

## Research Article

# Examining Built Environment Effects on Metro Ridership at Station-to-Station Level considering Circle Heterogeneity: A Case Study from Xi'an, China

Siyi Zhang <sup>1</sup>, Ziwei Li <sup>2</sup>, and Zixuan Liu <sup>2</sup>

<sup>1</sup>College of Transportation Engineering, Chang'an University, Xi'an 710064, Shaanxi, China

<sup>2</sup>College of Architecture, Chang'an University, Xi'an 710064, Shaanxi, China

Correspondence should be addressed to Zixuan Liu; [jzxy\\_liuwx@chd.edu.cn](mailto:jzxy_liuwx@chd.edu.cn)

Received 22 February 2023; Revised 25 August 2023; Accepted 9 October 2023; Published 23 October 2023

Academic Editor: Ren-Yong Guo

Copyright © 2023 Siyi Zhang 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.

Transit-oriented development is described as a geographic unit with multicircle structures. Most studies have analysed the impact of the built environment within station catchment areas on metro passenger flows from a macro perspective and have lacked analysis of the circle heterogeneity. Few relevant studies have independently investigated the impact of the built environment on the passenger flow in each circle and indeed neglected the systematic interaction between inner circles and circles in the TOD area. In this study, the 800 m buffer from the station was equally divided into four circles. Based on the gravity model, the representative built environment features around the metro stations on both sides were extracted using the block attention module (BAM). Subsequently, Shapley Additive exPlanation (SHAP) was used to explore the influence of different built environment variables on passenger flow at each circle between the origin and destination stations. The results indicate the following: (1) the station-to-station passenger flow is significantly affected by the availability of transfers and the distance between the origin and destination stations; (2) the impact of different built environments on ridership significantly varies within different circles; and (3) the built environment has a similar impact on average daily passenger flow on both sides. Therefore, this study proposes strategies to optimize the metro passenger flow by developing different land use in different circles and updating the urban spatial structure.

## 1. Introduction

The metro system has been prioritized in China to address the problems of traffic congestion and environmental pollution owing to automobile-based transportation. Between 2015 and 2021, the number of cities with metros in China increased from 26 to 49, and the total length of the network exceeded 9000 km. Despite the impact of the anti-epidemic policy, the average daily traffic intensity still reached 4800 persons/km. Meanwhile, transit-oriented development (TOD), which integrates the metro system and land use development, was applied in many cities. Many studies in this area focused on the relationship between the built environment and metro ridership at the station or station-to-station levels [1–3].

Since its inception, the TOD catchment area has been considered as a geographical unit with a multicircle structure to enhance the use of public transport [4]. The previous literature has highlighted the significant important roles of density, diversity, and nonmotorized friendly design on metro ridership [5, 6]. However, one critical question that has surfaced is how the land use layout within the catchment area affects ridership. Exploring the impact of different land use in different circles on ridership is key to addressing the question. Furthermore, while some studies about the delineation of TOD catchment area have involved exploring the circle heterogeneity of the built environment's influence on passenger flows, there are two gaps. On the one hand, the studies generally developed multiple models to explore the impact of the built environment of different circles on

passenger flow, respectively, ignoring the results of the mutual influence between different circles in the TOD area as an independent geographical unit. On the other hand, due to the black box nature of deep learning models, most of the studies have had to abandon their predictive power and revert to traditional statistical or machine learning models, which offer better explanatory capabilities but poorer predictive performance.

Against this background, this study aims to investigate the circle heterogeneity of the impact of different land use types on station-to-station passenger flow, considering the built environment factors of the station catchment area. To achieve this aim, the convolutional block attention module was selected due to its requirement for less sample data while ensuring prediction accuracy [7]. This type of attention mechanism was improved to extract the combined characteristics of the originating and destination stations, respectively, and modelled the average weekday metro passenger flow within the framework of a gravity model. Shapley Additive exPlanation (SHAP) was used to interpret the model results.

The main contributions of this study include the following aspects. First, this study strengthens the analysis of the impact of built environment factors on ridership at the station-to-station level and provides limited experience for deep learning models to analyze and interpret rather than just accurately predict travel behavior. Second, in view of the TOD multicircle structure, the circle heterogeneity of the impact of the built environment on ridership is studied to provide more feasible suggestions for planners. The remainder of this paper is organized as follows. The next section reviews the relevant literature. Section 3 presents an overview of the study area and data collection. The framework of the model is shown in Section 4. Section 5 presents and discusses the results. Finally, Section 6 provides the major conclusions and proposes potential applications of the study.

## 2. Literature Review

In the last three decades, TOD has become a focus in the field of urban planning and transportation [8, 9]. A better understanding of the impacts of station-area built environment factors on transit ridership can improve transit performance and inform land use within station catchment planning.

In terms of land use, almost all the existing studies have highlighted the influence of population [6, 10–12] and employment densities [11, 13, 14] on ridership. The findings on the relationship between the diversity and metro ridership are not consistent. The entropy index, a measure of diversity, positively influences metro ridership [15, 16]. By contrast, Cervero [17] found that the land use mix has no evident impact on metro passenger flow. Several TOD-related studies also focused on land use variables. Many scholars investigated the effects of the difference in land use or the different types of POIs on metro ridership [11, 18–24]. For example, An et al. [25] suggested that the commercial building is the most critical factor for the metro ridership prediction and the effect of residential factors was

inconclusive. Li et al. [26] concluded that only common residences effectively improve metro ridership and suggested that the residential and scenic spots have nontrivial effects on ridership.

In terms of transit service, variables such as road density, intersection density, number of bus stops (or lines), terminal station, and transfer station that are associated with ridership were explored. Some studies highlighted the influence of intersection density on ridership [13, 26]. However, the negative effect of intersection density was reported in another study [25, 27]. Some previous studies also concluded that the number of bus stops around the metro station could influence the metro passenger flow because the bus is one of the primary egress and access modes to metro stations [28–30]. E. Chen et al. [31] concluded that the effects of the built environment around the terminal or adjacent stations are more significant than those around the normal station.

Research on the station-to-station level can be considered an extension of that on the station level. The mentioned explanatory variables were calculated for both the origin and destination. Moreover, the station-to-station level ridership is affected by traffic impedance factors, including transfer times, detours, and route distance [32–36].

Most TOD development guides and related studies described the TOD as an area with a multicircle spatial structure (shown in Figure 1) and emphasized the development of differences according to the different circles of TOD. Several studies have divided the station catchment into multiple buffer bands and modelled the impact of the built environment on passenger flows within each buffer band. Gutiérrez et al. [37] developed distance-decay weighted regression using the mobility survey to forecast the Madrid metro ridership. They found that the effect of built environment factors on ridership changed with different buffers. Similar experiments conducted by Manout et al. [38] in Lyon also yielded the same conclusion. The built environment factors in the buffer zones of 0–300 m, 300–600 m, and 600–900 m were counted by Jun et al. [15], and the geographically weighted regression models were calibrated to explore the impact of land use characteristics in these buffer zones on ridership at the station level. The results showed that only the population and land use mix significantly affect the ridership in the 0–300 m and 300–600 m buffers. Pan et al. [21] conducted a questionnaire survey of 33 sites and 11 neighborhood units in Beijing and collected 300 responses to analyze the influence of shopping facilities within seven circles on the willingness of residents to shop around the nearest metro station. The results from the model revealed that the influence of different shopping facilities on residents' shopping trip willingness varies with different circles in three aspects: significance, sign, and value.

Regarding the models that examine the relationship with ridership, the direct ridership models (DRMs) are popular owing to the minimal data requirement and easy application. Initially, the relationship between the built environment and metro ridership was assumed to be linear or log-linear [15, 37, 39]. With the development of data mining methods, the traditional linear model has gradually been supplemented by machine learning models. These models

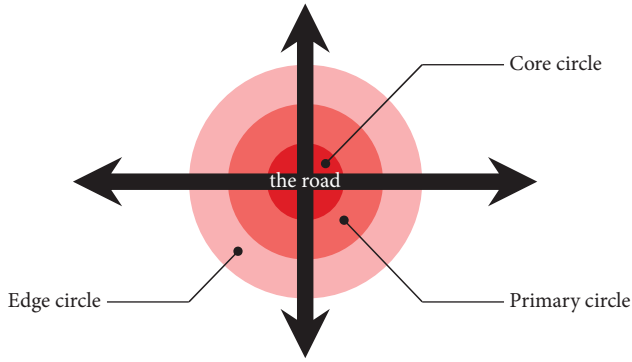


FIGURE 1: The TOD multicircle spatial structure.

have better predictive and explanatory capabilities. For example, tree-based models have been developed in related studies because they are more adept at dealing with nonlinear relationships [28, 34, 40, 41]. However, successful machine learning models require extensive expertise in capturing highly accurate features whenever possible. For metro ridership, it is difficult to achieve high-quality feature engineering manually. Thus, deep learning models are widely used in transportation because of their excellent performance in feature extraction. For example, recurrent neural network (RNN) models, which are suitable for modelling dynamic temporal dependency occurring in time series, were widely used to predict ridership in isolated metro stations or lines [42, 43]. As convolutional kernels of different sizes can extract spatial dependencies of features by automatically learning from the data, convolutional neural networks are often used for ridership prediction at large spatial scales [44].

