


## Article

# Study on the Influence Mechanism of Machine-Learning-Based Built Environment on Urban Vitality in Macau Peninsula

Chen Pan <sup>1,2</sup> , Jiaming Guo <sup>1,\*,†</sup>, Haibo Li <sup>1,\*,†</sup>, Jiawei Wu <sup>1</sup>, Nengjie Qiu <sup>1</sup> and Shengzhen Wu <sup>3</sup>

<sup>1</sup> Architecture and Civil Engineering Institute, Guangdong University of Petrochemical Technology, Maoming 525000, China; jourdan29@gdupt.edu.cn (C.P.); wayne@gdupt.edu.cn (J.W.); qnj906@gdupt.edu.cn (N.Q.)

<sup>2</sup> Faculty of Innovation and Design, City University of Macau, Macau 999078, China

<sup>3</sup> College of Arts and Design, Jimei University, Xiamen 361021, China; 2361000218@jmu.edu.cn

\* Correspondence: gjm11@gdupt.edu.cn (J.G.); hbl@gdupt.edu.cn (H.L.)

† These authors contributed equally to this work.

**Abstract:** Clarifying the mechanisms by which the micro-scale built environment influences urban vitality is an important scientific challenge, to guide precise urban planning in the context of urban renewal. In this study, we quantify the intensity of human activities through Baidu heat maps, analyze social interaction patterns using social media check-in data, and integrate built environment elements such as road network topology, 3D building morphology, and the spatial distribution of points of interest (POIs). A machine learning technique combining a real-encoded Accelerated Genetic Algorithm-Projective Pathfinding Model (RAGA-PPM) and Shapley Additive Projection for Interpretability (SHAP) for Interpretability Analysis (IPA) was used to investigate the nonlinear mechanisms of 17 factors affecting urban vitality in Macau Peninsula, China. Firstly, the explanatory power of the built environment for comprehensive vitality was significantly better than the other dimensions. Two factors, population vitality and microblogging check-in vitality, contributed the most to the composite vitality value. Secondly, road network density was the most important built environment factor affecting urban vitality in Macau Peninsula (SHAP = 0.025). Finally, the impacts of built environment factors on urban vitality showed nonlinearities, and the threshold effects of the core factors (road network density, spatial fractal dimension, and openness to the sky) showed a consistent neighborhood-level pattern. This study establishes a framework for micro-vitality mechanisms in high-density cities, addressing the limitations of traditional methods in modeling complex nonlinear relationships. The methodological integration of RAGA-PPM and SHAP advances the innovative paradigm of applying interpretable machine learning to the study of urban form.



Academic Editor: Paulo Santos

Received: 8 April 2025

Revised: 1 May 2025

Accepted: 3 May 2025

Published: 5 May 2025

**Citation:** Pan, C.; Guo, J.; Li, H.; Wu, J.; Qiu, N.; Wu, S. Study on the Influence Mechanism of Machine-Learning-Based Built Environment on Urban Vitality in Macau Peninsula. *Buildings* **2025**, *15*, 1557. <https://doi.org/10.3390/buildings15091557>

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

**Keywords:** built environment; urban vitality; projection pursuit; genetic algorithm; Macau Peninsula

## 1. Introduction

With the acceleration of global urbanization, the relationship between urban vitality and the built environment has increasingly become a crucial research topic in the fields of urban planning and management. Urban vitality generally refers to the comprehensive performance of a city in economic, social, cultural, and environmental aspects, directly influencing residents' quality of life, social interactions, and economic development levels [1]. The built environment, as a core component supporting urban functions, encompasses multiple dimensions such as urban spatial structure, transportation facilities, green space

distribution, and building density. Its quality and layout play a vital role in the generation and development of urban vitality [2]. Traditional research has often simplified the relationship between urban vitality and the built environment into linear correlations, employing statistical methods like regression analysis for exploration [3]. However, the relationship between urban vitality and the built environment is inherently nonlinear, influenced by the intertwining and interaction of various factors, and this interaction is highly complex [4]. Existing linear models fail to effectively capture this complex nonlinearity, leading to certain limitations in the prediction and optimization of urban vitality. In the context of sustainable urban development, there is an urgent need for more refined research on how to balance high-density development with ecological preservation, optimize transportation networks to reduce carbon emissions, and enhance environmental comfort. Although existing studies have recognized the nonlinear association between the built environment and urban vitality, there are still significant gaps in the analysis of specific nonlinear mechanisms of action. Complex relationships such as the inverse inhibitory characteristics of built environment elements for urban vitality in the threshold effect, the differential response mechanism of the same element for vitality under spatial heterogeneity, as well as the delayed and cumulative impacts of environmental changes on vitality in the dynamic time lag effect have not yet been adequately quantified and deciphered [5]. Traditional linear models have difficulty capturing multidimensional nonlinear features, due to the restriction of fixed coefficient assumptions, leading to insufficient guidance on key issues such as synergistic paths for high-density development and ecological protection, marginal benefit thresholds for transportation network optimization, and scale effects of green infrastructure implementation. Therefore, exploring a modeling method that can systematically identify and reveal the multiple types of nonlinear relationships between the built environment and urban vitality is an important direction, to promote the progress of urban research theory and practice.

In recent years, with the rapid development of artificial intelligence and big data technologies, modeling methods based on complex systems have gradually become important tools in urban research [6,7]. In particular, the Real-coded Adaptive Genetic Algorithm (RAGA) can achieve global optimization in complex, multidimensional, and multivariate data, while the Probabilistic Potential Model (PPM) reveals deep-seated correlations among various factors in urban systems through latent factor modeling [8]. Although existing research has made some progress in theoretical construction and model application, there has been a lack of exploration into urban vitality measurement based on the RAGA-PPM model [9]. The integration of machine learning algorithms with Shapley Additive Explanations (SHAP) provides intuitive interpretability for complex nonlinear models. By quantifying the contribution of each feature to the model's output, it can reveal how built environment elements nonlinearly influence urban vitality, further enhancing the model's transparency and practicality [10].

In this study, we aim to construct a hybrid computational framework capable of characterizing the nonlinear dynamics of the built environment and urban vitality by coupling the global optimization capability of the real-encoded adaptive genetic algorithm (RAGA) with the potential factor inference advantages of the probabilistic potential model (PPM). On this basis, SHAP is introduced to decouple the differentiated paths of the contribution of multi-dimensional built environment elements to vitality generation and their interaction effects. Based on multi-source high-precision urban data, a cross-scale empirical study was carried out to quantitatively verify the superiority of the model for vitality prediction accuracy, generation of spatial intervention strategies, and resolution of nonlinear relationships, forming a complete closed loop of "theory construction–mechanism analysis–decision-making support", and providing a combination of machine learning

and explanatory planning and decision-making capabilities for the optimization of the built environment in the context of a smart city. We provide a methodological tool for the optimization of built environment in the context of smart cities, which combines machine learning, sophistication, and interpretability of planning and decision-making.

## 2. Literature Review

The term “vitality” in Chinese was originally used to describe and express the state of activity of living organisms. Later, it was understood as the vigorous vitality of living organisms or the activity of individual organisms in their actions, communication and expression, and thinking. The term “vitality” was initially mainly applied in the field of physics. With the research and exploration of the natural world, it further was expanded to the field of biology. Later, it was introduced into various scientific research fields and has also developed different connotations in various research processes within different disciplinary systems. Ultimately, the term “vitality” has a rich meaning that transcends its original meaning, and there are certain differences in its representation and interpretation in different fields and specific research issues. The concept of “vitality” in this study is mainly in reference to cities and urban areas. It mainly represents the “vitality” within the city itself or the urban block space. The definition of “urban vitality” begins with understanding the nature of a city’s continuous operation and regards the overall framework of the city as a complex organic life or physical system. Urban vitality, as a crucial indicator for assessing urban health and sustainable development, has long attracted attention from the academic community. At the macro level, urban statistical data, interviews, and questionnaires have been utilized as data sources [11,12]. However, these data sources are often plagued by issues such as insufficient timeliness, limited sample sizes, and being time-consuming and labor-intensive, making them difficult to apply to fine-grained perspectives. At the meson and micro levels, Jacob’s explanatory indicator framework, which approaches from various dimensions, has been widely accepted [13–15]. Scholars have explored urban vitality from aspects such as urban morphology, design, land use, population density, mobility, and transportation accessibility [16–21]. Most research on urban vitality has focused on the exploration of indicators and their impact levels, with relatively limited studies on the formation mechanisms of vitality itself and comprehensive evaluations. Furthermore, more granular indicators have also not been fully developed [22,23].

### 2.1. Research on the Relationship Between Urban Vitality and Built Environment

Early studies on urban vitality primarily focused on the role of urban economic development, social structure optimization, and infrastructure improvement in promoting urban vitality. Scholars have extensively investigated the impact of built environment variables such as traffic density [1], green space [14], public facility provision [20], and building density [24] on urban vitality through regression analysis and spatial econometric models [24]. In her seminal work *“The Death and Life of Great American Cities”*, Jacobs defined urban vitality as a self-regenerating force capable of nurturing and promoting diverse pedestrian activities and the connections between neighborhood and street-level activities [25].

With the rapid advancement of information technology, a new data environment centered around big data and open data has gradually emerged. The availability of diverse data sources, including GPS data, mobile phone signaling, social media check-ins, and street view images, has provided richer possibilities for measuring urban vitality, thereby expanding the depth and breadth of built environment research [26–32]. Building upon this foundation, studies conducted at various urban scales—from administrative districts [30,33,34] to neigh-

borhoods [12,35], streets [36–38], and public spaces [39,40]—have consistently demonstrated the significant role of the built environment in fostering urban vitality.

## 2.2. Application of Nonlinear Models and Complex Systems Approache

Existing studies have mainly used linear regression models, spatial measurement models, and geographically weighted regression models to analyze the effects of built environment factors on urban vitality and their significance [20,31,33,34,36,39–41]. Although these methods can verify the statistical significance among variables, their linear assumptions have difficulty capturing the nonlinear nature of multi-factor interactions within urban systems. In recent years, scholars have gradually introduced complex modeling techniques to overcome this limitation. For example, the adaptive genetic algorithm (RAGA) optimizes the global solution of a high-dimensional nonlinear problem by simulating biological evolutionary mechanisms (e.g., selection, crossover, and mutation), which significantly improves the efficiency of parameter finding [42]. Its advantage lies in its ability to process multi-objective optimization tasks in parallel and adaptively adjust the search interval to approximate the optimal solution [43]. However, the RAGA is sensitive to the initial parameter settings, and the computational complexity increases exponentially with the increase in the dimensionality of the variables, which may lead to a decrease in the convergence speed, or falling into a local optimum [42,44]. In addition, its “black-box” nature limits the interpretability of the model results, making it difficult to directly guide the planning practice. Gradient boosted decision tree (GBDT) is an iterative approach that builds weak classifiers and weighted combinations by integrating learning strategies, and it has been shown to perform well in identifying the threshold effects and interactions of built environments for urban vitality [35,44–47]. The GBDT method is particularly suitable for multi-source heterogeneous urban big data analysis, as it can deal with multiple covariations between features under loose assumptions on data distribution. However, its limitations are equally significant: on the one hand, the model tends to over-rely on high-frequency features and may ignore low-frequency but high-impact variables; on the other hand, the complexity of the tree structure undermines the interpretability of the results, and although methods such as SHAP (Shapley Additive Properties for Interpretation) have been used for post hoc interpretations, they are computationally costly and rely on sample independence assumptions [48]. In addition, GBDT is more sensitive to outliers and noisy data, which may affect the robustness of threshold determination. Future research could compensate for the shortcomings of a single algorithm through a hybrid modeling framework. The RAGA could be combined with local search algorithms to balance global exploration and local exploitation capabilities; or an attention mechanism could be introduced to improve the feature weight allocation logic of GBDT and enhance the ability to capture sparse features. Meanwhile, the integration path between Explainable ML and Spatially Explicit Modeling should be further explored to reveal the spatial heterogeneity and spillover effect of built environment elements, to provide improved operational theoretical support for refined urban planning.

