panter, "Longitudinal associations between built environment characteristics and changes in active commuting," *BMC Public Health*, vol. 17, no. 1, Article ID 839, 458 pages, 2017.
[19] T. Tao, X. Wu, J. Cao, Y. Fan, K. Das, and A. Ramaswami, "Exploring the nonlinear relationship between the built environment and active travel in the Twin Cities," *Journal of Planning Education and Research*, Article ID 0739456X2091576, 2020.
[20] X. Ma, J. Yang, C. Ding, J. Liu, and Q. Zhu, "Joint analysis of the commuting departure time and travel mode choice: role of the built environment," *Journal of Advanced Transportation*, vol. 2018, Article ID 4540832, 13 pages, 2018.
[21] M. A. Beenackers, S. Foster, C. B. Kamphuis et al., "Taking up cycling after residential relocation: built environment factors," *American Journal of Preventive Medicine*, vol. 42, no. 6, pp. 610–615, 2012.
[22] B. Sun, A. Ermagun, and B. Dan, "Built environmental impacts on commuting mode choice and distance: evidence from Shanghai," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 441–453, 2017.
[23] K. M. Leung, P. K. Chung, D. Wang, and J. D. Liu, "Impact of physical and social environments on the walking behaviour of Hong Kong's older adults," *Journal of Transport & Health*, vol. 9, pp. 299–308, 2018.
[24] M. Moniruzzaman and A. Paez, "An investigation of the attributes of walkable environments from the perspective of seniors in Montreal," *Journal of Transport Geography*, vol. 51, pp. 85–96, 2016.
[25] C. Yin, J. Zhang, and C. Shao, "Relationships of the multi-scale built environment with active commuting, body mass index, and life satisfaction in China: a GSEM-based analysis," *Travel Behavior and Society*, vol. 21, pp. 69–78, 2020.
[26] C. Chen, H. Gong, and R. Paaswell, "Role of the built environment on mode choice decisions: additional evidence on the impact of density," *Transportation*, vol. 35, no. 3, pp. 285–299, 2008.
[27] M. Moniruzzaman, A. Paez, K. M. Nurul Habib, and C. Morency, "Mode use and trip length of seniors in Montreal," *Journal of Transport Geography*, vol. 30, pp. 89–99, 2013.
[28] R. Raman and U. K. Roy, "Taxonomy of urban mixed land use planning," *Land Use Policy*, vol. 88, Article ID 104102, 2019.
[29] C. Ding, X. Cao, and Y. Wang, "Synergistic effects of the built environment and commuting programs on commute mode choice," *Transportation Research Part A: Policy and Practice*, vol. 118, pp. 104–118, 2018.
[30] L. Cheng, J. D. Vos, P. Zhao, M. Yang, and F. Witlox, "Examining non-linear built environment effects on elderly's walking: a random forest approach," *Transportation Research Part D: Transport and Environment*, vol. 88, Article ID 102552, 2020.
[31] J. Liu, B. Wang, and L. Xiao, "Non-linear associations between built environment and active travel for working and shopping: an extreme gradient boosting approach," *Journal of Transport Geography*, vol. 92, Article ID 103034, 2021.
[32] H. Yang, Z. Zhao, C. Jiang, Y. Wen, and M. Muneeb Abid, "Spatially varying relation between built environment and station-level subway passenger-distance," *Journal of Advanced Transportation*, vol. 2022, Article ID 7542560, 18 pages, 2022.
[33] L. Yang, J. Liu, Y. Liang, Y. Lu, and H. Yang, "Spatially varying effects of street greenery on walking time of older adults," *ISPRS International Journal of Geo-Information*, vol. 10, no. 9, p. 596, 2021.
[34] S. Srinivasan and J. Ferreira, "Travel behavior at the household level: understanding linkages with residential choice," *Transportation Research Part D: Transport and Environment*, vol. 7, no. 3, pp. 225–242, 2002.
[35] F. M. Dieleman, M. Dijst, and G. Burghouwt, "Urban form and travel behaviour: micro-level household attributes and residential context," *Urban Studies*, vol. 39, no. 3, pp. 507–527, 2002.
[36] W. Tu, R. Cao, Y. Yue, B. Zhou, Q. Li, and Q. Li, "Spatial variations in urban public ridership derived from GPS trajectories and smart card data," *Journal of Transport Geography*, vol. 69, pp. 45–57, 2018.
[37] W. Zhang, Y. Zhao, X. Cao, D. Lu, and Y. Chai, "Nonlinear effect of accessibility on car ownership in Beijing: pedestrian-

scale neighborhood planning," *Transportation Research Part D: Transport and Environment*, vol. 86, Article ID 102445, 2020.
[38] S. Zhong, Z. Wang, Q. Wang, A. Liu, and J. Cui, "Exploring the spatially heterogeneous effects of urban built environment on road travel time variability," *Journal of Transportation Engineering Part A Systems*, vol. 147, no. 1, Article ID 04020142, 2021.
[39] C. Ding, X. Cao, B. Yu, and Y. Ju, "Non-linear associations between zonal built environment attributes and transit commuting mode choice accounting for spatial heterogeneity," *Transportation Research Part A: Policy and Practice*, vol. 148, pp. 22–35, 2021.
[40] M. A. Abas, S. Rajoo, and S. F. Zainal Abidin, "Development of Malaysian urban drive cycle using vehicle and engine parameters," *Transportation Research Part D: Transport and Environment*, vol. 63, pp. 388–403, 2018.
[41] J. Große, A. S. Olafsson, T. A. Carstensen, and C. Fertner, "Exploring the role of daily "modality styles" and urban structure in holidays and longer weekend trips: travel behaviour of urban and peri-urban residents in Greater Copenhagen," *Journal of Transport Geography*, vol. 69, pp. 138–149, 2018.
[42] M. T. Ballestar, P. Grau-Carles, and J. Sainz, "Customer segmentation in e-commerce: applications to the cashback business model," *Journal of Business Research*, vol. 88, pp. 407–414, 2018.
[43] A. Tkaczynski, S. Rundle-Thiele, and N. Beaumont, "Destination segmentation: a recommended two-step approach," *Journal of Travel Research*, vol. 49, no. 2, pp. 139–152, 2010.
[44] T. Chiu, D. Fang, J. Chen, Y. Wang, and C. Jeris, "A robust and scalable clustering algorithm for mixed type attributes in large database environment," in *Proceedings of the 7th ACM SIGKDDD International Conference on Knowledge Discovery and Data Mining*, pp. 263–268, San Francisco, August 2001.
[45] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
[46] G. Ridgeway, *Generalized Boosted Models: A Guide to the GBM Package*, 2005.
[47] X. J. Cao, "Examining the impacts of neighborhood design and residential self-selection on active travel: a methodological assessment," *Urban Geography*, vol. 36, no. 2, pp. 236–255, 2015.
[48] Y. Fan, K. V. Das, and Q. Chen, "Neighborhood green, social support, physical activity, and stress: assessing the cumulative impact," *Health & Place*, vol. 17, no. 6, pp. 1202–1211, 2011.
[49] J. Wu, B. Wang, R. Wang, T. Na, and Y. Chai, "Active travel and the built environment: a theoretical model and multidimensional evidence," *Transportation Research Part D: Transport and Environment*, vol. 100, Article ID 103029, 2021.
[50] P. Naess, "Meta-analyses of built environment effects on travel: no new platinum standard," *Journal of Planning Education and Research*, vol. 42, no. 2, pp. 199–205, 2019.