



OPEN ACCESS

EDITED BY

Tao Liu,
Peking University, China

REVIEWED BY

Mohamed Salah Ezz,
Associate Professor of Architecture,
Saudi Arabia
Ting Wang,
Beijing University of Civil Engineering and
Architecture, China

*CORRESPONDENCE

Linshen Wang,
✉ cea_wangls@ujn.edu.cn

RECEIVED 09 June 2025

ACCEPTED 01 October 2025

PUBLISHED 17 October 2025

CITATION

Zhao J, Jiang Y, Zhang X, Ye Q, Zhao Q, Wu X
and Wang L (2025) From data to decision:
empirical application of machine learning in
public space planning along the Grand Canal,
Shandong Province, China.
Front. Built Environ. 11:1643104.
doi: 10.3389/fbuil.2025.1643104

COPYRIGHT

© 2025 Zhao, Jiang, Zhang, Ye, Zhao, Wu and
Wang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

From data to decision: empirical application of machine learning in public space planning along the Grand Canal, Shandong Province, China

Jing Zhao¹, Yuan Jiang¹, Xiuhua Zhang², Qing Ye³, Qiang Zhao⁴,
Xianhua Wu¹ and Linshen Wang^{1*}

¹School of Civil Engineering and Architecture, University of Jinan, Jinan, China, ²Department of Architecture and Urban Planning, Shandong Urban Construction Vocational College, Jinan, China,

³School of Architecture and Art Design, Hebei University of Technology, Tianjin, China, ⁴Tianjin Municipal Bureau of Planning and Natural Resources, Tianjin, China

Introduction: In the process of urbanization, public space plays an increasingly important role in improving the livability and sustainability of cities. However, effectively understanding the preferences of different groups for public space and conducting reasonable planning integrated with environmental and infrastructure elements remains a challenge in urban planning. This is because traditional planning methods often fail to fully capture the detailed behavior of residents. Therefore, the purpose of this study was to explore the empirical application of machine learning technology to public space planning along the Grand Canal in Shandong Province (China), analyze the behavior patterns and preferences of residents regarding different public spaces, and thereby provide support for data - driven public space planning.

Methods: Based on survey data from 1008 respondents across 4 cities, this study employed machine learning methods such as K - means clustering, association rule mining, and correlation analysis to investigate the relationships between visitor behavior and the environmental characteristics of public spaces.

Results: The application of these methods yielded several important results. Cluster analysis identified three distinct groups: young and middle - aged local residents with a preference for accessibility, middle - aged and elderly groups enthusiastic about cultural engagement, and diverse transportation users with mixed spatial preferences. Additionally, association rule mining uncovered strong correlations between location types and perceived attributes such as cleanliness and aesthetics. Moreover, correlation analysis indicated statistically significant positive correlations between aesthetics and cleanliness, as well as between safety and cleanliness.

Discussion: This research offers valuable data - driven insights for public space planning and management. It demonstrates that machine learning can effectively identify and quantify key factors influencing public space use. As a result, it provides more accurate policy recommendations for urban planners and ensures that public space planning better meets the needs of different groups. For urban planners, the findings can guide the optimization of facility layouts for specific groups. For instance, adding canal cultural display nodes for cultural engagement groups and improving barrier - free facilities for groups

with high accessibility needs, thereby enhancing the inclusiveness and utilization efficiency of public spaces.

KEYWORDS

machine learning, public space, clustering analysis, association rule mining, correlation analysis

1 Introduction

Urban space, especially public space, has been widely recognized as an important element for improved livability and sustainability of cities (Sletto and Palmer, 2017; Rui and Othengrafen, 2023). With the acceleration of urbanization and the increasingly complex nature of the urban environment, the demand for evidence-based planning strategies that can meet the diverse needs of urban residents is growing (Dirsehan, 2024). As an important place for social interaction, leisure, and cultural expression, public space plays a key role in improving the quality of urban life (Faka et al., 2021). Thus, the rational design and management of public space can help encourage the formation of a sense of community, enhance social cohesion, and promote personal wellbeing.

As a UNESCO World Cultural Heritage Site (Li, 2023), the Grand Canal occupies an important position in the country's cultural heritage and history. Shandong Province represents one of the most dynamic sections along the Grand Canal in China, and the Grand Canal has exerted considerable influence on the configuration and development of public spaces in cities along its route within Shandong. The various public spaces distributed along the Grand Canal, including riverside parks, green spaces, and commercial blocks (Cheng et al., 2022), constitute a microcosm reflecting the diverse needs of urban residents (Marshall et al., 2019; Randrup et al., 2021). These public spaces provide a unique opportunity to study the interactions between the environmental elements of public space and visitor behaviors in historical cities along the Grand Canal.

However, understanding the preferences of different demographic groups, and determining how those preferences interact with environmental and infrastructural elements, remains a notable challenge (Wang et al., 2020). Although traditional urban planning methods offer certain guidance, it is often difficult to fully capture the detailed and dynamic behavioral characteristics of urban residents (Motomura et al., 2022). In recent years, widespread application of machine learning techniques in various fields has presented new opportunities for urban planning (Ferreira et al., 2025; Maté-Sánchez-Val et al., 2025). Compared with traditional research methods, machine learning can process massive and multidimensional data, uncover complex relationships hidden in the data (Durowoju et al., 2025), and thus provide a more accurate and data-driven decision-making foundation for public space planning (Robi and George, 2025).

Internationally, European historical canal cities such as Amsterdam and Venice have attempted to optimize public spaces through spatial syntax and behavioral observation (Pourbahador and Brinkhuijsen, 2023; Cabigiosu, 2025), but with heavy reliance on traditional statistical methods and limited application of machine learning in mining complex patterns. In

North America, studies have emphasized machine learning in relation to traffic and safety analysis (Almukhalifi et al., 2024; Alwahedi et al., 2024; Nourbakhsh et al., 2024) but rarely with a focus on refined public space planning in heritage contexts.

This study focused on empirical application of machine learning techniques to public space planning along the Grand Canal in Shandong Province, China. Using machine learning techniques comprising a clustering algorithm, association rule mining, and correlation analysis, the latent patterns and relationships between visitor behaviors and the environmental characteristics of public spaces were investigated. By analyzing the movement patterns and preferences of the respondents, and considering the background characteristics of the space, the objective of this study was to elicit data-driven insights for optimization of public space configuration. The ultimate goal was to improve the effectiveness of public space planning, improve resource allocation, better meet the needs of different groups, and improve the efficiency and comfort of space utilization.

2 Literature review

Machine learning technology has been widely used in different fields, such as urban science (Balogun et al., 2021; Yang et al., 2023), traffic flow prediction (Aljuaydi et al., 2023; Pérez Moreno et al., 2023; Zhao et al., 2024), healthcare (Rajagopal et al., 2024; Jentzer et al., 2025), biology (Rai et al., 2024; Jin et al., 2025), archaeology (Mahmoud et al., 2025), finance (Zhang et al., 2025), and even art (Garcia-Moreno et al., 2024). In recent years, application of machine learning technology to urban public space research has increased markedly, such as in relation to intelligent management and planning, spatial perception, and healthy cities. Machine learning has proven to be a valuable tool in the analysis and design of public space (Zhu et al., 2025).

2.1 Technical applications and case progress

In recent years, application of the combination of machine learning and public space analysis to urban planning and development has gradually increased, improving the efficiency of data collection and analysis. For example, a machine learning model was used to analyze the spatial distribution of street vendors' activities in Mexico, generating a spatial classification map of their activities, which served as a basis for policymaking (Barreda Luna et al., 2022). A hybrid method was developed to assess the quality and accessibility of public spaces in marginalized communities through geographic and network analysis, laying the

groundwork for future application of machine learning models to similar urban environments (Medina et al., 2023). Similarly, a recent study proposed using the “cyber walk” method combining unsupervised learning (K-means) and a deep learning model (YOLOv5) to classify public spaces, which substantially enhanced the efficiency of data collection and analysis (Valenzuela-Levi et al., 2024). These studies highlight the growing importance of combining machine learning with public space analysis, making urban planning and development a more efficient and data-driven process.

The security of public space, especially the social perception of security, is another important area of machine learning application. In addition to traditional data sources, such as questionnaires, some studies have also added new data sources such as social media, GPS, and user-generated content, thereby strengthening the analysis of public space use (Ramírez et al., 2021). Other studies combined deep learning models (such as RankNet and FCN-8s) with geographically weighted regression methods to analyze women's perception of public space security, revealing the influence of sky visual factors and green visual factors (Chen et al., 2024). This method provides urban planners with guidance regarding an improved sense of spatial security, especially for vulnerable groups. Moreover, some studies have also explored the relationship between urban density, spatial grammar, and residents' stress perception, revealing that urban characteristics such as building density and road network design have considerable impact on residents' stress perception, and providing new ideas for creating a more sympathetic urban environment (Le et al., 2024). These earlier studies emphasized the role of machine learning in understanding the psychology and social dynamics of public space, which can help planners develop more targeted and inclusive urban design strategies.

Multisource data integration is a major trend in research of modern urban spatial management. For example, a recent study that combined geospatial technology and machine learning techniques to analyze multisource crime data in Porto (Portugal) identified crime hotspots, conducted spatiotemporal pattern mining, and proposed a crime prediction method combining machine learning and spatial analysis (Saraiva et al., 2022). Another study used Strava cycling data to analyze relationships between green space areas and active travel modes, which provided valuable insights for sustainable urban planning (Yang et al., 2024). Studies have also used big data and unsupervised machine learning technologies to analyze the cultural ecosystem services of Beijing Xiangshan Park, and to identify key areas that enhance the public space experience (Fu et al., 2025). These studies demonstrate the potential of combining multiple data sources and enhancing the scope and accuracy of urban spatial management and optimization.

Machine learning has also played an important role in the development of smart cities. Some studies considered the application of digital twin technology in urban space management, emphasizing how best to optimize space utilization and improve environmental sustainability through intelligent systems, especially in the post-epidemic era (Piras et al., 2024). Another study used a machine learning model to classify global smart cities, focusing on quality of life, technology, and sustainability indicators, which provided urban planners with practical insights on how to improve the performance of smart cities (Alkhereibi et al., 2025). These previous studies reflect the increasing importance of application of machine learning

and other advanced technologies to improving urban sustainability, which is crucial to the development of resilient and adaptive cities.

2.2 Synthesis

The application of machine learning in urban public space research has expanded from improving data collection and analysis efficiency to areas such as security perception analysis, multisource data integration (Ta and Furuya, 2022), and smart city development. These studies not only highlight the value of technology in revealing the relationship between spatial behavior and the environment, but also provide data-driven methodological support for building more inclusive and sustainable urban planning strategies. Moreover, they indicate that cross-disciplinary technology integration and multidimensional data integration are core trends for future development.

2.3 Addressing research gaps

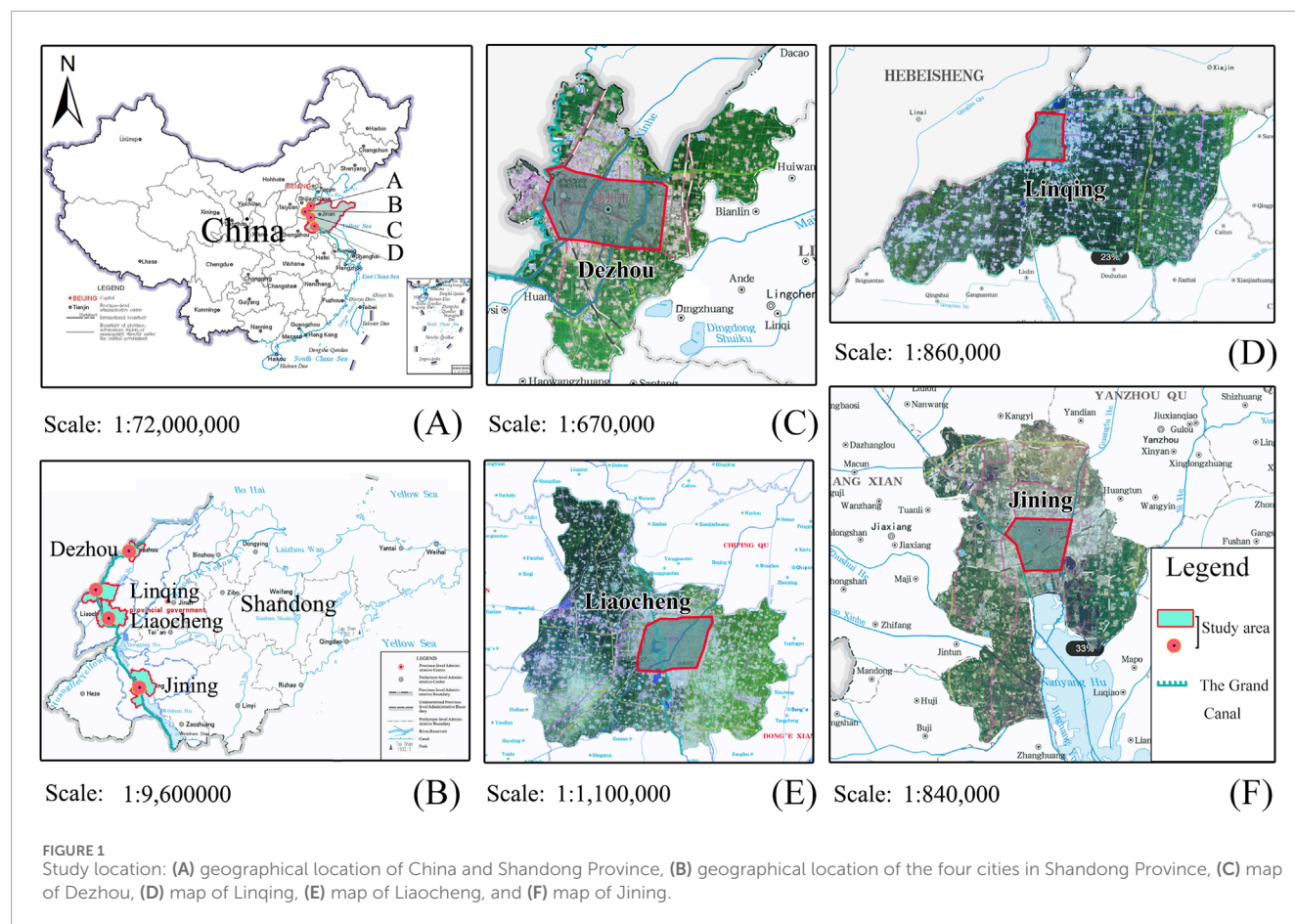
Three key gaps exist in current research. (1) Insufficient targeting of cultural heritage scenes. Most studies focus on the general urban environment or single functional space, thereby lacking systematic exploration of the Grand Canal heritage corridor that combines historical value and dynamic functions. (2) Weakness in the integration of multidimensional elements. Insufficient collaborative analysis of aesthetic values, cultural atmosphere, accessibility, and other factors makes it difficult to support refined planning of complex public spaces. (3) Lack of depth in analysis of the interaction mechanism between behavior and the environment. Traditional methods have difficulty in capturing the nonlinear correlation between dynamic behavioral characteristics and environmental factors.