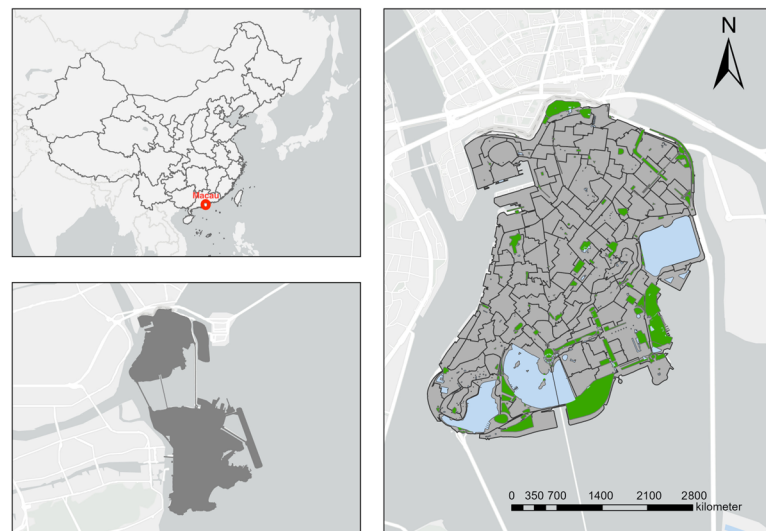
In summary, significant progress has been made in the current literature regarding the exploration of nonlinear relationships between urban vitality and the built environment. However, there remains considerable room for improvement regarding the complexity of model construction and the breadth and depth of multifactor analysis in existing studies. Most research still relies on linear assumptions or simplified variable settings, failing to comprehensively capture the complex interactive effects between the built environment and urban vitality. Future research should focus on developing more sophisticated and refined modeling methods, while emphasizing the integration of interdisciplinary theories, particularly the incorporation of sociological and behavioral economic theories. This will

not only enhance the explanatory power of existing models but also provide crucial support for advancing innovation in urban planning theory and practice.

### 3. Study Area and Data Sources

#### 3.1. Study Area

As one of the three core areas of the Macau Special Administrative Region, the Macau Peninsula is located on the west bank of the Pearl River Delta (latitude  $22^{\circ}11'$  to  $22^{\circ}13'$  N, longitude  $113^{\circ}32'$  to  $113^{\circ}35'$  E), adjacent to the Gongbei Border Crossing of Zhuhai City in the north, and connected to the islands of Taipa and Coloane in the south by crossing the Sai Van Bridge, making it the political, economic, and cultural hub of Macau (Figure 1). According to the 2024 data of the Statistics and Census Bureau of the Macau SAR Government [47], the region has a land area of  $9.1 \text{ km}^2$  and a resident population of about 452,700, with a population density of 49,747 inhabitants/ $\text{km}^2$ , which is among the most densely populated regions in the world. The topographic features are dominated by hilly terrain, with Pine Hill as the highest point (93 m above sea level), and artificial land areas such as the New Border Crossing and Nam Van Lake formed along the coast through reclamation projects, constituting a spatial superposition of the characteristics of a historical urban area and a modern new district. The Macau Peninsula is divided into 159 neighborhoods, among which the Historic District (which concentrates 22 historical building clusters such as the Ruins of St. Paul's, the A-Ma Temple, the Senado Square, etc.) demonstrates the unique urban texture of Sino-Portuguese cultural fusion.



**Figure 1.** Map of the study area.

As an important node of the Guangdong–Hong Kong–Macau Greater Bay Area, the Macau Peninsula has close transportation and economic links with cities such as Zhuhai and Hong Kong through the Hong Kong–Zhuhai–Macau Bridge and the Guangzhou–Zhuhai Intercity Railway. The urban space has undergone a transformation from a single-core agglomeration in the historical city center to a multi-core expansion in the peripheral new reclamation area, forming a “dual-center” structure: first, a historical, cultural, and tourism center with the Senado Square as the core, concentrating 80% of the world cultural heritage and traditional commercial districts, and facing contradictions between heritage preservation and high-density development; and second, a modern business and entertainment center in the Outer Harbor New Reclamation Area. The second area is the modern business and entertainment center of the Outer Harbor New Reclamation Area, which contains integrated resorts such as MGM MACAU and Wynn Palace, but suffers from an imbalance



in daytime and nighttime vitality due to its single function. The transportation network in the area is dense but limited by the narrow street pattern of the historical city center, where walking and public transportation are the main modes of travel. The Macau Peninsula, with its high-density urbanization, cultural diversity, and spatial compactness, has become a typical sample for exploring the non-linear interaction mechanisms between the built environment and urban vitality, and its experience has important theoretical value and practical significance for the sustainable regeneration and vitality enhancement of historic cities around the world.

### 3.2. Variables and Data

#### 3.2.1. Vitality Measurement

The measurement of urban vitality is conducted from three perspectives: human activity, online interaction, and built environment. Among these, human activity serves as a crucial indicator for assessing urban vitality, with the distribution and movement patterns of people being key drivers of regional prosperity [48]. Baidu's heatmap population data provide real-time population density and activity information, revealing areas of high human concentration and dynamic population movements, thereby offering scientific support for urban planning and management. As behavioral records from a mainstream social media platform, Weibo check-in data reflect peak demand and distribution characteristics of users of public facilities and services [46,49–52], unveiling the activity patterns and preferences of both residents and tourists, which provide valuable references for urban operations and decision-making.

#### 3.2.2. Measurement of the Built Environment

In this study, we adopted the “5D” built environment indicator system proposed by Ewing et al. [2], which includes density, diversity, design, destination accessibility, and distance to transit. Building upon previously established indicator systems [33,35,45], and considering data availability and indicator suitability, we selected 15 indicators to measure the built environment across seven dimensions: urban vitality, spatial form, functional characteristics, socio-economic environment, psychological perception, transportation accessibility, and street quality. Given that geographical elements exhibit spatial dependency effects, empirical models using machine learning algorithms are unable to capture the dependency between spatial units or measure the direct effects and spatial spillover effects between variables [53]. The specific calculation methods for each indicator are detailed in Table 1.

**Table 1.** Variable definitions and descriptive statistics.

Variable	Detection Metric	Calculation Formula	Formula Description
Urban Vitality	Population Density Weibo Check-in Density	$R = \frac{H}{A}$	R represents the density value of a specific neighborhood unit; H denotes the kernel density value of a particular neighborhood unit; A is the area of a given neighborhood unit.
Spatial Configuration	Spatial Compactness	$c = \frac{\sqrt[2]{\pi^* A_{era}}}{l}$	C represents the compactness of a block, which is determined by the area of the land parcel within the block unit and the lengths of the boundaries of the parcel.
	Fractal Dimension	$D = \frac{2\ln(\frac{h}{4})}{\ln(A)}$	D value typically ranges between 1.0 and 2.0, with a higher D value indicating greater complexity in the spatial configuration of land parcels within the region.

Table 1. Cont.

Variable	Detection Metric	Calculation Formula	Formula Description
Functional Attribute	Functional Density	$fd_i = \frac{N_i}{S_i}$	$fd_i$ represents the total number of Points of Interest (POI) within the $i$ -th urban block, $S_i$ refers to the area of the $i$ -th category of urban blocks.
	Functional Mix	$I = -\sum_{i=1}^n (p_i * \ln p_i)$	$I$ represents the functional mix of the block, “ $i$ ” denotes the number of POIs within the $i$ -th block unit, and “ $P_i$ ” signifies the ratio of the $i$ -th category of POIs to the total number of POIs within the block unit.
Socio-Economic Environment	Cultural Facility Density	$D = \frac{N}{A}$	$D$ represents the ratio of the total number of cultural facilities, public facilities, and commercial facilities within a specific region to the total area of that region.
	Public Facility Density		
Psychological Perception	Building Density	$P = \frac{\sum_{i=1}^n m_i}{s}$	$P$ represents the building density, $i$ denotes the identification number of different buildings, $m_i$ refers to the base area of the building with the $i$ -th identification number, and $s$ stands for the usable area.
	Hydrophilic index	$WI = K - NER\_DIST$	$WI$ denotes the hydrophilicity index of spatial blocks, while $K$ is a constant term representing the search radius value for neighborhood analysis.
Transportation Accessibility	Road Network Density	$D = \frac{N}{A}$	$D$ represents the ratio of the total number of commercial facilities and bus stops to the total area within the study region.
	Bus Stop Density		
Street Quality	Green View Index	$GVI_i = P_{Vegetation}$	The visibility of vegetation as perceived by the human eye was measured to assess the level of greenery on urban streets.
	sky view factor	$SVF_i = P_{sky}$	The term refers to the proportion of the sky visible within the field of view at a specific location, typically expressed as a value ranging from 0 to 1.
	Enclosure	$Enclosure_i = \frac{P_{Building} + P_{Wall} + P_{Fence} + P_{Pole} + P_{Light} + P_{Sign} + P_{Vegetation}}{P_{Vegetation}}$	The sense of enclosure measures the extent to which an individual feels surrounded by the surrounding environment within a space.
	Walkability	$Walkability_i = \frac{P_{Fence} + P_{Sidewalk}}{P_{road}}$	Refers to the level of pedestrian-friendliness of a street or urban area.

### 3.3. Data Sources and Preprocessing

The data were categorized into two main dimensions: comprehensive urban vitality and built environment. The datasets utilized primarily consisted of fundamental geographic data and open-source data obtained from the Internet, as detailed in Table 2.

Table 2. Data sources and acquisition timeline.

