

Research Article

How Does Built Environment Affect Metro Trip Time of Elderly? Evidence from Smart Card Data in Nanjing

Zhuangbin Shi ¹, Yang Liu ¹, Mingwei He ¹ and Qiyang Liu ²

¹Faculty of Transportation Engineering, Kunming University of Science & Technology, Jingming Road 727, Kunming 650500, China

²School of Urban Planning and Design, Peking University, Shenzhen Graduate School, Shenzhen 518055, China

Correspondence should be addressed to Mingwei He; hmwei01@163.com

Received 31 January 2022; Revised 10 July 2022; Accepted 16 July 2022; Published 22 August 2022

Academic Editor: Jaeyoung Lee

Copyright © 2022 Zhuangbin Shi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Understanding the determinants of elderly people's public transport usage patterns can offer new insights into elderly mobility issues and provide policy implications for planning toward an aging-friendly and sustainable city. However, few studies have examined the impact of the built environment on the trip time of the elderly using big data. Moreover, the elderly's trip time has been mostly investigated by the multivariate linear regression model (MLR), ignoring the non-linear association between explanatory variables and trip time. Using smart card data from Nanjing in 2019, this study employs a gradient boosting regression trees (GBRT) model to probe into the correlations between the built environment and the elderly's trip time. The results show that significant non-linear relationships exist between trip time and the selected explanatory variables, which cannot be captured by the MLR model. It suggests that relevant policy implementations should be carried out in conjunction with the elderly's travel environment by regarding their threshold effects. Besides, interaction effects of spatial attributes on trip time are identified in our study. For example, elderly people living in the exurban area are more likely to take long-distance metro travel for their physical exercise. These findings demonstrate that planners and policymakers should not only consider one single built-environment factor, but also the interactions of various factors to enhance elderly mobility.

1. Introduction

Due to the accelerating global aging process, the social life and well-being of the elderly have been attracting increasingly academic attention [1–8]. Besides, a positive correlation between population aging and carbon emissions has been proved [9]. Mobility is the fundamental guarantee of elderly people's quality of life (e.g., [10]). In most developing countries, as well as some countries in the New Europe, where driving is rarely a travel mode for the elderly, elderly mobility is highly dependent on walking and public transport. Therefore, their outdoor activities are usually associated with their physical abilities [11, 12]. With the rapid pace of urbanization in emerging economies, elderly public transport users have been increasing dramatically. In China, about 56.36% of elderly people travel by public transit, 19.41% of which are metro users [13]. Public transit, therefore, plays a critical role in the elderly's daily mobility

and wellbeing [14]. However, nearly 20% of the elderly are still experiencing various degrees of mobility disability due to the inconvenience of transportation.

Although a large body of literature delved into the determinants of elderly mobility, these determinants are still under scrutiny. First, previous studies mainly focused on either elderly's driving or active travel (i.e., walking and cycling), while public transport, especially metro, which, as aforementioned, has become increasingly the main travel mode for Chinese elderly people, is still received inadequate attention. Compared with young adults who usually use the car for long-distance trips, older people with the significantly lower physical condition are likely to give up their trips [15]. To enable the elderly to travel longer distances, local governments have provided corresponding preferential policies to encourage the elderly's public transport use [16]; however, the result can hardly reach policymakers' expectations because the determinants of older people's travel distance are

still unclear [17]. Second, the elderly's travel distance has been mostly investigated by a multivariate linear regression model, but in practice, the relationship between explanatory variables and travel distance may be non-linear [18]. Some other scholars used hazard-based models, logit models, or other non-linear regression models (e.g., log-linear) to examine the determinants of older people's travel; however, the relationship between explanatory variables and dependent variables in these models is also pre-specified. As a result, it cannot efficiently capture the unexpected relationship caused by spatial attribute thresholds. Third, most of these studies assume that the explanatory variables are independent and uncorrelated. However, there may be a certain degree of correlation between multiple explanatory variables even if the variables pass the test of multi-collinearity, especially those related to the built environment [19]. Thus, the influence of explanatory variables on the dependent variable in the model may be much greater than their actual effects. In addition, most existing studies were based on data from traditional cross-sectional surveys, which is hard to obtain the mobility of a large volume of elderly in transit networks over time.

To fill the above research gaps, this study attempts to investigate the non-linear associations between the built environment and the elderly's travel distance using smart card data and other spatial data. Travel distance has been widely used to understand travel behavior in the literature (e.g., [20–22]), because of not only its direct correlation with trip satisfaction (e.g., [23]) but its impacts on urban development, social-spatial segregation and transport-related social inequities [24]. This study, therefore, aims to understand the determinants of the elderly's travel distance by public transit, which can offer new insights into elderly mobility issues and provide policy implications for planning for an aging-friendly city.

This study contributes to the current literature in the following three aspects. First, it adds to the existing literature on the elderly's travel behavior through data mining methods. Compared with traditional travel survey data, smart card data can trace individual travel during a long period, and can reflect the habitual behavior of specific groups [25, 26]. Thus, using smart card data to investigate the mobility of older people as metro riders can provide new insights to planners and policymakers. Second, gradient boosting regression trees (GBRT), which have advantages in investigating the nonlinear relationship and the interaction of variables, are employed to consider the correlations between explanatory variables while examining factors influencing the elderly's trip time. Finally, through the case study of the elderly in Nanjing, we propose policy implications to improve public transport services for the elderly which are generalizable to wider transit-oriented contexts.

Despite adults aged 65+ being widely considered to be the elderly in most Western countries, in this study, people above 60 years old are defined as the elderly because 60 is the common retirement age in Mainland China. The rest of the study is organized as follows: Section 2 reviews the literature on factors influencing the elderly's travel behavior. Section 3 introduces the study area and describes the data set. Section

4 discusses the methodology used, while the modeling results are reported in Section 5. The discussion of policy implications and conclusions of the study are provided in Section 6.

2. Literature Review

There is a growing literature focusing on the travel behavior of elderly people, but few attempts have been made on the determinants of the elderly's travel behavior in a given transport system, especially for the travel distance in the metro system. Therefore, the literature review mainly focuses on the elderly's travel behavior and influencing factors in general.

Previous studies have intensely discussed the elderly's travel behavior in terms of travel purpose, travel mode, travel frequency, and distance/time (e.g., [2, 15, 21, 27–29]). Because of the different cultural backgrounds and car/driving license ownership rates, the elderly's travel mode in developed and developing countries varies significantly [7]. In most Western countries such as the United States, Canada, and the Netherlands, a car is the most popular travel mode for the elderly, with the lowest proportion of public transport (bus, coaches, train, and rail). Thus, previous studies in these countries have been mainly concerned with factors that influence elderly's travel behavior without focusing on elderly people's public transport use [12]. In such contexts, researchers emphasize investigating travel distance and frequency to understand the elderly's travel behavior. For example, empirical evidence revealed that elderly women are less likely to use cars than their counterparts men, thereby resulting in a shorter travel distance [30, 31]. With the increase of age, the elderly tend to decrease their travel frequency and distance in general [32], but their travel distance may increase for recreational purpose [33]. Some elderly people tend to shift from car to public transit, mainly due to functional status decline [14, 32, 34, 35]. This indicates that public transport is still fundamental to maintaining elderly mobility, especially for those who do not drive, even in car-dominant contexts. Consequently, researchers discussed the influence of the concessionary bus travel policy on elderly people (e.g., [16, 36–40]). Such interventions aimed to encourage the elderly's public transport use so as to increase multimodally and enhance the elderly's functional abilities; however, it is not necessarily applicable in developing countries where car ownership is low [41, 42].

In China, less than 1% of the elderly traveled by car [11]. Thus, numerous studies have been conducted to investigate the active travel behavior of the elderly such as travel frequency, walking/cycling duration, walking/cycling distance to transit, and its determinants (e.g., [3, 5, 18, 43–45]). However, only a few studies have focused on the elderly's public transport use, although more than half of elderly people in China are regular public transport users. Zhang et al. [42] examined the elderly's travel behavior during the morning peak in the context of the free bus program in Beijing. Shao et al. [25] analyzed the elderly's bus usage patterns between weekdays and weekends based on GPS data and smart card data in Qingdao. These studies focus more on

travel mode choice or travel characteristics of the elderly, but the underlying causes of the behavior, such as travel distance, were not examined.

The determinants of the elderly's travel behavior have been extensively discussed in the literature. There is no doubt that traveler socio-demographics attributes (including age, gender, income, racial ethnicity, and medical conditions) and trip characteristics (including travel mode and travel purpose) are important factors affecting travel distance [46]. However, when using large-scale smart card data, travel-related personal information is often difficult to obtain. On the other hand, several studies have confirmed that built environment attributes have highly significant direct and indirect effects on residential travel distance/time [47–49]. Specifically, empirical evidence demonstrates that the built environment attributes such as population density, building density, land-use mixture, traffic conditions, and destination accessibility have significant influences on the elderly's travel [3, 39, 50, 51]. Neighborhoods with higher population density and higher degrees of land-use mixture decrease the travel distance of the elderly [21]. Similarly, older people living in core areas are less likely to travel long distances and prefer to use public transport [42, 50], but those living in rural areas usually have a much longer travel distance [2, 39, 52]. In addition, if the elderly live close to the bus stops, metro stations, and bikeshare stations, they are inclined to use public transport [21, 39, 53–56].

Aiming at the Chinese elderly, Feng [21] found that wet markets, open spaces, and parks as well as chess and card rooms show profound impacts on the elderly's travel distance. Seniors living near wet markets increase their total and shopping distance, and closer to open spaces/parks and chess/card rooms increase their leisure activities. Similar results were reported by Cheng et al. [3], who also found that the elderly in China are more concerned with the distribution of some specific facilities, for instance, shopping facilities, convenience stores, parks/squares, and chess/card rooms instead of supermarket or gym/sports center.

Overall, most existing studies assume that the associations between built environment attributes and the elderly's travel behavior are linear, only a few recent studies considered nonlinear associations between variables in individual travel decisions. For instance, Tao et al. [18] examined the importance of spatial attributes to walking distance to transit and its threshold effect. They found that some spatial attributes such as population density have clear threshold effects on transit users' walking distance to stops, with different nonlinear patterns. Gan et al. [19] used smart card data to investigate the nonlinear relationship between the built environment attributes and station-to-station ridership. The discontinuous nonlinear effects across different values of the built environment attributes on metro ridership have been found in their study. Similar findings have been reported by Shao et al. [57] and Yang et al. [58]. However, these studies are not focused on the topic of elderly travel.

