A Novel Framework for Exploring the Spatial Characteristics of Leisure Tourism Using Multisource Data: A Case Study of Qingdao, China

Yiqun Shang, Caiyun Wen, Yangchun Bai, and Dongyang Hou

Abstract—Spatial characteristics of leisure tourism resources are essential for human life, the urban economy, and tourism planning. This article presents a novel framework to explore these characteristics based on multisource data, such as points of interest, OpenStreetMap roads, Sentinel-2 multispectral instrument images, and other data, and proposes a new tourism area identification method by integrating the attractiveness of attractions with term frequency–inverse document frequency. The roles of the influencing factors were measured by using the geodetector and related statistical analyses. The results showed that the resources were centered on Jiaozhou Bay, and their axial direction was “northeast to southwest.” The distribution of the overall resources was characterized by “one cluster with multiple core points,” and different types of resources had different aggregation distributions. The recreational recreation and cultural leisure zones were more likely to be distributed in and near the center of each district, and their numbers were high, while the shopping leisure and natural recreation (NR) zones were the opposite. The distribution of each type of resource was the result of a combination of factors working together, except for NR resources, which were mainly influenced by natural factors, while others were mainly affected by socioeconomic factors. The study findings are instructive for tourism planning.

Index Terms—Geodetector, improved TF–IDF method, leisure tourism, multisource data, spatial characteristics.

I. INTRODUCTION

With the progress of urbanization and the increase in people’s living standards, leisure travel has become a more common form of travel behavior; thus, leisure tourism has gradually become an essential recreational activity in people’s lives [1], [2]. Compared with the traditional tourism, leisure tourism pays more attention to recreation and relaxation, which generally refers to activities that people participate in to rid themselves of stress and to relax their minds during leisure time [3]. Leisure tourism resources can attract people to carry out leisure activities and can be exploited by the industry to produce benefits [4]. Leisure tourism resources are regarded as a special form of tourism resources; so, their research methods are similar to that of tourism resources. The majority of previous studies on spatial characteristics of leisure tourism resources tended to use statistical data, including yearbooks and interview data on tourists, to study the spatial characteristics and attractive factors of the resources. For instance, Zakariya et al. [5] explored the attractiveness of tourists in urban square leisure destinations by using survey data. Rainer [6] used interview data to explore leisure tourism in Salta’s wine regions and to analyze industrial development. Statistical yearbook data can provide a reflection of tourism resources if used to calculate the statistics therein; however, such data are often too macroscopic and not detailed. The interview data are more detailed; however, the data tend to be poorly representative at larger spatial scales. The process of collecting this data is also time-consuming and labor-intensive.

With the big data era, a large number of new geographical data sources have emerged [7]–[10], such as geotagged photos [11]–[13], cell phone signaling data [14]–[16], and smart card data [17]–[19]. They provide a new direction for the study of urban tourism space. For instance, García-Palomares et al. [20] used geotagged photographs on Panoramio to analyze several cities’ spatial distribution patterns. Abbasi et al. [21] used Twitter data from tourists and analyzed their travel patterns for better urban planning. Mou et al. [22] analyzed the spatial patterns of tourist flows based on tourists’ digital footprint data. Furthermore, Wang et al. [23] used cell phone data to identify tourists’ travel patterns, explore influencing factors, and predict travel behavior. Although these studies provide new technical methods for analyzing spatial characteristics or patterns of tourism, there are two major drawbacks in the study of tourism using these data. First, they are generated from specific human activities (e.g., taking photos, tweeting), so they are based on the perspective of a portion of travelers and cannot represent mass tourism; second, the complexity of data acquisition and processing makes the data collection and analysis particularly challenging.
less usable in practical work; for example, smart card data are internal to the company and difficult to access.

As a relatively easy-to-access emerging open-source data, points of interest (POIs) [24]–[26] can comprehensively, accurately, and meticulously reflect the spatial distribution of various geographic entities in the city and can be obtained relatively easily and objectively from map providers. Thus, they provide a new data source for the spatial characteristics of urban tourism resources. Xu et al. [27] used leisure tourism POIs in the Nanjing city of China to explore the spatial characteristics of the resources and their influencing factors, but the selected factors were more general and were not well defined. Li et al. [28] utilized the geodetector to quantify the strength of the factors. However, this study lacked the exploration of natural factors, and its spatial scale was relatively homogeneous. Wang and Hou [29] evaluated the roles of various factors in rural touristization, such as the distance to the main road; this factor could be a representative of traffic conditions, but it was oversimplified and cannot better express the complex structure of urban road space. With the development of space syntax, the content of the accessibility of road conditions has been enriched, precisely providing a new perspective for expressing road conditions.

Identifying leisure tourism attributes of different zones of a city is important for urban planning and tourism guidance. However, previous studies using POI data have mainly focused on urban functional zones; fewer studies have used POI data for tourism zone identification. Some natural language processing methods are used to identify functional zones, such as the latent Dirichlet allocation (LDA) model, the term frequency–inverse document frequency (TF–IDF) method, and the topic modeling method. For instance, Du et al. [30] used the LDA and SVM methods to classify segmented urban functional zones. Zhang et al. [31] used the TF–IDF method to identify urban functional zones by utilizing station-based public bicycle rental records and POI data. Du et al. [32] utilized the topic modeling method and combined it with multimodal data to improve the accuracy of the identified functional zones; given the similarity between identifying functional zones and leisure tourism zones, it is also reasonable to introduce these methods to identify leisure tourism zones innovatively.

Understanding the spatial distribution of leisure tourism resources is also important; specifically, the application of remote sensing may be an acceptable method because of its significant approach to comprehending complex interactions between social and natural systems on a large scale, and to some extent, the distribution of the resources is relevant to the natural resources. For example, people tend to visit sites that are close to forests or an ocean, but there have been few studies quantitatively exploring the relationship between these factors (e.g., land cover and landscape patterns) and tourism resources.

Therefore, we develop a novel research framework that utilizes multisource data to explore leisure tourism resources’ spatial characteristics and propose an improved TF–IDF method to identify the leisure tourism zones. We also explore the roles of various influencing factors, including socioeconomic and natural factors. In summary, our contributions in this article are as follows.

1) We construct a new framework for exploring the spatial characteristics of urban leisure tourism resources using multisource data, including mixed-scale determination, leisure tourism area identification, and influence factor detection. Compared with the traditional framework, this framework can effectively utilize the advantages of each data source, prompting it to reveal the spatial characteristics and influence factors of leisure tourism areas. In addition, we also design a hybrid scale determination method based on grids and blocks in this framework, which can compensate for the shortcomings of a single scale.