Data Type	Data Name	Data Source	Time	Preprocessing
Fundamental Geospatial Data	Macau Road Network	Tianditu ( <a href="https://www.tianditu.gov.cn/">https://www.tianditu.gov.cn/</a> accessed on 3 April 2024)	2024	The data were uniformly projected onto the WGS84 coordinate system to classify road and building categories, and metrics such as road density, building area, and water system area were subsequently calculated.
	Macau Architectural Structures			
Open-source Web Data	Hydrological Data	Baidu Map Smart Eye	2024	The Kernel Density analysis tool within ArcGIS was employed to convert the data into a continuous raster density map.
	Baidu Heatmap	Population Big Data Platform ( <a href="https://huiyan.baidu.com/">https://huiyan.baidu.com/</a> accessed on 3 April 2024)		
	Population Data	Estate Information Network of Macau ( <a href="https://www.malimalihome.net/">https://www.malimalihome.net/</a> accessed on 3 April 2024)	2023	The core fields, including unit price, total price, gross floor area, and property type, were collected to calculate key metrics such as price per square meter and average housing prices by region.

Table 2. Cont.

Data Type	Data Name	Data Source	Time	Preprocessing
Open-source Web Data	Points of Interest (POI) Data	GaoDe Map	2024	The dataset underwent comprehensive data cleaning, including the handling of missing values and outliers. Classification and standardization processes were implemented to ensure field uniformity across all 45,708 records.
	Weibo Check-in Data	Sina Weibo ( <a href="https://weibo.com/">https://weibo.com/</a> accessed on 3 April 2024)	2020	The data cleaning process involved handling missing values and outliers, followed by categorizing the data based on check-in locations. The total number of Weibo check-ins processed was 145,370. Sampling points were generated at 100 m intervals based on road network data.
	Street View Imagery	Baidu Street View Map Open Platform ( <a href="https://lbs.baidu.com/">https://lbs.baidu.com/</a> accessed on 3 April 2024)	2023	Utilizing the Baidu API endpoint with a pitch angle of 0 degrees, images were captured at four different azimuth angles: 0°, 90°, 180°, and 270°. A total of 10,442 images were collected.

### 3.3.1. Basic Geographic Data

The basic geographic data included the administrative division data of Macau Peninsula, road network data, water system data, building data, and road network data, all of which came from the Tianditu.

### 3.3.2. Open-Source Data

#### (1) Baidu Heat Population Data

Baidu thermal population data were obtained based on the Baidu Map Wise Eye demographic geographic big data platform, whose location information covers almost every aspect of daily life. Compared with “check-in” or census data, this heat map dataset is almost unbiased [32], can provide almost real-time demographic data, and can well demonstrate the direction of urban population movement and spatial aggregation status. The platform includes functions within a certain spatial range to provide online data analysis and visualization services, including statistical and dynamic visualization information such as regional population, occupancy and residence, passenger flow, facilities, brands, ODs, and heat maps. Baidu Map Wise Eye population location data are mainly derived from calling Baidu Map Positioning SDK terminal positioning data. It has been demonstrated that crowd movement conforms to similar distribution and evolution laws based on Baidu heat map data from Monday–Friday (weekdays) and Saturday and Sunday (weekends) [48]. In this paper, the statistical information of the heat map in Baidu Map Wise Eye urban population data platform was used to obtain the data of 24 moments on May 15 (Monday) and May 20 (Saturday), 2024 as the basic data.

#### (2) Macau Property Price Data

Macau house price data were obtained based on the preferred real estate information platform of Macau (Malimali Home), which integrates real-time transaction information of residential and commercial properties in the Macau Peninsula, covering core fields such as the unit price of the property, the total price, the floor area, the type of property ownership, and the time of the transaction. Compared with the government’s periodic statistical reports, such market transaction data can capture the spatial and temporal characteristics of house prices more dynamically and avoid the lagging bias of traditional statistics [47]. The platform provides historical transaction records and regional price indices through a



standardized data interface, and its data cover more than 90% of the real estate agencies in the Macau Peninsula. It has been demonstrated that such data have high reliability and validity in parsing the spatial differentiation of urban settlements and the distribution of commercial vitality [51]. In this study, we extracted transaction data for the whole year of 2023, cleaned the outliers (excluding records with unit prices exceeding three times the standard deviation of the mean or with missing key fields), and calculated the average house price, price gradient, and the proportion of the distribution of the property types in each neighborhood, in order to quantify the mechanism by which the built-up environment affects the spatial differentiation of house prices.

### (3) POI data

POI data, as a kind of point data representing geographic entities, include attribute information, such as name, type, location, etc., which is mainly used for density calculation and functional area diversity calculation. According to the Urban Land Use Classification and Planning and Construction Land Use Standards [49] and the actual situation of Nanjing, POIs are classified into seven major categories, which are residential land, commercial land, industrial land, public service land, science, education and culture land, green space and square land, and transportation facility land. In this paper, the POI data points of MSAR in 2024 obtained from Gaode Map totaled 145,370.

### (4) Weibo check-in data

The Weibo check-in function allows users to record their whereabouts in real time and show their life status and activity experience. Users can express their favorites, and support or comment on a certain place or activity, by checking in, and at the same time, they can also browse the check-in information of others to understand their activity trajectories and locations. By analyzing Weibo check-in data, we could understand the distribution of people flow, heat index, and activity density of different locations in the city, revealing the population flow patterns and social activity hotspots in the city. In this paper, a total of 145,370 Weibo check-ins of Macau SAR in 2020 were obtained from Sina Weibo.

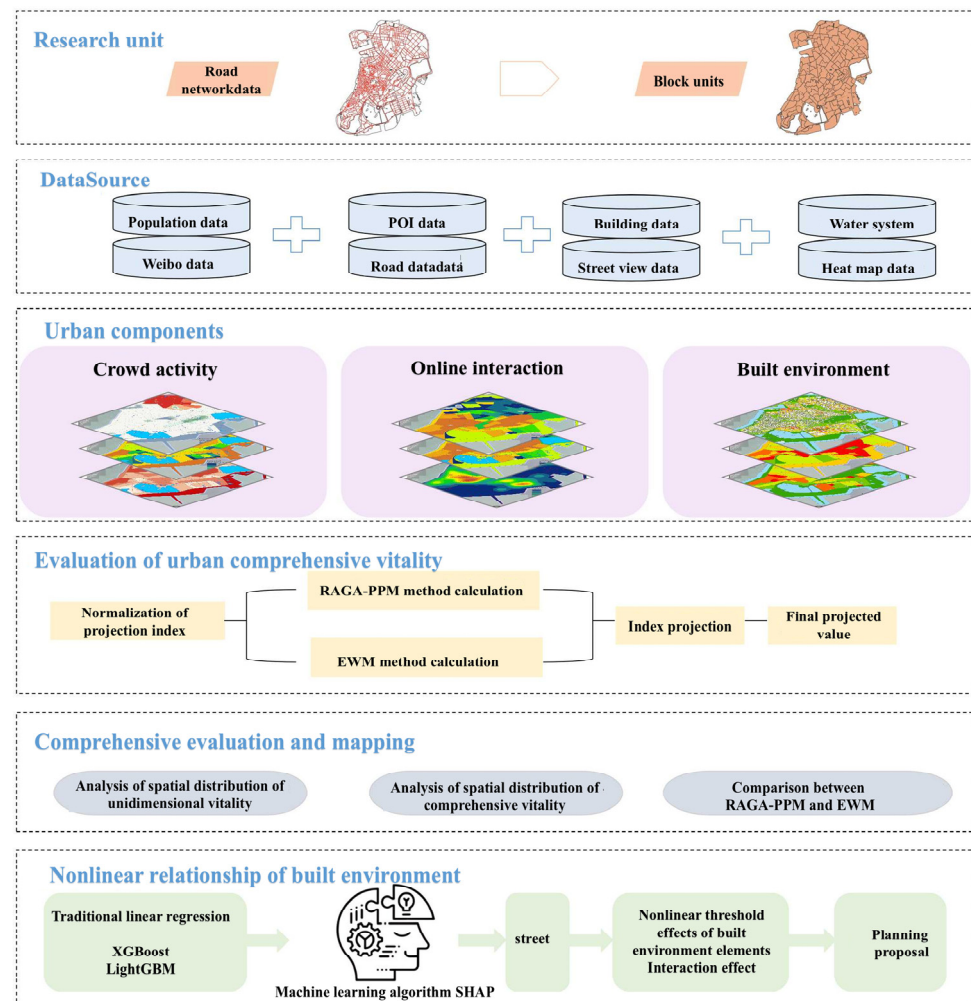
### (5) Street View Image

Street View image data were acquired based on the Baidu Street View map open platform, which covers urban street space through high-resolution panoramic photography technology and can objectively present the material form and spatial enclosure characteristics of street façades. Compared with traditional field research, this dataset has the advantages of standardized collection process and simultaneous acquisition from multiple viewpoints, which can effectively avoid the subjective bias of human observation [50]. The platform provides API interface services to support researchers in batch acquiring street view images of specified locations based on geographic coordinates or road network nodes, and it allows customization of parameters such as shooting angle, elevation angle, and image resolution. It has been shown that streetscape images have significant application value in the fields of street greening-rate measurement [51], building interface continuity assessment [52], and walking environment quality quantification [46]. In this study, based on the road network data of Macau Peninsula, sampling points were set up at 100 m intervals, and street view images were obtained from four horizontal viewpoints, ( $0^\circ$  pitch angle  $0^\circ$ ) at  $0^\circ$  (due north),  $90^\circ$  (due east),  $180^\circ$  (due south), and  $270^\circ$  (due west), of each point location through the Baidu Street View API, and 10,442 effective street view images were finally collected for the year 2023.

## 4. Research Methodology

This study employed the Macau Peninsula as a case study, dividing the research area into 159 urban blocks based on road network data. A comprehensive urban vi-

tality assessment framework was constructed, encompassing three dimensions: human activity, network interaction, and built environment. The projection pursuit model optimized by the accelerated genetic algorithm (RAGA-PPM) was utilized to reduce the dimensionality of block indicators, extract optimal projection directions, and calculate both single-dimensional and comprehensive vitality values across various aspects. Subsequently, vitality centers were identified and compared with results from the entropy weight method (EWM) to validate the effectiveness of RAGA-PPM. Furthermore, the study investigated the nonlinear impact mechanisms of the built environment on urban vitality by integrating machine learning with Shapley additive explanations (SHAP), providing scientific support for urban planning. The technical framework is illustrated in Figure 2.



**Figure 2.** Technical roadmap.

#### 4.1. Calculation of Projection Pursuit Model Based on Real-Coded Genetic Algorithm

##### 4.1.1. Projection Pursuit Model (PPM)

In urban vitality assessment, traditional methods typically rely on expert scoring and multi-indicator comprehensive evaluation, such as factor analysis, the entropy [54] weight method, and TOPSIS method. Although these methods have achieved certain results, they are limited by strong subjectivity, fewer dimensions, and difficulty in handling high-dimensional nonlinear data. Given that urban vitality involves multiple high-dimensional and nonlinear dimensions, 15 indicators were selected for calculation. The Projection Pursuit Model (PPM) [55] can effectively identify the optimal projection direction in high-

dimensional data. This method reveals the structural characteristics of data by mapping high-dimensional data into a low-dimensional space [26]. The specific steps are as follows:

(1) Since the data sources of projection indicators have different numerical types and dimensions, in order to eliminate the dimensions and unify the range of indicator changes, the following formula is used for standardization; for defined positive indicators, as shown in Formula (1), and for negative indicators, as shown in Formula (2):

$$X(i, j) = \frac{x^*(i, j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad (1)$$

$$X(i, j) = \frac{x_{\min}(j) - x^*(i, j)}{x_{\max}(j) - x_{\min}(j)} \quad (2)$$

In the given formula,  $x^*(i, j)$  represents the value of the  $j$  indicator for the  $i$  sample. Additionally,  $x_{\max}(j)$  and  $x_{\min}(j)$  correspond to the minimum and maximum values, respectively, of the  $j$  evaluation indicator within the  $i$  sample.