In theory, the nonlinear associations between the elderly's travel behavior and the determinants may be attributed to two mechanisms. The first is the law of diminishing marginal utility, which describes how the establishment of

facilities at a low-density level yields more utility than subsequent establishment [59, 60]. It also can explain why the density of the built environment has a threshold effect [61, 62] and the effect produced by densification is usually not good as expected in East Asian cities [63], especially compared with European cities. The second one is the nonlinear associations between built environment attributes and walking distance may result from physiological limitations [61, 64]. Since walking is the most common mode of undertaking the first/last mile for elderly's metro travel, the interactions of walking distance to metro station/final destination and metro travel distance may present a more complicated nonlinear relationship. Therefore, this study focuses on the determinants of travel distance for the elderly metro users, adopting a gradient boosting regression trees mode (GBRT) to explore the importance of the built environment attributes on elderly' travel distance in a Chinese context.

3. Study Area and Data

We adopted Nanjing as a case study. Nanjing is an important economic, educational center, and transit metropolis of Eastern China. The permanent residents of Nanjing reached 8.50 million, with the urbanization rate of 83.2% [65]. As the urban spatial structure continues to expand, Nanjing's metro has also been developing rapidly. Since the first metro line officially launched in 2005, 10 metro lines have been operating in Nanjing at the end of 2018, namely Metro Line 1, Line 2, Line 3, Line 4, Line 10, Line S1, S3, S7, S8, and Line S10, with a total of 159 metro stations. The total length of metro lines climbed to 378 km and the average daily passenger flow has reached 3.53 million, accounting for 19.2% of the total urban travel. In this study, Nanjing was divided into three regions, namely urban, suburban, and exurban regions (shown in Figure 1). The urban area consists mainly of mixed-use land, where jobs and educational resources are densely distributed and the quality of metro facilities is reasonably high; whereas the exurban region is made up of single-use, low-density areas where transport infrastructures, especially metro, are underdeveloped. In 2019, there were 17 metro stations in the urban region, 77 in the suburban region, and 65 in the exurban region.

Since July 2010, the elderly in Nanjing over 60-years-old can apply for the Senior Citizen Concession Card (SCCC). SCCC can be used on most of the public transport services in Nanjing City, including buses, metros, and ferries. When applying, CNY 20 needs to be paid as the card fee. Using the card, seniors aged 60–69 can enjoy a 50% discount on the usage of public transport in Nanjing, while those who are aged over 70 (including 70 years old) are free of charge. There is no time limit for this discount. Under the implementation of the elderly's public transport concessions program, more and more of Nanjing's elderly people prefer to choose the bus and metro as their main travel modes. Consequently, bus/metro-based travel is largely affecting their quality of daily life. Notably, since bus passengers in most cities of China don't need to swipe their cards when

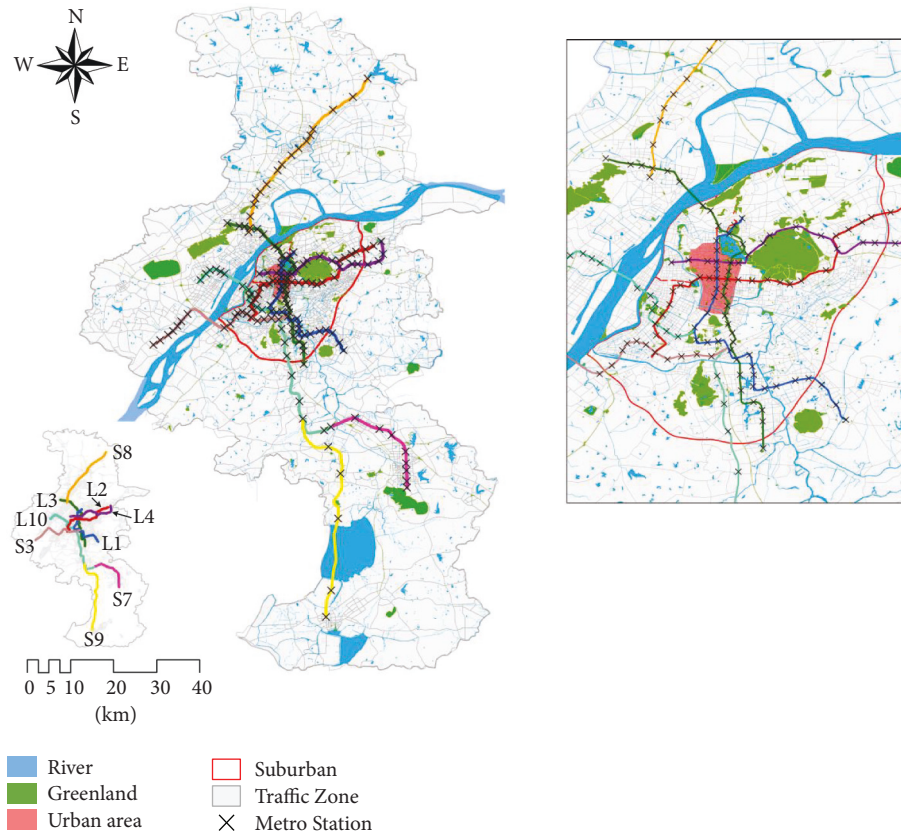


FIGURE 1: Study area and the existing metro lines in Nanjing, China.

they get off the bus, it is difficult to obtain the travel distance of bus riders. Therefore, this study only takes metro riders as the research object.

The data used for this study consists of two parts, one of which is the smart card data of Nanjing Metro from March 1, 2019 to June 28, 2019 including 120 consecutive days. The data were used to extract the travel characteristics of the elderly (i.e., trip time, travel frequency, etc.). As shown in Table 1, information used from smart card data includes date of arrival, time of arrival, date of departure, time of departure, card type (No. 85 represents the elderly card holder), arrival station number, departure station number, and card number. Only the elderly card holders were investigated in this study. After the collection of the original data, the data was then cleaned to filter out invalid data such as data with the same arrival and departure station, and those who arrived and departed on different days. Besides, the trips with a trip time longer than (or shorter than the mean of OD trip time plus (or minus) three times of the variance) were also removed. Finally, we got a total of 5.81 million valid metro smart card use records by 387 thousand elderly metro users in 120 continuous days 2019, with an average of 48.40 thousand trips per day. Considering the trade-off between time-consuming costs and quantity demand, we randomly selected 200 thousand records from the entire data set with `pandas.DataFrame.sample` function in Python, and set up random sampling without putting back and with equal probability weighting, which is enough to

investigate the relationship between elderly trip time and the determinants (total 25 independent variables) in subsequent modeling and analysis.

To investigate the effects of the built environment around metro stations on the elderly's travel distance, the point-of-interest (POIs) data of Nanjing in 2019 from the Baidu Map application programming interface (API) was also used in this study. This data covers all POIs categories related to the elderly's travel, mainly including shopping malls, convenience stores, supermarkets, wet markets, restaurants, open spaces/parks, places of interest, chess/card/tea rooms, healthcare, gym/sports centers, bus stops, metro stations, bikesharing stations, residential sites, and employment sites. Besides, socioeconomic attributes around metro stations may also affect the travel behavior of the elderly who use the metro. However, population and employment data are not available from the station level. Therefore, the population density and employee density of traffic zones in 2019 are employed in this paper. The traffic zones of Nanjing and the urban road network in the same year were respectively obtained from Nanjing Urban Planning Bureau (NUPB) and OpenStreetMap (<https://www.openstreetmap.org/>).

4. Methodology

4.1. Descriptive Statistics. The distributions of the variables used in the analysis are shown in Table 2. Since the speed of

TABLE 1: Metro smart card data samples on March 2019.

Card ID	Arrival data	Departure data	Station ID at entry	Arrival time	Station ID at departure	Departure time	Card type
970070****81	2019/3/9	2019/3/9	5	15:55:51	28	16:06:34	51
970071****56	2019/3/9	2019/3/9	61	7:25:07	1	8:02:51	52
970071****23	2019/3/11	2019/3/11	5	9:54:11	1	10:13:56	52
970074****73	2019/3/11	2019/3/11	26	15:03:54	98	15:56:05	85
...

TABLE 2: Descriptive statistics for the independent and dependent variables.

Variable	Definitions	Min	Max	Mean	Std. dev
Trip time	The trip time on the metro of each trip (min)	2.22	288.55	29.73	20.59
<i>Built environment</i> ¹					
Employment density_O	The density of jobs in 800 m buffer of boarding station (person/m ²)	0.00	0.40	0.06	0.09
Employment density_D	The density of jobs in 800 m buffer of alighting station (person/m ²)	0.00	0.40	0.06	0.09
Distance to the nearest bike station_O	Network distance to the nearest bike station of boarding station (km)	0.03	18.19	0.44	1.30
Distance to the nearest bike station_D	Network distance to the nearest bike station of alighting station (km)	0.03	18.19	0.44	1.29
Distance to the nearest bus stop_O	Network distance to the nearest bus stop of boarding station (km)	0.02	12.08	0.30	0.28
Distance to the nearest bus stop_D	Network distance to the nearest bike station of alighting station (km)	0.02	12.08	0.30	0.27
Distance to the nearest chess/card room_O	Network distance to the nearest chess/card room of boarding station (km)	0.06	11.83	0.57	0.79
Distance to the nearest chess/card room_D	Network distance to the nearest chess/card room of alighting station (km)	0.06	11.83	0.57	0.78
Distance to the nearest convenience store_O	Network distance to the nearest convenience store of boarding station (km)	0.01	2.98	0.29	0.22
Distance to the nearest convenience store_D	Network distance to the nearest chess/card room of alighting station (km)	0.01	2.98	0.29	0.22
Distance to the nearest park/open square_O	Network distance to the nearest park/open square of boarding station (km)	0.01	7.38	0.59	0.64
Distance to the nearest park/open square_D	Network distance to the nearest park/open square of alighting station (km)	0.01	7.38	0.59	0.64
Distance to the nearest restaurant_O	Network distance to the nearest restaurant of boarding station (km)	0.01	4.23	0.13	0.21
Distance to the nearest restaurant_D	Network distance to the nearest restaurant of alighting station (km)	0.01	4.23	0.14	0.21
Number of Sport/gym_O	Number of sport/fitness centres in 800 m buffer of boarding station	0.00	128.00	24.89	25.31
Number of Sport/gym_D	Number of sport/fitness centres in 800 m buffer of alighting station	0.00	128.00	25.17	25.51
Number of tourist attractions_O	Number of tourist attractions in 800 m buffer of boarding station	0.00	129.00	15.65	24.55
Number of tourist attractions_D	Number of tourist attractions in 800 m buffer of alighting station	0.00	129.00	16.04	24.97
<i>Travel characteristics</i>					
Daily frequency	Average number of trips per day	0.00	4.12	0.50	0.53
Departure time	MonPeak = 0, other = 0	0.00	1.00	0.18	0.38
	NightPeak = 1, other = 0	0.00	1.00	0.10	0.31
<i>Network attributes of metro stations</i>					
Betweenness centrality_O	The betweenness of boarding station	0.00	0.44	0.14	0.10
Betweenness centrality_D	The betweenness of alighting station	0.00	0.44	0.14	0.10
Location of boarding station	Urban area = 1, suburban area = 2 and exurban area = 3	1.00	3.00	1.88	0.67
Location of alighting station		1.00	3.00	1.88	0.67

