

Article

Identifying Land Use Functions in Five New First-Tier Cities Based on Multi-Source Big Data

Wangmin Yang ¹ , Yang Ye ¹, Bowei Fan ¹, Shuang Liu ²  and Jingwen Xu ^{1,*}

¹ College of Resources, Sichuan Agricultural University, Chengdu 611130, China; yangwangmin@stu.sicau.edu.cn (W.Y.); yeyang@stu.sicau.edu.cn (Y.Y.); 201608171@stu.sicau.edu.cn (B.F.)

² Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610299, China; liushuang@imde.ac.cn

* Correspondence: 11893@sicau.edu.cn; Tel.: +86-182-15673015; Fax: +86-028-82690983

Abstract: With the continuous development of big data technology, semantic-rich multi-source big data provides broader prospects for the research of urban land use function recognition. This study relied on POI data and OSM data to select the central urban areas of five new first-tier cities as the study areas. The TF-IDF algorithm was used to identify the land use functional layout of the study area and establish a confusion matrix for accuracy verification. The results show that: (1) The common feature of these five cities is that the total number and area of land parcels for residential land, commercial service land, public management and service land, and green space and open space land all account for over 90%. (2) The Kappa coefficients were all in the range [0.61, 0.80], indicating a high consistency of accuracy evaluation. (3) Chengdu and Tianjin have the highest land use function mixing degree, followed by Xi'an, Nanjing, and Hangzhou. (4) Among the five new first-tier cities, Hangzhou and Nanjing have the highest similarity in land use function structure layout. This study attempts to reveal the current land use situation of five cities, which will provide a reference for urban development planning and management.

Keywords: POI data; TF-IDF; urban function identification; spatial analysis



Citation: Yang, W.; Ye, Y.; Fan, B.; Liu, S.; Xu, J. Identifying Land Use Functions in Five New First-Tier Cities Based on Multi-Source Big Data. *Land* **2024**, *13*, 271. <https://doi.org/10.3390/land13030271>

Academic Editors: Zongcai Wei, Yuting Liu and Andrzej Biłozor

Received: 24 December 2023

Revised: 8 February 2024

Accepted: 15 February 2024

Published: 21 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Thanks to the high-speed development of the economy driven by reform and open policy, China has experienced the largest urbanization process in human history [1]. Although the rapid urbanization process improves people's material conditions, it also brings many problems to urban planning and development, such as the single function of urban fringe areas and excessive traffic congestion within the city, which brings inconvenience to people's daily life and travel [2]. Disorderly urban expansion also poses threats to the natural ecological environment and disrupts ecological balance, resulting in economic, social, and ecological problems [3]. Therefore, the functional zoning of cities is essential for their scientific and reasonable construction and planning. China's latest policies emphasize the importance of urbanization and stress the need to identify and divide the functions of urban land use. This approach not only facilitates rational construction and planning of the city but also contributes to the optimization of urban layout, efficient distribution of urban resources, and the formation of a more efficient and orderly urban structure layout. This provides a solid foundation for the steady progress of urbanization [4–7].

The traditional methods of identifying urban land use functions rely on data from population census, land census, industrial statistics, and socio-economic statistics. However, these sources often involve confidentiality, making them challenging to access and resulting in lengthy update cycles. This greatly reduces the practicality of experiments, leading to research that often remains at a macro level, and is difficult to widely apply. With the advent of the network era and the rapid development of communication technologies, big data has emerged as a valuable resource. This includes various types of data such as remote

sensing images and environmental and meteorological data obtained through professional scientific research equipment. These big data have advantages such as high accuracy, wide coverage, strong timeliness, and low acquisition cost [8]. Compared with traditional remote sensing images, the network big data obtained through communication devices has better timeliness and microscale, including POI (point-of-interest) data, GPS (global positioning system) trajectory data, mobile phone signaling data, and bus card swiping data that can be obtained from users' daily lives [9]. Currently, more and more scholars are analyzing urban space by mining big data, focusing on exploring urban economic activities, resident mobility, and resident traffic trajectories from a micro perspective. For example, Frias Martinez et al. used Twitter's social media activity data to analyze urban land use functions in Manhattan, London, and Madrid using clustering analysis [10]; Jiang et al. used the frequency of POI data in the Boston metropolitan area to identify land types in cities and estimate employment distribution density based on this data [11]; Hu et al. used Landsat remote sensing image data and POI data to achieve the division of functional land in Beijing [12]; Ye et al. combined social media data with street level remote sensing image data to identify urban land use functions in Beijing [13]; Liu et al. used time series data of taxi travel trajectories to identify the functional layout of land use in Chengdu City [14]; Sun et al. achieved recognition of land use functions in Beijing by analyzing text information on Sina Weibo [15]; and Huang et al. divided the urban functional areas of Beijing and Wuhan by combining nighttime lighting data and daytime multi-perspective remote sensing image data [16]. With the progress and development of technology, the acquisition and analysis methods of big data are becoming increasingly perfect, which can further explore urban spatial layout and intuitively reflect urban spatial structure.

With the maturation of big data mining technology, many sample data and precise positioning information have been excavated in cities. The integration of big data with urban spatial analysis and the comprehensive analysis of hidden information has emerged as a research hotspot [17]. For example, Yuan et al. used the time, location, and type of crowd activity to represent the potential activity trajectory of the crowd and combined it with POI data kernel density analysis to identify the urban function of Beijing [18]; Feng et al. combined the Logic regression model with POI data to extract the urban functional zoning of the Wuhan urban area [19]; Zhai et al. combined the Place2vec model with the K-means clustering algorithm to partition the functional layout of land use in Wuxi City, with an overall accuracy of 74.24% [20]; Yan et al. used the KD-Tree clustering algorithm and Tyson polygon algorithm combined with POI data to accurately partition urban functional areas [21]; and Sun et al. used Word2vec, LDA, and Block2vec models to identify land use functions in the central urban area of Wuhan, which also confirmed that the Word2vec model has the best recognition accuracy [22]. In addition, the TF-IDF (term frequency inverse document frequency) quantification method based on machine learning has also received the attention of some scholars [23–25]. The TF-IDF algorithm is a weighting technique used for key information in data mining. This method has unique advantages in analyzing policy texts at the macro level and extracting keywords of different importance levels [26].

Theoretical research on urban land use functions has reached a relatively mature stage, and technical methods have undergone significant transformation. However, current research on identified urban functional zoning has primarily focused on applying new urban planning technologies, evaluating current layout features, and identifying functional issues. There is a lack of in-depth analysis of the identification results and insufficient connection with regional urban planning policies and development models. As a result, there is a challenge in providing feedback for future planning and design, as well as offering guidance for urban formulation and management. Urban function identification based on POI data often stems from surveying and mapping, cartography, and related disciplines. In the current era of land spatial planning, it is essential to conduct research on land use functions from the perspective of urban planning [27]. Moreover, current research primarily relies on manual comparative statistical analysis for the accuracy evaluation of

the final recognition results, which is significantly influenced by the researcher's personal subjective will. The new first-tier cities are a list of 337 prefecture-level-and-above cities in China, evaluated and analyzed by First Financial News based on five major indicators: commercial resource agglomeration, urban hub, urban human activity, lifestyle diversity, and future plasticity. These new first-tier cities have strong representativeness, and their development process can be seen as a microcosm of China's urbanization process to a certain extent. However, current research on new first-tier cities predominantly focuses on cultural, commercial, financial, and other fields, with limited attention to the functional layout of urban land use. Furthermore, existing research on the functional structure layout of urban land use is primarily confined to the study of individual cities, with inadequate comprehensive analysis and research on the functional layout of land use across multiple cities.

This study is grounded in OSM data and POI data to identify the current land use functions in the central urban areas of five new first-tier cities. The aim is to enhance understanding of the current situation and distribution pattern of land use functions in major Chinese cities and to provide a reference basis for optimizing the allocation of urban and national spatial planning. Furthermore, building upon the results of urban land use function identification, this study conducts an analysis from three perspectives: urban functional spatial elements, urban functional mixing degree, and urban structural similarity, to ascertain whether there is a certain regularity in the land use functions of these five new first-tier cities. The objective is to establish a data foundation and provide theoretical support for the formulation of planning policies for underdeveloped cities in China, thereby ensuring the steady and smooth progress of urbanization in the country.

2. Overview of the Study Area

The research areas are the central urban areas within the ring highways of Chengdu, Nanjing, Hangzhou, Xi'an, and Tianjin (Figure 1). These five cities have outstanding performance in terms of GDP (gross domestic product), number of permanent residents, concentration of commercial resources, and urban hub, all of which are extremely strong new first-tier cities. The selected central urban area concentrates most of the city's construction land and permanent population, making it a key area for urban planning and construction. And the urban functional distribution within this area is concentrated, with high spatial heterogeneity and suitable block scale, making it an ideal area for identifying and analyzing the current urban land use functions.

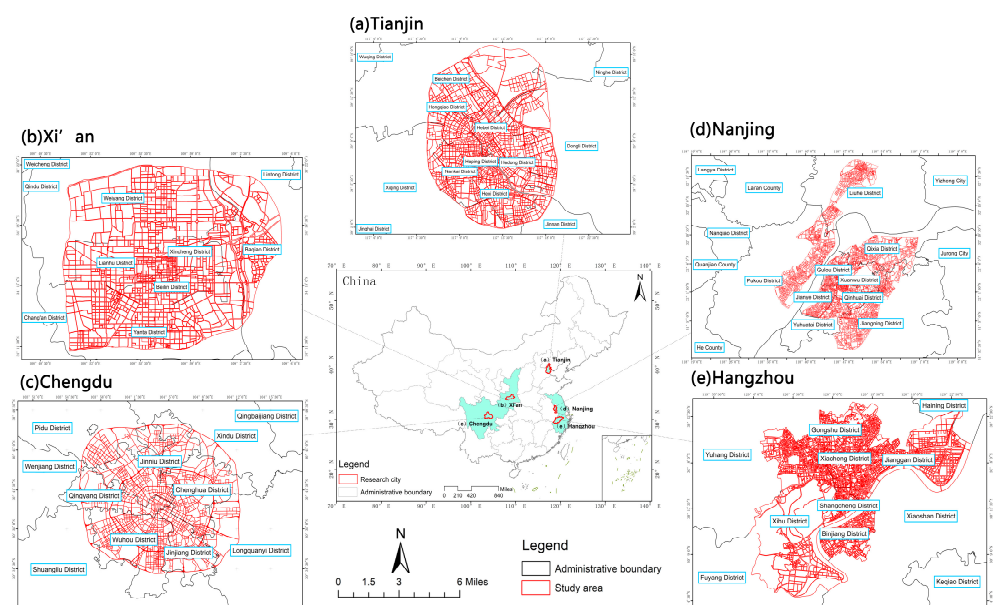


Figure 1. Research areas of the five cities.