(2) Construct the projection index function  $z(i)$ , project  $x(i, j)$  onto the projection direction  $a$ , let  $a = (a(1), a(2), \dots, a(j))$ , where  $a$  is the unit length vector, and the projection value  $z(i)$  is

$$z(i) = \sum_{j=1}^p (a_j \times x_{ij}) \quad (3)$$

where  $z(i)$  is the projection vector of the  $i$ th sample,  $x_{ij}$  is the  $i$ th value in the  $j$ th index, and  $a_j$  is the projection direction of the projection vector  $z(i)$ .

(3) Construct the projection objective function  $\rho(a)$  to identify the structural combination features of the data within the multidimensional indicators. During the comprehensive projection process, the projection value  $Z(i)$  should maximize the extraction of variation information from  $x(i, j)$ , ensuring that the standard deviation  $S_z$  of  $Z(i)$  is as large as possible. Simultaneously, the local density  $D_z$  of the projection values  $Z(i)$  should be maximized. Based on these principles, the projection objective function can be formulated as  $\rho$ .

$$\rho(a) = S_z D_z \quad (4)$$

$$S_z = \sqrt{\frac{\sum_{i=1}^p (Z(i) - Z^*)^2}{n - 1}} \quad (5)$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^n (R - r_{ij}) u(R - r_{ij}) \quad (6)$$

where  $S_z$  is the standard deviation of  $Z$ ;  $D_z$  is the local density of  $Z$ ;  $Z^*$  is the mean of  $Z(i)$  ( $i = 1 \sim n$ );  $r_{ij}$  is the window radius of the local density, which is related to the search radius of the projection point;  $r_{ij}$  refers to the difference between  $Z(i)$  and  $Z(j)$ ;  $u(R - r_{ij})$  is the unit step function. Its function value is 1 when  $R - r_{ij} \geq 0$ , and its function value is 0 when  $R - r_{ij} < 0$ .

(4) Optimizing the projection objective function to determine the optimal projection direction, different projection directions reflect various structural characteristics of the data. The optimal projection direction is the one that maximizes the likelihood of revealing specific structural features of high-dimensional data. To estimate the optimal projection direction, the projection objective function is maximized, ensuring that the most informative structural properties are exposed.

$$\text{Max} = (\rho(a)) = S_z D_z \quad (7)$$

$$s.t. \sum_{j=1}^p a^2(j) = 1, -1 \leq a(j) \leq 1 \quad (8)$$

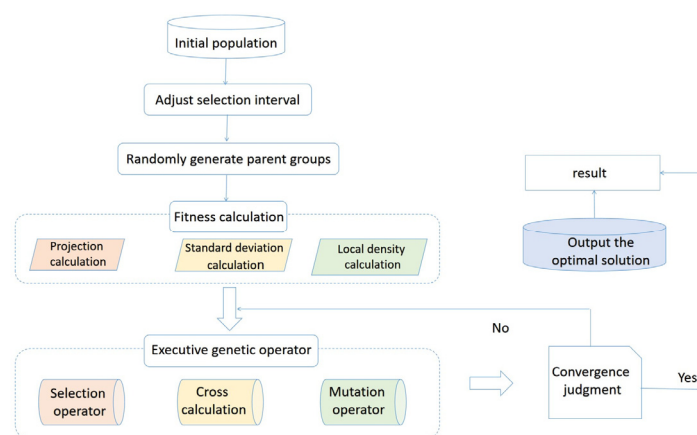
(5) Substituting the optimal projection direction into Equation (3), the projection values for each sample can be obtained, enabling a unified evaluation of the sample set. Given that Equations (7) and (8) represent a complex nonlinear optimization problem and involve parameter settings, conventional optimization methods may struggle to provide an effective solution. Therefore, a genetic algorithm (GA), a general global optimization approach that incorporates the principles of biological selection and intra-population chromosome information exchange, is introduced to address the high-dimensional global optimization problem.

#### 4.1.2. Accelerated Genetic Algorithm Based on Real-Coded Encoding (RAGA)

In this study, an improved genetic algorithm—the Accelerated Genetic Algorithm based on Real-Coded Encoding (RAGA)—was employed to optimize the Projection Pursuit Model (PPM) and address challenges in the solution process. RAGA integrates the global search capability of genetic algorithms with the continuous optimization advantages of real-coded encoding, significantly enhancing optimization performance [41–44]. By simulating the biological evolution process, this algorithm executes selection, crossover, and mutation operations in parallel, expanding the search space and increasing the likelihood of identifying the global optimal solution. The adaptive mechanism of RAGA progressively narrows the search range of optimization variables, thereby improving the solution accuracy. As the number of iterations increases, the precision of the solution continues to improve [42].

#### 4.1.3. Projection Pursuit Model Optimized by Accelerated Genetic Algorithm with Real-Coded Encoding (RAGA-PPM)

The Projection Pursuit Model (PPM) effectively handles complex nonlinear relationships, while the Real-coded Adaptive Genetic Algorithm (RAGA) enhances PPM's capability in processing high-dimensional nonlinear data through global search and adaptive adjustment. Compared to traditional methods, RAGA autonomously optimizes parameters, minimizing human intervention, improving the efficiency and accuracy of the evaluation process, and mitigating the risk of local optima. Notably, RAGA has demonstrated outstanding performance in multimodal function optimization. However, the model has certain limitations, such as a high computational complexity and sensitivity to parameter settings. By optimizing the calculation of 17 indicator factors, this study provides a more accurate representation of urban vitality, as illustrated in Figure 3.



**Figure 3.** Solution Process of the RAGA-PPM Model.

#### 4.2. Machine Learning Models and SHAP Interpretability Method

Both XGBoost [34] and LightGBM [35] are based on the Gradient Boosting Decision Tree (GBDT) algorithm, which iteratively adds new trees to optimize the model. Although the two algorithms differ in specific details and optimization strategies, they share a similar fundamental formulation. Their objective functions consist of two main components: a loss function and a regularization term.

$$\text{Objective} = \sum_{i=1}^n L(y_i, \check{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (9)$$

The loss function is defined as the squared error loss, which measures the squared error between the model predicted value  $\check{y}_i$  and the true value  $y_i$ .

$$L(y_i, \check{y}_i) = (y_i - \check{y}_i)^2 \quad (10)$$

The regularization term controls the complexity of the model and prevents overfitting.

$$\Omega(f_k) = \gamma M + \frac{1}{2} \lambda \sum_{j=1}^M \omega_j^2 \quad (11)$$

where  $\Omega(f_k)$  is the regularization term of the  $k$  th leaf in the tree;  $\gamma$  is the penalty coefficient of the number of leaf nodes, used to limit the complexity of the model and reduce overfitting;  $M$  is the number of leaf nodes;  $\lambda$  is the penalty coefficient of the leaf node fraction, used to control the L2 regularization strength of the leaf node fraction in the model; and  $\omega_j$  is the fraction of the  $j$  th leaf node.

For the  $t$  th tree, the objective function is

$$\text{Objective} = \sum_{i=1}^n \left\{ y_i - [\check{y}_i^{(t-1)} + f_t(x_i)] \right\}^2 + \gamma M + \frac{1}{2} \lambda \sum_{j=1}^M \omega_j^2 \quad (12)$$

where  $\check{y}_i^{(t-1)}$  is the predicted value of the  $t - 1$  tree for the  $i$  th sample,  $n$  is the number of training samples, and  $f_t(x_i)$  is the predicted value of the  $t$  th tree for the sample  $x_i$ .

The final prediction model is obtained by summing up of the iteratively generated trees:

$$\check{y}_i = \sum_{t=1}^T f_t(x_i) \quad (13)$$

where  $\check{y}_i$  is the predicted value of the  $i$  th sample, which is the predicted vitality of the block sample;  $T$  is the total number of trees; and  $f_t(x_i)$  is the predicted value of the  $t$  th tree for  $x_i$ .

The SHAP interpretability method [36] aims to quantify the contribution of each feature to the model's prediction, expressed as the SHAP value. This method considers all possible feature combinations and calculates the average contribution of each feature across all possible combinations. The calculation formula is as follows:

$$\text{SHAP}_{ij} = \sum_{S \subseteq \{x_1, x_2, \dots, x_n\} \setminus \{x_j\}} \frac{|S|!(n-|S|-1)!}{n} \times [f(S \cup \{x_j\}) - f(S)] \quad (14)$$

## 5. Results and Analysis

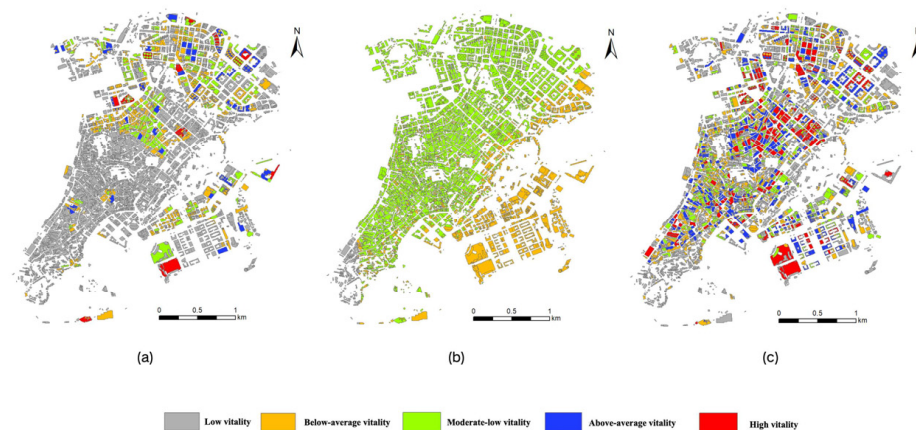
The projection pursuit model optimized by the accelerated genetic algorithm based on real-number encoding (RAGA-PPM) was applied to calculate the comprehensive vitality of the streets in the Macau Peninsula. A database consisting of 15 indicators was used as the foundation, and multiple experiments were conducted to determine the optimal projection direction when  $\alpha = 0.05$ . The resulting best projection direction was found to be  $a = (-0.038, 0.002, 0.003, -0.056, -0.056, -0.097)$ . The values of each component represent the contribution of the corresponding indicators to the overall evaluation goal. The order



of the indicators represented by each component was as follows: population density, Weibo check-in density, spatial compactness, functional density, and functional mix. It is evident that the largest coefficients in the projection direction correspond to population density and Weibo check-in density. These two factors contribute the most to the comprehensive vitality value, indicating that they have the greatest impact on the overall vitality of urban neighborhoods. For visualization, the vitality was classified into five levels using the natural breaks classification method, arranged from high to low vitality: high-vitality areas, relatively high-vitality areas, moderate-vitality areas, relatively low-vitality areas, and low-vitality areas.

