land

Special Issue Reprint

Feature Papers for “Land Innovations—Data and Machine Learning” Section

3rd Edition

Edited by
Chuanrong Zhang

mdpi.com/journal/land

MDPI

Feature Papers for "Land Innovations—Data and Machine Learning" Section: 3rd Edition

Feature Papers for "Land Innovations—Data and Machine Learning" Section: 3rd Edition

Guest Editor

Chuanrong Zhang

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor
Chuanrong Zhang
Department of Geography &
Center for Environmental
Sciences and Engineering
University of Connecticut
Storrs, CT
USA

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Land* (ISSN 2073-445X), freely accessible at: https://www.mdpi.com/journal/land/special_issues/9ML34931E9.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-9017-0 (Hbk)
ISBN 978-3-7258-9018-7 (PDF)
https://doi.org/10.3390/books978-3-7258-9018-7

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editor . . . . . **vii**

Darshana Athukorala, Yuji Murayama, Siri Karunaratne, Rangani Wijenayake, Takehiro Morimoto, S. L. J. Fernando and N. S. K. Herath
Spatio-Temporal Changes in Mangroves in Sri Lanka: Landsat Analysis from 1987 to 2022
Reprinted from: *Land* **2025**, *14*, 1820, https://doi.org/10.3390/land14091820 . . . . . **1**

Changwei Linghu, Zhipeng Qian, Weizhe Chen, Jiaren Li, Ke Yang, Shilin Zou, et al.
Assessment of Landslide Susceptibility and Risk in Tengchong City, Southwestern China Using Machine Learning and the Analytic Hierarchy Process
Reprinted from: *Land* **2025**, *14*, 1966, https://doi.org/10.3390/land14101966 . . . . . **27**

Yuxi Liu, Xinyu Du, Yu Bai, Qibing Chen and Shiliang Liu
Bridging Heritage Systems: Multi-Scale Spatial Coupling Between Tangible and Intangible Cultural Heritage in China Using Hierarchical Bayesian Model and Causal Inference
Reprinted from: *Land* **2025**, *14*, 2386, https://doi.org/10.3390/land14122386 . . . . . **49**

Yusheng Zheng, Xinying Huang and Huanmei Yao
Towards Trustworthy Urban Land Use Classification: A Synergistic Fusion of Deep Learning and Explainable Machine Learning with a Nanning Case Study
Reprinted from: *Land* **2026**, *15*, 158, https://doi.org/10.3390/land15010158 . . . . . **81**

Arnob Bormudoi and Masahiko Nagai
Disentangling Complexity and Performance: A Comparative Study of Deep Learning and Random Forest Models for Cropland Vulnerability Assessment in Bangladesh
Reprinted from: *Land* **2026**, *15*, 174, https://doi.org/10.3390/land15010174 . . . . . **108**

Kangcheng Zhu, Yuzhong Kong, Xiangyun Kong, Sen Hu, Junmeng Zhao, Ciren Pu, et al.
Landslide Susceptibility Assessment Based on a TSPF-BiLSTM Model: A Case Study of Sangzhi County, Hunan Province
Reprinted from: *Land* **2026**, *15*, 579, https://doi.org/10.3390/land15040579 . . . . . **131**

Yingkai Tang, Wangping Liu, Xi Yao, Liangzhao Chen and Min Li
Beyond Sprawl: How Urban Morphology Shapes Carbon Emission Intensity Categories via SHAP-PDP Framework
Reprinted from: *Land* **2026**, *15*, 738, https://doi.org/10.3390/land15050738 . . . . . **155**

Yunlong Wei and Hu Li
Analysis of Factors Influencing Fire Risk in High-Density Urban Areas Based on the CatBoost-SHAP Model
Reprinted from: *Land* **2026**, *15*, 796, https://doi.org/10.3390/land15050796 . . . . . **184**

Xiaoyu Ying, Shenbo Ni, Jiajing Wu, Yujie Zhao, Haiqiang Liu, Rongxin Qiu, et al.
Generative Regeneration of Historic Urban Fabric: A Framework Based on Deep Learning and Multi-Objective Optimization
Reprinted from: *Land* **2026**, *15*, 976, https://doi.org/10.3390/land15060976 . . . . . **215**

About the Editor

Chuanrong Zhang

Chuanrong (Cindy) Zhang is a professor in the Department of Geography and the Center for Environmental Science and Engineering at the University of Connecticut. With over 20 years of experience, her research focuses on GIS, remote sensing, and spatial analysis. She has conducted extensive work on the spatial analysis of environmental and socioeconomic data using spatial statistics and GIS methodologies. Dr. Zhang has published three books and over 200 peer-reviewed articles, and has collaborated on numerous research projects supported by major funding agencies, including the National Science Foundation, the National Institutes of Health, the U.S. Department of Energy, the U.S. Geological Survey, and NASA.