Although previous studies explored the relationship between built environment factors and ridership at the station or station-to-station level, they have some limitations. The influence of various built environment factors in the station catchment area on passenger flow is not independent, and it is the result of cross-influence within the TOD area as a system of independent geographical units. In other words, the spatial structure and land use layout of the catchment area also significantly impact passenger flows. Furthermore, the results of studies that focus only on global built environment factors are undoubtedly biased. However, although deep learning models performed well in extracting the implicit features of these interactions, they are rarely used to explain the passenger flow correlation owing to their black box. Modelling the effective extraction of features and their rational interpretation is extremely important for TOD planning and construction practice.

Overall, this study aims to fill the gaps in the literature by (1) quantifying the circle heterogeneity of the built environment's nonlinear effect on ridership and (2) using deep learning models to capture potential land use features within the station catchment and explaining the model results.

### 3. Dataset Description

As shown in Figure 2, the Xi'an metropolitan area was selected for the study. From 2011 to 2019, four metro lines and 57 stations were built and operated in Xi'an, making the

city rank eighth among 40 Chinese cities with metro lines. A total of 3192 pairs of the average daily station-to-station passenger flow on the weekday during November 2019 were counted using Auto Fare Collection (AFC) data. Although the number of metro users is staggering, the development of the metro system is still confronted by the uneven distribution of passenger flow. In particular, the Zhonglou-Xiaozhai route has the largest average daily passenger flow with 2836 riders, whereas the passenger flow for Xinjiamiao-Daminggongbei route is one rider. To optimize travel demand management, it is important to study the influencing factors of OD passenger flow.

It should be noted that this study considered the land use factors and transit service within the 800 m buffers of both origin and destination stations. Moreover, considering the average scale of plots in Xi'an and the sample size of the dataset, the 800 m buffer was divided into four circles: 0–200 m, 200–400 m, 400–600 m, and 600–800 m. The number of bus stops (S) within the four circles between the origin and destination stations was counted as the transit service factor. The areas of the administration buildings (A), residential buildings (R), and commercial buildings (B) within the different circles were calculated to represent the land use factors of the origin and destination sides. The parks, squares, and scenic spots were classified into one category (G) to calculate the land use area. Some other studies also calculated building areas according to industrial and warehousing logistics [23]. However, most of the two types of land use are located outside the study area.

To preserve the location information of different circles, a matrix was constructed to represent the five features of the four circles, considering the built environment of the origin or destination stations (shown in Figure 3). In addition, the number of transfer times and interval (the number of stops) between the origin and destination stations was used to determine the impact of travel impedance factors on the metro ridership. Table 1 describes the details and source of the variables.

### 4. Method

As shown in equation (1), the original gravity model is expressed in the multiplicative form to examine the relationship with station-to-station ridership.

$$T_{OD} = \partial \cdot O \cdot D \cdot C, \\ \ln(T_{OD}) = \partial + \partial_O \cdot \ln(O) + \partial_D \cdot \ln(D) + \partial_C \cdot \ln(C), \quad (1)$$

$$O, D, C = \prod_{i=1}^I X_{O_i}^{\partial_O}, \prod_{i=1}^I X_{D_i}^{\partial_D}, \prod_{i=1}^I X_{C_i}^{\partial_C},$$

where  $T_{OD}$  denotes the station-to-station ridership.  $X_{O_i}$ ,  $X_{D_i}$ , and  $X_{C_i}$  represent the  $i$ th independent variable of the station  $O$ ,  $D$ , or the traffic impedance  $C$ , respectively.  $\partial$ ,  $\partial_O$ ,  $\partial_D$ , and  $\partial_C$  are coefficients that could be estimated via logarithmic transformation.

Different from the original gravity model, the block attention module-gravity (BAM-Gravity) model proposed in this study uses the block attention module to extract built

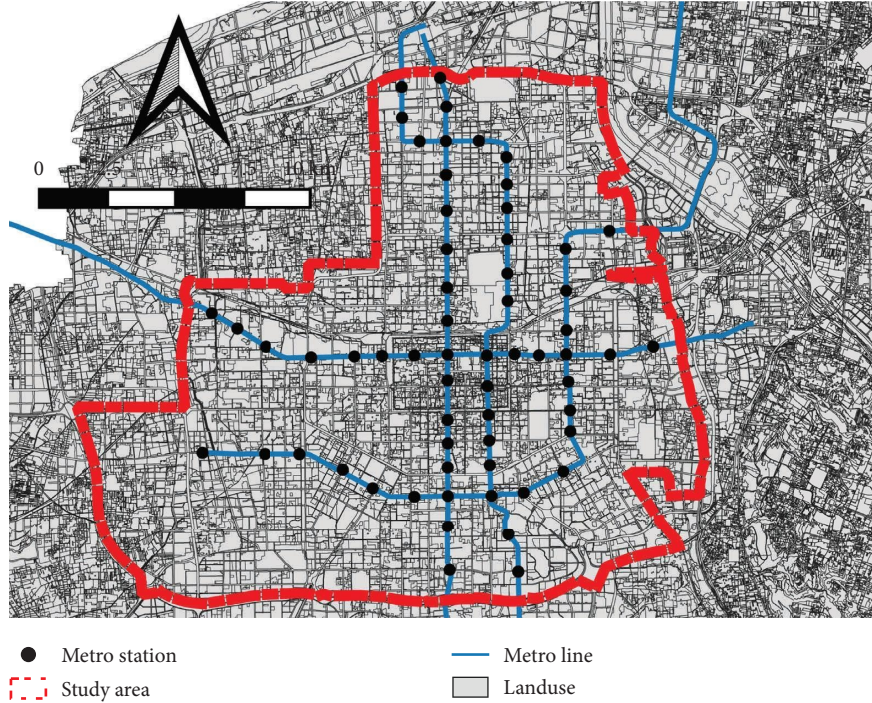


FIGURE 2: Metro lines and stations in the study area.

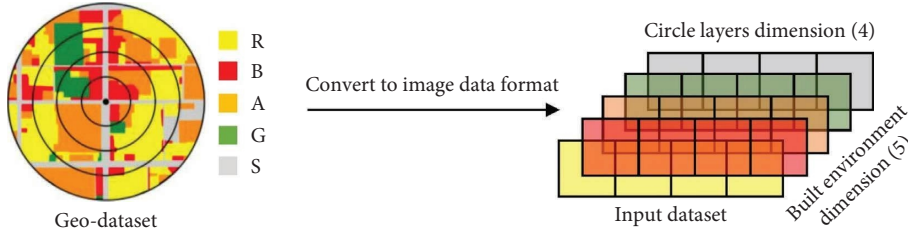


FIGURE 3: Illustration of data processing.

environment features around the stations based on the multiplicative form of the original gravity model. As shown in Figure 4, the overall structure of the BAM-Gravity model can be divided into three stages: (1) block attention module stage, (2) fully connected stage for extracting the built environment representing the metro station, and (3) original gravity model stage for examining the relationship with the station-to-station level ridership.

As shown in Figure 5 and equation (2), along the circle layer dimension, the independent variables of the origin/destination station  $I^{5 \times 4}$  are forwarded to the built environment dimension neural network  $NN_b$  to produce the built environment attention weights  $W_b^{5 \times 1}$ . The built environment attention refined data  $B^{5 \times 4}$  are generated by  $W_b^{5 \times 1} \otimes I^{5 \times 4}$ . Subsequently, another neural network  $NN_c$  is connected along the built environment dimension of  $B^{5 \times 4}$  to compute the circle attention weights  $W_c^{4 \times 1}$ . Finally, the attention refines data  $R^{5 \times 4}$  are defined as  $W_c^{4 \times 1} \otimes B^{5 \times 4}$ .

$$R = NN_c(NN_b(I) \otimes I) \otimes (NN_b(I) \otimes I), \quad (2)$$

where  $R$  denotes the attention refined variables,  $I$  is the input dataset of the metro stations, and  $\otimes$  denotes element-wise multiplication. The structures of  $NN_c(\bullet)$  and  $NN_b(\bullet)$  are shown in Figure 5.