Note 1: Due to a large number of built environment variables, we only show the variables with the top 50% importance according to equation (8).

the metro can be considered constant, the travel distance can be reflected, to a certain extent, by the trip time between boarding and alighting at the metro station, which could be

easily obtained from smart card data. Thus, trip time on the metro is the dependent variable in this study. After calculating the trip time of all elderly passengers during 120

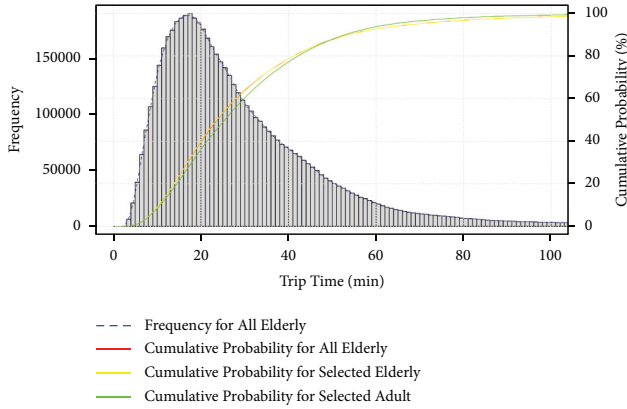


FIGURE 2: Frequency distribution of the metro-based trip time.

continuous days, we found that their mean trip time was about 29.77 minutes. As shown in Figure 2, 85% of the elderly metro users traveled less than 46.7 min. As trip time increases, the frequency sharply decreases after more than 46.7 minutes. Thus, 46.7 min can be defined as the acceptable longest travel duration of the elderly on the metro. The average number of trips an elderly passenger generated during these 120 continuous days was 15.01, including 10.74 trips on weekdays and 4.27 trips on weekends. Moreover, we used a binary variable to represent the elderly's departure time, with 1 for morning peak or evening peak and 0 for other time. We found that the elderly mostly traveled during non-peak hours—only about 18.42% of these trips were made in the morning peak (7:00 am–9:00 am), while 10.67% were in the evening peak (5:00 pm–7:00 pm).

It can be seen from Figure 2 that there is little or no difference between the cumulative probability distributions of all elderly samples and selected elderly samples, which indicates that the selected samples are sufficiently representative to reflect the behavior patterns of the elderly in Nanjing Metro. Besides, we also compared the features of metro-based trip time between elderly and normal adults (200 thousand records of adults that were randomly selected). As shown in Figure 2 and Table 3, there is no significant difference in the mean trip time between these two groups, whereas the trip time of adults is more concentrated than that of the elderly. Comparing the cumulative probability distributions of elderly and normal adults, it can be found that the proportion of elderly people who travel less than 20 minutes by metro is higher than that of adults, and at the same time, the proportion of elderly people who travel for a long time (longer than 60 minutes) is also higher than that of adults. This may be due to the relatively few time and work constraints of the elderly.

Because the socioeconomic attributes of card holders are missing due to privacy concerns, we are mainly concerned with factors of the built environment that influence elderly's travel distance. In addition, the elderly's travel characteristics such as metro usage frequency and departure time, and network attributes of metro stations were included in the analysis because they may also affect the trip time of the elderly.

TABLE 3: The features of trip time across different datasets.

Dataset	All elderly samples	Selected elderly samples	Selected adult samples
Mean trip time (min)	29.77	29.73	30.03
Std. deviation trip time (min)	20.61	20.59	18.6

Based on the aforementioned discussion, we assumed that built environment characteristics near elderly's boarding and alighting stations such as population/employment density, building density, facilities, and public transport availability potentially affect their travel behavior. Thus, we extracted the built environment characteristics from the data described in Section 3 by using ArcGIS 10.4.1 software. The number of tourist attractions, sports/gym centers, chess/card rooms, convenience stores, restaurants, park/open squares, wet markets, bus stops, and bikeshare stations within a radius of 800 m of the elderly's boarding and alighting stations were measured. It is worth noting that the number of population and jobs were measured at the traffic zone level, where metro stations are located. Similarly, the shortest route on the road network distance from the boarding/alighting station to the buildings and facilities was also measured to indicate the availability of elderly's shopping and leisure activities as well as public transport.

As for network attributes of metro stations, the region where the boarding and alighting station are located and betweenness centrality were adopted in this study. Among them, the region where the boarding and alighting station is located is a categorical variable, with 1 for an urban area, 2 for a suburban area, and 3 for an exurban area. A mean of 1.88 indicates that most of the elderly enter the metro in the urban area. The same result can be found when leaving the metro. On the other hand, betweenness centrality was used to quantify the importance of a given station on the connectivity between other stations in a metro network. Betweenness is a centrality measure based on shortest paths, which is widely used in metro network analysis. The betweenness centrality of the given station i can be defined as

$$C_i^B = \sum_{i \neq k \neq m} \frac{\phi_{km}(i)}{\phi_{km}}, \quad (1)$$

where ϕ_{km} is the total number of shortest paths from station k to station m , and $\phi_{km}(i)$ is the number of these paths that pass through station i .

4.2. Gradient Boosted Regression Trees Model. The gradient boosted regression trees model (GBRT) was applied in this study to investigate the nonlinear associations between the trip time of the elderly and its affecting factors. The GBRT model boosting belongs to the ensemble learning method, which is a combination algorithm of regression tree and gradient boosting [66]. GBRT has been increasingly adopted in various research fields, including the transport field because of its powerfulness of investigating the nonlinear relationship and the interaction of variables [19, 63, 64, 67].

GBRT integrates decision trees in an additive manner based on a machine learning algorithm, and can be expressed as follows:

$$\text{tree } f(\mathbf{x}) = f_M(\mathbf{x}) = \sum_{m=1}^M h_m(\mathbf{x}, \Theta), \quad (2)$$

where $f(\mathbf{x})$ is the estimated value of the GBRT model corresponding to a set of explanatory variable vectors \mathbf{x} , $h_m(\mathbf{x}, \Theta)$ is the estimated value of the m th regression tree, Θ is the parameter of the m th regression tree, and M is the number of regression trees.

Unlike the traditional ensemble learning method, namely Adaboost which is improved by adjusting the weight of misclassified data samples, the GBRT model is improved by calculating the negative gradient of previous trees. In general, $L(y, f_m(\mathbf{x}))$ can be used to represent the loss function of the model, where y is the actual observed value of the dependent variable. In the case of a regression tree, the loss function is the sum of the squared errors as follows:

$$L(y, f_m(\mathbf{x})) = (y - f_m(\mathbf{x}))^2. \quad (3)$$

We expanded the loss function in (2) by its first-order Taylor expansion at \mathbf{x} , and the following approximate equation can be obtained:

$$\begin{aligned} L(y, f_m(\mathbf{x})) &= L(y, f_{m-1}(\mathbf{x}) + h_m(\mathbf{x}, \Theta)), \\ &\approx L(y, f_{m-1}(\mathbf{x})) + \frac{\partial L(y, f_{m-1}(\mathbf{x}))}{\partial f_{m-1}(\mathbf{x})} h_m(\mathbf{x}, \Theta). \end{aligned} \quad (4)$$

In order to reduce the loss function, we need to construct the m th regression tree by learning the direction of the negative gradient of the loss function. According to (3), the negative gradient of the loss function can be expressed as

$$\nabla_m = - \left[\frac{\partial L(y, f(\mathbf{x}))}{\partial f(\mathbf{x})} \right]_{f(\mathbf{x})=f_{m-1}(\mathbf{x})}. \quad (5)$$

In order to improve the robustness of the model and avoid overfitting, one effective way is to introduce a learning rate so as to shrink the contribution of each tree. There is a trade-off between the learning rate λ ($0 \leq \lambda \leq 1$) and the number of estimators M . Thus, the GBRT model can be expressed as follows:

$$\begin{aligned} f(x) &= f_m(x) = f_{m-1}(x) + \lambda h_m(x, \Theta) \\ &= f_{m-1}(x) + \lambda \sum_{j=1}^J c_{m,j} I, \end{aligned} \quad (6)$$

where $R_{m,j}$ is the j th leaf node region of the m th regression tree ($j = 1, 2, \dots, J$), $c_{m,j}$ is the output value of the leaf node region $R_{m,j}$, and I is an indicator variable,

$$I = \begin{cases} 1, & \text{if } x \in R_{m,j} \\ 0, & \text{otherwise} \end{cases}$$

The implementation steps of the GBRT model are shown in Figure 3. Since the GBRT model is a machine learning method, it is difficult to directly interpret the estimated parameters. But fortunately, the GBRT model is able to quantify the relative importance of explanatory variables based on the estimated GBRT model. The importance of explanatory variables is the increase in the squared error of the model after we permuted the values of the variable. Mathematically, the importance of the i th explanatory variable is defined as the mean of its importance in each regression tree:

$$\begin{aligned} \tilde{I}_i &= \frac{1}{M} \sum_{m=1}^M \tilde{I}_i^m = \frac{1}{M} \sum_{m=1}^M (OOB_{MSE,perm-i}^m - OOB_{MSE}^m), \\ I_i &= \frac{1}{K} \sum_{k=1}^K \tilde{I}_k * 100\%, \end{aligned} \quad (7)$$

where \tilde{I}_i and I_i are the initial and the normalized variable importance, respectively, OOB_{MSE}^m is the mean square error of tree m before permuting variable x_i , $OOB_{MSE,perm-i}^m$ is the mean square error after permuting variable x_i .