land

MDPI

Article

Spatio-Temporal Changes in Mangroves in Sri Lanka: Landsat Analysis from 1987 to 2022

Darshana Athukorala [1,*], Yuji Murayama [1], Siri Karunaratne [2], Rangani Wijenayake [3], Takehiro Morimoto [1], S. L. J. Fernando [4] and N. S. K. Herath [5]

[1] Institute of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8572, Japan; murayama.yuji.gb@u.tsukuba.ac.jp (Y.M.); morimoto.takehiro.fe@u.tsukuba.ac.jp (T.M.)
[2] Land Use Policy Planning Department, Ministry of Lands, Colombo 00500, Sri Lanka; slskarukanlup@gmail.com
[3] National Institute for Environmental Studies, Tsukuba 305-8506, Japan; wijenayake.rangani@nies.go.jp
[4] Department of Geography, Faculty of Humanities and Social Sciences, University of Ruhuna, Matara 81000, Sri Lanka; fernando@geo.ruh.ac.lk
[5] Department of Environmental Management, Rajarata University of Sri Lanka, Mihintale 50300, Sri Lanka; samankherath4@ssh.rjt.ac.lk
* Correspondence: darshana.athukora.gf@u.tsukuba.ac.jp

Abstract

Mangroves in Sri Lanka provide critical ecosystem services, yet they have undergone significant changes due to anthropogenic and natural drivers. This study presents the first national-scale assessment of mangrove dynamics in Sri Lanka using remote sensing techniques. A total of 4670 Landsat images from Landsat 5, 7, 8, and 9 were selected to detect mangrove distribution, changes in extent, and structure and stability patterns from 1987 to 2022. A Random Forest classification model was applied to elucidate the spatial changes in mangrove distribution in Sri Lanka. Using national-scale data enhanced mapping accuracy by incorporating region-specific spectral and ecological characteristics. The average overall accuracy of the maps was over 96.29%. The total extent of mangroves in 2022 was 16,615 ha, representing 0.25% of the total land of Sri Lanka. The results further indicate that, at the national scale, mangrove extent increased from 1989 to 2022, with a net gain of 1988 ha (13.6%), suggesting a sustained and continuous recovery of mangroves. Provincial-wise assessments reveal that the Eastern and Northern Provinces showed the largest mangrove extents in Sri Lanka. In contrast, the Colombo, Gampaha, and Kalutara districts in the Western Province showed persistent declines. The top mangrove spatial structure and stability districts were Jaffna, Trincomalee, and Gampaha, while the most degraded mangrove districts were Batticaloa, Colombo, and Kalutara. This study offers critical insights into sustainable mangrove management, policy implementation, and climate resilience strategies in Sri Lanka.

Keywords: mangrove; Landsat; remote sensing; spatial analysis; fragmentation; Sri Lanka

1. Introduction

Mangroves primarily consist of salt-tolerant plants in natural transitions between coastal marine environments and dry terrestrial ecosystems [1–3]. They are important to humanity, providing numerous ecosystem services such as carbon sequestration, coastal protection, fisheries, biodiversity, water purification, sediment stabilization, livelihoods, and economic benefits [4–6]. Mangroves also play an essential role in the Sustainable Development Goals (SDGs), especially SDG 1: End poverty in all its forms everywhere;

SDG 13: Take urgent action to combat climate change and its impacts; SDG 14: Conserve and sustainably use the oceans, seas, and marine resources for sustainable development; and SDG 15: Life on Land [7]. However, increasing anthropogenic activities [8–10] and the growing pressure of climate change influence mangroves [11–13] and their combined ecosystem services in numerous ways [14–17].

Many studies have reported mangrove losses [18–20]. However, mangroves can respond to environmental changes at local and regional scales, increasing their extent via physical propagation—sedimentation, encroachment, and segmentation [21,22]. Large-scale mangrove rehabilitation activities further propagate and recover these natural processes [23,24]. Previous studies have reported that mangroves are particularly vulnerable to ecological and anthropogenic change [11,25,26]; however, our understanding of their impact on mangroves remains unclear.