The independent variables of the origin and destination station are refined by the attention mechanism and then connected to the two fully connected layers ( $FCN_O$ ,  $FCN_D$ ), separately. After another fully connected layer ( $FCN_C$ ), the two traffic impedance variables become a one-dimension indicator representing traffic impedance. The overall BAM-Gravity model can be summarized as follows:

$$\begin{aligned} \ln(T_{OD}) &= \partial + \partial_O \bullet \ln(O) + \partial_D \bullet \ln(D) + \partial_C \bullet \ln(C) \\ &= FCN_O(A_O(I_O)), FCN_D(A_D(I_D)), FCN_C(I_C), \end{aligned} \quad (3)$$

where  $T_{OD}$  and  $I_C$  denote the ridership and the traffic impedance variables from the origin station to the destination station, respectively.  $I_O$  and  $I_D$  represent the built environment variables of the stations on both sides.  $A_O(\bullet)$

TABLE 1: Description of the independent and dependent variables.

| Variables*            | Mean                                                                                                                                                                                                                                                                                                                                            | S.D       | Min       | Max        |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|-----------|------------|
| $R$ (m <sup>2</sup> ) | The area of the residential buildings, including villas, residences, and apartments, is calculated from the third national land use survey data provided by Xi'an Planning Department in 2020                                                                                                                                                   |           |           |            |
| $R_{O/D}^1$           | 114620.21                                                                                                                                                                                                                                                                                                                                       | 99915.72  | 0.00      | 546399.12  |
| $R_{O/D}^2$           | 443583.68                                                                                                                                                                                                                                                                                                                                       | 217174.80 | 69839.74  | 1002606.26 |
| $R_{O/D}^3$           | 771613.44                                                                                                                                                                                                                                                                                                                                       | 343141.17 | 231646.25 | 1977813.50 |
| $R_{O/D}^4$           | 1067472.38                                                                                                                                                                                                                                                                                                                                      | 393039.58 | 381132.04 | 2212369.41 |
| $B$ (m <sup>2</sup> ) | The floor area of the commercial/business buildings, including business and financial buildings, commercial buildings, and entertainment buildings, is calculated from the third national land use survey data provided by Xi'an Planning Department in 2020                                                                                    |           |           |            |
| $B_{O/D}^1$           | 87132.59                                                                                                                                                                                                                                                                                                                                        | 89982.64  | 0.00      | 344311.12  |
| $B_{O/D}^2$           | 146724.61                                                                                                                                                                                                                                                                                                                                       | 143021.37 | 0.00      | 742934.06  |
| $B_{O/D}^3$           | 163076.45                                                                                                                                                                                                                                                                                                                                       | 166621.83 | 0.00      | 770465.13  |
| $B_{O/D}^4$           | 199881.19                                                                                                                                                                                                                                                                                                                                       | 172886.75 | 1516.81   | 803319.89  |
| $A$ (m <sup>2</sup> ) | The area of the administration buildings, including administrative office buildings, scientific research office buildings, schools, hospitals, sports facilities, libraries, art exhibition halls, and other social public facilities, is calculated from the third national land use survey data provided by Xi'an Planning Department in 2020 |           |           |            |
| $A_{O/D}^1$           | 19701.96                                                                                                                                                                                                                                                                                                                                        | 26427.84  | 0.00      | 133262.08  |
| $A_{O/D}^2$           | 62220.86                                                                                                                                                                                                                                                                                                                                        | 70693.21  | 0.00      | 347153.65  |
| $A_{O/D}^3$           | 95564.15                                                                                                                                                                                                                                                                                                                                        | 87008.39  | 0.00      | 360969.24  |
| $A_{O/D}^4$           | 138564.22                                                                                                                                                                                                                                                                                                                                       | 106691.86 | 1303.81   | 512148.79  |
| $G$ (m <sup>2</sup> ) | The land use area of parks/squares/scenic spots is calculated from the third national land use survey data provided by Xi'an Planning Department in 2020                                                                                                                                                                                        |           |           |            |
| $G_{O/D}^1$           | 12173.56                                                                                                                                                                                                                                                                                                                                        | 17420.53  | 0.00      | 81784.07   |
| $G_{O/D}^2$           | 36089.56                                                                                                                                                                                                                                                                                                                                        | 44896.74  | 0.00      | 177826.38  |
| $G_{O/D}^3$           | 57921.12                                                                                                                                                                                                                                                                                                                                        | 69070.18  | 0.00      | 281231.12  |
| $G_{O/D}^4$           | 80201.81                                                                                                                                                                                                                                                                                                                                        | 90885.71  | 0.00      | 420670.35  |
| $S$                   | The number of bus stops is calculated from the POI data in 2020                                                                                                                                                                                                                                                                                 |           |           |            |
| $S_{O/D}^1$           | 1.37                                                                                                                                                                                                                                                                                                                                            | 1.15      | 0.00      | 5.00       |
| $S_{O/D}^2$           | 1.93                                                                                                                                                                                                                                                                                                                                            | 1.31      | 0.00      | 6.00       |
| $S_{O/D}^3$           | 2.79                                                                                                                                                                                                                                                                                                                                            | 1.90      | 0.00      | 7.00       |
| $S_{O/D}^4$           | 4.11                                                                                                                                                                                                                                                                                                                                            | 2.29      | 0.00      | 12.00      |
| Travel impedance      | The number of transfer times and interval (the number of stops) between the origin station and the destination station is calculated by ArcGIS                                                                                                                                                                                                  |           |           |            |
| Transfer time         | 0.89                                                                                                                                                                                                                                                                                                                                            | 0.69      | 0.00      | 2.00       |
| Interval              | 7.60                                                                                                                                                                                                                                                                                                                                            | 3.71      | 1.00      | 18.00      |

\*The variable  $X_{O/D}^i$  denotes the land use area/floor area/number of  $X$  within the  $i$ th circle at the origin or destination station.

and  $A_D(\bullet)$  denote the attention mechanisms for both stations. As shown in Figure 6,  $FCN_O(\bullet)$  and  $FCN_D(\bullet)$  refer to fully connected layers used to extract multidimensional built environment variables refined by the attention mechanism in one dimension. Similarly,  $FCN_C(\bullet)$  is used to extract the

two-dimensional traffic impedance variables.  $\partial$ ,  $\partial_O$ ,  $\partial_D$ ,  $\partial_C$ , and other weights within the model are estimated by the MSE loss function and the Adam optimizer [45].

Shapley Additive exPlanation (SHAP), a method from coalitional game theory that is as important as permutation features, is an inspection technique that can be used for any model [46]. The SHAP values measure a features importance by calculating the average of its effect on the prediction under different circumstances. Specifically, it can be defined as

$$I_i = \frac{1}{n} \sum_{j=1}^n |\theta_i^j|, \quad (4)$$

where  $I_i$  represents the importance of the  $i$ th built environment variable on the station-to-station ridership and  $n$  is the number of the samples.  $\theta_i^j$  denotes the effect of the  $i$ th variable of the  $j$ th sample on ridership and is defined as follows:

$$\theta_i^j = \sum_{s \subseteq \{x_1^j, \dots, x_p^j\} \setminus x_i^j} \frac{|s|!(p-|s|-1)!}{p!} [F(s \cup x_i^j) - F(s)], \quad (5)$$

where  $s \subseteq \{x_1^j, \dots, x_p^j\} \setminus x_i^j$  refers to the variable combination that does not contain the  $i$ th variable of the  $j$ th sample,  $p$  is the number of the variables, and  $F(\bullet)$  represents the trained BAM-Gravity model.

## 5. Results and Discussion

To detect overfitting, 25% of the total samples were randomly selected as a test subset, and the 5-fold cross validation was used during the training process. Before applying the BAM-Gravity model, all variables were standardized. After tuning hyperparameters, when the learning rate and number of training times were 0.005 and 1200, the  $R^2$  for the BAM-Gravity model was 0.88.

**5.1. Importance of the Independent Variables.** The contributions of the independent variables for station-to-station ridership are shown in Table 2. Overall, there are two interesting findings that should be noted. First, the contribution of the commercial/business buildings (374.66) to the passenger flow is maximum, followed by the residential buildings (368.22) and traffic impedance (276.83). This result is consistent for most studies, and most metro trips on weekdays are to and from residential and commercial/business buildings that provide primary residence, daily shopping, and job opportunities [6, 14, 18, 20, 24].