### 5.1. Spatial Distribution Analysis of Unidimensional Activity

The distribution of the vitality of people's activities in the Macau Peninsula presents a significant spatial characteristic of “high value at the core and low value at the periphery”, which reflects a high degree of agglomeration in the functional layout and spatial utilization of the city, as shown in Figure 4a. High-vitality areas are mainly concentrated in the integrated commercial and entertainment functional areas centered on the Grand Lisboa Hotel, Wynn Macau, and MGM MACAU, and at the same time cover important historical and cultural landmarks such as the Senado Square, the Ruins of St. Paul's, and Our Lady of the Rosary Church. The Central Vitality Zone is distributed in Hac Sa Wan and New Boundary Control Point, showing the transitional characteristics of function and vitality. Hac Sa Wan is predominantly residential in function but has the advantage of strong spatial connectivity due to its proximity to the core business district. Low-vitality areas are mainly located in traditional residential districts such as Chopstick Key and Taishan. These areas are single-function, mainly residential, and lack large public spaces and highly attractive commercial facilities, resulting in a low frequency of crowd activity.



**Figure 4.** Activity diagram of one-dimensional aspect. (a) Vitality of crowd activities; (b) network interaction vitality; (c) built environment vitality.

The spatial pattern of network interaction vitality in the Macau Peninsula generally showed “predominantly medium-vitality and scattered low-vitality”, as shown in Figure 4b. Medium-vitality areas are mainly concentrated in functionally diversified areas such as the New Boundary Crossing Area and Hac Sa Wan. As a combination of business and life service functions, the New Port area has created active online participation through digital activities such as check-ins and reviews, while the Hac-Sa-Huan area, although mainly residential, shows a certain degree of online interaction activity due to its proximity to the core business district and good transportation conditions. Areas with low online interaction vitality are mainly located in traditional residential areas such as Chopsticks

Key, Tsing Chau, and Ha Wan. These districts are significantly less active, due to their single function and lack of large public spaces and commercial facilities.

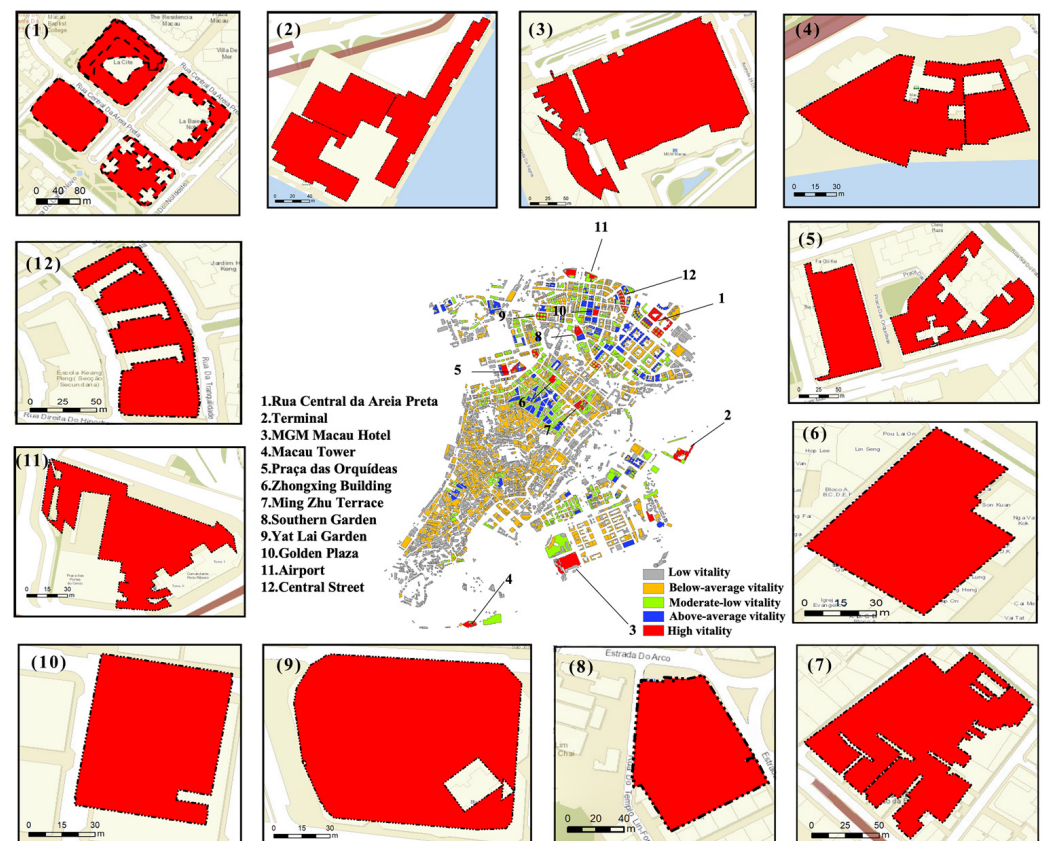
The vibrancy of the built environment of the Macau Peninsula shows a clear hierarchical distribution, closely related to the functional layout of the area, the land use pattern, and the transportation conditions, as shown in Figure 4c. Highly vibrant areas are mainly concentrated around the Outer Harbour, Nam Van, Senado Square, and Ruins of St. Paul's. These areas are characterized by commercial, tourism, and cultural functions. These areas attract many people due to the high integration of commercial, tourism, and cultural functions. With Hotel Lisboa and Wynn Macau as the core, these landmark districts have become the main driving centers of the vibrancy of Macau's built environment by virtue of their high-density building layouts, diversified forms of land use, and well-developed transportation networks. The central vitality areas are distributed at the periphery of the core area, mainly covering the New Border Crossing and Hac Sa Wan areas. These areas are mainly residential, with certain commercial and living service functions, moderate building density, and convenient transportation conditions. Low-vitality areas are concentrated in the urban fringe areas such as Chopstick Key, Taishan, and Xia Huan.

In terms of unidimensional vitality, high-vitality areas in the Macau Peninsula are primarily concentrated around locations such as the New Lisboa, Ruins of St. Paul's, and A-Ma Temple. These areas exhibit significant advantages in terms of human activity, network interaction, and built environment. In contrast, low-vitality areas are primarily found in the northern region, mainly due to the lack of diverse functions and insufficient infrastructure. This distribution of vitality provides important data support for optimizing urban functions and adjusting the spatial structure of Macau.

## 5.2. Comprehensive Vitality Space Analysis

The comprehensive vitality evaluation results based on the RAGA-PPM algorithm indicate that the urban spatial structure of Macau exhibits significant hierarchical and central aggregation characteristics in Figure 5. The values for the spatial distribution pattern of vitality show a gradual transition from high-vitality areas in the core region to low-vitality areas at the periphery, forming a distinct central-peripheral hierarchical structure. The core area, characterized by its highly concentrated commercial activities, convenient transportation network, and high-quality spatial environment, emerged as the most vibrant zone. High- and relatively high-vitality areas on the Macau Peninsula are primarily concentrated in the urban center, which includes landmarks such as the Ruins of St. Paul's, Senado Square, and the areas surrounding the A-Ma Temple and Nam Van. These regions, where the city's historical, cultural, and commercial resources converge, exhibit a prosperous commercial atmosphere and dense pedestrian traffic, thus serving as the main sources of urban vitality. Additionally, areas such as Praça de Amizade, Wynn Macau, Fisherman's Wharf, Casino Lisboa, and the New Porto Area, owing to their rich entertainment, accommodation, and dining functions, also constitute relatively high-vitality zones. In contrast, low-vitality areas are mainly found in peripheral regions such as the Lower Ring, Ka I, and Taipa, where high building density, complex road networks, and a lack of diversified functions result in relatively lower vitality levels. Compared with previous studies, this research identified more potential vitality centers. Key vitality regions, such as Hác Sá, Ka Ho, Senado Square, New Porto Area, and Nam Van, were recognized as the main vitality hubs, forming the core vitality zone of the Macau Peninsula. These areas not only concentrate rich historical and cultural resources, but also feature bustling commercial districts and diverse transportation networks, making them the largest gathering points for crowds and economic activities. Specifically, notable vitality centers in various districts include the Vista Bay (1) and Hai Ming Ju (2) in the

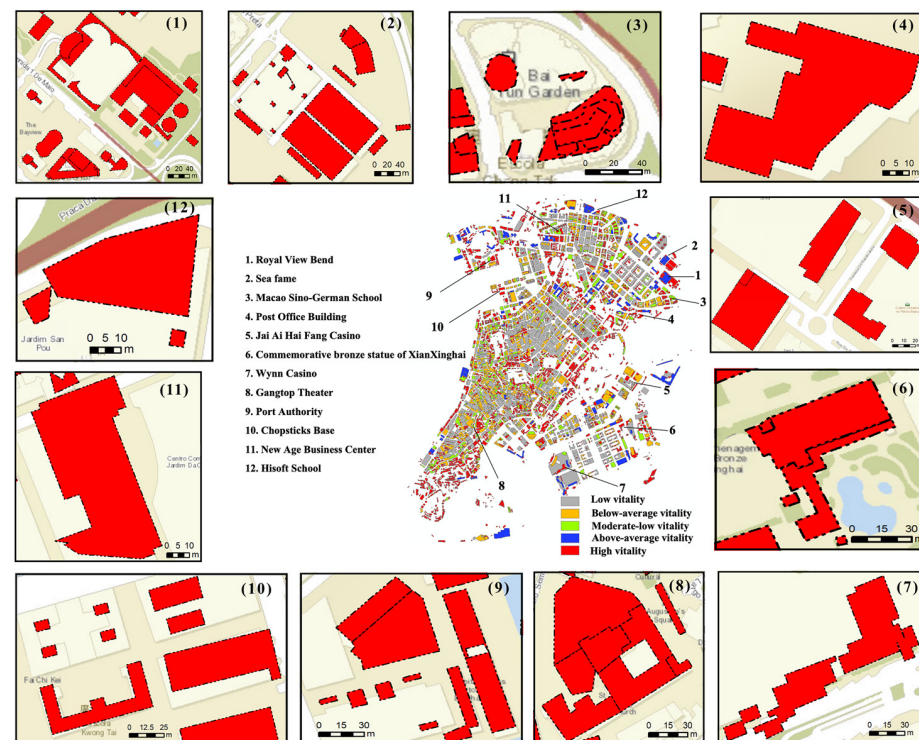
Our Lady of Fátima Parish, the Macau German School (3), the Post Office Building (4) in the Senado Square area, Casino Lisboa (5) and the Xian Xinghai Memorial (6) in the New Porto Area, Wynn Macau (7) and the Gung Ding Theatre (8) in Praça de Amizade, the Port Authority (9) in the Lower Ring, Ka I Square (10), the New Era Commercial Center (11) in Nam Van, and the Haihui School (12) in the Border Gate area. These vitality centers reflect the close relationship between the spatial distribution and functional layout of Macau's urban vitality, highlighting the intensity of regional commerce and population flows.