This study addressed the identified knowledge gaps via three innovations. First, machine learning systems were applied for the first time to public space planning of the Grand Canal cultural heritage corridor, with a balance between historical preservation and modern functional needs. Second, expanded data dimensions improved analytical comprehensiveness, as the research integrated questionnaire data and GIS spatial data covering multidimensional information including demographic characteristics, behavioral patterns and spatial attributes. Third, integration of methods by combining K-means clustering, association rule mining, and correlation analysis, while processing behavioral data and environmental features, revealed complex relationships between multiple factors. Through the above framework, this study broke through the limitations of traditional methods and provided a new data-driven model for public space planning in cultural heritage areas.

3 Materials and methods

3.1 Study area

Along the Grand Canal in Shandong Province (China), the central cities of Dezhou, Linqing, Liaocheng, and Jining (Figure 1)



are crossed by the canal. Historically, these cities have held relatively high administrative status. Among them, Liaocheng was once the seat of a prefecture, while Dezhou, Linqing, and Jining served as secondary prefectures (Jiang et al., 2022). As listed in Table 1, these locations were selected as research subjects owing to their rich historical heritage, unique urban landscapes, and integration of waterfront public spaces. This study specifically focused on public spaces near the Grand Canal within these four cities, including riverside parks, public green spaces, and commercial districts (Figure 2). The selection criteria for these areas included accessibility, cultural importance, and diversity of urban spatial characteristics (Senik and Uzun, 2022). The research covered both highly urbanized areas and historically important regions, thereby providing a comprehensive context for analysis of the interactions between different urban environments and the preferences of residents.

3.2 Data sources and collection

This study used a combination of questionnaire survey and geospatial data. The primary data source consisted of questionnaire survey responses gathered from local residents and visitors to the public spaces of the Grand Canal. The surveys were designed to capture detailed information about respondents' preferences, behaviors, and perceptions of the public spaces. The survey was

distributed to a diverse sample of residents, ensuring representation across different demographic groups, with consideration of age, gender, and occupation.

3.2.1 Variable set composition and functional role

The data involved in the study is divided into two types, namely, questionnaire survey data and GIS spatial data. The former covers 21 indicators and is the core data of machine learning models. The latter includes four indicators, with the first three used for analysis and the last one used for validation (Table 2).

3.2.2 Stratification

Questionnaire surveys were conducted using a purposive sampling method as part of a cross-sectional study from October 2023 to April 2025 in Dezhou, Linqing, Liaocheng, and Jining. A total of 1008 individuals completed valid questionnaires, and the basic information on the respondents is presented in Figure 3. The eligibility criteria adopted were as follows: (a) being aged 18 years and above, (b) being conscious and able to communicate normally, and (c) being willing to participate in this study.

Eleven explanatory variables were identified and categorized into three groups: environmental quality, canal cultural characteristics, and supporting facilities (Joseph et al., 2016). The explanatory or predictor variables used in this research were safety, cleanliness, beauty, convenience, leisure facilities, catering facilities, entertainment facilities, tourist facilities, canal landscape

TABLE 1 Designation and codification of sample points across various locations.

Code	Sample point name	Code	Sample point name
A1	Longyun Lake Park	C4	Pearl of Water City
A2	People's Park of Dezhou	C5	Longwan Cultural Square
A3	Xinhu Scenic Area	C6	Jinfeng Square
A4	Jinembroidery Scenic Spot	C7	Rose Park
A5	Changhe Park	C8	Sports Park
A6	Dezhou Sports Park	D1	Guanghe Park
A7	Dongziyuan Scenic Spot	D2	Xianying Green Space
A8	Meihe Wetland Park Beach	D3	Tieta Temple
B1	Galaxy Park	D4	Huanbiquan Road
B2	Public Park	D5	Guiheyuan
B3	Longshan Park	D6	People's Park of Jining
B4	Linqing Canal Culture Square	D7	Kuaihuolin Riverside Green Space
B5	Linqing Theatre Square	D8	Yunhe Mall
B6	Linqing Sports Park	D9	Wanda Plaza
C1	Water City Square	D10	Huijingyuan
C2	The River Waterfront Park	D11	Yunhe Shengshi Riverside Green Space
C3	Xiaomei Riverside Park	D12	Nanchi Park

conservation, canal cultural atmosphere, and canal publicity. These variables were chosen following literature review, field observation, and discussion with local experts (Wilson and Kelly, 2011; Cao et al., 2019; Manta et al., 2019).

Safety is one of the basic elements reflecting environmental quality. It refers specifically to the security situation and emergency facilities (Karacor and Parlar, 2017; D'Amico et al., 2024).

Cleanliness is another factor important in assessing environmental quality (Poleykett, 2022). It means that the open space is a clean, quiet, and pollution-free area (Luo et al., 2022).

Beauty is one of the factors used to qualify environmental quality. It reflects scenery, architecture, and the green environment pleasing to the eye (Ubani et al., 2023).

Convenience is another factor that reflects environmental quality. It is a measure of the accessibility of the open space (Dadpour et al., 2016).

Leisure facilities comprise integral components of supporting infrastructure, such as public seating, pavilions, fitness apparatus, areas shaded by trees, and potable water systems (Kim et al., 2022).

Catering facilities are supporting facilities that mainly include convenience stores, restaurants, and bars (Yuen et al., 2019).

Entertainment facilities refer to other important supporting facilities (Song et al., 2024). These include mainly chess

and card rooms, ballrooms, karaoke televisions, and swimming pools (Zhao et al., 2022).

Tourist facilities, which mainly serve users who travel from afar, include hotels, health resorts, and parking lots (Das and Maitra, 2024).

Canal landscape conservation refers to the inheritance of the traditional landscape features of the Grand Canal (Sas-Bojarska et al., 2024).

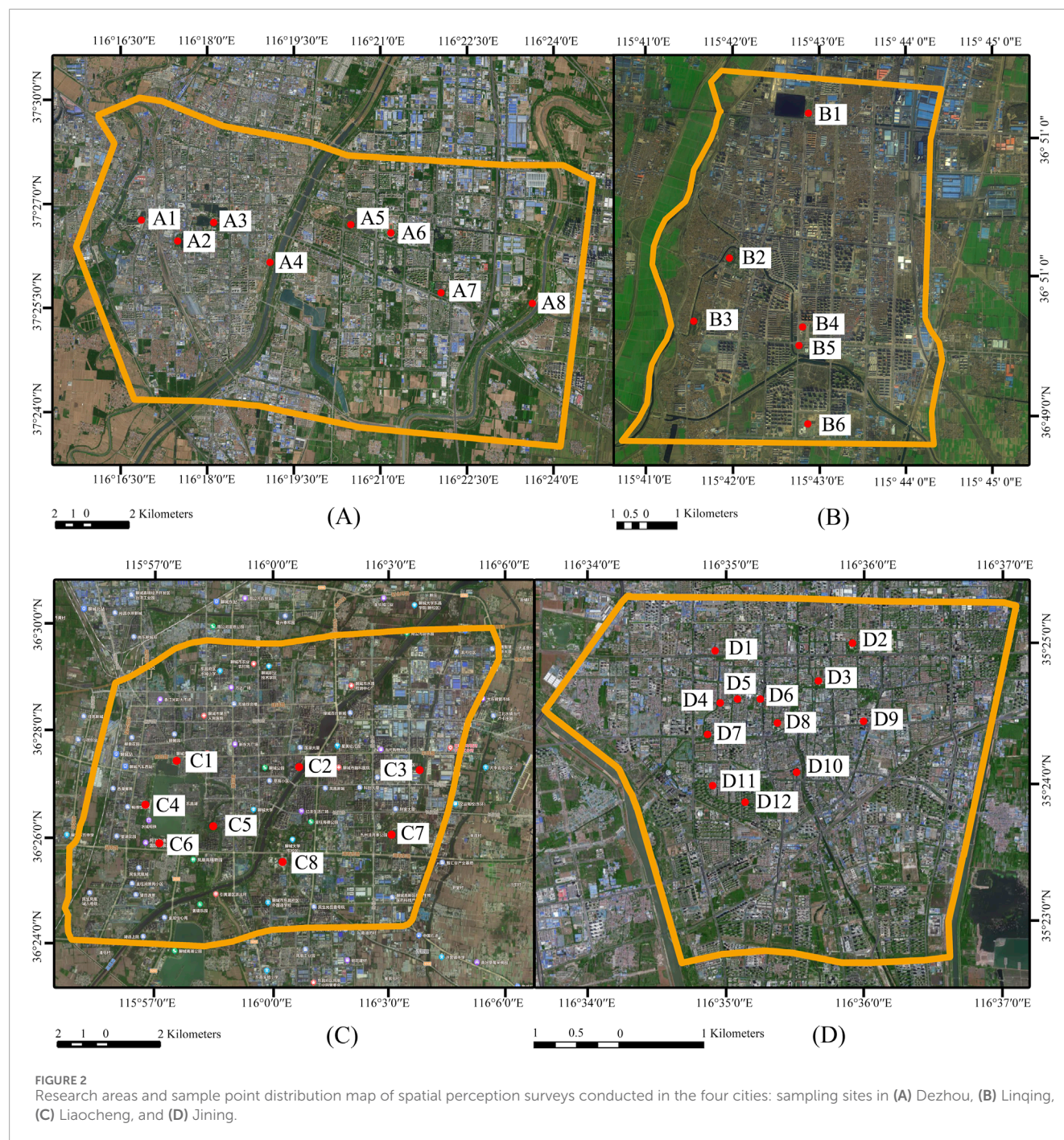
Canal cultural atmosphere means that the open space reflects the unique canal cultural characteristics of the city, epitomized by folk culture and traditional artistic forms (Flemsæter et al., 2020).

Canal publicity refers to the existence of some form of advertising of the Grand Canal by local governments, institutions, and/or other organizations (Wang and Stevens, 2020), such that users can be made aware of the public spaces of the Grand Canal.

In addition to the survey data, Geographic Information Systems (GIS) data were utilized to map the physical characteristics of the study area, which helped in the analysis of spatial variables such as proximity to transportation hubs, environmental quality, and facility availability.

3.2.3 Data mapping

Since the original expressions of categorical variables are non-numerical and cannot be directly input into machine



learning models, data mapping is required to convert them into standardized numerical values, laying the foundation for subsequent analysis (Macieira et al., 2024). These categorical variables included behavior purpose, transportation mode, occupation, household registration, frequency of behavior, encounter with traffic congestion, gender, age, education level, and monthly income. For example, numbers 1–7 in the behavior purpose category corresponded to physical exercise, cultural and artistic activities, meeting friends, taking children out to play, sightseeing, passing through, and other purposes, respectively. Similar mapping

strategies were applied to other variables such as transportation mode and occupation.

3.2.4 Data classification

This study involved various spatial categories such as riverfront spaces, public green spaces, and commercial streets. A mapping dictionary was employed to convert location names into the corresponding spatial categories. Additionally, One-Hot Encoding (OHE) was applied to the categorical variables to convert the data into a format suitable for subsequent statistical analysis

TABLE 2 Composition, data provenance and functional assignment of variables.

No.	Variable	Data source	Role
1	Behavior purpose	Questionnaire	Analysis
2	Mode of transportation	Questionnaire	Analysis
3	Occupation	Questionnaire	Analysis
4	Household registration	Questionnaire	Analysis
5	Visit frequency	Questionnaire	Analysis
6	Encounter congestion	Questionnaire	Analysis
7	Gender	Questionnaire	Analysis
8	Age group	Questionnaire	Analysis
9	Education level	Questionnaire	Analysis
10	Monthly income	Questionnaire	Analysis
11	Safety	Questionnaire	Analysis
12	Cleanliness	Questionnaire	Analysis
13	Beauty	Questionnaire	Analysis
14	Convenience	Questionnaire	Analysis
15	Leisure facilities	Questionnaire	Analysis
16	Catering facilities	Questionnaire	Analysis
17	Entertainment facilities	Questionnaire	Analysis
18	Tourist facilities	Questionnaire	Analysis
19	Canal landscape conservation	Questionnaire	Analysis
20	Canal cultural atmosphere	Questionnaire	Analysis
21	Canal publicity	Questionnaire	Analysis
22	Transportation hubs	GIS	Analysis
23	Environmental quality	GIS	Analysis
24	Facility accessibility	GIS	Analysis
25	Kernel density of visits	GIS	Validation

and machine learning model training (Klimo et al., 2021). The OHE technique transforms each category of a categorical variable into a new binary (0 or 1) column, which is required by most machine learning algorithms because they can process only numerical data (Al-Shehari and Alsowail, 2021). The purpose of OHE is to categorize variables that lack ordinal relationships. However, because behavioral frequency, educational attainment, and monthly income exhibit inherent order, OHE was not applied to these variables.