3. Data Sources and Processing

3.1. Data Sources

This study uses OSM (OpenStreetMap) data as the base map of the study area, which is an open-source world map that can be freely edited by Internet users. Users can access and download data for free through an open license agreement (www.openstreetmap.org). This study uses POI data from Amap (<https://lbs.amap.com>), which was crawled on its API platform in December 2022 to ensure that the POI data utilized in this research is highly current. The total number of raw POI data for the five cities reached more than 1,735,000 items. And the land use status of the five cities on the Gaode satellite map was referenced as the real value to construct the confusion matrix to test the precision of identifying urban functions. The technical pathway is illustrated in Figure 2.

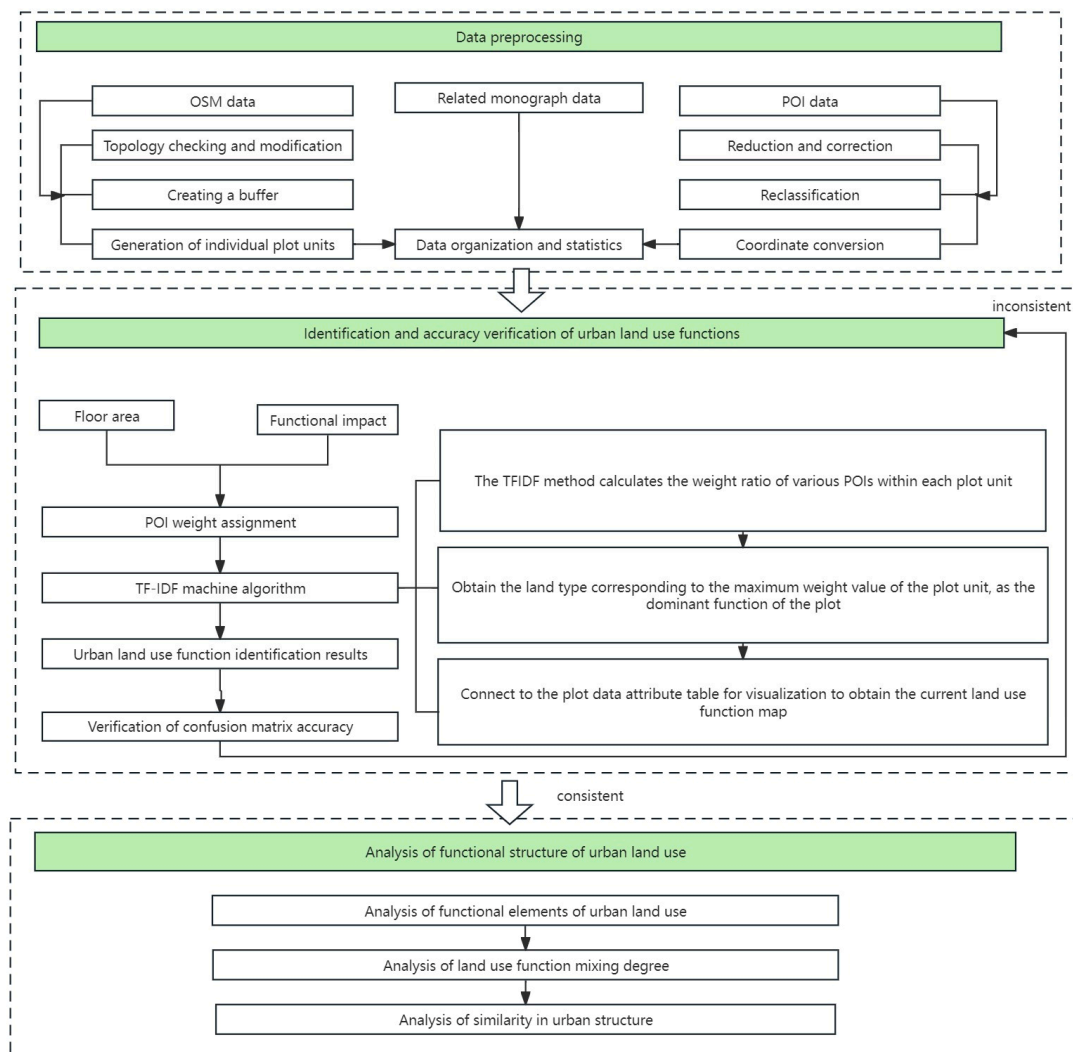


Figure 2. Flowchart of the research.

3.2. Data Preprocessing

3.2.1. OSM Data Preprocessing

(1) Topology checking and modification.

The original data volume of OSM data is huge and contains many disconnected, cluttered, and duplicated roads. Topology checking and modification of road line elements in ArcMap is performed to remove and fix the problematic roads by establishing topology rules such as no overlapping, no suspension points, no pseudo-nodes, no self-intersections, and so on.

(2) Creating a buffer.

Delete suburban rural roads, pedestrian roads, bicycle lanes, unknown roads, and some internal roads based on road attribute information. Then, use highways, expressways, urban main roads, urban secondary roads, and some internal roads as the main line elements. Refer to China's "Urban Road Engineering Design Specification" (CJJ37-2019) to determine the grade width of different roads, and establish buffer zones on this basis. The specific width of the buffer is shown in Table 1.

Table 1. Urban road buffer widths.

Road Class	Name of Road	Buffer Width (m)
Primary Road	freeway	60
Secondary Road	speedway	45
	trunk road of city	
Tertiary roads	secondary urban road	30
Fourth-class road	Internal roads	20

(3) Generation of individual plot units.

After generating buffer zones for different roads, independent plot units are obtained by removing the road space. The smaller and fragmented parcel units are removed by combining the actual city conditions and satellite electronic maps. Finally, the base map of the five new first-tier cities is obtained for this study.

3.2.2. Preprocessing of POI Data

(1) Data Reduction and Correction.

The original POI data of Amap is huge in volume, and there are problems such as missing information and duplicated information, and some information features are not significant. Firstly, POI data with missing coordinate information is directly deleted. Then, POI data with the same coordinate information and name are regarded as duplicated data, and only one piece of data is simplified and retained. Finally, POI data such as public toilets and bus stops, which are not significant in terms of functional features, are selected for deletion.

(2) Data Reclassification.

The classification standard of POI data of Amap is relatively standardized among the POI data of many websites, but it also has the problems of rigid classification, duplication of classification, and overly detailed criteria of subtables, which lead to the fact that the POI data of Amap cannot be effectively connected with the actual land use function of the city, and it is challenging to establish a standardized framework for the future planning and construction of the city. Therefore, this study reclassified the POI data of five new first-tier cities based on multiple classification criteria, totaling eight primary classifications [28]. For details, see Table 2.

(3) Data coordinate conversion.

The original POI data obtained directly used the GCJ-02 coordinate system. It is necessary to unify the spatial coordinates and convert the coordinate system of POI data to the coordinate system of OSM data. After the conversion is completed, the POI data are vectorized and then converted to projected coordinates. Finally, a shape file of POI data points is generated.

Table 2. List of classifications after POI reclassification.

Functional Site Type	Secondary Category	POI Detailed Category
Public administration and public service facilities	Administrative offices	Party and government organs; social organizations; police stations; institutions
	Cultural facilities	Book exhibitions; cultural events
	Healthcare	Hospitals; primary care; public health
	Education	Higher education; secondary vocational; primary and secondary schools; kindergartens; other education
		Research organizations
	Sports	Public gymnasiums
Commercial service land	Commercial land	Shopping centers; supermarkets; convenience stores
		Regular restaurants; fast food restaurants
		Hotels; hostels
	Commercial and financial land	Commercial office buildings; corporate companies
	Recreational land	Theaters
Serviced land	Other commercial service land	Hairdressing and cosmetology; training institutions
	Supply of land for facilities	Land for water supply; land for electricity supply; land for communications
	Land for environmental facilities	Drainage land; sanitation land
	Land for security facilities	Firefighting land; flood control land
Green spaces and open spaces	—	Parks and squares; heritage sites
Industrial and mining area	Industrial site	Industrial parks
Residential land	—	Living area
Transportation	Transportation site	Train station; bus station
Warehousing land	Land for logistics and warehousing	Logistics warehousing

4. Research Methodology

4.1. Urban Land Use Function Identification

4.1.1. POI Weight Assignment Methods

Due to the lack of information on the area occupied by physical objects and the level of public understanding in POI data, it is difficult to objectively reflect the importance of physical objects in their respective regions, which in turn affects the determination of dominant functions in that region [29,30]. In the process of quantifying the importance of different types of POI data, it is necessary to comprehensively consider indicators such as their quantity, function, scale, and public understanding. Given the difficulty in collecting relevant information and quantifying some indicators, this study used two indicators, land area and functional impact, to assign weight values to POI data [31]. The land area refers to the designated construction land within the land use boundary, requiring measurement, approval, and planning by the relevant management authorities. It primarily encompasses the land requisition area and the net land area of construction projects, including the main buildings, supporting facilities, activity areas, roadways, and green spaces. Additionally, the environmental awareness of urban residents can reflect the functional impact and significance of various types of POI entities over a specific period, which can be gauged through public awareness. This study determines the land area of different POI data and their functional impact based on standards such as the “Retail Format Classification Standard (GB/T18106-2021)” and the relevant research by Zhao et al. [32]. The

functional impact of each type of POI data is determined (Table 3), and then weighted values are calculated.

Table 3. POI data area and functional influence reference table.