**Figure 5.** Comprehensive activity and activity center map based on RAGA-PPM.

The comprehensive vitality evaluation results based on the EWM entropy weight method indicate that the spatial distribution of urban vitality in the Macau Peninsula presents a hierarchical pattern of “core-functional nodes-periphery” in Figure 6. The core area of the Gongbei Port exhibits the highest vitality level, due to its high commercial density, convenient transportation, and strong population mobility. The functional node areas, such as the MGM Macau Hotel District and the harbor area, show concentrated vitality, due to their unique service functions and high foot traffic, though their radiation effect is limited. The peripheral areas, such as Heshawan and Haoshi, primarily serve ecological and residential functions, displaying relatively weaker vitality but with significant potential for development. Overall, the urban vitality distribution shows a trend of multi-core development, with the core areas maintaining a dominant advantage.

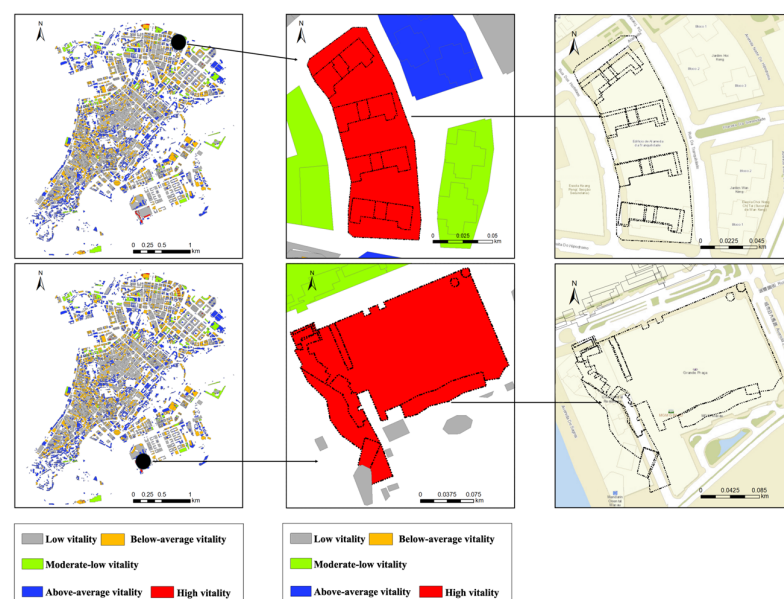




**Figure 6.** Comprehensive vitality and vitality center map based on EWM.

### 5.3. Analysis of Spatial Differences in Vitality Using RAGA-PPM and EWM

A comparative analysis of the spatial distribution of urban vitality in Macau using the RAGA-PPM and EWM revealed significant differences in vitality assessment between the two methods. The RAGA-PPM approach primarily relies on a genetic algorithm to optimize the weights of multivariable data. Its results predominantly reflect the vitality distribution patterns of urban functional zones, emphasizing the vitality contributions from functional nodes and ecological areas. In contrast, the EWM method uses objective weighting based on information entropy, focusing on the spatial distribution characteristics of regional vitality under a comprehensive set of indicators. It placed greater emphasis on the performance of core areas and functional complexity, as illustrated in Figure 7.



**Figure 7.** Differences in activity levels between RAGA-PPM and EWM.

The differences between the two methods in vitality evaluation are also reflected in the level of spatial scale detail. RAGA-PPM provides a more refined description of vitality distribution in small-scale functional areas, effectively identifying small commercial and public spaces around hotel districts. In contrast, EWM focuses more on the urban vitality patterns at a macro scale, making it suitable for holistic planning analysis.

#### 5.4. The Impact of the Built Environment on Urban Vitality

In this study, an urban vitality prediction model was constructed based on a Bayesian optimization framework and the LightGBM algorithm, with 14 built environment elements of 159 neighborhoods as feature variables, systematically balancing the model accuracy and generalization ability in Table 3. A hierarchical random partitioning strategy was used to construct a training set (80%) and independent test set (10%), and the optimal hyperparameter combinations were determined by Bayesian optimization (5-fold cross-validation, 25 iterations): learning rate (0.262), number of trees (364), maximal depth (9), and feature/test sub-sampling rate (0.860/0.823). The early stopping method (terminate without improvement in 10 consecutive rounds of validation set) was enabled in training, the best model converged in the 30th round and the independent test set showed significant performance ( $RMSE = 0.082$ ,  $MAE = 0.065$ ,  $R^2 = 0.555$ ), which confirmed its effective capture of nonlinear relationships for urban vitality, without overfitting. The method was further combined with SHAP interpretability analysis to reveal the threshold effect and nonlinear influence mechanism of key elements such as BC, BFD, FD, etc., which provided a quantitative decision-making basis for the optimization of the built environment. The method achieves a synergistic breakthrough in accuracy, efficiency, and mechanism analysis for multi-source urban data modeling through an efficient search of hyperparameters, an anti-overfitting strategy, and interpretability enhancement.

**Table 3.** The fitting accuracy of the indicators.

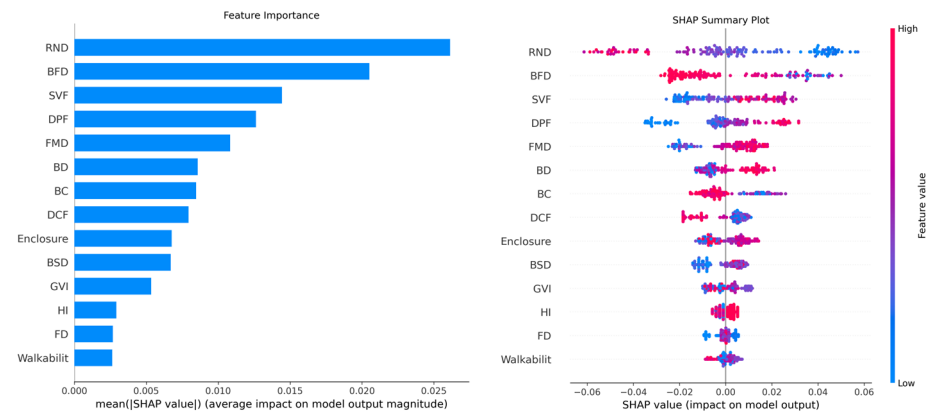
Dataset	MAE	RMSE	$R^2$
Training Set Results	0.0256	0.0375	0.8170
Test Set Results	0.0650	0.0817	0.5548

##### 5.4.1. Relative Importance of Built Environment Elements

This study revealed the hierarchical characteristics of the impact of built environment elements on vitality in Macau's districts through SHAP value analysis (Figure 8). Panel (a) of Figure 8 shows the contribution and importance ranking of the 14 built environment elements, while panel (b) presents the specific influence patterns of each element on vitality prediction through SHAP value distribution. In the feature importance ranking, the top six built environment elements by relative importance were road network density (0.025), spatial fractal dimension (0.020), sky openness (0.015), public facility density (0.012), functional mix (0.010), and building density (0.008). Road network density, spatial fractal dimension, and sky openness were identified as key factors influencing district vitality. Road network density enhances traffic connectivity and accessibility, facilitating economic activities and the walking experience, thereby attracting more pedestrian flow. The spatial fractal dimension strengthens spatial diversity and social interaction through complex spatial structures, improving environmental attractiveness. Sky openness, by offering a sense of open space and environmental comfort, promotes outdoor activities and the gathering of crowds. These three factors work synergistically to improve traffic convenience, spatial diversity, and environmental quality, significantly enhancing the vitality and attractiveness of a district. Public facility density, functional mix, and building density are secondary key factors influencing district vitality. The high rankings of sky



openness and green visibility indicate that the micro-scale street design plays a crucial role in the overall vitality of the district. In comparison, factors such as enclosure and green visibility have a relatively smaller impact, yet still influence district vitality to some extent. A greater sense of enclosure may enhance walking comfort, while districts with higher green visibility are likely to attract more outdoor activities. Walkability, waterfront index, and functional density have a relatively minor impact on vitality, possibly due to the specific environmental characteristics of the study area.



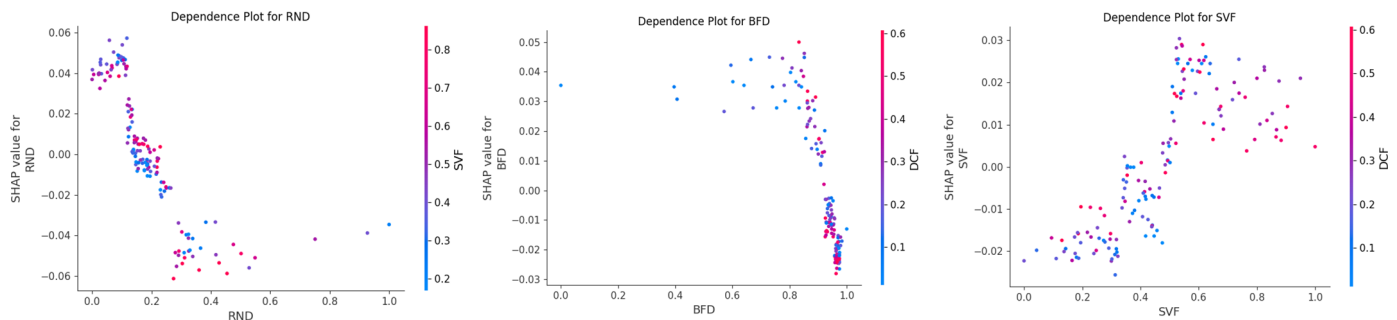
**Figure 8.** The relative importance of built environment elements on vitality in the Macau Peninsula districts.

#### 5.4.2. Nonlinear Relationship Between Built Environment Elements and Urban Vitality

To further investigate the localized nonlinear relationship between urban vitality and the built environment, we focused on the top three most globally significant built environment elements. Figure 9 illustrates the nonlinear relationship between these top three elements at the block level and their contributions to predicting vitality, represented by SHAP values. The color legend indicates the value of the interacting element that had the strongest influence on each sample. The impact of road network density on urban vitality exhibited a complex nonlinear relationship. When the road network density was less than 0.2, the SHAP values were relatively high. As the density increased, the SHAP values decreased rapidly and reached negative values within the range of road network density from 0.2 to 0.4. This suggests that overly dense road networks may reduce block vitality, possibly due to street fragmentation or excessive road occupation of public spaces. However, when the road network density exceeded 0.6, the SHAP values stabilized, indicating that the marginal effect of extremely high road network density on vitality had diminished. The SHAP values for spatial fractal dimension generally showed a downward trend, particularly when the spatial fractal dimension exceeded 0.8, where the SHAP values decreased sharply. This suggests that excessively high spatial fractal dimensions may negatively impact urban vitality. Such an effect could be attributed to overly complex spatial forms, which may hinder block vitality due to factors like visual obstruction or decreased accessibility. Sky openness demonstrated a positive relationship with urban vitality, especially when the sky openness increased from 0.2 to 0.8, during which the SHAP values steadily rose. This suggests that appropriately increasing sky openness could help enhance block vitality. However, when sky openness exceeded 0.8, a decrease in SHAP values for some points was observed, possibly reflecting the negative impact of excessively open environments on block vitality.