In this study, the following categorical variables underwent the OHE process (Figure 4):

Behavior purpose (e.g., physical exercise or cultural activity), Mode of transportation (e.g., taxi, driving or public transport), Occupation, Household registration, Frequency of behavior (e.g., multiple times a day, daily, or weekly), Encounter with congestion (yes or no), Gender, Age group, Education level, Monthly income.

The OHE process was implemented using the `get_dummies` function from the Pandas library (Elsner, 2023). This function automatically creates new columns for each category and assigns a value of 1 to the corresponding column based on the label in the original data, while assigning 0 to the other columns. For example, if the behavior purpose of a sample is “physical exercise”, the “behavior_purpose_physical_exercise” column will be marked as 1, while all other behavior-related columns will be marked as 0. Through this encoding, the original categorical data were transformed into a numerical format that could be analyzed further and used for model training.

3.2.5 Data conversion

The original data have scale differences and inconsistent data types. Therefore, data type conversion and standardization are required to unify data format and scale, ensuring analysis accuracy. All data were converted into integer type data to satisfy the requirements of subsequent analytical tools. Additionally, to mitigate the impact on the analysis of features with different scales, the data underwent standardization. Several common standardization techniques were compared, including MinMaxScaler, StandardScaler, Normalizer, and RobustScaler (Lee, 2024). These methods transform the data into a uniform scale, thereby ensuring that the data are prepared appropriately for subsequent cluster analysis.

Contour coefficients were used to evaluate the clustering performance of each standardized method. The silhouette coefficient can measure the similarity between data points within a cluster and their dissimilarity from the data points of other clusters. The higher the score, the better the clustering effect. As listed in Table 3, the Normalizer method achieved the highest average silhouette score (0.14) across the K-means, Agglomerative Clustering, and Bisecting K-means algorithms, outperforming MinMaxScaler (0.05), StandardScaler (0.03), and RobustScaler (0.04). The clustering results of the Normalizer method demonstrated the best overall performance under multiple classifications and various clustering algorithms.

3.3 Analytical framework and study design

This study developed a multilayered analytical framework integrating multisource data acquisition and machine learning algorithms to systematically identify and quantify key factors influencing public space usage along the Shandong section of the Grand Canal (Figure 5). The framework consists of five layers: Data Acquisition and Preprocessing, Feature Engineering and Modeling, Behavioral Profiling and Preference Mining, Validation, and Output.

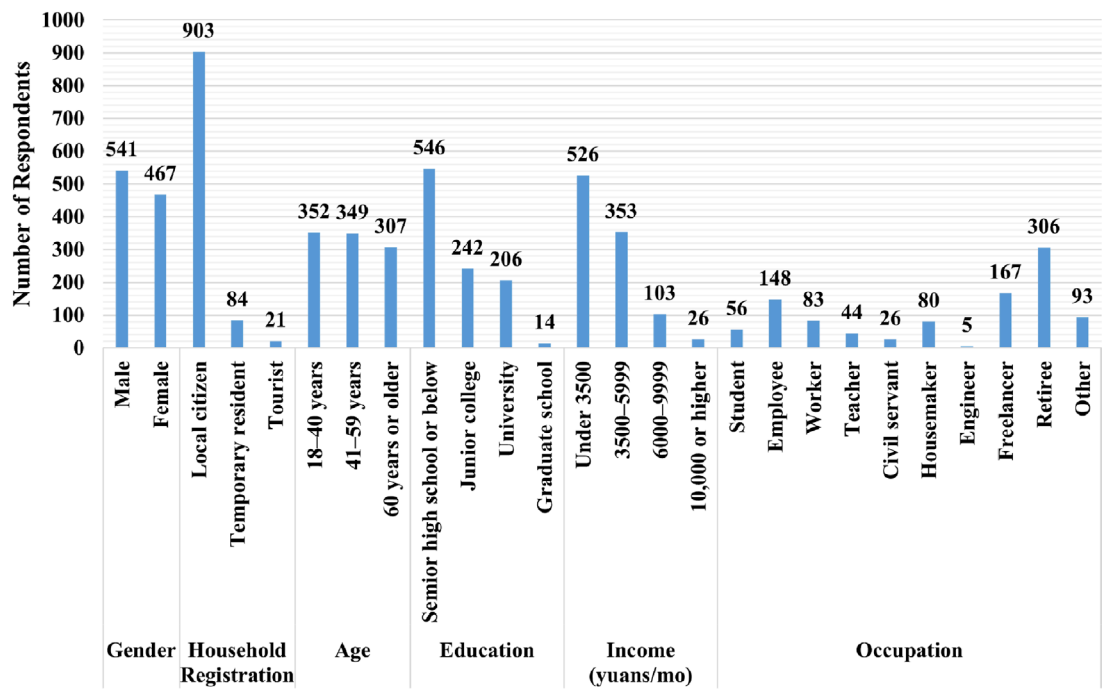


FIGURE 3
Basic information on respondents to the open space questionnaire survey conducted in the four cities along the Grand Canal in Shandong Province, China.

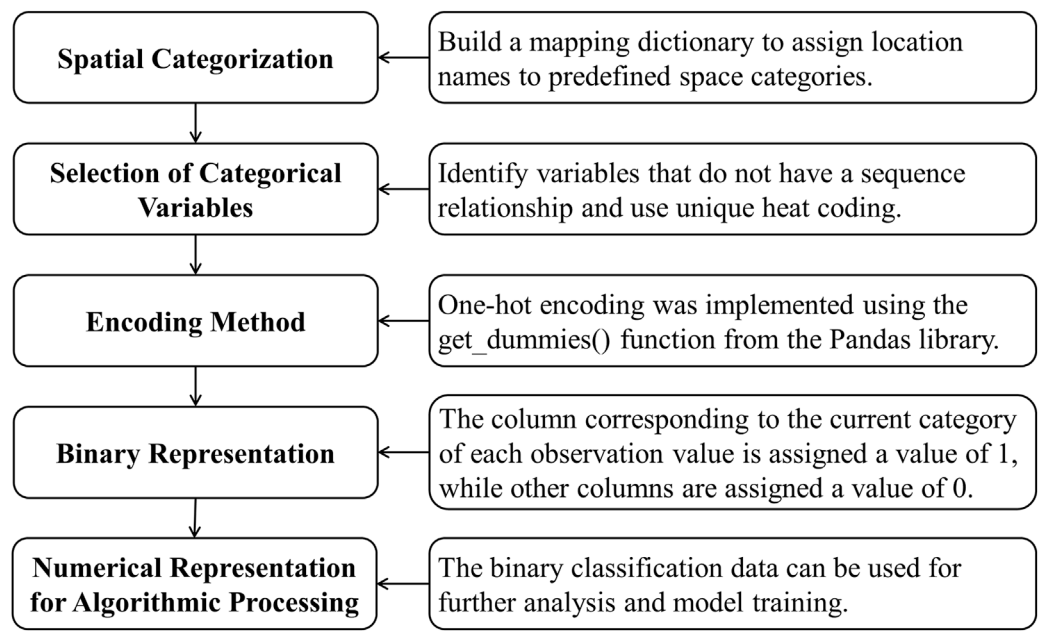


FIGURE 4
Flowchart of the categorical data encoding process.

3.3.1 Justification for adopting K-means in heritage-oriented public space mining

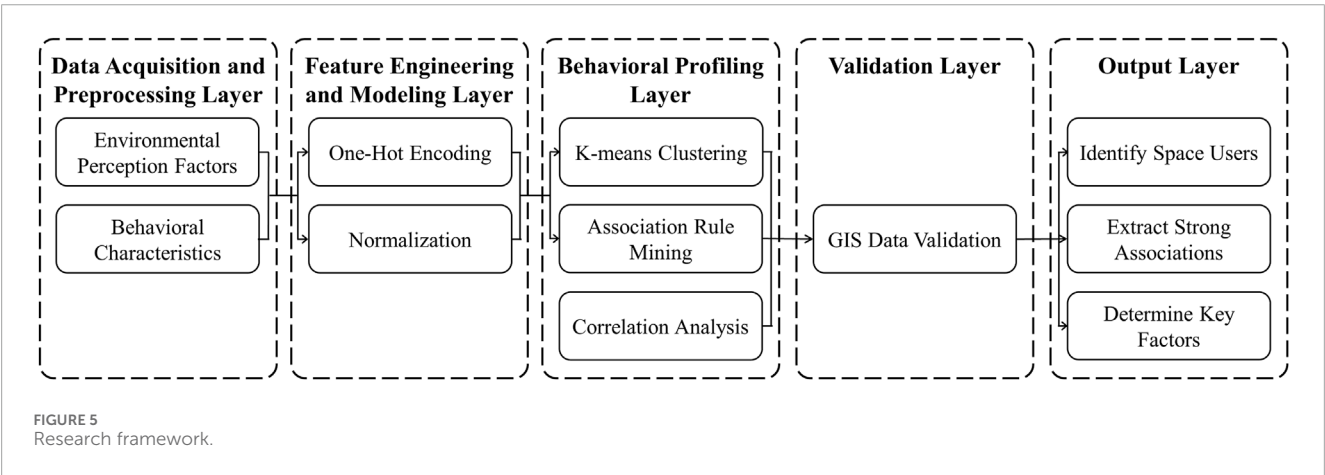
Firstly, K-means clustering is adept at handling multidimensional data in public space research, which is highly

in line with the need to integrate multiple variables in the analysis of visitor groups in the Grand Canal Heritage Corridor. Secondly, compared to other clustering methods, the K-means algorithm has higher computational efficiency and can quickly process 1008

TABLE 3 Comparison of clustering algorithm performance with different standardization methods.

Clustering algorithm	MinMaxScaler	StandardScaler	Normalizer	RobustScaler
K-means	0.06	0.04	0.13	0.05
Agglomerative Clustering	0.04	0.03	0.13	0.03
Bisecting K-means	0.05	0.02	0.15	0.03
Average	0.05	0.03	0.14	0.04

Note: 0.14 indicates the best average performance of clustering algorithms with Normalizer standardization.



valid questionnaire samples while maintaining stable clustering results. Finally, the clustering centers generated by the K-means algorithm have good interpretability and can clearly reflect the core preferences of various visitor groups, providing direct support for the formulation of targeted planning strategies for heritage public spaces.

3.3.2 Data acquisition and preprocessing layer

Environmental perception factors and behavioral characteristics were collected through structured surveys, including perceived safety, facility accessibility, cultural atmosphere, behavioral purposes, and visit frequency. A stratified sampling design was adopted that yielded 1008 valid questionnaires from the 4 studied cities along the Grand Canal, forming the empirical foundation for model construction.

3.3.3 Feature engineering and modeling layer

Categorical variables were standardized using mapping dictionaries and then transformed into numerical features through OHE. Among the four normalization techniques tested (MinMaxScaler, StandardScaler, Normalizer, RobustScaler), the Normalizer method exhibited optimal performance across the clustering algorithms and was therefore adopted in this study.

3.3.4 Behavioral profiling and preference mining layer

K-means clustering was used to segment respondents into distinct behavioral groups. The optimal number of clusters was

determined to be three based on Silhouette Scores and the Elbow Method. Association rule mining via the Apriori algorithm uncovered strong associations between environmental attributes and behavioral preferences. Furthermore, correlation analysis quantified the relationships between key factors that influence public space utilization.

3.3.5 Output layer

The output phase identifies prototypical respondents' groups, extracts highly confident and statistically significant behavior–environment associations, and determines the key factors driving space utilization.

3.4 Machine learning models

To analyze the data, machine learning techniques were used to identify the patterns of respondents' behavior and preferences (Yao et al., 2024). To classify respondents into different clusters based on their behaviors and environmental preferences, this study primarily adopted K-means clustering, which is a widely used unsupervised learning algorithm (Sinaga and Yang, 2020). To determine the optimal number of clusters, two common clustering quality assessment methods were used: the Silhouette Score and the Elbow Method. The Silhouette Score is used to evaluate the quality of clustering, with higher values indicating better clustering results (Yu et al., 2020). The Elbow Method calculates the sum of squares error for different numbers of clusters to identify the

optimal cluster number (Shi et al., 2021). Additionally, association rule mining was used to explore relationships between different environmental factors and respondents' preferences. The Apriori algorithm was applied to generate frequent itemset (Bashir, 2020) and association rules (Luo and Miao, 2023) to help uncover notable patterns in the usage of various public spaces.

3.4.1 Clustering analysis

Clustering analysis is an unsupervised learning technique used to group data objects based on similarity. It is possible to briefly explain how this algorithm identifies similar groups by comparing data features, and to highlight its main differences from other classification methods. K-means clustering was used to classify respondents into distinct groups based on their behavioral patterns in this study (Harris and De Amorim, 2022). This approach supported identification of different types of space respondents, each with unique characteristics and preferences (Vera et al., 2022). To determine the optimal number of clusters, silhouette coefficients were calculated for different cluster numbers. The silhouette coefficient is a metric used for assessing clustering performance, with higher values indicating better clustering quality (Bagirov et al., 2023).

K-means clustering: A centroid-based clustering method that optimizes cluster assignment by minimizing the within-cluster sum of squares error:

$$\min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2$$

where k represents the number of clusters, S_i denotes the set of data points in the i th cluster, and μ_i is the centroid of the i th cluster.

As shown in Figure 6, by comparing the silhouette coefficients and the sum of squares errors across different numbers of clusters, we found that the silhouette coefficient—after normalization using Normalizer—reaches its maximum value. Therefore, we determined that three was the optimal number of clusters. After determining the optimal clustering number and standardization method, the K-means algorithm was used for the final clustering analysis, adopting a fixed random state to ensure the reproducibility of the results. By choosing the optimal number of clusters, we could more accurately identify the characteristics of visitor groups and thereby provide urban planners with targeted recommendations.

3.4.2 Association rule mining

The association rules generated using the Apriori algorithm were employed to explore the relationship between environmental features and visitor behaviors (Du et al., 2024). The factors of lift and confidence were used to evaluate the strength of these associations (Altay and Alatas, 2020).