2) We introduce Ctrip’s attractions into leisure tourism area identification and propose a new tourism area identification method by integrating attractiveness of attractions with TF–IDF method. Compared with the traditional method, the method can improve the role of popular attractions.

II. CASE-STUDY AREA: QINGDAO, CHINA

The case-study area included the whole Qingdao city (see Fig. 1), with a total area of 11,282 km². Qingdao has a long history; as a result, there are many famous historical properties, such as the Eight Great Passes and St. Michael’s Cathedral. It is also a thriving city: it is the economic center of Shandong province and one of the first open coastal cities in China. At the end of 2018, its permanent population was 9.39 million, and the urbanization rate exceeded 70%.

Excellent natural conditions and strong socioeconomic development have enabled the rapid growth of leisure tourism industry in this city and have also led Qingdao to develop its own unique leisure tourism culture (e.g., annual beer festival and marine cultural events) and to generate abundant relevant tourism resources [33]. These factors contributed to Qingdao being considered an ideal area to test the applicability of multisource data for spatial analysis of urban leisure tourism resources.
III. METHODOLOGY

A. Research Framework for Analysis of Tourism Spatial Pattern and Influencing Factors

To enhance the analysis of the tourism spatial pattern and influencing factors, we propose a novel research framework, which is shown in Fig. 2.

This research framework is based on multisource basic data and its corresponding preprocessing data. It includes POI data, OpenStreetMap (OSM) road data, and vector boundary data as the main data, but the other data (e.g., Sentinel-2 multispectral instrument (MSI) images, statistical data, and DEM) are also included to produce the new data to represent the influencing factors. The analysis of the tourism spatial pattern and influencing factors is at the core of this framework. To analyze the spatial characteristics more systematically, we divided the research contents into four sections:

1) data preprocessing;
2) point pattern analysis: analysis of the resources’ distribution pattern and spatial agglomeration by using spatial point pattern analysis based on POI data;
3) identification of leisure tourism zones by utilizing the improved TF–IDF method to map the leisure tourism zones on both block and grid scales (this process innovatively combines the online information of Ctrip’s popular attractions to upgrade the importance of some best-known attractions);
4) factors analysis: to explore the relationship between the factors (such as socioeconomic and natural factors) and the resources; we use GIS, space syntax, and landscape pattern analysis to generate new data to represent these factors based on the multisource basic data and then apply the geodetector to quantify the strength of the factors.

B. Data Sources

The basic data used in this study included:

1) Amap POI data for July 2018, from the Amap service;
2) OSM road data acquired in 2018;
3) boundary data and waterline data from the Chinese basic geographic database v2017;
4) DEM image and Sentinel-2 MSI images between June 2018 and September 2018, which were obtained from the google earth engine (GEE) platform;
5) statistical data from the Statistical Yearbook of Qingdao City in 2019;
6) online information of Ctrip’s popular attractions and their corresponding rankings, from Trip.com.

C. Data Preprocessing

1) Amap POIs: As a type of point geospatial big data reflecting real geographical entities, POIs can effectively extract geographic information and explore the spatial information relationship contained therein [34]. Their positioning accuracy is relatively high, and the categories are multilevel, which gives them great application value for tourism geography. In this study, the POIs were fetched from the Amap service in July 2018 (see Table I). Amap is one of the most widely used map service providers in China, and the coordinate system of Amap is GCJ-02. After applying the correction algorithm to correct the data to the WGS-84 coordinate system and project the data to the World Mercator system, the back-coded verification of the corrected data revealed that the maximum offset distance was within 10 m; so, its accuracy could meet the research needs.

There is no mature classification standard for leisure tourism resources; therefore, referring to the existing research’s classification system and combining it with the actual situation of the resources, the initial dataset was obtained. Its class was similar or consistent with the classification system. We then used the following process on the dataset.

1) Remove duplicate data: As the initial dataset contained many duplicate POIs, we had to export the POIs with high similarity in attributes, such as name and category, and then merge them after the manual screening.
2) Eliminate the data unrelated to leisure tourism: Since there are POI data that are literally related to attractions but semantically irrelevant (e.g., the parking lot in the Badaguang Scenic Area and homestay in the Laoshan Scenic Area of Qingdao City), this portion of data also needs to be excluded.
3) Check out small-volume sites, such as local snack bars, then remove them from the dataset.

After preprocessing the POI data, we randomly selected and examined 1000 of the POIs, focusing on their name, attributes, and location. The check result showed an accuracy of 100%, indicating that the used POI data are reliable and accurate. Ultimately, 25 450 POIs were obtained (see Table II), which were divided into 4 main classes and 15 subclasses. They were equivalent to the corresponding type’s leisure tourism resources.

2) OSM Roads Data: OSM is a voluntary project that provides users with free and easily accessible map data and is known as the most prevailing volunteered geographic information [35]. In addition to the fundamental information, the OSM road data contain additional attributes (such as road type), which are helpful for preprocessing and establishing the road network. The OSM roads’ raw data used in this study were acquired in 2018. We then clipped the part of Qingdao on the ArcGIS 10.2 software platform. The processing steps are as follows.

1) Trim the individual roads with a threshold of 200 m to remove hanging lines and then extend the individual road segments at both ends for 20 m to connect nearby but topologically separated lines.

2) Grade the different roads according to attributes, and then treat the roads with the attributes: motor, trunk, primary, secondary, and tertiary roads as major roads and other roads as secondary roads.

3) As shown in Fig. 3(a) and (b), binarize vector roads to raster and operate the closing operation in morphology to extract the road skeleton by merging spatially adjacent roads (such as bidirectional roads), where the major road’s parameter is set to 30 m, the secondary road’s parameter is set to 10 m, and then the road skeleton and boundary data are used to generate blocks, as shown in Fig. 3(c).

Spatial design network analysis (sDNA) is a spatial network analysis software that extends space syntax tools and can measure the betweenness and closeness of roads [36]. Network quantity penalized by distance in radius angular (NQPDA) is one of the outputs that represent the road’s closeness; we calculated it to represent the road proximity by using sDNA and exploited Kernel density estimation (KDE) to generate the NQPDA image on ArcGIS, as shown in Fig. 4.