Today, mangroves are recognized as one of the most vulnerable ecosystems and have been affected by two main processes: land-direction and sea-direction processes [27,28]. Land-direction processes include human activities, hydrological alteration, and tidal flooding hampers such as dikes and roadways [2,27]. Accelerated sea-level rise, strong storm waves, tsunamis, and other climate phenomena are examples of sea-level processes [2,27,28]. Impacted mangrove forests show reduced ecosystem services [5,20,29,30], potentially modifying their role in the global carbon cycle from a carbon sink to a carbon source [11,31,32]. Therefore, detecting mangrove changes is crucial for restoring and enhancing carbon sequestration, as well as supporting national and international climate initiatives.

Mangrove change detection is required to identify their future trajectories. Generally, traditional surveying methods are challenging to apply in mangrove detection due to the large physical environment, and these techniques are often used for relatively small areas over a short period [33,34]. Furthermore, due to the challenges of processing large quantities of Earth observation data, traditional mangrove detection systems generally use fewer images at a particular time [35,36]. Conversely, the continuous advancement of state-of-the-art technology (remote sensing) and the increasing availability and accessibility of the data offer a significant opportunity for the spatio-temporal monitoring of large-scale and long-term changes in mangroves [37,38]. For example, Global Mangrove Watch (GMW) aims to provide comprehensive, up-to-date data on the distribution, status, and changes in mangroves worldwide [39]. The GMW provides valuable insights at the global level; however, direct application of data at the regional and national levels remains challenging [40]. The Google Earth Engine (GEE) cloud-based data computing system has emerged as a prompt and suitable solution to this problem [41].

Previous studies have revealed patterns of mangrove loss and gain due to coastal development, aquaculture expansion, climate change, and conservation efforts worldwide [42,43]. These changes are not consistent, with distinct trends often observed between the Global North and the Global South [44,45], as well as between tropical and subtropical regions [46,47]. For instance, some areas in the Global North exhibit stable or expanding mangrove cover due to protection and restoration initiatives [44]. Conversely, many regions in the Global South, particularly Latin America, Southeast Asia, and South Asia, have experienced substantial degradation of their mangrove forests. Tropical mangroves include the majority of the global mangrove area and are especially vulnerable to anthropogenic pressures and sea-level rise. For example, extensive deforestation is observed in the Mekong Delta [48], the Caribbean mangroves [49], and in more localized yet significant areas, such as parts of West Africa, where clearing occurs for urban expansion [50]. Sri Lanka, situated in the tropical zone near the equator, possesses significant mangroves that are subject to similar pressures, including encroachment for aquaculture, agriculture, coastal infrastructure development, urbanization, pollution, and climate impacts, mak-

ing it a crucial area for focused mangrove change detection studies to inform national conservation and management strategies [51].

Sri Lanka stretches between 6° N and 10° N latitude and 79° E and 82° E longitude in the Indian Ocean—the total area of the country is approximately 65,610 km^2 [52]. The coastline stretches approximately 1340 km. The island features a mountainous area in the southern and central regions, as well as a surrounding coastal plain with rich ecosystems, including mangroves. Mangroves in Sri Lanka are distributed across fourteen coastal districts in five provinces (Table S1). The wet and dry climate conditions of Sri Lanka are favorable for propagating mangroves [53]. Sri Lanka is particularly vulnerable to climate change due to its high precipitation and temperatures, complex hydrological cycle, and exposure to catastrophic weather events, especially cyclonic activities from the Bay of Bengal and the Arabian Sea [53].

Figure 1 shows the selected locations of mangroves along the coastal areas of Sri Lanka. Sri Lanka is home to six Ramsar wetlands, recognized globally for their exceptional biological and ecological significance [54]. Moreover, due to its rich wetland ecology, the capital, Colombo, was declared a Ramsar wetland city by the Ramsar in 2018 [54,55].



Figure 1. Mangroves in Sri Lanka. (**a**) Mangroves at Trincomalee Bay; (**b**) Mangroves at Uppu Aru; (**c**) Mangroves at Kala Oya Estuary; (**d**) Mangroves at Muthurajawela Wetland; and (**e**) Mangroves near Chilaw Lagoon. Images from Google Earth Pro [56]. The location presented in Figure 1 is shown in Figure 2.

Despite having a complex mangrove ecology, it is surprising that there are no high-resolution, high-accuracy, and long-term mangrove distribution data to track mangrove dynamics in Sri Lanka. The International Union for Conservation of Nature (IUCN) Sri Lanka and the Central Environmental Authority (CEA) of Sri Lanka published the national wetland directory in 2006 [51]. However, there has yet to be a concrete program to monitor wetland changes in the country, including those of mangroves. Many studies have focused on wetland mapping and change over the past decades in Sri Lanka [57–59]. However, these studies are considered relatively small in terms of wetland areas [58–60]. Therefore, an overall view of the mangroves of Sri Lanka is currently needed to understand past and present changes and their future trajectories.