Clustering analysis revealed no notable patterns in the visits to locations by the three groups of individuals. To further explore the associations between the groups and the visited locations, association rule mining was conducted on the existing data. First, continuous variables in the dataset were binarized into Boolean values (Michalak, 2024). This was achieved by comparing each feature value with 60% of the range of that feature (Marcondes et al., 2018). If the feature value exceeded 60% of the range added to the minimum value, it was considered true (True); otherwise, it

was considered false (False). For generating a frequent itemset, the Apriori algorithm was applied with a minimum support threshold of 0.4 to ensure that the discovered itemset had a high frequency of occurrence in the data. The generated frequent itemset included various combinations such as “cleanliness, safety, and aesthetics” with occurrence frequencies exceeding 40%. Based on the frequent itemset, association rules were extracted using lift as a measure. Lift reflects the positive correlation of the itemset, with a lift value of >1 indicating positive association. For example, the analysis revealed positive correlation between “cleanliness” and “safety.”

3.4.3 Correlation analysis

The machine learning model was used to conduct more accurate and comprehensive analysis of the correlation between respondents' behavior and public space environmental elements. Through application of advanced machine learning techniques, we systematically evaluated the relationship between the key factors affecting the use of public space along the Grand Canal in Shandong.

The correlation analysis process involved using supervised learning models to predict the strength of the correlations between urban space features and respondents' behaviors. These models were trained on datasets that included variables such as aesthetics, cleanliness, safety, convenience, and the presence of services like rest facilities and tourism amenities.

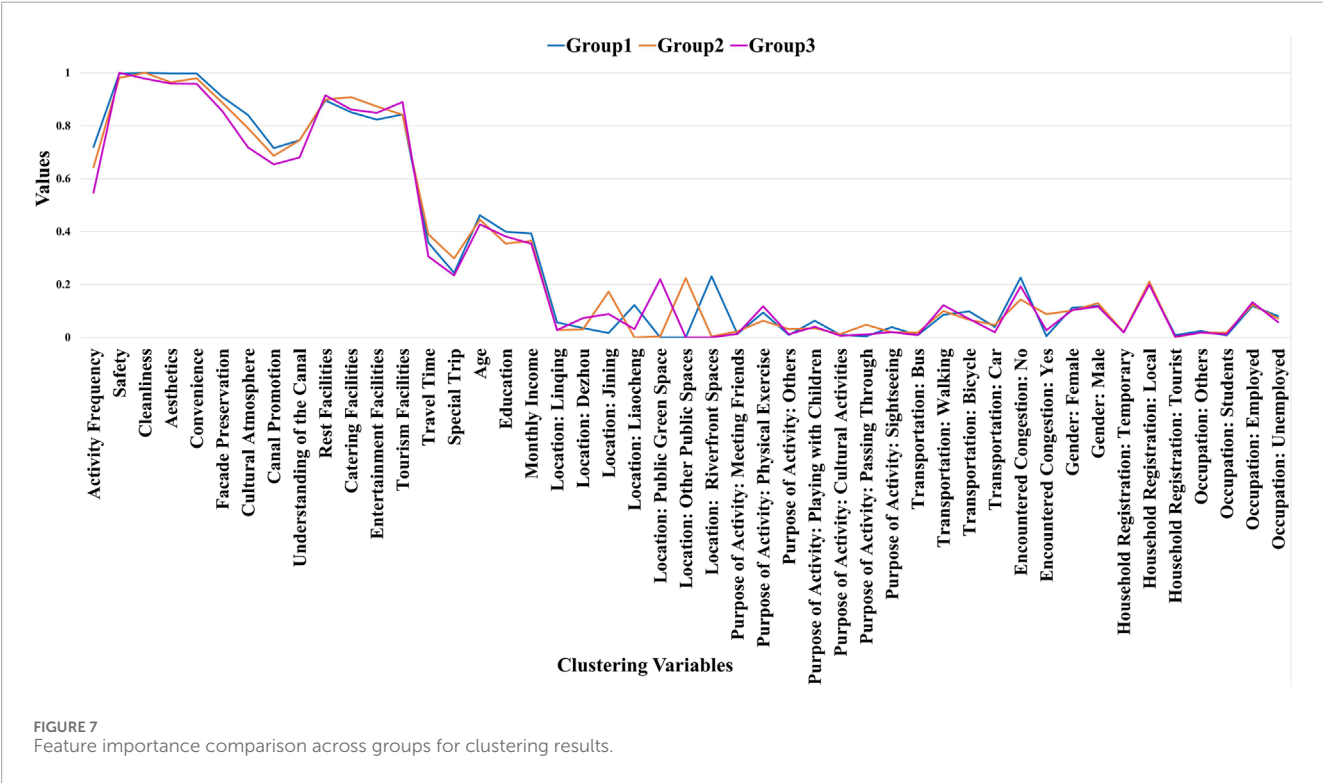
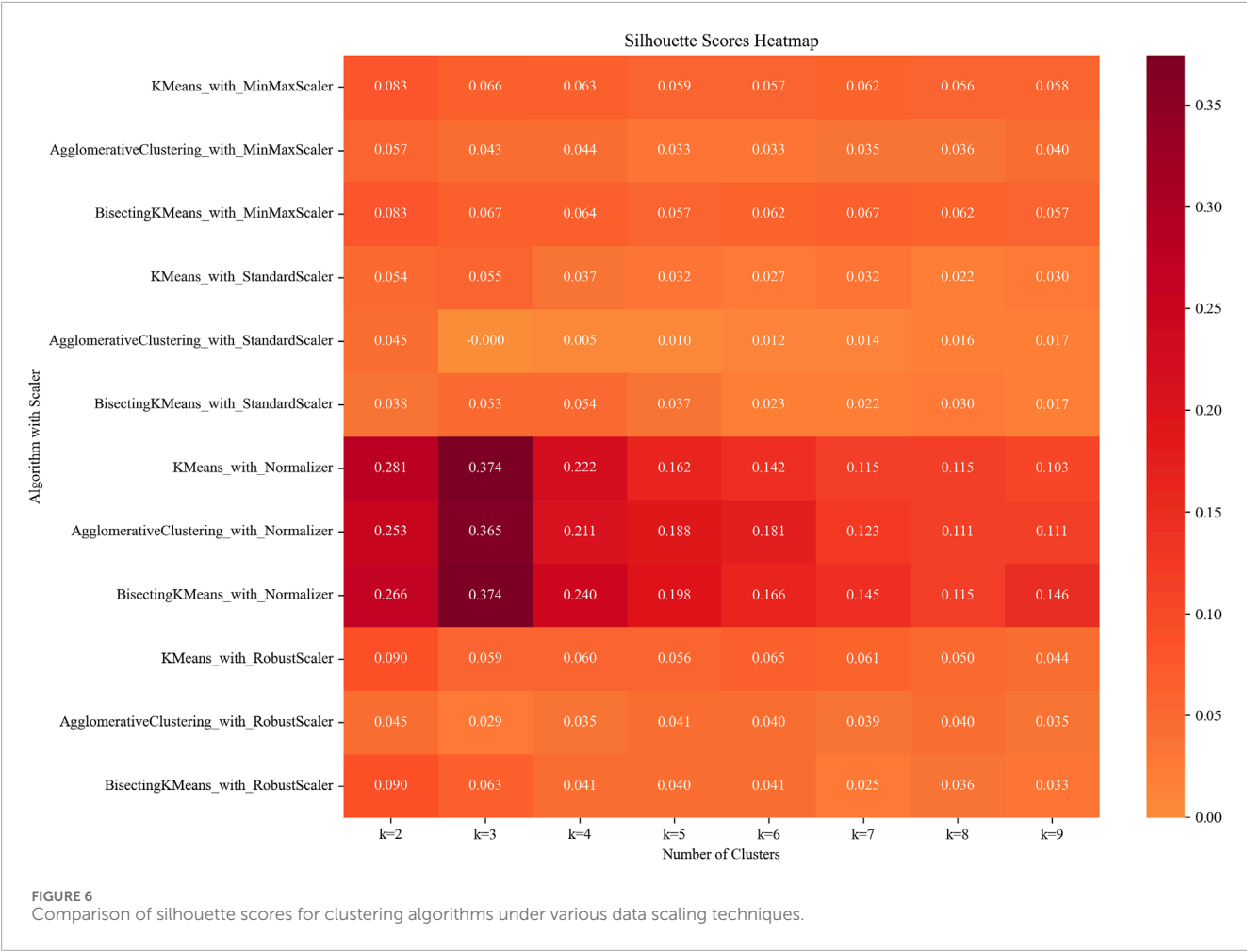
4 Results

We used a combination of clustering, association rule mining, and correlation analysis to investigate the behaviors and preferences of respondents in relation to public spaces along the Grand Canal. The results revealed several key patterns in public space usage and respondents' preferences.

4.1 Clustering results

The mean values of the clustering results for the three groups were calculated to determine the cluster centers for each group, which were then plotted as a line chart (Figure 7). The 72% Confidence Intervals for the cluster centroids were obtained via 200-times bootstrap resampling; none of the pairwise intervals overlapped, confirming the statistical distinctiveness of the three visitor groups. This indicated that the results of the cluster analysis were stable. The chart clearly illustrates marked differences in the cluster centers, particularly in the first half of the dataset. These differences are most apparent in the clustering variables such as “Activity Frequency,” “Cleanliness,” “Aesthetics,” “Convenience,” “Cultural Atmosphere,” “Canal Promotion,” “Rest Facilities,” “Catering Facilities,” “Entertainment Facilities,” and “Tourism Facilities.” Additionally, notable distinctions are observed in factors related to behavioral patterns such as walking behavior, and local household registration, further emphasizing the behavioral disparities across the groups.

Cluster analysis identified three groups of users with notable differences. There are obvious differences in the preferences and behavior patterns of the various groups in terms of the use of public space. These differences are not only attributable to



basic demographic characteristics but are also closely related to the form, functional layout, and accessibility of the public space in the cities along the Grand Canal in Shandong. From the perspective of urban/rural planning and social behavior, the following part systematically explains the behavior characteristics of the three groups of people and their relationship with environmental factors, emphasizing the spatial logic of “respondents environment interaction.”

4.1.1 Young and middle-aged local residents with high accessibility preferences

Group 1 is composed mainly of young and middle-aged local residents, with high activity frequency (0.72). Their preferred environmental factors are cleanliness (1.00), safety (1.00), and convenience (0.99). Members of this group have high requirements for the public space environment, and they pay great attention to the infrastructure and their quality of life.

In terms of transportation mode, more people in Group 1 walk (0.09) rather than travel by motor vehicle (0.04) to access public spaces, and most of them visit public spaces regularly.

Group 1 respondents prefer to use waterfront spaces with clear functions, easy accessibility, and open vistas. They usually choose to access riverside spaces with both leisure and commuting functions.

In terms of the purpose of the activity, Group 1 members prioritize physical exercise (0.09) and meeting friends (0.01), indicating that physical exercise and social interaction are the main motivations for this group of people to access public space.

The spatial behavior distribution of this group is highly consistent with the characteristics of the ribbon spatial structure along the Grand Canal. In cities along the canal, such as Dezhou, Linqing, Liaocheng, and Jinan, residential areas and employment areas are often connected by the canal in a linear distribution, which is conducive to the combination of exercise and commuting for this group.

The behavioral patterns of Group 1 are indicative of daily users whose spatial utilization is strongly shaped by the rhythms of work and the routines of everyday life. Their high attention to safety, cleanliness, and spatial order reflects their trust and dependence on urban infrastructure.

4.1.2 Retired and middle-aged individuals with a strong interest in cultural engagement

Group 2 comprises mainly middle-aged and elderly people, especially retired residents, with similar numbers of males and females. In terms of their preference for environmental factors, cleanliness (1.00) and convenience (0.98) are highly valued, indicating that the respondents of this group value environments that are both clean and easily accessible. Safety (0.98) is also one of the important features, while catering facilities (0.91) and rest facilities (0.90) are also valued. This indicates that Group 2 members are inclined more toward convenient facilities and infrastructure that can enhance comfort and happiness.

The transportation modes of this group are more diversified, with walking (0.10) and cycling (0.07) being the main modes of transportation, indicating that Group 2 respondents are concerned about the accessibility of the urban walking system. Interviews revealed that this group of people usually use walking or cycling to indirectly achieve their physical exercise goals.

Different from Group 1, whose preference is for waterfront spaces, the members of Group 2 prefer other types of public space (0.22). The trends of their activity objectives indicate that the members of this group have greater interest in physical exercise (0.06), meeting friends (0.02), and cultural activities (0.01).

This result is consistent with the on-site interviews, with older respondents indicating high dependence on facilities such as dining, leisure, and cultural participation in public spaces. They strongly prefer spaces with historical or cultural symbolic importance, such as historic blocks and green spaces with cultural activities.

The preferred type of public space for this group has a commonality, i.e., cultural facility intensive areas. In the public spaces along the Jining Canal, historical attractions coexist with modern leisure facilities, forming a spatial complex that combines “cultural memory + comfortable experience.” The development of these spaces is suitable for low-intensity activities such as slow walking and stopping experiences that accord with the behavioral needs of this group.

From the behavioral perspective, Group 2 demonstrates characteristics emblematic of the “Communicative User” typology. Group 2 members are more inclined to achieve social interaction, emotional comfort, and cultural identity in public spaces, and therefore have a higher sensitivity to cultural atmosphere, convenience, and facility completeness. The low-intensity walking mode and preference for cultural participation indicate that Group 2 members are more concerned with the context and meaning that public spaces can provide.

4.1.3 Diverse transportation respondents with mixed spatial preferences

Group 3 has a relatively diverse composition that includes both tourists and local residents. The members of this group attach great importance to rest facilities (0.91) and tourism facilities (0.89), and they prefer public spaces that provide a relaxing environment and rich entertainment. Their use of public space is more exploratory and diverse, and the frequency of activities is not fixed. They often experience the changes brought by different spaces at different times.