Functional Site Type	Secondary Category	POI Detailed Category	Reference Footprint (hm ²)	Functional Influence
Public administration and public service facilities	Administrative offices	Party and government organs; social organizations; police stations; institutions	0.8; 0.1; 0.5	0.3550
	Cultural facilities	Book exhibitions; cultural events	0.2; 1	0.6760
	Healthcare	Hospitals; primary care; public health	6; 2; 0.1	0.7460
	Education	Higher education; secondary vocational; primary and secondary schools; kindergartens; other education	10; 3; 1.5	0.6760
	Scientific research	Research organizations	0.5	0.5010
	Sports	Public gymnasiums	1	0.5060
Commercial service land	Commercial land	Shopping centers; supermarkets; convenience stores	1; 0.2	0.5220
		Regular restaurants; fast food restaurants	0.1; 0.05	0.8140
	Commercial and financial land	Hotels; hostels	5; 0.3	0.9260
		Commercial office buildings; corporate companies	0.2	0.3520
	Recreational land	Theaters	2	0.2260
	Other commercial service land	Hairdressing and cosmetology; training institutions	0.2; 0.1	0.1250
Serviced land	Supply of land for facilities	Land for water supply; land for electricity supply; land for communications	0.5	0.0000
	Land for environmental facilities	Drainage land; sanitation land	0.5	0.0000
	Land for security facilities	Firefighting land; flood control land	0.5	0.0000
Green spaces and open spaces	—	Parks and squares; heritage sites	10	0.7520
Industrial and mining area	Industrial site	Industrial parks	5	0.2690
Residential land	—	Living area	8	0.7890
Transportation	Transportation site	Train station; bus station	50	0.0000
Warehousing land	Land for logistics and warehousing	Logistics warehousing	2	1.0000

Referring to the method adopted by Li et al. [33] in urban land functional zoning, first assign values to the reference land area corresponding to different types of POIs (Table 4), and then assign values to the functional impact of each type of POI (Table 5).

Table 4. POI floor area assignment table.

Area (m ²)	0–100	100–1000	1000–3000	3000–5000	5000–10,000	>10,000
Assignment	1	10	30	50	80	100

Table 5. Functional impact assignment table.

Functional Impact	0	0.3057	0.3350	0.5010	0.5069	0.6548	0.7460	0.8146
Assignment	0	30.6	33.5	50.1	50.7	65.5	74.6	81.4

Five different weighting ratios of 1:9, 3:7, 5:5, 7:3, and 9:1 were set for the floor space and functional influence. Through the precision test of the sampled plots, it was found that the average precision was the highest when the weighting ratio was 5:5. The weighting ratio of 5:5 was set for the POI for the assignment of weights with the weighting formula as follows:

$$W_{ij} = 0.5 \times A_{ij} + 0.5 \times I_{ij} \quad (1)$$

where W_{ij} is the weight value of the i th POI data in parcel j , A_{ij} is the footprint assignment of the i th POI data in parcel j , and I_{ij} is the functional influence assignment of the i th POI data in parcel j .

4.1.2. TF-IDF Machine Algorithm

In order to highlight the unique POI types within each plot unit and reduce the interference of general POI types, the identification method used in this study is the TF-IDF statistical method. TF-IDF machine algorithm is a statistical method commonly used in information retrieval and text processing [34]. In the study of urban functional land use, all the constituent land parcels of the study area are considered as a collection of documents, with each individual land parcel being treated as a document, and the POI data distributed in the land parcels are regarded as different words. By utilizing the TF-IDF algorithm to calculate and determine which type of POI data has the highest TF-IDF value in the individual land parcels, it signifies the most “important” type of POI and also reflects the dominant function of the land parcel [35]. Considering that the simple structure of IDF does not effectively reflect the importance and distribution of words in the specific calculation process, this study adopts POI data endowed with weight values for calculation for the purpose of enhancing the algorithm’s effectiveness in identifying the urban land use functions. The calculation can be performed using the following formula.

$$TF_{ij} = \frac{n_{ij}}{\sum_k n_{kj}} \quad (2)$$

$$IDF_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (3)$$

$$TF - IDF_{ij} = TF_{ij} \times IDF_i \quad (4)$$

where i denotes the POI data and j denotes the parcel unit where the POI data is located. n_{ij} denotes the frequency of POI data i in parcel unit j ; $\sum_k n_{kj}$ denotes the total frequency of all the POI data occurrences in parcel unit j ; D denotes the total number of parcel units; and $|\{j : t_i \in d_j\}|$ denotes the total number of parcel units containing POI data i .

The specific method includes three key steps. Firstly, the preprocessed POI data and road network are spatially connected in ArcGIS to obtain the corresponding number of various POIs for each plot unit. The spatially connected attribute table is exported, and POI categories are merged according to the methods in the literature. Then, the names and quantities of each POI category for each spatial unit are read and written into the document. Secondly, according to the TF-IDF algorithm, read the documents corresponding to each spatial unit and calculate and output the weight results for each category of each document (Namely, spatial unit). Finally, the category with the highest weight can be considered as the dominant function of the plot.

4.2. Verification of the Accuracy of Urban Land Use Function Identification Results

After obtaining the land use function layout of the five new first-tier cities, if we want to analyze and evaluate them, the analysis and evaluation need to be based on the recognition results with sufficient accuracy. This study randomly selected a certain number of parcel units for each different land use type as the identification results and the corresponding Gaode satellite map of the cities as the actual results; meanwhile, it used the method of constructing confusion matrix to test the accuracy of the land use function identification results of the five new first-tier cities, supplemented by the manual visual comparison test method, in which the land use function identification results of the five new first-tier cities’ downtowns are contrasted with the land use functions of the Gaode satellite electronic maps by manual visual comparison.

A confusion matrix is constructed to contain four indexes: user accuracy (UA), Producer accuracy (PA), overall accuracy (OA), and Kappa coefficient, in which the Kappa coefficient is used to assess the precision of the recognition results [36]. The confusion matrix consists of N rows and N columns, where the rows represent the recognition results and the columns of the matrix represent the actual results. In this study, a certain number of randomly selected parcel units for each different land use type are used as the recognition results, and their city counterparts of the Gaode electronic satellite map are used as the actual results. The following is the calculation formula:

$$UA = \frac{X_{ii}}{\sum_{j=1}^N X_{ij}} \times 100\% \quad (5)$$

$$PA = \frac{X_{ii}}{\sum_{j=1}^N X_{ji}} \times 100\% \quad (6)$$

$$OA = \frac{\sum_{i=1}^N X_{ii}}{M} \times 100\% \quad (7)$$

$$Kappa = \frac{M \times \sum_{i=1}^N X_{ii} - \sum_{i=1}^N \left(\sum_{j=1}^N X_{ij} \times \sum_{j=1}^N X_{ji} \right)}{M^2 - \sum_{i=1}^N \left(\sum_{j=1}^N X_{ij} \times \sum_{j=1}^N X_{ji} \right)} \quad (8)$$

where M denotes the overall number of sampled parcel units, N denotes the number of land types, X_{ij} represents the total number of correctly identified parcels, $\sum_{j=1}^N X_{ij}$ denotes the sum of the number of parcels in row i , and $\sum_{j=1}^N X_{ji}$ denotes the sum of the number of parcels in column i .

4.3. Functional Mix of Urban Land Use

With the rise of the concept of intensive land use, government departments will also consider the degree of functional mixing in functional areas when formulating urban planning [37]. The mixing of land use functions within a certain range helps to realize the efficient use of urban land. The degree of land use mixing reflects the degree of mixing of functions in a parcel of land by calculating the proportion of various types of POI data within the parcel unit. This study further analyzes the current status of urban land use functions in the urban centers of five new first-tier cities by calculating the degree of mixing of land use functions in the downtown of the five new first-tier cities. The equation used to determine the extent of land use mixing is shown below [38]:

$$M = - \sum_{i=1}^n (p_i \times \ln p_i) \quad (9)$$

where M is the functional mixing degree of the site, n denotes the number of POI data types in the parcel, and p_i denotes the proportion of POI data of type i in the parcel to all POI data. For each parcel unit, the site mixing degree values were categorized into low, lower, general, high, and higher.

4.4. Similarity of Urban Land Use Structure

VSM (vector space mode) was developed in the late 1960s when G Salton et al. [39] proposed a model that was first used in SMART information retrieval systems and has now become a commonly used model in natural language processing [40]. This model transforms text into high-dimensional vectors, with feature terms as its basic unit. Each dimension of a vector corresponds to a feature item of the text, and the dimension also represents the weight value of the feature item in the text, and the degree of the weight value also indicates the importance of the feature item in its text. When comparing the similarity between two documents, the inner product between vectors is usually used in calculations. In this study, the document was considered as a city, and the feature items were considered as the type of land use. Referring to the algorithm proposed by Li et al. [41],

the similarity of land use functional structure between two pairs of five cities was calculated. The formula is as follows:

$$\text{Similarity}_{vz} = \sqrt{\sum_{i=1}^n (w_{vi} \times c_{vi} - w_{zi} \times c_{zi})^2} \quad (10)$$

where v and z denote the name of the city, similarity_{vz} denotes the spatial distribution structure similarity between v city and z city, w_{vi} denotes the value of the weight of the functional point of category i of the v city, c_{vi} denotes the number of the functional point of category i in the v city, and n denotes the number of types of the POI data of the v and z city.

5. Results and Analysis

5.1. Urban Land Use Function Identification Results

The study area of Chengdu City has a significant proportion of residential, green space, commercial service land, and public administration and public service land use in the study area of Chengdu City (Figure 3). Further observation of the first four types of land use reveals that residential is mostly distributed in the Jinniu, Wuhou, and Qingyang districts. Public management and public service land is concentrated in Chenghua District and Jinjiang District. Commercial service land is concentrated in Qingyang District, Wuhou District, and Jinjiang District. Green space is distributed in all administrative districts and is concentrated in the outermost part of the study area.