Second, the contribution of the built environment variables in the third circle layer to ridership (403.32) is the highest, followed by the second (353.22), fourth (298.43), and the first circles (258.57). To explain this counter-intuitive result, the relationship between built environment variables in each of the circles and ridership is discussed and analyzed in detail in the following subsections.

**5.2. Relationship between Ridership and the Residential Buildings.** Figure 7 displays the relationship between station-to-station ridership and the residential buildings in

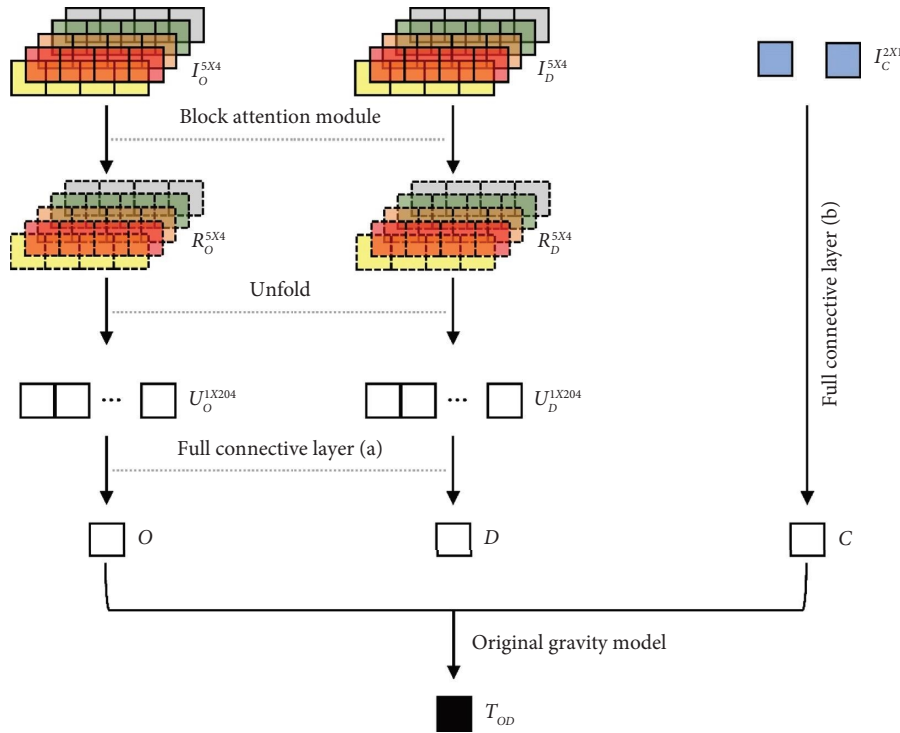


FIGURE 4: The overall structure of BAM-Gravity model.

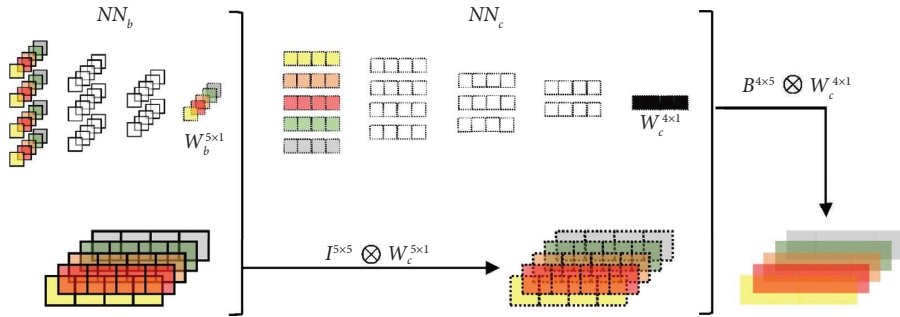


FIGURE 5: The structure of the block attention module.

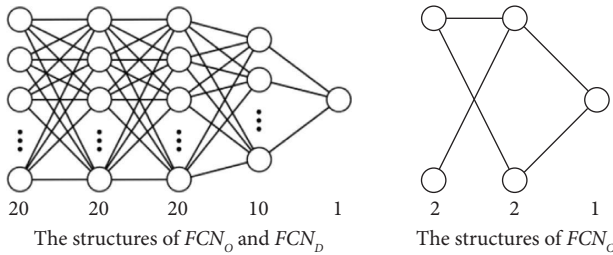


FIGURE 6: The structures of the fully connected layers.

different circles of origin and destination sides. In addition to the first circle, residential buildings in the other circles were found to significantly contribute to passenger flow. In particular, as shown in Figures 7(a) and 7(b), the residential buildings in the second circle of the origin and destination stations affecting the ridership show a similar trend of

exponential contribution to passenger flow. When the residential buildings are below 50,000 m<sup>2</sup>, the ridership is below the average and flat values in the 0–50 range. With the increase in residential floor area, ridership exceeds the average by a maximum of 150 persons. The residential buildings in the third circle of the origin (Figure 7(c)) and destination (Figure 7(d)) stations promote ridership with logarithmic trends. Gradual and slow increase in ridership are observed when residential buildings are greater than 125,000 m<sup>2</sup>. Figures 7(e) and 7(f) depict the increase in ridership with increasing number of residential buildings in the fourth circle on both sides. When the residential buildings are less than 100,000 m<sup>2</sup>, the ridership is below average.

Almost all the studies claimed that residential buildings or population density promote growth in passenger flow [25, 47, 48]. However, several further findings are worth noting; the residential buildings in the first circle do not significantly

TABLE 2: The importance of variables (mean of SHAP values).

| Variable      | SHAP             |
|---------------|------------------|
| $R_O^1$       | 13.68 (Rank 32)  |
| $R_D^1$       | 8.50 (Rank 36)   |
| $A_O^1$       | 33.64* (Rank 18) |
| $A_D^1$       | 31.77* (Rank 20) |
| $B_O^1$       | 32.51* (Rank 19) |
| $B_D^1$       | 31.08* (Rank 21) |
| $G_O^1$       | 5.43 (Rank 40)   |
| $G_D^1$       | 9.49 (Rank 35)   |
| $S_O^1$       | 40.79* (Rank 13) |
| $S_D^1$       | 51.68* (Rank 9)  |
| $R_O^2$       | 46.41* (Rank 12) |
| $R_D^2$       | 52.31* (Rank 8)  |
| $A_O^2$       | 25.35 (Rank 25)  |
| $A_D^2$       | 12.54 (Rank 33)  |
| $B_O^2$       | 13.80 (Rank 31)  |
| $B_D^2$       | 19.14 (Rank 28)  |
| $G_O^2$       | 48.39* (Rank 11) |
| $G_D^2$       | 66.48* (Rank 6)  |
| $S_O^2$       | 30.78* (Rank 22) |
| $S_D^2$       | 38.02* (Rank 14) |
| $R_O^3$       | 103.39* (Rank 2) |
| $R_D^3$       | 73.21* (Rank 5)  |
| $A_O^3$       | 21.21 (Rank 27)  |
| $A_D^3$       | 11.04 (Rank 34)  |
| $B_O^3$       | 60.48* (Rank 7)  |
| $B_D^3$       | 35.62* (Rank 16) |
| $G_O^3$       | 27.53* (Rank 24) |
| $G_D^3$       | 30.51* (Rank 23) |
| $S_O^3$       | 18.15 (Rank 29)  |
| $S_D^3$       | 22.18 (Rank 26)  |
| $R_O^4$       | 36.67* (Rank 15) |
| $R_D^4$       | 34.05* (Rank 17) |
| $A_O^4$       | 7.15 (Rank 38)   |
| $A_D^4$       | 5.95 (Rank 39)   |
| $B_O^4$       | 96.98* (Rank 3)  |
| $B_D^4$       | 85.05* (Rank 4)  |
| $G_O^4$       | 5.07 (Rank 42)   |
| $G_D^4$       | 14.12 (Rank 30)  |
| $S_O^4$       | 5.15 (Rank 41)   |
| $S_D^4$       | 8.24 (Rank 37)   |
| Transfer time | 228.42* (Rank 1) |
| Interval      | 48.41* (Rank 10) |

For reasons of space, the top 24 variables (with \*) are considered to significantly affect ridership.

affect passenger flow. Residential buildings in the third circle have the most significant impact on ridership (103.39 for O, 73.21 for D), followed by the second (46.41 for O, 52.31 for D) and fourth circles (36.67 for O, 34.05 for D). The area of the first circle is too small relative to the other circles, resulting in the development of a smaller amount of residential floor area within its boundaries (a maximum of approximately

50,000 m<sup>2</sup>). Several studies demonstrated that residential buildings or population densities within an appropriate threshold have a significant impact on passenger flow [34, 41, 47]. Fewer residential buildings in the first circle do not fall within the effective threshold for influencing ridership. Moreover, the relationships between residential buildings in different circles and passenger flow are slightly different, exhibiting exponential, logarithmic, and linear growth in the second, third, and fourth circles, respectively. The combination of distance to the metro station and the area of land available for residential development leads to these results.