The transportation mode adopted by Group 3 members is usually walking or cycling (0.07), and the proportion that drives to access the public space is very low (0.02). Typically, members of this group explore public spaces in a relaxed way and avoid complex transportation choices.

This group is more inclined to engage in activities in areas with dense green spaces and tourism facilities, especially focusing on the completeness of spatial facilities and visual quality. There is high demand for the overall atmosphere of the space, together with visual and sensory experience it provides.

Group 3 members prefer waterfront areas that combine ecological and tourism functions, especially those with strong openness and easy accessibility. This type of space emphasizes landscape coherence and the tourist experience, which is a core node of a city’s “soft attraction.”

The behavioral patterns of Group 3 members are characterized predominantly by experiential tendencies. Their spatial use does not rely on daily commuting or long-term settlement, but is driven mainly by novelty, environmental diversity, and short-term convenience. Compared with Group 1 and 2 members, their behavior patterns emphasize the flexibility and sensory

stimulation of public spaces, showing higher spatial fluidity and freedom of choice.

Cluster analysis clearly identified three groups of public space users with significant differences in behaviors and preferences, whose characteristics are highly consistent with the spatial structure and functional layout of cities along the Grand Canal, providing a basis for group division in subsequent differentiated planning.

4.2 Association rule analysis results

4.2.1 Analysis of the association between location visits and respondents groups based on association rule mining

The dataset consists of various location types, including riverside spaces and public green spaces, and the characteristics of these locations as perceived by respondents, such as cleanliness, aesthetics, safety, and other facilities. Initially, the visits to different locations did not exhibit any clear patterns related to respondent type. Therefore, to delve deeper into the connections between respondents' groups and the locations they frequent, we applied association rule mining techniques.

4.2.2 Data transformation and methodology

Continuous variables in the dataset were binarized using a threshold-based method, where the threshold was set at 60% of the range between the minimum and maximum values of each feature. This means that if a particular feature's value exceeded 60% of the difference between the minimum and maximum values plus the minimum value, the feature was considered "True" (i.e., the respondents visited that type of location or exhibited that characteristic); otherwise, the feature was marked as "False".

The environmental features involved in this study are largely dependent on the subjective perceptions of the respondents, which often exhibit a distribution pattern of "moderate preference with the highest proportion." The threshold of 60% broadly corresponds to the subjective evaluation range of "good to excellent," which is four points or above on the Likert scale. It can effectively distinguish whether the scene is strongly recognized by users and is consistent with the general public's judgment criteria for the quality of public spaces. This study verified through pre-experiments that the binarization results at the 60% threshold highly match the grouping features of other machine learning models such as K-means clustering, further supporting the rationality of this threshold.

The primary goal of this transformation was to identify specific relationships between respondents and the locations they visit, focusing on the patterns of associations that emerge when considering specific attributes of the locations such as cleanliness, aesthetics, and safety. The association rules revealed several notable relationships between these location features and the respondents' preferences (Table 4).

4.2.3 Key findings from the association rules

4.2.3.1 Environmental perception and emotional connection

Respondents associated waterfront spaces with the features of cleanliness, aesthetic value, and social activities, indicating a

strong emotional and perceptual bond between them and such environments. The perfect confidence level (1.0) between Riverside Space and Cleanliness indicates that cleanliness is a fundamental expectation, and might even be prerequisite for people to be willing to approach such spaces. Similarly, the improvement value between Riverside Space, Cleanliness, and Aesthetics is 1, highlighting people's overall perception of waterfront space as "beautiful and orderly." This shows that visitors to waterfront areas are often emotionally driven, seeking psychological recovery, a beautiful experience, and social vitality.

4.2.3.2 Functional use and safety awareness in public green spaces

Public green spaces are often perceived through a more functional and pragmatic lens. The association between Public Green Space and Safety (confidence: 0.94; lift: 1.04) shows that safety plays a critical role in determining whether individuals choose to engage with these areas. The strong correlation with rest facilities indicates that such respondents pay more attention to comfort, convenience, and predictability. They prefer to choose a safe, pleasant, and relaxing environment rather than an exploratory or visually impactful space.

4.2.3.3 Activity purpose and behavioral typologies

The association rules also reveal the specific preferences of respondents in terms of the purpose of their activities. For riverside spaces, the strong link between "Riverside Space" and "Special Trip" (support = 40.30%, confidence = 100.00%, lift = 1.00) indicates these areas primarily attract intentional visits, likely driven by their scenic and cultural significance. In public green spaces, the dominant association with "Rest Facilities" (support = 43.09%, confidence = 95.27%, lift = 1.01) highlights their role as functional zones for relaxation, complemented by high confidence in safety (93.82%) and cleanliness (94.91%).

Association rule mining revealed strong correlations between different public space types and environmental attributes as well as usage behaviors, providing an accurate attribute matching basis for spatial function optimization.

4.3 Correlation analysis results

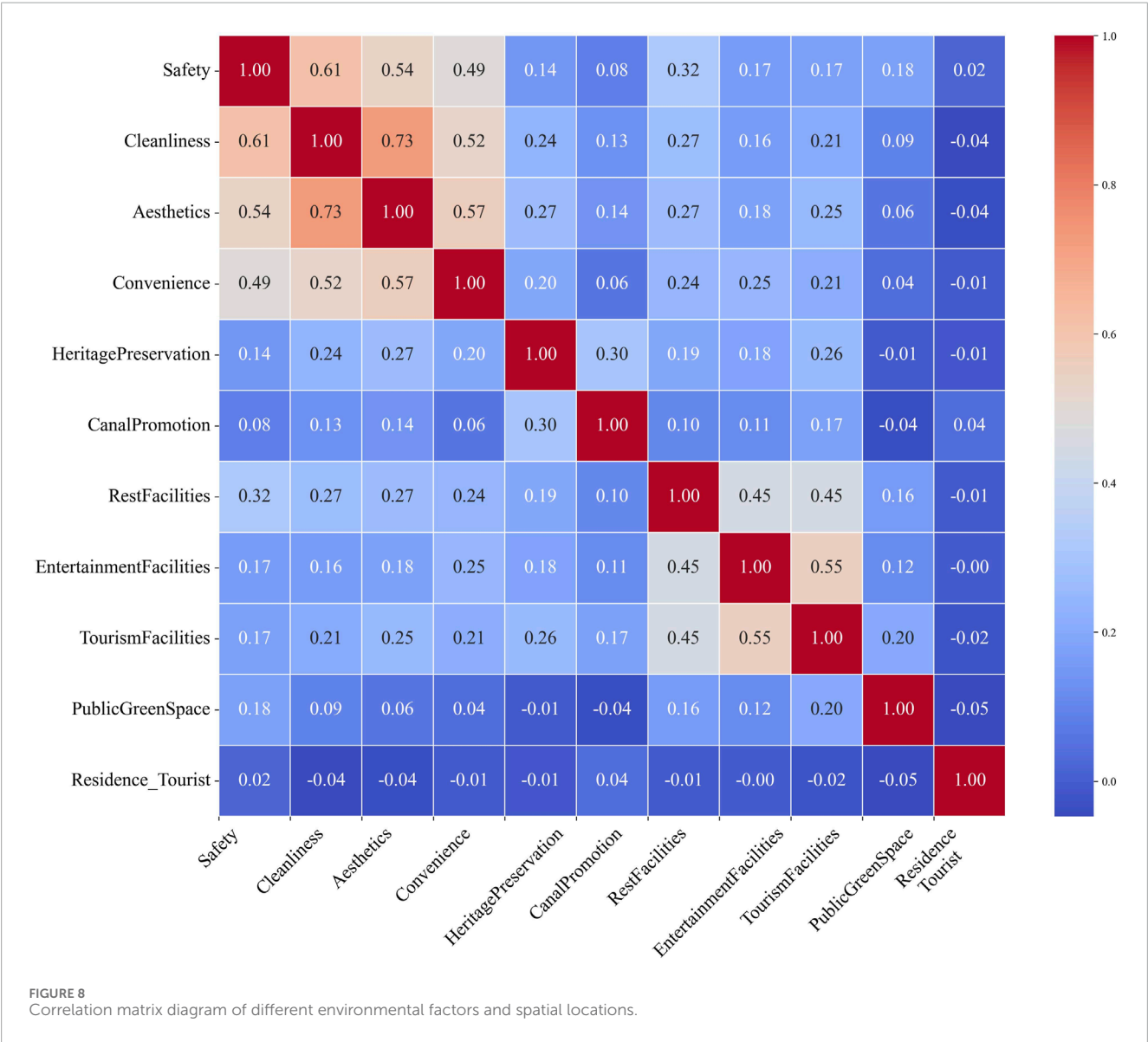
This study conducted in-depth analysis of the correlations between various variables in relation to urban and rural planning, and the results are presented in the form of a heatmap for clearer understanding of the relationships between these variables (Figure 8). The correlation coefficients range from -1.0 to $+1.0$, where $+1$ indicates perfect positive correlation, -1 indicates perfect negative correlation, and 0 indicates no correlation. Several key insights can be drawn from analysis of the heatmap.

Aesthetics and Cleanliness: The substantial correlation (0.73) between Aesthetics and Cleanliness indicates that aesthetically pleasing environments are often perceived as cleaner. This relationship suggests that enhancing the aesthetic appeal of public spaces, through landscaping or design, might improve the perception of cleanliness, thereby attracting more respondents.

Safety and Cleanliness: The strong positive correlation (0.61) between Safety and Cleanliness suggests that locations perceived

TABLE 4 Binary association rules between locations and features.

Antecedent	Consequent	Support (%)	Confidence (%)	Lift
(Riverside Space)	(Cleanliness)	40.30	100.00	1.00
(Riverside Space)	(Aesthetics)	40.30	100.00	1.00
(Riverside Space)	(Special Trip)	40.30	100.00	1.00
(Public Green Space)	(Safety)	42.43	93.82	1.04
(Public Green Space)	(Cleanliness)	42.93	94.91	1.04
(Public Green Space)	(Aesthetics)	42.11	93.09	1.07
(Public Green Space)	(Convenience)	40.63	89.82	1.03
(Public Green Space)	(Rest Facilities)	43.09	95.27	1.01



as safe tend also to be viewed as clean. This interdependence highlights the importance of integrating safety features, such as lighting and surveillance, within well-maintained areas to enhance the perception of cleanliness and safety.

Convenience and Aesthetics: The moderate correlation (0.57) between Convenience and Aesthetics underscores the link between accessibility and the visual appeal of public spaces. The ease of access to spaces enhances their perceived aesthetic value, highlighting the need for urban planners to prioritize both accessibility and visual quality in the design of public areas.

Rest Facilities and Tourism Facilities: The strong correlation (0.45) between Rest Facilities and Tourism Facilities implies that public spaces with sufficient rest amenities are more likely to be associated with tourism-oriented features. This relationship suggests that the development of tourism infrastructure should prioritize the inclusion of comfortable resting spaces alongside other tourism facilities to enhance the overall visitor experience.

Correlation analysis quantified the strong positive correlations between key environmental factors such as aesthetics-cleanliness and safety-cleanliness. The strong correlation between aesthetics and cleanliness provides a new perspective for the coordinated optimization of aesthetics and cleanliness in the planning of public spaces along the Grand Canal waterfront. This discovery breaks through the limitations of traditional planning where the two are designed separately, and can guide planners to simultaneously embed dynamic cleaning management nodes in waterfront landscape creation, improving the overall perceived quality of the space. The findings of the correlation analysis provide valuable guidance for urban planning, particularly in terms of facility layout and functional optimization, offering empirical support for better integration of spatial elements in urban development.

4.4 Validation

4.4.1 Model validation methods

This study employed various model validation techniques to ensure the reliability and effectiveness of the results of the machine learning analysis, including cluster quality assessment, multimethod cross-validation, spatial distribution validation, and population preference statistical validation (Figure 9). The clustering quality evaluation used the contour coefficient and elbow method to determine that three was the optimal number of clusters for the K-means clustering. By comparing the clustering performance under different standardization methods, Normalizer was selected as the optimal standardization method (Table 3). Multimethod cross-validation combines clustering analysis, association rule mining, and correlation analysis to cross-validate the relationship between user behavior and environmental factors. For example, strong correlation was confirmed between cleanliness and safety in both the association rules and the correlation analysis.

4.4.2 Respondents' preferences and spatial distribution

Using kernel density analysis in GIS, we compared the distinct public space respondents' groups identified through

clustering analysis with the actual spatial distribution of public spaces (Figure 10). The results reveal strong spatial correlation between the preferences of the three groups of respondents and spatial characteristics such as cleanliness, safety, and aesthetics. For example, the members of Group 1 prefer safety and cleanliness, and they tend to frequently visit parks and green spaces near the city center, while Group 2 members enjoy cultural activities and are more likely to choose waterfront areas with historical and cultural importance. This finding is further supported by field observation data, confirming that our machine learning model accurately captures the actual preferences of public space respondents.

4.4.3 Group-specific preferences and perceptions of public space facilities

The analysis of public space facility preferences among Groups 1, 2, and 3 reveals marked differences across various indicators, including safety, cleanliness, aesthetics, convenience, heritage preservation, cultural atmosphere, canal promotion, canal knowledge, leisure facilities, dining facilities, entertainment facilities, and tourism facilities (Table 5).

Among the three groups, Group 1 consistently exhibited the lowest scores across all environmental indicators. Moderate preferences were observed for safety (3.86) and cleanliness (3.86), indicating a certain level of concern for fundamental environmental conditions. However, relatively low demands were recorded for functional facilities such as leisure facilities (3.72) and dining facilities (3.66). Additionally, Group 1 assigned notably low scores for canal promotion (2.78) and cultural atmosphere (3.30), suggesting limited interest in cultural dissemination and environmental ambiance.

In contrast, Group 2 demonstrated the highest scores across all environmental attributes, particularly for safety (4.89), cleanliness (4.86), and aesthetics (4.79), indicating extremely high expectations for the basic quality of public spaces. This group also showed substantial demand for convenience (4.77) and heritage preservation (4.04). The relatively high scores for leisure facilities (4.26) and tourism facilities (4.00) further emphasized their strong emphasis on spatial comfort and cultural enrichment.