Besides, the partial dependence plot (PDP) produced by the GBRT model displays the marginal effect one or two variables that have on the predicted outcome, whether the association is linear, exponential, or more complex [66]. The partial dependence function for GBRT is formulated as follows:

$$\hat{f}_{x_S}(x_S) = E_{x_C}[\hat{f}_{x_S}(x_S, x_C)] = \int \hat{f}_{x_S}(x_S, x_C) dP(x_C), \quad (8)$$

where x_S are the variables for which the PDP should be plotted and x_C are other explanatory variables in the GBRT model. Usually, the integral formulation in (9) is approximately estimated by a Monte Carlo method, namely the average of the training data $1/N \sum_{i=1}^N \hat{f}_{x_S}(x_S, x_{i,C})$.

5. Results and Discussions

5.1. Baseline Model and the GBRT Model Regularization. For the purpose of comparison, we firstly built a multiple linear regression (MLR) model as a base line, and calculated the variance inflation factors (VIF) for each explanatory variable (as shown in Table 4). The R^2 of the MLR model is 0.21. Except for the employment density and the distance to the nearest chess/card room, most of the explanatory variables are significant with a confidence level above 99%. The results of the multi-collinearity test show that there is no significant multi-collinearity between the built environment, travel characteristics, and network attributes because all of the explanatory variables have a VIF value lower than 4. Thus, they will be retained in the GBRT model according to the suggested threshold (larger than five) for excluding variables [68]. Nevertheless, medium correlations were found among the built environment attributes and the location of the station with some Pearson's correlation coefficients lying between ± 0.30 and ± 0.49 .

According to the estimated coefficients in Table 4, the more convenient other destinations can be accessed by using

the metro, the shorter the trip time for elderly is. The number of jobs and tourist attractions has a positive impact on trip time. In addition, metro stations with higher connectivity and located in the city center could lead to shorter trip times. The above results demonstrate that the MLR model explains the determinants of the elderly's trip time from a global average point-of view, without considering the possible nonlinear associations between the dependent variable and explanatory variables. To explore the impact of spatial thresholds on the elderly's trip time, we report the results of the GBRT model.

First, in order to avoid over-fitting, we developed the GBRT model with a 5-fold cross-validation procedure. The training dataset was randomly divided into five subsets with the same sample size. Then, four of the subsets were used to train a GBRT model in proper sequence, and the remaining subset was used as validation data to evaluate the performance of the model. Following the basis of the general experience of machine learning, we set the learning rate λ as 0.005, and the depth of the tree as 5. The model converged after 30,000 iterations and achieved the minimum deviance on the validation subset at 55,968 boosting iterations. The testing pseudo- R^2 (the square of Pearson correlation coefficient between the predicted value and real value) for the GBRT model is 0.86, which is far larger than the R^2 of the MLR model.

5.2. The Relative Importance of Explanatory Variables. Table 5 presents the relative importance of explanatory variables in predicting trip time among elderly travelers for both the GBRT model and the baseline MLR model. In the MLR model, the relative importance can be defined as the improvement of the model goodness-of-fit R^2 by adding one explanatory variable [18, 64]. A strong correlation was found between the relative importance of explanatory variables for the GBRT and the MLR - the value of the Pearson correlation coefficient is 0.69. However, Table 5 shows that the rank is quite different between the relative importance of explanatory variables (GBRT) and % R^2 (MLR). It is probably because of the weak hypotheses of the MLR model that the relationships are linear for log-linear and the explanatory variables are irrelevant to each other.

The relative importance indicates the relative increase in the impurity while permuting some explanatory variable. According to the result of the GBRT model, the locations of metro stations—both boarding and alighting—are the most important explanatory variables, which contribute to a total of 15.18% of explanation power. The joint contribution of the built environment attributes is 72.01%, whereas the three travel characteristics only account for 3.70% of explanation power. It suggests that the trip times of elderly's metro use mainly depend on the location of the stations and surrounding environment attributes. Besides, the departure time (morning peak, evening peak, or non-peak) also slightly influences the travel duration.

As for the relative importance of location and surrounding environment attributes, no obvious significant difference between the boarding and alighting stations can be observed. The aggregated contribution of explanatory

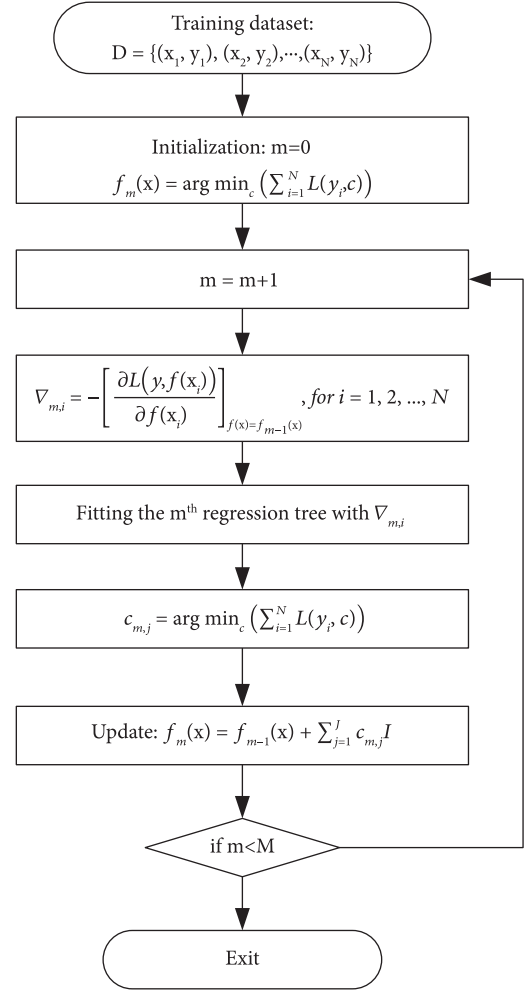


FIGURE 3: Solution process of GBRT model.

variables is 46.51% and 50.32% for the boarding and alighting stations, respectively. It is probably because the metro is usually used for closed-loop travel (e.g., an elderly may go shopping at a convenience store from home, and then get back home after the activity is finished).

The relative importance of single-build environment attribute collected from both boarding and alighting stations for a trip are ordered from high to low: distance to the nearest bike station (10.75%), distance to the nearest convenience store (10.12%), distance to the nearest restaurant (8.48%), distance to the nearest bus stop (8.20%), distance to the nearest chess/card room (8.11%), employment density (7.20%), distance to the nearest park/open square (6.73%), sport/gym density (6.70%), tourist attractions density (5.69%). It indicates that the transfer facilities near metro stations, including bikesharing facilities and bus stops, play a major role in the elderly's trip time. Especially, distance to the nearest convenience store around the boarding station is the most important explanatory variable among the built environment attributes. This is expected because Chinese elderly people usually undertake the responsibility of acquiring family daily necessities. Compared with a healthy lifestyle such as exercise and outdoor travel, the trip time is

TABLE 4: Linear regression model results.

Description	Coefficient	<i>t</i> -test	<i>p</i> value	VIF
<i>Constant</i>				
<i>Built environment</i>				
Employment density_O	0.627	1.222	0.222	1.343
Employment density_D	-0.056	-0.107	0.915	1.362
Distance to the nearest bike station_O	0.948	26.192	0	1.323
Distance to the nearest bike station_D	1.042	28.518	0	1.316
Distance to the nearest bus stop_O	2.673	15.23	0	1.385
Distance to the nearest bus stop_D	3.144	17.837	0	1.384
Distance to the nearest chess/card room_O	-0.042	-0.526	0.599	2.407
Distance to the nearest chess/card room_D	0.062	0.768	0.443	2.415
Distance to the nearest convenience store_O	1.008	3.617	0	2.248
Distance to the nearest convenience store_D	2.247	8.066	0	2.317
Distance to the nearest park/open square_O	2.406	24.651	0	2.283
Distance to the nearest park/open square_D	2.677	27.592	0	2.293
Distance to the nearest restaurant_O	5.662	21.125	0	1.884
Distance to the nearest restaurant_D	4.736	17.757	0	1.926
Number of sport/gym_O	0.008	3.817	0	1.839
Number of sport/gym_D	0.012	5.342	0	1.842
Tourist attractions density_O	0.047	20.363	0	1.893
Tourist attractions density_D	0.053	23.21	0	1.907
<i>Travel characteristics</i>				
Daily frequency	-0.035	-53.882	0	1.018
Morning peak	0.937	8.598	0	1.044
Evening peak	-1.643	-12.006	0	1.036
<i>Network attributes of metro stations</i>				
Betweenness centrality_O	-14.964	-32.487	0	1.351
Betweenness centrality_D	-13.811	-30.241	0	1.346
Location of boarding station	5.238	46.978	0	3.281
Location of alighting station	5.251	47.153	0	3.307

more easily affected by indoor sedentary entertainment—playing chess/card. This result is consistent with the results of Feng [21] and Cheng et al. [64].

5.3. The Nonlinear Associations between Explanatory Variables and Trip Time. Partial dependence plots were used to demonstrate the nonlinear associations between the explanatory variables and the elderly's metro trip time. In addition to the predicted curves produced by the GBRT model, we drew smooth curves for easily observing the general trend of the association. Notably, all the explanatory variables related to distance are intercepted within 2 km which is different from the general service range (800 m), because the buffer zone of the metro station can be extended with bike-sharing mode as a feeder.

5.3.1. Built Environment Characteristics. As for the built environment variables, distance to the nearest bike station is the most important factor influencing the elderly's metro trip time (10.74%), followed by distance to the nearest convenience store (10.11%), restaurant (8.47%) and bus stop (8.20%). We only analyze the four most important variables due to the limitation of length. Figure 4(a) presents the effects of distance to the nearest bike station on predicting the trip time of elderly metro users, controlling for all the other explanatory variables. The general trend is that

distance to the nearest bike station is negatively associated with trip time, while a significant nonlinear association exists on both boarding and alighting stations. Taking distance to the nearest bike station on the boarding (origin) side as an example, distance to the nearest bike station on the boarding side has a linear effect on trip time when it increases from 0 to around 200 m; then the effect becomes positive (negative sometimes) when the distance to the nearest bike station is from 200 m to 1,000 m; trip time decreases from 34 min to 28 min when this variable exceeds 1,000 m. Perhaps this result occurs because long-distance elderly metro users are more likely to accept longer walking transfer distances which can be considered as sunk costs of metro trips. But metro stations surrounded by accessible bike-sharing stations (within 250 m) are usually located in areas with higher degrees of the land-use mixture, which, however, decreases the likelihood of long-distance travel. Similar results can be found from the alighting (destination) side.