*5.3. Relationship between Ridership and the Business\Commercial Buildings.* As shown in Figure 8, the business\commercial buildings have significantly impact on ridership in the first (32.51 for O, 31.08 for D), third (60.48 for O, 35.62 for D), and fourth circles (96.98 for O, 85.05 for D). In particular, the business\commercial buildings in the first and third circles are positively correlated with ridership, whereas in the fourth circle, ridership is inhibited, regardless of the origin or destination station.

The results of this study differed slightly from those of previous studies [25, 49], and this difference is attributed to a combination of reasons. First, the development of large commercial facilities mainly for leisure and entertainment has a certain agglomeration effect and is mainly concentrated in the first circle, whereas the commercial facilities scattered in the fourth circle are mostly small retail businesses providing daily services within a walking distance for residents; this not only reduces the dependence of surrounding residents on the metro but also makes it less attractive to outside residents. Second, the first circle has higher land prices owing to its transportation advantages, and most of the large commercial or business office facilities within its boundaries are geared towards the middle and upper-income groups. The well-equipped parking facilities and the preference of this group to travel by private car result in a less-than-expected contribution of the commercial buildings in the first circle. As a result, business\commercial buildings have the most significant negative effect on ridership in the fourth circle, while in the first circle, they do not have the expected degree of impact.

*5.4. Relationship between Ridership and the Public Green Land Area\Public Administration Buildings.* As shown in Figures 9(a)–9(d), there are significant negative correlations between public green space and passenger flow in the second (48.39 for O, 66.48 for D) and third (27.53 for O, 30.51 for D) circles, regardless of whether it is at the origin or destination stations.

Although Du et al. [47] and similar studies [25, 50] concluded that the public green spaces would attract more passengers and the impact has no significant difference between weekdays and weekends, the results of this study are significantly different. Despite the fact that Xi'an is a well-known tourist destination with many places of interest, the corresponding AFC data analyzed for this study show that during low tourist seasons such as November, when

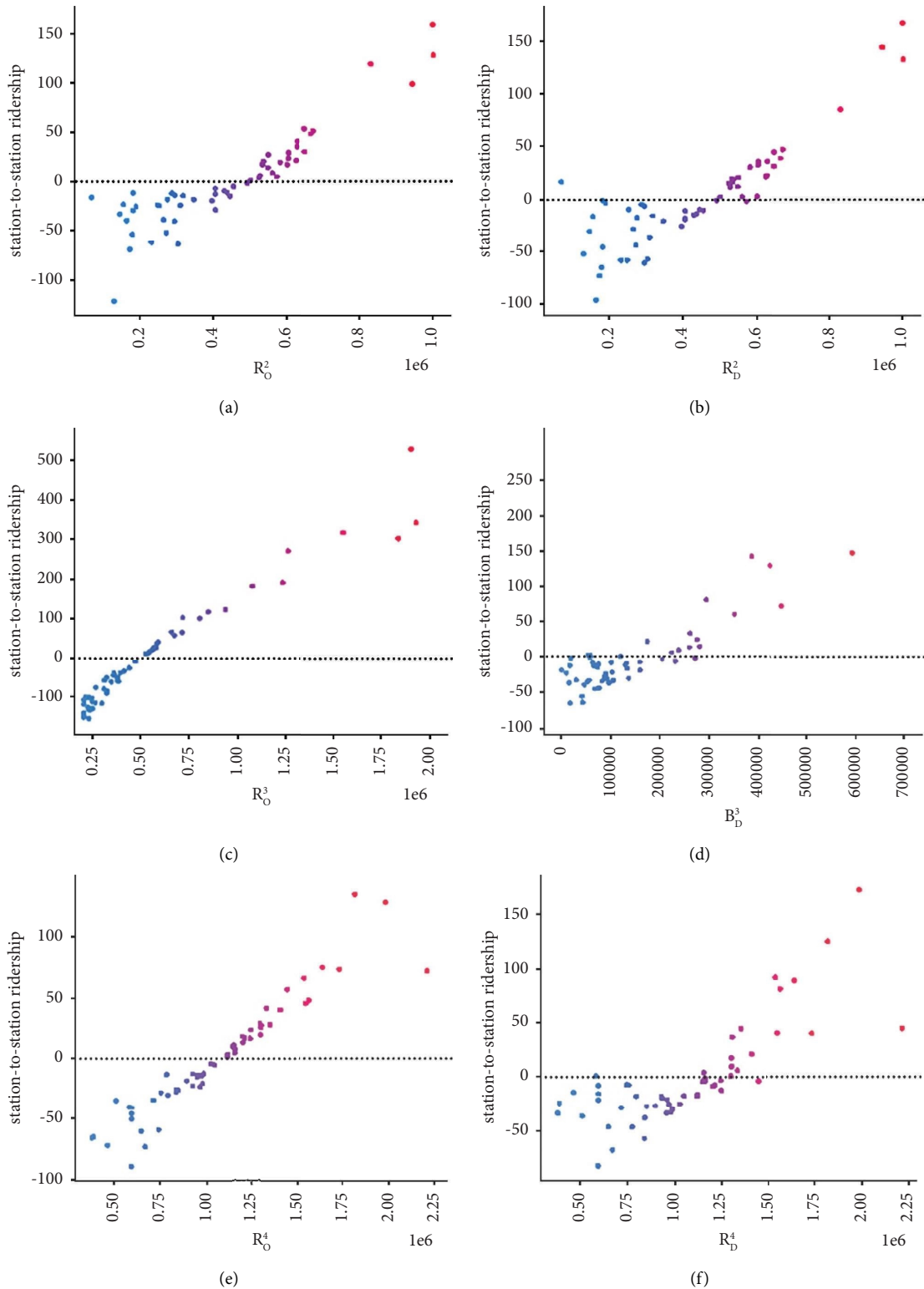


FIGURE 7: The relationship between ridership and the residential buildings.

temperatures are cooler, fewer people exercise in parks or visit scenic areas. Therefore, the larger public open space compresses the size of other facilities, leading to a reduction in ridership.

The public administration buildings significantly affect ridership in the first circle only (33.64 for O, 31.77 for D). The effective thresholds for public administration buildings on passenger flows at the origin (Figure 9(e)) and destination