Group 3 presented intermediate scores across the indicators, while still maintaining relatively high preferences for safety (4.62) and cleanliness (4.51), reflecting considerable concern for basic environmental quality. The score of cultural atmospheres (3.59) and heritage preservation (3.99) in this group was higher than that in Group 1, indicating a certain degree of need for cultural experience. Moreover, Group 3 exhibited notable preferences for leisure facilities (4.16) and entertainment facilities (3.86), highlighting a tendency toward seeking recreational and entertainment functions.

Based on the analysis results of the three groups of data, the different groups show notable differences in their functional needs and environmental perception of public space. These differences are largely consistent with the results obtained through clustering analysis, thereby further validating the effectiveness and reliability of the clustering outcomes.

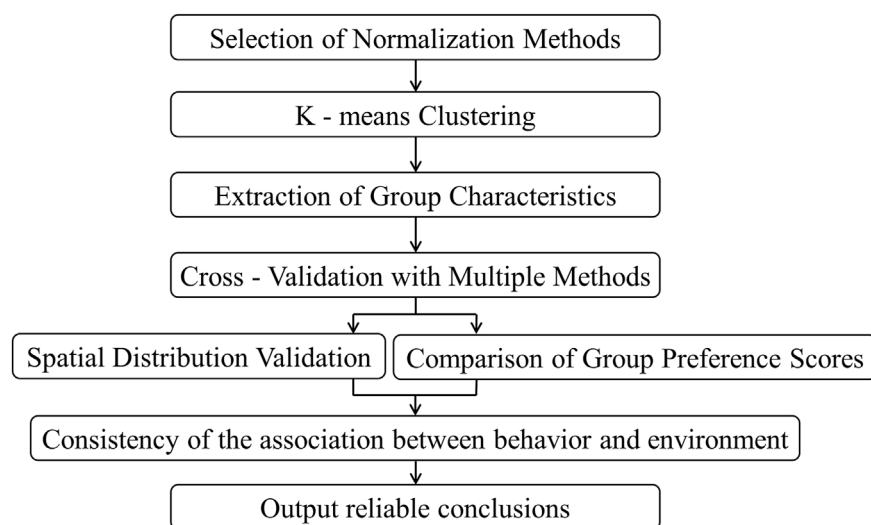


FIGURE 9
Flowchart of model validation process.

5 Discussion

5.1 Innovation of the study

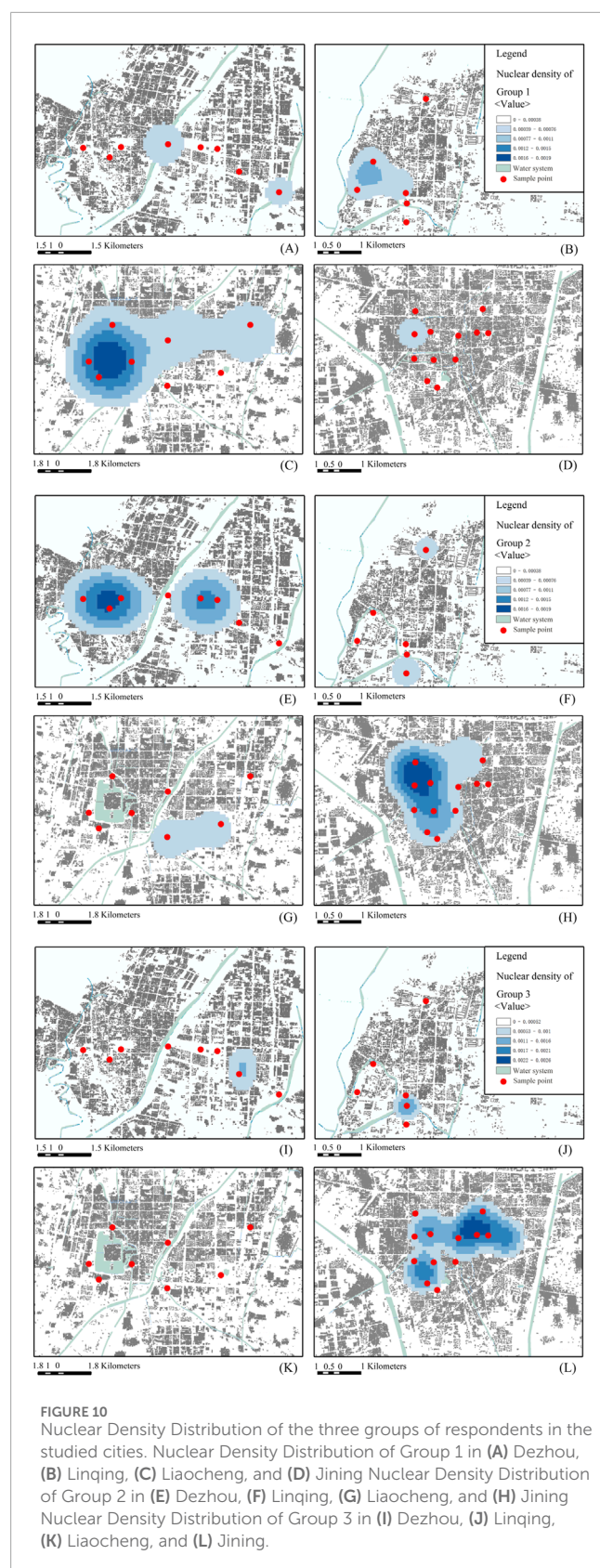
The application of a machine learning model to public space planning has proven to be an efficient method, which can analyze complex datasets and identify patterns that are difficult to detect using traditional methods (Dabra and Kumar, 2023; Koutra and Ioakimidis, 2023; Riva et al., 2024; Schmitt et al., 2024). For example, traditional methods usually rely on qualitative means such as questionnaire surveys and field observations, which have limitations when dealing with large-scale datasets or revealing the complex relationships between multiple variables. This study expanded the technical boundary of urban spatial analysis by introducing K-means clustering, association rule mining, and correlation analysis algorithms.

In recent years, some studies have tried to apply machine learning to the analysis of urban public space use. However, most such studies focused on the broader urban environment (Riva et al., 2024; Alkhereibi et al., 2025; Durowoju et al., 2025; Robi and George, 2025), or on a specific functional space such as green space or street vendors (Barreda Luna et al., 2022; Piras et al., 2024). Compared with most other related existing studies, the innovation of this study lies in its comprehensive analysis method, which integrates aesthetic value, cultural atmosphere, accessibility, and other factors, focusing on China's unique Grand Canal cultural heritage corridor. Additionally, this study divided the respondents into different behavioral preference groups through cluster analysis to deepen the understanding of the characteristics of public space use, which is still a rare approach in related research conducted in China.

5.2 Limitations of the study

Although the application of machine learning to urban planning has realized many advantages, this study had some limitations. One of the main problems is the dependence on questionnaire data, which might have sample bias or include inaccurate information provided by the respondents. Although it was hoped that the sample would include a sufficient number of a wide variety of people, time and resource limitations meant that it was difficult to fully cover all social strata and regional types, especially low-income groups and migrant workers. Therefore, the data might not fully represent the public space use characteristics of the sample area (Delnevo et al., 2023). Additionally, the subjective expression of environmental perception by respondents might have certain bias, sometimes resulting in over glorification or underestimation of specific spatial features that could affect the objectivity of the data.

There are still several ethical limitations regarding the process of data collection and machine learning model application in this type of research. First, although the questionnaire survey followed the principle of voluntary participation, owing to the timeliness of on-site research, some respondents might not have had sufficient understanding of the purpose of the data. Especially among the elderly population, there are differences in their understanding of digital technology, which might lead to implicit informed consent bias. Second, the results of machine learning models depend on the representativeness of the input data. The proportions of different socioeconomic groups in the sample might have been imbalanced, resulting in bias in the model's preference recognition of marginalized groups, which in turn might have affected the fairness and inclusiveness of policy recommendations (Liu and Li, 2025; Liu et al., 2025).



5.3 Comparative analysis and integration with traditional methods

There are many advantages to using machine learning methods in urban public space research. Most related studies use machine learning because of its high data processing efficiency (Plunz et al., 2019; Rossetti et al., 2019). Because when processing large amounts of data, the cost of manual data processing might be too high for a study to be feasible. However, it must be recognized that the combination of machine learning methods and traditional urban planning survey methods is more effective (Hamarash et al., 2024). Traditional methods, such as manual surveys and field observations, continue to play an irreplaceable role in capturing qualitative insights, local context, and community-specific knowledge that might be overlooked in purely data-driven analyses (Lak et al., 2020).

Integrating machine learning techniques with conventional approaches can thus result in a more holistic and scientifically grounded planning process. For example, K-means clustering facilitates the identification of groups of respondents based on behavioral similarities across multidimensional datasets, offering a data-driven perspective on spatial usage segmentation. Although these models are statistically significant, they still need to be interpreted contextually through field interviews and observations to ensure their social and cultural relevance (Smith et al., 2023).

Similarly, association rule mining can reveal the nonlinear relationships between environmental characteristics and user behaviors, which are often difficult to identify by traditional descriptive statistical methods. Without field verification, these associations might be misread or could lack practical application value.

This study confirmed that traditional methods such as on-site interviews and questionnaire surveys are indispensable in the sociocultural context. Therefore, it is very important to build a collaborative model between machine learning and traditional methods. Such an approach not only uses the analysis ability of machine learning processing and structured complex urban data, but also uses traditional methods to provide interpretation depth, which not only improves the accuracy of urban planning decision-making but also enhances its social legitimacy and practical feasibility.

5.4 Spatial design guidelines for different visitor groups based on data-driven insights

Guidelines for Group 1: For high accessibility groups, it is recommended to add “barrier-free passages” and “high-frequency cleaning routes” in green areas within city centers to enhance the coverage of accessibility facilities, and improve the cleanliness and convenience of public spaces. Additionally, the layout of facilities should be optimized by integrating walking indices to ensure that the locations of walkways and public amenities best meet the needs of

the population. Data analysis results derived from machine learning models can help identify areas with high crowd density, enabling dynamic adjustment of cleaning frequencies and facility layouts.

Guidelines for Group 2: For cultural engagement groups, it is advisable to establish “Canal Cultural Display Nodes” and “Traditional Art Experience Spaces” in historical riverside districts. This design would not only satisfy the cultural needs of the group but also enrich the respondents’ cultural experience through enhanced landscape preservation and cultural storytelling. In spatial planning, GIS and machine learning analysis results can be used to flexibly adjust the frequency and form of cultural activities based on crowd density and the cultural and historical context of the area, thereby enhancing the respondents’ sense of involvement and belonging.

Guidelines for Group 3: For respondents using multiple modes of transportation, it is recommended to establish “shared bicycle parking points” and “green shuttle buses” around large parks, together with supporting facilities such as rest areas and tourist information stations. This would not only meet the needs of various modes of transportation but also enhance the convenience of the respondents. By integrating GIS-based kernel density analysis, dynamic analyses of traffic flow in densely visited areas and changes in visitor behavior can be conducted, allowing adjustments to be made to the layout of transportation facilities and rest areas, thereby improving the efficiency of public space utilization.

6 Conclusion

6.1 Summary of findings from the machine learning models

This study demonstrated the effectiveness of machine learning models in elucidating public space usage and visitor preferences along the Grand Canal. By employing K-means clustering, we identified three distinct visitor groups with varying preferences for factors such as cleanliness, safety, and cultural experiences. Association rule mining uncovered notable correlations between public space features and respondents’ behavior, providing valuable insights for urban planners. Furthermore, correlation analysis highlighted the impact of environmental factors such as cleanliness and safety on respondents’ engagement with those spaces.

The machine learning model revealed the potential to promote the transformation of urban spatial planning. Data-based insight can effectively improve the design quality and functional adaptability of public space, assist urban planners make more scientific and reasonable decisions in resource allocation, and ensure that intervention measures can better meet the diverse needs of different groups.

6.2 Implications for spatial policy, design, and future research perspectives

6.2.1 Policy and design impact

Based on machine learning analysis of three types of group behavior characteristics and spatial preferences, a differentiated intervention system with precise adaptation can be constructed for

TABLE 5 Group behavioral preference statistics.

Serial number	Safety	Clean lines	aesthetics	Convenience	Heritage preservation	Cultural atmosphere	Canal promotion	Canal knowledge	Leisure facilities	Dining facilities	Entertainmen facilities	Tourism facilities
1	3.86	3.86	3.78	3.79	3.70	3.30	2.78	3.05	3.72	3.66	3.59	3.66
2	4.89	4.86	4.79	4.77	4.04	3.44	3.00	3.16	4.26	3.99	3.86	4.00
3	4.62	4.51	4.46	4.50	3.99	3.59	3.39	3.42	4.16	3.89	3.86	3.91

urban public spaces along the Grand Canal to improve the efficiency and inclusiveness of public space use.

For the group comprising mainly young and middle-aged local residents who value convenience and safety, the environmental department should require the establishment of barrier-free access networks and dynamic clean paths in urban green spaces and open spaces along the Grand Canal. Specifically, a linkage adjustment mechanism between cleaning frequency and peak usage can be established based on the crowd density thermal data generated by machine learning models. The frequency of cleaning during the morning commuting period on weekdays should be increased, and the spatial coupling between trails and public facilities should be optimized through walking indices to ensure that the basic needs of this group for safety, cleanliness, and convenience are systematically met.

For the group comprising mainly middle-aged and elderly residents who value cultural experience, policymakers should designate canal cultural protection and utilization demonstration areas in historical districts along the Grand Canal, and organically integrate traditional art experience spaces with aged-friendly facilities. This suggestion echoes the strong correlation between cultural identity and spatial selection discovered in this research, by strengthening the cultural symbolic meaning of space through landscape narrative and dynamic cultural activities, and enhancing the sense of belonging and participation of this group.