In summary, urban vitality is influenced by the complex and nonlinear effects of built environment factors. A moderate spatial fractal dimension, an appropriate road network density, and a balanced sky openness may be more conducive to enhancing vitality. Future

planning and design should consider the interactions between different elements, in order to optimize urban spatial forms.



**Figure 9.** Nonlinear relationship between built environment elements and vitality in the Macau Peninsula district.

## 6. Conclusions and Discussion

### 6.1. Conclusions

This paper proposes a new framework for evaluating urban vitality that integrates crowd activities, network interactions, and the built environment. By introducing the RAGA-PPM method, it provides a quantitative assessment of the overall vitality of the Macau Peninsula. Unlike traditional methods that introduce biases due to subjective weighting, this approach identifies the contribution of each indicator to overall vitality based on the inherent structural characteristics of the data, resulting in a more objective and accurate identification of vitality outcomes. Using machine learning algorithms combined with the SHAP interpretability method, this study explored the nonlinear impacts and interactive effects of the built environment on urban vitality in the urban areas of the Macau Peninsula. The findings were as follows:

- (1) There are certain spatial distribution differences in the urban vitality across the three dimensions—crowd activity, network interaction, and the built environment—on the Macau Peninsula. However, a common feature across all dimensions is that they exhibit high values at the center, with urban vitality values gradually decreasing toward the periphery. The vitality of crowd activities is concentrated around tourism, commerce, entertainment, and cultural functions, with a high value at the center and secondary vitality clusters in each district. The vitality of network interactions is centered around the New Port and Hac-Sa Bay, presenting a spatial pattern of “moderate vitality as the dominant feature, with scattered areas of low vitality”. The spatial distribution of built environment vitality shows clear hierarchical patterns, closely related to the regional functional layout, land use patterns, and transportation conditions, with high-vitality areas primarily concentrated around the Outer Harbor, Nanhai, Senado Square, and the area surrounding the Ruins of St. Paul’s.
- (2) The spatial distribution of overall vitality on the Macau Peninsula is like that of the individual dimensions. Specifically, the largest vitality center is located around the historic district, with vitality decreasing as it moves outward, and the lower vitality areas are distributed at the urban periphery. The RAGA-PPM method used in this study provided a more accurate identification of vitality areas, revealing more potential vitality centers. Each district contains a vitality center, with the historic district—comprising the Ruins of St. Paul’s, New Road, A-Ma Temple, and Senado Square—serving as the main vitality hub for the entire Macau Peninsula. The areas around New Road and Nanhai exhibited the highest vitality, while the areas around Lower Harbor and Taishan showed the lowest vitality. The spatial distribution of

urban vitality on the Macau Peninsula is influenced by various factors, showing heterogeneity in distribution.

- (3) This study reveals that, at the neighborhood level, the six built environment factors most significantly influencing urban vitality in the Macau Peninsula are, in order of importance, street network density, spatial fractal dimension, sky openness, public facility density, functional mix, and building density. These factors were ranked based on their relative significance to urban vitality and form a key set of built environment indicators that impact the urban vitality of the Macau Peninsula.
- (4) The built environment on the Macau Peninsula has a nonlinear effect on urban vitality. Regarding the key factors that influence the vitality of streets and blocks, road network density, spatial fractal dimension, and sky openness had similar threshold values, with all demonstrating a positive impact on vitality when the value was below 0.2.

## 6.2. Discussion

Exploring the similarities and differences in the spatial distribution of vitality across different dimensions and the overall vitality of the central urban area of the Macau Peninsula can contribute to refining the urban comprehensive vitality assessment system and offer targeted strategies for urban governance. The urban vitality of the Macau Peninsula is predominantly concentrated in the historic districts and commercial centers, while areas with lower vitality are mainly located at the outskirts of the city. Therefore, greater emphasis should be placed on fostering and developing a polycentric structure, to prevent excessive resource concentration in a single center. By clearly defining the developmental direction of each area, identifying distinct regional growth paths, and optimizing the layout of public service facilities, commercial outlets, and cultural and entertainment amenities, the vitality of each district can be enhanced, ultimately achieving a more balanced distribution of urban vitality. Additionally, lower vitality areas are primarily situated at the urban periphery. It is necessary to further strengthen the radiating role of transportation hubs and implement specific development policies for low-vitality regions. Improving the quality and coverage of public infrastructure and transportation services will facilitate better connections between peripheral and central areas. By leveraging the flow of resources and population, both within and outside the city, industries can be guided toward development, employment opportunities increased, and living conditions improved, thus gradually elevating the vitality levels of these areas. Furthermore, the historic districts play a central role in the vitality of the Macau Peninsula. In the planning and construction process, it is essential to protect and utilize the rich historical and cultural resources, while preserving the historical features and cultural characteristics of these areas. Through appropriate updates and renovations, the livability and attractiveness of the old districts can be enhanced. Lastly, the distribution of urban spatial vitality in the Macau Peninsula is influenced by multiple factors and exhibits significant heterogeneity. Smart city development should be promoted, utilizing big data, the Internet of Things, and other technologies for refined management and the establishment of innovative platforms. Real-time monitoring and analysis of dynamic changes in urban vitality can provide scientific evidence for urban planning and decision-making. Based on the development status and potential of different areas, targeted policies and measures should be formulated to improve urban governance levels and efficiency.

Based on the traditional Geographic Information System (GIS) and general map tools (Google Maps), this study overcomes the limitations of existing tools in the analysis of complex urban systems through multidimensional data fusion, nonlinear modeling approaches, and interpretability enhancement techniques in the following three respects. First, traditional GIS and map tools mainly rely on static spatial data (road networks,

building contours) and basic statistical functions, while this study constructed a more comprehensive framework for urban vitality assessment by fusing heterogeneous big data from multiple sources. Secondly, compared with the linear regression or spatial autocorrelation models relied on by traditional GIS, this study combined machine learning algorithms and optimization methods to achieve an accurate resolution of complex nonlinear relationships. This paper constructed a systematic and refined set of independent variables using multi-source data science, analyzing the distribution of urban vitality in the Macau Peninsula at the micro-scale of street block units. By integrating machine learning algorithms and the SHAP interpretability method, it explored the nonlinear effects of built environment factors on urban vitality. This approach provided new insights into the mechanisms of how the built environment influences urban vitality, offering practical and actionable guidance for more detailed urban planning. It can contribute to the coordinated design of street and block planning, strengthening the connection and interaction between the urban skeleton and its “muscles”. Although this study has made progress in methodological innovation and data analysis, it still has some limitations. The study mainly focused on the measurement of dynamic population density, and future research could incorporate Jiang Difei’s theory to comprehensively assess vitality from economic, social, and cultural perspectives. Furthermore, this study investigated the spatial heterogeneity of the built environment’s impact on urban vitality at the block level; however, future studies could explore temporal heterogeneity to deepen the understanding of the dynamic mechanisms of urban vitality. Lastly, the conclusions drawn from the Macau Peninsula may need further validation in other regions. Future research could expand data sources and incorporate more dimensions, such as economic, social, and cultural data, to build a more comprehensive urban vitality evaluation system. Additionally, time-series data could be integrated to investigate the dynamic changes in urban vitality, providing more accurate decision-making support for urban planning.

**Author Contributions:** C.P., Conceptualization; Supervision; Validation; Writing—original draft; Writing—review and editing. J.G., Writing—original draft; Validation; Investigation. H.L., Writing—original draft; Data curation; Resources; Validation. J.W., Writing—original draft; Investigation; Software. N.Q., Writing—review and editing; Visualization. S.W., Writing—review and editing; Software. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by (1) Guangdong Philosophy and Social Science Planning Project (GD24XSH06), (2) Guangdong Philosophy and Social Science Planning Project (GD24XYS018), (3) 2023 Maoming Philosophy and Social Science Planning Project (2023YB18), (4) Projects of Talents Recruitment of GDUPT (2023rcyj2015), (5) Projects of PhDs’ Start-up Research of GDUPT (2023bsqd1008) (2022bsqd2004), and (6) Science and Technology Programme of Maoming of Guangdong Province of China (2023398), (2023411), (2024055), (2024061).

**Data Availability Statement:** The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Conflicts of Interest:** The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

1. Ye, Y.; Zhuang, Y.; Zhang, L.Z. Morphological Study of Vitality Creation in Urban Design: Based on Quantitative Analysis of Urban Spatial Morphology and Resident Activity Verification. *Int. J. Urban Plan.* **2016**, *31*, 26–33.
2. Ewing, R.; Cervero, R. Travel and the Built Environment: A Synthesis. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1780*, 87–114. [[CrossRef](#)]
3. Montgomery, J. Making a City: Urbanity, Vitality, and Urban Design. *J. Urban Des.* **1998**, *3*, 93–116. [[CrossRef](#)]
4. Wang, Z.M.; Liu, Y.F.; Luo, X. Nonlinear Relationship Between Urban Vitality and Built Environment Based on Multi-source Data: A Case Study of the Main Urban Area of Wuhan on Weekends. *Prog. Geogr.* **2023**, *42*, 716–729. [[CrossRef](#)]