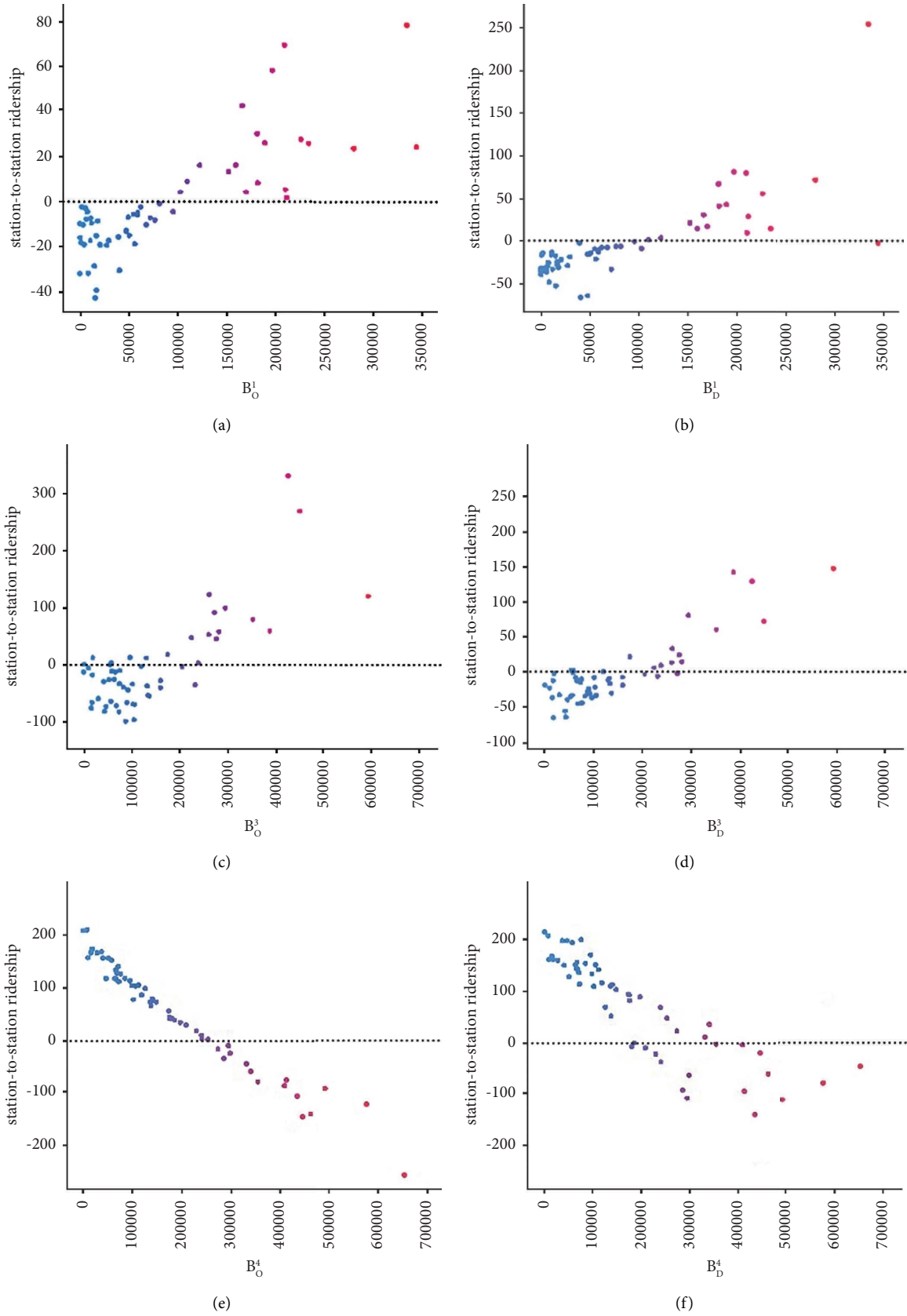


FIGURE 8: The relationship between ridership and the business\commercial buildings.

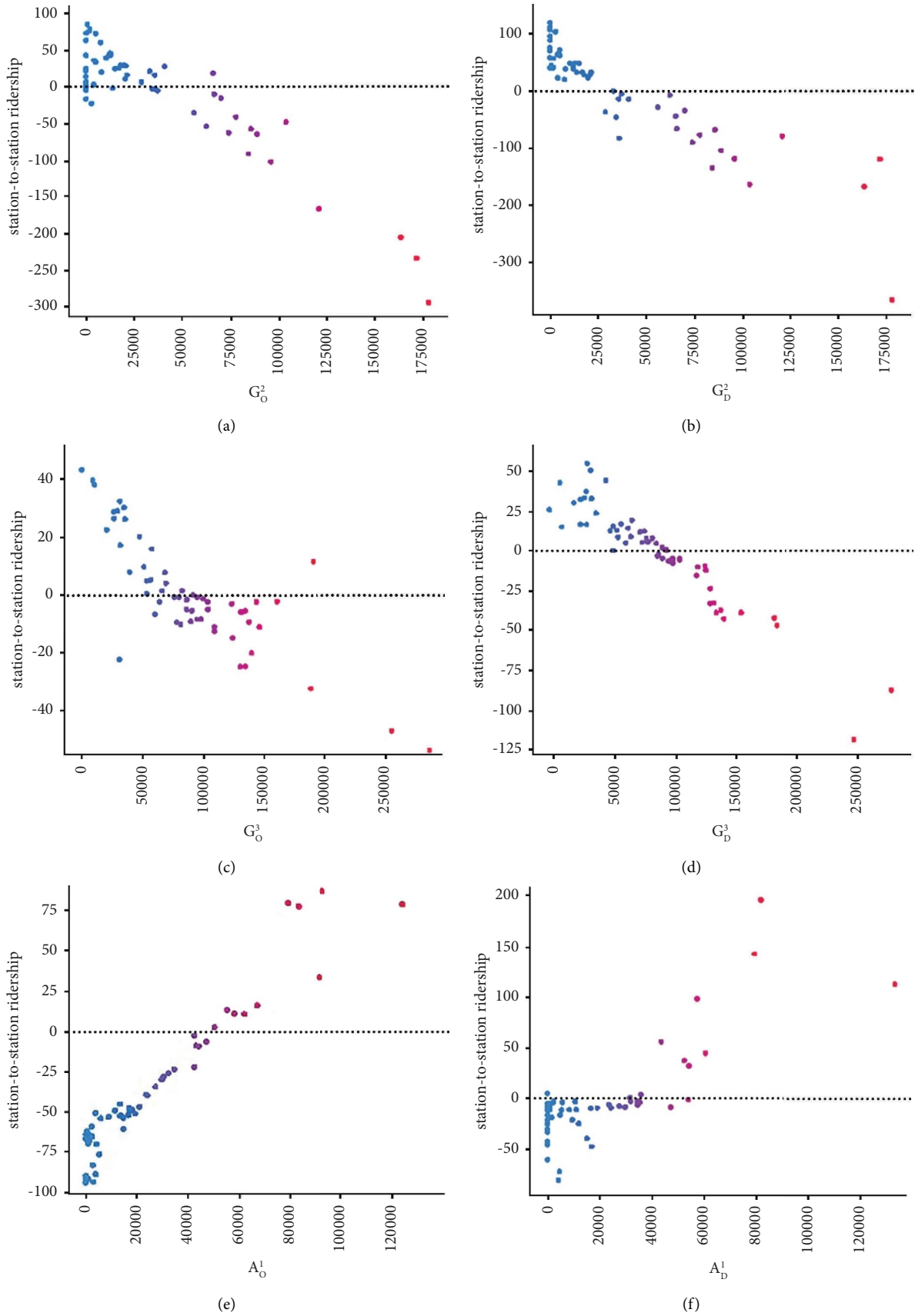


FIGURE 9: The relationship between ridership and the public green land area\public administration buildings.