Figure 4(b) presents the partial dependence plots for distance to the nearest convenience store on boarding and alighting sides. It is clear that distance to the nearest convenience store has a significant nonlinear association with trip time. Distance to the nearest convenience store has almost no effect when it is less than about 700 m on both sides, while the trip time differs between boarding and alighting stations when the distance to the nearest convenience store exceeds 700 m. As for the boarding station,

TABLE 5: Relative importance of all the variables.

Description	The relative importance of MLR (%)	Total (%)	Relative importance of GBRT (%)	Total (%)
<i>Constant</i>				
<i>Built environment</i>				
Employment density_O	0.412		3.650	
Employment density_D	0.497		3.553	
Distance to the nearest bike station_O	4.120		5.013	
Distance to the nearest bike station_D	4.862		5.734	
Distance to the nearest bus stop_O	1.711		3.868	
Distance to the nearest bus stop_D	2.287		4.333	
Distance to the nearest chess/card room_O	3.266		3.957	
Distance to the nearest chess/card room_D	3.941		4.154	
Distance to the nearest convenience store_O	3.309		3.677	
Distance to the nearest convenience store_D	4.466	72.00	6.440	56.44
Distance to the nearest park/open square_O	7.203		3.366	
Distance to the nearest park/open square_D	8.239		3.397	
Distance to the nearest restaurant_O	4.327		4.247	
Distance to the nearest restaurant_D	4.127		4.235	
Sport/gym density_O	1.726		3.565	
Sport/gym density_D	1.580		3.131	
Tourist attractions density_O	0.226		2.682	
Tourist attractions density_D	0.139		3.006	
<i>Travel characteristics</i>				
Daily frequency	6.567		2.899	
Morning peak	0.238	3.17	0.196	7.27
Evening peak	0.467		0.074	
<i>Network attributes of metro stations</i>				
Betweenness centrality_O	4.211		4.271	
Betweenness centrality_D	3.806		5.369	
Location of boarding station	14.087	24.82	8.214	36.29
Location of alighting station	14.185		6.968	

1500 m is an important turning point: before which a significant decline of the trip time is observed, whilst after which the trip time increases rapidly again. However, the curve is much more flattened on the alighting site. This may be because 700 m is a walking distance threshold for the elderly to use the metro for shopping, and the elderly seem to be not sensitive to the change of walking distance to convenience stores within this acceptable threshold. It is less likely for the elderly to use the metro to access convenience stores that are not within a walkable distance. Some elderly people are likely to ride shared bikes as a feeder mode from convenience stores to their boarding station because 1,500 m is a cycling distance threshold for the feeder in general. This explains the elevation of trip time after 1500 m on the boarding side. That is, bike-sharing provides opportunities for the elderly to acquire daily necessities from more alternatives.

Figure 4(c) presents the partial dependence plots for distance to the nearest restaurant on both sides. The curves indicate the association between the variable and trip time differs between the boarding and alighting station. Trip time is relatively stable from 0 to 1500 m but increases rapidly after 1500 m on the boarding side, while it has a general

growth trend on the alighting side except for the first 300 m. This suggests that 1,500 m is a spatial threshold on the boarding site. Elderly metro users tend to use a shared bike as a feeder mode to travel long distances for meals, as the distance from the restaurant to the metro increases. This result cannot be found in the linear model (see Table 4).

Figure 4(d) demonstrates the nonlinear effects of distance to the nearest bus stop on trip time. Specifically, the effect is negative when the variable is less than about 300 m, while it becomes positive when the distance to the nearest bus stop is between 300 m and 1250 m. After that, trip time becomes stable. This phenomenon may be the consequence of both land use and travel behavior. On the one hand, metro station with high-density bus stop is usually located in areas with higher degrees of land-use mixture, and hence elderly people can participate in their activities within a short travel range. On the other hand, 1,250 m may be the distance threshold for elderly people taking the bus from home to the boarding station, and those who use the bus as a feeder mode tend to make long-distance trips, where the walking transfer distance can also be taken as a sunk cost. Therefore, increasing bus stops within 1,250 m around the metro station and the residential areas could encourage elderly people to

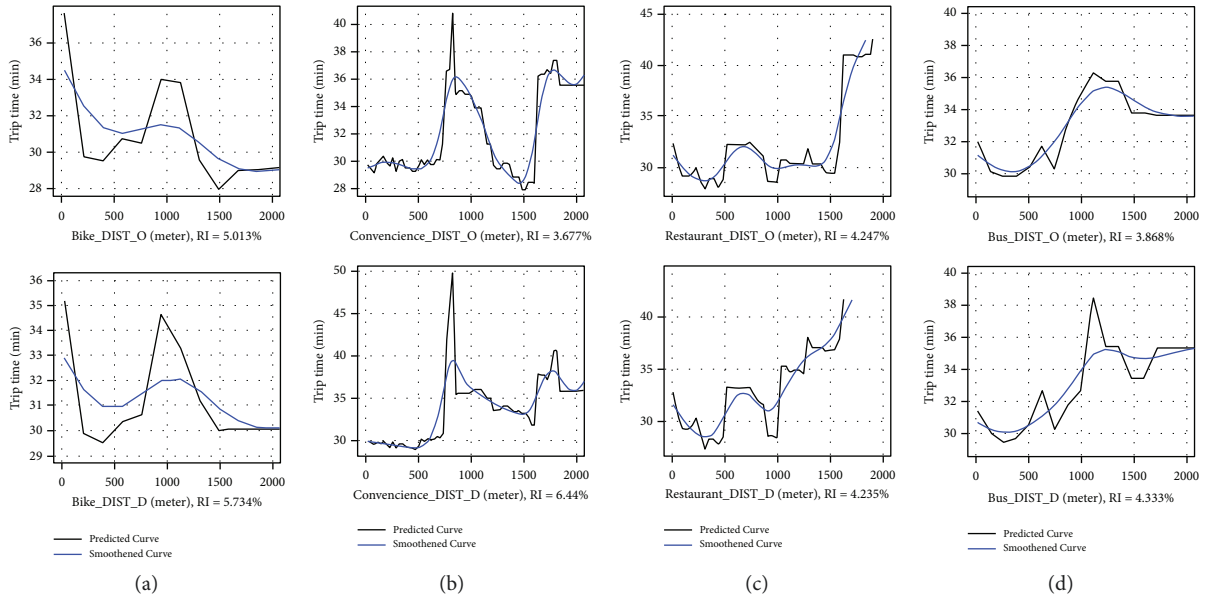


FIGURE 4: The effects of built environment on trip time (the top four variables of importance).

travel by metro-bus integrated mode, especially for the elderly who travel a short distance.

5.3.2. Other Explanatory Variables. Other explanatory variables we considered in this study are travel characteristics and network attributes of metro stations. As for travel characteristics, the daily usage frequency of the metro is the most important variable. Figure 5(a) presents the effects of daily usage frequency on predicting the trip time of elderly metro users, controlling for all other explanatory variables. The elderly's daily usage frequency of the metro is negatively associated with the trip time when they travel less than three times per day. However, this effect turns positive when the daily usage frequency of the metro exceeds three times per day. Therefore, three times per day can be deemed as an important usage frequency threshold of the metro, before which the trip time decreases from 30.5 min to 27 min, after which it increases from 27 min to more than 31 min. This finding indicates that with the increase in usage frequency of the metro, the trip time of the elderly decreases in general. However, those who use the metro more than three times per day are more likely to travel long distances. Since the average number of elderly's daily metro use is about 0.5 times (Table 2), the daily usage frequency of metro is negatively correlated with trip time in general. This result is consistent with the literature (e.g., [29, 43]).

As for network attributes of metro stations, both betweenness centrality on the boarding and alighting stations and the region where the stations are located play key roles in trip time. Among them, the partial dependence plots for station location are three straight lines because station location variables are categorical variables, and we will analyze them in the next section in detail. Figures 5(b) and 5(c)

present the partial dependence plots for betweenness centrality on the boarding and alighting side, respectively. Taking Figure 5(b) as an example, we found that trip time decreases sharply (from 34 min to 29 min) when the betweenness degree is within the range of 0–0.16; and trip time becomes stable at about 0.16. This indicates that the betweenness degree has a notable threshold effect on trip time, which may be caused by the law of diminishing marginal utility. Since the index reflects the proportion of shortest paths going through the boarding station in the metro network, a larger betweenness value indicates a more important boarding station for the elderly. Therefore, within the threshold range, trip time decreases with the increase of station importance. However, there are still many stations in Nanjing whose betweenness centrality is far less than 0.16. Supply buses and shared bikes should be added around these metro stations so as to increase the betweenness degrees to 0.16.

5.4. The Interaction Effects of Explanatory Variables on Trip Time. The GBRT model can not only have advantages in fitting nonlinear and nonregular relationships between the explanatory variables and trip time, but also can automatically capture the interaction effects among explanatory variables [58, 69]. Figure 6 presents an intelligible example to understand the phenomenon of variable interaction. If the location of the boarding and alighting station does not interact, the partial dependence function can be decomposed as follows:

$$\hat{f}_{L_b, L_L}(i+1, j+1) = \hat{f}_{L_b, L_L}(i, j) + \Delta_{i,j}^{i+1,j} \hat{f}_{L_b, L_L} + \Delta_{i,j}^{i,j+1} \hat{f}_{L_b, L_L}, \quad (9)$$

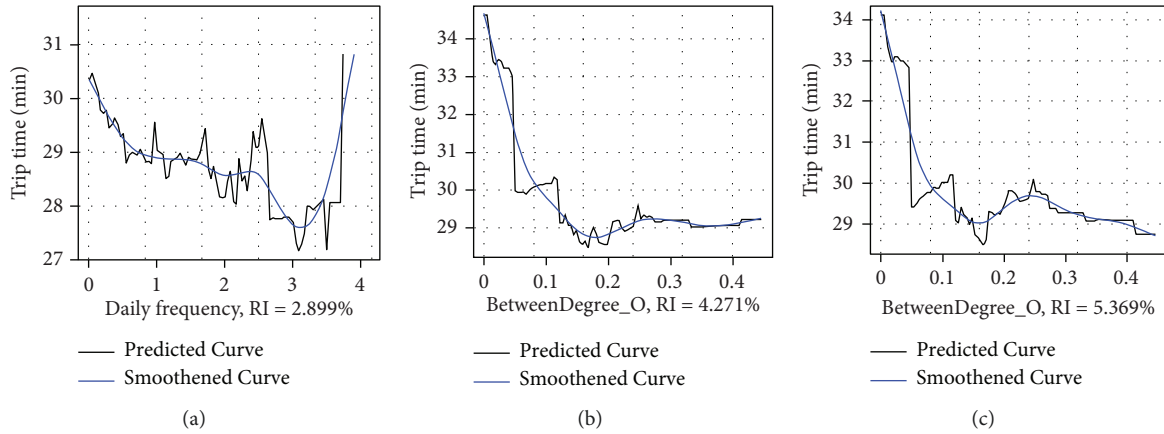


FIGURE 5: The effects of daily travel frequency and betweenness degree on trip time.

