

Article

Exploring the Determinants of Spatial Vitality in High-Speed Rail Station Areas in China: A Multi-Source Data Analysis Using LightGBM

Pengpeng Liang, Xu Cui *, Jiexi Ma *, Wen Song and Yao Xu 

School of Architecture, Southwest Jiaotong University, 999 Xi'an Road, Chengdu 611756, China; liang_pengpeng@my.swjtu.edu.cn (P.L.); songwen@my.swjtu.edu.cn (W.S.); xuyao@my.swjtu.edu.cn (Y.X.)

* Correspondence: cuixu@swjtu.edu.cn (X.C.); mjx@my.swjtu.edu.cn (J.M.)

Abstract: High-speed rail (HSR) station areas play a vital role in shaping urban form, stimulating economic activity, and enhancing spatial vitality. Understanding the factors that influence this vitality is key to supporting sustainable urban development and transit-oriented planning. This study investigates 66 HSR station areas in 35 Chinese cities by integrating multi-source data—Sina Weibo check-in records, urban support indicators, station attributes, and built environment variables—within a city–node–place analytical framework. Using Multiple Linear Regression (MLR) and Light Gradient Boosting Machine (LightGBM) models, we identify key drivers of spatial vitality, while SHAP analysis reveals nonlinear and interaction effects. The results show that city population size, urbanization level, commercial land use, transit accessibility, and parking facilities significantly enhance station area vitality. However, diminishing returns are observed when commercial land and bus stop densities exceed certain thresholds. The station location index shows a negative correlation with spatial vitality. The analysis of interaction effects highlights strong synergies between urban development and functional configuration, as well as between accessibility and service infrastructure. Different station types exhibit varied spatial patterns and require differentiated strategies. This study offers empirical insights for aligning transport infrastructure and land use planning, supporting the development of vibrant, accessible, and sustainable HSR station areas.

Keywords: high-speed rail; station area spatial vitality; multi-source data; Light Gradient Boosting Machine (LightGBM)



Academic Editor: Thomas Panagopoulos

Received: 3 May 2025

Revised: 4 June 2025

Accepted: 6 June 2025

Published: 12 June 2025

Citation: Liang, P.; Cui, X.; Ma, J.; Song, W.; Xu, Y. Exploring the Determinants of Spatial Vitality in High-Speed Rail Station Areas in China: A Multi-Source Data Analysis Using LightGBM. *Land* **2025**, *14*, 1262. <https://doi.org/10.3390/land14061262>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent decades, the rapid expansion of high-speed rail (HSR) networks has become a transformative force reshaping the spatial organization of cities and regions [1,2]. As an efficient, high-capacity mode of intercity transport, HSR enhances regional connectivity, compresses travel time, and drives urban integration and spatial restructuring across scales [3–5]. In China, the development of the world's largest HSR system has not only fulfilled key transportation demands but has also profoundly impacted urban form, land use patterns, and population distribution [6–8].

HSR stations, serving as both transport nodes and urban centers, play a pivotal role in promoting economic vitality and guiding development direction [9–11]. However, the success of these stations extends beyond their transport functionality—it is increasingly reflected in the intensity, diversity, and persistence of human activities they attract. This has given rise to the concept of spatial vitality, which refers to the dynamic state of human

presence and interaction within urban space over time. Spatial vitality is a multidimensional construct encompassing economic vitality (e.g., business activity, consumption), social vibrancy (e.g., pedestrian flows, diversity of social interactions), and environmental engagement (e.g., the use of public and green spaces). In the context of HSR station areas, spatial vitality serves as a key indicator of how effectively transportation infrastructure integrates with urban life [12–16].

To explore what drives spatial vitality in HSR station areas, this study focuses on three categories of influencing factors—urban support capacity, station attributes, and built environment characteristics—based on a city–node–place analytical framework. These dimensions are selected because they collectively reflect the macro-level development background, the meso-level transportation functions of the station, and the micro-level spatial qualities experienced by users. Specifically, urban support indicators such as population size and urbanization level represent the demand base and economic potential of the host city. Station-level variables, including service frequency, location accessibility, and connectivity, determine how well the station functions as a mobility hub. Built environment indicators—such as commercial land use ratio, transportation facilities, and functional diversity—describe the spatial configuration and service provision in the immediate station area. By jointly analyzing these three levels, the study aims to reveal not only which factors significantly affect spatial vitality, but also why and how they do so under different conditions and in combination.

Understanding what drives this vitality is essential for enhancing land use efficiency, optimizing functional configurations, and promoting transit-oriented, sustainable development. More importantly, spatial vitality plays a fundamental role in improving the broader quality of urban life. Economically, vibrant station areas can stimulate commercial development and attract investment. Socially, they facilitate inclusiveness, accessibility, and human interaction. Environmentally, vital station areas can promote compact development, reduce vehicle dependence, and encourage the use of public and green spaces. Therefore, enhancing spatial vitality around HSR stations is not only a matter of functional efficiency, but also a strategic pathway toward inclusive, resilient, and environmentally sustainable cities.

Despite growing academic attention, current research on HSR station area vitality faces several notable limitations [17–19]. First, many studies adopt a static perspective, relying on instantaneous indicators such as heatmaps (e.g., Baidu thermal maps), which capture short-term activity snapshots but fail to reflect the cumulative vitality that develops over time—a key aspect for assessing the long-term spatial dynamics of station areas. Second, traditional statistical models typically assume linear relationships and struggle to capture complex nonlinear interactions or threshold effects among influencing variables, limiting their explanatory power in multifactorial urban contexts. Third, conventional data sources often involve survey-based or aggregated statistics, which lack spatial precision and temporal continuity. Such data mask local variations and individual behaviors within station areas. In contrast, emerging location-based big data, like social media check-ins, provide fine-grained, real-time insights, enabling more accurate mapping of human activity patterns and overcoming the coarse granularity and static nature of prior data.

With the emergence of geospatial big data and social media-based human mobility records, there is a new opportunity to conduct high-resolution analyses of spatial vitality [20–22]. In particular, the use of location-based check-in data provides valuable insights into real-time activity patterns, enabling more dynamic assessments of how people interact with station environments. Moreover, combining classical regression models with machine learning algorithms offers a robust framework to capture both the interpretable and nonlinear aspects of population behavior [23–27].

Recent academic efforts have increasingly leveraged multi-source data and advanced analytical methods to study the dynamics of HSR station areas. For instance, Wang et al. focused on the Yangtze River Delta, developing an integrated framework to assess urban vitality through indicators such as concentration, accessibility, livability, and functional diversity [6]. Yue et al. analyzed mobile phone signaling data in Jiangsu Province, clustering 71 HSR stations by passenger flow time series and employing geographically weighted multinomial logit models to examine the roles of entertainment POI density, population, GDP, and built area in shaping station classifications [9]. Additionally, recent research by Doan et al. applied machine learning models combined with SHAP (Shapley Additive Explanations) to explore the nonlinear and threshold effects of built environment, traffic, and air quality factors on urban vitality in Manhattan, using street-view pedestrian presence data [28]. Building on our earlier exploratory work, which employed panel data and a difference-in-differences (DID) approach to examine the temporal effects of HSR on commercial agglomeration, this study makes a significant conceptual and methodological advancement. We shift from a temporal to a spatial comparative perspective, analyzing the vitality mechanisms of 66 HSR station areas across 35 cities. Through interpretable machine learning and a city–node–place framework, we aim to uncover the multiscale drivers of station area vitality and provide actionable planning insights [29]. Together, these studies highlight the growing potential of integrating diverse datasets and interpretable computational models to uncover complex spatial patterns.

Therefore, this study aims to comprehensively examine the spatial vitality of HSR station areas across China by integrating multi-source data and applying both multiple linear regression (MLR) and Light Gradient Boosting Machine (LightGBM) models. Drawing on check-in data from Sina Weibo, urban statistical indicators, and detailed built environment metrics, we analyze 66 representative HSR stations in 35 cities to identify key drivers of population clustering. We further explore the threshold and interaction effects of influencing factors using SHAP (Shapley Additive Explanations) analysis [30,31].

To comprehensively understand the factors shaping spatial vitality in HSR station areas, this study adopts a city–node–place analytical framework and combines both traditional statistical models and advanced machine learning techniques. By integrating multi-source data including social media check-ins, statistical indicators, and built environment features, we aim to explore the nonlinear, threshold, and interaction effects underlying population clustering. The following sections detail the study area, data sources, variable design, modeling methods, and empirical results, culminating in theoretical discussions and policy implications.

2. Study Area and Data

2.1. Study Area

As of now, HSR services have been introduced in 269 cities across China, covering approximately 72.3% of all cities nationwide. To ensure the representativeness of our research, we selected 35 cities through a stratified sampling framework. The selection considered five key criteria: (1) regional diversity (eastern, central, western, and ethnic minority regions); (2) administrative level (including municipalities, provincial capitals, and sub-provincial cities); (3) urban scale and population density; (4) level of economic development and functional importance; and (5) the maturity and intensity of HSR operations. This approach ensures that the sample reflects China's diverse urban contexts and developmental stages.

From these cities, 66 representative HSR stations were selected based on comprehensive criteria, including station classification, operational scale, geographic location, service capacity, and data availability [11,29,32,33]. In order to avoid bias and maintain consistency, stations co-located with airports or lacking adequate data—such as Chengdu Xipu Station

and Guangzhou Baiyun Airport North Station—were excluded. We also avoided stations with weak service capabilities or low integration with the surrounding urban fabric. This filtering process ensures the analytical validity and robustness of our sample, making it suitable for comparative analysis.

The final sample includes a wide range of station types (e.g., central city hubs, regional gateways, peripheral nodes), covering diverse geographic and functional contexts. These stations span metropolitan cores such as Beijing and Shanghai, as well as key inland centers like Xi'an and Chengdu. This distribution enables a comprehensive understanding of how HSR integrates with varying urban structures [34,35]. Overall, the selected 66 stations in 35 cities provide a robust empirical foundation for analyzing spatial vitality in different socioeconomic and planning environments across China (Figure 1).



Figure 1. Study on the spatial distribution of HSR stations. Note: The original map (No. GS(2019)1686) is from the China National Natural Resources Standard Map Service website (<http://bzdt.ch.mnr.gov.cn/download.html?searchText=1686>, accessed on 5 June 2025). The authors mapped the spatial distribution of HSR stations based on it.

2.2. HSR Station Areas

In the existing literature, scholars have adopted various approaches to delineate the spatial extent of areas influenced by HSR stations, depending on their specific research objectives [36–42]. Among these, distance-based and time-based criteria are widely applied. For example, Schütz (1998) classified the surrounding areas into three development zones based on station accessibility: a 5–10 min access zone, a zone accessible within 15 min, and a zone beyond 15 min [43]. Building on this, Zheng et al. (2024) investigated development patterns and multi-level spatial interactions within a 1500 m radius around HSR stations, shedding light on the mechanisms of station area development and spatial structure [1]. Similarly, Wang Lan et al. (2014) proposed a concentric zoning approach based on spatial distance, dividing the station's influence area into a core zone (2000 m), an impact zone (4000 m), and a peripheral zone (8000 m) [44].

Findings from these studies indicate that the functional intensity of facilities around HSR stations generally decreases with increasing distance from the station, forming a layered, centripetal spatial structure [45–47]. Based on this insight, the present study

adopts a 1500 m fixed-radius buffer centered on each HSR station as the primary area of investigation. This decision is supported by both existing literature and the practical focus of this study. On one hand, the 1–2 km range is widely used in the existing literature as a typical zone of transit influence, ensuring methodological consistency. On the other hand, since this research focuses on spatial vitality, which is typically concentrated within walkable distances from the station—especially during the early and middle stages of development—a 1500 m range effectively captures the core activity zone.

Additionally, using a uniform fixed-radius buffer allows for standardized comparison across 66 HSR stations in 35 cities, enhancing the robustness of cross-sectional analysis. While network-based travel distances may offer a more realistic representation of accessibility, such data are often unavailable or inconsistent across large samples due to differences in local infrastructure quality and data availability.

2.3. Data Sources and Processing

Population agglomeration is a critical indicator for evaluating the vibrancy and economic consumption potential of HSR station areas [48,49]. Meanwhile, users' check-in behaviors at various spatial locations reflect their preferences for different types of activity spaces. Therefore, to gain deeper insights into the spatial behavior patterns of individuals within HSR station areas, this study employed a web crawler to collect check-in data from Sina Weibo within defined station boundaries.

The acquired check-in data were first matched with corresponding spatial coordinates. Subsequently, the spatial locations were categorized based on a point-of-interest (POI) classification system. This processing enabled the identification of the spatial distribution patterns of active individuals and the characterization of activity hotspots within the station areas.

Notably, Sina Weibo initially provided an open API for check-in data access in 2012 [50–52]. However, the original interface was later discontinued due to overwhelming access traffic. In response to this limitation, this study improved upon previous methods by developing a targeted data acquisition strategy. Specifically, we conducted a detailed search of the Sina Weibo source pages, applied precise filtering techniques, and utilized both the location service API and user service API in an integrated manner. This optimized workflow significantly enhanced both the accuracy and efficiency of data collection, thereby laying a solid foundation for subsequent analysis.

To ensure the integrity of the dataset, we implemented a multi-stage preprocessing workflow. First, we removed duplicate entries based on user ID, timestamp, and spatial coordinates. Second, we filtered out invalid check-ins lacking essential attributes such as geolocation or content. Third, we eliminated advertising-related records using keyword-based content filtering to exclude promotional or irrelevant posts. These steps collectively ensured that only high-quality, user-generated activity records were retained for analysis.

The data acquisition process involved several key steps (Figure 2): identifying available interface paths, constructing query addresses for HSR station areas, sending requests to retrieve POI collections, iteratively accessing URLs and parsing the returned JSON files, and finally, performing data validation and storage. The collected check-in data covered the period up to 31 December 2024, and each entry was precisely matched with the latitude and longitude of its corresponding HSR station area (Table 1).

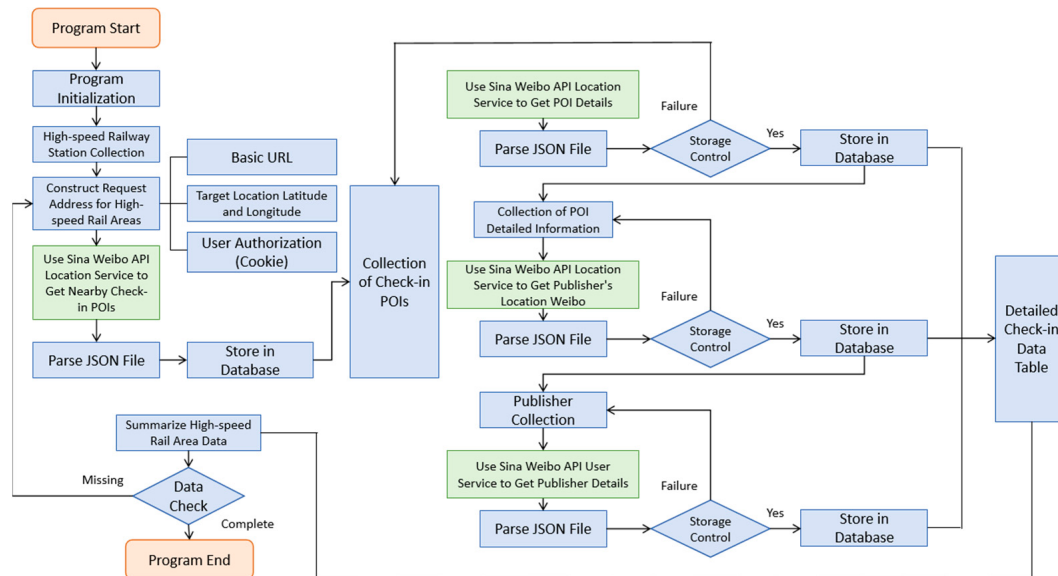


Figure 2. Workflow for acquiring check-in data from Sina Weibo.

Table 1. Sample records from Weibo check-in dataset.

Field Name	Data Type	Description	Example
record_index	int	Sequential index of the data record	1
station_name	string	Name of HSR station	Shanghai Hongqiao Station
poi_name	string	Name of the point of interest (POI)	Haidilao Hotpot (Zhouhong Road Branch)
poi_address	string	Full address of the POI	799 Shenchang Road, Minhang District, Shanghai
checkin_frequency	int	Number of check-in occurrences	126
poi_category	string	Category/type of the POI	Hotpot Restaurant
longitude	double	Geographic longitude of the POI	121.30929749345
latitude	double	Geographic latitude of the POI	31.193352869547752

2.4. Variables

The influencing factors of population agglomeration represent a multidimensional and complex research domain, involving the interaction of various spatial and functional components. At the urban level, the capacity of a city to support population concentration serves as a fundamental condition. This capacity encompasses urban scale, economic development, transportation network structure, and the orientation of planning policies (Table 2). Together, these factors determine the city's attractiveness and its ability to accommodate large populations [53–55].

In selecting urban support variables, we prioritized those with solid theoretical foundations and empirical support in the field of urban vitality research. Specifically, we identified “Population Scale,” “Economic Development,” and “Transportation Network” as core dimensions, as they comprehensively reflect a city's development stage and population attraction capacity. These macro-level factors play a fundamental role in determining the intensity of population activity and the spatial vitality of HSR station areas.

First, the scale and level of economic development define the availability of employment, education, and living resources, which directly enhance the city's capacity to attract people. Second, the layout and efficiency of the transportation network facilitate the flow

of people and goods, making the city a key node in regional mobility. Additionally, urban planning strategies and policy support optimize resource allocation, thus further reinforcing population concentration. The data for these indicators are primarily derived from municipal statistical yearbooks, government bulletins, and open-access databases.

Table 2. Urban support variables.

Category	Dimension	Variables
Urban Support (C)	Population Scale (C-POP)	Population of the city where the HSR station is located (C1), population of the district/county (C2)
	Economic Development (C-ED)	City GDP (C3), district GDP (C4), tertiary industry GDP of the city (C5) and district (C6), urbanization rate of the city (C7) and district (C8), general public budget revenue of the city (C9) and district (C10)
	Transportation Network (C-TN)	Road network density of the city (C11) and district (C12), city-level HSR connection frequency index (C13), one-stop reachable cities index (C14), direct HSR link index (C15), city-level degree centrality (C16), city-level closeness centrality (C17), city-level betweenness centrality (C18), city-level eigenvector centrality (C19)

At the node level, the characteristics of HSR stations also exert a significant influence on the degree of population aggregation. Factors such as the rationality of station location, the suitability of its design and scale, and the completeness of supporting facilities jointly determine the attractiveness of the station as a transportation hub [56,57]. Stations located in central urban areas or major traffic corridors are more likely to attract large flows of people. A well-organized spatial layout and functional zoning can enhance user experience, while high-quality supporting facilities—including waiting areas, commercial amenities, and accessibility features—encourage frequent use and promote further agglomeration. Data on HSR station characteristics are sourced from official railway reports, open databases, and transportation planning documents. Table 3 presents the variables used to assess the HSR station conditions.

Table 3. HSR station condition variables.

Category	Dimension	Variables
HSR Station Condition (S)	Opening Time and Level (S-TL)	HSR service start time (S1), station classification level (S2)
	Station Location (S-Z)	Number of HSR connections (S3), number of one-stop connections (S4), number of direct links (S5), station-level degree centrality (S6), station-level closeness centrality (S7), station-level betweenness centrality (S8), station-level eigenvector centrality (S9), distance to the city center (S10), location index (S11), public transit time (S12), driving time (S13), total train services (S14), originating/terminating services (S15), percentage of originating/terminating services (S16), percentage of through services (S17)
	Station Scale and Design (S-SD)	Number of platforms (S18), number of rail lines (S19), station building area (S20), square area in front of the station (S21)
	Connecting Transport (S-JT)	Number of metro entrances (S22), metro lines (S23), bus stops (S24), bus lines (S25), long-distance bus stations (S26)
	Commercial and Leisure Facilities (S-CF)	Number of commercial and leisure facilities (S27)

At the place level, the built environment of the station area serves as the immediate spatial carrier for population activity and thus plays a key role in influencing agglomeration [16,58–61]. Core dimensions of the built environment include spatial density, functional diversity, design quality, transportation accessibility, and facility availability. Appropriate population density ensures urban vitality; functional diversity refers to the mix of land uses, commercial types, and social activities; and high-quality design enhances visual appeal and usability. Furthermore, the accessibility of the station area to the broader urban context—through both physical connections and service coverage—affects whether people choose it as a destination for travel or daily activity. Relevant data are collected from urban planning documents, remote sensing imagery, and POI data from Gaode Maps. Table 4 presents the built environment variables of station areas.

Table 4. Station areas built environment variables.

Category	Dimension	Variables
Station Area Built Environment (E)	Development Scale (E-DS)	Building density (E1), floor area ratio (E2), development proportion (E3), underground space density (E4)
	Land Use (E-LU)	Proportion of residential land (E5), commercial/service land (E6), public administration/service land (E7), green space (E8), transport infrastructure (E9), public utility land (E10), land use mix index (E11)
	Spatial Design (E-SD)	Road network density (E12), number of intersections (E13), building proximity index (E14), green view index (E15), sky openness (E16), enclosure index (E17)
	Transportation Connectivity (E-TC)	Number of metro entrances (E18), number of metro lines (E19), number of bus stops (E20), number of bus lines (E21), number of parking facilities (E22)
	Functional Facilities (E-FF)	Number of commercial service facilities (E23), residential facilities (E24), public facilities (E25), business facilities (E26)

3. Methods

3.1. Multiple Linear Regression Model

The Multiple Linear Regression (MLR) model is a widely used statistical method that explores the linear relationship between a dependent variable and multiple independent variables. In the context of high-speed rail (HSR) station area studies, this model is particularly valuable as it enables the quantification of how various factors—such as urban scale, transportation networks, and station-area facilities—affect spatial vitality.

Specifically, the MLR model helps identify and evaluate the extent to which different urban and infrastructural variables contribute to population agglomeration and activity intensity around HSR stations. By establishing a regression equation, it becomes possible to measure the individual contribution of each explanatory variable to the dependent variable and to reveal the underlying mechanisms that drive spatial dynamics.

In this study, we adopt the MLR approach by using spatial vitality in HSR station areas as the dependent variable and a set of multidimensional influencing factors as independent variables. The basic mathematical form of the model is expressed as follows:

$$\text{Vitality} \left(\text{check-ins/km}^2 \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad (1)$$

In this equation, Vitality (check-ins/km²) denotes the dependent variable, representing the intensity of human activity per unit area within the 1500 m buffer zone of each HSR station during its operational period. The variables $X_1, X_2, \dots, X_n$ are the independent

variables, corresponding to influencing factors such as transportation accessibility, functional diversity, and station-area infrastructure. β_0 is the intercept term, and $\beta_1, \beta_2, \dots, \beta_n$ are the regression coefficients, each representing the marginal effect of a corresponding independent variable on the dependent variable. The term ϵ captures the random error, accounting for unexplained variability in the model.

Each regression coefficient has a clear interpretive meaning. For instance, the coefficient β_1 indicates the magnitude and direction of change in spatial vitality Y when the corresponding factor X_1 increases by one unit, holding all other variables constant. This allows for a detailed understanding of the relative importance of different influencing factors.

3.2. LightGBM Model

Unlike traditional multiple linear regression models, LightGBM (Light Gradient Boosting Machine) is an ensemble learning method based on the gradient boosting framework. It builds multiple decision trees to model data and iteratively optimizes the objective function to enhance prediction accuracy. As a result, LightGBM demonstrates significant advantages in handling complex nonlinear relationships and large-scale datasets, making it an effective tool for uncovering hidden patterns and influential features in data.

The core algorithm underlying LightGBM is the Gradient Boosting Decision Tree (GBDT). The main idea is to construct a strong predictive model by sequentially training a series of weak learners, typically shallow decision trees, and minimizing the loss function in each iteration.

The general form of the loss function is defined as

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i) + \Omega(T) \quad (2)$$

In this expression, L represents the total loss, y_i is the actual value of the i -th sample, $\hat{y}_i$ is the corresponding predicted value, $l(\cdot)$ denotes the loss function for an individual sample (e.g., squared error or log loss), and $\Omega(f)$ is a regularization term that controls the complexity of the model, such as tree depth or the number of leaf nodes.

During each iteration, LightGBM leverages both the first-order gradient (i.e., the derivative of the loss function with respect to the prediction) and the second-order gradient (i.e., the second derivative) to update the model. This use of gradient information enables LightGBM to improve computational efficiency and convergence speed while maintaining high prediction accuracy.

To further improve training speed and reduce memory consumption, LightGBM introduces a histogram-based algorithm to construct decision trees (Figure 3). Instead of evaluating every possible split point, LightGBM first buckets continuous feature values into discrete bins (i.e., histograms). Then it finds the optimal split based on these bins, significantly reducing the number of comparisons needed during training.

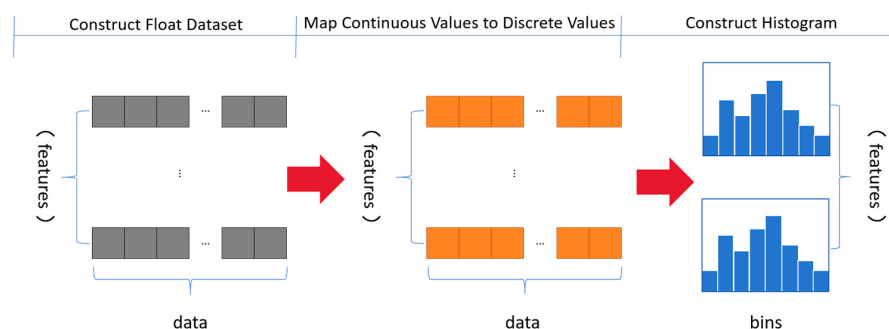


Figure 3. Illustration of LightGBM histogram optimization principle.

3.3. Shapley Additive Explanation (SHAP)

Although LightGBM and other machine learning methods can produce highly accurate predictions, their “black-box” nature often makes it difficult to understand how the models arrive at specific decisions. To address this issue, SHAP (Shapley Additive Explanations) was developed as a model-agnostic interpretability framework that offers transparent explanations for complex machine learning models.

The SHAP model is theoretically grounded in the Shapley value from cooperative game theory. It quantifies the contribution of each feature to the prediction by computing its marginal effect across all possible feature combinations. Therefore, SHAP not only maintains the predictive power of advanced models but also reveals the underlying mechanisms driving their outputs.

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

In this equation, ϕ_i denotes the Shapley value of feature i , representing its contribution to the model’s prediction. N is the set of all features, and S is any subset that does not include feature i . The function $f(S)$ refers to the model output using only the features in subset S , while $f(S \cup \{i\})$ refers to the output when feature i is added to the subset.

By averaging the marginal contributions over all possible subsets, the Shapley value provides a fair and comprehensive measure of feature importance. As a result, SHAP serves as a powerful tool for enhancing the interpretability of machine learning models without sacrificing predictive accuracy.

3.4. Model Selection and Comparative Justification

Given the heterogeneous and high-dimensional nature of the data used in this study—including structured socioeconomic indicators, spatial POI distributions, and user-generated check-in behaviors—we adopt a dual-modeling strategy using both Multiple Linear Regression (MLR) and Light Gradient Boosting Machine (LightGBM).

The MLR model provides robust interpretability and is well suited for structured variables with assumed linear effects. It supports inference of marginal impacts and is ideal for testing policy-relevant hypotheses. However, it falls short in capturing complex nonlinearity and variable interactions.

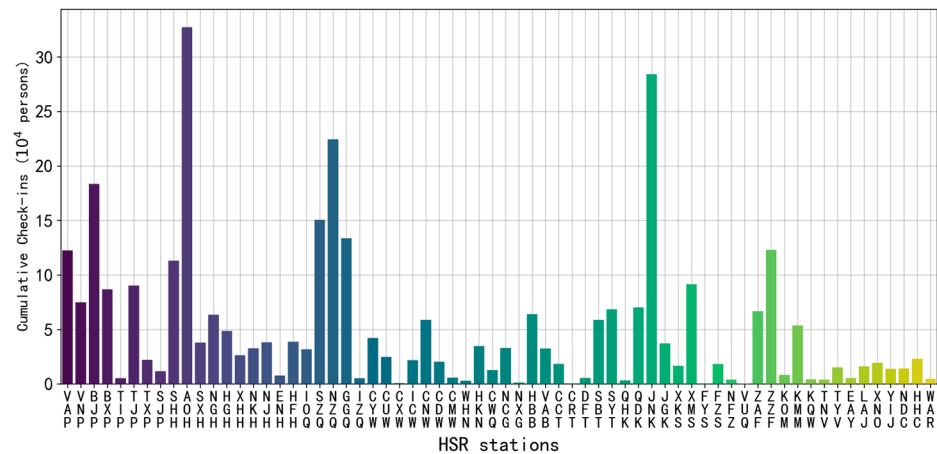
LightGBM, on the other hand, excels in handling nonlinearities, feature interactions, and sparse behavioral data—common in social sensing applications. It is especially valuable for uncovering threshold effects and high-order dependencies, albeit at the cost of interpretability. This is addressed by incorporating SHAP.

Therefore, the combined use of MLR and LightGBM allows us to bridge explanatory clarity and analytical depth, leading to a more comprehensive understanding of what drives spatial vitality in HSR station areas.

4. Results

4.1. Spatiotemporal Variations in the Vitality of HSR Station Areas

From the cumulative check-in bar chart, it is evident that human activity in HSR station areas exhibits a significant spatial clustering pattern, particularly concentrated in economically developed metropolitan regions (Figure 4). For instance, Shanghai Hongqiao Station, Futian Station, Beijing Station, and Shenzhen Station rank among the top nationwide in terms of cumulative check-ins, reaching 327,000, 224,000, 183,000, and 151,000 check-ins, respectively. Notably, Shanghai Hongqiao Station recorded the highest number of check-ins, highlighting its strong attractiveness as a core transportation hub.



In addition, stations such as Jinan Station (284,000), Guangzhou East Station (134,000), Zhengzhou Station (123,000), and Shanghai Station (113,000) also recorded high levels of activity. These stations are typically located near city centers and benefit from convenient transportation links, which effectively attract large crowds. It is also worth noting that popular tourist destinations like Xiamen Station have drawn substantial numbers of active users, indicating that HSR station areas also possess strong tourism appeal.

Figure 4 illustrates the distribution of cumulative check-ins across different HSR station areas. Based on functional attributes, the active spaces surrounding the stations can be categorized into commercial service space, business space, public space, leisure space, transportation space, residential space, green space, industrial space, and administrative space. The activity levels in each space type were analyzed (Figure 5). Among them, commercial service spaces exhibited the highest level of activity, with approximately 1.37 million check-ins, reflecting their strong commercial attractiveness. Transportation and residential spaces followed, with 610,000 and 600,000 check-ins, respectively. This pattern not only underscores the critical role of HSR stations as transportation hubs but also demonstrates the effective development of residential functions in adjacent areas.

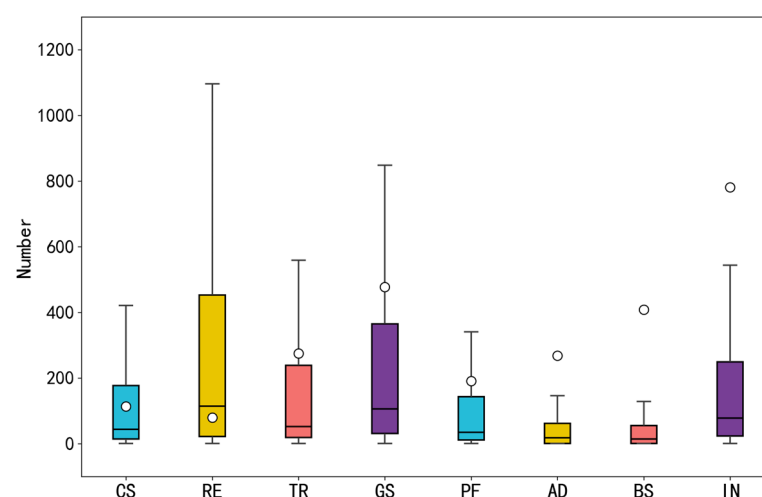


Figure 5. Number of check-ins in different types of spaces around high-speed rail stations. Note: For clarity and brevity, space types are represented using abbreviations: GS stands for Green Space, CS for Commercial Services, PF for Public Facilities, TR for Transportation, RE for Residential, AD for Administrative, BS for Business, and IN for Industrial. These abbreviations correspond to major urban functional zones.

Furthermore, public and green spaces recorded 420,000 and 160,000 check-ins, respectively, showing a certain degree of vitality. This suggests that in addition to transportation and commercial services, HSR station areas also provide public amenities and green leisure environments that meet the diverse needs of residents and travelers.

According to the spatial distribution of check-in data, the activity patterns of populations in HSR station areas can be generally categorized into three types: balanced distribution, fan-shaped clustering, and star-shaped dispersion (Figure 6). First, stations with a balanced distribution are typically located near city centers. These areas tend to develop evenly, with commercial and business functions highly concentrated, which generates strong attractiveness and leads to relatively uniform population aggregation. Typical examples include Beijing Station, Tianjin Station, Shanghai Station, Futian Station, Guangzhou East Station, Chengdu South Station, and Zhengzhou Station. These stations represent the typical development pattern of central urban HSR stations.

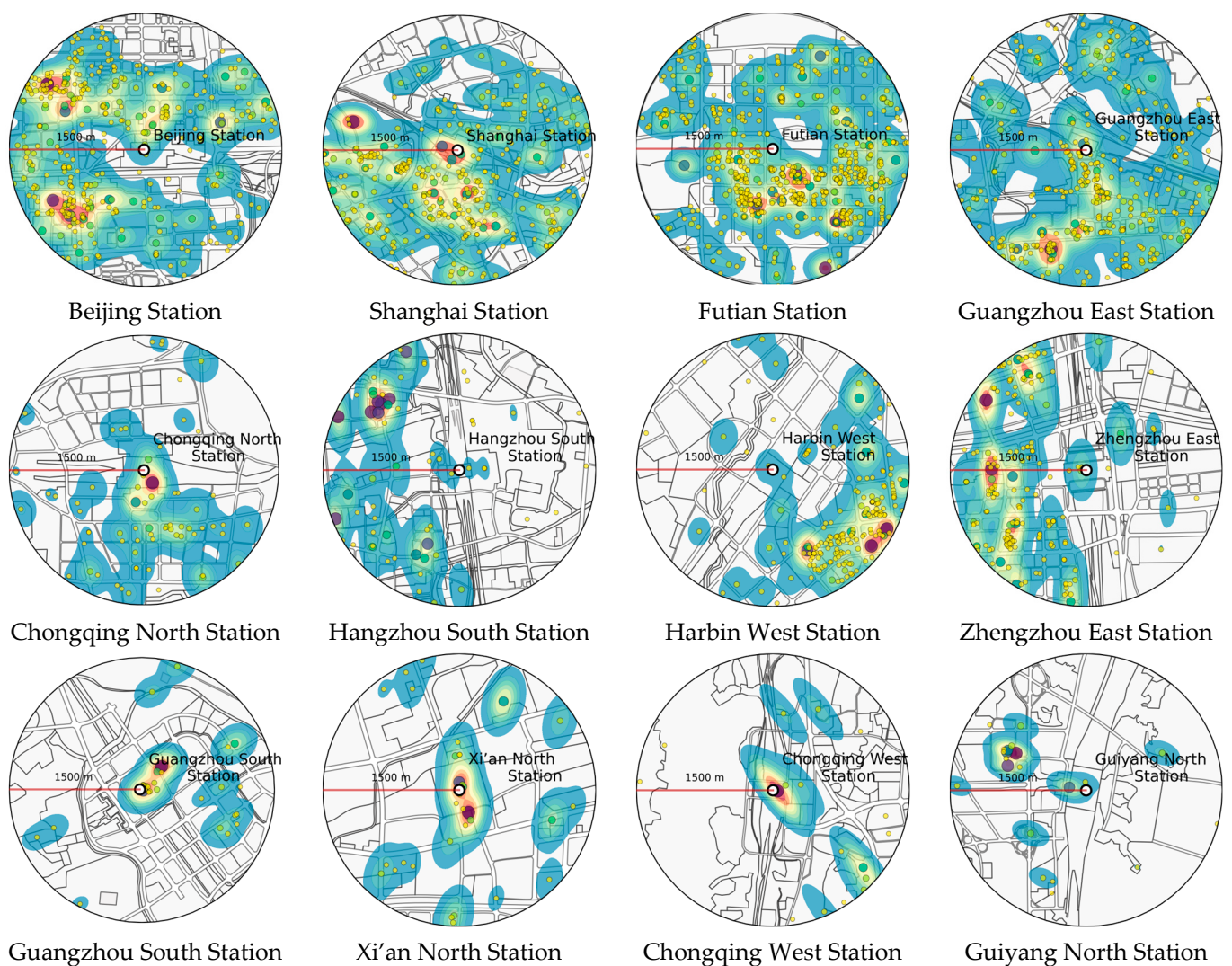


Figure 6. Spatial distribution map of crowd check-in activities in HSR station areas.

Second, fan-shaped clustering occurs in HSR stations that expand along existing urban development axes. Influenced by the built environment, these station areas extend in a directional pattern that mirrors urban spatial growth. Stations such as Chongqing North, Hangzhou South, Harbin West, and Zhengzhou East fall into this category, and their development aligns closely with the city's expansion trajectory.

Lastly, star-shaped dispersion is observed in HSR stations located farther from urban centers. These stations often feature undeveloped land or lack adequate supporting facilities, resulting in weaker integration with the city and a more scattered pattern of population activity. Despite their current lower levels of aggregation, these stations hold the potential to become key nodes in future urban development. Representative examples include Guangzhou South, Xi'an North, Chongqing West, Guiyang North, Kunming South, and Changchun West stations.

4.2. Results of Linear Regression

In this study, all explanatory variables were subjected to a multicollinearity test. The results showed that the Variance Inflation Factors (VIFs) for all variables were below 5, indicating that multicollinearity among the variables was not significant and the model estimates are statistically reliable. The MLR model achieved an R-squared value of 0.3975, a Root Mean Squared Error (RMSE) of 5.014, and a Mean Absolute Error (MAE) of 3.712. Table 5 presents the results of the MLR model.

Table 5. Results of the multiple linear regression model.

Variable Code	Variable Description	Coef.	Std Err.	t	p-Value
C1	Population size of the HSR station city	7.0798	3.017	2.346	0.022 *
C7	Urbanization level of the HSR station city	8.0828	3.646	2.217	0.03 **
S11	Station location index	−9.5416	2.821	−3.383	0.001 ***
S15	Number of originating and terminating train services	6.5358	2.996	2.181	0.033 **
S24	Number of bus stops near the station	12.4662	3.202	3.893	0.001 ***
E6	Land area for commercial service facilities	9.9389	4.107	2.42	0.018 **
E22	Number of parking facilities in the station area	8.7181	3.224	2.704	0.009 ***

Note: Significance levels—* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Several factors related to urban support capacity exhibited significant effects on the spatial clustering of active populations. Specifically, both the population size of the city (C1) and the level of urbanization (C7) were strongly correlated with the degree of population aggregation, showing statistically significant positive relationships. The regression coefficient for city population size (C1) was 7.0798 with a p -value of 0.022, suggesting that larger urban populations tend to attract more active individuals to gather around HSR station areas. Similarly, the urbanization level (C7) had a coefficient of 8.0828 and a p -value of 0.03, indicating that more urbanized environments promote higher levels of population clustering.

Regarding station-level characteristics, the station location index (S11) had a significant negative effect on population aggregation, with a coefficient of −9.5416 and a p -value of 0.001. This result implies that stations with lower location index values—i.e., more favorable or central locations—tend to attract higher levels of population activity. In contrast, the number of originating and terminating train services (S15) positively influenced population clustering, with a coefficient of 6.5358 and a p -value of 0.033, demonstrating that increased service frequency enhances the station's ability to attract people. Additionally, the number of bus stops near the station (S24) also had a statistically significant positive effect, with a coefficient of 12.4662 and a p -value of 0.001, underscoring the importance of public transportation connectivity in reinforcing population gathering.

Furthermore, several elements of the built environment around the station area were found to significantly influence the level of population activity. The area designated for commercial service facilities (E6) had a positive and significant impact, with a coefficient of 9.9389 and a p -value of 0.018. This suggests that expanding commercial land use

around stations can substantially boost their attractiveness and crowd-gathering capacity. Similarly, the number of parking facilities (E22) had a positive and significant effect, with a coefficient of 8.7181 and a p -value of 0.009, indicating that adequate parking infrastructure supports both accessibility and functional completeness, thereby contributing to greater population clustering.

These findings align with urban economic theory, which suggests that larger urban populations create stronger consumption bases and attract higher levels of pedestrian activity. The positive effect of commercial land use reflects the principle of agglomeration economies, where mixed-use developments increase land efficiency and social interaction.

In contrast, the negative effect of the station location index suggests that stations located in peripheral or poorly integrated zones suffer from low spatial synergy, as predicted by urban spatial mismatch theory. The accessibility factor, such as the number of bus stops, supports the notion that multi-modal connections enhance last-mile connectivity, which is a cornerstone of transit-oriented development (TOD).

4.3. Results of the LightGBM Regression Model and SHAP Analysis

According to the convergence evaluation plot of the active population prediction model, the model's loss value showed a significant decrease during the initial phase of training (Figure 7). In particular, within the first 50 iterations, the loss dropped sharply from a high value near 7.0 to approximately 5.75. This indicates that the optimization algorithm quickly adjusted the hyperparameters in the early training stage and substantially reduced the prediction error. As the number of iterations increased, the rate of loss reduction gradually slowed, and the curve began to flatten. After around 250 iterations, the loss value stabilized at approximately 5.2, suggesting that further training did not lead to noticeable improvement, and the model reached convergence with stable performance.

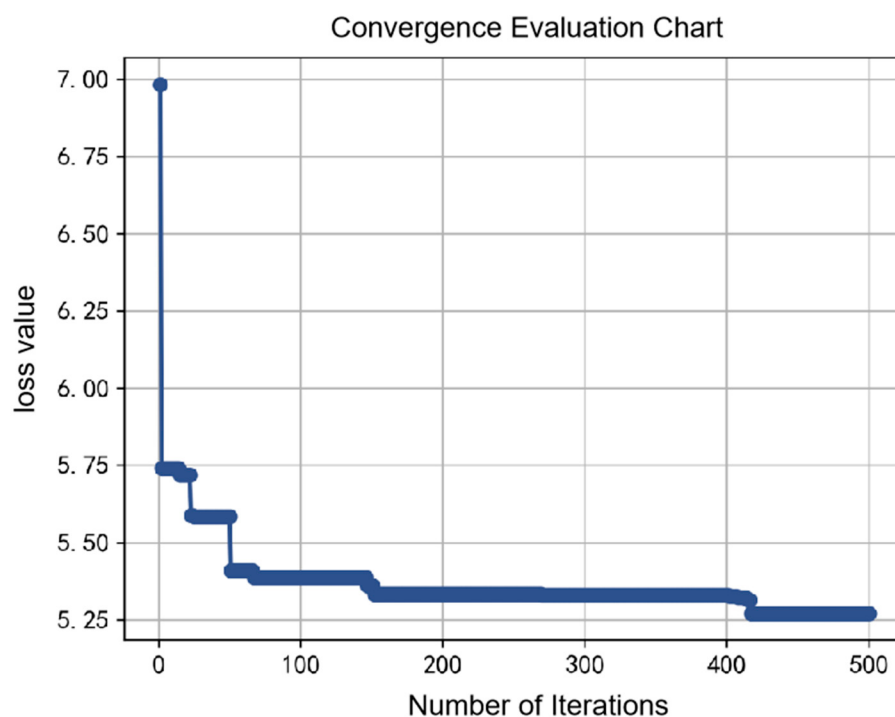


Figure 7. Convergence evaluation chart of the crowd aggregation model in HSR station areas.

Based on the hyperparameter tuning results, a scatterplot matrix was used to visualize the interactions between parameters and their impact on model performance (Figure 8). The selected optimal parameter configuration achieved a well-balanced trade-off between model fitting and generalization. Specifically, the learning rate was set to 0.01 to ensure

stable optimization and reduce the risk of overfitting. The number of trees was set to 1041 to improve accuracy without introducing excessive computational burden. The maximum tree depth was limited to 10, effectively controlling model complexity and enhancing generalization capability. Each tree was allowed a maximum of 15 leaf nodes, increasing fitting capacity while mitigating overfitting. The subsample ratio was set at 0.648, helping introduce randomness and reduce overfitting. Additionally, the feature subset ratio was set to 0.624, which further enhanced model diversity. Lastly, the minimum number of samples per leaf was set to 10, ensuring node stability and robust decision splits. Collectively, these hyperparameter settings enabled the model to achieve strong and stable performance in predicting active population clustering. The LightGBM model achieved an R-squared value of 0.7347, a Root Mean Squared Error (RMSE) of 3.3273, and a Mean Absolute Error (MAE) of 1.9621. Figure 9. Learning curve of the LightGBM model showing the loss trajectories for both the training and validation sets. As the number of iterations increases, both losses decrease and gradually stabilize, with no upward trend in validation loss—indicating good model convergence and no significant overfitting.

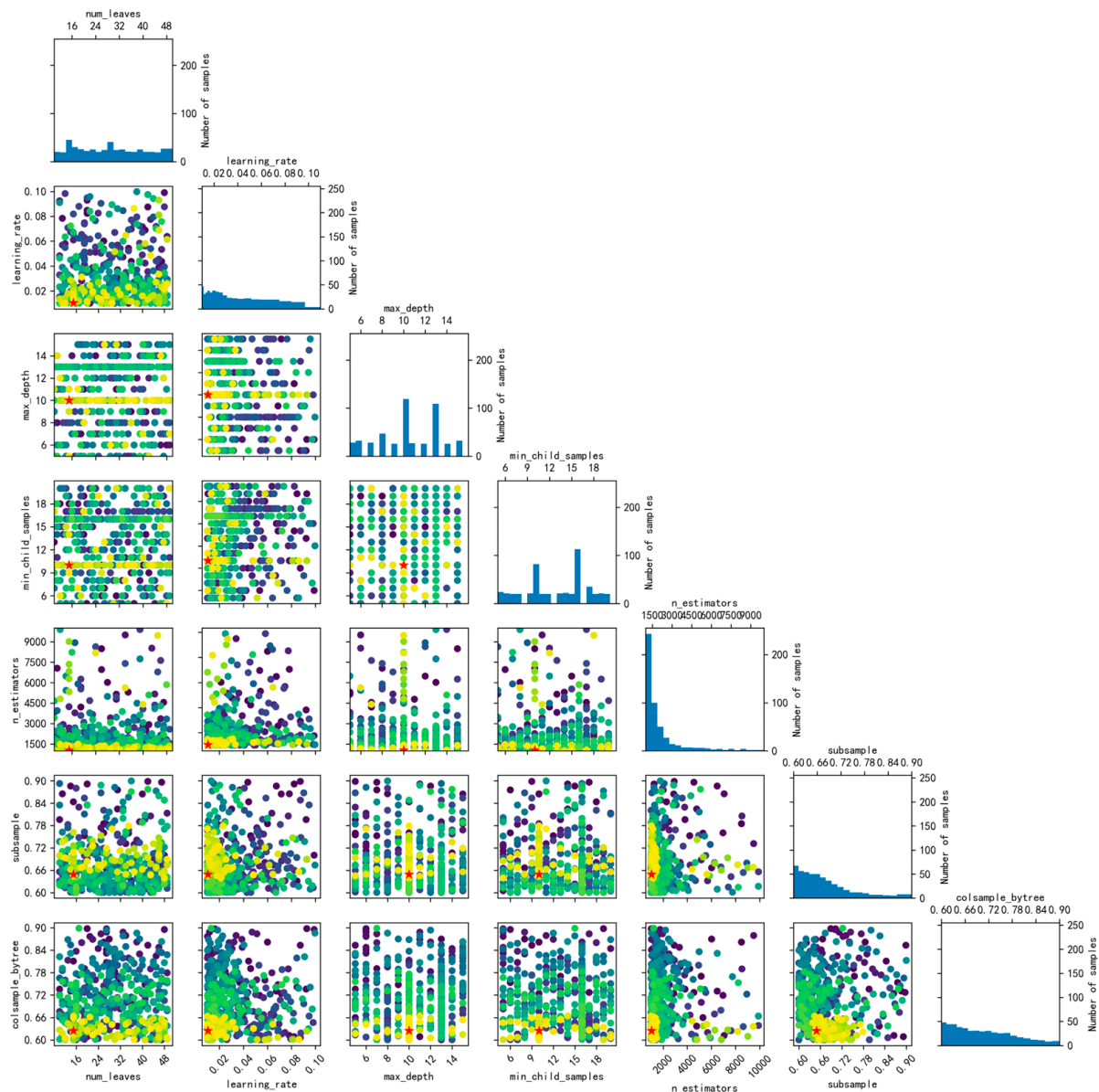


Figure 8. Model parameter optimization.

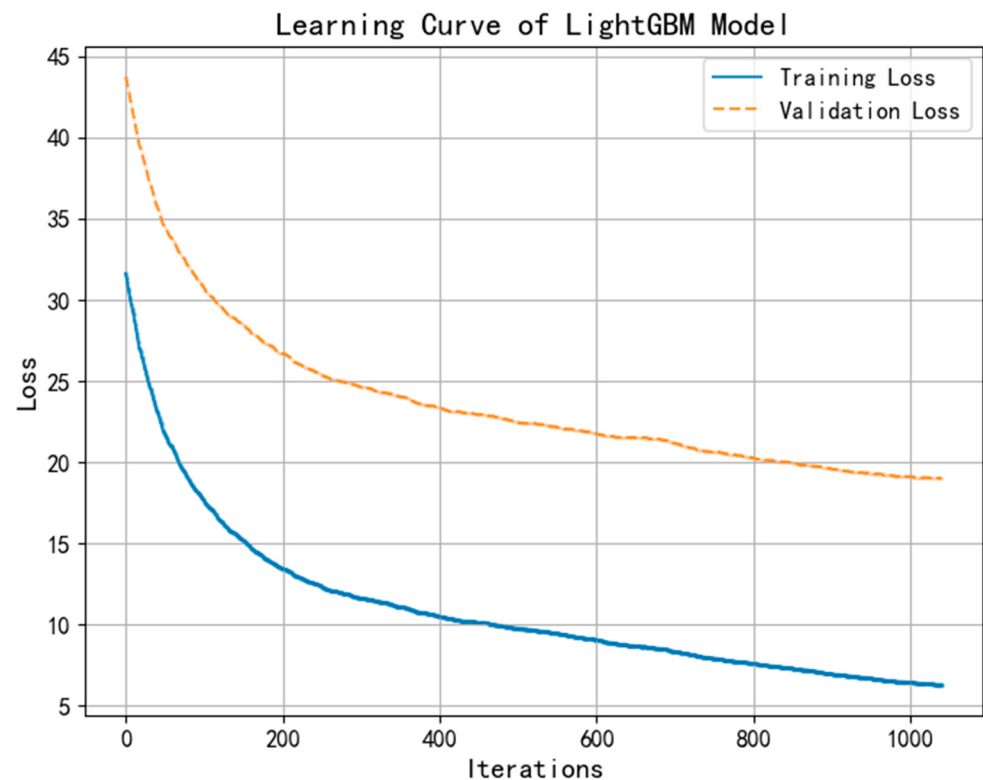


Figure 9. Learning curve of the LightGBM model.

After training the model, SHAP analysis was conducted to interpret the contribution of each input variable to the prediction outcomes. The resulting variable importance ranking (Figure 10) and SHAP value distribution plots (Figure 11) revealed that the factors influencing active population clustering differed from those affecting passenger flow or dwellers. In particular, active population distribution more strongly reflected perceptions of spatial design, economic vitality, and local consumption potential in HSR station areas.

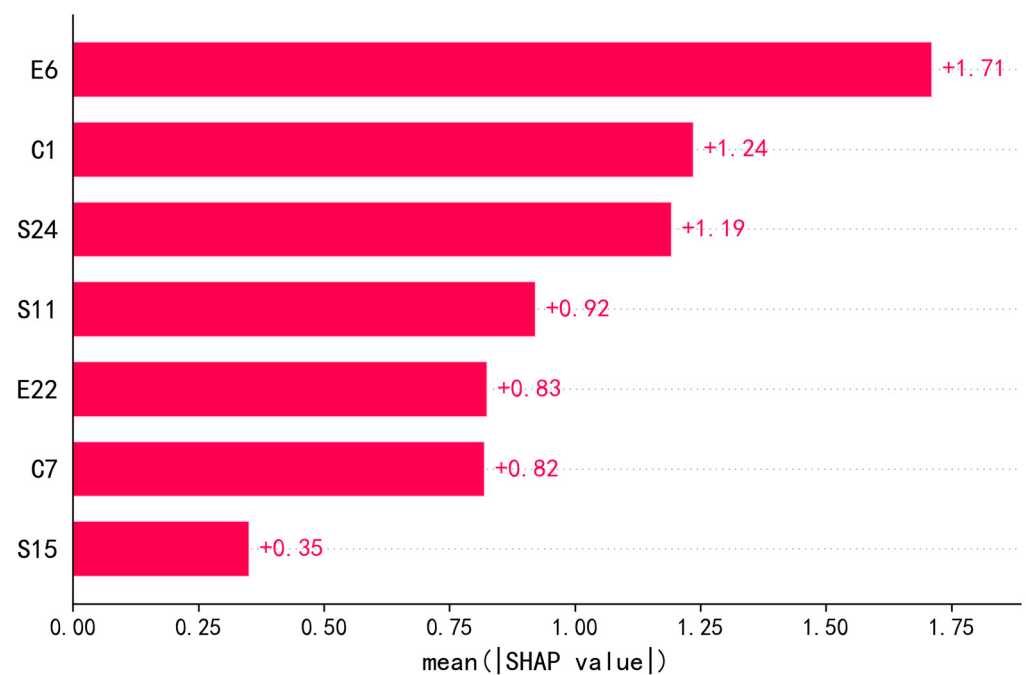


Figure 10. Feature importance ranking based on mean SHAP values.

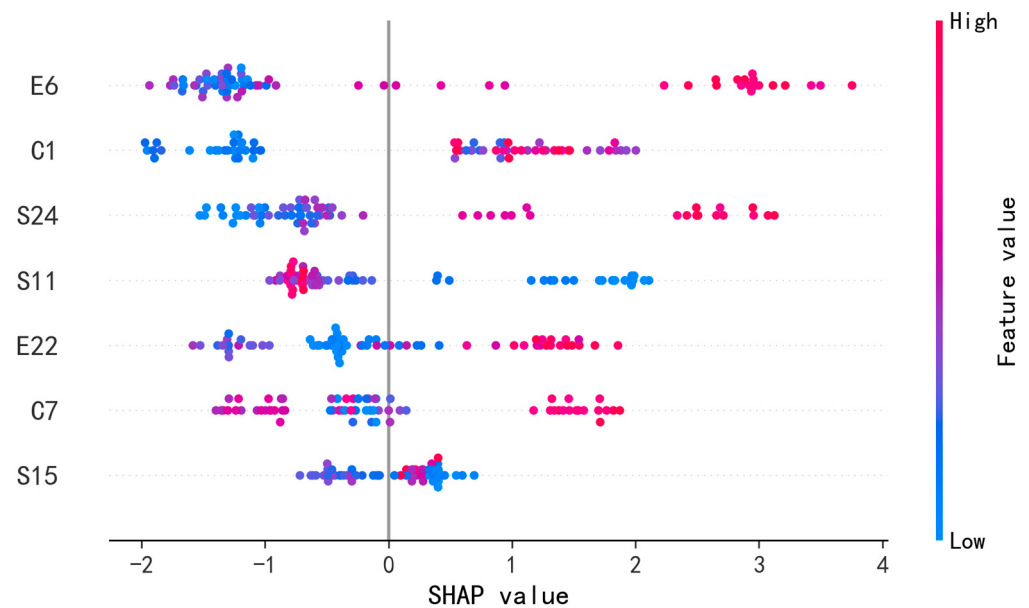


Figure 11. SHAP summary plot showing feature effects and value distributions.

Among all features, the proportion of land used for commercial service facilities within the station area (E6) exhibited the highest SHAP value of +1.71, indicating it had the most substantial positive impact on active population clustering. In addition, the population size (C1) and urbanization level (C7) of the HSR station's host city had SHAP values of +1.24 and +0.92, respectively, suggesting that both city scale and development level significantly promote human activity near the stations. The station location index (S11) also showed a notable SHAP value of +0.92, confirming that favorable geographic positioning and transit hub advantages play a key role in attracting people. Moreover, the number of bus stops near the station (S24) and the number of parking facilities in the station area (E22) had SHAP values of +0.83 and +0.82, respectively, underscoring the importance of well-developed transit infrastructure in enhancing population aggregation.

4.4. Threshold Effects

The primary purpose of the univariate driving model is to examine the independent effect of a single variable on population clustering, thereby identifying and understanding the specific degree and pattern of influence each factor has on high-speed rail (HSR) station areas. Through the quantitative assessment of each variable, the model reveals how these factors operate under different conditions and how their effects may change across value ranges.

The results indicate a strong relationship between commercial facilities and population clustering in HSR station areas. Specifically, the partial dependence plot for the proportion of commercial service land use in the station area (E6) shows that when the share increases from a low baseline to around 12%, the predicted value of active population rises sharply. However, once the proportion exceeds 15%, the predicted value tends to stabilize (Figure 12). This pattern reflects the classical economic law of diminishing marginal utility and suggests that while moderate commercial land allocation enhances attractiveness and pedestrian density, excessive concentration may cause functional redundancy, spatial congestion, or underutilization—consistent with the “resource saturation” theory in urban land use planning. This finding suggests that a moderate increase in commercial land use can significantly enhance the vibrancy of the station area by attracting more consumers and stayers. Therefore, in planning areas surrounding HSR stations, commercial, cul-

tural, and office functions should be strategically allocated to match actual demand, thus strengthening the overall attractiveness and functional diversity of the district.

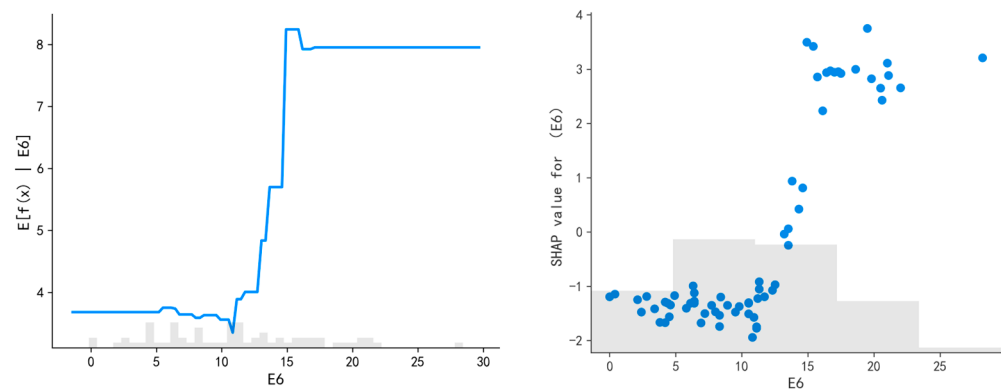


Figure 12. Partial dependence plot for feature E6.

In addition, the station location index (S11) exhibits a significant negative correlation with the predicted value of active population clustering (Figure 13). The partial dependence plot shows that when S11 is low—indicating a superior spatial location within the city—the predicted population level is high. As S11 increases, the predicted value drops rapidly and levels off, particularly after the value exceeds 0.3. This result highlights the critical role of spatial positioning in attracting human activity. And then levels off, reinforcing the spatial mismatch theory which states that urban facilities located far from core areas often suffer from reduced accessibility and weaker socio-economic integration. Consequently, the planning and development of HSR station areas should prioritize their integration within the broader urban spatial framework. Strengthening connectivity with central urban zones and major functional areas, improving transport accessibility, and ensuring the availability of high-quality public services can collectively enhance the locational advantages of HSR stations, thereby promoting greater population clustering and urban vitality.

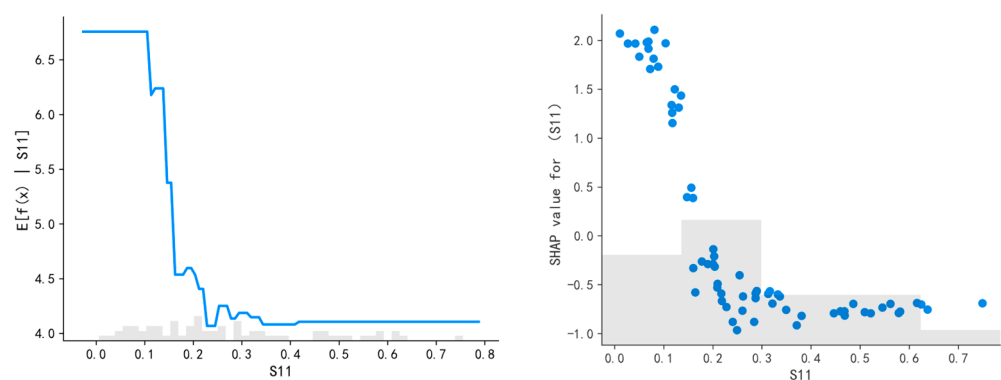


Figure 13. Partial dependence plot for feature S11.

Furthermore, the number of bus stops (S24) near the station has a clear and positive effect on population activity (Figure 14). According to the partial dependence plot, as the number of bus stops increases, the predicted value of active population rises notably—especially after the number surpasses 10, at which point the prediction exhibits a step-like increase and gradually stabilizes. This pattern confirms that a well-developed bus transfer system is a key factor in enhancing the vitality of HSR station areas. Increasing the number of bus stops not only improves transportation accessibility but also expands the service coverage of the station area, thus attracting more pedestrian flow and encouraging longer stays. Accordingly, future planning of station-area transit infrastructure should prioritize

the density and coverage of the bus network, support multimodal integration, and promote the construction of efficient and convenient public transportation systems to better support human activity and sustainable station-area development.

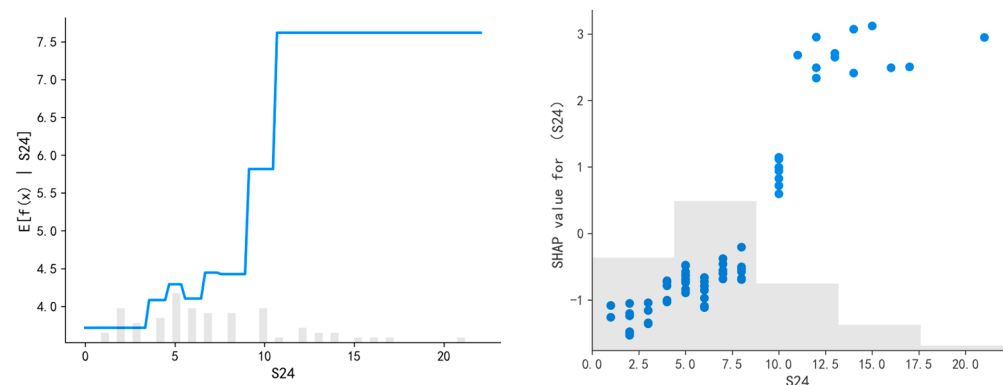


Figure 14. Partial dependence plot for feature S24.

These patterns reflect the classical economic concept of diminishing marginal returns, which states that beyond a certain threshold, the additional benefit of increasing an input (e.g., commercial land ratio or bus stop density) decreases.

From a theoretical standpoint, the observed turning points align with threshold theory, which emphasizes the presence of critical values beyond which system responses change nonlinearly. Additionally, the phenomenon is related to resource saturation theory, where excessive infrastructure or functional allocation may exceed local demand or spatial capacity, leading to diminishing efficiency or even congestion effects.

These findings suggest that urban planning around HSR stations must consider not only increasing supply but also optimizing it to avoid over-concentration and spatial redundancy.

4.5. Interaction Effects

The main objective of the multi-factor interaction model is to assess the combined influence of multiple variables on population clustering. This model emphasizes the interaction and synergy among different variables, revealing how their joint effects can significantly enhance clustering outcomes. By analyzing the composite influence of variable combinations, the model helps identify which interactions are most impactful in shaping human activity around HSR station areas. Figure 15 shows the interaction effects of influencing factors on spatial vitality in HSR station areas.

To quantify interaction effects, we used SHAP's built-in method, which separates each prediction into the effect of individual features and the added effect from feature pairs. For any two features, SHAP calculates how much their combined influence differs from the total of their separate effects. If the combined effect is stronger, this indicates a synergy—what we call a “synergy bonus.” These observed synergies are entirely data-driven and reflect the nonlinear relationships captured by the LightGBM model. We did not assign any fixed values or assumptions; instead, the SHAP interaction values reflect the complex relationships automatically learned by the LightGBM model.

The interaction plot reveals that the urbanization level of the HSR station's host city (C7) and the proportion of commercial service land use in the station area (E6) jointly affect the predicted intensity of population activity (Figure 16). The results show that when the urbanization level is high (above approximately 85%) and the proportion of commercial service land use exceeds 15%, the predicted value of population activity rises significantly, peaking at around 8.9. This indicates a strong clustering effect. Conversely, when either the

urbanization level is low or the proportion of commercial land is insufficient, the predicted values remain at a relatively low level. This supports the “urban function–infrastructure synergy” theory, where a mature urban context provides sufficient population density and mobility demand, while commercial land facilitates consumption and social interaction.

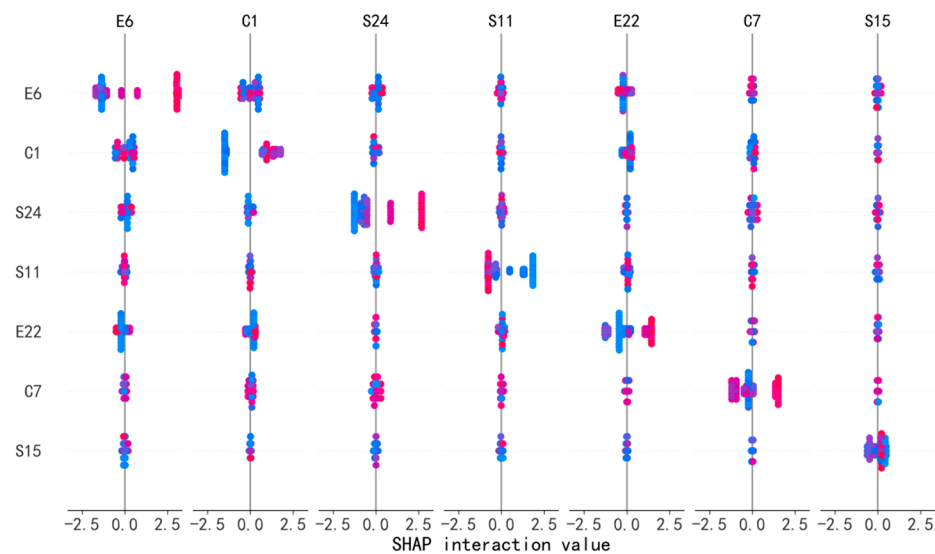


Figure 15. SHAP interaction values for spatial vitality in HSR station areas.

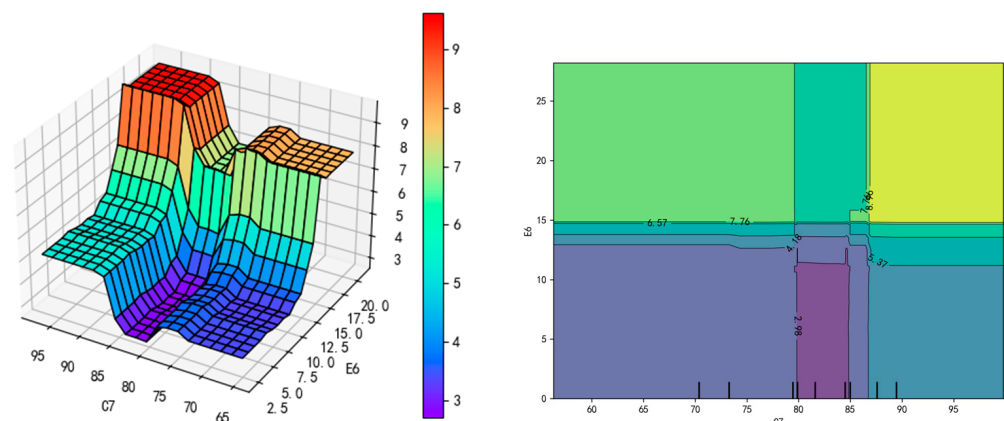


Figure 16. Three-dimensional and two-dimensional joint partial dependence plots for features C7 and E6.

This suggests a clear synergistic relationship between a city’s development level and the functional capacity of the station area. A high level of urbanization typically brings higher population density and better infrastructure, while commercial services offer spaces for consumption and social engagement. The combination of these two factors maximizes the vitality of HSR station areas. Therefore, improving station-area vibrancy depends not only on local planning and design but also on aligning with the broader developmental context of the city. In cities with high urbanization levels, special attention should be paid to enhancing functional diversity within the station area—particularly by increasing the supply and quality of commercial service facilities—to ensure efficient land use and foster a positive cycle of population clustering.

Another interaction plot shows the joint effect of the number of bus stops (S24) and the proportion of commercial service land use (E6) on population activity (Figure 17). The plot reveals that when the number of bus stops is relatively high and the commercial land use ratio is moderate to high (between approximately 10% and 20%), the predicted value

of population activity increases significantly, peaking at around 10.3. This demonstrates a strong positive synergy. In contrast, if the number of bus stops is low or the commercial land share is insufficient, the predicted activity level drops considerably. This aligns with transit-oriented development (TOD) principles: multimodal connectivity enhances last-mile accessibility, while commercial functions support destination attractiveness.

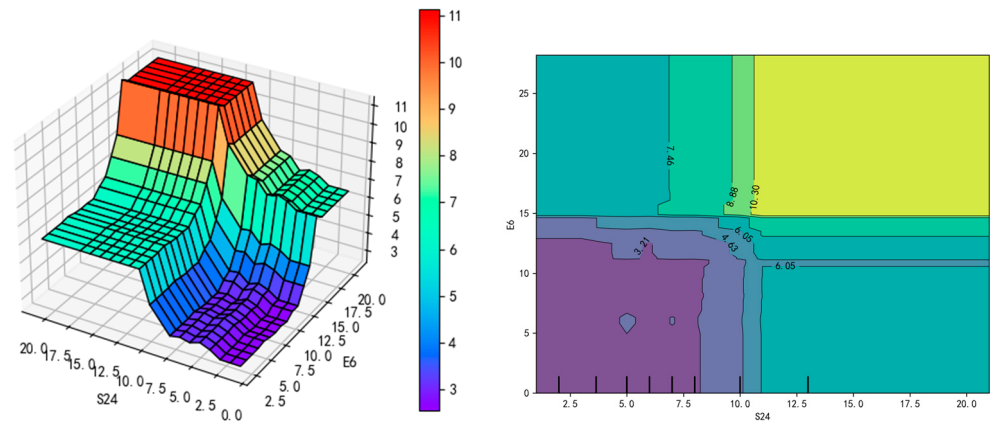


Figure 17. Three-dimensional and two-dimensional joint partial dependence plots for features S24 and E6.

Additionally, the interaction between the station location index (S11) and the number of parking facilities (E22) also influences both population clustering and mobility within the station area. As shown in Figure 18, when the station location index is low (approximately 0.1) and the number of parking facilities increases substantially (e.g., up to 400 spaces), the predicted value of population activity rises to 7.87, reflecting a pronounced clustering effect. A higher location index typically means the station is farther from the city center, leading to weaker transport connections. In such cases, sufficient parking becomes critical—especially for car users—by lowering access costs and improving station reachability. This interaction illustrates a compensatory mechanism: parking facilities can mitigate the disadvantage of poor location by reducing access costs for car users, thus broadening the station’s functional catchment. This finding resonates with access–cost theory and supports the importance of integrated planning in peripheral station areas.

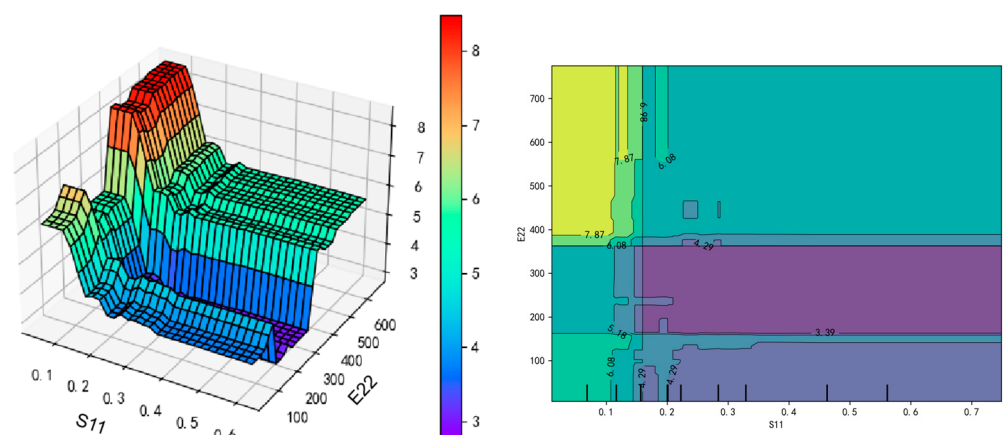


Figure 18. Three-dimensional and two-dimensional joint partial dependence plots for features S11 and E22.

As the number of parking facilities increases, the area’s transportation support capacity improves, which in turn promotes population aggregation. Particularly when the station’s spatial location is less advantageous, optimizing parking infrastructure can help offset

locational disadvantages and attract more users. The interaction between station location and parking capacity highlights the importance of coordinated planning. Enhancing transportation linkages and parking management can significantly improve accessibility and appeal, contributing to sustainable human activity and functional development in HSR station areas.

The observed synergies between urbanization and commercial land use suggest that spatial vitality is a product of both macro-level urban development and micro-level spatial design. This supports theories from urban systems and complexity science, which emphasize the interdependence between urban form and function. The enhanced activity observed at well-connected, well-equipped station areas validates the role of integrated planning in maximizing land value and urban accessibility.

4.6. Individual-Level Influences

Due to differences in land use types, development intensity, and spatial configuration patterns, high-speed rail (HSR) station areas exhibit diverse developmental characteristics within the urban spatial structure. Different types of station areas show distinct patterns in how they attract populations, which factors most influence clustering, and what development strategies are most effective. Therefore, analyzing individual station areas helps identify key influencing factors and enables the formulation of tailored planning approaches.

For HSR station areas dominated by residential functions, population clustering is primarily driven by the availability of lifestyle amenities, commercial services, and transportation accessibility (Figure 19). For example, at Beijing South Station, several variables were found to positively influence population aggregation, including the number of commercial and leisure facilities within the station area (S27), general public budget revenue of the host city (C10), centrality within the urban network (C16), and the number of directly connected stations via one-transfer routes (S4). These factors collectively indicate that the station attracts people by offering robust commercial amenities and strong connectivity. However, certain constraints were also identified, such as limited sky openness (E16), high road density (C12), and a lower number of metro lines (S23), which may hinder further population clustering. Therefore, the development strategy for such residential-oriented station areas should focus on enhancing everyday service facilities, strengthening metro and bus integration, and improving land use efficiency to increase the convenience and livability of the area, thereby enhancing its overall attractiveness.

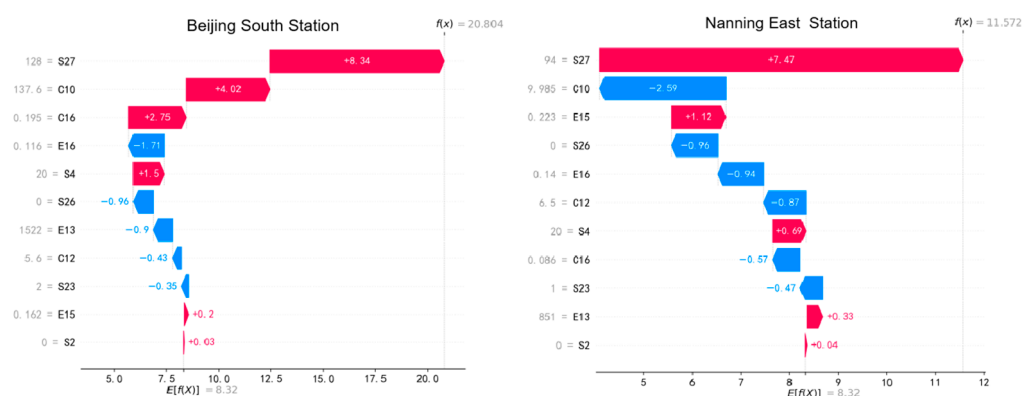


Figure 19. SHAP feature contribution for Beijing South Station and Nanning East Station.

In commercially oriented station areas, such as Zhengzhou Station, population clustering is mainly influenced by the quality of commercial facilities, the radiation effect of the city center, and transportation connectivity (Figure 17). Key promoting factors include

urban centrality (C16), the number of station-based commercial and leisure facilities (S27), and the number of directly connected stations (S4). These variables demonstrate that Zhengzhou Station effectively attracts passengers and dwellers through its commercial vitality and strong transport hub functions, enhancing its clustering effect (Figure 20). However, several limiting factors were also identified, including general public budget revenue (C10), sky openness (E16), road network density (C12), number of road intersections (E13), and the number of metro lines (S23). These constraints indicate that although Zhengzhou Station performs well in commercial and transportation dimensions, it still has room for improvement in environmental comfort, traffic network optimization, and vertical spatial design. To further promote clustering effects in this station area, it is recommended to integrate surrounding resources through comprehensive and future-oriented planning and renewal efforts.

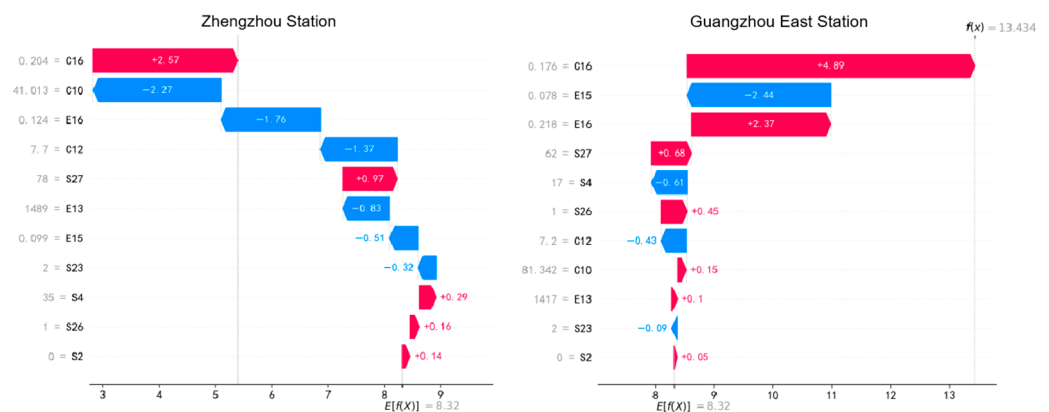


Figure 20. SHAP feature contribution for Zhengzhou Station and Guangzhou East Station.

Overall, these station-level differences reaffirm the necessity of differentiated planning strategies. Different types of HSR station areas exhibit their own unique spatial characteristics and development potential. Residential-, transport-, commercial-, and public service-oriented station areas each have their own advantages and challenges. Therefore, development strategies should be tailored according to the station's functional orientation, geographic context, infrastructure quality, and socioeconomic background. For each station type, enhancing infrastructure, optimizing spatial layout, and reinforcing inter-functional connectivity are key strategies for fostering population aggregation and achieving sustainable development of HSR station areas.

5. Discussion

5.1. Rethinking Station Area Vitality Through Nonlinear Urban Dynamics

This study's findings challenge the conventional linear assumptions in urban spatial analysis and contribute to a more nuanced understanding of how high-speed rail (HSR) station areas evolve. The revealed nonlinear relationships—particularly the saturation effects of commercial land and bus stop density—indicate that vitality does not increase indefinitely with infrastructural inputs. This echoes recent shifts in urban theory emphasizing complex systems thinking, where urban space responds to incremental interventions in nonlinear, often threshold-dependent ways.

In comparison with existing literature, which primarily views commercial land and transit accessibility as linear enablers of urban vitality, our results add empirical weight to the idea that “more” is not always “better.” The discovery of diminishing marginal returns in facility input highlights the importance of optimal functional balance in station-area development. From a theoretical standpoint, this finding extends the urban spatial equilib-

rium framework by introducing nonlinear feedback effects triggered by over-saturation or spatial crowding.

5.2. Spatial Thresholds and the Redefinition of Planning Efficiency

The identification of clear spatial thresholds—such as the optimal commercial land ratio (10–15%) and ideal bus stop density—offers a critical redefinition of what constitutes planning efficiency in station area contexts. Rather than maximizing infrastructure, planners should aim for functionally calibrated interventions based on local capacity and contextual needs. This perspective moves beyond traditional engineering-driven logic and embraces performance-based spatial planning.

Moreover, these thresholds provide actionable tools for resource prioritization in cities facing financial or spatial constraints. For instance, peripheral stations may not benefit from excessive commercial zoning or transit lines, whereas moderate-scale, context-sensitive development could yield more sustainable outcomes. This insight has practical value for transit-oriented development (TOD), supporting phased, evidence-based investment models that align with evolving population flows and spatial usage.

5.3. Theoretical and Policy Implications for Node–Place Synergy

The findings reinforce and refine the “node–place” framework by revealing that spatial vitality emerges not from the node (HSR infrastructure) or the place (urban environment) alone, but from their synergy, particularly when modulated by city-level conditions (e.g., urbanization rate, centrality). The interaction effects between macro-structural variables (e.g., C7, urbanization) and micro-spatial attributes (e.g., E6, commercial land) suggest that infrastructure performance is contingent upon urban context compatibility.

This insight has clear policy relevance. For national and provincial planning bodies, it calls for differentiated station development models based on regional typologies. For instance, stations in highly urbanized zones should prioritize function mixing and pedestrian-centric design, while those in transition zones may benefit from mobility-enhancing infrastructure like park-and-ride systems. Furthermore, integrating social sensing data (e.g., Weibo check-ins) into urban monitoring can facilitate adaptive governance and promote real-time evaluation frameworks for major transport infrastructure.

To further contextualize the findings, we compare two representative cases: Chengdu and Beijing. Chengdu, as a rapidly expanding city, features HSR stations like Chengdu East and Chengdu South that are still in the process of urban integration. Here, planning should emphasize balanced commercial land development, robust multimodal access, and staged infrastructure investment to prevent premature saturation. In contrast, mature hubs such as Beijing South or Beijing West require strategies that alleviate congestion, optimize transfer systems, and elevate service quality. These differentiated approaches underscore that the node–place synergy must be adapted to each city’s growth stage and spatial characteristics, reaffirming the importance of localized, context-aware planning in HSR station area development.

6. Conclusions

This study systematically examined the spatial vitality of high-speed rail (HSR) station areas in China by integrating multi-source data and applying both statistical and machine learning methods. Based on a city–node–place framework, we evaluated 66 HSR station areas across 35 cities using Sina Weibo check-in data, urban support capacity, station attributes, and built environment indicators.

Importantly, this study considered both the broader urban context in which each station is embedded and the detailed environmental features within the surrounding

area. City-level factors such as population size, urbanization level, and transit network configuration form the underlying support conditions, while local-scale features—including commercial land use ratio, bus stop density, and parking facilities—shape the functional performance of each station area. Our use of a 1500 m buffer allowed for consistent comparison across cities, capturing meaningful differences in how both urban structure and immediate surroundings influence vitality.

The results show that city population size, urbanization level, commercial land use ratio, transit accessibility, and parking facilities are key positive contributors to station area vitality. SHAP analysis revealed nonlinear threshold effects and interaction relationships—such as diminishing returns in commercial land use and bus stop density, and strong synergies between urban development and functional configuration.

These findings have several policy implications. First, land use and transit planning can be integrated to enhance functional synergy around HSR stations. Second, moderate commercial land allocation (around 10–15%) is most effective in attracting population activity. Third, improving multimodal transport infrastructure—including bus stops and parking—can strengthen last-mile connectivity. Fourth, tailored development strategies are needed for different station types (e.g., residential vs. commercial). Lastly, data-driven evaluation tools like social media analytics and interpretable machine learning can support real-time monitoring and adaptive policy formulation.

Nonetheless, this study has some limitations. The use of Sina Weibo check-in data, while providing high spatial-temporal granularity, may introduce demographic and geographic biases. Younger users and residents of more developed cities are overrepresented, while the behavior of older adults and residents in less developed or peripheral areas may be underrepresented. These biases could influence the comprehensiveness of vitality assessment. Future research should further integrate diverse datasets—such as mobile phone signaling, smart card records, and public Wi-Fi logs—to improve representativeness. Additionally, the present study focuses on a single temporal snapshot. Future studies could explore how spatial vitality evolves over time using longitudinal data sources. Such efforts may reveal seasonal rhythms, lifecycle stages, or long-term shifts in station activity patterns, thereby enriching the understanding of station area development dynamics. Overall, this study offers empirical insights and methodological innovations to guide the sustainable, accessible, and vibrant development of HSR station areas in China and beyond.

Author Contributions: Conceptualization, P.L. and X.C.; Investigation, W.S. and Y.X.; Methodology, P.L. and J.M.; Software, P.L. and J.M.; Writing—original draft, P.L.; Writing—review and editing, X.C., J.M. and W.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 52472331, 52278081 and U20A20330.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. HSR Stations, Corresponding Codes, and Key Independent Variable Values.

No.	Station Name	Station Code	C1 (Million)	C7 (%)	S11	S15 (Count)	S24 (Count)	E6 (%)	E22 (Count)
1	Beijing North	VAP	21.84	87.6	0.1	109	5	11.3	590
2	Beijing South	VNP	21.84	87.6	0.08	577	3	9.5	399

Table A1. Cont.

No.	Station Name	Station Code	C1 (Million)	C7 (%)	S11	S15 (Count)	S24 (Count)	E6 (%)	E22 (Count)
3	Beijing	BJP	21.84	87.6	0.03	81	8	21	719
4	Beijing West	BXP	21.84	87.6	0.12	289	2	17.3	562
5	Tianjin South	TIP	13.63	84.5	0.29	1	2	4.5	41
6	Tianjin	TJP	13.63	84.5	0.13	143	7	13.5	631
7	Tianjin West	TXP	13.63	84.5	0.2	158	5	7.7	261
8	Shijiazhuang	SJP	11.22	71.44	0.19	106	8	13.5	188
9	Shanghai	SHH	24.76	89.46	0.21	295	12	14.9	659
10	Shanghai Hongqiao	AOH	24.76	89.46	0.25	736	14	19.5	158
11	Shanghai West	SXH	24.76	89.46	0.24	0	4	11.3	247
12	Ningbo	NGH	9.62	79.9	0.2	94	10	10.9	427
13	Hangzhou East	HGH	12.38	84	0.05	195	6	14.6	294
14	Hangzhou South	XHH	12.38	84	0.58	0	3	12.5	198
15	Nanjing South	NKH	9.49	87.01	0.32	226	2	18.6	192
16	Nanjing	NJH	9.49	87.01	0.07	155	8	8	238
17	Hefei South	ENH	12.04	84.6	0.23	229	5	8.3	115
18	Hefei	HFH	12.04	84.6	0.22	5	10	21.1	315
19	Shenzhen North	IOQ	17.66	99.8	0.16	562	10	8.3	224
20	Shenzhen	SZQ	17.66	99.8	0.12	225	10	13.8	621
21	Futian	NZQ	17.66	99.8	0.01	53	15	28.2	668
22	Guangzhou East	GGQ	18.73	86.48	0.21	356	5	16.4	600
23	Guangzhou South	IZQ	18.73	86.48	0.62	804	5	11.7	103
24	Shapingba	CYW	32.13	71	0.37	57	12	11.1	451
25	Chongqing North	CUW	32.13	71	0.07	237	2	7.2	243
26	Chongqing West	CXW	32.13	71	0.55	164	3	3.8	72
27	Chengdu East	ICW	21.27	79.89	0.16	657	5	5.8	165
28	Chengdu South	CNW	21.27	79.89	0.12	103	10	16.7	549
29	Chengdu	CDW	21.27	79.89	0.09	4	6	17	248
30	Chengdu West	CMW	21.27	79.89	0.18	40	10	4.2	133
31	Wuhan	WHN	13.74	84.66	0.33	248	1	6.9	55
32	Hankou	HKN	13.74	84.66	0.16	287	2	17.5	376
33	Changsha South	CWQ	7.68	83.27	0.32	256	7	4.2	97
34	Nanchang	NCG	6.54	79.58	0.22	73	14	10.5	362
35	Nanchang West	NXG	6.54	79.58	0.51	260	12	3.4	56
36	Harbin	HBB	9.4	56.3	0.07	138	13	22	571
37	Harbin West	VAB	9.4	56.3	0.2	117	7	12.3	133
38	Changchun	CCT	9.07	67.1	0.26	166	16	14.3	464
39	Changchun West	CRT	9.07	67.1	0.46	27	2	0	7
40	Dalian North	DFT	7.53	73	0.45	137	3	11.2	149
41	Shenyang North	SBT	9.15	84.99	0.13	66	6	16.1	478
42	Shenyang	SYT	9.15	84.99	0.04	142	8	15.7	775
43	Qingdao North	QHK	10.34	77.32	0.29	157	4	2.4	67
44	Qingdao	QDK	10.34	77.32	0.16	156	17	6.4	363
45	Jinan	JNK	9.42	74.3	0.07	42	21	15.4	421
46	Jinan West	JGK	9.42	74.3	0.34	126	6	6.1	128
47	Xiamen North	XKS	5.31	90.2	0.47	155	6	4.9	61
48	Xiamen	XMS	5.31	90.2	0.08	162	11	10.5	495
49	Fuzhou South	FYS	8.45	73.27	0.58	120	7	8.9	49
50	Fuzhou	FZS	8.45	73.27	0.47	220	10	4.6	152
51	Nanning East	NFZ	8.89	70.36	0.49	306	5	2.1	46
52	Haikou	VUQ	8.89	82.7	0.64	17	1	0.4	9
53	Zhengzhou East	ZAF	12.83	79.4	0.56	345	5	13.2	231
54	Zhengzhou	ZZF	12.83	79.4	0.15	29	4	19.8	374
55	Kunming South	KOM	8.5	80.5	0.75	273	4	2.8	45
56	Kunming	KMM	8.5	80.5	0.28	90	13	20.5	376
57	Guiyang North	KQW	6.22	80.3	0.29	263	6	6.4	49
58	Taiyuan South	TNV	5.44	89.3	0.35	269	6	10.5	123
59	Taiyuan	TYV	5.44	89.3	0.25	7	5	20.6	393
60	Xi'an North	EAY	13	79.6	0.62	474	2	10.8	51

Table A1. Cont.

No.	Station Name	Station Code	C1 (Million)	C7 (%)	S11	S15 (Count)	S24 (Count)	E6 (%)	E22 (Count)
61	Lanzhou West	LAJ	4.42	84.07	0.2	133	4	6.3	75
62	Xining	XNO	5.95	61.43	0.52	55	7	4.4	100
63	Yinchuan	YIJ	7.28	66.34	0.38	88	8	6.4	102
64	Hohhot East	NDC	3.55	79.8	0.31	98	5	9.8	67
65	Hohhot	HHC	3.55	79.8	0.12	2	8	8.4	226
66	Urumqi	WAR	4.08	96.5	0.26	77	7	11.1	87

References

- Zheng, W.; Wei, S. A 'node-place-network-city' framework to examine HSR station area development dynamics: Station typologies and development strategies. *J. Transp. Geogr.* **2024**, *120*, 103993. [\[CrossRef\]](#)
- Li, J.; Sun, X.; Cong, W.; Miyoshi, C.; Ying, L.C.; Wandelt, S. On the air-HSR mode substitution in China: From the carbon intensity reduction perspective. *Transp. Res. A Policy Pract.* **2024**, *180*, 103977. [\[CrossRef\]](#)
- Shao, S.; Tian, Z.; Yang, L. High speed rail and urban service industry agglomeration: Evidence from China's Yangtze River Delta region. *J. Transp. Geogr.* **2017**, *64*, 174–183. [\[CrossRef\]](#)
- Chen, Z.; Haynes, K.E. Impact of high-speed rail on regional economic disparity in China. *J. Transp. Geogr.* **2017**, *65*, 80–91. [\[CrossRef\]](#)
- Xu, W.A.; Zhou, J.; Yang, L.; Li, L. The implications of high-speed rail for Chinese cities: Connectivity and accessibility. *Transp. Res. A Policy Pract.* **2018**, *116*, 308–326. [\[CrossRef\]](#)
- Wang, F.; Wei, X.; Liu, J.; He, L.; Gao, M. Impact of high-speed rail on population mobility and urbanisation: A case study on Yangtze River Delta urban agglomeration, China. *Transp. Res. A Policy Pract.* **2019**, *127*, 99–114. [\[CrossRef\]](#)
- Jia, S.; Zhou, C.; Qin, C. No difference in effect of high-speed rail on regional economic growth based on match effect perspective? *Transp. Res. A Policy Pract.* **2017**, *106*, 144–157. [\[CrossRef\]](#)
- Shaw, S.; Fang, Z.; Lu, S.; Tao, R. Impacts of high speed rail on railroad network accessibility in China. *J. Transp. Geogr.* **2014**, *40*, 112–122. [\[CrossRef\]](#)
- Yue, Y.; Chen, J.; Feng, T.; Ma, X.; Wang, W.; Bai, H. Classification and determinants of high-speed rail stations using multi-source data: A case study in Jiangsu Province, China. *Sustain. Cities Soc.* **2023**, *96*, 104640. [\[CrossRef\]](#)
- Deng, T.; Gan, C.; Perl, A.; Wang, D. What caused differential impacts on high-speed railway station area development? Evidence from global nighttime light data. *Cities* **2020**, *97*, 102568. [\[CrossRef\]](#)
- Diao, M.; Zhu, Y.; Zhu, J. Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Stud.* **2017**, *54*, 2249–2267. [\[CrossRef\]](#)
- Wei, S.; Wang, L. Classifying High-Speed Rail Stations in the Yangtze River Delta, China: The Node-Place Modelling Approach. *Appl. Spat. Anal. Policy* **2022**, *16*, 625–646. [\[CrossRef\]](#)
- Wang, L.; Zheng, W.; He, S.; Wei, S. Assessing urban vitality and its determinants in high-speed rail station areas in the Yangtze River Delta, China. *J. Transp. Land Use* **2022**, *15*, 333–354. [\[CrossRef\]](#)
- Lan, F.; Gong, X.; Da, H.; Wen, H. How do population inflow and social infrastructure affect urban vitality? Evidence from 35 large- and medium-sized cities in China. *Cities* **2020**, *100*, 102454. [\[CrossRef\]](#)
- Chen, Y.; Yu, B.; Shu, B.; Yang, L.; Wang, R. Exploring the spatiotemporal patterns and correlates of urban vitality: Temporal and spatial heterogeneity. *Sustain. Cities Soc.* **2023**, *91*, 104440. [\[CrossRef\]](#)
- Li, Q.; Cui, C.; Liu, F.; Wu, Q.; Run, Y.; Han, Z. Multidimensional Urban Vitality on Streets: Spatial Patterns and Influence Factor Identification Using Multisource Urban Data. *ISPRS Int. J. Geo-Inf.* **2021**, *11*, 2. [\[CrossRef\]](#)
- Wang, B.; de Jong, M.; van Bueren, E.; Ersoy, A.; Meng, Y. Transit-Oriented Development in China: A Comparative Content Analysis of the Spatial Plans of High-Speed Railway Station Areas. *Land* **2023**, *12*, 1818. [\[CrossRef\]](#)
- Zhang, X.; Wang, L.; Wei, S.; Cui, Y. Multifunctional development patterns and driving mechanisms of high-speed rail station areas in the Yangtze River Delta, China. *Transp. Policy* **2025**, *167*, 264–275. [\[CrossRef\]](#)
- Yao, Y.; Liu, X.; Li, X.; Zhang, J.; Liang, Z.; Mai, K.; Zhang, Y. Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 1220–1224. [\[CrossRef\]](#)
- Kim, S.; Lee, K.Y.; Shin, S.I.; Yang, S. Effects of tourism information quality in social media on destination image formation: The case of Sina Weibo. *Inf. Manag.* **2017**, *54*, 687–702. [\[CrossRef\]](#)
- Huang, H.; Long, R.; Chen, H.; Sun, K.; Li, Q. Exploring public attention about green consumption on Sina Weibo: Using text mining and deep learning. *Sustain. Prod. Consum.* **2022**, *30*, 674–685. [\[CrossRef\]](#)

22. Li, S.; Dragicevic, S.; Castro, F.A.; Sester, M.; Winter, S.; Coltekin, A.; Pettit, C.; Jiang, B.; Haworth, J.; Stein, A.; et al. Geospatial big data handling theory and methods: A review and research challenges. *ISPRS J. Photogramm. Remote Sens.* **2016**, *115*, 119–133. [\[CrossRef\]](#)
23. Yang, L.; Lu, Y.; Cao, M.; Wang, R.; Chen, J. Assessing accessibility to peri-urban parks considering supply, demand, and traffic conditions. *Landsc. Urban Plan.* **2025**, *257*, 105313. [\[CrossRef\]](#)
24. Yang, L.; Yang, H.; Yu, B.; Lu, Y.; Cui, J.; Lin, D. Exploring non-linear and synergistic effects of green spaces on active travel using crowdsourced data and interpretable machine learning. *Travel Behav. Soc.* **2024**, *34*, 100673. [\[CrossRef\]](#)
25. Shafiq, S.M.; Tian, Z.; Bashir, A.K.; Jolfaei, A.; Yu, X. Data mining and machine learning methods for sustainable smart cities traffic classification: A survey. *Sustain. Cities Soc.* **2020**, *60*, 102177. [\[CrossRef\]](#)
26. Liu, L.; Silva, E.A.; Wu, C.; Wang, H. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* **2017**, *65*, 113–125. [\[CrossRef\]](#)
27. Jing, Y.; Hu, H.; Guo, S.; Wang, X.; Chen, F. Short-Term prediction of urban rail transit passenger flow in external passenger transport hub based on LSTM-LGB-DRS. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 4611–4621. [\[CrossRef\]](#)
28. Doan, Q.C.; Ma, J.; Chen, S.; Zhang, X. Nonlinear and threshold effects of the built environment, road vehicles and air pollution on urban vitality. *Landsc. Urban Plan.* **2025**, *253*, 105204. [\[CrossRef\]](#)
29. Liang, P.; Cui, X.; Lin, M.; Yang, T.; Wu, B. High-speed rail effects on station area-level business commercial agglomeration: Evidence from 110 stations in China. *Front. Environ. Sci.* **2022**, *10*, 1045959. [\[CrossRef\]](#)
30. Yin, H.; Xiao, R.; Fei, X.; Zhang, Z.; Gao, Z.; Wan, Y.; Tan, W.; Jiang, X.; Cao, W.; Guo, Y. Analyzing “economy-society-environment” sustainability from the perspective of urban spatial structure: A case study of the Yangtze River delta urban agglomeration. *Sustain. Cities Soc.* **2023**, *96*, 104691. [\[CrossRef\]](#)
31. Li, X.; Shi, L.; Shi, Y.; Tang, J.; Zhao, P.; Wang, Y.; Chen, J. Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Appl. Geogr.* **2024**, *166*, 103264. [\[CrossRef\]](#)
32. Loo, B.P.Y.; Huang, Z. Location matters: High-speed railway (HSR) stations in city evolution. *Cities* **2023**, *139*, 104380. [\[CrossRef\]](#)
33. Deng, Y.; Bai, Y.; Cui, L.; He, R. Travel Mode Choice Behavior for High-Speed Railway Stations Based on Multi-Source Data. *Transp. Res. Rec.* **2023**, *2677*, 525–540. [\[CrossRef\]](#)
34. Liu, Y.; Xu, S.; Tian, J.; Liu, T.; Dong, T. What matters in promoting new town by High-Speed Railway station? Evidence from China. *Transp. Policy* **2024**, *159*, 241–253. [\[CrossRef\]](#)
35. Deng, T.; Wang, D.; Yang, Y.; Yang, H. Shrinking cities in growing China: Did high speed rail further aggravate urban shrinkage? *Cities* **2019**, *86*, 210–219. [\[CrossRef\]](#)
36. Lin, C. Research of impact factors of High-speed railways Hub area development in China. *Urban Plan. Int.* **2011**, *26*, 72–77.
37. Wang, X.; Pan, H. Land Change Pattern in High-Speed Rail Station Area: Empirical Research on Yangtze River Delta Region in China from 2010 to 2020. In *Socioeconomic Impacts of High-Speed Rail Systems, Proceedings of the 2nd International Workshop on High-Speed Rail Socioeconomic Impacts*, online, 12–13 September 2023; Springer: Cham, Switzerland, 2023; pp. 315–334. [\[CrossRef\]](#)
38. Niu, B.; Yin, P.; Shen, P. Industrial Spatio-Temporal Distribution of High-Speed Rail Station Area from the Accommodation Facilities Perspective: A Multi-City Comparison. *Land* **2023**, *12*, 332. [\[CrossRef\]](#)
39. Sugie, L.; Sung, H.; Joong, C.M. Analysis of Users’ Facility Preference for Mixed Land Use Development in the High-Speed Rail Station Areas. *J. Korea Plan. Assoc.* **2014**, *49*, 211–224.
40. Ribalaygua, C.; Perez-Del-Caño, S. Assessing spatial planning strategy in high-speed rail station areas in Spain (1992–2018): Towards a sustainable model. *Eur. Plan. Stud.* **2019**, *27*, 595–617. [\[CrossRef\]](#)
41. Wang, B.; Ersoy, A.; van Bueren, E.; de Jong, M. Rules for the Governance of Transport and Land use Integration in High-speed Railway Station Areas in China: The Case of Lanzhou. *Urban Policy Res.* **2022**, *40*, 122–141. [\[CrossRef\]](#)
42. Doronzo, D.; Caggianelli, A.; Dellino, P. A ghost eruption behind “Le vulcanoclastiti di Craco” ghost town, Basilicata, Italy. *Int. J. Earth Sci.* **2019**, *108*, 545–546. [\[CrossRef\]](#)
43. Schütz, E. Stadtentwicklung durch Hochgeschwindigkeitsverkehr: Konzeptionelle und methodische Ansätze zum Umgang mit den Raumwirkungen des schienengebundenen Personen-Hochgeschwindigkeitsverkehrs (HGV) als Beitrag zur Lösung von Problemen der Stadtentwicklung. *Informationen Zur Raumentwickl.* **1998**, *6*, 369–383.
44. Lan, W.; Can, W.; Chen, C.; Hao, G. Development and Planning of the Surrounding Area of High-Speed Rail Stations: Based on Empirical Study of Beijing-Shanghai Line. *Urban Plan. Forum* **2014**, *4*, 31–37.
45. Zou, Z.; Tang, Y. Evaluation of Sustainable Development Potential of High-Speed Railway Station Areas Based on “Node-Place-Industry” Model. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 349. [\[CrossRef\]](#)
46. Du, J.; Druta, O.; van Wesemael, P. Place quality in high-speed rail station areas: Concept definition. *J. Transp. Land Use* **2021**, *14*, 1165–1186. [\[CrossRef\]](#)
47. Wang, L.; Yuan, F.; Duan, X. How high-speed rail service development influenced commercial land market dynamics: A case study of Jiangsu province, China. *J. Transp. Geogr.* **2018**, *72*, 248–257. [\[CrossRef\]](#)

48. Jiao, J.; Zhang, Q.; Jiang, R.; Lyu, G. Does high-speed rail contribute to cross-boundary agglomeration of migrant workers? Evidence from China. *J. Transp. Geogr.* **2025**, *125*, 104170. [[CrossRef](#)]
49. Wang, F.; Liu, Z.; Xue, P.; Dang, A. High-speed railway development and its impact on urban economy and population: A case study of nine provinces along the Yellow River, China. *Sustain. Cities Soc.* **2022**, *87*, 104172. [[CrossRef](#)]
50. Jendryke, M.; Balz, T.; Liao, M. Big location-based social media messages from China's Sina Weibo network: Collection, storage, visualization, and potential ways of analysis. *Trans. GIS* **2017**, *21*, 825–834. [[CrossRef](#)]
51. Lu, Y.; Li, J.; Yang, Z.; Ou, X.; Xie, W. Data mining and social networks processing method based on support vector machine and k-nearest neighbor. *J. Comput. Methods Sci. Eng.* **2021**, *21*, 435–447. [[CrossRef](#)]
52. Zhou, M.; Wang, M.; Hu, Q. A POI Data Update Approach Based on Weibo Check-in Data. In Proceedings of the 2013 21st International Conference on Geoinformatics (Geoinformatics), Henan University, Kaifeng, China, 20–22 June 2013.
53. Deng, T.; Wang, D.; Hu, Y.; Liu, S. Did high-speed railway cause urban space expansion?—Empirical evidence from China's prefecture-level cities. *Res. Transp. Econ.* **2020**, *80*, 100840. [[CrossRef](#)]
54. Diao, M. Does growth follow the rail? The potential impact of high-speed rail on the economic geography of China. *Transp. Res. A Policy Pract.* **2018**, *113*, 279–290. [[CrossRef](#)]
55. Guo, Y.; Cao, L.; Song, Y.; Wang, Y.; Li, Y. Understanding the formation of City-HSR network: A case study of Yangtze River Delta, China. *Transp. Policy* **2022**, *116*, 315–326. [[CrossRef](#)]
56. Xu, M.; Shuai, B.; Wang, X.; Liu, H.; Zhou, H. Analysis of the accessibility of connecting transport at High-speed rail stations from the perspective of departing passengers. *Transp. Res. A Policy Pract.* **2023**, *173*, 103714. [[CrossRef](#)]
57. Zhang, X.; Loo, B.P.Y.; Wang, L.; Wei, S. Disparities in high-speed rail accessibility in the Yangtze River Delta, China: A door-to-door travel time perspective. *Res. Transp. Bus. Manag.* **2025**, *59*, 101308. [[CrossRef](#)]
58. Yang, J.; Cao, J.; Zhou, Y. Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. *Transp. Res. A Policy Pract.* **2021**, *144*, 74–88. [[CrossRef](#)]
59. Mouratidis, K.; Poortinga, W. Built environment, urban vitality and social cohesion: Do vibrant neighborhoods foster strong communities? *Landsc. Urban Plan.* **2020**, *204*, 103951. [[CrossRef](#)]
60. Li, X.; Li, Y.; Jia, T.; Zhou, L.; Hijazi, I. The six dimensions of built environment on urban vitality: Fusion evidence from multi-source data. *Cities* **2022**, *121*, 103482. [[CrossRef](#)]
61. He, Q.; He, W.; Song, Y.; Wu, J.; Yin, C.; Mou, Y. The impact of urban growth patterns on urban vitality in newly built-up areas based on an association rules analysis using geographical 'big data'. *Land Use Policy* **2018**, *78*, 726–738. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.