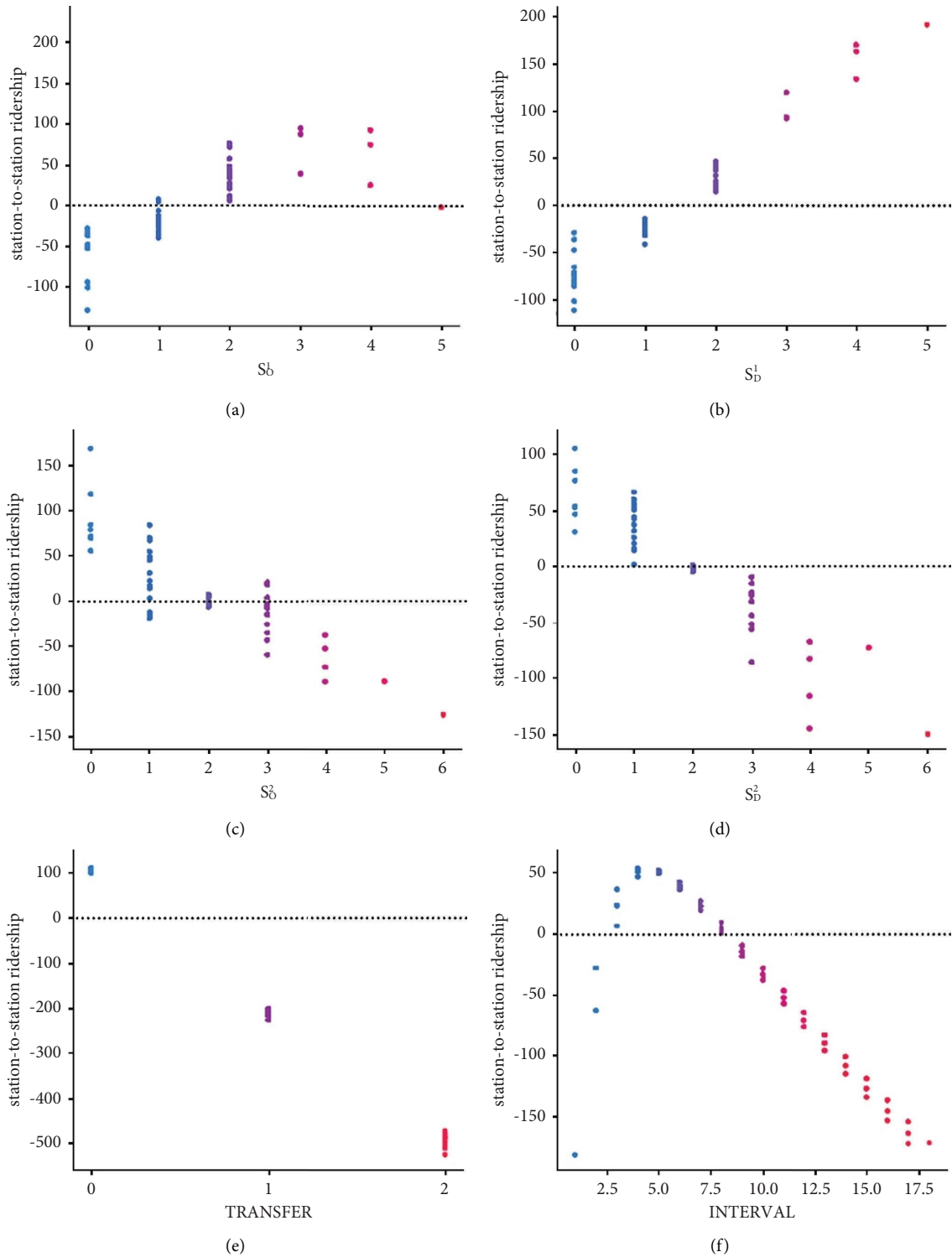


FIGURE 10: The relationship between ridership and the bus stop\travel impedance.

station (Figure 9(f)) are 2,000–8,000 m<sup>2</sup> and 4,000–8,000 m<sup>2</sup>, respectively. The public administration buildings include primary and secondary schools, government offices, university campuses, hospitals, stadiums, and exhibition halls. Students and patients are reluctant to walk long distances to their destinations after reaching the metro

station owing to safety and health concerns. Government offices only provide a small number of jobs, and therefore, they have little impact on passenger flow. During the working day, student travel on the university campus is largely spread around the university, with few long journeys by metro. At the same time, facilities such as stadiums and

exhibition halls are hardly attractive to residents with busy daily work schedules.

*5.5. Relationship between Ridership and the Traffic Impedance\Bus Stops.* The number of bus stops in the first circle at the origin station has a different impact on ridership from that at the destination station. Figure 10(a) shows that when the number of bus stops in the first circle at the origin station is 3, the maximum number of passengers is reached, exceeding the average by approximately 70 persons. Figure 10(b) shows that the more the bus stops in the first circle on the destination side, the higher the passenger flow. The results are attributed to three reasons. First, the existing metro system does not yet fully meet Xi'an's travel demand because a significant proportion of trips are made using the metro-bus connection or the bus-metro connection. Second, travelers prefer the metro-bus connection to the bus-metro connection. Eighty random interviews on the preference to metro-bus or bus-metro connections were conducted to explore the underlying reasons for this result. The majority of responses received include better punctuality and speed of the metro than bus and guaranteed subsequent travel arrangements after taking the metro first. Finally, some people would consider abandoning the metro if there was a direct bus to their destination, or if they do not intend to save time. Therefore, the optimal number of bus stops on the origin side can promote an increase in ridership; however, overly developed bus routes can lead to a decrease in ridership.

Other studies claimed that the number of bus stops on both sides positively affects passenger flow [11, 26, 41]. However, Figures 10(c) and 10(d) describe the impact of the number of bus stops in the second circle on ridership as negative, regardless of whether they are on the origin or destination side. The second circle of bus stops competes with the metro, with a greater number of stops leading to a higher chance of abandoning the metro for the bus.

The number of transfers (228.42, Rank 1) is the most important feature that affects OD flows. It corresponds to the findings of the other studies [34, 36]; people are more likely to abandon the metro because of the transfers (Figure 10(e)). Figure 10(f) shows that the relationship between the intervals and passenger flow is a parabolic curve. In particular, when the interval between two stations is in the range of 0–2 or 10–18, the OD passenger flow is lower than average. Furthermore, when it is in the range of 3–9, ridership is above average. It is consistent with the results of Gan et al. [34], as both studies conclude that most metro trips are taken to cover medium distances, and considerably short travel distance cannot reflect the advantages of metro reliability, thereby resulting in more alternative modes of travel options.

## 6. Conclusion

To investigate the influence of built environment factors in different circles of the origin and destination stations on passenger flow, five types of built environment variables (administration buildings, residential buildings,

commercial/business buildings, land area of the parks/squares/scenic spots, and the number of bus stops) in four circles (0–200, 200–400, 400–600, and 600–800 m) of the origin and destination stations and the travel impedance variables (transfer times and intervals) were used for modelling the metro station-to-station ridership. The BAM-Gravity model was employed to detect the relationship with ridership, and SHAP was used to explain the modelling results. The results of this study are expected to provide planners with better actionability.

First, the transfer is the most critical factor affecting the metro station-to-station passenger flow. The result indicates that the coupling between urban spatial structure planning and rail transit network planning should be strengthened. In other words, it suggests that policymakers can improve connective efficiency by directly connecting two important areas or gradually reducing the functional connection between the two regions in the process of urban renewal. Moreover, the result shows that people are more inclined to ride the metro for medium distances. Based on this result, planners should pay more attention to the spatial connection between the two areas at this scale in optimizing the urban structure.

Second, the bus stops within the first circle at the origin and destination exhibit a parabola and a positive correlation with the ridership, respectively. They exhibit negative correlations in the second circle, regardless of the origin or destination station. The results show that the existing metro network in Xi'an is not adequate to cover the city's travel demands. The results also highlight the role of integrated development of bus and metro systems in regulating the metro passenger flow. When the utility of metro travel needs to be increased, the connection between the metro and the bus systems should be strengthened in the first circle on the destination side and the number of bus stops (at most 3) on the origin side should be appropriately increased. When the metro passenger flow pressure needs to be relieved, the number of bus stops within the second circle should be increased. Furthermore, transportation planners and operators can adjust metro ridership by optimizing the bus schedules of the first and second circles on both sides.

Finally, and most importantly, the results of this study reveal the differences and similarities in the impact of different land use factors within different circles on passenger flow. The findings provide some reference for the optimization of land use around the metro station, whether it is around the origin or destination station. The residential buildings contribute to noticeable improvements in passenger flow, and the improvement level arranged in the sequence is as follows: third, second, and fourth circles. It suggests that planners should focus more on the circle-level heterogeneity of the impact of residential buildings on ridership and rationally develop residential land in each circle. Similarly, planners can adjust the passenger flow by planning commercial or business office buildings in different circles. Although the land use of parks and squares has an inhibitory effect on the passenger flow in the second and third circles, it does not mean that the development of land use such as parks should be prohibited. Planners should

make the most of broken spaces by arranging public open spaces along major pedestrian routes and enhancing the landscaping of pedestrian spaces.

Although this study enriches the existing research on TOD, some aspects should be further explored. First, the results show that the built environment has essentially the same impact on average daily ridership at the origin and destination stations, but there is no discussion regarding the impact on passenger flows at different periods of the day. Second, this study does not investigate the competition between travel modes and the metro in terms of traffic impedance. Third, based on the density of the road network in Xi'an, 200 m was used to delineate the station catchment area circles; however, a smaller buffer band would have led to more detailed results. These limitations of this study can be addressed in the future.

### Data Availability

The datasets generated and analyzed in the current study are not publicly available due to privacy reasons but are available from the corresponding author upon reasonable request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

This work was partially supported by the Fundamental Research Funds for the Central Universities (300102210657).