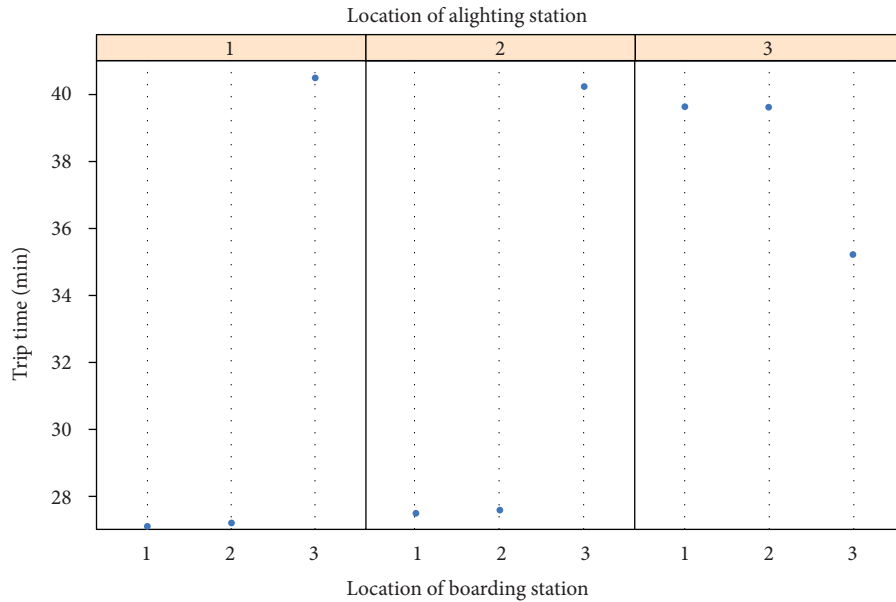


FIGURE 6: Interaction effects of boarding and the alighting stations' locations on metro trip time.

where $\hat{f}_{L_b, L_L}(i, j)$ is the partial dependence function with $L_b = i$ and $L_L = j$, $\hat{f}_{i, j}^{i+1, j} \hat{f}_{L_b, L_L}$ denotes the deviation as $\hat{f}_{L_b, L_L}(i + 1, j) - \hat{f}_{L_b, L_L}(i, j)$.

Thus, $\hat{f}_{L_b, L_L}(3, 3)$ should be the sum of $\hat{f}_{L_b, L_L}(3, 2)$ and the deviation $\hat{f}_{L_b, L_L}(2, 3) - \hat{f}_{L_b, L_L}(2, 2)$ that is obviously larger than zero. However, the actual value of $\hat{f}_{L_b, L_L}(3, 3)$ is even lower than $\hat{f}_{L_b, L_L}(3, 2)$. It means that the location of the boarding station and the location of the alighting station interact with each other. As an additive model, the MLR model will provide an overestimation for the trip time while the origin and destination are all located in an ex-urban area.

We used H-statistic mathematically method proposed by Friedman and Popescu [70] to measure the interaction between explanatory variables. The H-statistic is dimensionless and always in the range of 0 to 1—an interaction statistic of 0 indicates no interaction between the two

explanatory variables, and the statistic is 1 when the estimated trip time only comes through the interaction. The total H-statistic (variable k vs. all other variables) is quite difficult to evaluate, especially in our case whose data has high variable dimensions and large sample size. Therefore, we only compute the two-way H-statistic (variable j vs. variable k). The average H-statistic for each explanatory variable with other variables is displayed in Figure 7. It shows that the distance to the nearest chess/card room surrounding the alighting station has the strongest interaction with other explanatory variables. The value of the H-statistic (=18.5%) means that an average of 18.5 percent variance (difference between observed and no-interaction PD) can be explained by the interaction effects between the nearest chess/card room on the original side with other explanatory variables.

A high relative interaction effect on trip time was found between distances to the nearest restaurant and chess/card

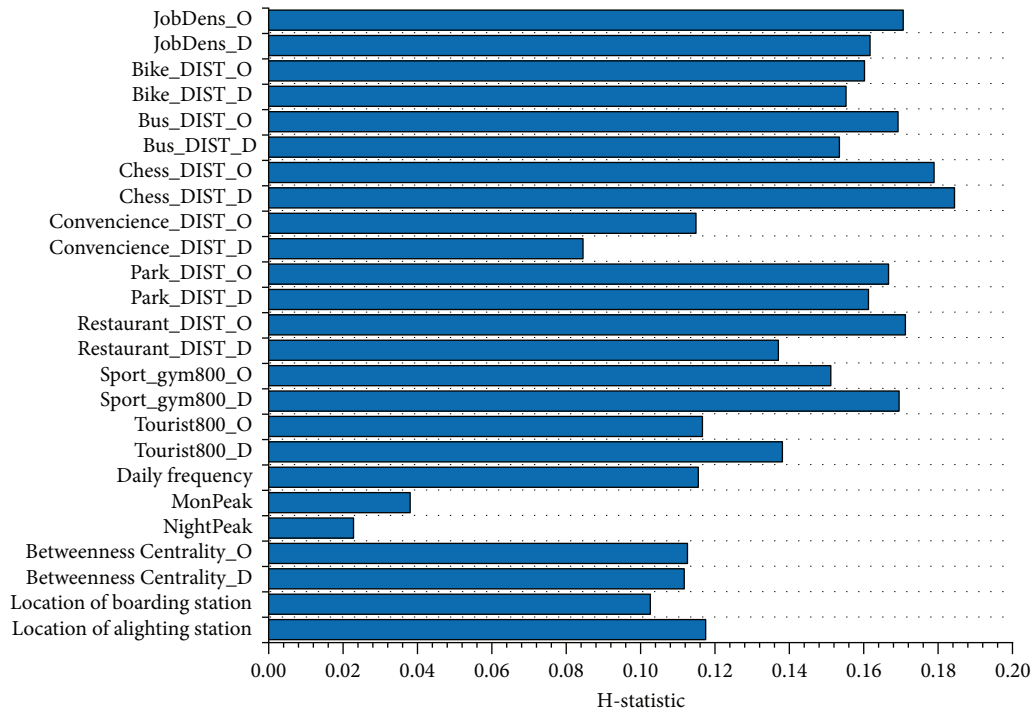


FIGURE 7: The average interaction strength (H-statistic) for each explanatory variable with other explanatory variables for the GBRT model predicting the elderly trip time.

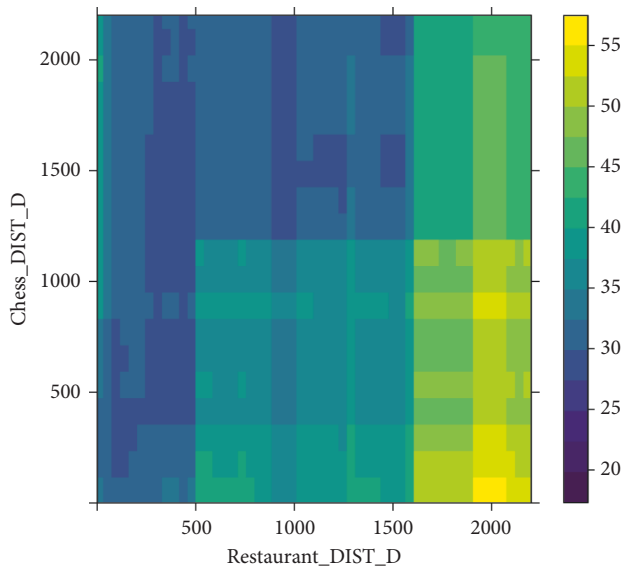


FIGURE 8: Interactive effects on trip time between distances to the nearest restaurant and chess/card room on the alighting side (H-statistic = 41.6%, the color bar stands for trip time).

room on the alighting side, for which the interaction accounts for 41.6% of the variance of the output of the partial dependence (as shown in Figure 8). A general upward trend is found in trip time with the increase of the distance to the nearest restaurant on the alighting side, which is consistent with the aforementioned findings. A downward trend can be observed in trip time with the increase of the distance to the nearest chess/card room on the alighting side. However, this

downward trend is moderated by the nearest restaurant on the alighting side. When the distance to the nearest restaurant on the alighting side grows from 0 to 2.2 km, the trip time reduction varies by the distance to the nearest restaurant on the alighting side. There is no observable change when the distance to the nearest restaurant on the alighting side is less than 500 m; then the trip time falls by about 8 minutes as the distance to the nearest restaurant moves to the range 500–1500 m. The trip time drops to the lowest point (about 15 minutes) when the distance to the nearest restaurant reaches 2 km. This variation may be attributable to two potential explanations. First, playing chess or card is usually an incidental leisure activity after dinner with friends. Second, the chess/card rooms are relative more familiar to those elderly people who live nearby, which is different from restaurants that can be easily found with web-based takeout meal services (e.g., the Meituan app).