For the experiential group that includes tourists and local residents, urban public space design should focus on strengthening the functional integration of waterfront ecology and tourism corridors. Rest stations, smart tourism information kiosks, and shared transportation hubs should be integrated to form a one-stop comprehensive service system. This layout needs to match the multimode travel characteristics and exploratory behavior preferences of the group, dynamically monitor the distribution of pedestrian flow through GIS kernel density analysis, optimize the spatial configuration efficiency of facilities, and ultimately achieve the dual goals of ecological landscape protection and tourism experience improvement.

6.2.2 Future research and tools

To promote intelligent planning and management of public spaces along the Grand Canal, this study proposed potential technological solutions for further research in the future, aiming to upgrade the model from data-driven analysis to dynamic governance. It is recognized that a planning and management system should be built to integrate prediction, decision-making, and monitoring, and break through the limitations of traditional static planning. Through integrated application of advanced algorithms and intelligent tools, the speed of response of public spaces to changes in user needs and the adaptation accuracy of heritage protection could be improved.

It is recommended that a multidimensional prediction system based on Long Short-Term Memory neural networks be developed in the future, with a focus on incorporating exogenous variables such as seasonal fluctuations and cultural and tourism policy adjustments, to dynamically predict the spatial use intensity and preference transfer patterns of the three groups. This model could decompose historical data into time series and strengthen the weights of key influencing factors such as holiday effects through

attention mechanisms, thereby providing a quantitative basis for forward-looking adjustments in facility layout.

It is recommended that the spatial analysis function of GIS be integrated with the clustering results and association rule database of this study, and an interactive platform with a scene simulation function be developed. The core module should include a user behavior simulation engine and a visual decision-making interface. The user behavior simulation engine could output the marginal impact on satisfaction of different groups by inputting planning variables such as adding cultural nodes and optimizing accessible paths. The visual decision-making interface could present the spatial spillover effects of intervention measures through heat maps and network correlation graphs, thereby providing a verifiable evidence chain for planners.

It is recommended to deploy IoT sensor networks in key public spaces along the Grand Canal to collect real-time environmental parameters such as cleanliness, light intensity, and crowd density data. By training lightweight machine learning classifiers to perform real-time parsing of data, an automatic warning mechanism could be constructed. When cleanliness is deemed to be below the threshold or when the population density exceeds the safe range, the system could trigger a graded response, and through measures such as scheduling cleaning personnel and initiating flow restriction guidance, achieve a management mode transition from passive response to active intervention.

These tools and strategies would transform static planning into adaptive, user-centric governance, ensuring public spaces along the Grand Canal balance heritage preservation, functional efficiency, and diverse user needs. They will promote the evolution of public space planning for the Grand Canal from a static blueprint to an adaptive governance system, ultimately achieving collaborative optimization of cultural heritage protection, functional efficiency improvement, and meeting diverse user needs.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the [patients/participants OR patients/participants legal guardian/next of kin] was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

JZ: Software, Methodology, Writing – review and editing, Conceptualization, Writing – original draft, Funding acquisition. YJ: Conceptualization, Software, Writing – original draft, Investigation. XZ: Writing – review and editing, Data curation, Visualization,

Funding acquisition. QY: Writing – original draft, Investigation. QZ: Writing – original draft, Visualization. XW: Data curation, Validation, Writing – original draft. LW: Writing – review and editing, Validation, Methodology.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was financially supported by the Shandong Province Social Science Planning Research Project (No. 23CLYJ03).

Acknowledgments

The authors would also like to thank the respondents to surveys conducted in Dezhou, Linqing, Liaocheng, and Jining for their generous support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Al-Shehari, T., and Alsowail, R. A. (2021). An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques. *Entropy* 23 (10), 1258. doi:10.3390/e23101258
- Aljuaydi, F., Wiwatanapatapee, B., and Wu, Y. H. (2023). Multivariate machine learning-based prediction models of freeway traffic flow under non-recurrent events. *Alexandria Eng. J.* 65, 151–162. doi:10.1016/j.aej.2022.10.015
- Alkhereibi, A. H., Abulibdeh, R., and Abulibdeh, A. (2025). Global smart cities classification using a machine learning approach to evaluating livability, technology, and sustainability performance across key urban indices. *J. Clean. Prod.* 503, 145394. doi:10.1016/j.jclepro.2025.145394
- Almukhalifi, H., Noor, A., and Noor, T. H. (2024). Traffic management approaches using machine learning and deep learning techniques: a survey. *Eng. Appl. Artif. Intell.* 133, 108147. doi:10.1016/j.engappai.2024.108147
- Altay, E. V., and Alatas, B. (2020). Intelligent optimization algorithms for the problem of mining numerical association rules. *Phys. a-Statistical Mech. Its Appl.* 540, 11. doi:10.1016/j.physa.2019.123142
- Alwahedi, F., Aldhaheri, A., Ferrag, M. A., Battah, A., and Tihanyi, N. (2024). Machine learning techniques for IoT security: current research and future vision with generative AI and large language models. *Internet Things Cyber-Physical Syst.* 4, 167–185. doi:10.1016/j.iotcps.2023.12.003
- Bagirov, A. M., Aliguliyev, R. M., and Sultanova, N. (2023). Finding compact and well-separated clusters: clustering using silhouette coefficients. *Pattern Recognit.* 135, 109144. doi:10.1016/j.patcog.2022.109144
- Balogun, A.-L., Tella, A., Baloo, L., and Adebisi, N. (2021). A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Clim.* 40, 100989. doi:10.1016/j.uclim.2021.100989
- Barreda Luna, A. A., Kuri, G. H., Rodríguez-Reséndiz, J., Zamora Antuñano, M. A., Altamirano Corro, J. A., and Paredes-García, W. J. (2022). Public space accessibility and machine learning tools for street vending spatial categorization. *J. Maps* 18 (1), 43–52. doi:10.1080/17445647.2022.2035836
- Bashir, S. (2020). An efficient pattern growth approach for mining fault tolerant frequent itemsets. *Expert Syst. Appl.* 143, 113046. doi:10.1016/j.eswa.2019.113046
- Cabigiosu, A. (2025). The adoption of electric vehicles in public transport services: space, path dependency and embeddedness: the venice case. *Technol. Forecast. Soc. Change*, 216. doi:10.1016/j.techfore.2025.124132
- Cao, Y. X., Heng, C. K., and Fung, J. C. (2019). Using walk-along interviews to identify environmental factors influencing older adults' out-of-home behaviors in a high-rise, high-density neighborhood. *Int. J. Environ. Res. Public Health* 16 (21), 4251. doi:10.3390/ijerph16214251
- Chen, S., Lin, S., Yao, Y., and Zhou, X. (2024). Urban public space safety perception and the influence of the built environment from a female perspective: combining street view data and deep learning. *Land* 13 (12), 2108. doi:10.3390/land13122108
- Cheng, S. Y., Zhai, Z. R., Sun, W. Z., Wang, Y., Yu, R., and Ge, X. Y. (2022). Research on the satisfaction of beijing waterfront green space landscape based on social media data. *Land* 11 (10), 1849. doi:10.3390/land11101849
- D'Amico, A., Sparvoli, G., Bernardini, G., Bruno, S., Fatiguso, F., Curra, E., et al. (2024). Behavioural-based risk of the built environment: Key performance indicators for sudden-onset disaster in urban open spaces. *Int. J. Disaster Risk Reduct.* 103, 27. doi:10.1016/j.ijdr.2024.104328
- Dabra, A., and Kumar, V. (2023). Evaluating green cover and open spaces in informal settlements of mumbai using deep learning. *Neural Comput. and Appl.* 35 (16), 11773–11788. doi:10.1007/s00521-023-08320-7
- Dadpour, S., Pakzad, J., and Khankeh, H. (2016). Understanding the influence of environment on adults' walking experiences: a meta-synthesis study. *Int. J. Environ. Res. Public Health* 13 (7), 731. doi:10.3390/ijerph13070731
- Das, P., and Maitra, S. (2024). Priority areas of intervention for improving pedestrian infrastructure and facilities at tourist destinations in India. *Transp. Policy* 145, 126–136. doi:10.1016/j.tranpol.2023.10.018
- Delnevo, G., Mirri, S., Prandi, C., and Manzoni, P. (2023). An evaluation methodology to determine the actual limitations of a TinyML-based solution. *Internet Things* 22, 100729. doi:10.1016/j.iot.2023.100729
- Dirsehan, T. (2024). Why do citizens not prefer to use e-scooters? Views of the public in the Netherlands. *Travel Behav. Soc.* 37, 100863. doi:10.1016/j.tbs.2024.100863
- Du, L. J., Huang, F. S., Lu, H., Chen, S. J., and Guo, Q. W. (2024). An association rule mining-based modeling framework for characterizing urban road traffic accidents. *Sustainability* 16 (23), 10597. doi:10.3390/su162310597

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbuil.2025.1643104/full#supplementary-material>