5. Wu, W.; Ma, Z.; Guo, J.; Zhao, K. Nonlinear effect of built environment on street vitality: A multi-source big data analysis based on XGBoost model. *Chin. Landsc. Archit.* **2022**, *38*, 82–87.
6. Hong, Y.M.; Wang, S.Y. Big Data, Machine Learning, and Statistics: Challenges and Opportunities. *J. Econom.* **2021**, *1*, 17–35.
7. Feng, Z.M.; Zheng, H.X.; Liu, B.Q. Comprehensive Evaluation of Agricultural Water Resource Utilization Efficiency Based on Genetic Projection Pursuit Model. *Trans. Chin. Soc. Agric. Eng.* **2005**, *21*, 66–70.
8. Jin, J.L.; Yang, X.H.; Ding, J. An Accelerated Genetic Algorithm Based on Real Coding. *J. Sichuan Univ. (Eng. Sci. Ed.)* **2000**, *32*, 20–24.
9. Long, Y.; Huang, C.C. Does Block Size Matter? The Impact of Urban Design on Economic Vitality in Chinese Cities. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 406–422. [[CrossRef](#)]
10. Yang, D.F.; Wang, X.M.; Han, R.N. Nonlinear Effects and Interaction of Built Environment on Street Vitality: A Case Study of Shenyang. *Urban Plan. J.* **2023**, *05*, 93–102.
11. Sung, H.G.; Go, D.H.; Choi, C.G. Evidence of Jacobs' Street Life in Great Seoul: Identifying the Association Between Physical Environment and Walking Activity on Streets. *Cities* **2013**, *35*, 164–173. [[CrossRef](#)]
12. Wu, J.Y.; Ta, N.; Song, Y.; Lin, J.; Chai, Y. Urban Form Breeds Neighborhood Vibrancy: A Case Study Using a GPS-Based Activity Survey in Suburban Beijing. *Cities* **2018**, *74*, 100–108. [[CrossRef](#)]
13. Delclòs-Alió, X.; Gutiérrez, A.; Miralles-Guasch, C. The Urban Vitality Conditions of Jane Jacobs in Barcelona: Residential and Smartphone-Based Tracking Measurements of the Built Environment in a Mediterranean Metropolis. *Cities* **2019**, *86*, 220–228. [[CrossRef](#)]
14. Jiang, D.F. *Urban Morphology and Vitality Theory*; Southeast University Press: Nanjing, China, 2007.
15. Tu, W.; Zhu, T.T.; Xia, J.Z.; Zhou, Y.; Lai, Y.; Jiang, J.; Li, Q. Portraying the Spatial Dynamics of Urban Vibrancy Using Multisource Urban Big Data. *Comput. Environ. Urban Syst.* **2020**, *80*, 101428. [[CrossRef](#)]
16. He, Q.S.; He, W.S.; Song, Y.; Wu, J.Y.; Yin, C.H.; Mou, Y.C. The Impact of Urban Growth Patterns on Urban Vitality in Newly Built-up Areas Based on Association Rules Analysis Using Geographical Big Data. *Land Use Policy* **2018**, *78*, 726–738. [[CrossRef](#)]
17. Zhang, A.Q.; Li, W.F.; Wu, J.Y.; Lin, J.; Chu, J.; Xia, C. How Can the Urban Landscape Affect Urban Vitality at the Street Block Level? A Case Study of 15 Metropolises in China. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 1245–1262. [[CrossRef](#)]
18. Gómez-Varo, I.; Delclòs-Alió, X.; Miralles-Guasch, C. Jane Jacobs Reloaded: A Contemporary Operationalization of Urban Vitality in a District in Barcelona. *Cities* **2022**, *123*, 103565. [[CrossRef](#)]
19. Hajrasouliha, A.; Yin, L. The Impact of Street Network Connectivity on Pedestrian Volume. *Urban Stud.* **2015**, *52*, 2483–2497. [[CrossRef](#)]
20. Lan, F.; Gong, X.Y.; Da, H.L.; Wen, H. How Do Population Inflow and Social Infrastructure Affect Urban Vitality? Evidence from 35 Large- and Medium-Sized Cities in China. *Cities* **2020**, *100*, 102454. [[CrossRef](#)]
21. Li, X.; Li, Y.; Jia, T.; Zhou, L.; Hijazi, I.H. The Six Dimensions of Built Environment on Urban Vitality: Fusion Evidence from Multisource Data. *Cities* **2022**, *121*, 103482. [[CrossRef](#)]
22. Yue, W.Z.; Chen, Y.; Thy, P.T.M.; Fan, P.; Liu, Y.; Zhang, W. Identifying Urban Vitality in Metropolitan Areas of Developing Countries from a Comparative Perspective: Ho Chi Minh City versus Shanghai. *Sustain. Cities Soc.* **2021**, *65*, 102609. [[CrossRef](#)]
23. Zhang, Z.R.; Liu, J.P.; Wang, C.Y.; Zhao, Y.; Zhao, X.; Li, P.; Sha, D. A Spatial Projection Pursuit Model for Identifying Comprehensive Urban Vitality on Blocks Using Multisource Geospatial Data. *Sustain. Cities Soc.* **2024**, *100*, 104998. [[CrossRef](#)]
24. Zhang, Y. Study on the vitality measurement and influence mechanism of urban blocks—A case study of Wuhan main city. Diss. Master's Thesis, Wuhan University, Wuhan, China, 2019.
25. Jacobs, J. *The Death and Life of Great American Cities*; Random House: New York, NY, USA, 1961.
26. Ye, Y.; Van Nes, A. Quantitative Tools in Urban Morphology: Combining Space Syntax, Space Matrix, and Mixed-Use Index in a GIS Framework. *Urban Morphol.* **2014**, *18*, 97–118. [[CrossRef](#)]
27. Cao, Z.M.; Zhen, F.; Li, Z.X. Study on Urban Temporal Vitality Patterns and Influencing Factors Based on Mobile Phone Signaling Data: A Case Study of the Central Urban Area of Nanjing. *Hum. Geogr.* **2022**, *37*, 109–117.
28. Liu, S.J.; Zhang, L.; Long, Y. Urban Vitality Area Identification and Pattern Analysis from the Perspective of Time and Space Fusion. *Sustainability* **2019**, *11*, 4032. [[CrossRef](#)]
29. Wu, G.S.; Dang, Y.T.; Zhao, K. Study on Urban Vitality Spatial Characteristics Based on Multi-dimensional Perception. *J. Earth Inf. Sci.* **2022**, *24*, 1867–1882.
30. Wang, B.; Zhen, F.; Zhang, H. Study on the Temporal and Spatial Dynamics of Urban Activities Based on Check-in Data and Area Division. *Geogr. Sci.* **2015**, *35*, 151–160.
31. Liu, Y.S.; Zhao, P.J.; Liang, J.S. Study on Urban Vitality Based on Location-Based Service Data: A Case Study of the Sixth Ring Area in Beijing. *Reg. Res. Dev.* **2018**, *37*, 64–69.
32. Wu, Z.Q.; Ye, Z.N. Study on Urban Spatial Structure Based on Baidu Map Heat Maps: A Case Study of the Central Urban Area of Shanghai. *Urban Plan.* **2016**, *40*, 33–40.



33. Jia, J.Y.; Song, J.F. Research on the Relationship Between Urban Vitality and Built Environment “3D” Characteristics: A Case Study of Wuhan. *Mod. Urban Res.* **2020**, *35*, 59–66.
34. Wang, N.; Wu, J.S.; Li, S. Research on the Spatial Characteristics of Urban Vitality and the Impact Mechanisms of the Built Environment Based on Multisource Data: A Case Study of Shenzhen. *Trop. Geogr.* **2021**, *41*, 1280–1291.
35. Long, Y.; Zhou, Y. Quantitative Evaluation and Influencing Factors Analysis of Street Vitality: A Case Study of Chengdu. *New Archit.* **2016**, 52–57.
36. Niu, X.Y.; Wu, G.S.; Li, M. Study on the Impact of Built Environment on Street Vitality and Its Temporal and Spatial Characteristics Based on LBS Location Data. *Int. J. Urban Plan.* **2019**, *34*, 28–37. [[CrossRef](#)]
37. Zhang, L.M.; Zhang, R.X.; Yin, B. The Impact of Built Environment on Pedestrian Activities in Historical Areas. *Alex. Eng. J.* **2021**, *60*, 285–300. [[CrossRef](#)]
38. Zou, S.C.; Zhang, S.Q.; Zhen, F. Study on Community Daily Activity Space Measurement and Vitality Influencing Factors Based on Residents’ Spatiotemporal Behavior: A Case Study of Shazhou and Nanyuan Streets in Nanjing. *Prog. Geogr.* **2021**, *40*, 580–596. [[CrossRef](#)]
39. Luo, S.Z.; Zhen, F. Study on Urban Public Space Vitality Evaluation Method Based on Mobile Phone Data: A Case Study of Parks in Nanjing. *Geogr. Res.* **2019**, *38*, 1594–1608.
40. Liu, S.; Lai, S.Q. Study on Influencing Factors of Urban Public Space Vitality Based on Multisource Data: A Case Study of the Huangpu River Waterfront Area in Shanghai. *Landsc. Archit.* **2021**, *28*, 75–81.
41. Shi, Y.; Li, T.T.; Yang, J.Y. Study on Urban Waterfront Space Vitality Based on Mobile Phone Signaling Data: A Case Study of Jinji Lake in Suzhou. *Landsc. Archit.* **2021**, *28*, 31–38.
42. Wang, M.W.; Jin, J.L.; Li, L. Application of Accelerated Genetic Algorithm-Based Projection Pursuit in Liquefaction Potential Evaluation of Sandy Soil. *Rock Soil Mech.* **2004**, *23*, 631–634.
43. Fu, Q.; Fu, H.; Wang, L.K. Application of Accelerated Genetic Algorithm-Based Projection Pursuit Model in Water Quality Evaluation. *Geogr. Sci.* **2003**, *23*, 236–239.
44. Ta, N.; Zeng, Y.T.; Zhu, Q.Y.; Wu, J.Y. Relationship Between Built Environment and Urban Vitality in Central Shanghai Based on Big Data Analysis. *Geogr. Sci.* **2020**, *40*, 60–68.
45. Wang, C.G.; Wang, B.; Wang, Q.Z. Nonlinear Relationship and Threshold Effect Between Urban Vitality and Built Environment: A Case Study of Guangzhou’s Central Urban Area. *Prog. Geogr.* **2023**, *42*, 79–88. [[CrossRef](#)]
46. Han, Y.; Qin, C.P.; Xiao, L.Z.; Ye, Y. The Nonlinear Relationships Between Built Environment Features and Urban Street Vitality: A Data-Driven Exploration. *Environ. Plan. B Urban Anal. City Sci.* **2024**, *51*, 195–215. [[CrossRef](#)]
47. Macau Statistics and Census Service. Available online: <https://www.dsec.gov.mo/zh-MO/Statistic/Database> (accessed on 3 April 2024).
48. Fang, L.; Huang, J.L.; Zhang, Z.Y.; Nitivattananon, V. Data-Driven Framework for Delineating Urban Population Dynamic Patterns: A Case Study on Xiamen Island, China. *Sustain. Cities Soc.* **2020**, *62*, 102365. [[CrossRef](#)]
49. Huang, B.; Zhou, Y.L.; Li, Z.G.; Song, Y.; Cai, J.; Tu, W. Evaluating and Characterizing Urban Vibrancy Using Spatial Big Data: Shanghai as a Case Study. *Environ. Plan. B Urban Anal. City Sci.* **2020**, *47*, 1543–1559. [[CrossRef](#)]
50. Ding, J.; Luo, L.; Shen, X.; Xu, Y. Influence of built environment and user experience on the waterfront vitality of historical urban areas: A case study of the Qinhuai River in Nanjing, China. *Front. Archit. Res.* **2023**, *12*, 820–836. [[CrossRef](#)]
51. Yu, L.; Sheng, Y.H.; Liu, X.Y. Research on urban vitality evaluation method based on RAGA-PPM model: A case study of central urban area in Nanjing. *J. Geo-Inf. Sci.* **2024**, *26*, 2567–2582.
52. Cao, Q.F.; Jiang, Q.X.; Zheng, L.L. Improving the compactibility of Gardeniae Fructus extract by HPMC based on systematic optimization with QbD concept and FAHP-entropy weight method. *Chin. Tradit. Herb. Drugs* **2024**, *55*, 5448–5458.
53. Liu, Y.; Wang, K.L.; Xing, X.Y. Spatial Effects in Geographic Analysis. *Acta Geogr. Sin.* **2023**, *78*, 517–531.
54. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [[CrossRef](#)]
55. Friedman, J.H.; Tukey, J.W. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **1974**, *23*, 881–890. [[CrossRef](#)]

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