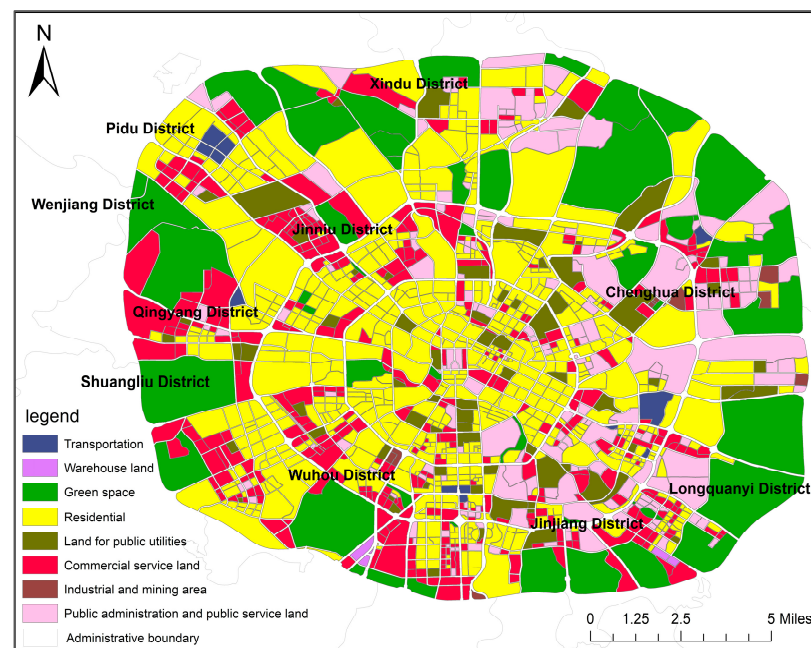


Figure 3. Results of functional identification of land use in Chengdu's central urban area.

As depicted in Figure 4, residential land is the most prevalent land type in the study area of Nanjing, evenly dispersed across all districts. Upon closer examination, it becomes apparent that, akin to commercial service land, residential, public management, and public service land are distributed across all administrative districts, while green space is primarily concentrated in Xuanwu District, Qixia District, and Pukou District.

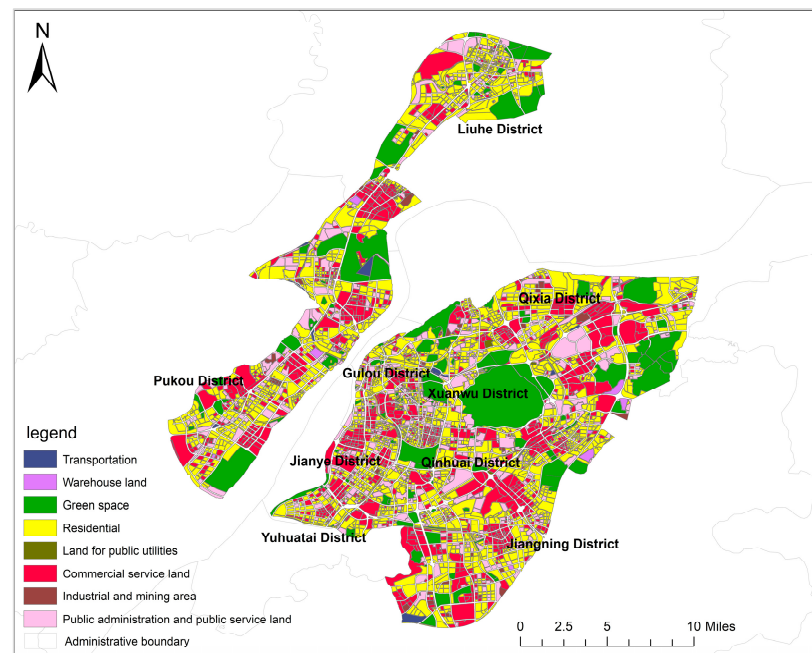


Figure 4. Results of functional identification of land use in Nanjing's central urban area.

As can be seen from Figure 5, residential land and green space occupy most of the study area in Hangzhou. In addition, most of the green space, represented by Hangzhou West Lake Scenic Spot, Wuchao Mountain National Forest Park, and Xixi National Wetland Park, is concentrated in the West Lake district, while residential land and commercial service land are concentrated in Gongshu District, Shangcheng District, and Binjiang District.

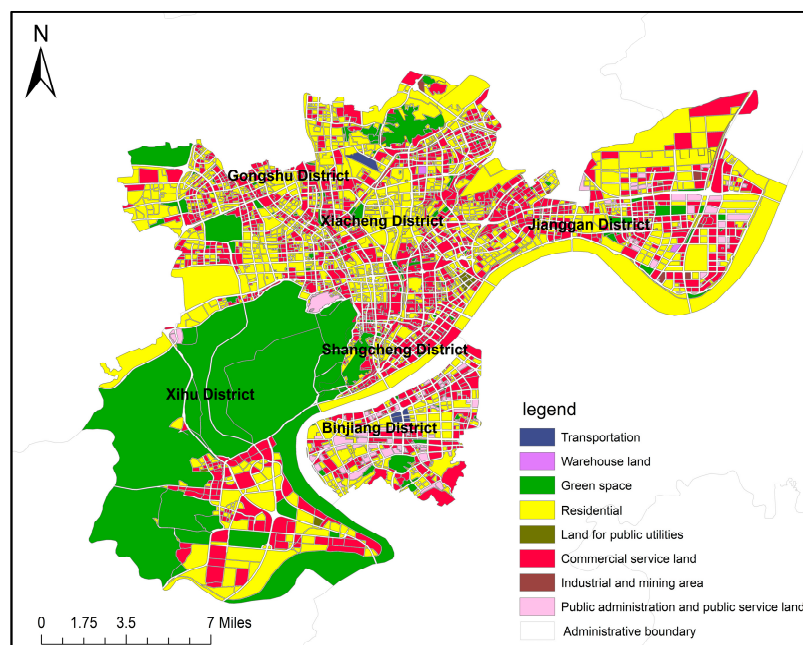


Figure 5. Results of land use function identification in Hangzhou's central urban area.

Figure 6 shows that residential land occupies most of the land in the center of Xi'an. Further observation reveals that residential land, commercial service land, and public administration and public service land are distributed in each administrative district, while green space is mainly distributed in Weiyang District, Lianhu District, Beilin District, and Baqiao District.

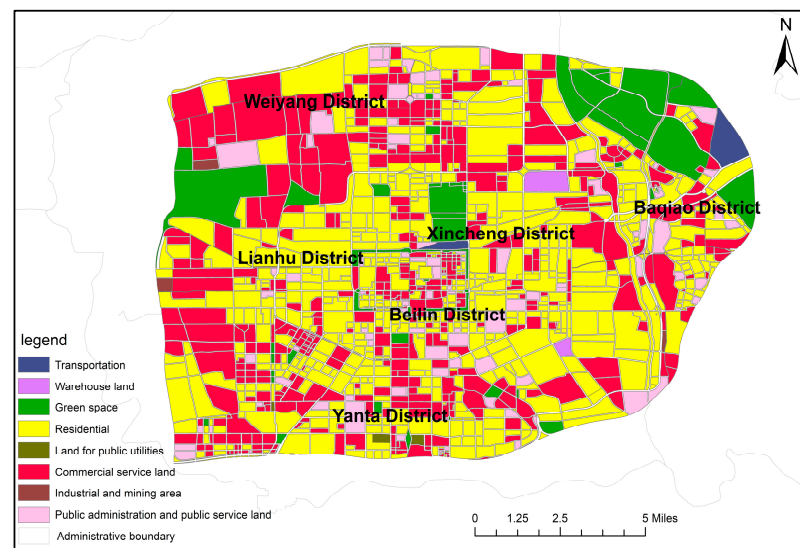


Figure 6. Results of land use function identification in Xi'an's central urban area.

As can be seen from Figure 7, residential land occupies most of downtown Tianjin. Meanwhile, residential land, commercial service land, and public management and public service land are more evenly distributed, while green space is mostly distributed in Dongli District, Nankai District, and Beichen District and concentrated at the edges of the study area. Meanwhile, the number of internal areas is relatively small.

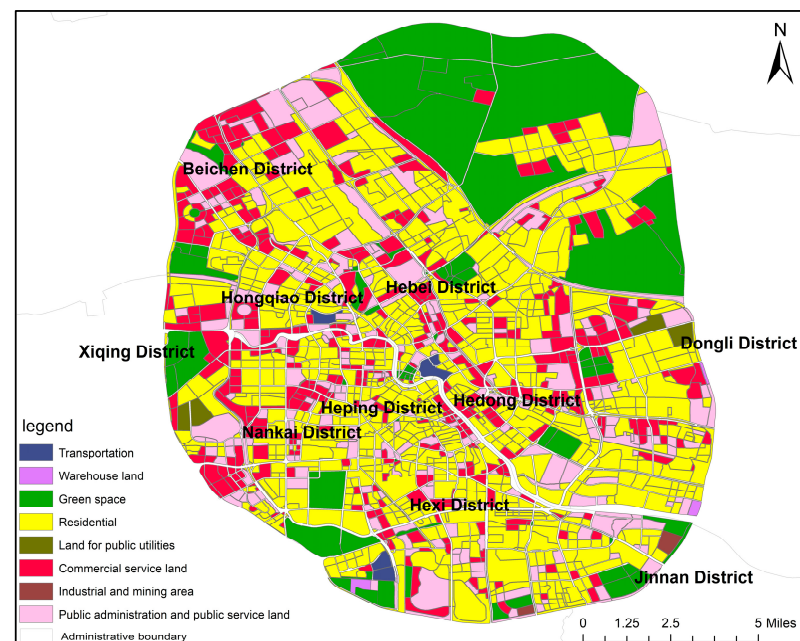


Figure 7. Results of land use function identification in Tianjin's central urban area.

5.2. Verification of the Accuracy of the Results of Site Function Identification

After using the TF-IDF machine algorithm on the OSM data and POI data of the five new first-tier cities, their respective land use function identification results are obtained. To verify the accuracy of the identification results, 30 plots are indiscriminately picked for each land use function to establish the confusion matrix. The Kappa coefficients of the confusion matrices of the five new first-tier cities are 0.68, 0.70, 0.73, 0.73, and 0.72, respectively (Table 6). They are in the range of 0.61 to 0.80, which means that the results of

the identification results show high consistency in comparison with the actual situation of the cities and can be continued to carry out the subsequent analysis and evaluation.

Table 6. Overall accuracy of the five new first-tier cities and the Kappa coefficient.