- Durowoju, O. S., Obateru, R. O., Adelabu, S., and Olusola, A. (2025). Urban change detection: assessing biophysical drivers using machine learning and google Earth engine. *Environ. Monit. Assess.* 197 (4), 441. doi:10.1007/s10661-025-13863-4
- Elsner, J. (2023). Taming the panda with python: a powerful duo for seamless robotics programming and integration. *Software* 24, 101532. doi:10.1016/j.softx.2023.101532
- Faka, A., Kalogeropoulos, K., Maloutas, T., and Chalkias, C. (2021). Urban quality of life: spatial modeling and indexing in Athens metropolitan area, Greece. *Isprs Int. J. Geo-Information* 10 (5), 347. doi:10.3390/ijgi10050347
- Ferreira, Z., Almeida, B., Costa, A. C., Fernandes, M. D., and Cabral, P. (2025). Insights into landslide susceptibility: a comparative evaluation of multi-criteria analysis and machine learning techniques. *Geomatics Nat. Hazards and Risk* 16 (1), 2471019. doi:10.1080/19475705.2025.2471019
- Flemsæter, F., Stokowski, P., and Frisvoll, S. (2020). The rhythms of canal tourism: synchronizing the host-visitor interface. *J. Rural Stud.* 78, 199–210. doi:10.1016/j.jrurstud.2020.06.010
- Fu, L., Fu, H., and Xiong, C. (2025). Evaluating perceived cultural ecosystem services in urban green spaces using big data and machine learning: insights from fragrance hill park in Beijing, China. *Sustainability* 17 (4), 1725. doi:10.3390/su17041725
- García-Moreno, F. M., Alcaraz, J. C., Del Castillo De La Fuente, J. M., Rodríguez-Simón, L. R., and Hurtado-Torres, M. V. (2024). ARTDET: machine learning software for automated detection of art deterioration in easel paintings. *Software* 28, 101917. doi:10.1016/j.softx.2024.101917
- Hamarash, M. Q., Ibrahim, R., Yaas, M. H., Abdulghani, M. F., and Al Mushhadany, O. (2024). Comparative effectiveness of health communication strategies in nursing: a mixed methods study of internet, mHealth, and social media Versus traditional methods. *JMIR Nurs.* 7, e55744. doi:10.2196/55744
- Harris, S., and De Amorim, R. C. (2022). An extensive empirical comparison of *k*-means initialization algorithms. *Ieee Access* 10, 58752–58768. doi:10.1109/access.2022.3179803
- Jentzer, J. C., Patel, S., Gajic, O., Herasevich, V., Lopez-Jimenez, F., Murphree, D. H., et al. (2025). Early prediction of shock in intensive care unit patients by machine learning using discrete electronic health record data. *J. Crit. Care* 88, 155093. doi:10.1016/j.jcrc.2025.155093
- Jiang, A. H., Cai, J., Chen, F. L., Zhang, B. L., Wang, Z. W., Xie, Q. Y., et al. (2022). Sustainability assessment of cultural heritage in Shandong Province. *Sustainability* 14 (21), 13961. doi:10.3390/su142113961
- Jin, W., Wang, F., Chen, L., and Zhang, W. (2025). Machine learning-assisted synthetic biology of Cyanobacteria and microalgae. *Algal Res.* 86, 103911. doi:10.1016/j.algal.2025.103911
- Joseph, A., Choi, Y. S., and Quan, X. B. (2016). Impact of the physical environment of residential health, care, and support facilities (RHCSF) on staff and residents: a systematic review of the literature. *Environ. Behav.* 48 (10), 1203–1241. doi:10.1177/0013916515597027
- Karacor, E. K., and Parlar, G. (2017). Conceptual model of the relationship between neighbourhood attachment, collective efficacy and open space quality. *Open House Int.* 42 (2), 68–74. doi:10.1108/ohi-02-2017-b0010
- Kim, E. J., Park, S. M., and Kang, H. W. (2022). Changes in leisure activities of the elderly due to the COVID-19 in Korea. *Front. Public Health* 10, 966989. doi:10.3389/fpubh.2022.966989
- Klimo, M., Lukác, P., and Tarábek, P. (2021). Deep neural networks classification via binary error-detecting output codes. *Appl. Sciences-Basel* 11 (8), 18. doi:10.3390/app11083563
- Koutra, S., and Ioakimidis, C. S. (2023). Unveiling the potential of machine learning applications in urban planning challenges. *Land* 12 (1), 83. doi:10.3390/land12010083
- Lak, A., Aghamolaei, R., Baradaran, H. R., and Myint, P. K. (2020). A framework for elder-friendly public open spaces from the Iranian older adults' perspectives: a mixed-method study. *Urban For. and Urban Green.* 56, 126857. doi:10.1016/j.ufug.2020.126857
- Le, Q. H., Kwon, N., Nguyen, T. H., Kim, B., and Ahn, Y. (2024). Sensing perceived urban stress using space syntactical and urban building density data: a machine learning-based approach. *Build. Environ.* 266. doi:10.1016/j.buildenv.2024.112054
- Lee, J. (2024). Behind and beyond the standard(ization) trap: diversifying power sources by differentiating centralization and standardization. *Int. Rev. Public Adm.* 29 (1), 1–19. doi:10.1080/12294659.2024.2310899
- Li, L. (2023). Cultural communication and diversity along the grand canal of China: a case study of folk songs in intangible cultural heritage. *Herit. Sci.* 11 (1), 66. doi:10.1186/s40494-023-00911-w
- Liu, Y., and Li, G. (2025). Inequities in thermal comfort and urban blue-green spaces cooling: an explainable machine learning study across residents of different socioeconomic statuses in hangzhou, China. *Sustain. Cities Soc.* 127, 106427. doi:10.1016/j.scs.2025.106427
- Liu, Z., Huang, L., Fan, C., and Mostafavi, A. (2025). Generating equitable urban human flows with a fairness-aware deep learning model. *Cities* 167, 106296. doi:10.1016/j.cities.2025.106296
- Luo, L., and Miao, X. M. (2023). Retracted article: study on the exploration of poverty index's association rules based on CBCM-apriori algorithm. *Ann. Operations Res.* 326 (Suppl. 1), 157. doi:10.1007/s10479-022-04607-5
- Luo, Y. Y., Liu, Y. F., Xing, L. J., Wang, N. N., and Rao, L. (2022). Road safety evaluation framework for accessing park green space using active travel. *Front. Environ. Sci.* 10, 864966. doi:10.3389/fenvs.2022.864966
- Macieira, T. G. R., Yao, Y. W., Marcelle, C., Mena, N., Mino, M. M., Huynh, T. M. L., et al. (2024). Standardizing nursing data extracted from electronic health records for integration into a statewide clinical data research network. *Int. J. Med. Inf.* 183, 105325. doi:10.1016/j.ijmedinf.2023.105325
- Mahmoud, A. M. A., Sheldrick, N., and Ahmed, M. (2025). A novel machine learning automated change detection tool for monitoring disturbances and threats to archaeological sites. *Remote Sens. Appl. Soc. Environ.* 37, 101396. doi:10.1016/j.rsase.2024.101396
- Manta, S. W., Reis, R. S., Benedetti, T. R. B., and Rech, C. R. (2019). Public open spaces and physical activity: disparities of resources in florianopolis. *Rev. De. Saude Publica* 53, 112. doi:10.11606/s1518-8787.2019053001164
- Marcondes, D., Simonis, A., and Barrera, J. (2018). Feature selection based on the local lift dependence scale. *Entropy* 20 (2), 97. doi:10.3390/e20020097
- Marshall, A. J., Grose, M. J., and Williams, N. S. G. (2019). From little things: more than a third of public green space is road verge. *Urban For. and Urban Green.* 44, 126423. doi:10.1016/j.ufug.2019.126423
- Maté-Sánchez-Val, A., Hernández, F. a. L., and Maté-Sánchez-Val, M. (2025). City for people, city for cars. Analyzing maximum walkability areas through machine learning algorithms and open-source data. *Cities* 162, 105895. doi:10.1016/j.cities.2025.105895
- Medina, A., Mosquera, D., and Gallegos, F. A. (2023). A Methodological Approach for Data Collection and Geospatial Information of Healthy Public Spaces in Peripheral Neighborhoods—Case Studies: la Bota and Toctiucio, Quito, Ecuador. *Sustainability* 15 (21), 15553. doi:10.3390/su152115553
- Michalak, M. (2024). Searching for continuous n-Clusters with boolean reasoning. *Symmetry-Basel* 16 (10), 1286. doi:10.3390/sym16101286
- Motomura, M., Koohsari, M. J., Lin, C. Y., Ishii, K., Shibata, A., Nakaya, T., et al. (2022). Associations of public open space attributes with active and sedentary behaviors in dense urban areas: a systematic review of observational studies. *Health Place* 75, 102816. doi:10.1016/j.healthplace.2022.102816
- Nourbakhsh, A., Jadidi, M., and Shahriari, K. (2024). Clustering bike sharing stations using quantum machine learning: a case study of Toronto, Canada. *Transp. Res. Interdiscip. Perspect.* 27, 101201. doi:10.1016/j.trip.2024.101201
- Pérez Moreno, F., Gómez Comendador, V. F., Delgado-Aguilera Jurado, R., Zamarreño Suárez, M., Janisch, D., and Arnaldo Valdés, R. M. (2023). Methodology of air traffic flow clustering and 3-D prediction of air traffic density in ATC sectors based on machine learning models. *Expert Syst. Appl.* 223, 119897. doi:10.1016/j.eswa.2023.119897
- Piras, G., Muzi, F., and Tiburcio, V. A. (2024). Enhancing space management through digital twin: a case study of the Lazio region headquarters. *Appl. Sci.* 14 (17), 7463. doi:10.3390/app14177463
- Plunz, R. A., Zhou, Y., Carrasco Vintimilla, M. I., Mckeown, K., Yu, T., Ugucioni, L., et al. (2019). Twitter sentiment in New York city parks as measure of well-being. *Landsc. Urban Plan.* 189, 235–246. doi:10.1016/j.landurbplan.2019.04.024
- Poleykett, B. (2022). A broom to the head: 'cleaning Day' and the aesthetics of emergence in Dakar. *Urban Stud.* 59 (2), 381–396. doi:10.1177/0042098021993357
- Pourbahador, P., and Brinkhuijsen, M. (2023). Municipal strategies for protecting the sense of place through public space management in historic cities: a case study of amsterdam. *Cities* 136, 104242. doi:10.1016/j.cities.2023.104242
- Rai, K., Wang, Y., O'connell, R. W., Patel, A. B., and Bashor, C. J. (2024). Using machine learning to enhance and accelerate synthetic biology. *Curr. Opin. Biomed. Eng.* 31, 100553. doi:10.1016/j.cobme.2024.100553
- Rajagopal, A., Ayanian, S., Ryu, A. J., Qian, R., Legler, S. R., Peeler, E. A., et al. (2024). Machine learning operations in health care: a scoping review. *Mayo Clin. Proc. Digit. Health* 2 (3), 421–437. doi:10.1016/j.mcpdig.2024.06.009
- Ramírez, T., Hurtubia, R., Lobel, H., and Rossetti, T. (2021). Measuring heterogeneous perception of urban space with massive data and machine learning: an application to safety. *Landsc. Urban Plan.* 208, 104002. doi:10.1016/j.landurbplan.2020.104002
- Randrup, T. B., Svänell, J., Sunding, A., Jansson, M., and Sang, Å. (2021). Urban open space management in the nordic countries. Identification of current challenges based on managers' perceptions. *Cities* 115, 12. doi:10.1016/j.cities.2021.103225
- Riva, M., Kienast, F., and Grêt-Regamey, A. (2024). Mapping open spaces in Swiss Mountain regions through consensus-building and machine learning. *Appl. Geogr.* 165, 13. doi:10.1016/j.apgeog.2024.103237
- Robi, R. K., and George, J. K. (2025). Application of machine learning algorithms to predict urban expansion. *J. Urban Plan. Dev.* 151 (2), 03125001. doi:10.1061/jupddm.upeng-5466

- Rossetti, T., Lobel, H., Rocco, V., and Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: a massive data approach. *Landsc. Urban Plan.* 181, 169–178. doi:10.1016/j.landurbplan.2018.09.020
- Rui, J., and Othengrafen, F. (2023). Examining the role of innovative streets in enhancing urban mobility and livability for sustainable urban transition: a review. *Sustainability* 15 (7), 5709. doi:10.3390/su15075709
- Saraiva, M., Matijošaitienė, I., Mishra, S., and Amante, A. (2022). Crime prediction and monitoring in Porto, Portugal, using machine learning, spatial and text analytics. *Isprs Int. J. Geo-Information* 11 (7), 400. doi:10.3390/ijgi11070400
- Sas-Bojarska, A., Orzechowska-Szajda, I., Puzdrakiewicz, K., and Kiejzik-Głowinska, M. (2024). Landscape, EIA and decision-making. A case study of the vistula spit canal, Poland. *Impact Assess. Proj. Apprais.* 42 (1), 2–29. doi:10.1080/14615517.2023.2273612
- Schmitt, R. H., Wolfschla, D., Woltersmann, J. H., and Stohrer, L. (2024). Measurability of quality characteristics identified in latent spaces of generative AI models. *Cirp Annals-Manufacturing Technol.* 73 (1), 389–392. doi:10.1016/j.cirp.2024.04.073
- Senik, B., and Uzun, O. (2022). A process approach to the open green space system planning. *Landsc. Ecol. Eng.* 18 (2), 203–219. doi:10.1007/s11355-021-00492-5
- Shi, C. M., Wei, B. T., Wei, S. L., Wang, W., Liu, H., and Liu, J. L. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Eurasip J. Wirel. Commun. Netw.* 2021 (1), 31. doi:10.1186/s13638-021-01910-w
- Sinaga, K. P., and Yang, M. S. (2020). Unsupervised K-Means clustering algorithm. *Ieee Access* 8, 80716–80727. doi:10.1109/access.2020.2988796
- Sletto, B., and Palmer, J. (2017). The liminality of open space and rhythms of the everyday in jallah town, Monrovia, Liberia. *Urban Stud.* 54 (10), 2360–2375. doi:10.1177/0042098016643475
- Smith, L., Burgoine, T., Ogilvie, D., Jones, A., Coombes, E., and Panter, J. (2023). Demonstrating the applicability of using GPS and interview data to understand changes in use of space in response to new transport infrastructure: the case of the Cambridgeshire Guided Busway, UK. *J. Transp. and Health* 30, 101620. doi:10.1016/j.jth.2023.101620
- Song, X. Y., Du, L. Z., and Wang, Z. Y. (2024). Correlation analysis of urban road network structure and spatial distribution of tourism service facilities at multi-scales based on tourists' travel preferences. *Buildings* 14 (4), 914. doi:10.3390/buildings14040914
- Ta, D. T., and Furuya, K. (2022). Google street view and machine learning—useful tools for a street-level remote survey: a case study in Ho chi minh, Vietnam and ichikawa, Japan. *Land* 11 (12), 2254. doi:10.3390/land11122254
- Ubani, O. J., Alabi, M. O., Chiemelu, E. N., Okosun, A., and Sam-Amobi, C. (2023). Influence of spatial accessibility and environmental quality on youths' visit to green open spaces (GOS) in akure, Nigeria. *Sustainability* 15 (17), 13223. doi:10.3390/su151713223
- Valenzuela-Levi, N., Gálvez Ramírez, N., Nilo, C., Ponce-Méndez, J., Kristjanpoller, W., Zúñiga, M., et al. (2024). A cyborg walk for urban analysis? From existing walking methodologies to the integration of machine learning. *Land* 13 (8), 1211. doi:10.3390/land13081211
- Vera, C., Lucchini, F., Bro, N., Mendoza, M., Loebel, H., Gutierrez, F., et al. (2022). Learning to cluster urban areas: two competitive approaches and an empirical validation. *Epj Data Sci.* 11 (1), 62. doi:10.1140/epjds/s13688-022-00374-2
- Wang, Z. G., and Stevens, Q. (2020). How do open space characteristics influence open space use? A study of Melbourne's southbank promenade. *Urban Res. and Pract.* 13 (1), 22–44. doi:10.1080/17535069.2018.1484152
- Wang, Q., Lu, M., and Li, Q. Q. (2020). Unfolding the city: spatial preference based on individual demographic characteristics. *Ieee Access* 8, 43455–43465. doi:10.1109/access.2020.2977673
- Wilson, J. S., and Kelly, C. M. (2011). Measuring the quality of public open space using google Earth. *Am. J. Prev. Med.* 40 (2), 276–277. doi:10.1016/j.amepre.2010.11.002
- Yang, E. J., Fulton, J., Swarnaraja, S., and Carson, C. (2023). Machine learning to support citizen science in urban environmental management. *Heliyon* 9 (12), e22688. doi:10.1016/j.heliyon.2023.e22688
- Yang, L., Yang, H., Yu, B., Lu, Y., Cui, J., and Lin, D. (2024). Exploring non-linear and synergistic effects of green spaces on active travel using crowdsourced data and interpretable machine learning. *Travel Behav. Soc.* 34, 100673. doi:10.1016/j.tbs.2023.100673
- Yao, T. N., Xu, Y., Sun, L., Liao, P., and Wang, J. (2024). Application of machine learning and multi-dimensional perception in urban spatial quality evaluation: a case study of shanghai underground pedestrian street. *Land* 13 (9), 1354. doi:10.3390/land13091354
- Yu, J., Zhong, H., and Kim, S. B. (2020). An ensemble feature ranking algorithm for clustering analysis. *J. Classif.* 37 (2), 462–489. doi:10.1007/s00357-019-09330-8
- Yuen, J. W. M., Chang, K. K. P., Wong, F. K. Y., Wong, F. Y., Siu, J. Y. M., Ho, H. C., et al. (2019). Influence of urban green space and facility accessibility on exercise and healthy diet in Hong Kong. *Int. J. Environ. Res. Public Health* 16 (9), 1514. doi:10.3390/ijerph16091514
- Zhang, Y., Liu, Y., Li, Y., Chu, J., and Yang, Q. (2025). Nonlinear and spatial non-stationary effects of land finance on urban expansion at the county level in China: insights from explainable spatial machine learning. *Cities* 160, 105850. doi:10.1016/j.cities.2025.105850
- Zhao, J., Wang, L., Ye, Q., Zhao, Q., and Wei, S. (2022). Association of environmental elements with respondents' behaviors in open spaces using the direct gradient analysis method: a case study of jining, China. *Int. J. Environ. Res. Public Health* 19 (14), 8494. doi:10.3390/ijerph19148494
- Zhao, K., Guo, D., Sun, M., Zhao, C., Shuai, H., and Shao, C. (2024). Short-term traffic flow prediction based on hybrid decomposition optimization and deep extreme learning machine. *Phys. A Stat. Mech. its Appl.* 647, 129870. doi:10.1016/j.physa.2024.129870
- Zhu, Y., Zhang, Y., and Biljecki, F. (2025). Understanding the user perspective on urban public spaces: a systematic review and opportunities for machine learning. *Cities* 156, 105535. doi:10.1016/j.cities.2024.105535