### References

- [1] R. Ewing and R. Cervero, "Travel and the built environment: a meta-analysis," *Journal of the American Planning Association*, vol. 76, no. 3, pp. 265–294, 2010.
- [2] D. Wang and M. Zhou, "The built environment and travel behavior in urban China: a literature review," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 574–585, 2017.
- [3] K. Kepaptsoglou, A. Stathopoulos, and M. G. Karlaftis, "Ridership estimation of a new LRT system: direct demand model approach," *Journal of Transport Geography*, vol. 58, pp. 146–156, 2017.
- [4] P. Calthorpe, *The Next American metropolis: Ecology, Community, and the American Dream*, Princeton architectural press, New York, NY, USA, 1993.
- [5] 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.
- [6] H. Fitzová, M. Matulová, and Z. Tomeš, "Determinants of urban public transport efficiency: case study of the Czech Republic," *European Transport Research Review*, vol. 10, no. 2, 2018.
- [7] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," *Lecture Notes in Computer Science*, vol. 11211, pp. 3–19, 2018.
- [8] A. Nasri and L. Zhang, "The analysis of transit-oriented development (TOD) in Washington, DC and Baltimore metropolitan areas," *Transport Policy*, vol. 32, pp. 172–179, 2014.
- [9] X. J. Cao, "Examining the effect of the hiawatha LRT on auto use in the twin cities," *Transport Policy*, vol. 81, pp. 284–292, 2019.
- [10] Q. Hou, Y. Xing, D. Wang, J. Liu, X. Fan, and Y. Duan, "Study on coupling degree of rail transit capacity and land use based on multivariate data from cloud platform," *Journal of Cloud Computing*, vol. 9, no. 1, 4 pages, 2020.
- [11] X. Ma, J. Zhang, C. Ding, and Y. Wang, "A geographically and temporally weighted regression model to explore the spatio-temporal influence of built environment on transit ridership," *Computers, Environment and Urban Systems*, vol. 70, pp. 113–124, 2018.
- [12] D. Zhang, X. Wang, and Cara), "Transit ridership estimation with network Kriging: a case study of Second Avenue Subway, NYC," *Journal of Transport Geography*, vol. 41, pp. 107–115, 2014.
- [13] M. Durning and C. Townsend, "Direct ridership model of rail rapid transit systems in Canada," *Transportation Research Record*, vol. 2537, no. 1, pp. 96–102, 2015.
- [14] Y. Zhu, F. Chen, Z. Wang, and J. Deng, "Spatio-temporal analysis of rail station ridership determinants in the built environment," *Transportation*, vol. 46, no. 6, pp. 2269–2289, 2019.
- [15] M.-J. Jun, K. Choi, J.-E. Jeong, K.-H. Kwon, and H.-J. Kim, "Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul," *Journal of Transport Geography*, vol. 48, pp. 30–40, 2015.
- [16] C. E. Vergel-Tovar and D. A. Rodriguez, "The ridership performance of the built environment for BRT systems: evidence from Latin America," *Journal of Transport Geography*, vol. 73, pp. 172–184, 2018.
- [17] R. Cervero, "Alternative approaches to modeling the travel-demand impacts of smart growth," *Journal of the American Planning Association*, vol. 72, no. 3, pp. 285–295, 2006.
- [18] V. Chakour and N. Eluru, "Examining the influence of stop level infrastructure and built environment on bus ridership in Montreal," *Journal of Transport Geography*, vol. 51, pp. 205–217, 2016.
- [19] Z. Gan, T. Feng, M. Yang, H. Timmermans, and J. Luo, "Analysis of metro station ridership considering spatial heterogeneity," *Chinese Geographical Science*, vol. 29, no. 6, pp. 1065–1077, 2019.
- [20] X. Li, Y. Liu, Z. Gao, and D. Liu, "Decision tree based station-level rail transit ridership forecasting," *Journal of Urban Planning and Development*, vol. 142, no. 4, 2016.
- [21] H. Pan, J. Li, Q. Shen, and C. Shi, "What determines rail transit passenger volume? Implications for transit oriented development planning," *Transportation Research Part D: Transport and Environment*, vol. 57, pp. 52–63, 2017.
- [22] L. S. Sun, S. W. Wang, L. Y. Yao, J. Rong, and J. M. Ma, "Estimation of transit ridership based on spatial analysis and precise land use data," *Transportation Letters*, vol. 8, no. 3, pp. 1–8, 2016.
- [23] 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.
- [24] S. Wang, L. Sun, J. Rong, S. Hao, and W. Luo, "Transit trip distribution model considering land use differences between catchment areas," *Journal of Advanced Transportation*, vol. 50, no. 8, pp. 1820–1830, 2016.

- [25] D. An, X. Tong, K. Liu, and E. H. W. Chan, "Understanding the impact of built environment on metro ridership using open source in Shanghai," *Cities*, vol. 93, pp. 177–187, 2019.
- [26] P. Li, P. Zhao, and T. Schwanen, "Effect of land use on shopping trips in station areas: examining sensitivity to scale," *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 969–985, 2020.
- [27] M. Pasha, S. M. Rifaat, R. Tay, and A. De Barros, "Effects of street pattern, traffic, road infrastructure, socioeconomic and demographic characteristics on public transit ridership," *KSCE Journal of Civil Engineering*, vol. 20, no. 3, pp. 1017–1022, 2016.
- [28] 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, Article ID 102878, 2020.
- [29] K. Sohn and H. Shim, "Factors generating boardings at Metro stations in the Seoul metropolitan area," *Cities*, vol. 27, no. 5, pp. 358–368, 2010.
- [30] G. Sun, C. Webster, and A. Chiaradia, "Objective assessment of station approach routes: development and reliability of an audit for walking environments around metro stations in China," *Journal of Transport & Health*, vol. 4, pp. 191–207, 2017.
- [31] E. Chen, Z. Ye, C. Wang, and W. Zhang, "Discovering the spatio-temporal impacts of built environment on metro ridership using smart card data," *Cities*, vol. 95, Article ID 102359, 2019.
- [32] J. Choi, Y. J. Lee, T. Kim, and K. Sohn, "An analysis of Metro ridership at the station-to-station level in Seoul," *Transportation*, vol. 39, no. 3, pp. 705–722, 2012.
- [33] M. Duncan, "To park or to develop: trade-off in rail transit passenger demand," *Journal of Planning Education and Research*, vol. 30, no. 2, pp. 162–181, 2010.
- [34] Z. Gan, M. Yang, T. Feng, and H. J. P. Timmermans, "Examining the relationship between built environment and metro ridership at station-to-station level," *Transportation Research Part D: Transport and Environment*, vol. 82, Article ID 102332, 2020.
- [35] H. Iseki, C. Liu, and G. Knaap, "The determinants of travel demand between rail stations: a direct transit demand model using multilevel analysis for the Washington D.C. Metrorail system," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 635–649, 2018.
- [36] J. Zhao, W. Deng, Y. Song, and Y. Zhu, "Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models," *Transportation*, vol. 41, no. 1, pp. 133–155, 2014.
- [37] J. Gutiérrez, O. D. Cardozo, and J. C. García-Palomares, "Transit ridership forecasting at station level: an approach based on distance-decay weighted regression," *Journal of Transport Geography*, vol. 19, no. 6, pp. 1081–1092, 2011.
- [38] O. Manout, P. Bonnel, and L. Bouzouina, "Transit accessibility: a new definition of transit connectors," *Transportation Research Part A: Policy and Practice*, vol. 113, pp. 88–100, 2018.
- [39] S. Chan and L. Miranda-Moreno, "A station-level ridership model for the metro network in Montreal, Quebec," *Canadian Journal of Civil Engineering*, vol. 40, no. 3, pp. 254–262, 2013.
- [40] E. Chen, Z. Ye, and H. Wu, "Nonlinear effects of built environment on intermodal transit trips considering spatial heterogeneity," *Transportation Research Part D: Transport and Environment*, vol. 90, Article ID 102677, 2021.
- [41] 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.
- [42] Y. Liu, Z. Liu, and R. "D.P. F. Jia, "DeepPF: a deep learning based architecture for metro passenger flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 101, pp. 18–34, 2019.
- [43] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [44] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, p. 818, 2017.
- [45] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, Academia, San Francisco, CA, USA, 2014.
- [46] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4768–4777, Red Hook, NY, USA, December 2017.
- [47] Q. Du, Y. Zhou, Y. Huang, Y. Wang, and L. Bai, "Spatio-temporal exploration of the non-linear impacts of accessibility on metro ridership," *Journal of Transport Geography*, vol. 102, Article ID 103380, 2022.
- [48] F. Gao, L. Yang, C. Han, J. Tang, and Z. Li, "A network-distance-based geographically weighted regression model to examine spatiotemporal effects of station-level built environments on metro ridership," *Journal of Transport Geography*, vol. 105, Article ID 103472, 2022.
- [49] P. Lin, J. Weng, D. K. Brands, H. Qian, and B. Yin, "Analysing the relationship between weather, built environment, and public transport ridership," *IET Intelligent Transport Systems*, vol. 14, no. 14, pp. 1946–1954, 2020.
- [50] L. M. Pfister, R. G. Thompson, and L. Zhang, "Spatiotemporal exploration of Melbourne pedestrian demand," *Journal of Transport Geography*, vol. 95, Article ID 103151, 2021.