City	Overall Accuracy (%)	Kappa
Chengdu	78.00	0.72
Xi'an	74.37	0.68
Tianjin	76.00	0.70
Hangzhou	78.67	0.73
Nanjing	78.67	0.73

5.3. Comprehensive Analysis of Urban Functional Land Use Elements

The statistical results of land use function identification for eight types of land parcels in the central urban areas of five cities are as follows:

From Figure 8a, in the research area of Chengdu, among the 1420 plot units and a total area of 485.02 km², residential land has the highest number of plot units, with 685 (185.08 km²). The commercial land plot is second, with 387 (72.46 km²). Additionally, there are 199 (64.64 km²) and 49 (128.36 km²) plots of land for public management and public services, as well as for green spaces and open spaces.

As described in Figure 8b, in the research area of Nanjing, among the 5289 plot units and a total area of 649.92 km², residential land has the highest number of plot units, with 2691 (251.78 km²). The commercial land plot is second, with 1368 (153.55 km²). Furthermore, there are 944 (90.98 km²) and 165 (140.61 km²) plots of land for public management and public services, as well as for green spaces and open spaces.

As shown in Figure 8c, among the 3959 plot units and a total area of 541.30 km² in the Hangzhou research area, residential land has the highest number of plot units, with 2086 (212.42 km²). The commercial land parcels are second, with 1499 (122.75 km²). There are 166 (190.15 km²) and 145 (12.38 km²) plots of land for green spaces and open spaces, as well as for public management and services.

In Figure 8d, among the 1396 plot units and a total area of 335.87 km² in the research area of Xi'an, residential land has the highest number of plot units, with 738 (163.76 km²). The commercial land parcels come in second place, with 443 (109.21 km²). There are 104 (20.32 km²) and 66 (36.05 km²) plots of land for public management and public services, as well as for green spaces and open spaces.

As illustrated in Figure 8e, among the 1920 plot units and a total area of 374.23 km² in the Tianjin research area, residential land has the highest number of plot units, with 1070 (173.01 km²). The commercial land parcels come in second place, with 439 (57.68 km²). There are 280 (53.33 km²) and 74 (84.43 km²) plots of land for public management and public services, as well as for green spaces and open spaces. In addition, the number of land parcels for public facilities, industrial and mining, warehousing, and transportation in the five cities is relatively small and decreasing in sequence.

Comparing the number of plots and area of the eight different land use types in these five new-tier cities, it is found that their common characteristics, which are the total number of plots and area of the four land use types, namely, residential, commercial service land, public management and public service land, and green space, account for more than 90%, which is also in line with the functions that are carried by the downtown of the cities (Figure 9). Looking at the ratio of the number of plots and area of these four types of land in the five new first-tier cities, it is clear that residential land is the first land type in the five cities in terms of the number of plots and area of land. As for the proportion of land, the situation varies from city to city, but in general, it can be seen that the commercial service land and green space are larger, followed by the public management and public service land.

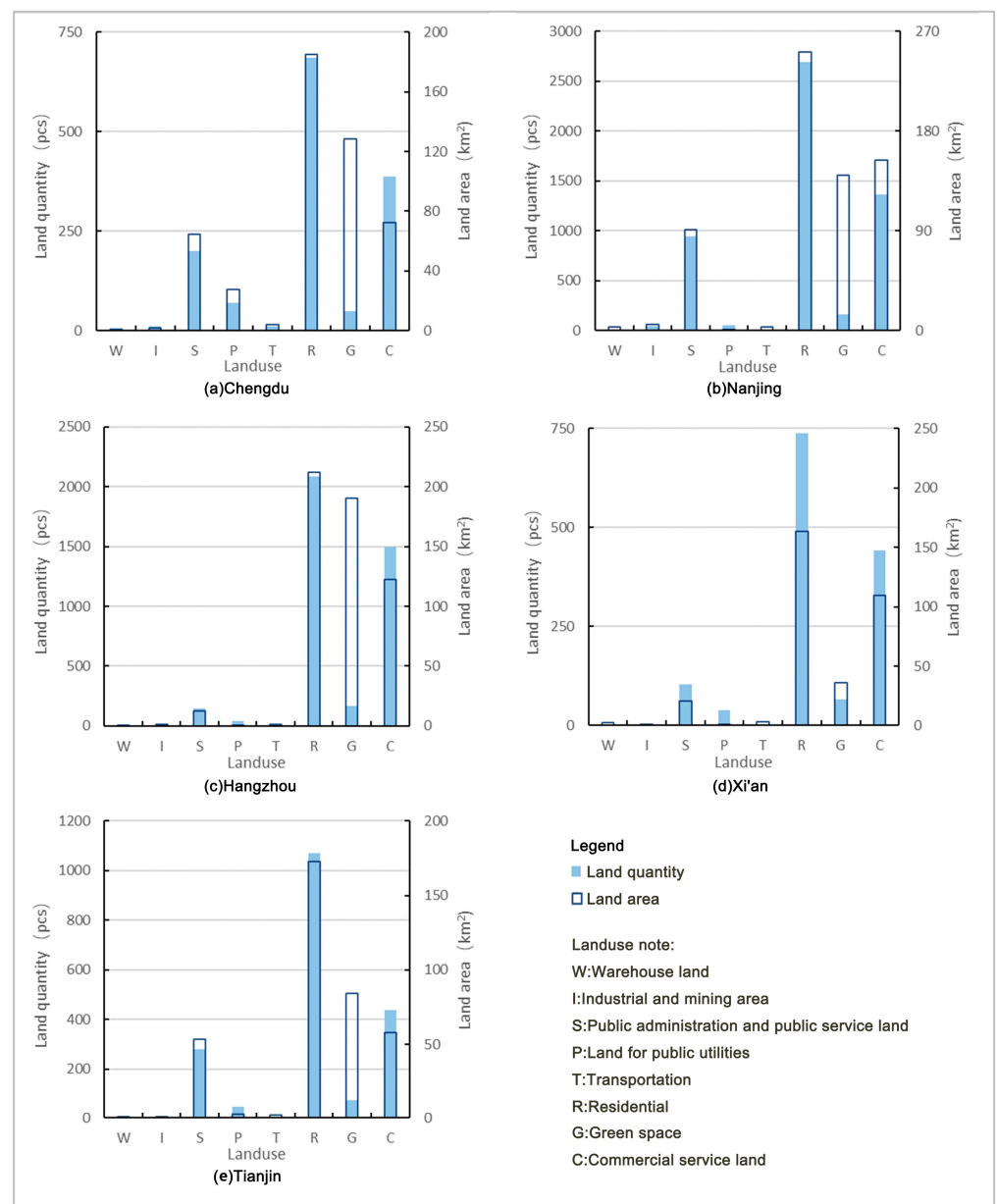


Figure 8. Quantity and area of land use in five new first-tier cities.

5.4. Comprehensive Analysis of the Functional Mix of Urban Land Use

In Figure 10a, the mixed degree of functional land in the study area of Chengdu is relatively high. The land use units with mixed degrees of “high” and “higher” account for nearly 50% of the study area, mainly concentrated in Wuhou District, Qingyang District, Jinniu District, Chenghua District, and Jinjiang District, which are part of the five districts in the city, as well as Xindu District and Pidun District. Due to the policy of constructing urban forest parks, the land mix in Longquanyi District is relatively low. The plot units with “low” and “lower” mixing degrees account for about 20% of the study area and are relatively scattered, mainly carrying the functions of residential land.

In Figure 10b, the overall land use function mixing degree in Nanjing is relatively low, with only about 20% of the study area being occupied by land units with “high” and “higher” mixing degrees, mainly concentrated in Gulou District and Qinhuai District. The land mix of land units in Xuanwu District, Pukou District, Qixia District, and Liuhe District is mostly “low” and “average”, mainly carrying green space and open space land and

residential land, such as Zhongshan Scenic Area, Xuanwu Lake Park, Pearl Spring Scenic Area, Niushoushan Tourism Area, etc.

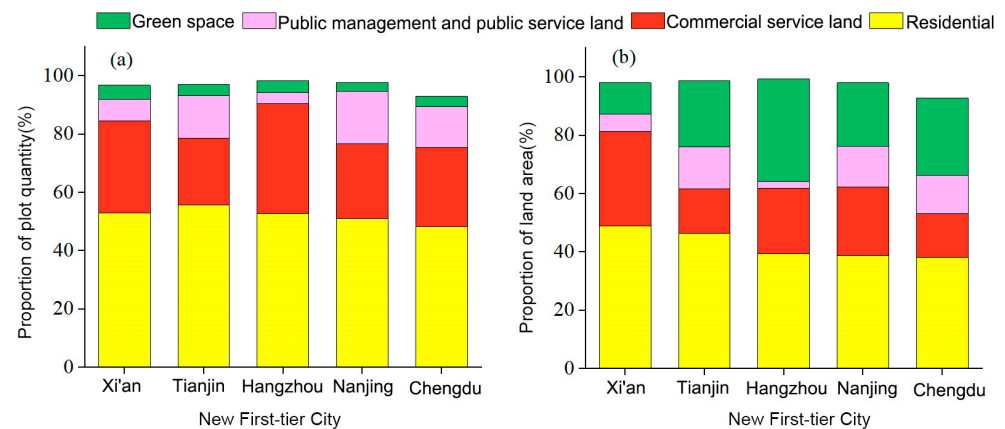


Figure 9. Number and size of land parcels for major land use functions in new first-tier cities. Note: (a) represents the proportion of land parcels with main land use functions in new first-tier cities; (b) represents the proportion of land area for the main land use functions in new first-tier cities.

In Figure 10c, the overall land mix in Hangzhou City is not high, especially in the West Lake District, as well as the riverside areas of Shangcheng District and Jianggan District. The vast majority of the West Lake area is the West Lake Scenic Area, while the remaining riverside areas follow the ecological protection policies of Hangzhou City, and their land use functions are relatively single. The north side of the research area in Hangzhou basically only carries residential and commercial service functions, so the land mix is also relatively low.