In order to more intuitively observe the interactive effects of the explanatory variable on trip time, we took the interaction between the categorical variable and one of the other continuous variables as an example. Figure 9 displays how the location of boarding stations moderates the effect of the number of sports/gyms around the alighting station on trip time. When the original station is located in an exurban area, the increase in the number of sports/gyms around the alighting station, ranging from 5 to 30, has a salient influence on the trip time of the elderly, which is larger than the original station located in urban or suburban area. It suggests that the elderly living in the exurban area are more likely to take long-distance metro travel for their healthy lifestyle. This may be resulted from that the elderly people living in the city center are more likely to live

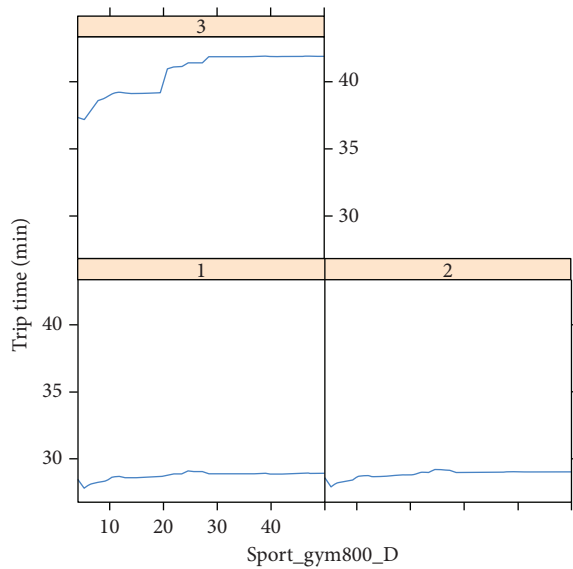


FIGURE 9: Interactive effects between sport/gyms density and the location of the boarding station on trip time (H-statistic = 26.9%).

with their children and have to take on more household responsibilities.

6. Conclusions

Taking Nanjing, China as a case study, this study explored the influence of built environment attributes, travel-related characteristics, and network attributes of metro stations on metro trip time of urban elderly and their nonlinear associations using smart card data. Some interesting findings of the paper can be summarized as follows. First, an elderly's acceptable longest metro trip time is 46.7 minutes; long-distance travel (more than 46.7 minutes) is the main barrier for the elderly to use the metro. However, about 7.5% of the elderly in Nanjing spend over 60 minutes on the metro currently, which is much longer than the acceptable longest trip time. Given that the average speed of the Nanjing metro is 35 km/h, some elderly metro users live about 4 km farther than the ideal trip time. This is also one of the reasons why the elderly use the metro less frequently (0.5 times per day on average). Therefore, land-use mixture planning and the last-mile issue should be paid more attention to, for example, bike-sharing or minibuses with well-designed facilities for cyclists and pedestrians. About 29.09% of the elderly use metro during the peak hours. This may aggravate the congestion of the metro during the peak commute period, especially in the early peak period. Since departure time of the elderly has little effect on their trip time based on our modeling results, the implementation of staggered shifts for the elderly can be encouraged.

Second, our results show that built environment attributes to play the most important role in predicting trip time, followed by network attributes of metro stations and travel-related characteristics. More importantly, all of these variables are associated with trip time in a nonlinear instead of a linear way. Therefore, the GBRT mode can provide better

and more reasonable results than the MLR model. In the model, the specific nonlinear patterns vary among variables. Distance to the nearest bike station, for example, has both positive and negative effects on trip time: first decreasing then increasing and then increasing again. The effect thresholds are 200 m, 1,000 m, and 1,500 m, respectively. The results of the analysis suggest that if we want to decrease the elderly's trip time for those who have to travel by metro, relevant policy implementation should be carried out in conjunction with the elderly's travel environment by regarding their threshold effect. For instance, adding shared bike stations and related services within 200 m and 1,000 m to 1,500 m around the metro stations. Due to the similar thresholds of distance to the nearest bus stop, planners can also add bus stops and routes within 200 m around the metro station to decrease the elderly's trip time. The above-mentioned measures can not only increase the accessibility of the station but also increase the betweenness centrality of the station, to increase elderly mobility in Nanjing. In addition, since shopping, dining, and playing chess or card are the main travel purpose of elderly metro users, the corresponding facilities should also be laid out according to the impact threshold analyzed in Section 5.3.1.

Third, the nonlinear interaction effects of explanatory variables on trip time have been identified in our study. By using the H-statistic method, we found that distance to the nearest chess/card room on the alighting station has the strongest interaction with other explanatory variables, in which a high relative interaction effect on trip time is found between distances to the nearest restaurant and chess/card room on the alighting station. Besides, we also found that the number of sports/gyms near alighting stations in urban core areas, suburbs, and rural areas has different effects on trip time. The elderly living in the exurban area is more likely to take long-distance metro travel for their physical exercise. These findings demonstrate that planners and policymakers should not only consider one single factor, but also the interactions of various factors when planning or making policies. For example, a chess/card room could be provided for elderly people near restaurants around the metro.

Although this study provides some interesting findings and new implications with a large set of smart card data, it still can be further extended from the following aspects. First, the smart card data lack some important information, such as the socio-demographic attributes of elderly people. Future studies could supplement the survey data or phone signaling data with smart card data to implement a complementary analysis. Second, the spatial attributes of different anchor points (e.g. home, recreation) have been proved to have different effects on activity-based travel behavior. To investigate the effects of the exact spatial attributes of stations, a panel analysis with longitudinal data should be conducted to identify the role of metro stations that play in the trip chain. Third, we only developed the GBRT model and the baseline MLR model, which could just offer insight into the relevance across multiple factors, but not the causal relationships. We still do not exactly understand what causal link results in such a significant nonlinear correlation between the built environment around

metro stations and travel distance. This may influence the performance of the models in another context. Similar studies based on other methods such as path analysis and structural equation model with comprehensive data, including the residential location, destination location, trip purpose, etc., can be an important supplement. It will help to explore the logical explanation of how the built environment attributes influence elderly travel distance with metro, and provide valuable planning suggestions for improving the quality of life and travel equity for the elderly. Moreover, the large number of floating populations led to wide variations in the aging stages between Chinese cities/provinces. The findings and the policy implications in this study may be not transferrable to a different context but nonetheless, it would be helpful for practitioners to extend this modeling approach to other cities, as well as other age groups.

Data Availability

The smart card data used to support the findings of this study were supplied by Nanjing Metro under license and so cannot be made freely available.

Conflicts of Interest

No potential conflicts of interest were reported by the authors.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 52102378 and 71861017), Yunnan Provincial Department of Education Science Research Fund (Grant nos. 2022J0058 and 2022J0075), and Yunnan Fundamental Research Projects (Grant nos. 202201AU070148, 202201AU070109, and 202001AT070030).

References

- [1] D. Banister and A. Bowling, "Quality of life for the elderly: the transport dimension," *Transport Policy*, vol. 11, no. 2, pp. 105–115, 2004.
- [2] L. Bocker, P. van Amen, and M. Helbich, "Elderly travel frequencies and transport mode choices in Greater Rotterdam, The Netherlands," *Transportation*, vol. 44, no. 4, pp. 831–852, 2017.
- [3] 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.
- [4] P. P. Koh, B. W. Leow, and Y. D. Wong, "Mobility of the elderly in densely populated neighbourhoods in Singapore," *Sustainable Cities and Society*, vol. 14, pp. 126–132, 2015.
- [5] Z. Liu, A. Kemperman, and H. Timmermans, "Social-ecological correlates of older adults' outdoor activity patterns," *Journal of Transport & Health*, vol. 16, p. 100840, 2020.
- [6] P. Pettersson and J.-D. Schmöcker, "Active ageing in developing countries? – trip generation and tour complexity of older people in Metro Manila," *Journal of Transport Geography*, vol. 18, no. 5, pp. 613–623, 2010.
- [7] W. Y. Szeto, L. Yang, R. C. P. Wong, Y. C. Li, and S. C. Wong, *Spatio-temporal travel characteristics of the elderly in an ageing society*, *Travel Behaviour and Society*, vol. 9, pp. 10–20, 2017.
- [8] L. T. Truong and S. V. C. Somenahalli, "Exploring frequency of public transport use among older adults: a study in Adelaide, Australia," *Travel Behaviour and Society*, vol. 2, no. 3, pp. 148–155, 2015.
- [9] Q. Wang and L. Wang, "The nonlinear effects of population aging, industrial structure, and urbanization on carbon emissions: a panel threshold regression analysis of 137 countries," *Journal of Cleaner Production*, vol. 287, p. 125381, 2021.
- [10] T. Schwanen and A. Paez, "The mobility of older people - an introduction," *Journal of Transport Geography*, vol. 18, no. 5, pp. 591–595, 2010.
- [11] X. Hu, J. Wang, and L. Wang, "Understanding the Travel Behavior of Elderly People in the Developing Country: A Case Study of Changchun, China," *Procedia - Social Behavioral Sciences*, vol. 96, pp. 873–880, 2013.
- [12] M. Šimeček, V. Gabrhel, M. Tögel, and M. Lazor, "Travel behaviour of seniors in eastern Europe: a comparative study of brno and bratislava," *European Transport Research Review*, vol. 10, no. 1, 2018.
- [13] DiDi Company, "Investigation Report on the Travel Habits of the Chinese Elderly: Bus Is the First Choice for the Elderly," 2016, https://www.sohu.com/a/63211175_115412.
- [14] R. Wong, W. Szeto, L. Yang, Y. C. Li, and S. Wong, "Elderly users' level of satisfaction with public transport services in a high-density and transit-oriented city," *Journal of Transport & Health*, vol. 7, pp. 209–217, 2017.
- [15] C. Morency, A. Paez, M. J. Roorda, R. Mercado, and S. Farber, "Distance traveled in three Canadian cities: spatial analysis from the perspective of vulnerable population segments," *Journal of Transport Geography*, vol. 19, no. 1, pp. 39–50, 2011.
- [16] E. J. Shin, "Exploring the causal impact of transit fare exemptions on older adults' travel behavior: evidence from the Seoul metropolitan area," *Transportation Research Part A: Policy and Practice*, vol. 149, pp. 319–338, 2021.
- [17] J. Kandt and A. Leak, "Examining inclusive mobility through smartcard data: what shall we make of senior citizens' declining bus patronage in the West Midlands?" *Journal of Transport Geography*, vol. 79, p. 102474, 2019.
- [18] T. Tao, J. Wang, and X. Cao, "Exploring the non-linear associations between spatial attributes and walking distance to transit," *Journal of Transport Geography*, vol. 82, p. 102560, 2020.
- [19] Z. Gan, M. Yang, T. Feng, and H. J. Timmermans, "Examining the relationship between built environment and metro ridership at station-to-station level," *Transportation Research Part D: Transport and Environment*, vol. 82, p. 102332, 2020.
- [20] J. De Vos, D. Ettema, and F. Witlox, "Changing travel behaviour and attitudes following a residential relocation," *Journal of Transport Geography*, vol. 73, pp. 131–147, 2018.
- [21] J. X. Feng, "The influence of built environment on travel behavior of the elderly in urban China," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 619–633, 2017.
- [22] L. Frändberg and B. Vilhelmson, "More or less travel: personal mobility trends in the Swedish population focusing gender and cohort," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1235–1244, 2011.
- [23] J. De Vos and F. Witlox, "Travel satisfaction revisited. On the pivotal role of travel satisfaction in conceptualising a travel