3) Boundary Data: The boundary data in this study were from the Chinese basic geographic database v2017. We extracted the initial boundaries for each district of Qingdao, but they were too cursory and insufficiently detailed in some areas (such as Zhanqiao Park and Qingdao International Sailing Center) to meet study needs. Therefore, they were manually edited to reference the satellite image and resulted in the final data that

### Table II

| Main Class      | Subclass                      | Quantity | Proportion |
|-----------------|-------------------------------|----------|------------|
| CTL             |                               | 10,211   | 34.15%     |
| HC              |                               | 5389     | 18.02%     |
| SL              |                               | 4648     | 15.54%     |
| RR              | Sports and Fitness (SF)       | 2206     | 7.38%      |
|                 | Entertainment Leisure (EL)    | 1705     | 5.70%      |
|                 | Public Recreation (PR)        | 419      | 1.40%      |
| CL              | Cultural Venues (CV)          | 1545     | 5.17%      |
|                 | Humanities Synthesis (HIS)    | 895      | 2.99%      |
|                 | Historical Heritage (HHH)     | 504      | 1.69%      |
| Natural Recreation (NR) | Land Scenery (LS) | 561      | 1.88%      |
|                 | Water Recreation (WR)         | 474      | 1.59%      |
|                 | Natural Synthesis (NS)        | 390      | 1.30%      |
|                 | Biological Leisure (BL)       | 125      | 0.42%      |
| Special Leisure (SL) | Industrial Leisure (IL) | 409      | 1.37%      |
|                 | Other Special Leisure (OS)    | 423      | 1.41%      |
met the study needs. Based on the boundary data, the 3 km × 3 km grids were generated.

4) Sentinel-2 MSI Images: Sentinel-2 satellite mission systematically collects multispectral land surface imagery by two satellites (i.e., Sentinel-2A and Sentinel-2B) with a revisit cycle of 5 days at 10 m, 20 m, and 60 m spatial resolutions. These satellites have MSI that collects 13 bands: coastal blue, blue, green, red, three vegetation red-edge bands, near-infrared (NIR), narrow NIR, water vapor, shortwave infrared (SWIR) cirrus, and two SWIR bands.

To map the land cover of the city, we selected Sentinel-2 MSI Level-1C Top-Of-Atmosphere reflectance images between June 2018 and September 2018 based on the GEE platform. Five frequently used bands, blue, green, red, one NIR, and one SWIR, were selected for classification. Those bands were chosen because of the spatial resolution of 10 m (blue, green, red, and one NIR) and 20 m (one SWIR), where the SWIR band was resampled to 10 m resolution.

To obtain the low cloud image of all the selected images, we obtained the least cloudy image by applying the algorithm provided by GEE. Land cover types were divided into cropland, forest, grassland, water, built-up area, and barren land. Then, based on the random forest classifier built into GEE and sample points selected in advance from google earth, the land cover was mapped, and the overall accuracy (OA) reached 90.23%, as shown in Fig. 5.

The land cover image had to be further processed by Fragstats 4.0 using the moving window analysis to calculate landscape metrics, including Simpson’s diversity index (SIDI), Shannon’s diversity index (SHDI), mean perimeter area ratio (PARA_MN), perimeter area fractal dimension (PFRAC), number of patches (NP), largest patch index (LPI), edge density (ED), and Contagion index (CONTAG).

5) Statistical Data: The statistical data were from the Statistical Yearbook of Qingdao in 2019, which recorded statistics for 2018. We mainly used each district’s gross domestic product (GDP) data for the primary, secondary, and tertiary industries, as well as resident population data.

To simulate the gridded GDP dataset and the gridded population dataset to facilitate discussions of further research, we first created 1 × 1 km grids based on the boundary data. For the gridded GDP dataset, the land cover types corresponding to the primary industry of GDP were cropland, forest, grassland, and water, and the type corresponding to the secondary and tertiary industries of GDP was built-up area. We used (1) to simulate the dataset, as shown in Fig. 6(a), and we used (2) to simulate the gridded population dataset, as shown in Fig. 6(b).

\[
GDP_k = \sum_{i=1}^{3} \left( \frac{GDP_{ij}}{\text{Land}_{ij}} \cdot \text{Land}_{ik} \right)
\]

where \(i\), \(j\), and \(k\) denote the GDP’s industry ranked \(i\), the serial number of each district, and the serial number of the grid, respectively; \(\text{Land}_{ij}\) denotes the area of the corresponding land cover types for GDP’s industry, ranked \(i\) in district \(j\); and \(\text{Land}_{ik}\) denotes the area of the corresponding land cover types for GDP’s industry, ranked \(i\) of grid \(k\).

\[
POP_k = \frac{\text{POP}_i \cdot \text{Builtuparea}_k}{\text{Builtuparea}_i}
\]

where \(i\) and \(k\) denote the serial number of each district and the serial number of the grid, respectively; \(\text{Builtuparea}_i\) denotes the area of the built-up area in district \(i\); and \(\text{Builtuparea}_k\) denotes the area of the built-up area in grid \(k\).

6) Ctrip’s Popular Attractions: The online information of Ctrip’s popular attractions and their corresponding rankings were from Trip.com, which is the leading travel service website in China. To obtain the information on Qingdao, we wrote a Python program to crawl the information and then compared these spots to the leisure tourism POIs’ names, and a total of 968 spots were successfully matched with the POIs.

D. Point Pattern Analysis

1) Standard Deviational Ellipse (SDE): SDE is a classical method for analyzing spatial characteristics [37], [38]. It allows simultaneous analysis of the direction and distribution trends of the point data. SDE has four parameters to analyze point data: the
ellipse center, major axis, minor axis, and azimuth angle. The center indicates the central position of the entire data and the central trend of the data. The major axis of the ellipse indicates the direction of data distribution, the minor axis indicates the range of data distribution, and their proportional relation denotes the degree of the ellipse’s flattening; the larger the degree of flattening, the more pronounced the directionality of the data. The azimuth angle reflects the predominant trend direction of the data distribution.

In this study, the standard deviations of the major and minor axes were \( \sigma_x \) and \( \sigma_y \), respectively, which are calculated as follows:

\[
\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (\bar{x}_i \cos \theta - \bar{y}_i \sin \theta)^2}{n}}
\]  

(3)

\[
\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (\bar{x}_i \sin \theta - \bar{y}_i \cos \theta)^2}{n}}
\]  

(4)

where \( \bar{x}_i \) and \( \bar{y}_i \) denote the difference between the mean center coordinates, \( x \) and \( y \), and the point coordinates, \( x_i \) and \( y_i \), respectively.

The azimuth angle \( \theta \) is the angle of clockwise rotation from due north to the major axis of the ellipse, and it is calculated as follows:

\[
\tan \theta = \frac{\sqrt{\sum_{i=1}^{n} \bar{x}_i^2 - \sum_{i=1}^{n} \bar{y}_i^2 + \left( \sum_{i=1}^{n} \bar{x}_i^2 - \bar{y}_i^2 \right)^2 + 4 \left( \sum_{i=1}^{n} \bar{x}_i^2 \bar{y}_i^2 \right)^2}}{\sum_{i=1}^{n} \bar{x}_i \bar{y}_i}.
\]  

(5)

2) Kernel Density Estimation: KDE is a function used to estimate the unknown density in probability theory. It uses spatial density analysis of events to represent spatial point patterns [39]. Compared with the quadrat count methods, KDE is more suitable for visualizing the spatial distribution pattern. It reflects the degree of aggregation of the POIs. As shown in (6), \( x_1, x_2, \ldots, x_n \) is an independent sample of the same distribution extracted from a population with the distribution density function \( f \) given by

\[
f_h(x) = \frac{1}{n} \cdot \frac{1}{h} \sum_{i=1}^{n} k \left( \frac{x - x_i}{h} \right)
\]  

(6)

where \( k(*) \) is the sum function; \( h \) is the bandwidth, and \( h > 0 \); \( x - x_i \) is the distance from the evaluation point \( x \) to the sample \( x_i \); the larger the \( k(x) \), the denser the point distribution.

E. Improved TF–IDF Method for Identifying Leisure Tourism Zones

As a typical spatial scale, blocks in a city usually imply rich semantic messages related to people’s impressions [40], and they can be well depicted and defined by the leisure tourism sites of blocks. For example, people familiar with Qingdao would think of the Eight Great Passes block as offering many historical and cultural attractions. To further enrich the identification of spatial scales, we also introduced the grid scale into the identification to compare with the block scale.

To simplify the description of the identification method, each block or grid was treated as a zone. Therefore, this study aimed to understand the semantic messages of zones by their leisure tourism POI types. A commonly used method is to calculate the proportions of POI types in each zone; however, some common POI types, such as catering leisure (CTL) POIs and shopping leisure (SL) POIs, would take up high proportions in many zones, which would fail to reflect the uniqueness of each.

To emphasize the famous and unique POIs and minimize the role of general POIs, we propose an improved TF–IDF method: dividing the zones into two parts, one with a spatial area containing Ctrip’s popular attractions and the other part without. For the zones with popular attractions, each zone’s type is the attraction’s type with a ranking that is the highest among the included attractions in this zone; for the zones without the attractions, each area is treated as a document, and the weight of the POI type \( j \) in the zone \( Z_i \) can be calculated as follows:

\[
\text{TFIDF}_{ij} = \frac{\text{TF}_{ij} \cdot \log_2 \frac{\text{DF}_{j}}{\text{DF}_i}}
\]  

(7)

where \( \text{TF}_{ij} \) is the proportion of POIs with type \( j \) among all POIs in \( Z_i \); \( \text{DF}_j \) is the number of zones in Qingdao, and \( \text{DF}_j \) is the number of zones that contain type \( j \). The POI type with the highest TFIDF value will be selected as the type of the zone.

F. Factor Analysis Method

The geodetector is a statistical tool developed by Wang et al. [41]. It is a set of methods used to reveal stratified heterogeneity and detect the driving factors behind it. In this study, we mainly selected two factor detectors to reveal the influencing mechanism of the leisure tourism resources’ distribution: First, the factor detector measured the degree of factor’s explanatory power to the variable \( Y \); second, the interaction detector revealed whether factors \( X_1 \) and \( X_2 \) had an interactive influence on a variable \( Y \). Both detectors quantify the explanatory power of \( Y \) through the \( q \)-value, as follows:

\[
q = 1 - \frac{1}{N \sigma^2} \sum_{k=1}^{L} N_k \sigma^2_k
\]  

(8)

where \( q \) represents the explanatory power of the influencing factor \( X \) to the resources and lies between 0 and 1; \( N \) and \( \sigma^2 \) denote the number of units and the variance of the whole area, respectively; \( L \) is the postdiscrete classification number of the factor; and \( N_k \) and \( \sigma^2_k \) denote the number of units and the variance of the resources in class \( h \), respectively [41]. The datasets of the factors were quantitatively characterized in 3 km \( \times \) 3 km grids that were previously generated.

Factors leading to the distribution of leisure tourism resources are complex and diverse. Following the literature review, we selected two main types of factors: natural and socioeconomic. The natural factors included landscape metrics, the biological abundance index (BAI), the vegetation coverage index (VCI), elevation (ELE), and distance to water (DIST_WATER). The socioeconomic factors include the NQPDA, GDP, population (POP), and the proportion of built-up area (BU_P), as shown in Table III. Among them, the landscape metrics, ELE, NQPDA,
GDP, and POP are averaged onto the $3 \times 3$ km grids based on their original data, and the remaining are generated directly based on the grids.

Specifically, for the factor DIST_WATER, we exploited the near analysis to calculate the distance between each grid and the waterline on ArcGIS; the BAI, VCI, and BU_P are calculated by (9)–(11), respectively; and the remaining factors were used to calculate the average value for each unit from the grids.

$$BAI = A_1 \cdot \frac{0.35 \cdot \text{Forest} + 0.21 \cdot \text{Grass} + 0.28 \cdot \text{Water} + 0.11 \cdot \text{Crop} + 0.04 \cdot \text{Builtup} + 0.01 \cdot \text{Barren}}{\text{grid area}}$$  
$$VCI = A_2 \cdot \frac{0.38 \cdot \text{Forest} + 0.34 \cdot \text{Grass} + 0.19 \cdot \text{Crop} + 0.07 \cdot \text{Builtup} + 0.02 \cdot \text{Barren}}{\text{grid area}}$$  
$$BU_P = \frac{\text{Builtup}}{\text{grid area}}$$  

where $A_1$ and $A_2$ are the normalization factors; Forest, Grass, Water, Crop, Builtup, and Barren denote the corresponding land type’s area in the grid, the weights of the land types in (9) and (10) are referred to [42] and [43], they are generally fixed values, and gridarea is the area of the grid.

The treatment processes were as follows:

1) extract the mean Kernel density value of different types’ leisure tourism POIs (dependent variable $Y$) and the values of selected factors in the 2153 grids;
2) discretize all the numerical factors by the quantile method in Python;
3) input the 2153 samples to the geodetector, including values of the dependent variable and 16 kinds of stratified factors;
4) click the run button to obtain the results;
5) plot the box plot in Python and analyze the distribution of numerical variables in different types of grids.

### IV. RESULTS

#### A. Distribution Pattern of Leisure Tourism Resources Based on the SDE Method

The spatial distribution patterns of the overall and different types of leisure tourism resources were recognized by this method, and the results obtained are shown in Fig. 7. The major axis denotes the direction, and the minor axis denotes the range of the point data distribution. In general, the spatial pattern of leisure tourism resources in Qingdao was characterized by “dense middle and sparse edges,” and the resources were centered on Jiaozhou Bay in space, expanding to both the northeast and southwest directions. This result coincided with the direction of the economic focus areas’ distribution, such as Huangdao and Shinnan districts.

The values of features are shown in Table IV. Based on the value of the minor axis, NR had the widest distribution, and SL had the smallest range. Based on the flatter, the directional
trend of SL was the most obvious, followed by recreational recreation (RR) and cultural leisure (CL). The distribution of NR was more scattered and less directional. From the azimuth angle, the axes of different types of tourism resources were all in the northeast–southwest direction, with the NR angle being the largest, followed by SL, CL, and RR.

B. Hotspot Identification of Leisure Tourism Resources Based on KDE

Using the KDE, the higher the density value, the higher the level of territorial spatial agglomeration. As can be seen in Fig. 8, the spatial distribution of Qingdao’s total leisure tourism resources was characterized by “one cluster with multiple core points.” “One cluster” refers to hotspots with the highest density value, including Shinan and Shibei, and “multiple core points” refer to subhotspots, such as Huangdao, Chengyang, Jiaozhou, and Laixi. Shinan, Shibei, and other hotspots are the core areas of Qingdao’s economy and culture and contain many popular attractions, such as Zhanqiao Park, the Eight Great Passes, and May Fourth Square. Thus, powerful agglomeration of tourism resources made these areas hotspots with highest density values. Although the above-mentioned subhotspot areas also contain famous attractions—such as Huangdao Golden Beach and Fantawild Dreamland Park—the clustering effects of their attraction resources were not as strong as those of the attractions in the hotspots. Therefore, these areas form subhotspot areas.

Judging from the density values of different types of leisure tourism resources, which are shown in Fig. 9, the distribution patterns of the RR and CL resources were consistent with that of the overall tourism resources. Districts, such as Shinan and Shibei, formed the hotspots and spread to neighboring areas, such as Huangdao and other districts, forming subhotspots. In terms of the NR resources, Lao Mountain had a large number of natural tourism resources, thereby forming a large high-density area. The location of NR’s hotspots was different from that of the other types; they were located at the junction of Chengyang and Laoshan because this area includes Qingdao Horticulture Expo Park and some spots in Laoshan Scenic Area. Additionally, there were fewer POIs of NR than of other types because these POIs mostly depend on areas with better natural conditions, such as mountains, and their overall distribution was not as the RR and CL types. Similar to the NR resources, the SL resources had few POIs, and their overall agglomeration was weak. The hotspots of SL were mainly concentrated in Shinan, Shibei, Huangdao, and the east coast of Laoshan because these areas have many resort hotels and farm family resorts.

C. Identification of Leisure Tourism Zones Based on Improved TF–IDF Method

The improved TF–IDF method was applied to identify the leisure tourism zones from the grid and block scales, respectively. The results are shown in Fig. 10. Spatially, the RR and CL zones were more likely to be distributed at the center of each district, while the number of SL zones was relatively small, and the distribution of these zones was scattered.
the number of NR zones was also small, but they were more concentrated than the SL zones and were mainly distributed in suburbs, such as Lao Mountain in Laoshan, Daze Mountain in Pingdu, and Xiaozhu Mountain in Huangdao.

Fig. 11 shows the proportions of leisure tourism POIs of all types on the block scale. In terms of the main class, the RR type had the largest proportion, accounting for 45.07% of all the blocks, followed by the CL, NR, and SL. In terms of the subclass blocks included in the RR type, the CTL type accounted for 12.49% of all the blocks, and then, in descending order of proportion were SL, health care (HC), and others. This indicates that the leisure tourism blocks in Qingdao are mainly of the RR type, especially the blocks with gourmet restaurants and shopping malls.

As for the scale of block or grid, each has its own advantages and disadvantages in identifying tourism zones. While identification in block scale can better portray tourism zones with dense road networks, it ignores many details in the large blocks formed by sparse road networks. The identification in grid scale can be presented in grid units of the same size throughout the whole study area, but in daily life, people do not use grids to describe a certain zone, and their practical significance is not as strong as the block scale.

To compensate for the limitations of the two scales, we replaced large blocks with grids to form a hybrid scale. Specifically, the blocks are sorted in descending order of area, and the top $k$ percent of blocks are selected to integrate the grid. Fig. 12 shows the hybrid scale of the top 5% and 1% block integrating with the grid.

In order to evaluate the identification performances at the four scales, we selected 30 samples of each class using a random stratified sampling method, which means a total of 480 samples ($30 \times 4 \times 4$). Then, the actual leisure tourism categories of the samples were finally determined through the manual analysis of multiple sources, including urban tourism planning maps, high-resolution remote sensing images, and online maps. Finally, the OA and Kappa coefficient were selected as evaluation metrics, and their results are shown in Table V.

It can be seen that the OAs of the four scales are all above 80%. It means that our framework works better on all four scales. Specifically, the OA at the two hybrid scales achieve the best performances. The Kappa coefficient at the hybrid scale of the top 1% is 88.89%, which is 8.89% and 5.56% higher than those of grid and block scales, respectively. These results indicate that our framework with the hybrid scale is more consistent with the perception of leisure tourism planning.

D. Relationship Between the Factors and Leisure Tourism Resources’ Distribution

The factor detector and interaction detector were used to calculate the explanatory power of the influencing factors on the distribution of leisure tourism resources. Here, we mainly discuss the most important effects.

In terms of the overall leisure tourism POIs, the top six factors with the highest explanatory power could be ranked as NQPDA (0.428) > POP (0.401) > BU_P (0.380) > GDP (0.331) > VCI (0.237) > BAI (0.234), as shown in Fig. 13.
power of the four factors (NQPDA, POP, BU_P, and GDP) exceeded 0.3. Moreover, the four factors were closely related to human activity. The explanatory power of ELE, DIST_WATER, and the landscape metrics was weak, at no more than 0.1.

In addition, we calculated the explanatory power of the factors on the distribution of each of the main types of POIs, and the results are shown in Fig. 14. For the RR POIs, the result was similar to that of the overall POIs. The power of the socioeconomic factors was stronger than that of the natural factors, and the reason for this is that the RR POIs occupied the highest percentage of the overall POIs. For the CL POIs, the explanatory power of all the factors was relatively weak, at no more than 0.3. Among these factors, the top six factors with the highest power could be ranked as POP (0.241) > NQPDA (0.234) > GDP (0.218) > BU_P (0.183) > VCI (0.106) > BAI (0.090). The power of socioeconomic factors was also stronger than that of the natural factors. In terms of the NR POIs, the result was unlike that of the above-mentioned two main types; the top factors could be ranked as VCI (0.203) > ELE (0.189) > BAI (0.188) > PAFRAC (0.080) > PARA_MN (0.065) > NP (0.061). The explanatory power of some natural factors was stronger than that of the socioeconomic factors, which was in line with what we know from daily life. For the SL POIs, the top factors could be ranked as GDP (0.227) > POP (0.226) > NQPDA (0.188) > BU_P (0.141) > VCI (0.112) > BAI (0.093). This ranking was also strongly influenced by socioeconomic factors.

As shown in Fig. 15, compared with the results of the factor detector, we found that the explanatory power of the interaction between any two factors was greater than that of any single factor. Among these interactions, the explanatory power of interaction between NQPDA or POP and any other factor was maintained at more than 0.4, and the interaction between NQPDA and CONTAG reached the strongest explanatory power (0.680). Interestingly, the power of CONTAG was greater than the factor detection result and was partly evidence of the potential role of this landscape metric.

The interaction results of the four main types of leisure tourism POIs are shown in Fig. 16. For the result of the RR POIs, the overall situation was generally consistent with that of the overall POIs because of the highest proportion. For the CL POIs, the explanatory power of interaction between NQPDA and CONTAG was also the highest (0.49) among the interactions; moreover, the interactions between socioeconomic factors played a major explanatory power. In addition, the PAFRAC also played a larger role among these interactions than the result of the factor detector. In terms of the NR POIs, the explanatory power of interaction between BAI or VCI or ELE and any other factor exceeded 0.2, which indicates the importance of natural factors to the NR POIs; however, this does not mean that socioeconomic factors are not important to the NR POIs. The interaction between ELE and GDP showed the highest explanatory power (0.42). For the SL POIs, the role of the socioeconomic factors was stronger than that of the natural factors; interestingly, a relatively high explanatory power was also achieved by the interaction between some landscape metrics, such as PAFRAC, LPI, and ED.
As the geodetector cannot indicate whether the values of the factors are positive or negative for the resources, we selected the nine most important factors, calculated the mean values from different types of zones on the grid scale, and then plotted the box plots by ignoring the outliers (see Fig. 17).

For the PAFRAC, the fluctuation degree of the NR zones was greater than that of the other three types, while the RR zones had a relatively small range and low level of values. For the CONTAG, the four types had a similar degree, with a slightly higher overall degree for the CL zones. As for the four socioeconomic factors of NQPDA, GDP, POP, and BU_P, their situations were more consistent across the four types in the four figures, with greater fluctuations in the three types of the RR, CL, and SL, while the fluctuations in the NR zones were smaller and the overall degree remained low, indicating that compared with the other three types of resources, the NR zones were less dependent on the socioeconomic factors. In terms of the three natural factors of BAI, VCI, and ELE, it was clear that the overall degree of NR zones was higher than the other three types of zones, showing that they were more distributed in areas with a higher BAI, VCI, and ELE, which aligned with people’s conceptual perceptions.

V. DISCUSSION

Big data are changing the discovery of the information on urban tourism resources. In this study, we used new data POIs to represent leisure tourism sites and based on the advantages of data and the existing classification system, we divided the types of leisure tourism resources that could reveal the characteristics in a detailed way. Qingdao’s overall leisure tourism resources were unevenly distributed, with most of them being concentrated in the area around Jiaozhou Bay. Specifically, the RR and CL resources were mainly distributed in the economic centers of each district, while the SL and NR resources tended to be distributed in the suburbs. To investigate the importance of different factors for the distribution of leisure tourism, various factors were selected and their corresponding data were generated based on the basic data. Then, we used the geodetector to explore the relationship between factors and this phenomenon. The result indicated that, owing to different socioeconomic and natural conditions between zones, this phenomenon is not influenced by these factors acting alone but is the result of a combination of factors working together.

Moreover, the TF–IDF method, which is used to identify functional zones, was improved and applied to identify different leisure tourism zones in a novel manner. It is beneficial to understand the distribution of urban tourism and resources at different spatial scales. In many cases, visitors are not limited to one attraction type but are likely to focus together on leisure tourism zones adjacent to that type of attraction [44]. For example, when a person wants to go shopping, an RR zone containing many shops or malls would be expected, and when a person plans to visit historical or cultural places and natural scenic spots, the CL zone, which is closer to the NR zone, would be optional.

These findings have potentially important implications for the sustainable management of leisure tourism resources because this research framework allows us to obtain a better understanding of how these resources are distributed spatially and to detect the role of factors quantifiably. This is particularly crucial in the case-study area in view of the huge changes in infrastructure construction and land use in recent years. Improved knowledge of the different ways in which tourism resources can be influenced by these factors can help the government to better understand the characteristics of leisure tourism and, thus, promote tourism-related decision making.

VI. CONCLUSION

With the development of big data entering a new stage, tourism-related data tend to be multisourced and large. Therefore, finding useful information from massive data has become challenging. This article presents a novel framework to explore these characteristics based on the multisource data. In this study, the POIs, OSM roads, satellite images, and other sources of data were used to study the spatial characteristics of leisure tourism. Besides, an improved TF–IDF method, the geodetector method, and other technical methods were also exploited to accomplish this process.

Taking Qingdao as the case-study area, the following conclusions can be drawn by using the proposed framework.

1) The spatial pattern of the overall leisure tourism resources was characterized by “dense middle and sparse edges,” the resources were mainly concentrated on Jiaozhou Bay, and the distribution trend was along the “northeast to southwest” direction. The spreading range of the NR POIs was the largest, while the SLs range was the smallest.

2) The spatial distribution of the overall resources could be characterized by “one cluster with multiple core points,” and this feature was related to the spatial pattern; moreover, the distributions of the RR and CL POIs were consistent with that of the overall resources. The NR and SL types had a small number of POIs, and their respective agglomeration effects were also weaker than that of the other types.

3) Spatially, the RR and CL zones were more likely to be distributed at the center of each district, while the number of SL zones was relatively small, and the distribution of these zones was scattered. Intuitively, the number of NR
zones was also small, but they were more concentrated than the SL zones and were mainly distributed in suburbs, such as Lao Mountain in Laoshan, Daze Mountain in Pingdu, and Xiaoazhu Mountain in Huangdao. According to the identification result on block scale, the leisure tourism blocks in Qingdao were mainly the RR type, especially the blocks with gourmet restaurants and shopping malls.

4) As shown in the detection results, the distribution of the overall resources was mainly influenced by the socioeconomic conditions, such as traffic condition, economic level, and others. The POIs of the RR type, CL type, and SL type were also similar to those of the overall resources, while the NR resources were mainly controlled by the natural conditions. Although the role of some factors in the factor detection results was not obvious, this does not mean they did not work; they may have explanatory power in interactions with other factors. For instance, GDP’s role is not obvious in the NR type’s factor detection result, but the interaction between ELE and GDP showed the highest explanatory power in the result of the NR POIs.

Although the achievements in this study, there are still some limitations that are deserving of being investigated in the future study. For instance, the POI data represent leisure tourism sites, which can reflect the spatial characteristics of the tourism resources but not the behavioral characteristics of tourists. And we used online information on Ctrip’s popular attractions to make up for this shortcoming. A more accurate analysis should use inbound tourists’ mobile phone signaling data or data from social media sites (such as Twitter and Facebook) to sense the tourists’ behavior. In addition, we only identified the dominant leisure tourism types for different zones and did not identify the mixed-tourism types. However, identifying the mixed-tourism types of the zones will help us understand the characteristics of urban leisure tourism resources in more detail. Therefore, we plan to identify the mixed-tourism types for the zones to fill this gap in the following studies.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments to improve this article.

REFERENCES

[1] L. Yan, F. Duarte, D. Wang, S. Zheng, and C. Ratti, “Exploring the effect of air pollution on social activity in China using geotagged social media check-in data,” Cities, vol. 91, pp. 116–125, Aug. 2019.
[2] J. Lu, L. Hu, K. Hung, and Z. Mao, “Assessing servicescape of cruise tourism: The perception of Chinese tourists,” Int. J. Contemporary Hospitality Manage., vol. 29, no. 10, pp. 2556–2572, Oct. 2017.
[3] Z.-F. Huang, Y. Zha, L.-W. Yuan, X.-H. Hu, and F.-D. Cao, “The intention, classification and evaluation of leisure tourism resources: A case study of Changzhou, Jiangsu province,” Geographical Res., vol. 30, no. 9, pp. 1543–1553, 2011.
[4] J. Tribe, The Economics of Recreation, Leisure and Tourism. London, U.K.: Routledge, 2020.
[5] K. Zakariya, N. Z. Harun, and M. Mansor, “Place meaning of the historic square as tourism attraction and community leisure space,” Procedia-Social Behav. Sci., vol. 202, pp. 477–486, Aug. 2015.
[6] G. Rainer, “Constructing globalized spaces of tourism and leisure: Political ecologies of the Salta wine route (NW-Argentina),” J. Rural Stud., vol. 43, pp. 104–117, Feb. 2016.
[7] S. Y. Han, M.-H. Tsou, and K. C. Clarke, “Revisiting the death of geography in the era of big data: The friction of distance in cyberspace and real space,” Int. J. Digit. Earth, vol. 11, no. 5, pp. 451–469, May 2018.
[8] F. E. A. Horita, J. P. de Albuquerque, V. Marchezini, and E. M. Mendiondo, “Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil,” Decis. Support Syst., vol. 97, pp. 12–22, May 2017.
[9] L. Ahmed et al., “The role of big data analytics in Internet of Things,” Comput. Netw., vol. 129, pp. 459–471, Dec. 2017.
[10] F. M. Simon and R. Schroeder, “Big data goes to Hollywood: The emergence of big data as a tool in the American film industry,” in Second International Handbook of Internet Research. Dordrecht, the Netherlands: Springer, 2019, pp. 549–567.
[11] A. Bujari, M. Ciman, O. Gaggi, and C. E. Palazzi, “Using gamification to discover cultural heritage locations from geo-tagged photos,” Pers. Ubiquitous Comput., vol. 21, no. 2, pp. 235–252, Nov. 2017.
[12] G. Cai, K. Lee, and I. Lee, “Itinerary recommender system with semantic trajectory pattern mining from geo-tagged photos,” Expert Syst. Appl., vol. 94, pp. 32–40, Mar. 2018.
[13] T. Osawa, Y. Ueno, T. Nishida, and J. Nishihiro, “Do both habitat and species diversity provide cultural ecosystem services? A trial using geo-tagged photos,” Nature Conserv., vol. 38, pp. 61–77, Mar. 2020.
[14] D. Wang, W. Zhong, Z. Yin, D. Xie, and X. Luo, “Spatio-temporal dynamics of population in Shanghai: A case study based on cell phone signaling data,” in Big Data Support of Urban Planning and Management. New York, NY, USA: Springer, 2018, pp. 239–254.
[15] Z. Lu, Z. Long, J. Xia, and C. An, “A random forest model for travel mode identification based on mobile phone signaling data,” Sustainability, vol. 11, no. 21, Oct. 2019, Art. no. 5950.
[16] K. Chin, H. Huang, C. Horn, I. Kasanicky, and R. Weibel, “Inferencing fine-grained transport modes from mobile phone cellular signaling data,” Comput., Environ. Urban Syst., vol. 77, Sep. 2019, Art. no. 101348.
[17] Y. Long and J.-C. Thill, “Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing,” Comput., Environ. Urban Syst., vol. 53, pp. 19–35, Sep. 2015.
[18] Y. Zhou, Z. Fang, Q. Zhan, Y. Huang, and X. Fu, “Inferencing social functions available in the metro station area from passengers’ staying activities in smartphone card data,” ISPRS Int. J. Geo-Inf., vol. 6, no. 12, 2017, Art. no. 394.
[19] N. S. Aslam, T. Cheng, and J. Cheshire, “A high-precision heuristic model to detect home and work locations from smartphone card data,” Geo-Spatial Inf. Sci., vol. 22, no. 1, pp. 1–11, Nov. 2019.
[20] J. C. García-Palomares, J. Gutiérrez, and C. Mínguez, “Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS,” Appl. Geogr., vol. 63, pp. 408–417, Sep. 2015.
[21] A. Abbasi, T. H. Rashidi, M. Maghreb, and S. T. Waller, “Utilising location based social media in travel survey methods: Bringing twitter data into the play,” in Proc. 8th ACM SIGSPATIAL Int. Workshop Location-Based Social Netw., 2015, pp. 1–9.
[22] N. Mou, R. Yuan, T. Yang, H. Zhang, J. Tang, and T. Makkonen, “Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China,” Tourism Manage., vol. 76, Feb. 2020, Art. no. 103955.
[23] Z. Wang, S. Y. He, and Y. Leung, “Applying mobile phone data to travel behaviour research: A literature review,” Travel Behav. Soc., vol. 11, pp. 141–155, Apr. 2018.
[24] F. He and P. Wei, “Research on comprehensive point of interest (POI) recommendation based on spark,” Cluster Comput., vol. 22, no. 4, pp. 9049–9057, Feb. 2018.
[25] S. Xu, L. Qing, L. Han, M. Liu, Y. Peng, and L. Shen, “A new remote sensing images and point-of-interest fused (RPF) model for sensing urban functional regions,” Remote Sens., vol. 12, no. 6, Mar. 2020, Art. no. 1032.
[26] Y. Yue, Y. Zhuang, A. G. O. Yeh, J.-Y. Xie, C.-L. Ma, and Q.-Q. Li, “Measurements of POI-based mixed use and their relationships with neighborhood vibrancy,” Int. J. Geographical Inf. Sci., vol. 31, no. 4, pp. 658–675, 2017.
[27] D. Xu, Z. Huang, H. Sun, X. Shi, H. Liu, and L. Tan, “The spatial characteristics and its influencing factors of leisure tourism resources in Nanjing,” J. Nanjing Normal Univ. (Natural Sci. Ed.), vol. 40, no. 1, pp. 127–133, 2017.
[28] L. Li, G.-L. Hou, S.-Y. Xia, and Z.-F. Huang, “Spatial distribution characteristics and influencing factors of leisure tourism resources in Chengdu,” J. Natural Resour., vol. 35, no. 3, pp. 683–697, 2020.
[29] X. Wang and J. Hou, “Spatial distribution features and influence factors of rural leisure tourism destinations in Shandong province,” *Scientia Geographica Sinica*, vol. 36, no. 11, pp. 1706–1714, 2016.

