

Research on the Mechanism of Multi-Scale Urban Morphology Influencing Accessibility Equity and Machine Learning-Based Evaluation Methods

Xiaoying Ou

*Faculty of Architecture, The University of Hong Kong, Hong Kong, China
ouxiaoying02409@outlook.com*

Abstract. Urban planning and development governance are shifting from scale-oriented expansion to equity-oriented spatial resource optimization. Accessibility equity has therefore become an important indicator for evaluating public service allocation and urban spatial inclusiveness. Existing studies have often assessed accessibility through travel distance, service coverage, or facility supply, yet the predictive relationship between multi-scale urban morphology and accessibility equity remains insufficiently explained. Taking the central urban area of Chengdu as the empirical case, this study integrates remote sensing images, OSM road networks, POI data, building footprints, bus and metro stations, community boundaries, and gridded population data. It constructs an urban morphology indicator system across 500 m grids, street-units, and district scales. Accessibility to education, healthcare, public transportation, and public green space is first measured separately, then standardized into a comprehensive accessibility equity index. Random forest, XGBoost, support vector regression, and SHAP interpretation are used to identify key predictive variables. The results show that XGBoost performs better than the other models. Transit station density, intersection density, land-use mix, healthcare facility density, and population demand intensity show higher predictive contributions. This study provides a computable and interpretable evaluation path for urban renewal, transport facility improvement, and public service spatial governance.

Keywords: urban morphology, accessibility equity, multi-scale analysis, machine learning, urban governance

1. Introduction

Urban planning and development governance are shifting from expansion-oriented growth to equity-oriented spatial optimization, making accessibility equity a key criterion for evaluating the balanced allocation of public services. Existing studies have measured opportunity differences through travel time, service radius, and facility coverage, yet they have not sufficiently explained how urban morphology shapes these differences across spatial scales [1]. In real urban space, grid units describe basic spatial attributes, street units reflect road-enclosed pedestrian environments, and district scales are associated with public transit organization and cross-regional resource connections

[2]. If accessibility is assessed from only one scale, the combined effects of road connectivity, land-use mix, facility density, and population demand may be overlooked. Therefore, this paper integrates remote sensing data, POI data, road networks, gridded population data, transit stations, and public service facilities to build a multi-scale urban morphology indicator system. It further applies an accessibility equity index, machine learning models, and SHAP interpretation to identify key predictive factors and provide quantitative support for urban renewal and public service governance.

2. Literature review

2.1. Multi-scale urban morphology and spatial governance

Multi-scale urban morphology refers to the organization of urban spatial structure across grid, block, community, district, and city scales, including road networks, land-use mix, building coverage, functional density, and centrality patterns [3]. Previous studies suggest that compact form and highly connected road networks often improve urban operational efficiency, but excessive reliance on macro-level density indicators may conceal differences within micro-spatial units. The block scale is closer to residents' daily walking experience, while the district scale corresponds more directly to public transit organization and urban governance boundaries. These two scales cannot be simply substituted for each other [4]. Grid analysis helps solve the problem of inconsistent spatial resolution among multi-source datasets, but it is essentially a standardized computational unit rather than a real urban block. Therefore, urban morphology analysis should combine grid-based calculation, street-unit identification, and district-level aggregation to maintain both computational comparability and planning interpretability.

2.2. Accessibility equity and public service distribution

Accessibility equity refers to the balanced opportunity of different spatial units or population groups to access public services such as education, healthcare, transportation, and green space. Traditional accessibility studies often focus on distance or time cost, but equity is not determined only by whether a facility can be reached. It also depends on service capacity, population demand, and the spatial distribution of opportunities [5]. The two-step floating catchment area method measures service opportunities by matching facility supply with population demand, which compensates for the limitation of simple service-radius analysis. The Gini coefficient and Theil index are suitable for measuring overall inequality, while local spatial autocorrelation can identify low-service areas and clusters of accessibility disadvantage [6]. Since these methods reflect different dimensions of equity, accessibility should first be calculated separately for education, healthcare, public transit, and green space, and then standardized and weighted into a comprehensive accessibility equity index.

2.3. Machine learning applications in urban spatial evaluation

Machine learning in urban spatial evaluation refers to the use of algorithms such as random forest, XGBoost, and support vector regression to identify nonlinear relationships between high-dimensional spatial variables and evaluation outcomes [7]. Compared with traditional regression, machine learning does not require fully linear variable relationships, making it more suitable for handling complex interactions among urban morphology, transport conditions, facility distribution, and population demand. However, predictive accuracy does not automatically produce planning interpretability [8]. If a model only outputs predicted results without explaining variable

contributions, it is difficult to translate the findings into governance strategies. Explainable machine learning uses feature importance and SHAP values to reveal the predictive contribution of different indicators to the accessibility equity index, allowing model outputs to correspond to specific planning variables [9]. In this framework, multi-scale urban morphology indicators serve as independent variables, the accessibility equity index serves as the dependent variable, and model evaluation metrics are used to compare predictive stability and explanatory reliability.

3. Experimental methods

3.1. Data sources and multi-scale urban morphology indicator construction

This study selects the central urban area of Chengdu as the empirical case. The study area covers the main built-up areas of Jinjiang District, Qingyang District, Jinniu District, Wuhou District, Chenghua District, and Chengdu High-tech Zone, with an approximate area of 430 km². This area includes high-density old urban blocks, rail transit corridors, newly developed districts, and peripheral residential clusters. It can therefore support a comparative analysis of the relationship between multi-scale urban morphology and accessibility equity. The data sources include Sentinel-2 remote sensing images, OpenStreetMap road networks, POI data from Gaode or Baidu, building footprints, bus and metro stations, community boundaries, and 100 m gridded population data. The experiment first uses the GIS Fishnet tool to generate 500 m × 500 m standard grids, with approximately 1,720 basic computational units. The grid is only used to unify the spatial scale of multi-source data. It is not treated as a real urban block. Road density, intersection density, building coverage ratio, POI functional mix, public service facility density, and transit station density are then calculated within each grid. Street-units are identified through road-enclosed boundaries. Grid-based indicators are aggregated into street-units through area weighting and spatial overlay. They are further aggregated to the district or subdistrict scale. This process forms a multi-scale indicator system based on grid calculation, street-unit aggregation, and district-level comparison. It provides explanatory variables for the subsequent calculation of the accessibility equity index and machine learning modelling.

3.2. Accessibility equity measurement and spatial evaluation model

After spatial units and morphological indicators are constructed in Section 3.1, Section 3.2 measures service accessibility for four categories, namely education, healthcare, public transportation, and public green space. The accessibility of each spatial unit is first calculated separately. These results are then integrated into a comprehensive accessibility equity index. The two-step floating catchment area method is used to describe the matching relationship between facility supply and population demand [10]. Its formula as shown in Equation (1):

$$A_{ik} = \sum_{j\{d_{ij} \leq d_0\}} \frac{S_{jk}}{\sum_{i\{d_{ij} \leq d_0\}} P_i} \quad (1)$$

In this formula, A_{ik} represents the accessibility of spatial unit i to service category k . S_{jk} represents the service capacity of facility j . P_i represents population demand. d_{ij} represents network-based travel distance or travel time. d_0 represents the service threshold. School capacity can be used for education services. Hospital grade or bed numbers can be used for healthcare services. Station service frequency can be used for public transportation. Green space area can be used as the service capacity of public green space. The accessibility values of the four service categories are then

normalized through min-max normalization. A comprehensive accessibility equity index as shown in Equation (2):

$$AEI_i = \sum_{k=1}^4 w_k \cdot \frac{A_{ik} - \min(A_k)}{\max(A_k) - \min(A_k)} \quad (2)$$

In this formula, AEI_i represents the comprehensive accessibility equity index of spatial unit i . w_k represents the weight of service category k . The Gini coefficient, Theil index, and local spatial autocorrelation are mainly used to test the spatial inequality of the comprehensive index. They are not directly used as the dependent variable in the machine learning model.

3.3. Machine learning modeling and explainable analysis process

In Section 3.3, the multi-scale urban morphology indicators extracted in Section 3.1 are used as independent variables. The comprehensive accessibility equity index AEI_i calculated in Section 3.2 is used as the dependent variable. This setting is used to build a predictive relationship between urban morphology variables and accessibility equity levels. The input variables include road density, intersection density, land-use mix, building coverage ratio, public service facility density, transit station density, central node proximity, and population demand intensity. Random forest, XGBoost, and support vector regression are compared in the modelling process. The dataset is divided into training and testing sets at a ratio of 7:3. Five-fold cross-validation is used to control model stability [11]. The prediction function as shown in Equation (3):

$$\widehat{AEI}_i = f(X_i) + \varepsilon_i \quad (3)$$

In this formula, X_i represents the set of multi-scale urban morphology variables of spatial unit i . $\widehat{AEI}_i$ represents the predicted value of the model. ε_i represents the error term. Model performance is evaluated using R^2 , MAE, and RMSE. The RMSE as shown in Equation (4):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (AEI_i - \widehat{AEI}_i)^2} \quad (4)$$

In the interpretation stage, this study uses feature importance and SHAP values to explain the predictive contribution of variables. The model results are not interpreted as strict causal effects. SHAP results are reported through the Mean absolute SHAP value, which represents the average contribution intensity of each variable to the prediction of AEI_i . This helps identify the urban morphology indicators that are most important for predicting accessibility equity.

4. Results

4.1. Effects of multi-scale urban morphology on accessibility equity

The results show that the mean value of the comprehensive accessibility equity index at the 500 m grid scale in the central urban area of Chengdu is 0.586, with a standard deviation of 0.142. This indicates clear differences in public service opportunities across spatial units. By service category, public transportation has the highest mean accessibility value, at 0.648. Education services reach 0.603, healthcare services reach 0.557, and public green space reaches 0.529. Spatially, the mean AEI of the old urban core reaches 0.692. The value around rail transit corridors is 0.631. The value

in peripheral new districts is 0.448. These results suggest that low-density development areas and areas with insufficient public transit coverage are more likely to become weak accessibility zones. Figure 1 is titled Multi-scale Mean Accessibility Equity Index across Grid, Street-Unit and District Scales. The data come from the AEI calculated in Section 3.2. The mean AEI values of the 500 m grid, street-unit, and district scale are 0.586, 0.604, and 0.621, respectively. The index slightly increases when the scale becomes larger. This is mainly because spatial aggregation smooths some local low values. It does not mean that low-accessibility units have disappeared. Therefore, the interpretation of the results still needs to return to the grid scale and the street-unit scale to identify specific spatial weaknesses.

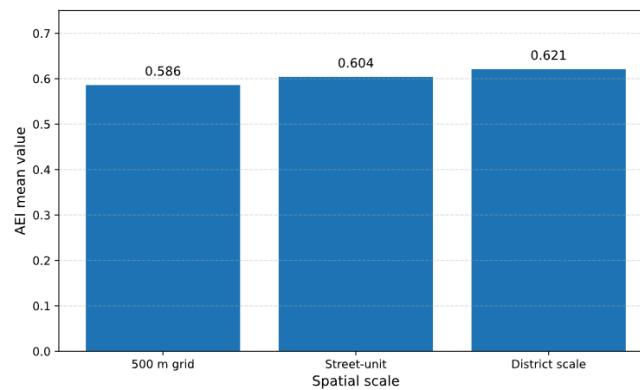


Figure 1. Multi-scale mean accessibility equity index across grid, street-unit and district scales

4.2. Model performance and identification of key factors

As shown in Table 1, the model comparison indicates that XGBoost performs best on the testing set, with an R^2 of 0.812, an MAE of 0.046, and an RMSE of 0.063. Random forest achieves an R^2 of 0.776, an MAE of 0.052, and an RMSE of 0.071. Support vector regression achieves an R^2 of 0.694, an MAE of 0.064, and an RMSE of 0.086. These results indicate that ensemble learning models are more suitable for handling the nonlinear relationship between multi-scale urban morphology variables and AEI. SHAP interpretation further shows that transit station density has the highest Mean absolute SHAP value, at 0.083. Intersection density reaches 0.071, land-use mix reaches 0.064, healthcare facility density reaches 0.058, and population demand intensity reaches 0.052. These values represent the average contribution of each variable to model prediction. They do not represent strict causal strength. More specifically, transit station density and intersection density correspond to cross-district connectivity and walking accessibility, respectively. Land-use mix reflects the proximity of daily services. Population demand intensity tends to lower the predicted AEI when facility supply is insufficient.

Table 1. Model performance and SHAP-based predictive contribution

Category	Indicator	Value
Model performance	XGBoost R^2	0.812
	XGBoost MAE	0.046
	XGBoost RMSE	0.063
	Random Forest R^2	0.776

Table 1. (continued)

Model performance	Random Forest MAE	0.052
	Random Forest RMSE	0.071
	SVR R ²	0.694
	SVR MAE	0.064
	SVR RMSE	0.086
SHAP interpretation	Transit station density Mean absolute SHAP value	0.083
	Intersection density Mean absolute SHAP value	0.071
	Land-use mix Mean absolute SHAP value	0.064
	Healthcare facility density Mean absolute SHAP value	0.058
	Population demand intensity Mean absolute SHAP value	0.052

5. Discussion

The results indicate that multi-scale urban morphology has a scale-specific and nonlinear predictive relationship with accessibility equity. The grid scale helps identify local low-accessibility units. The street-unit scale better reflects walking environments and road connectivity. The district scale captures public transit connection and the distribution pattern of service resources. In terms of indicator importance, transit station density shows the highest predictive contribution, followed by intersection density, land-use mix, healthcare facility density, and population demand intensity. This suggests that transport accessibility and service-demand matching are key factors shaping predicted AEI values. Interventions should therefore be scale-specific. At the district scale, transit station density should be improved first. At the street-unit scale, intersection connectivity and pedestrian routes should be strengthened. At the community scale, land-use mix should be optimized. In medically underserved areas, primary healthcare facilities should be added. In high-demand areas, service capacity should be increased at the same time.

6. Conclusion

This study develops a machine learning-based predictive framework to examine the relationship between multi-scale urban morphology and accessibility equity in central Chengdu. The 500 m grid is used as a standardized computational unit, while street-units and district scales support planning interpretation. A comprehensive accessibility equity index is constructed through the two-step floating catchment area method and standardized integration. The results show that XGBoost performs well, and transit station density, intersection density, and land-use mix make higher contributions to AEI prediction. These findings indicate predictive relationships rather than strict causal effects. Future research can include dynamic traffic and real travel data to improve model accuracy.

References

- [1] Wu, Cai, et al. "Machine learning-based characterisation of urban morphology with the street pattern." *Computers, environment and urban systems* 109 (2024): 102078.
- [2] Chen, Yang, et al. "Refining urban morphology: An explainable machine learning method for estimating footprint-level building height." *Sustainable Cities and Society* 112 (2024): 105635.
- [3] Jiang, Yanxiao, et al. "Evaluating the impact of urban morphology on urban vitality: An exploratory study using big geo-data." *International Journal of Digital Earth* 17.1 (2024): 2327571.
- [4] Ma, Ding, et al. "Graph convolutional networks for street network analysis with a case study of urban polycentricity in Chinese cities." *International Journal of Geographical Information Science* 38.5 (2024): 931-955.
- [5] Song, Lu, Xuesong Kong, and Peng Cheng. "Supply-demand matching assessment of the public service facilities in 15-minute community life circle based on residents' behaviors." *Cities* 144 (2024): 104637.
- [6] Ellder, Erik. "Exploring socio-economic inequalities in access to the 15-minute city across 200 Swedish built-up areas." *Journal of Transport Geography* 122 (2025): 104060.
- [7] Hu, Liyang, et al. "Integration of e-scooter sharing with public transit on employment accessibility and equity." *Transportation Research Part D: Transport and Environment* 140 (2025): 104604.
- [8] Kharel, Subham, et al. "Impact of Zoning Regulations on Transportation Equity of Job Accessibility: A Spatial Autoregressive Approach for the Dallas–Fort Worth and Houston-Galveston Area Council Areas." *Transportation Research Record* 2679.1 (2025): 1259-1277.
- [9] Sulis, Patrizia, and Paola Proietti. "Who can access what? Uncovering urban inequality in access to service for senior citizens." *Environment and Planning B: Urban Analytics and City Science* 51.7 (2024): 1650-1665.
- [10] Weigand, Matthias, et al. "Are public green spaces distributed fairly? A nationwide analysis based on remote sensing, OpenStreetMap and census data." *Geocarto International* 38.1 (2023): 2286305.
- [11] Chukwu, Michaelmary, et al. "Urban park accessibility assessment using human mobility data: A systematic review." *Annals of GIS* 30.2 (2024): 181-198.