In Figure 10d, plot units with “high” and “higher” mixing degrees occupy nearly one-third of the research area in Xi’an and are mostly distributed in Lianhu District, Xincheng District, and Beilin District within the Second Ring Road. These areas mainly carry urban functions such as residential, commercial services, public management, and public services. In the research area of Xi’an City, most of the areas with “low” and “average” mixing degrees are distributed in the West Third Ring Road, where the development level is relatively backward and the functions carried by the plot units are relatively few. In addition, the land mix in the Chanba Ecological Area on the east side of Baqiao District is relatively low.

In Figure 10e, the degree of land use mixing in the study area of Tianjin shows a clear dispersion pattern. The old urban areas of Tianjin, Hexi District, Hedong District, Hebei District, Heping District, Nankai District, and Hongqiao District, have a reasonable planning layout, with land units of different degrees of mixing arranged in a staggered manner. While carrying land functions such as residential, commercial services, public management, and public services, they meet the living needs of citizens. It will not cause excessive concentration of land use functions, leading to traffic congestion problems. In addition, the plots with a low degree of mixing are mostly located in the northern expansion section of the outer ring road. Due to their low level of development, they mainly carry residential and green space functions, and the overall functional land mixing is the lowest.

In Figure 11, the similarities in land use function mix among the five new first-tier cities are as follows: (1) In the proportion of the number of land plots with varying degrees of land mix to the land area, the number of land plots with a high degree of land mix to the land area is mostly the smallest part. (2) Chengdu, Tianjin, and Xi’an show a clear pattern of dispersion in the functional mix of the five classes of land use, among which Chengdu and Tianjin have a higher overall functional mix of land use, and Xi’an has a lower overall functional mix of land use. (3) Nanjing and Hangzhou both have a low degree of functional mixing due to a relatively significant amount of scenic spots and cultural heritage sites in

their urban study areas. (4) The plot units with a high and higher degree of mixing only account for about 20% of the study area.

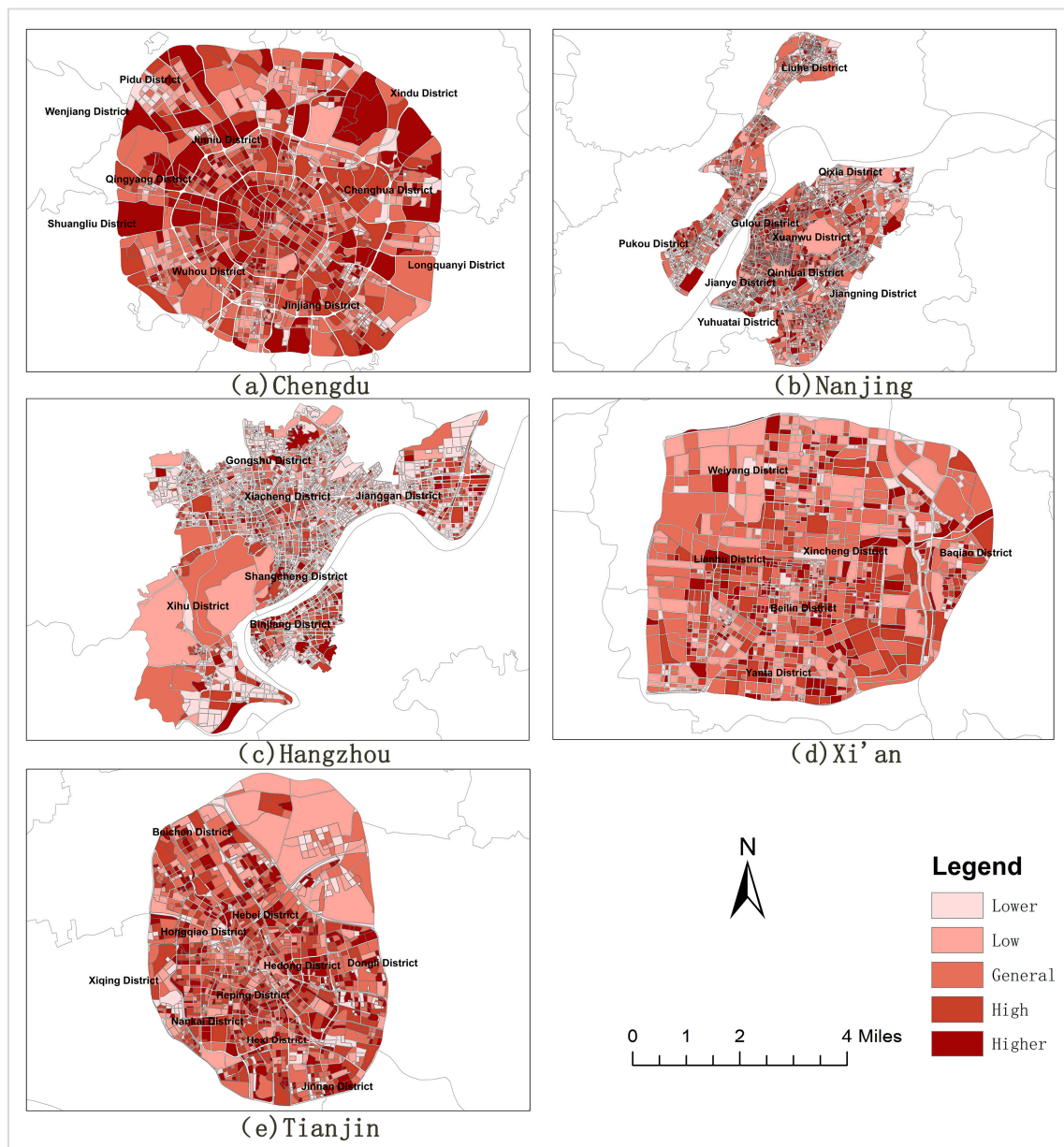


Figure 10. Land use mixing degree in five new first-tier cities.

5.5. Similarity Analysis of Urban Land Use Structure

From Table 7 and Figure 12, we can see that in the downtown of the five new first-tier cities, Hangzhou and Nanjing have the highest degree of similarity in the layout of the functional structure of urban land use. Xi'an and Tianjin have the second highest degree of similarity, and Chengdu and Nanjing have the lowest degree of similarity in the layout of functional structure of urban land use.

In addition, the similarity of the functional structure of land use between these five new first-tier cities, from high to low, is as follows: Hangzhou–Nanjing, Tianjin–Xi'an, Tianjin–Nanjing, Xi'an–Hangzhou, Xi'an–Nanjing, Chengdu–Xi'an, Tianjin–Hangzhou, Tianjin–Chengdu, Chengdu–Hangzhou, Chengdu–Nanjing.

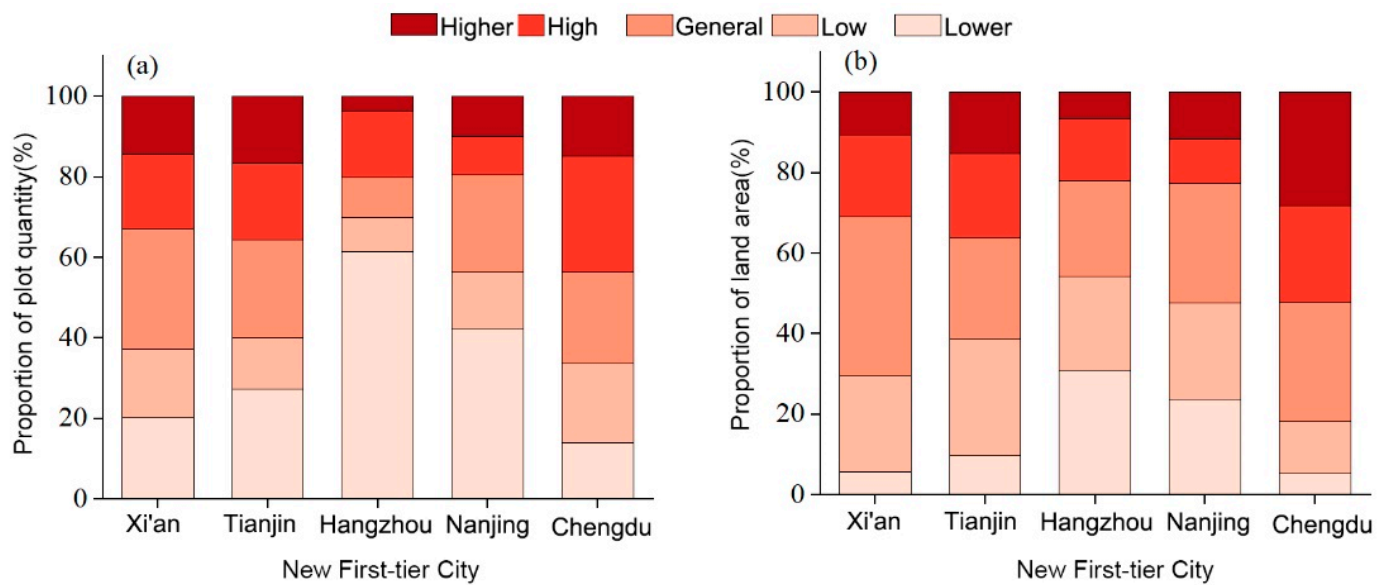


Figure 11. Number and Size of plots with different degrees of functional mix in new first-tier cities. Note: (a) represents the proportion of land parcels with varying degrees of functional mix in new first-tier cities; (b) represents the proportion of land use area in new first-tier cities with varying degrees of functional mix.

Table 7. Similarity of land use functional structure among the five new first-tier cities.

Cities	Chengdu	Xi'an	Tianjin	Hangzhou	Nanjing
Chengdu	-	-	-	-	-
Xi'an	0.59	-	-	-	-
Tianjin	0.36	0.76	-	-	-
Hangzhou	0.29	0.72	0.56	-	-
Nanjing	0.19	0.66	0.74	0.81	-

As the two cities with the highest structural similarity, Hangzhou and Nanjing exhibit a close resemblance in their POI data, particularly under the detailed category classification. The number of POI data in the two cities is very similar, contributing to their higher structural similarity in the final calculation result. Conversely, the structural similarity between the remaining cities diminishes as the proximity of the type and number of POI data diverges. Consequently, the evaluation index of urban land using function similarity can effectively illustrate the degree of similarity in land use functions between cities in terms of quantity and type, thereby facilitating a comprehensive exploration of the general pattern of land use function structure in the five new first-tier cities. As the four land use types occupy the majority of land use in the study area of the five new first-tier cities, residential land is similar in quantity and occupies the largest area, and commercial service land is similar in quantity, but there are differences in types. Green space is similar in type but different in number. Public management and public services land are similar in number and type.