Many studies have demonstrated the significance of analyzing mangrove spatial structure, fragmentation, and overall configuration for sustainable mangrove conservation, enhancing restoration efforts, and improving ecosystem services [37,57,61]. However, in

Sri Lanka, few studies have systematically addressed the structural and stability patterns of mangroves [57]. Therefore, a fundamental assessment is required to detect changes in spatial structure, fragmentation, and stability patterns, thereby improving core habitat integrity and ecosystem services across the entire country of Sri Lanka. Addressing this gap is crucial for developing evidence-based conservation strategies, enhancing ecosystem restoration efforts, and strengthening Sri Lanka's contributions to global climate mitigation. Therefore, our study area is defined by creating a coastal buffer that extends 10 km into the sea and the land area from the coastline of Sri Lanka (representing the extent of mangroves in Sri Lanka), because mangroves typically occur within low-lying coastal zones influenced by tidal action, salinity, and sedimentation (Figure 2).

This study aims to (i) detect mangrove distribution in Sri Lanka from 1987 to 2022, utilizing available Landsat imagery to ensure long-term temporal consistency and spatial coverage across the entire country, using the national-scale data by incorporating region-specific spectral and ecological characteristics, enabling a robust assessment of mangrove distribution and dynamics; (ii) monitor and analyze mangrove gain and loss in different scales, including national, provincial, district levels, and climatic zones; and (iii) analyze the structure and stability patterns of mangroves in Sri Lanka, emphasizing whether they are static or changing spatiotemporally over time. The results of this study can provide theoretical guidelines for future mangrove research and the sustainable management of mangroves in Sri Lanka. Scheme 1 shows the selected locations of mangroves along the coastal areas of Sri Lanka for visual interpretation.

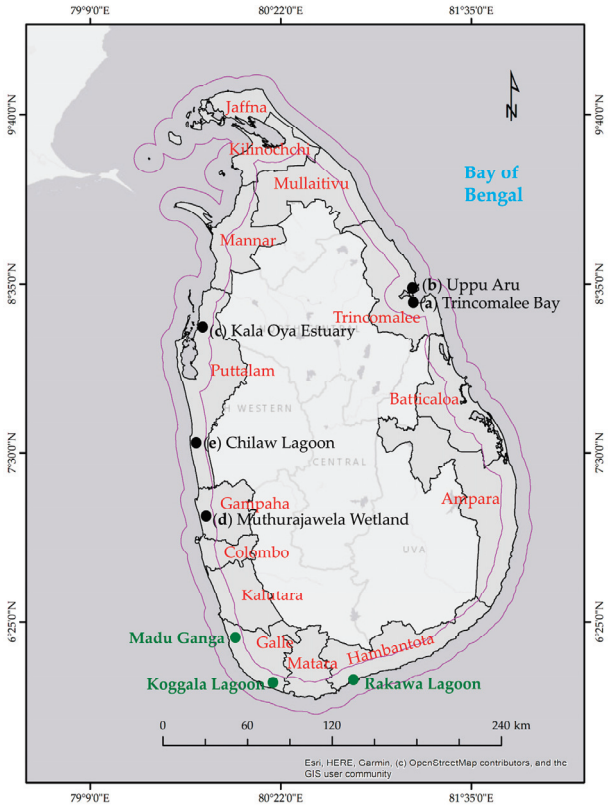


Figure 2. Sri Lanka and its surrounding environment. Red labels on the map indicate the coastal districts of Sri Lanka. The purple buffer represents the coastal zone used in this study, extending 10 km seaward and 10 km inland. Basemap source: Esri Map [62]. The black labels on the figure indicate the locations shown in Figure 1. The green labels on the figure indicate the locations shown in Scheme 1.



Scheme 1. Mangroves in Sri Lanka. Source: Photographs taken by the first author during the field survey near Madu Ganga (**a**–**c**), Rakawa Lagoon (**d**,**f**), and Koggala Lagoon (**e**) in Sri Lanka (2025).

2. Materials and Methods

2.1. Typology of Mangrove Mapping

This study used state-of-the-art technology (remote sensing) and other supplementary data to establish a framework for mangrove distribution dynamics and changes in Sri Lanka. In our study, mangroves are considered generally inundated throughout the tide cycle and comply with the definitions of the IUCN Global Ecosystem Typology (Typology code MFT1.2—intertidal forests and shrublands considered mangroves) [63]. Furthermore, we also considered the definition of Sri Lanka's coastal zone as outlined in the Coastal Conservation Act of Sri Lanka [64]. