WILEY | Hindawi

Research Article

Exploring the Spatial Impact of Multisource Data on Urban Vitality: A Causal Machine Learning Method

Zhixuan Xiao ,[1] Chengyi Li ,[1] Shihua Pan ,[1] Gaoda Wei ,[1] Mengmeng Tian ,[1] and Runjiu Hu [2]

[1]*School of Cyber Science and Engineering, Zhengzhou University, Henan, China*
[2]*Wuhan Eleho Co. Ltd., China*

Correspondence should be addressed to Runjiu Hu; hurunjiu@163.com

Identifying urban vitality is the key to optimizing the urban structure. Previous studies on urban multisource data and urban vitality often assume that they follow a predefined (linear or nonlinear in terms of parameters) relationship, and few studies have explored the causality of urban multisource data on urban vitality. The existing machine learning methods often pay attention to the correlation in the data and ignore the causality. With the continuous emergence of new needs, its disadvantages gradually begin to appear and face a series of urgent problems in interpretability, robustness, and fairness. In this paper, we use a combination of causal inference and machine learning to deeply explore and analyze the causal effects of multisource data on the 16 administrative districts of Shanghai, taking Shanghai as an example. The analysis results show that each data indicator has different degrees of influence on the urban vitality of the 16 administrative districts of Shanghai, resulting in different heterogeneous effects, and through the analysis result, each administrative district can better optimize urban resources and improve urban vitality according to its situation. This discovery guides urban planning and has enlightenment significance for cities seeking construction facility investment and facility construction-oriented development.

1. Introduction

The concept of "urban vitality" was proposed by Jane Jacobs in "The Death and Life of Great American Cities." With the unprecedented global urbanization since the 20th century, large cities have witnessed economic prosperity and cultural diffusion and due to the rapid growth of population in central cities, the severe shortage of urban infrastructure capacity has brought about traffic congestion, environmental pollution, traffic chaos, operational inefficiency, etc. In addition, the expansion of cities has also led to residential vacancy, population loss, and degradation of ecosystems and these problems have seriously constrained the sustainable development of cities. A growing number of studies have used urban vitality to understand the complex relationships of urban disruption, and creating urban vitality is crucial for developing urban centers. We need to understand the spatial-scale causal effects of urban multisource data on urban vitality in order to better optimize urban resource allocation to further stimulate urban vitality. Causal inference on the spatial extent of urban multisource data can help us analyze the essence of the problem and provide a new way to alleviate the problems of unbalanced resource allocation and blind urban expansion.

With the advent of the big data era, we can make full use of urban multisource data to explore the urban structure and the laws behind it. Compared with the data obtained from traditional questionnaires, the urban vitality analysis based on big data can allow us to obtain a wide variety of temporal and spatial data about the city, with the help of which, we can build a model from the traditional way to the identification of spatial causal effects. Causal inference is a process of identifying the causes that lead to an outcome. It refers to the relationship between quantities, where a change in one quantity is the cause of a change in another quantity. Causal reasoning is a broad scientific framework rather than a set of

methods and has gained wide popularity in several fields such as medicine, economics, and education.

In statistics, correlation is not the same as causality. The existence of a correlation between two variables in an event does not mean that a change in one of the variables is caused by a change in the other variable. Because there are often confounding factors in an event that may affect both variables, it is impossible to say that there is a causal relationship between the two variables. Therefore, we want to explore the degree of influence of one variable on the other and we must perform a causality analysis, not a correlation analysis.

Sung et al. (2015) tested the importance of built environment features and assessed their relative importance in promoting urban vibrancy; Liu et al. [1] proposed a method to extract vibrancy zones and integrate spatiotemporal feature clustering to analyze the spatiotemporal distribution patterns of vibrancy zones and their correlation with land use. Although there are papers studying the correlation of spatiotemporal distribution patterns of urban vitality, association Not Causation. In this paper, we design a causal inference method combined with machine learning to identify spatio causal effects using multiple sources of Shanghai data.

Thus, the contributions of this paper are as follows.

(1) The first proposal is to use causal machine learning models to explore the impact of multisource data on urban dynamics

(2) It is more robust and rational and interpretable than the traditional direction of correlation research

(3) Planning recommendations are given for each ward to improve urban vitality

The rest of the paper is organized as follows. Section 2 gives a brief review of urban vitality. Section 3 describes the causal machine learning method and analysis of the spatio causal effects of urban vitality. Section 4 detailed causal analysis for each administrative region. At last, Section 5 concludes the research and points to future research directions.

2. Related Work

Creating urban vibrancy has been the focus of urban planners [2]. The vitality of cities stems from people and their activities in space [3]. Attoe and Logan [4] argue that behaviors associated with everyday urban life are the basis and starting point of urban vitality. Lynch argues that vitality is the level of support for life and the demands on ecology and humanity [5]. These important theories lay the foundation for the study of urban dynamics.

Urban vitality is an important basis for evaluating urban development and spatial balance. In the era of big data, the quantitative analysis of urban vitality has become a research hotspot in the field of urban sustainable development and planning research [1]. Jacobs [6] and Lynch [7] define urban vitality as the ability of an urban system to sustain its internal survival, growth, and development, which evolves from a satisfactory urban form. Montgomery [2] describes urban

vitality as the degree to which the urban atmosphere is active.

Urban vitality is the source and key engine of urban development [8], which is commonly used in the quality assessment of urban development. It reflects human activities in different places and times [9]. It also reflects the essential elements for achieving quality of life in cities, which results from good urban form, developed urban functions, and adequate urban activities [10]. Urban morphology understands urban vitality as the urban life and activities shaped and influenced by the urban spatial form [11]. From the perspective of urban sociology, scholars understand cities as the reflection and existence of economy, society, and culture and urban vitality consists of three parts: economic vitality, social vitality, and cultural vitality [12] which demonstrate the ability of cities to provide facilities and spaces for the activities of their residents.

The concept of urban vitality states that urban vitality is shaped by the continuous human activity in the area during the day, rather than by the crowd gathering effect during peak hours [6]. Urban vitality can be measured in two ways: objectively by measuring the level of activity in a place and subjectively by measuring people's perceptions of the vitality of a place. Assessing urban vitality in a traditional data environment is mainly based on a qualitative perspective or through field observations and questionnaires [13]. For example, March et al. [14] noted that measuring vibrancy should take into account the different experiences required for a healthy life. Sung & Lee [13] conducted a telephone survey to examine the daily walking activities of Seoul residents, further revealing the link between the residential environment and walking activities. As the importance of cities as consumption centers has increased, the measurement of economic vitality has gradually shifted from macroproduction indicators to the level of the consumer economy and scholars have begun to experiment with POI [15]. The spatial distribution characteristics of economic dynamism have been characterized by the use of data such as POI and public reviews [16].

Urban spatial vitality is an urban activity based on the influence of the urban spatial form. A vibrant city can contribute to a better performance of the spatial layout and thus respond to the constant challenges faced by cities [17]. Rapid population expansion and a severe shortage of urban infrastructure capacity have brought about problems such as traffic congestion, environmental pollution, disorder, and operational inefficiency. These problems have seriously constrained the sustainable development of cities [18–22]. Therefore, it is urgent to deeply analyze the spatiotemporal dynamic characteristics of urban human spatial activities and their relationship with urban resource allocation, so as to provide a reliable basis for urban infrastructure construction and optimal resource allocation. The study of the urban spatiotemporal dynamics model provides a new way to alleviate the problems of blind urban expansion and unbalanced resource allocation [1].

Many scholars are currently researching the factors influencing urban vitality [23]. Yang et al. [24] illustrated the nonlinear association and synergistic effect relationship

between metro access and land use and urban vitality in Shenzhen. Lu et al. [25] studied the impact of the built environment on the city. Xia et al. [26] analyzed the spatial relationship between urban land-use intensity and urban vitality at the neighborhood level. Liu et al. [27] used a method based on principal component analysis to quantitatively evaluate the economic vitality of cities in the Yangtze River Delta. Lopes and Camanho [28] studied the impact of the public green space utilization on urban vitality. Lan et al. [29] investigated how population inflow and social infrastructure affect urban vitality. Yue and Zhu [30] conducted a study on the relationship between urban vitality and street centrality based on retrospective data of social networks in Wuhan. Sulis et al. [31] proposed a computational approach for urban vitality by analyzing new forms of spatial data to obtain quantitative measures of urban quality, which are often used to assess cities. Yue et al. [32] explored the relationship between neighborhood vitality and the mix and diversity of POIs using a linear regression model.

In general, the current directions are mainly (1) centering on the influence of certain factors on urban vitality, such as population flow, public green space, and built environment [13, 24–30] and (2) proposing a linear or nonlinear model to explore the connection between multisource data and urban vitality [1, 13, 31, 32]. Although these methods have their own advantages, the problem with these directions is that they cannot exclude the influence of confounding factors other than the study object on urban vitality and the results obtained lack interpretability and rationality.

3. Methodology

3.1. Data and Variables. The study used multisource data from Shanghai. Shanghai has a total land area of 6,340 square kilometers and a population of about 25 million. It is one of the four top-tier cities in China—Beijing, Guangzhou, and Shenzhen. This study focuses on the 16 administrative districts of Shanghai, Pudong New Area, Huangpu district, Xuhui district, Changning district, Jing'an district, Putuo district, Hongkou district, Yangpu district, Minhang district, Baoshan district, Jiading district, Jinshan district, Songjiang district, Qingpu district, Fengxian district, and Chongming district.

3.1.1. Socioeconomic Data. The socioeconomic data used in this study mainly include data on the resident population and its density and built environment. The resident population is the number of people who have lived in the city for more than six months. This indicator is different from the household population, which may not live in the city. The resident population is a better indicator of the actual population distribution in a city. Permanent population data and built environment data are obtained by consulting the yearbook of each region.

3.1.2. Point of Interest (POI). The POI data used in this study were obtained from AMAP (Gaode Map), one of the largest map search engines and providers in China. AMAP is part of the Beijing-based company AutoNavi, which provides a free application programming interface to collect data from different layers and functions. The POI data were labeled for all types of facilities. These POIs are divided into 14 categories: textile and food, restaurants, transportation, companies and businesses, retail and wholesale, scientific, educational and cultural services, government and organizations, residential, finance and insurance, sports, healthcare, public facilities, hotels and entertainment, and scenic spots. We selected the more representative points of interest of food and beverage services; scientific, educational, and cultural services; health care services, public life services; scenic spots; and transportation facility services. The variable name and variable description are shown in Table 1. The importance of variables and details is shown in Table 2.

The "city vitality index" goes beyond the "economy-centered" city development evaluation model and takes into account the development goals of building new cities and the existing development level of cities in China and designs the following calculation method for the city vitality index: point of population (Pop), point of land use (Polu), point of POI (Ppoi), and urban vitality score (UvS).

$$Pop = (Ba * Pd) * 0.00001, \tag{1}$$

$$Polu = \frac{(Fas * 0.05 + Ass * 0.25 + Sofs * 0.2 + Ssa * 0.15 + Sma * 0.15 + Saa * 3 + Asfa * 10 + Spa * 0.03)}{area\ of\ the\ district}, \tag{2}$$

$$Ppoi = (Fsd * 0.15 + Sed * 0.25 + Hsd * 0.25 + Lsd * 0.2 + Dss * 0.1 + Td * 0.05) * 100, \tag{3}$$

$$UvS = (Pop + Polu + Ppoi)^{0.5}. \tag{4}$$

We cannot always count all the poi points in a certain place when we count the POI of each district, but we think that the proportion of the number of poi points in each district is almost constant, so we use all the detected poi points in a certain district in a certain year as the base and divide it by the total number of such poi points in Shanghai, and the percentage that we get can be considered real and valid for calculation. Restaurants include general and fast-food restaurants, tea or coffee shops, pastry stores, and cold-drink stores. The health care service business includes all kinds of hospitals, clinics, pharmacies, emergency centers, etc. Scenic facilities include tourist attractions, parks and squares,

TABLE 1: Variable name and variable description.

Variable name	Variable description
Building area	Total floor area of the administrative district (10,000 square meters)
Population density	Total population of the administrative district/total floor area (people/km^2)
Land use	
Factory area share	Factory floor area/total floor area of the district
Area share of science and education zone	School floor area/total floor area of the district
Share of office floor space	Office building floor area/total floor area of the district
Share of shopping area	Floor area of the mall/gross floor area of the district
Share of medical area	Hospital floor area/gross floor area of the district
Share of accommodation area	Hotel floor area/gross floor area of the district
Area share of film and theater area	Theatre floor area/total floor area of the district
Share of park area	Park floor area/gross floor area of the district
POI	
Food service density	Number of restaurant POIs/total restaurant POIs
Science and education service density	Number of science and education service POIs/total Edu POIs
Health service density	Number of health service POIs/total number of health POIs
Life service density	Life service POI volume/total life POI volume
Density of scenic spots	Number of scenic POI/number of scenic POI
Transportation density	Number of transportation POIs/total number of transportation POIs

TABLE 2: Importance of variables and details.

Categories	Variables	Importance part	Details		
			Max	Min	Mean
Point of population	Building area (Ba)	0.001	1210.41	20.46	396.2831
	Population density (Pd)	0.01	34315	577	13024.688
Point of land use	Factory area share (Fas)	0.05	5986	123	1762.688
	Area share of school (Ass)	0.25	777	86	237.75
	Share of office floor space (Sofs)	0.2	2120	80	560.6042
	Share of shopping area (Ssa)	0.15	1820	102	509.6041
	Share of medical area (Sma)	0.15	130	9	47.8125
	Share of accommodation area (Saa)	0.03*100	384	23	94.91667
	Area share of film area (Asfa)	0.1*100	15	1	4.58333
	Share of park area (Spa)	0.03	790.38	16.99	194.0304
Point of interest (POI)	Food service density (Fsd)	0.15	0.234502	0.11793	0.065625
	Science and education service density (Sed)	0.25	0.212796	0.01	0.064312
	Health service density (Hsd)	0.25	0.23233	0.01724	0.063125
	Life service density (Lsd)	0.2	0.28812	0.001	0.06351
	Density of scenic spots (Dss)	0.1	0.19126	0.02	0.06291
	Transportation density (Td)	0.05	0.355389	0.01	0.106465

churches, temples, and other monuments. Life service facilities include courier services, business offices, offices, sales outlets, barber stores, banks, and travel agencies. Scientific, educational, and cultural services include schools, research institutions, training institutions, media institutions, and various knowledge and cultural pavilions. Transportation facility services include airports, railway stations, port terminals, various stations, and subway stations. Table 3 shows that the vitality index has been calculated and ranked for

the last three years for each administrative region. The heat map of urban vitality is shown in Figure 1.

3.2. Causal Machine Learning Preliminary.
The type of question that machine learning is currently very good at answering is the predictive one. Compared to simple linear models, machine learning models allow for more flexible relationships between dependent and independent variables and thus are better able to capture complex nonlinear

TABLE 3: Vitality index of each administrative region.

| Administrative district	2018		2019		2020	
	Score	Rank	Score	Rank	Score	Rank
Pu dong	8.809499	1	8.868721	1	8.952615	1
Huang pu	6.390004	2	6.311164	2	6.517	2
Min xing	6.146185	3	6.03454	3	5.835692	3
Jing an	5.791503	4	5.627998	4	5.667912	4
Xu hui	5.502869	5	5.326254	6	5.256	6
Bao shan	5.236984	6	5.078386	5	5.305133	5
Yang pu	5.200576	7	5.289908	7	5.028372	8
Hong kou	5.169065	8	5.062258	8	5.232655	7
Pu tuo	5.086245	9	4.972422	9	4.970099	9
Song jiang	5.027535	10	4.96695	10	4.941703	10
Jia ding	4.776406	11	4.776305	11	4.877614	11
Chang ning	4.660506	12	4.524272	12	4.7015	12
Qing pu	4.064888	13	4.21615	13	4.291804	13
Feng xian	3.957069	14	4.012427	14	4.267434	14
Jin shan	3.306435	15	3.35312	15	3.757067	15
Chong ming	3.011241	16	3.026724	16	3.069579	16
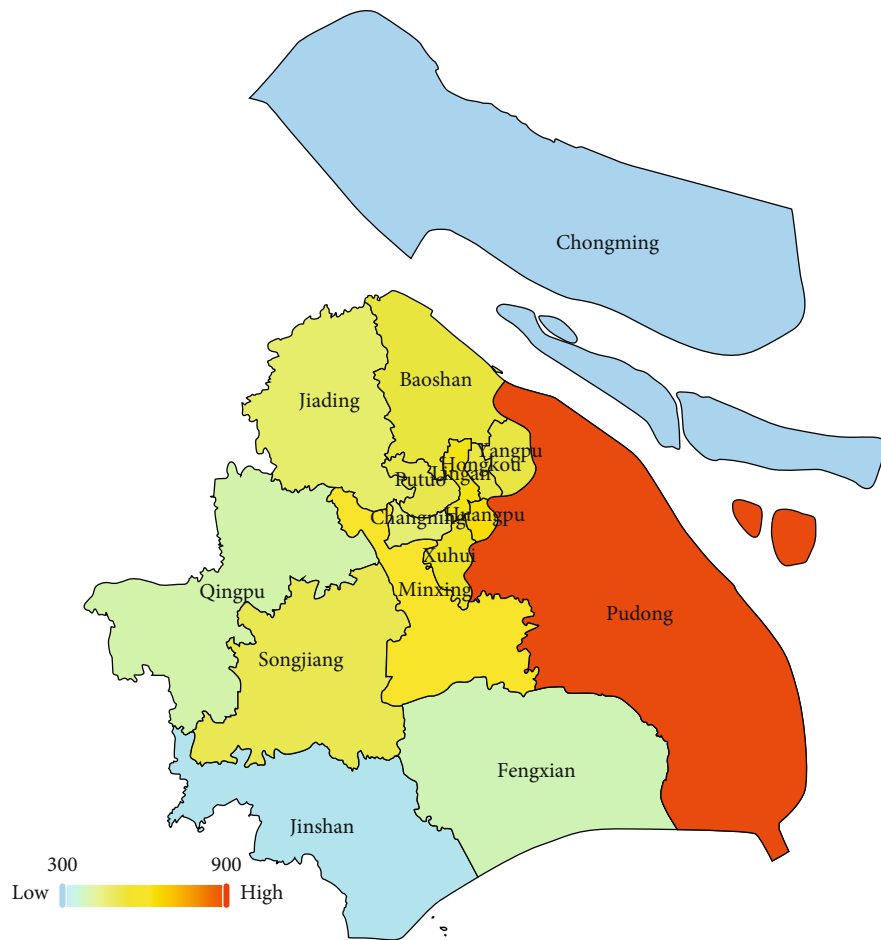
FIGURE 1: 2018 heat map of urban vitality index by district.

relationships between variables and are also better able to exploit the information embedded in nonnumerical variables to improve predictive accuracy [33] and we can do all kinds of wonderful things with machine learning things. The only requirement is that we define our problem as a prediction problem.

But the current machine learning is only used as a special function $f(x)$ that can be fitted to the data or to fit some polynomial distribution. But just a simple superfitting function is unable to understand the deep abstract relationship behind it like the human brain does and needs to be inferred from the logical relationship of facts to be more intelligent.

ML is not a panacea. It can perform miracles under very tight bounds, but it can be very ineffective if it uses data that deviates slightly from what the model is used to. ML is known to perform poorly on these types of reverse causality problems. They ask us to answer "what if" questions, which economists call counterfactuals. What would happen if instead of this price, I currently demanded a different price for a commodity? What would happen to urban vitality if I reduced the amount of green space in my city?

The core of the abovementioned question is causal inference, and we would like to know the answer. The causal effects estimated by causal inference studies are mainly obtained from the differences between the actual observations of the observed samples and their hypothesized or constructed counterfactual results, and the presence of the aforementioned influencing factors can affect the accuracy of the counterfactual results to varying degrees and consequently cause potential bias in the estimates of causal effects [34].

The reason why it is very difficult to solve the abovementioned problem is mainly that the link is not causal. For example, if a reduction in urban green space shows a decrease in urban vitality, is it true that the reduction in urban green space leads to a decrease in urban vitality? The answer is obviously not. If we need to study the effect of green space on urban vitality, we need to remove other urban data and estimate it unbiasedly.

Currently, in the fields of economic and financial policies, public policies, and health policies, there is an increasing interest in the response of individuals or subgroups to a particular policy or measure [35–37]. With the predictive superiority of machine learning methods, combined with causal inference, it can be better for unbiased estimation and be applied to the estimation of the abovementioned heterogeneity effects, which will be a very promising direction for theoretical research and practical application.

There are multiple different analytical frameworks for causal inference research, and the two main types of mainstream research nowadays are the "structural causal model (SCM)" [38] and the "potential outcome (PO) framework" [39]. The SCM framework can be described as a mathematical representation of a causal graph, where causality is a set of structural equations consisting of a series of nonlinear and nonparametric equations, and the causal effect is considered as the difference between the fact and the counterfactual. In the PO framework, the causal effect is viewed as the difference between the actual outcome of the sample in the experiment and the potential outcome that results from being subjected to randomization.

In terms of theoretical model representation, there are major differences between the SCM framework and the PO framework, with the SCM framework being more concerned with portraying complex relationships between variables and using cause-effect diagram tools to visually analyze and explain the relationships between various variables. However, as pointed out by [40], the PO framework has advantages over the SCM framework in five areas relative to the SCM framework: premise assumptions, economic theory relevance, model simplicity, heterogeneity assumptions, and experimental design. Therefore, this paper also intends to analyze the causal effects of urban multisource data on urban dynamics by using causal inference under the PO framework.

To analyze the impact of specific city data on urban vitality, we introduce a machine learning approach combined with causal inference to unbiasedly estimate the spatial impact of city-specific data on urban vitality in Shanghai.

To better illustrate the study that follows, we introduce some notation. The fundamental problem of causal inference is that something happens and we cannot observe its counterfactual because there are confounding variables. To solve this problem, we will conduct a study on potential outcomes, which we will call factual outcomes and nonoccurrence outcomes, counterfactuals. Use Y_{0i} to denote a potential outcome where event i does not occur; Y_{1i} denotes a potential outcome where event i occurs. With the potential outcomes, we can define the individual treatment effect:

$$\text{ITE} = Y_{1i} - Y_{0i}, \tag{5}$$

Usually, we focus more on the average treatment effect:

$$\text{ATE} = E[Y_1 - Y_0]. \tag{6}$$

The difference between supervised learning and causal inference is shown in Figure 2.

3.3. Construct the Causal Machine Learning Model.

The processing flow of causal inference is shown in Figure 3.

In social science, it is often difficult to conduct a truly "randomized" social experiment, so it is necessary to make full use of the actual observed data to create "randomized" conditions as much as possible. In reality, the observed data are biased, i.e., there are features X that affect both the target outcome OutcomeY and TreatmentT. Then, before causal modeling, we need to perform a debiasing process so that TreatmentT is independent of feature X, and only the effect of TreatmentT on OutcomeY is of interest.

3.3.1. Randomized Controlled Trial (RCT).

After that, causal modeling can be performed.

In causal inference, we use the do-operator to implement the randomized controlled experiment hypothesis, i.e., the hypothesis does $(T = t)$ represent the effect of using a TreatmentT on OutcomeY in a randomized social experiment without any confounding factors. After using the do-
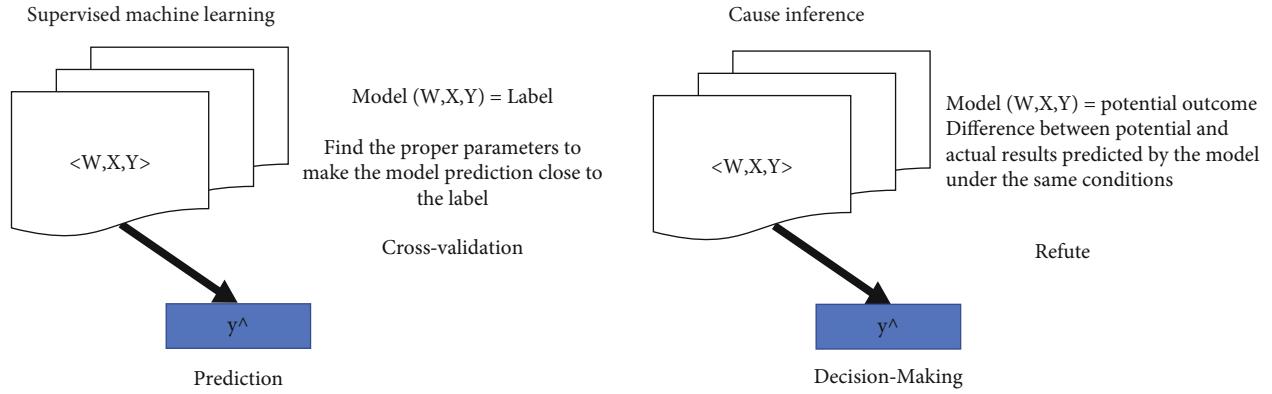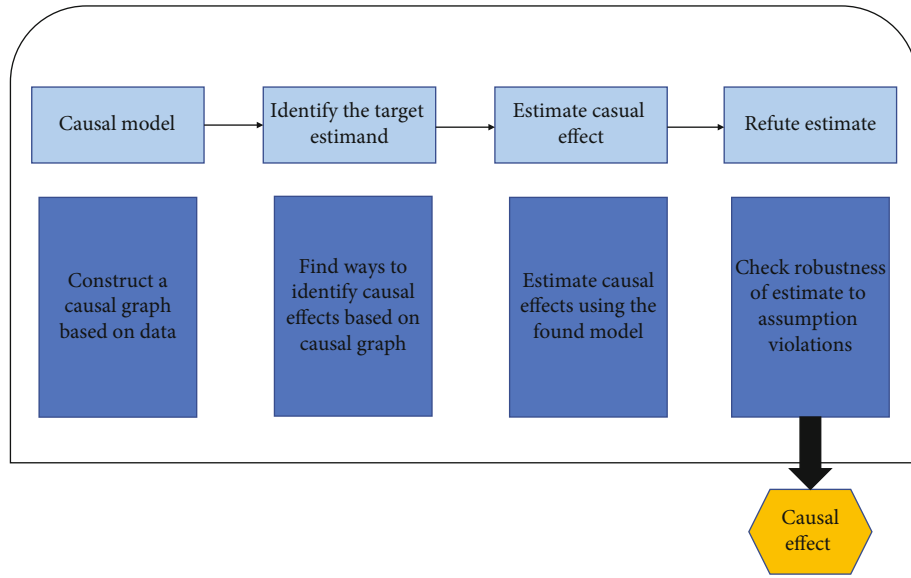
Supervised machine learning

Model (W,X,Y) = Label

Find the proper parameters to make the model prediction close to the label

Cross-validation

<W,X,Y>

y^

Prediction

Cause inference

Model (W,X,Y) = potential outcome
Difference between potential and actual results predicted by the model under the same conditions

Refute

<W,X,Y>

y^

Decision-Making

Figure 2: Supervised learning vs causal inference.

Causal model → Identify the target estimand → Estimate casual effect → Refute estimate

Construct a causal graph based on data

Find ways to identify causal effects based on causal graph

Estimate causal effects using the found model

Check robustness of estimate to assumption violations

Causal effect

Figure 3: The process of constructing causal inference.

operator again, ATE can be represented as follows:

$$\text{ATE} = E[Y|do(T=1) - Y|do(T=0)] \\ = E[Y|do(T=1)] - E[Y|do(T=0)]. \tag{7}$$

A central problem in many data-driven decision making scenarios is the estimation of the heterogeneous treatment effects (HTE). To briefly recall, our goal is to compute the average treatment effect (ATE), i.e., what is the causal effect of the intervention (T) on the output outcome (Y) for a sample with a particular feature set. In the analysis of the impact of urban multisource data on urban dynamics, since we need to analyze a specific variable, we need to calculate the CATE (conditional average causal effect conditional average treatment effect):

$$\text{CATE} = E[Y_1 - Y_0 | X = X_i], \\ \text{CATE} = E[Y(do(T=1)|X = x_i] - E[Y(do(T=0)|X = x_i]. \tag{8}$$

Condition X means that we now allow treatment effects to vary depending on the characteristics of each unit, because not all conditions respond well to treatment, and we want to take advantage of this heterogeneity to identify the causal effects of different conditions.

We can estimate the effect of conditional treatment in a certain way denoted by SY/ST; in this way, we can estimate the different CATE or elasticity of each partition, the elasticity being the slope of the function from T to Y. Thus, being able to generate partitions where the slope or elasticity is different means that entities on these partitions respond differently to the treatment.

Causal inference focuses on the unbiased estimation of feature effects on the target, and machine learning excels in giving accurate predictions. Double machine learning [41] combines the approach of causal inference with machine learning and uses a causal forest [42], which not only achieves a better sample than the traditional nearest neighbor matching method matching results, but also has point-

by-point consistency and asymptotic normality in the estimation of heterogeneous causal effects. In this paper, we use CausalForestDML abbreviated as DML to explore the impact of analyzing urban multisource data on urban vitality and we attempt to assess the causal effect of multisource data on urban vitality rather than the correlation between variables.

Next, we need to perform causal effects for estimation, which needs to be modeled first, using conditional outcome modeling: conditional outcome modeling (COM):

$$\tau = E_X[E[Y\, do(T=1)|X=x_i] - [E[Y\, do(T=0)|X=x_i], \quad (9)$$

where τ is the causal effect that we want.

To estimate the causal effect, the most straightforward idea is to model the two expectations shown using a causal forest. As mentioned before, we only care about the effect of TreatmentT on OutcomeY, so we can first regress T using X to get a residual of T, then regress Y using X to get a residual of Y, and finally regress Y using the residual of T residuals, and the estimated parameter is the CATE that we want.

$$(Y - (Y \sim X)) \sim (T - (T \sim X)),$$
$$Y_i - E[Y_i|X_i] = \tau * (T_i - E[T_i|X_i]) + \epsilon. \quad (10)$$

Thus, the specific DML method came out, the core idea of which is that the machine learning algorithm predicts T and Y based on X and then regresses the residuals of Y using the residuals of T.

$$Y_i - \widehat{M_y}(X_i) = \tau * \left(Y_i - \widehat{M_t}(X_i)\right) + \epsilon, \quad (11)$$

where $\widehat{M_y}(X_i)$ modeling $E[T|X]$ and $\widehat{M_t}(X_i)$ modeling $E[Y|X]$.

Why is it that DML can debias?

In fact, the key is $\widehat{M_t}(X_i)$, which implements the debiasing of treatment. The residual of T can be viewed as the amount left after removing the action of X on T from T. The residual of T is then independent of X.

And the role of $\widehat{M_y}(X_i)$ is to remove the variance of Y, i.e., the variance of Y caused by X is removed from Y.

DML is an algorithmic framework to obtain unbiased CATE estimation even if the W (nuisance parameter) estimation is biased by residual estimation moments (obeying Neyman orthogonality) during the study of HTE (heterogenous treatment effect).

Finally, the residual modeling is then regressed, i.e., CATE is obtained.

Next, we model the paper in the specific context of this paper.

We construct DML models to analyze the spatial causal effects of urban multiple-source data on urban dynamics. The following table is used to represent the labels of DML-related variables. The notations of DML are shown in Table 4.

The process for assessing urban vitality by causal inference is shown in Figure 4.

In this paper, the 2018 year and the 16 administrative districts of Shanghai are used as control variables because we need the same year and district to infer the heterogeneous treatment effect of urban multisource data on urban dynamics HTE. In addition to setting the other urban multisource data except as study variables to X, the goal is to estimate without any bias $\theta(X)$. If we can achieve this goal, the causal inference between Y and T can be captured as a way to better understand the heterogeneous causal effects of urban multisource data on urban dynamics.

First, we construct the linear regression DML model as equations (1) and (2) for discussion.

$$Y = \theta(X)T + g(X, W) + U, \quad \text{where } E[U|X, W] = 0, \quad (12)$$

$$T = f(X, W) + V, \quad \text{where } E[V|X, W] = 0, \quad (13)$$

where $g(X, W)$ and $f(X, W)$ are trained using two machine learning methods and U and V are random noise. θ is a parameter representing the causal effect. Given the exact parameters X and control variables W, the effect of urban research data on urban dynamics can be inferred.

In order to estimate $\theta(X)$, (12) can be converted to (14) as follows:

$$Y - E[Y|X, W] = \theta(X)(T - E[T|X, W]) + U. \quad (14)$$

After that, we use random forest to train two models $g(X, W)$ and $f(X, W)$ to represent the two expectations in (14).

$$g(X, W) = E[Y|X, W],$$
$$f(X, W) = E[T|X, W]. \quad (15)$$

In the next step, equation (14) can be expressed in terms of the residuals of the actual and predicted values, as shown in equation (16):

$$Y - g(X, W) = \theta(X)(T - f(X, W)) + U. \quad (16)$$

Finally, we can get $\theta(X)$ as in equation (17), the estimated value of $\theta(X)$ using random forest to train the final model.

$$\theta(\widehat{X}) = \operatorname{argmin} E[(Y - q(X, W)) - \theta(X)(T - f(X, W))]. \quad (17)$$

How can we assess the validity of the causal effects that we derived? We use the refutation method to verify.

3.3.2. Effect Estimates for Our Causal Estimator. The main rebuttal is made using the following methods.

(1) Adding random common causes: the robustness of causal inference is tested by estimating whether the method will change its estimates after adding a confounder as a common cause to the dataset. Expected

TABLE 4: Notations of DML.

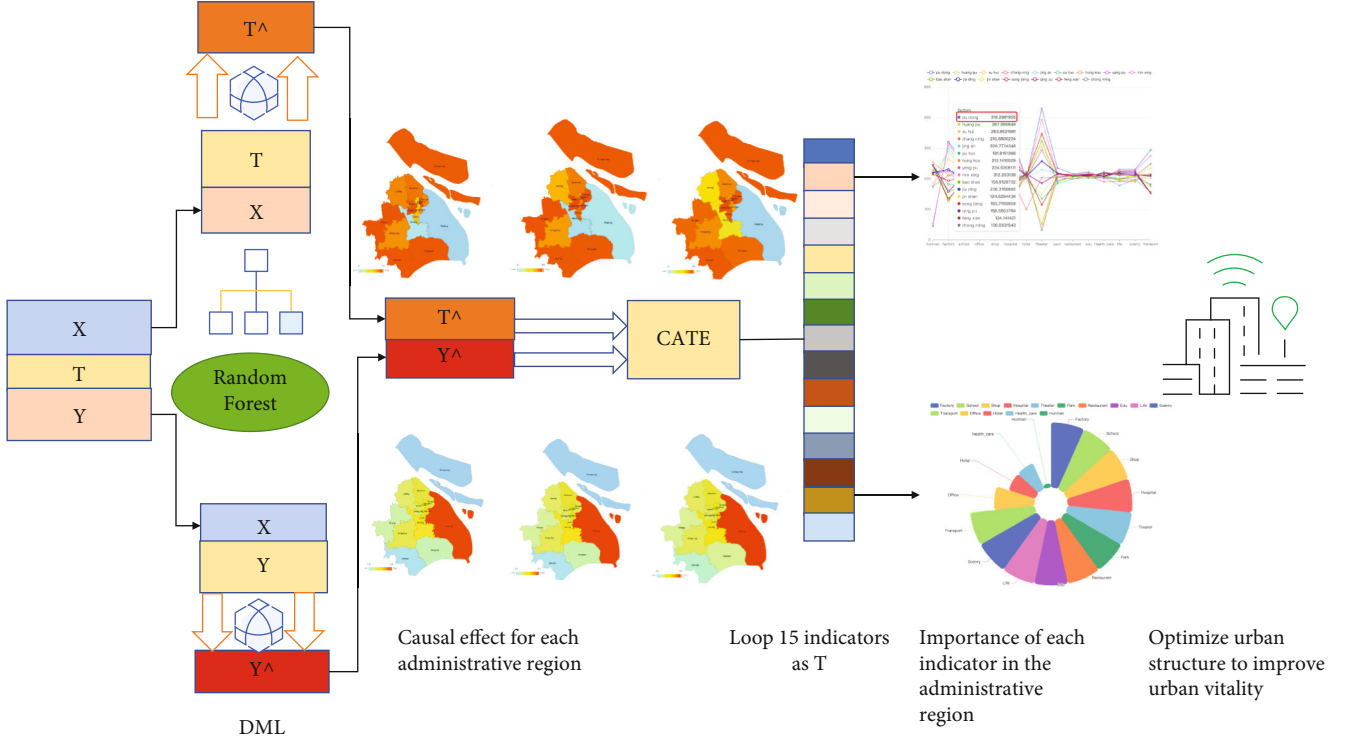| Notations	Variables	Descriptions
Y	Outcome variable	Urban vitality at Y at S (suburb)
T	Treatment variable	Study-specific city data at Y
X	Features	Excluding study-specific city data
W	Control variable	Precipitation at stage Y, S
$\theta(X)$	Treatment effect	Given X, how much influence Treatment T has on Outcome Y

FIGURE 4: Process for assessing the causal effects of urban multisource data on urban vitality.

value: the new causal effect is similar to the previous causal effect

(2) Adding unobserved variables: it adds an unobserved confounding variable to the dataset and whether the estimation method changes its estimate. Expected value: the new causal effect is similar to the previous causal effect

(3) Placebo treatment: what happens to the estimated causal impact when replacing real variables with independent random variables? Expected value: the new causal effect tends to be close to zero

The estimated values are shown in Table 5.

We can conclude that the overall robustness of causal inference is still relatively stable and there are some discrepancies in some individual indicators, so the heterogeneous causal effects that we derived using DML are more convincing.

To compare more clearly the differences between the previous work and our model, we have selected some representative work for comparison; the results are shown in Table 6.

4. Results

A heat map of causal effects is shown in Figure 5.

Taking the 2018 data as an example, we take 15 indicators in turn as T to analyze their causal effects on the 16 administrative regions of Shanghai. After debiased estimation by the abovementioned modeling method, the relative contribution of each indicator in each region to the region can be obtained, as a way to explore the analysis of which indicator in the region has a greater impact on the importance of the urban vitality index in the region. We conduct the analysis as follows:

4.1. Pudong New Area Analysis. Pudong New Area is a municipal district of Shanghai, named after its location east

TABLE 5: Urban multisource data robustness test for causal inference for each data (bold = poor effect).

Data refute	Random vs new random	Unobserved vs new unobserved	Placebo vs new placebo
Human	0.00239 vs 0.00144	0.00239 vs 0.001315	0.00239 vs 0.000488
Factory	0.0597 vs 0.0588	0.0597 vs 0.0600	0.0597 vs 0.000823
School	0.465 vs 0.444	0.465 vs 0.482	0.465 vs −0.129
Office	0.168 vs 0.175	0.168 vs 0.178	0.168 vs −0.000200
Shop	0.326 vs 0.311	0.326 vs 0.300	0.326 vs 0.0222
Hospital	1.700 vs 1.380	1.700 vs 2.00	1.700 vs −0.340
Hotel	1.0373 vs 0.980	1.0373 vs 0.966	1.0373 vs −0.0881
Theater	−4.0220 vs **−11.437**	−4.0220 vs **−11.722**	−4.0220 vs −1.40567
Park	0.358 vs 0.387	0.358 vs 0.358	0.358 vs −0.00456
Restaurant	2090.452 vs 1991.236	2090.452 vs **149.234**	2090.452 vs **102.148**
Edu	1944.911 vs 1780.390	1944.911 vs **154.735**	1944.911 vs **78.394**
Health care	2144.268 vs 2060.782	2144.268 vs **176.677**	2144.268 vs **−317.822**
Life	1461.368 vs 1520.191	1461.368 vs **230.915**	1461.368 vs **−100.435**
Scenery	1980.187 vs 2093.490	1980.187 vs **−36.0728**	1980.187 vs **−131.854**
Transport	249.075 vs 366.321	249.075 vs 83.696	249.075 vs −24.373

TABLE 6: Results of comparison with previous work.

Research name	Research area	Research variables	Methods	Relationships
[27]	Yangtze River Delta region	Economy	Principal component analysis	Correlation
[28]	174 European cities	Public green space	Data envelopment analysis technique	Correlation
Yue et al.,2019 [30]	Wuhan	Street centrality	Spatial regression models	Correlation
Sung et al.,2019 [13]	Seoul	Walking activity	Multilevel regression models	Correlation
Lu S et al.,2019 [25]	Beijing and Chengdu	Social media check-in data	Regression model	Correlation
Our research	Shanghai	Multisource data	Causal machine learning	*Causality*

of the Huangpu River. Pudong is bordered by Fengxian district and Minhang district in the south, Xuhui district, Huangpu district, Hongkou district, Yangpu district, and Baoshan district in the west and Chongming district in the north, across the Huangpu River. According to the data of the 7th census, Pudong New Area has a resident population of 5,681,512. Pudong New Area is an important transportation hub in Shanghai and an advanced international logistics port. Air transportation, railroad rail transportation, and intercity highway together build a trinity of water, land, and air transportation system. The Pudong River Bridge, submarine tunnel, Maglev trains, and subway lines form a dense transportation network.

The contribution of the causal effect values of each indicator in Pudong New Area relative to the 16 districts according to the model is as follows: human (16), factory (1), school (1), office (7), shop (1), hospital (1), hotel (10), theater (1), park (1), restaurant (1), Edu (1), health_care (10), life (1), scenery (1), and transport (1). The contribution of each indicator to regional vitality in Pudong New Area is shown in Figure 6.

Pudong New Area ranks first among the 16 administrative districts in the urban vitality. As one of the fastest-growing administrative districts in Shanghai since the reform and opening up, Pudong New Area's ranking of the

relative contribution to the multisource data, 11 indicators ranked first among the districts. However, Office is ranked 7th, and hotel, health_care is ranked 10th. Next, Pudong New Area should build on its strengths and find ways to improve some supporting facilities of office, hotel, and health_care to further stimulate the city's vitality.

4.2. Huangpu District Analysis. Huangpu district, which is part of Shanghai, is located at the southwest end of the confluence of Huangpu River and Suzhou River. According to the data of the 7th census, the resident population of Huangpu district was 662,030. As the urban origin and transportation hub center of Shanghai, Huangpu district has a very convenient three-dimensional transportation network.

The contribution of the causal effect values of each indicator in Huangpu district relative to the 16 districts according to the model is as follows: human (12), factory (4), school (4), office (12), shop (14), hospital (10), hotel (14), theater (10), park (6), restaurant (12), Edu (8), health_care (16), life (11), scenery (13), and transport (10). The contribution of each indicator to regional vitality in Huangpu district is shown in Figure 7.

Huangpu district ranks second among the 16 administrative districts in the urban vitality. Among the 15
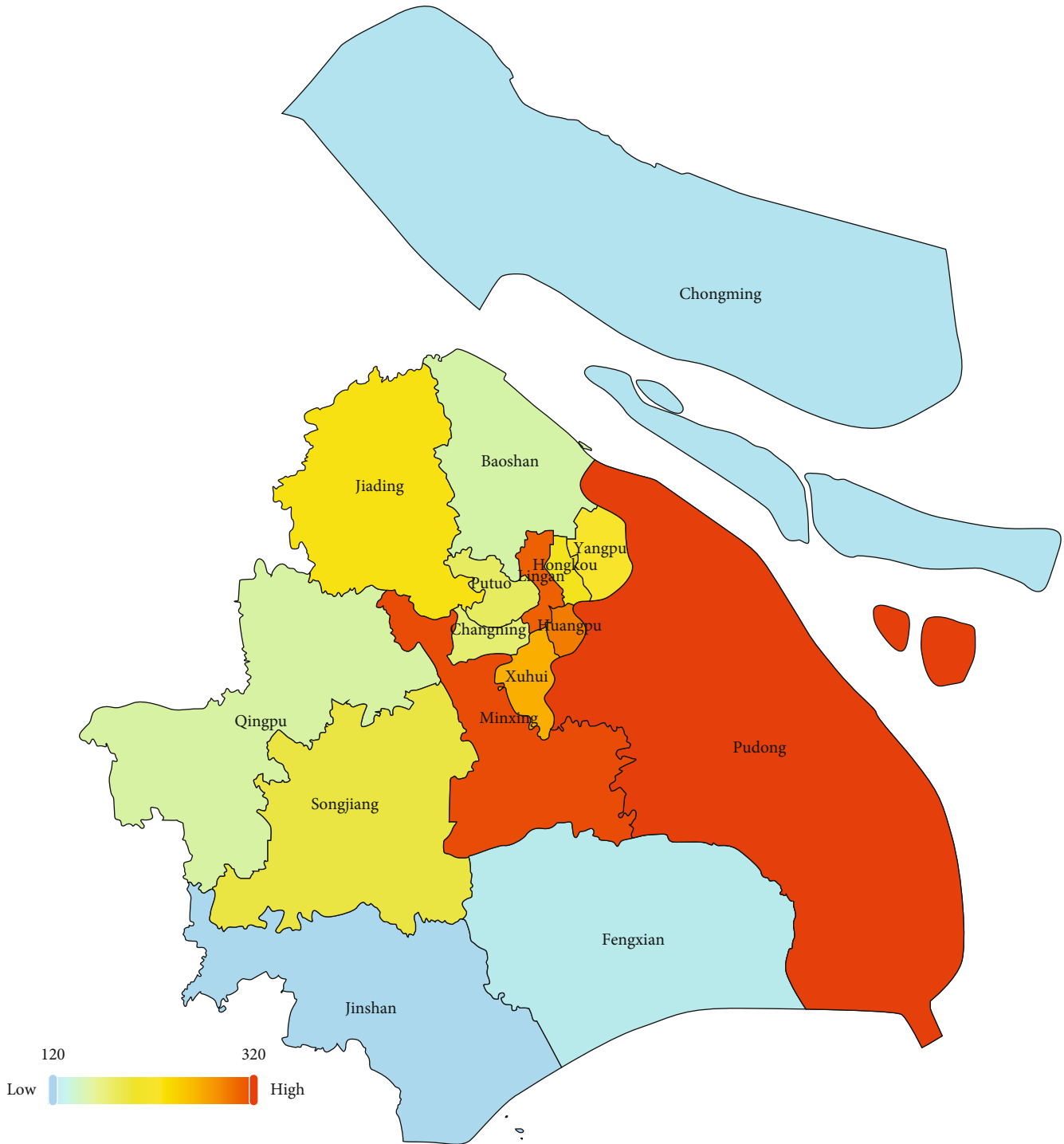
Figure 5: A heat map of causal effects by administrative regions of factory in 2018.

indicators of Huangpu district, factories, schools, theaters, and parks are highly competitive and play a major role in the district's urban vitality but the two indicators of hotels and health services play a smaller role in the district's urban vitality. Considering that Huangpu district is located in the city center and hotels are more expensive, we can consider the next two indicators to improve urban services and further stimulate urban vitality.

4.3. Xuhui District Analysis. Xuhui district, a district under the jurisdiction of the city of Shanghai, is located in the southwestern part of the central city of Shanghai, adjacent to Huangpu district in the northeast, facing the Huangpu River in the east and looking across the river from Pudong New Area, divided from Minhang district in the west, and bordering Jing'an district and Changning district in the north. According to the data of the 7th census, the Xuhui
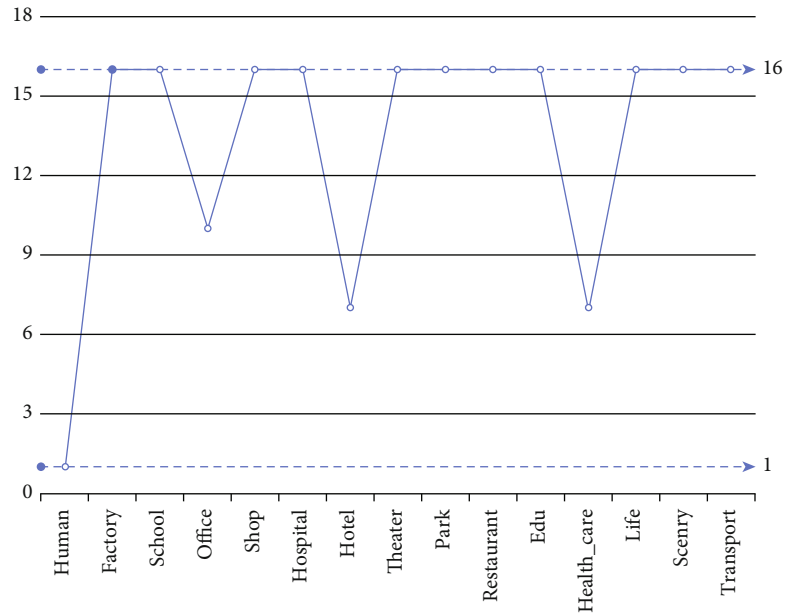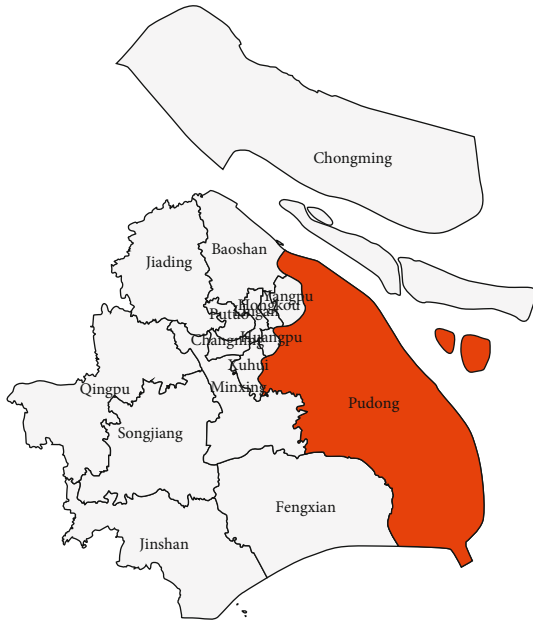
FIGURE 6: Contribution of each indicator to regional vitality in Pudong New Area in 2018.
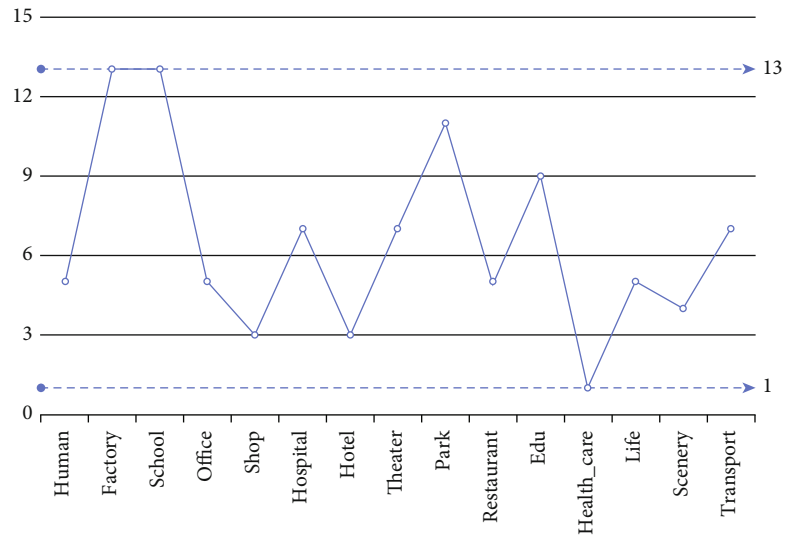
FIGURE 7: Contribution of each indicator to regional vitality in Huangpu district in 2018.

district resident population is 1,113,078. Xuhui is one of the central urban areas in Shanghai that completed the transformation of old areas earlier. Xuhui focuses on the development of information technology, modern biomedicine, new nanomaterials, and other high-tech industries, the formation of the district industrial electronic information industry, and biomedical industry development on an industrial scale. Xuhui territory has Shanghai Jiaotong University, East China University of Science and Technology, Shanghai Normal University, and other institutions of higher learning: Shanghai Botanical Garden, Shanghai Song Qingling Former Residence Memorial Hall, and Xujiahui source for the national 4A-level tourist attractions.

The contribution of the causal effect values of each indicator in Xuhui district relative to the 16 districts according to the model is as follows: human (13), factory (5), school (5), office (13), shop (10), hospital (8), hotel (15), theater (5), park (9), restaurant (9), Edu (6), health_care (13), life (10), scenery (10), and transport (7). The contribution of each indicator to regional vitality in Xuhui district is shown in Figure 8.
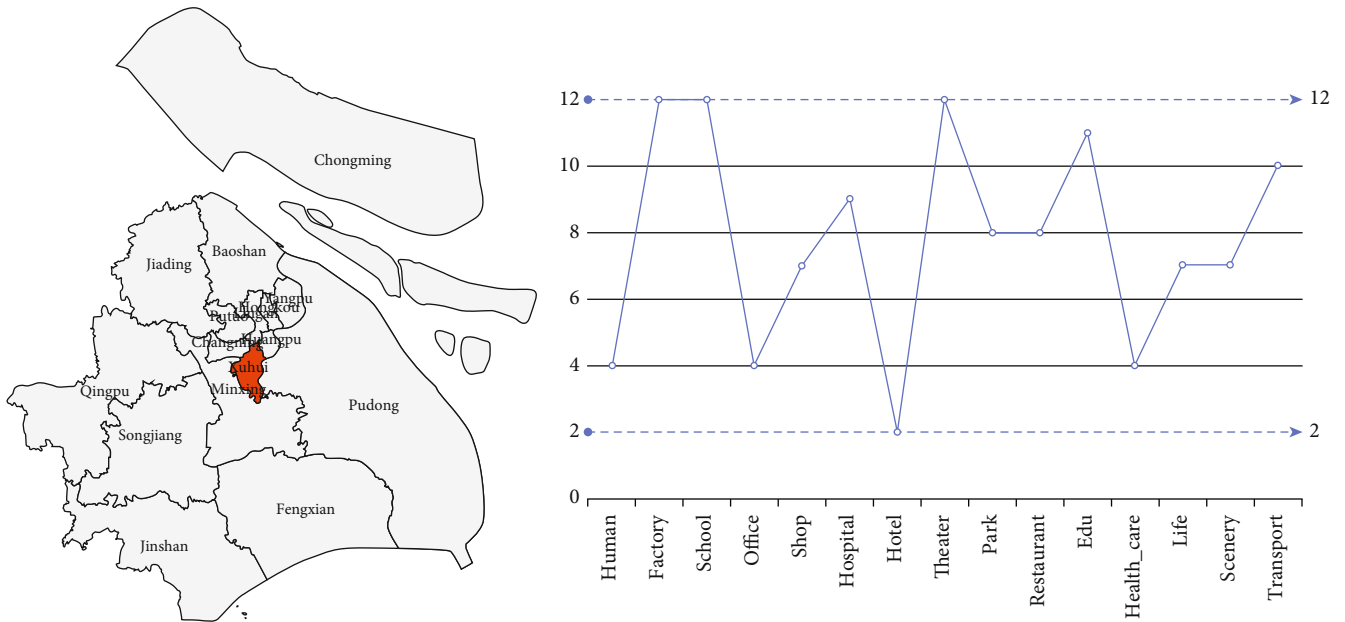
FIGURE 8: Contribution of each indicator to regional vitality in Xuhui district in 2018.

Xuhui district ranks fifth among the 16 administrative districts in the urban vitality. Xuhui district has many local colleges and universities, so schools and education are the main indicators to drive the district's urban vitality, and considering a large number of local students, the district can add more supporting infrastructure, increase cultural and recreational facilities, and develop high-tech industries to further stimulate urban vitality.

4.4. Changning District Analysis.

Changning district is located in the western part of the central city of Shanghai, connecting Jing'an district to the east, Minhang district to the west and southwest, Xuhui district to the southeast, and Putuo district to the north with the Wusong River (Suzhou Creek) as the boundary. According to the data of the 7th census, 693,051 people reside in Changning district. Changning has the most concentrated foreign-related high-standard residential business office complex in Shanghai, Gubei New District, Xinhua road, the traditional high-class residential area of old Shanghai, Hongqiao Garden Villa District, middle- and high-class residential areas along Suzhou River, and ordinary new village residential areas. Changning district is strategically located with convenient transportation and is the "Y" pivot point where the Shanghai-Nanjing and Shanghai-Hangzhou development axes converge and are the "bridgehead" connecting Shanghai to the Yangtze River Delta.

The contribution of the causal effect values of each indicator in Changning district relative to the 16 districts according to the model is as follows: human (2), factory (11), school (7), office (11), shop (13), hospital (14), hotel (5), theater (13), park (13), restaurant (15), Edu (16), health_care (6), life (15), scenery (16), and transport (11). The contribution of each indicator to regional vitality in Changning district is shown in Figure 9.

Changning district ranks 12th among 16 administrative districts in terms of urban vitality. Changning district is mainly a residential area, so schools, hotels, and office buildings with high population density are the indices that mainly stimulate the urban vitality of the district, and the district can develop more facilities for people's convenience to further optimize resources to enhance urban vitality.

4.5. Jing'an District Analysis.

Jing'an district is part of Shanghai, located in the center of Shanghai, with six districts adjacent to the east and Huangpu district, Hongkou district, Baoshan district as neighbors; west Changning district, Putuo district, and Baoshan district border; south of Changle road, and Xuhui District interface; and north and Baoshan district border. According to the data of the 7th census, Jing'an district's resident population is 975,707 people. Jing'an district is strategically located, with convenient transportation, and is well connected by rail, elevated, subway, bus, and other transportation networks and is known as Shanghai's "northern gate on land."

The contribution of the causal effect values of each indicator in Jing'an district relative to the 16 districts according to the model is as follows: human (14), factory (3), school (3), office (14), shop (12), hospital (5), hotel (16), theater (8), park (4), restaurant (10), Edu (4), health_care (15), life (8), scenery (9), and transport (6). The contribution of each indicator to regional vitality in Jing'an district is shown in Figure 10.

Jing'an district ranks fourth among the 16 administrative districts in the urban vitality. Jing'an district is located in the northern part of Shanghai's center and has an advantageous location, which has a strong positive effect on urban vitality in many indicators, but the contribution of office areas, shopping malls, hotels, and health care to the district's urban vitality is relatively weak and has some potential, so the district can further optimize these resources in depth.
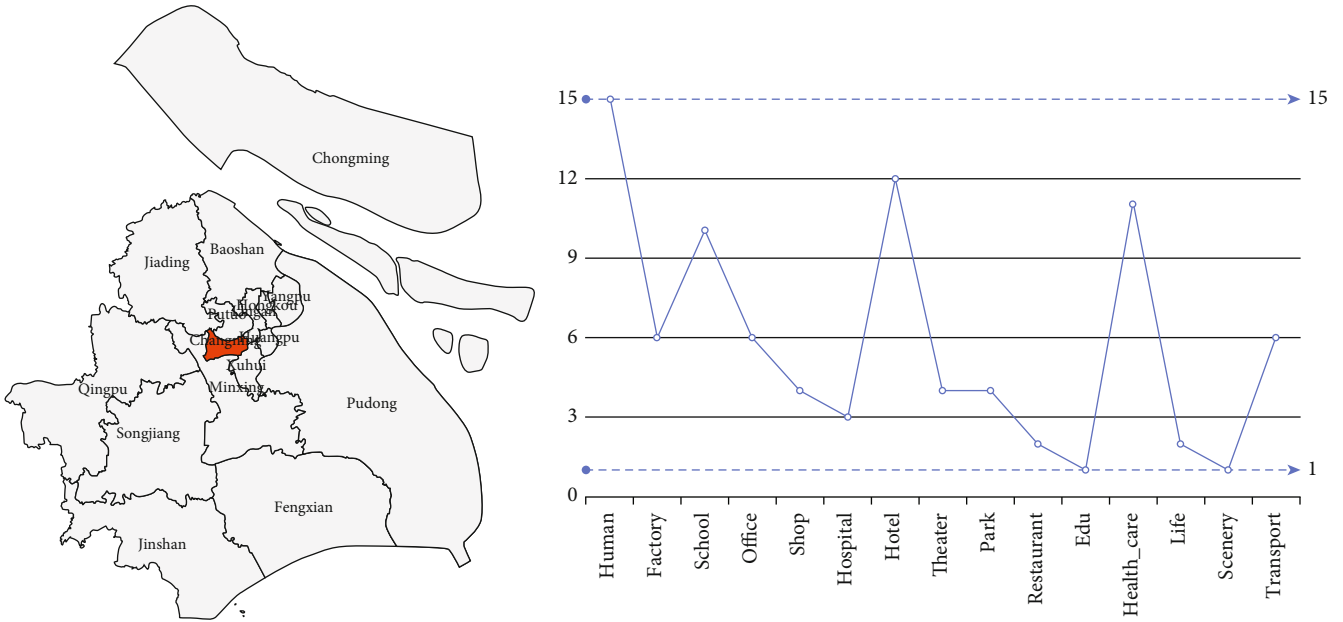
FIGURE 9: Contribution of each indicator to regional vitality in Changning district in 2018.
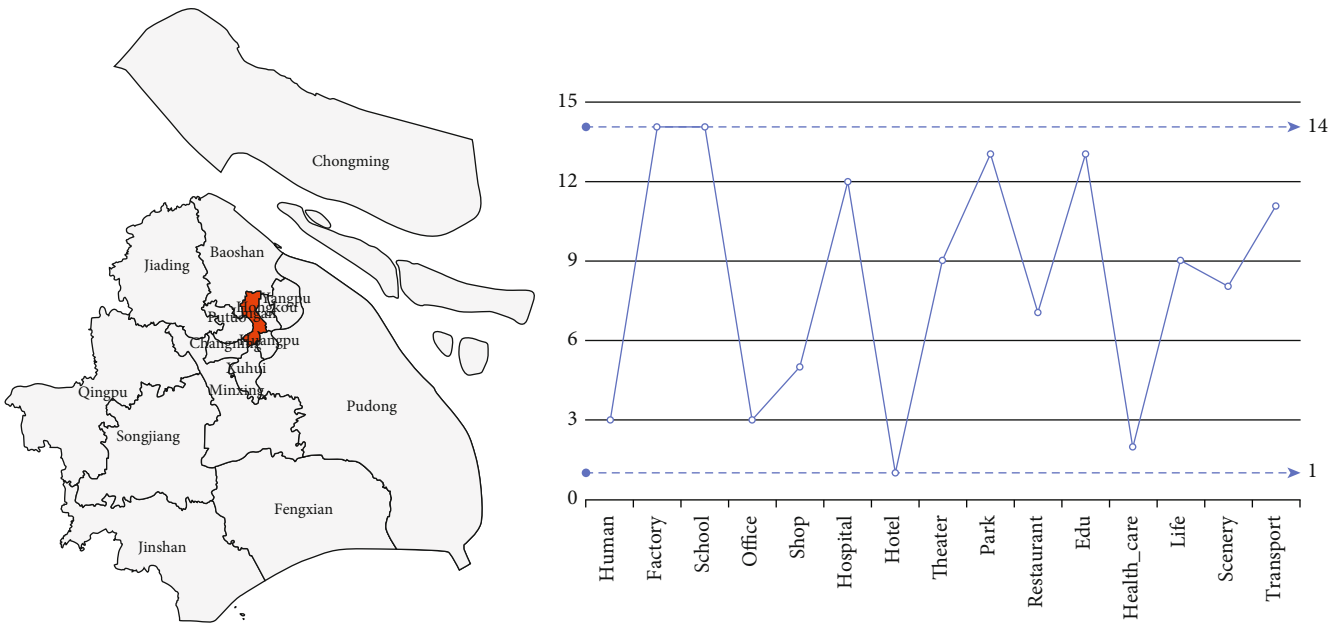

FIGURE 10: Contribution of each indicator to regional vitality in Jing'an district in 2018.

Therefore, the district can further optimize the allocation of these resources to better stimulate urban vitality.

4.6. Putuo District Analysis. Putuo district, which belongs to Shanghai, is located in the northwest of Shanghai's central urban area and is the starting point of the Shanghai-Nanjing development axis, as well as an important land gateway and transportation hub connecting Shanghai with the Yangtze River Delta and the Mainland. It is bordered by Jiading district and Baoshan district to the west and north and by Jing'an district and Changning district to the east

and south. According to the data of the 7th census, Putuo district had a resident population of 1,239,800. It is the starting point of the Shanghai-Nanjing development axis and an important land gateway and transportation hub connecting Shanghai to the Yangtze River Delta and the Mainland.

The contribution of the causal effect values of each indicator in Putuo district relative to the 16 districts according to the model is as follows: human (6), factory (10), school (16), office (16), shop (15), hospital (11), hotel (11), theater (11), park (16), restaurant (16) Edu (15), health_care (11), life (16), scenery (14), and transport (14). The contribution of
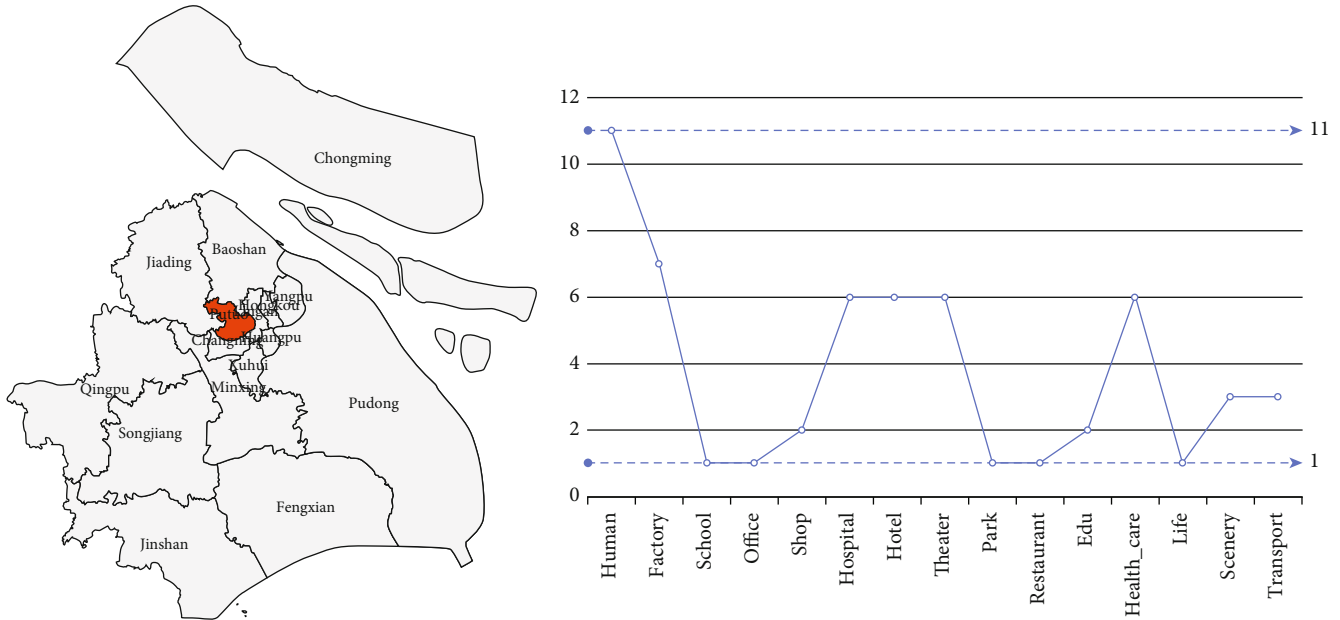
FIGURE 11: Contribution of each indicator to regional vitality in Putuo district in 2018.

each indicator to regional vitality in Putuo district is shown in Figure 11.

The urban vitality of Putuo district ranks ninth among the 16 administrative districts. The contribution of each of the 15 indicators of Putuo district to the urban vitality of the district is relatively average, with certain advantages in factories, hospitals, and theaters. The district should find advantages unique to the district, such as developing some industries with special characteristics and building more parks to stimulate the urban vitality of the district.

4.7. Hongkou District Analysis. Hongkou district, a district under the jurisdiction of the city of Shanghai, is located in the northeastern part of the central city of Shanghai, bordering Yangpu district in the east, adjoining Jing'an district in the west, looking across the river from Pudong New District and adjacent to Huangpu district in the south, and bordering Baoshan district in the north. According to the data of the 7th census, Hongkou district has a resident population of 757,498.

The contribution of the causal effect values of each indicator in Hongkou district relative to the 16 districts according to the model is as follows: human (1), factory (8), school (6), office (10), shop (16), hospital (15), hotel (7), theater (14), park (2), restaurant (13), Edu (12), health_care (9), life (14), scenery (15), and transport (8). The contribution of each indicator to regional vitality in Hongkou district is shown in Figure 12.

Hongkou district ranks eighth among the 16 administrative districts in the city vitality, with one indicator ranking first among the 16 administrative districts, which is human, and the second indicator of park, which is very competitive in these two indicators, but life, and scenery are relatively weak; the district should improve the supporting facilities

around park and increase recreational activities to further stimulate urban vitality.

4.8. Yangpu District Analysis. Yangpu district, a part of Shanghai, is located in the northeastern part of Shanghai's central city, on the northwestern bank of the lower Huangpu River, across the river from Pudong New Area to the west of Hongkou district, and the north from Baoshan. According to the data of the 7th census, 1,242,548 people reside in Yangpu District. Yangpu is located in one of the four major subcenters of Shanghai and one of the ten major commercial centers, the new Jiangwan city, the third-generation international community in Shanghai, the Tongji Knowledge-Economy Circle, the Dalian Road Headquarters R&D Cluster, and the East Bund, where the world's top 500 companies gather. Yangpu is rich in scientific and educational resources, with 14 colleges and universities located in the area, the number of which exceeds one-third of the total number of colleges and universities in Shanghai, and is known as the "central district of Shanghai's universities." It is known as the "central district of Shanghai", including Fudan University, Tongji University, Shanghai University of Finance and Economics, Shanghai University of Technology, Second Military Medical University, and other universities.

The contribution of the causal effect values of each indicator in Yangpu district relative to the 16 districts according to the model is as follows: human (10), factory (7), school (10), office (15), shop (11), hospital (7), hotel (12), theater (6), park (14), restaurant (14), Edu (11), health_care (12), life (13), scenery (12), and transport (12). The contribution of each indicator to regional vitality in Yangpu district is shown in Figure 13.

Yangpu district ranks seventh among the 16 administrative districts in the city vitality. The district has many
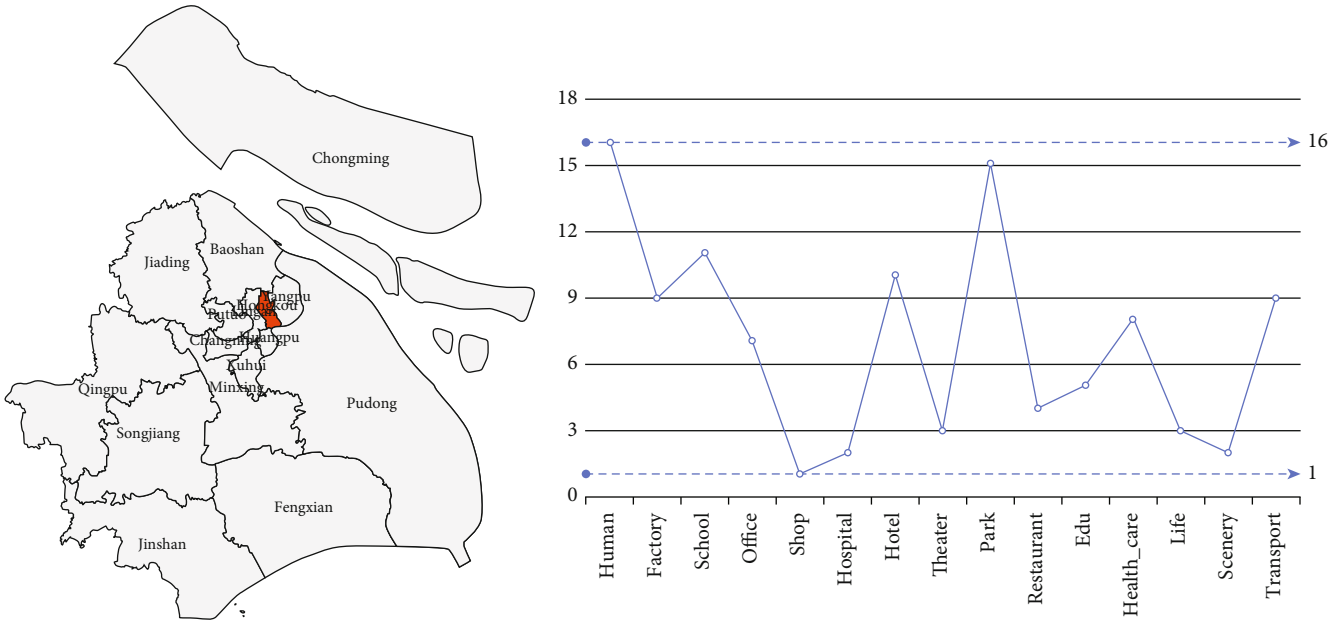
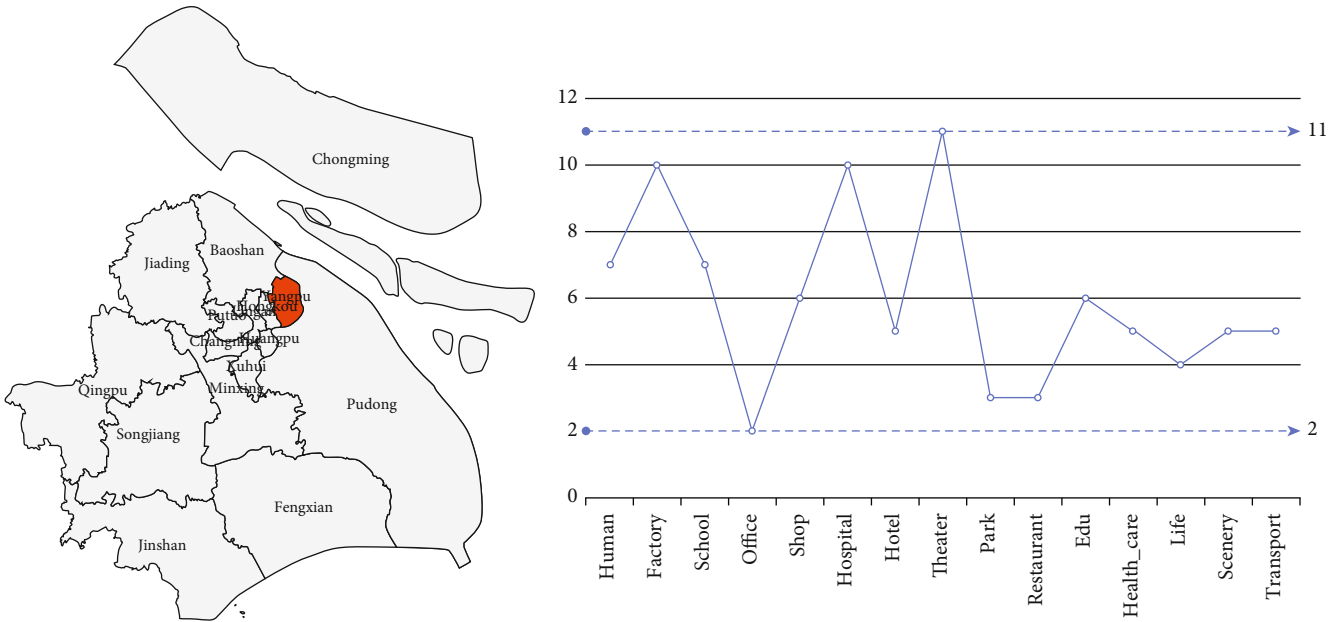FIGURE 12: Contribution of each indicator to regional vitality in Hongkou district in 2018.

FIGURE 13: Contribution of each indicator to regional vitality in Yangpu district in 2018.

colleges and universities and is highly competitive in factory, school, shop, and theater. Next, it is recommended that the district improves the supporting facilities around colleges and universities, develops diversified entertainment facilities, and increases innovation to further stimulate the city's vitality.

4.9. Minhang District Analysis. Minhang district is a part of Shanghai, located in the middle of Shanghai, shaped like a "key," bordering Xuhui district and Pudong New District in the east, Fengxian district in the south by Huangpu River,

Songjiang district and Qingpu district in the west, and Changning district and Jiading district in the north. According to the data of the 7th census, 2,653,489 people reside in Minhang district. The construction of the core area of the southern Shanghai Science and Innovation Center is at full speed. Focusing on the south of Minhang, the south riverside area around Jiaotong University and Huashi University, the "Big Zero Bay Global Innovation and Entrepreneurship Cluster" will be built, with 300,000 square meters of new science and innovation carriers along Jianchuan road and Canyuan road. After the incubation space for innovation

and entrepreneurship, four enterprises have landed. During the year, 81 projects were awarded above municipal level science and technology progress awards.

The contribution of the causal effect values of each indicator in Minhang district relative to the 16 districts according to the model is as follows: human (15), factory (2), school (2), office (8), shop (2), hospital (2), hotel (13), theater (2), park (8), restaurant (2), Edu (2), health_care (14), life (2), scenery (2), and transport (2). The contribution of each indicator to regional vitality in Minhang district is shown in Figure 14.

Minhang district urban vitality ranks third among 16 administrative districts, with a superior geographical location and a number of indicators in the top 15 indicators. Minhang district should take advantage of its superior geographical location to increase the development of transportation support facilities, but office is relatively weak, and the district can make up for the weakness by supporting more office with the advantage of transportation to further stimulate urban vitality.

4.10. Baoshan District Analysis.

Baoshan district is located in the north of Shanghai, bordering the Yangtze River in the northeast, the Huangpu River in the east, adjoining Yangpu, Hongkou, Jing'an, and Putuo in the south, bordering Jiading District in the west, and neighboring Taicang city in Jiangsu province in the northwest. According to the data of the 7th census, 2,235,218 people reside in Baoshan district. Located at the intersection of Huangpu River and Yangtze River, Baoshan is the "waterway gateway" of Shanghai, connecting more than 400 ports in 164 countries and regions by sea, with container throughput accounting for more than 70% of Shanghai's port, and well-developed intermodal transportation and inland river shipping; on land, it has formed a well-connected transportation network of railroads, railways, and highways. Hongqiao and Pudong international airports are within half an hour's drive away.

The contribution of the causal effect values of each indicator in Baoshan district relative to the 16 districts according to the model is as follows: human (8), factory (13), school (15), office (9), shop (4), hospital (6), hotel (9), theater (4), park (15), restaurant (11), Edu (10), health_care (7), life (12), scenery (6), and transport (13). The contribution of each indicator to regional vitality in Baoshan district is shown in Figure 15.

Baoshan district ranks sixth among the 16 administrative regions in the city vitality. Baoshan district has certain advantages in the 15 indicators of shop, hospital, theater, and scenery more forward, but the population is small, so Baoshan district should attract a large number of talents and improve the transportation infrastructure to further stimulate urban vitality.

4.11. Jiading District Analysis.

Jiading district, which belongs to Shanghai, is located in the northwestern part of Shanghai. According to the data of the 7th census, 1,834,258 people reside in Jiading district. In 2018, 21 new municipal science and technology small giant (incubation) enterprises were added to Jiading district and the number of municipal science and technology small giant (incubation) enterprises in the district reached 207 times.

The contribution of the causal effect values of each indicator in Jiading district relative to the 16 districts according to the model is as follows: human (9), factory (6), school (12), office (6), shop (9), hospital (4), hotel (8), theater (7), park (3), restaurant (8), Edu (5), health_care (8), life (4), scenery (4), and transport (16) (available for year analysis). The contribution of each indicator to regional vitality in Jiading district is shown in Figure 16.

The urban vitality of Jiading district ranks 11th among 16 administrative districts, and hospital, park, Edu, life, and scenery have a greater contribution to urban vitality in each index, but transport is at the last of all districts, which indicates that the transport system of Jiading district is relatively backward, which largely restricts the impact of urban vitality, and next, we need to increase the improvement of the transportation system to further enhance the city vitality.

4.12. Jinshan District Analysis.

Jinshan district is part of Shanghai, a distant suburb in the southwest of Shanghai, located south of the Yangtze River and on the south bank of the upper Huangpu River in China, bordering Fengxian district in the east, Songjiang district and Qingpu district in the north, and Pinghu and Jiashan in Zhejiang province in the west. According to the data of the 7th census, the resident population of the Jinshan district was 822,776.

The contribution of the causal effect values of each indicator in Jinshan district relative to the 16 districts according to the model is as follows: human (5), factory (16), school (14), office (3), shop (7), hospital (13), hotel (2), theater (15), park (12), restaurant (5), Edu (14), health_care (2), life (9), scenery (11), and transport (4). The contribution of each indicator to regional vitality in Jinshan district is shown in Figure 17.

Jinshan district urban vitality ranks 15 among 16 administrative districts, at the seaside so some tourist facilities have strong competitiveness, in shop, hotel, restaurant, health_care, and other indicators which contribute to the urban vitality of the district; the district should continue to improve tourism resources while strengthening the construction of transport infrastructure to further attract people to improve urban vitality.

4.13. Songjiang District Analysis.

Songjiang district, located in the southwestern part of Shanghai, has a long history and culture and is known as the "root of Shanghai." It is located in the upper reaches of Huangpu River, neighboring Minhang district and Fengxian district in the east, bordering Jinshan district in the south and southwest, and bordering Qingpu district in the west and north. According to the data of the 7th census, the resident population of Songjiang district was 1,909,713.

The contribution of the causal effect values of each indicator in Songjiang district relative to the 16 districts according to the model is as follows: human (11), factory (9), school (9), office (5), shop (3), hospital (3), hotel (6), theater (3), park (5), restaurant (4), Edu (3), health_care (5), life (3), scenery (3), and transport (15). The contribution of each
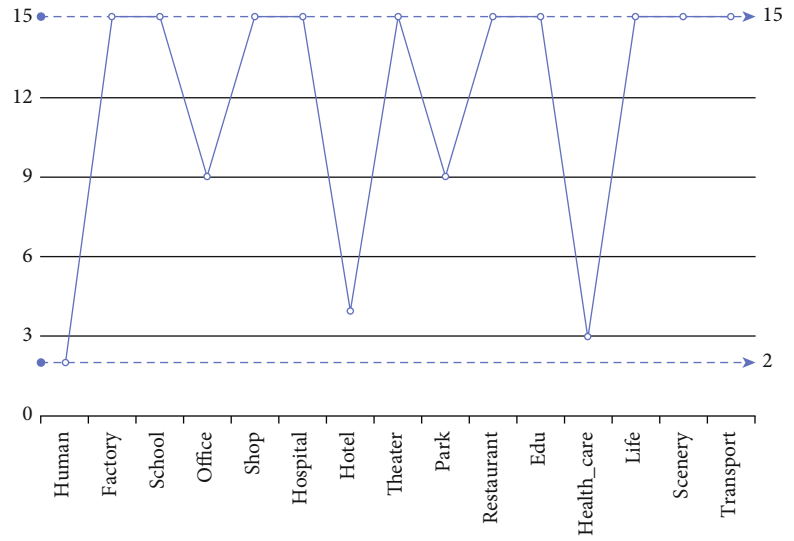
FIGURE 14: Contribution of each indicator to regional vitality in Minhang district in 2018.
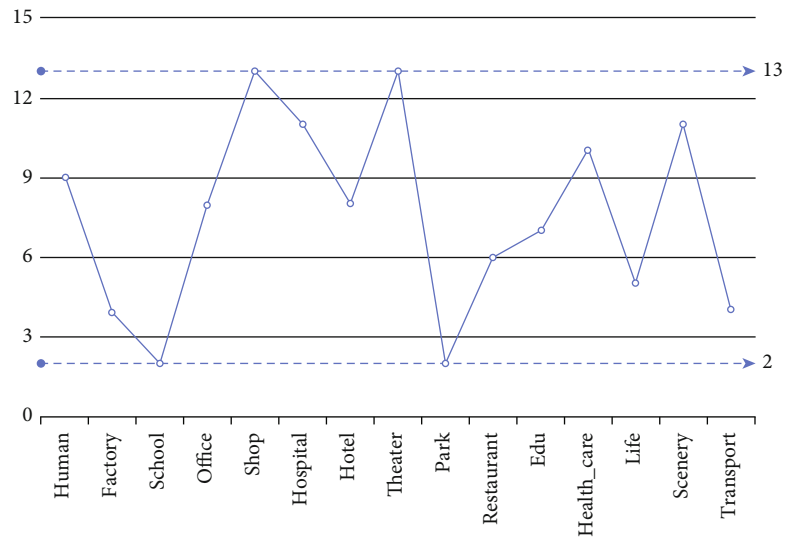

FIGURE 15: Contribution of each indicator to regional vitality in Baoshan district in 2018.

indicator to regional vitality in Songjiang district is shown in Figure 18.

Songjiang district urban vitality ranks 10 out of 16 administrative districts. Songjiang district has a notable contribution to urban vitality in 15 indicators, but the contribution of the transport indicator is relatively small because there are fewer subway lines and imperfect transportation facilities in Songjiang district, which limit urban vitality, so Songjiang district should further improve transportation facilities to enhance urban vitality.

4.14. Qingpu District Analysis. Qingpu district, a municipal district of Shanghai, China, is located in the west of Shanghai, downstream of Taihu Lake and upstream of Huangpu River. It is adjacent to Minhang district in the east, bordered by Songjiang district and Jinshan district in the south, and Jiashan county in Jiaxing city, Zhejiang province, connected to Wujiang district and Kunshan city in Suzhou city, Jiangsu province in the west, and Jiading district in the north. According to the data of the 7th census, the resident population of Qingpu district was 127,424.
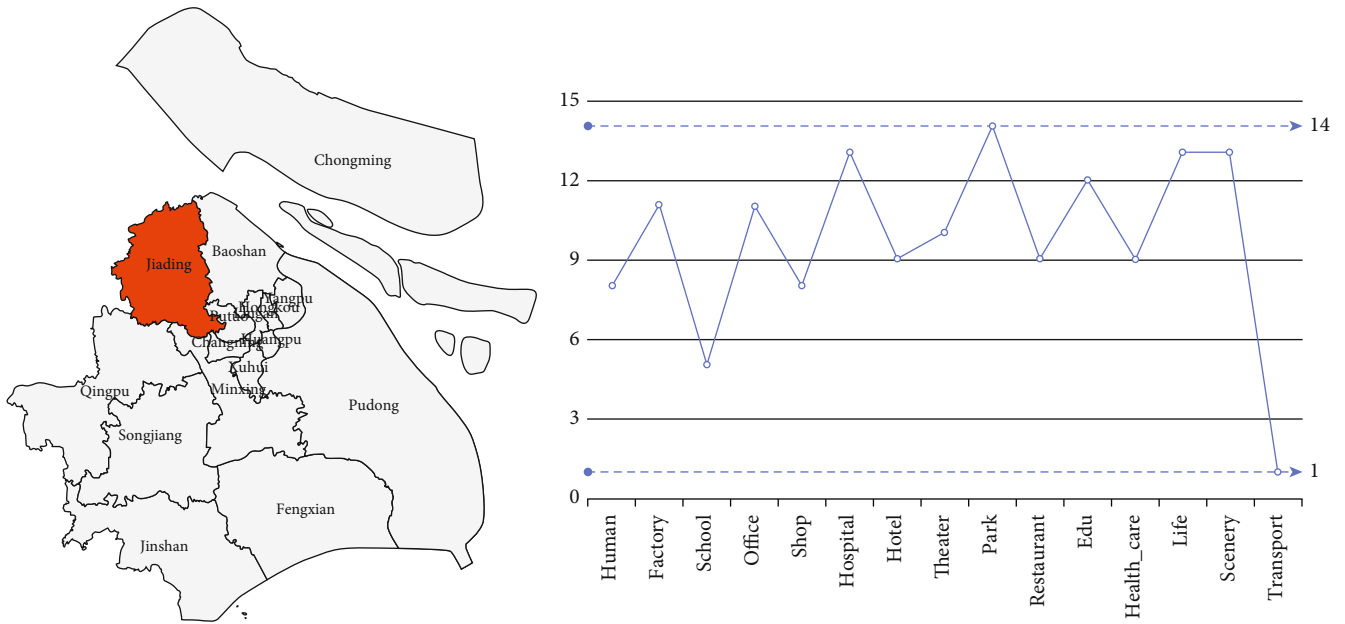
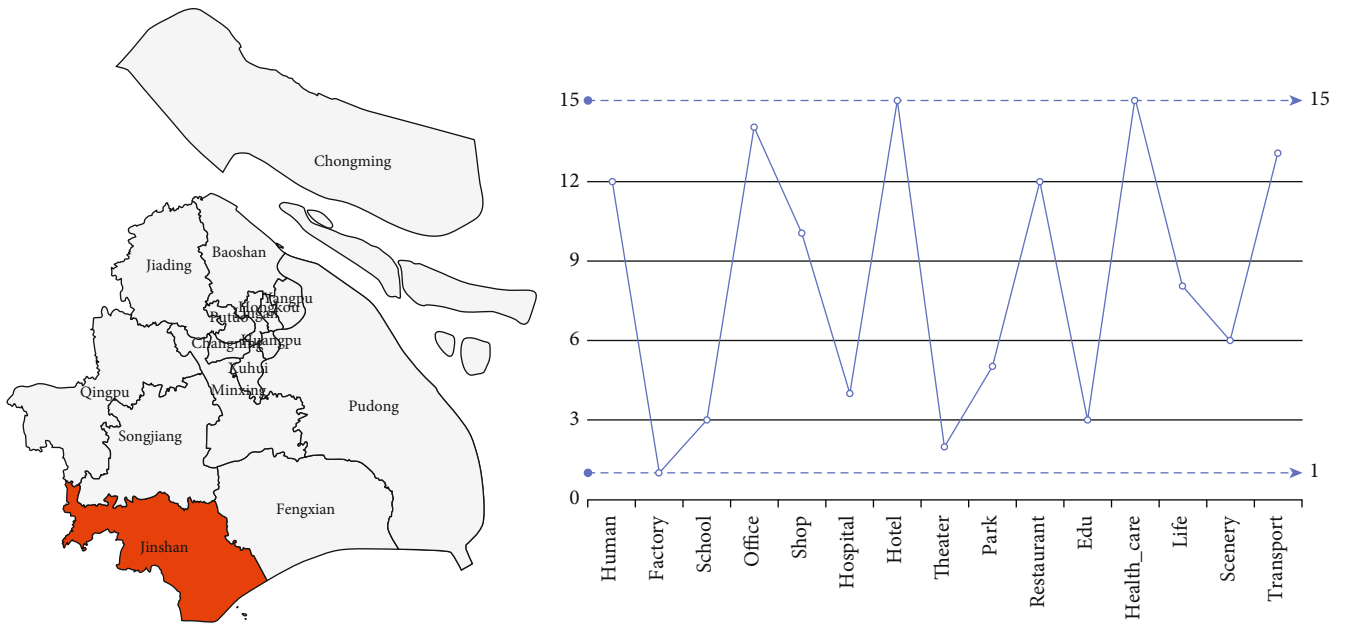FIGURE 16: Contribution of each indicator to regional vitality in Jiading district in 2018.

FIGURE 17: Contribution of each indicator to regional vitality in Jinshan district in 2018.

The contribution of the causal effect values of each indicator in Qingpu district relative to the 16 districts according to the model is as follows: human (3), factory (12), school (8), office (4), shop (5), hospital (9), hotel (4), theater (9), park (7), restaurant (6), Edu (7), health_care (4), life (5), scenery (5), and transport (9). The contribution of each indicator to regional vitality in Qingpu district is shown in Figure 19.

The urban vitality of Qingpu district ranks 13 out of 16 administrative districts. Qingpu district has a relatively aver-

age contribution of each index to the urban vitality of the district among 15 indexes, and since Qingpu district is in a more remote place in the west of Shanghai, where the population density contributes relatively more to the urban vitality of the district, the district needs to increase the improvement of the infrastructure and further attract talents to improve the urban vitality.

4.15. Fengxian District Analysis. Fengxian district, which belongs to Shanghai, is located at the southeastern end of
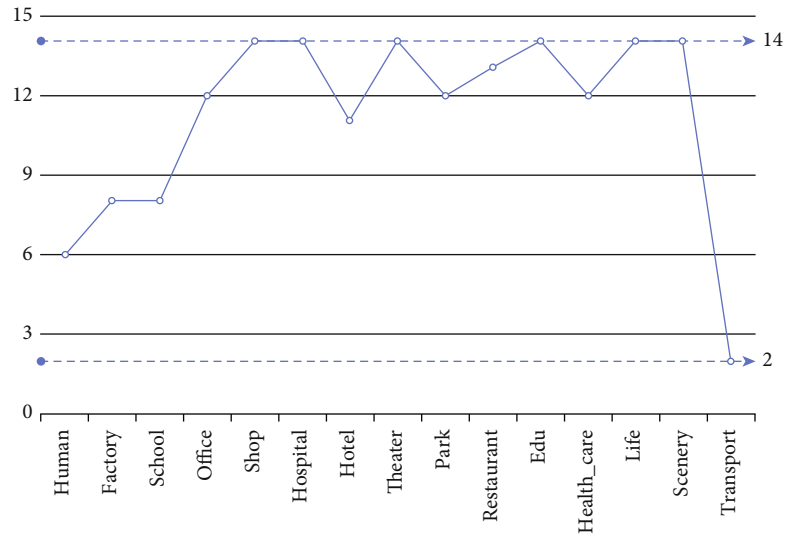
FIGURE 18: Contribution of each indicator to regional vitality in Songjiang district in 2018.
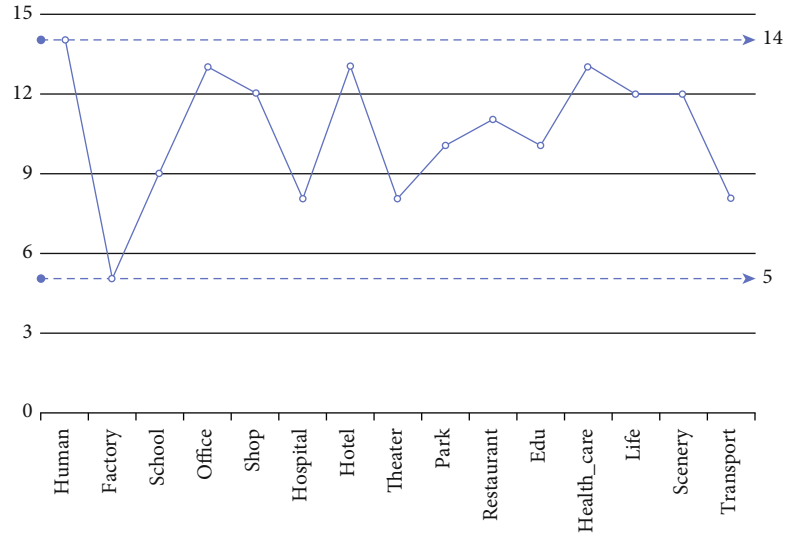

FIGURE 19: Contribution of each indicator to regional vitality in Qingpu district in 2018.

the Yangtze River Delta, in the south of Shanghai, bordering Pudong New Area to the east, Jinshan district and Songjiang district to the west, Hangzhou Bay to the south, and Minhang district to the north across the Huangpu River. According to the data of the 7th census, Qingpu district has a resident population of 1,140,872.

The contribution of the causal effect values of each indicator in Fengxian district relative to the 16 districts according to the model is as follows: human (4), factory (14), school (13), office (2), shop (6), hospital (12), hotel (3), theater (12), park (10), restaurant (7), Edu (9), health_care (3),

life (7), scenery (7), and transport (5). The contribution of each indicator to regional vitality in Fengxian district is shown in Figure 20.

Fengxian district's urban vitality ranks 14 out of 16 administrative districts. Fengxian district is located in the south of Shanghai near the sea, with good tourism resources and a good air environment, so human, office, shop, care, life, scenery, and transport contribute a lot to the urban vitality of this district, and next, Fengxian district should give full play to the advantages of tourism resources, and at the same time, further next, Fengxian district should make
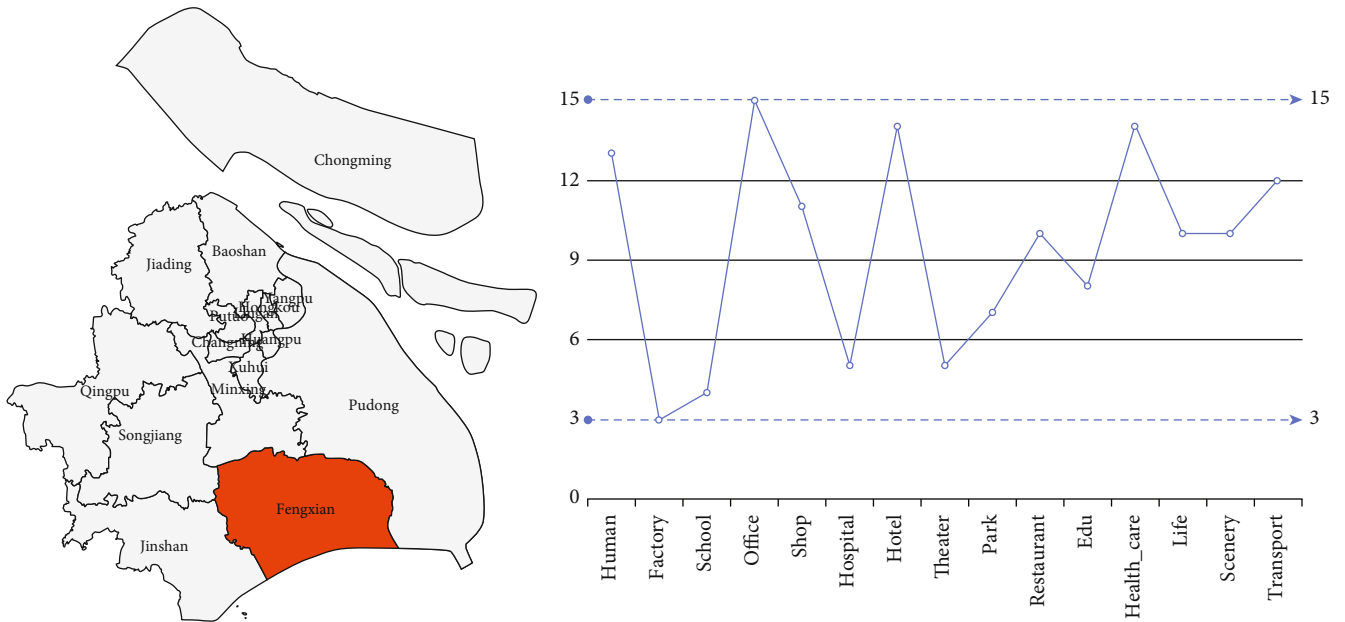
Figure 20: Contribution of each indicator to regional vitality in Fengxian district in 2018.
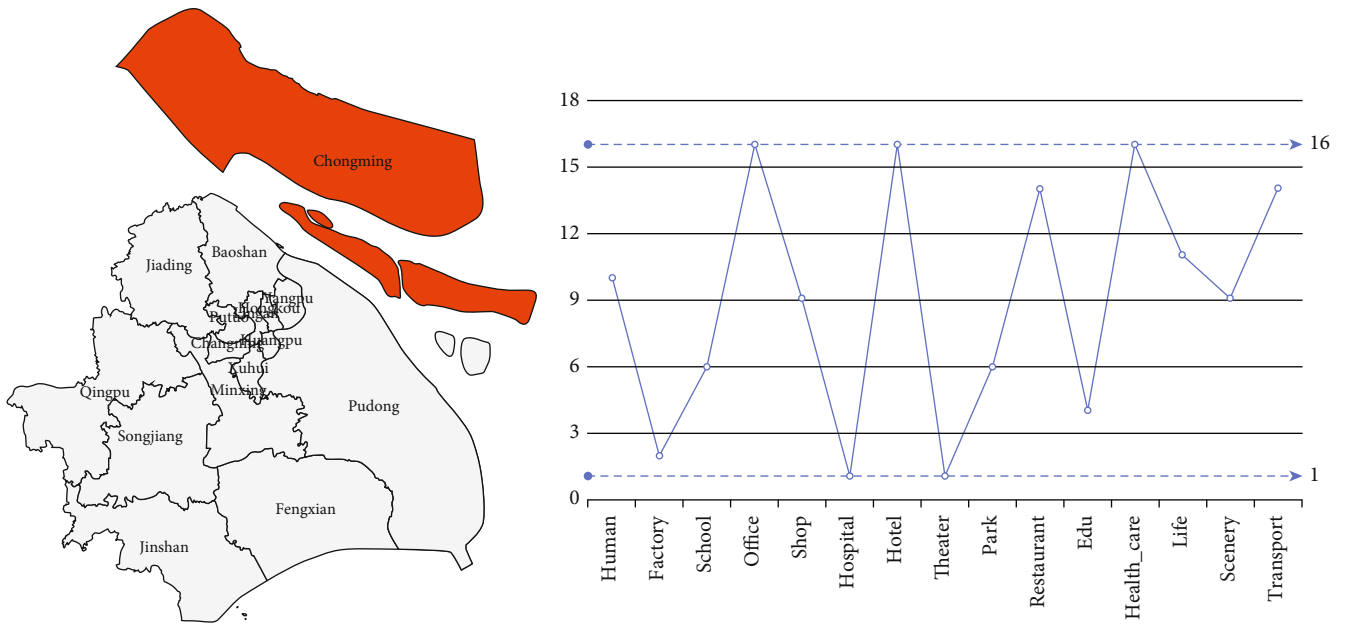

Figure 21: Contribution of each indicator to regional vitality in Chongming district in 2018.

use of the advantages of tourism resources and further improve transportation and other infrastructure to attract people and improve urban vitality.

4.16. Chongming District Analysis. Chongming district, under the Shanghai municipality, is located at the mouth of the Yangtze River. The whole district is flat and consists of Chongming, Changxing, and Heng Sha islands. According to the data of the 7th census, Chongming district has a resident population of 637,921.

The contribution of the causal effect values of each indicator in Chongming district relative to the 16 districts

according to the model is as follows: human (7), factory (15), school (11), office (1), shop (8), hospital (16), hotel (1), theater (16), park (11), restaurant (3), Edu (13), health_care (1), life (6), scenery (8), and transport (3). The contribution of each indicator to regional vitality in Chongming district is shown in Figure 21.