[30] S. Du, Shihong Du, B. Liu, X. Zhang, and Z. Zheng, “Large-scale urban functional zone mapping by integrating remote sensing images and open social data,” *GIScience Remote Sens.*, vol. 57, no. 3, pp. 411–430, Feb. 2020.

[31] X. Zhang, S. Du, and Q. Wang, “Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data,” *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 170–184, Oct. 2017.

[32] Z. Du, X. Zhang, W. Li, F. Zhang, and R. Liu, “A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou city, China,” *Trans. GIS*, vol. 24, no. 1, pp. 123–141, Oct. 2019.

[33] H. Shen, Q. Wang, C. Ye, and J. S. Liu, “The evolution of holiday system in China and its influence on domestic tourism demand,” *J. Tourism Futures*, vol. 4, pp. 139–151, Jul. 2018.

[34] Y. Yao et al., “Sensing spatial distribution of urban land use by integrating points-of-interest and google word2vec model,” *Int. J. Geographical Inf. Sci.*, vol. 31, no. 4, pp. 825–848, 2017.

[35] M. F. Goodchild and L. Li, “Assuring the quality of volunteered geographic information,” *Spacial Statist.*, vol. 1, pp. 110–120, May 2012.

[36] S. He, S. Yu, P. Wei, and C. Fang, “A spatial design network analysis of street networks and the locations of leisure entertainment activities: A case study of Wuhan, China,” *Sustain. Cities Soc.*, vol. 44, pp. 880–887, Jan. 2019.

[37] Q. Du, J. Zhou, T. Pan, Q. Sun, and M. Wu, “Relationship of carbon emissions and economic growth in China’s construction industry,” *J. Cleaner Prod.*, vol. 220, pp. 99–109, May 2019.

[38] S. Zhang, D. D. Zhang, J. Li, and Q. Pei, “Secular temperature variations and the spatial disparities of war in historical China,” *Climatic Change*, vol. 159, pp. 545–564, Jan. 2020.

[39] S. Yao, J. Wang, L. Fang, and J. Wu, “Identification of vehicle-pedestrian collision hotspots at the micro-level using network kernel density estimation and random forests: A case study in Shanghai, China,” *Sustainability*, vol. 10, no. 12, Dec. 2018, Art. no. 4762.

[40] K. Liu, P. Qiu, S. Gao, F. Lu, J. Jiang, and L. Yin, “Investigating urban metro stations as cognitive places in cities using points of interest,” *Cities*, vol. 97, Feb. 2020, Art. no. 102561.

[41] J.-F. Wang et al., “Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China,” *Int. J. Geographical Inf. Sci.*, vol. 24, no. 1, pp. 107–127, Feb. 2010.

[42] Y. Chang, K. Hou, Y. Wu, X. Li, and J. Zhang, “A conceptual framework for establishing the index system of ecological environment evaluation—A case study of the upper Hanjiang river, China,” *Ecol. Indicators*, vol. 107, 2019, Art. no. 105568.

[43] C. Wang, Q. Jiang, Y. Shao, S. Sun, L. Xiao, and J. Guo, “Ecological environment assessment based on land use simulation: A case study in the Hehe river basin,” *Sci. Total Environ.*, vol. 697, 2019, Art. no. 133928.

[44] S. Gao, K. Janowicz, and H. Couclelis, “Extracting urban functional regions from points of interest and human activities on location-based social networks,” *Trans. GIS*, vol. 21, no. 3, pp. 446–467, Jun. 2017.