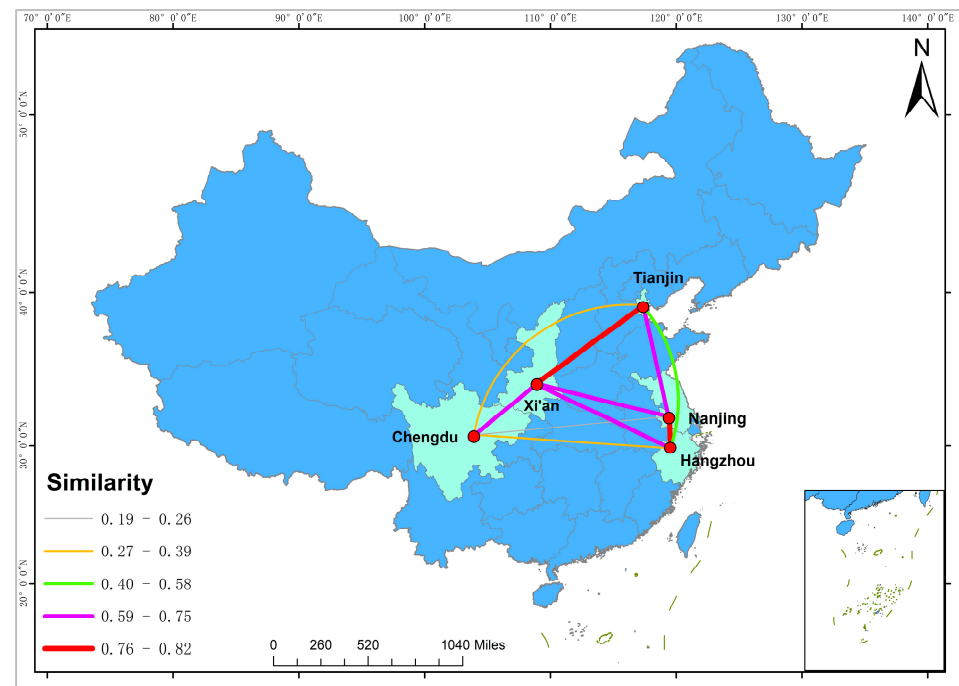


Figure 12. Similarity of land use structure in five new first-tier cities.

6. Discussion

6.1. Feasibility Analysis of Methods

Previous studies have demonstrated the advantages of multi-source big data in urban land use research and refinement [42]. Among these, POI data represents urban functions in daily life in spatial point form, encompassing spatial information such as longitude, latitude coordinates, and addresses of geographic entities, along with various attribute information like main categories, administrative divisions, and names. Compared to other big data sources, POI data is closely related to land use types, has fewer privacy concerns, is easily obtainable, and relatively straightforward to process. It significantly reduces the cost of preliminary research, effectively reflects urban spatial structure, and finds wide application in the identification and analysis of urban functional areas [11,12,43]. For a more accurate analysis of urban functional zoning, a need for refined research units arises. Li et al. used grid data to generate research units to identify functional areas in the central urban area of Wuhan, which is a simple and fast division method [33]. However, urban blocks typically have irregular polygonal shapes, and using grid data may not accurately reflect the actual block situation. OSM (OpenStreetMap) road network data, with its high positioning accuracy and basic spatial information, including longitude, latitude, road types, and names, provide a more realistic division of block units, making it a valuable data source in land use analysis [44–46]. This study, based on OSM and POI data, breaks through the limitations of traditional static data by utilizing the TF-IDF quantitative analysis method in machine learning to identify urban land use functions. This approach reduces manual intervention and provides a more scientific and objective reflection of urban land use. Comparison with previous studies reveals relatively consistent identification results for land use functions in these five cities. For example, Yu et al. identified the urban functional areas in Chengdu and observed that the primary functional zones were distributed in concentric circles centered around the city center. Residential land dominated the main urban areas, with mixed functional areas comprising 50% of the total area, aligning closely with the findings of this study [47]; Ding et al. identified the functional areas in the central urban area of Nanjing and discovered that residential land comprised 42.3% of the study area, a slight variance from the 50.88% residential land in this study. This difference may be attributed to the broader scope of our study area, leading to a greater diversity of

functional elements and a larger proportion of various land uses [48]. Wu et al. developed an urban functional area identification model using the kernel density estimation method and delineated six single/mixed functional areas within Tianjin City. The distribution of different land types exhibits a gradually dispersing pattern from the city center to the periphery, which is largely in line with the findings of this work [49]. In conclusion, the availability of OSM and POI data makes this method easily applicable to other cities.

6.2. Current Situation Analysis

The study found that residential land in five cities occupies most of the major urban areas. This is because the urbanization level of these cities is relatively high, and a large number of rural populations are gathering in the central urban areas, leading to an increasing demand for urban residential land. In addition, due to historical reasons, Hangzhou, Nanjing, and Chengdu implemented the plan of evacuating residential functions to the periphery earlier in the process of urban development. Therefore, their proportion of residential land is lower than that of the central urban areas of Xi'an and Tianjin. For the element of green space and open space land, as Hangzhou is located along the southern edge of the Yangtze River Delta and the Qiantang River Basin with abundant water and forest resources, it not only has natural landscapes such as West Lake and Xixi Wetland, but its unique natural endowment also facilitates the creation and development of various green spaces and parks between cities, resulting in the highest proportion of green space and open space land. Although Xi'an, as one of the four ancient capitals of China, has cultural relics such as the Daming Palace and the Weiyang Palace, its complex terrain and variable climate result in the lowest proportion of green space and open space land.

By integrating the findings of land use function identification, it was observed that the majority of land units with low functional diversity in the study area of the five new first-tier cities are characterized by diverse urban landscapes. This can be attributed to two potential reasons: (1) These land units may contain buildings and facilities with singular functions, such as commercial, residential, or industrial areas. The lack of comprehensive planning and design results in a singular and less dynamic urban landscape, contributing to the low degree of functional diversity in these land units; (2) Constraints imposed by land ownership and planning management. Divergent interests and planning objectives among different landowners and planning managers may lead to the fragmentation and decentralization of internal functions within land units.

6.3. Shortcomings and Prospects

First of all, there is still room for further improvement in the processing of POI data. From the initial deletion processing to the subsequent weight assignment, there are certain human influence factors. To address this problem, deep learning algorithms can be used to further reduce POI data to ensure its authenticity and objectivity and minimize the occurrence of duplicate superimposition. OSM has low density and a lack of data in suburban and rural areas, resulting in excessive division of functional units and poor local recognition results.

Secondly, when analyzing and evaluating the five new first-tier cities, there is potential to enrich the indicators used, particularly for comprehensive comparative analysis among these cities. In the future, it would be beneficial to include indicators such as location entropy and nearest-neighbor distance to further assess the current status of urban land use functions and explore their patterns. This approach will enable the formulation of more objective and comprehensive problem assessments and optimization suggestions.

Thirdly, when selecting research subjects, for a more extensive exploration of the current state of urbanization in China, it would be advantageous to encompass a broader range of cities for a more comprehensive analysis. This broader perspective will allow for an exploration of the current state of urbanization in China from a more macroscopic viewpoint.

Finally, this study identified certain similarities in their land use functions. However, due to the different policies and historical conditions implemented in different cities, it is difficult to promote this land use feature throughout China and even the world. Future research could involve a comprehensive evaluation of various land use functional layouts to determine the optimal layout for urban planning.

7. Conclusions

Identification of central urban land use functions in the five new first-tier cities involves processing OSM and POI data for each city and utilizing the TF-IDF machine algorithm for function identification. Influenced by geographical conditions, resource endowments, and planning policies, differences in the functional layout of land use are observed among these cities. However, common characteristics among the five cities include residential land, commercial service land, public management and public service land, green land, and open land, with the total number and area of these land types accounting for over 90%.

The accuracy verification of central urban land use function identification results of the five new first-tier cities involved constructing a confusion matrix to test the accuracy of land use function identification. The Kappa coefficients, falling within the range of [0.61, 0.80], indicate a high level of consistency in accuracy evaluation. This further verifies the feasibility of urban land function identification based on multi-source big data. Most of the plot units with low functional mixing in the research area of the five new first-tier cities are distributed with various urban landscapes.

The evaluation of central urban land use in the five new first-tier cities involves assessing the present situation of urban land use function through land use mix degree and urban structure similarity degree. Chengdu and Tianjin exhibit the highest land use function mixing degree, followed by Xi'an, Nanjing, and Hangzhou. In Chengdu, Tianjin, and Xi'an, the mixing degree of five levels of land use functions shows an obvious dispersion pattern, while Nanjing and Hangzhou have high mixed land use function plots concentrated in several administrative regions. According to the calculation results of urban structure similarity, Hangzhou and Nanjing have the highest similarity of land use function structure layout, followed by Xi'an and Nanjing, with the lowest similarity observed between Chengdu and Nanjing. The similarity of land use function structure layout is higher when cities are closer in terms of the type and quantity of land use functions.

Author Contributions: Conceptualization, methodology, and formal analysis, W.Y.; draft and funding support, Y.Y.; visualization and investigation, B.F.; resources and review, S.L.; supervision, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National University Student Innovation & Entrepreneurship Development Program of China (grant number: 202310626023), the National Natural Science Foundation of China (NSFC) project (grant number: 42305178), Sichuan Tianfu New Area Rural Revitalization Research Institute project of 'Announce and be in command' (grant number: XZY1-14), and China Telecom & Sichuan Agricultural University Smart Agricultural Innovation Laboratory project of 'Announce and be in command'.

Data Availability Statement: Data are available from the corresponding authors upon request.

Acknowledgments: Thanks to the anonymous experts for their suggestions. This work was supported by the National University Student Innovation & Entrepreneurship Development Program of China (202310626023).