Chongming district's urban vitality ranks 16 out of 16 administrative districts. Chongming is the largest estuarine alluvial island in the world and the third largest island in China after Taiwan and Hainan. The climate is suitable, the environment is beautiful, and the tourism resources are very competitive. Among the 15 indicators, human, office,
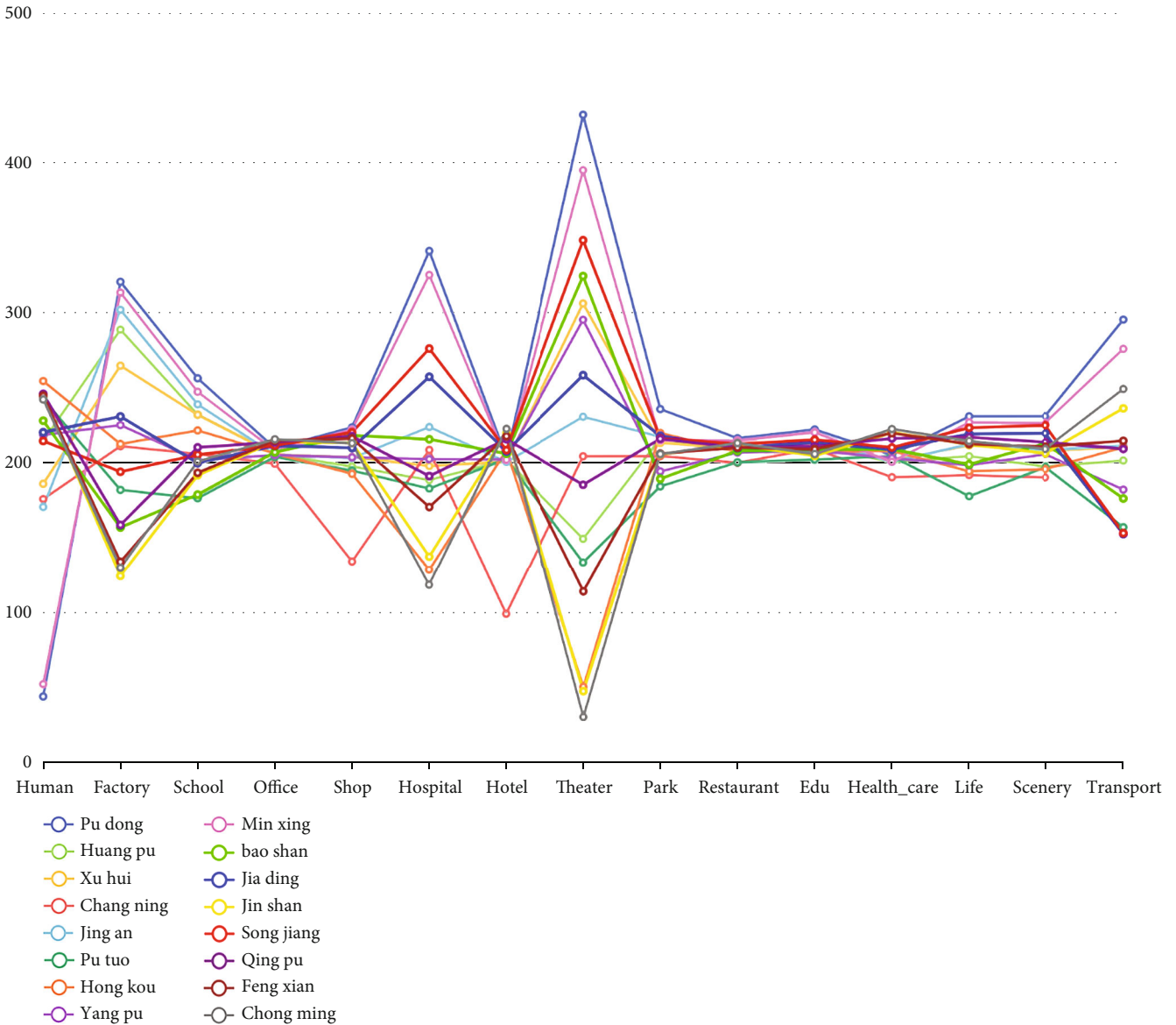
FIGURE 22: Details of Shanghai's administrative districts on all variables in 2018.

shop, hotel, restaurant, health_care, life, scenery, and transport contribute a lot to the city's vitality. Chongming district should give full play to the advantages of its tourism resources, do a good job of supporting facilities for tourism resources, and try to do more advertising, and at the same time, because it is located on the island, it can add more cruise services and find ways to increase the flow of people to enhance the vitality of the city.

In general, to improve their urban dynamics, boroughs should first focus on the top ranking of these indicator variables for themselves, as these are the indicators that contribute the most to them. Figure 22 shows the performance of each borough under each indicator. In the case of theater, for example, Pudong, Baoshan, Minhang, and Songjiang perform better, so these regions still need to focus on their strengths in the theater.

According to our model, the urban vitality indices of Pu dong, Huangpu, Minhang, and Jingan are in the top 4 of the 16 administrative regions, while the urban vitality indices of remote areas are lower, indicating that the urban vitality of Shanghai gradually decreases from the central city core to the urban periphery, while the urban vitality of several old urban areas is almost on par with that of the new ones, showing a polycentric. This indicates that Shanghai's urban vitality gradually decreases from the central core to the periphery of the city. Over the past three years, the overall index of Shanghai's urban vitality has been on an upward trend, reflecting the increase in urban vitality and Shanghai's competitiveness.

We find that Pudong, one of the centers of Shanghai, has strong economic vitality, but it is sparsely populated and has room for further growth, while Xuhui, Huangpu, Yangpu,

and Hongkou, even though they are still in the center of the city, are hampered in their development of urban vitality because they are sparsely populated and should further optimize their industrial structure and develop high-precision industries to enhance urban vitality. Changning district is mainly a residential area, so it can develop more facilities for people's convenience to Yangpu district; Hongkou, Putuo, Jing'an, Baoshan, and Jiading should improve the transportation, cultural, and recreational industries; Minhang district has strong transportation advantages and can support more commercial industries to increase vitality, and Chongming, a suburb, is far away from the city center, but it relies on tourism to make its leisure and entertainment industries not inferior to the city center. To sum up, we believe that Shanghai needs to make more efforts in developing the economy of noncenter areas, increasing investment in education, medical care, technology, and social security to attract the population, and build more urban green areas and public spaces in the future.

Shi et al. [43] concluded that Shanghai urban vitality shows a monocentric spatial pattern decreasing gradually from the center to the periphery and Shanghai urban vitality shows a significant positive spatial correlation in space and regional vitality is influenced by the surrounding areas. It is suggested to increase the vitality level of suburban areas, attract the population, and build more public spaces to meet people's activity needs.

Yue et al. [44] concluded that urban vitality in Shanghai declines from the urban core but also shows a polycentric vitality pattern; suburban large transportation infrastructure has a negative impact on urban vitality; it is suggested to strengthen the development of new subcenters and pay attention to the coordination of built environment planning and human activities.

Yue et al. [23] concluded that the urban development of Shanghai shows a "monocentric" pattern decreasing from the central city to the peripheral city; the built environment of Pudong New Area tends to be monotonous compared to the vibrant old city of Shanghai; the urban population activity tends to decrease from the center to the periphery. Shanghai's urban vitality seems to be closely related to urban planning, and it is recommended that urban planning should be carried out carefully.

Previous authors have well corroborated our view in different degrees and perspectives, proving again the validity of our model.

Different from previous studies on urban vitality, this paper proposes a new causal machine learning-based evaluation model for exploring urban vitality changes, which provides a causal analysis of urban vitality and unbiased estimation of the impact of multisource data indicators on each district as well as their relative contribution, to better optimize urban resources and improve urban vitality. The research in this paper provides a new perspective for the study of urban vitality, which can help us better understand the relationship between urban multisource data and urban vitality, and the results of the study can provide a decision basis for the government to formulate relevant policies.

5. Conclusion

In this paper, we design a framework combining causal inference and machine learning to evaluate the spatial causal effects of urban multisource data indicators on urban vitality. This paper is the first to combine causal inference with machine learning to unbiasedly estimate the contribution of each indicator to urban vitality, understand the situation of each region, and optimize urban resources for better urban vitality.

The shortcomings of this paper and future work only consider the spatial causal effects of urban indicators on urban vitality in isolation, without considering changes over time evolution. In addition, causal representation learning methods that integrate deep learning methods and causal inference have great potential in representing complex potential causal effects between urban multisource data and urban vitality, which can be explored in more aspects in the future deeper causal effects.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] S. Liu, L. Zhang, and Y. Long, "Urban vitality area identification and pattern analysis from the perspective of time and space fusion," *Sustainability*, vol. 11, no. 15, p. 4032, 2019.

[2] J. Montgomery, "Making a city: urbanity, vitality and urban design," *Journal of Urban Design*, vol. 3, no. 1, pp. 93–116, 1998.

[3] J. Gehl, *New City Spaces*, Danish Architectural Press, Copenhagen, Denmark, 2000.

[4] W. Attoe and D. Logan, *American Urban Architecture: Catalysts in the Design of Cities*, University of California Press, Berkeley, CA, USA, 1992.

[5] K. Lynch, *A Theory of Good City Form*, MIT Press, Boston, MA, USA, 1981.

[6] J. Jacobs, *The Death and Life of Great American Cities. Random House*, New York, NY, USA, 1961.

[7] K. Lynch, *Good City Form*, MIT press, 1984.

[8] C. Landry, "Urban vitality: a new source of urban competitiveness," *Archis*, vol. 12, pp. 8–13, 2000.

[9] M. Li, Z. Shen, and X. Hao, "Revealing the relationship between spatio-temporal distribution of population and urban function with social media data," *GeoJournal*, vol. 81, no. 6, pp. 919–935, 2016.

[10] X. Jin, Y. Long, W. Sun, Y. Lu, X. Yang, and J. Tang, "Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data," *Cities*, vol. 63, pp. 98–109, 2017.

[11] Y. Ye, Y. Zhuang, Z. H. Lingzhu, and A. van Nes, "Designing urban spatial vitality from morphological perspective–a study based on quantified urban morphology and activities' testing," *Urban Planning International*, vol. 1, no. 31, pp. 26–33, 2016.

[12] J. Difei, *The Theory of City Form Vitality*, Southeast University Press, Nanjing, 2007.

[13] H. Sung and S. Lee, "Residential built environment and walking activity: empirical evidence of Jane Jacobs' urban vitality," *Transportation Research Part D: Transport and Environment*, vol. 41, pp. 318–329, 2015.

[14] A. March, Y. Rijal, S. Wilkinson, and E. F. Özgür, "Measuring building adaptability and street vitality," *Planning Practice and Research*, vol. 27, no. 5, pp. 531–552, 2012.

[15] L. Gang, Y. Lan, H. Jianxiong, X. Chen, and S. Xue, "The spatial pattern and organizational relation of the pickup points based on POI data in Xi'an: focus on Cainiao stations," *Scientia Geographica Sinica*, vol. 38, no. 12, pp. 2024–2030, 2018.

[16] L. Weihua and N. Xin, "Spatial association analysis for urban service based on big data," *Scientia Geographica Sinica*, vol. 37, no. 9, pp. 1310–1317, 2017.

[17] X. U. Yingying, "Quantitative analysis of spatial vitality of urban built-up area: a case study of Shenzhen," *Landscape Research: English Edition*, vol. 12, no. 4, p. 5, 2020.

[18] M. Batty, "Empty buildings, shrinking cities and ghost towns," *Environment and Planning. B, Planning & Design*, vol. 43, no. 1, pp. 3–6, 2016.

[19] C. Deng and J. Ma, "Viewing urban decay from the sky: a multi-scale analysis of residential vacancy in a shrinking U.S. city," *Landscape and Urban Planning*, vol. 141, pp. 88–99, 2015.

[20] S. Hamidi and R. Ewing, "A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States," *Landscape and Urban Planning*, vol. 128, pp. 72–82, 2014.

[21] F. Wang, Q. Tang, and L. Wang, "Post-Katrina population loss and uneven recovery in New Orleans, 2000-2010," *Geographical Review*, vol. 104, no. 3, pp. 310–327, 2014.

[22] L. Lanfen, "The good life: criticism and construction of urban meaning," *Social Sciences in China*, vol. 31, no. 4, pp. 133–146, 2010.

[23] W. Yue, Y. Chen, Q. Zhang, and Y. Liu, "Spatial explicit assessment of urban vitality using multi-source data: a case of Shanghai," *China. Sustainability*, vol. 11, no. 3, p. 638, 2019.

[24] J. Yang, J. Cao, and Y. Zhou, "Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen," *Transportation Research Part A Policy and Practice*, vol. 144, pp. 74–88, 2021.

[25] S. Lu, C. Shi, and X. Yang, "Impacts of built environment on urban vitality: regression analyses of Beijing and Chengdu, China," *International Journal of Environmental Research and Public Health*, vol. 16, no. 23, p. 4592, 2019.

[26] C. Xia, G. O. Yeh, and A. Zhang, "Analyzing spatial relationships between urban land use intensity and urban vitality at street block level: a case study of five Chinese megacities," *Landscape and Urban Planning*, vol. 193, p. 103669, 2020.

[27] S. Liu, H. Liu, and Y. Long, "Exploring the quantitative assessment of urban economic vitality in the Yangtze River Delta of China based on PCA," *IOP Conference Series Earth and Environmental Science*, vol. 836, no. 1, article 012011, 2021.

[28] M. N. Lopes and A. S. Camanho, "Public green space use and consequences on urban vitality: an assessment of European cities," *A Concise Treatise on the Progress of Medicine, since the Year 1573*, vol. 113, no. 3, pp. 751–767, 2013.

[29] F. Lan, X. Gong, H. Da, and H. Wen, "How do population inflow and social infrastructure affect urban vitality? Evidence from 35 large- and medium-sized cities in China," *Cities*, vol. 100, p. 102454, 2020.

[30] H. Yue and X. Zhu, "Exploring the relationship between urban vitality and street centrality based on social network review data in Wuhan, China," *Sustainability*, vol. 11, no. 16, p. 4356, 2019.

[31] P. Sulis, E. Manley, C. Zhong, and M. Batty, "Using mobility data as proxy for measuring urban vitality," *Journal of Spatial Information Science*, vol. 16, no. 16, pp. 137–162, 2018.

[32] Y. Yue, Y. Zhuang, A. G. O. Yeh, J. Xie, C. Ma, and Q. Q. Li, "Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy," *International Journal of Geographical Information Systems*, vol. 31, no. 4, pp. 658–675, 2017.

[33] H. R. Varian, "Big data: new tricks for econometrics," *Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–28, 2014.

[34] L. Chen and H. J. Wu, "Research status and potential problems of differences-in-differences method in China[J]," *The Journal of Quantitative and Technical Economics*, vol. 32, no. 7, pp. 133–148, 2015.

[35] J. J. Heckman and J. L. Garcia, "Social policy: targeting programmes effectively," *Nature Human Behaviour*, vol. 1, no. 1, pp. 1-2, 2017.

[36] C. Louizos, U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling, "Causal effect inference with deep latent-variable models," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[37] A. Jesson, S. Mindermann, U. Shalit, and Y. Gal, "Identifying causal-effect inference failure with U ncertainty-aware models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11637–11649, 2020.

[38] S. Wright, *Correlation and Causation*1921, https://http://naldc.nal.usda.gov/catalog/IND43966364.

[39] J. Splawa-Neyman, D. M. Dabrowska, and T. P. Speed, "On the application of probability theory to agricultural experiments," in *Essay on Principles. Section 9*, pp. 465–472, Statistical Science, 1990.

[40] M. Bertanha and G. W. Imbens, "External validity in fuzzy regression discontinuity designs," *Journal of Business and Economic Statistics*, vol. 38, no. 3, pp. 593–612, 2020.

[41] V. Chernozhukov, D. Chetverikov, M. Demirer et al., *Double Machine Learning for Treatment and Causal Parameters*, 2016, https://arxiv.53yu.com/abs/1608.00060.

[42] S. Wager and S. Athey, "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, vol. 113, no. 523, pp. 1228–1242, 2018.

[43] J. Shi, W. Miao, H. Si, and T. Liu, "Urban vitality evaluation and spatial correlation research: a case study from Shanghai, China," *Landscape*, vol. 10, no. 11, p. 1195, 2021.

[44] W. Yue, Y. Chen, P. T. M. Thy, P. Fan, Y. Liu, and W. Zhang, "Identifying urban vitality in metropolitan areas of developing countries from a comparative perspective: Ho Chi Minh city versus Shanghai," *Sustainable Cities and Society*, vol. 65, p. 102609, 2021.