- behaviour process," *Transportation Research Part A: Policy and Practice*, vol. 106, pp. 364–373, 2017.
- [24] R. Mercado and A. Páez, "Determinants of distance traveled with a focus on the elderly: a multilevel analysis in the Hamilton CMA, Canada," *Journal of Transport Geography*, vol. 17, no. 1, pp. 65–76, 2009.
- [25] F. Shao, Y. Sui, X. Yu, and R. Sun, "Spatio-temporal travel patterns of elderly people—A comparative study based on buses usage in Qingdao, China," *Journal of Transport Geography*, vol. 76, pp. 178–190, 2019.
- [26] Y. Liu, Y. Ji, T. Feng, and H. Timmermans, "Understanding the determinants of young commuters' metro-bikeshare usage frequency using big data," *Travel behaviour and society*, vol. 21, pp. 121–130, 2020.
- [27] S. Y. He, Y. H. Y. Cheung, and S. Tao, "Travel mobility and social participation among older people in a transit metropolis: a socio-spatial-temporal perspective," *Transportation Research Part A: Policy and Practice*, vol. 118, pp. 608–626, 2018.
- [28] J.-D. Schmöcker, M. A. Quddus, R. B. Noland, and M. G. Bell, "Mode choice of older and disabled people: a case study of shopping trips in London," *Journal of Transport Geography*, vol. 16, no. 4, pp. 257–267, 2008.
- [29] D. Yang, H. Timmermans, and A. Grigolon, "Exploring heterogeneity in travel time expenditure of aging populations in The Netherlands: results of a CHAID analysis," *Journal of Transport Geography*, vol. 33, pp. 170–179, 2013.
- [30] S. Rosenbloom and C. Winsten-Bartlett, "Asking the right question: understanding the travel needs of older women who do not drive," *Transportation Research Record*, vol. 1818, no. 1, pp. 78–82, 2002.
- [31] S. Kim, "Assessing mobility in an aging society: personal and built environment factors associated with older people's subjective transportation deficiency in the US," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 5, pp. 422–429, 2011.
- [32] R. Alsnih and D. A. Hensher, "The mobility and accessibility expectations of seniors in an aging population," *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, pp. 903–916, 2003.
- [33] J.-D. Schmöcker, M. A. Quddus, R. B. Noland, and M. G. H. Bell, "Estimating trip generation of elderly and disabled people: analysis of London data," *Transportation Research Record*, vol. 1924, no. 1, pp. 9–18, 2005.
- [34] G. Currie and A. Delbosc, "Exploring public transport usage trends in an ageing population," *Transportation*, vol. 37, no. 1, pp. 151–164, 2010.
- [35] M. Tacken, "Mobility of the elderly in time and space in The Netherlands: an analysis of the Dutch National Travel Survey," *Transportation*, vol. 25, no. 4, pp. 379–393, 1998.
- [36] T. Rye and M. Carreno, "Concessionary fares and bus operator reimbursement in Scotland and Wales: No better or no worse off?" *Transport Policy*, vol. 15, no. 4, pp. 242–250, 2008.
- [37] R. Mackett, "Has the policy of concessionary bus travel for older people in Britain been successful?" *Case studies on transport policy*, vol. 2, no. 2, pp. 81–88, 2014.
- [38] A. M. Ricciardi, J. Xia, and G. Currie, "Exploring public transport equity between separate disadvantaged cohorts: a case study in Perth, Australia," *Journal of Transport Geography*, vol. 43, pp. 111–122, 2015.
- [39] Y. Yang, Y. Xu, D. A. Rodriguez, Y. Michael, and H. Zhang, "Active travel, public transportation use, and daily transport among older adults: the association of built environment," *Journal of Transport & Health*, vol. 9, pp. 288–298, 2018.
- [40] S. Mah and R. Mitra, "The effects of a free bus program on older adults travel behaviour: a case study of a Canadian suburban municipality," *Case Studies on Transport Policy*, vol. 5, no. 3, pp. 460–466, 2017.
- [41] R. Wong, W. Szeto, L. Yang, Y. Li, and S. Wong, "Public transport policy measures for improving elderly mobility," *Transport Policy*, vol. 63, pp. 73–79, 2018.
- [42] Y. Zhang, E. Yao, R. Zhang, and H. Xu, "Analysis of elderly people's travel behaviours during the morning peak hours in the context of the free bus programme in Beijing, China," *Journal of Transport Geography*, vol. 76, pp. 191–199, 2019.
- [43] M. Du, L. Cheng, X. Li, and J. Yang, "Factors affecting the travel mode choice of the urban elderly in healthcare activity: comparison between core area and suburban area," *Sustainable Cities and Society*, vol. 52, p. 101868, 2020.
- [44] C. Liu, E. Bardaka, R. Palakurthy, and L.-W. Tung, "Analysis of travel characteristics and access mode choice of elderly urban rail riders in Denver, Colorado," *Travel Behaviour and Society*, vol. 19, pp. 194–206, 2020.
- [45] Y. Zhang, Y. Li, Q. X. Liu, and C. Y. Li, "The built environment and walking activity of the elderly: an empirical analysis in the zhongshan metropolitan area, China," *Sustainability*, vol. 6, no. 2, pp. 1076–1092, 2014.
- [46] G. A. Jordan, P. C. Anastasopoulos, S. Peeta, S. Somenahalli, and P. A. Rogerson, "Identifying elderly travel time disparities using a correlated grouped random parameters hazard-based duration approach," *Research in Transportation Business & Management*, vol. 30, p. 100369, 2019.
- [47] M. Manoj and A. Verma, "Effect of built environment measures on trip distance and mode choice decision of non-workers from a city of a developing country, India," *Transportation Research Part D: Transport and Environment*, vol. 46, pp. 351–364, 2016.
- [48] D. Wang, Y. Chai, and F. Li, "Built environment diversities and activity-travel behaviour variations in Beijing, China," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1173–1186, 2011.
- [49] J. Hong, Q. Shen, and L. Zhang, "How do built-environment factors affect travel behavior? A spatial analysis at different geographic scales," *Transportation*, vol. 41, no. 3, pp. 419–440, 2014.
- [50] M. J. Figueroa, T. A. S. Nielsen, and A. Siren, "Comparing urban form correlations of the travel patterns of older and younger adults," *Transport Policy*, vol. 35, pp. 10–20, 2014.
- [51] L. F. Gómez, D. C. Parra, D. Buchner et al., "Built environment attributes and walking patterns among the elderly population in bogotá," *American Journal of Preventive Medicine*, vol. 38, no. 6, pp. 592–599, 2010.
- [52] M. Shirgaokar, B. Dobbs, L. Anderson, and E. Hussey, "Do rural older adults take fewer trips than their urban counterparts for lack of a ride?" *Journal of Transport Geography*, vol. 87, p. 102819, 2020.
- [53] L. Cheng, J. De Vos, K. Shi, M. Yang, X. Chen, and F. Witlox, "Do residential location effects on travel behavior differ between the elderly and younger adults?" *Transportation Research Part D: Transport and Environment*, vol. 73, pp. 367–380, 2019.
- [54] Y. J. Ji, Y. L. Fan, A. Ermagun, X. Cao, W. Wang, and K. Das, "Public bicycle as a feeder mode to rail transit in China: the role of gender, age, income, trip purpose, and bicycle theft experience," *International Journal of Sustainable Transportation*, vol. 11, no. 4, pp. 308–317, 2017.
- [55] A. Kemperman and H. Timmermans, "Influences of built environment on walking and cycling by latent segments of

- aging population,” *Transportation Research Record*, vol. 2134, pp. 1–9, 2009.
- [56] X. Ma, Y. Ji, M. Yang, Y. Jin, and X. Tan, “Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data,” *Transport Policy*, vol. 71, pp. 57–69, 2018.
- [57] Q. Shao, W. Zhang, X. Cao, J. Yang, and J. Yin, “Threshold and moderating effects of land use on metro ridership in Shenzhen: implications for TOD planning,” *Journal of Transport Geography*, vol. 89, p. 102878, 2020.
- [58] J. Yang, P. Su, and J. Cao, “On the importance of Shenzhen metro transit to land development and threshold effect,” *Transport Policy*, vol. 99, pp. 1–11, 2020.
- [59] Z. Li, “Unobserved and observed heterogeneity in risk attitudes: implications for valuing travel time savings and travel time variability,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 112, pp. 12–18, 2018.
- [60] C. R. Bhat, “The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions,” *Transportation Research Part B: Methodological*, vol. 42, no. 3, pp. 274–303, 2008.
- [61] H.-J. Eom and G.-H. Cho, “Exploring thresholds of built environment characteristics for walkable communities: empirical evidence from the Seoul Metropolitan area,” *Transportation Research Part D: Transport and Environment*, vol. 40, pp. 76–86, 2015.
- [62] C. Ding, X. Cao, and C. Liu, “How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds,” *Journal of Transport Geography*, vol. 77, pp. 70–78, 2019.
- [63] C. Ding, X. J. Cao, and P. Næss, “Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo,” *Transportation Research Part A: Policy and Practice*, vol. 110, pp. 107–117, 2018.
- [64] L. Cheng, J. De 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, p. 102552, 2020.
- [65] Nanjing Statistics Bureau, “2019 Nanjing National Economic and Social Development Statistics Bulletin,” 2019.
- [66] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [67] E. Chen and Z. Ye, “Identifying the nonlinear relationship between free-floating bike sharing usage and built environment,” *Journal of Cleaner Production*, vol. 280, p. 124281, 2021.
- [68] T. A. Craney and J. G. Surles, “Model-dependent variance inflation factor cutoff values,” *Quality Engineering*, vol. 14, no. 3, pp. 391–403, 2002.
- [69] J. Elith, J. R. Leathwick, and T. Hastie, “A working guide to boosted regression trees,” *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 2008.
- [70] J. H. Friedman and B. E. Popescu, “Predictive learning via rule ensembles,” *Annals of Applied Statistics*, vol. 2, no. 3, pp. 916–954, 2008.