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

References

1. Ni, P.F.; Xu, D.H. China's Urbanization for 2035. *Reform* **2022**, *8*, 98–110.
2. Yang, X.; Lu, X. Driving factors of urban land urbanization in China from the perspective of spatial effects. *China Popul. Resour. Environ.* **2021**, *31*, 156–164.
3. Yang, Y. A Study on Identification of Urban Functional Zones and Evolution of Spatial Structure—A Case of Changchun City. Ph.D. Dissertation, Jilin University, Changchun, China, 2022.

4. Pan, C.; Wu, S.; Li, E.; Li, H.; Liu, X. Identification of urban functional zones in Macau Peninsula based on POI data and remote information sensors technology for sustainable development. *Phys. Chem. Earth Parts A/B/C* **2023**, *131*, 103447. [\[CrossRef\]](#)
5. Wang, Q.; Liu, X.; Li, Y. Spatial Structure, City Size and Innovation Performance of Chinese Cities. *China Ind. Econ.* **2021**, *398*, 114–132.
6. Qiu, B.; Ye, M.; Ning, K. Twelve Inflection Points Focusing on the Second Half of Urbanization in China. *Urban Dev. Stud.* **2021**, *9*, 1006–3862.
7. Wang, K. Green Transformation and Development of China's Urbanization. *City Plan. Rev.* **2021**, *36*, 9–16.
8. Dan, X. Research on Functional Area Division and Spatial Distribution of Chengdu Based on Multi-source Data. Master's Thesis, Yunnan University, Kunming, China, 2021.
9. Xue, B.; Xiao, X.; Li, J.; Zhao, B.; Fu, B. Multi-source Data-driven Identification of Urban Functional Areas: A Case of Shenyang, China. *Chin. Geogr. Sci.* **2022**, *33*, 21–35. [\[CrossRef\]](#)
10. Frias-Martinez, V.; Frias-Martinez, E. Spectral clustering for sensing urban land use using Twitter activity. *Eng. Appl. Artif. Intell.* **2014**, *35*, 237–245. [\[CrossRef\]](#)
11. Jiang, S.; Alves, A.; Rodrigues, F.; Ferreira, J., Jr.; Pereira, F.C. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Comput. Environ. Urban Syst.* **2015**, *53*, 36–46. [\[CrossRef\]](#)
12. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping Urban Land Use by Using Landsat Images and Open Social Data. *Remote Sens.* **2016**, *8*, 151. [\[CrossRef\]](#)
13. Ye, C.; Zhang, F.; Mu, L.; Gao, Y.; Liu, Y. Urban function recognition by integrating social media and street-level imagery. *Environ. Plan. B Urban Anal. City Sci.* **2020**, *48*, 1430–1444. [\[CrossRef\]](#)
14. Liu, X.; Tian, Y.; Zhang, X.; Wan, Z. Identification of Urban Functional Regions in Chengdu Based on Taxi Trajectory Time Series Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 158. [\[CrossRef\]](#)
15. Sun, Y.; Yin, H.; Wen, J.; Sun, Z. Urban Region Function Mining Service Based on Social Media Text Analysis. *Int. J. Softw. Eng. Knowl. Eng.* **2021**, *31*, 563–586. [\[CrossRef\]](#)
16. Huang, X.; Yang, J.; Li, J.; Wen, D. Urban functional zone mapping by integrating high spatial resolution nighttime light and daytime multi-view imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 403–415. [\[CrossRef\]](#)
17. Liu, B.; Deng, Y.; Li, M.; Yang, J.; Liu, T. Classification Schemes and Identification Methods for Urban Functional Zone: A Review of Recent Papers. *Appl. Sci.* **2021**, *11*, 9968. [\[CrossRef\]](#)
18. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering Urban Functional Zones Using Latent Activity Trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [\[CrossRef\]](#)
19. Feng, R.; Dong, X.; Liang, T.; Wang, L. Identification of Urban Functional Areas Based on Logistic Regression Model. *Geomat. Spat. Inf. Technol.* **2018**, *41*, 109–112.
20. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.-R.; Gu, C. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [\[CrossRef\]](#)
21. Shicheng, Y.; Jingli, W.; Chengming, L.; Zhaoxin, D.; Hao, L. Urban Functional Area Division Considering POI and Land Use Data. *J. Geomat. Sci. Technol.* **2021**, *38*, 181–186.
22. Zhihao, S.; Hongzan, J.; Hao, W.; Zhenghong, P.; Lingbo, L. Block2vec: An approach for identifying urban functional regions by integrating sentence embedding model and points of interest. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 339.
23. Lai, G.; Zhao, G.; Yang, M. Identification of Urban Functional Areas and Measurement of Mixing Degree Based on Classification Modeling and TF-IDF Algorithm. *Geomat. Spat. Inf. Technol.* **2023**, *46*, 89–93.
24. Yang, M.; Cao, S.; Zhang, H.; Wu, S.; Zhang, D. Using Point of Interest Data and Satellite Observation for Urban Functional Zone Mapping. *Sens. Mater.* **2023**, *35*, 269–284. [\[CrossRef\]](#)
25. Miao, R.; Wang, Y.; Li, S. Analyzing Urban Spatial Patterns and Functional Zones Using Sina Weibo POI Data: A Case Study of Beijing. *Sustainability* **2021**, *13*, 647. [\[CrossRef\]](#)
26. Li, L.; Ma, S.J.; Wang, Y.P. Smart city and public cultural resources demand based on big data: A case study of Tianjin Library. *Int. J. Innov. Manag. Technol.* **2020**, *11*, 57–62. [\[CrossRef\]](#)
27. Gao, X. Identification and Optimization of Xi'an Urban Land Use Function Based on POI Data. Master's Thesis, Northwest University, Xi'an, China, 2021.
28. Cheng, Y.; Zhao, M. On the Land-use Classification System in the Context of Territorial Spatial Planning: Classification Structure and Application Rationale. *Urban Plan. Forum* **2021**, *4*, 367–377.
29. Bao, H.; Ming, D.; Guo, Y.; Zhang, K.; Zhou, K.; Du, S. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data. *Remote Sens.* **2020**, *12*, 1088. [\[CrossRef\]](#)
30. Renato, A.; Ana, A.; Carlos, B. POI mining for land use classification: A case study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 493.
31. Luo, G.; Ye, J.; Wang, J.; Wei, Y. Urban Functional Zone Classification Based on POI Data and Machine Learning. *Sustainability* **2023**, *15*, 4631. [\[CrossRef\]](#)
32. Zhao, W.; Li, Q.; Li, B. Extracting hierarchical landmarks from urban POI data. *J. Remote Sens.* **2011**, *15*, 973–988.
33. Qiang, L.; Xinqi, Z.; Yi, C. Research on function identification and distribution characteristics of Wuhan supported by big data. *Sci. Surv. Mapp.* **2020**, *45*, 119–125.
34. Zhuo, H.W.; Dong, W.; Qing, L.I. Keyword Extraction from Scientific Research Projects Based on SRP-TF-IDF. *Chin. J. Electron.* **2021**, *30*, 652–657. [\[CrossRef\]](#)

35. Zhou, L.; Liu, M.; Zheng, Z.; Wang, W. Quantification of Spatial Association between Commercial and Residential Spaces in Beijing Using Urban Big Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 249. [[CrossRef](#)]
36. Wei, X.; Hu, M.J.; Wang, X.J. The differences and influence factors in extracting urban green space from various resolutions of data: The perspective of blocks. *Remote Sens.* **2023**, *15*, 1261. [[CrossRef](#)]
37. Lv, Q.; Lin, Y.; Zhao, J.; Qin, B. The correlation between the prediction of land mixing degree and the evolution of land mixed use in Kunming. *J. Lanzhou Univ. Nat. Sci.* **2022**, *58*, 548–559.
38. Liu, X.; Long, Y. Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest. *Environ. Plan. B* **2015**, *43*, 498–510.
39. Salton, G.; Wong, A.; Yang, C.S. A vector space model for automatic indexing. *Commun. ACM* **1975**, *18*, 613–620. [[CrossRef](#)]
40. Yan, L.; Ge, L.N.; Hu, Z. Attribute-based multi-keyword ranking search scheme based on blockchain. *Appl. Res. Comput.* **2023**, *40*, 1952–1956+1963.
41. Li, T.; Zheng, X.; Zhang, C.; Wang, R.; Liu, J. Mining Spatial Correlation Patterns of the Urban Functional Areas in Urban Agglomeration: A Case Study of Four Typical Urban Agglomerations in China. *Land* **2022**, *11*, 870. [[CrossRef](#)]
42. Kuang, W.H. Spatio-temporal patterns of intra-urban land use change in Beijing, China between 1984 and 2008. *Chin. Geogr. Sci.* **2012**, *22*, 210–220. [[CrossRef](#)]
43. Krause, C.M.; Zhang, L. Short-term travel behavior prediction with GPS, land use, and point of interest data. *Transp. Res. Part B Methodol.* **2019**, *123*, 349–361. [[CrossRef](#)]
44. Kulawiak, M.; Dawidowicz, A.; Pacholczyk, M.E. Analysis of server-side and client-side Web-GIS data processing methods on the example of JTS and JSTS using open data from OSM and geoportal. *Comput. Geosci.* **2019**, *129*, 26–37. [[CrossRef](#)]
45. Brovelli, M.A.; Zamboni, G. A New Method for the Assessment of Spatial Accuracy and Completeness of OpenStreetMap Building Footprints. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 289. [[CrossRef](#)]
46. Jokar Arsanjani, J.; Mooney, P.; Zipf, A.; Schauss, A. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In *OpenStreetMap in GIScience: Experiences, Research, and Applications*; Springer: Cham, Switzerland, 2015; pp. 37–58.
47. Yu, B.; Wang, Z.; Mu, H.; Sun, L.; Hu, F. Identification of Urban Functional Regions Based on Floating Car Track Data and POI Data. *Sustainability* **2019**, *11*, 6541. [[CrossRef](#)]
48. Ding, Y.W.; Xu, H.W.; Wang, C.H. Research on Urban Functional Area Recognition Integrating OSM Road Network and POI Data. *Geogr. Geo-Inf. Sci.* **2020**, *36*, 57–63.
49. Wu, K.P. *Research on Spatial Structure Identification and Distribution Characteristics of Urban Functional Areas Based on POI Data*; Tianjin Chengjian University: Tianjin, China, 2023.

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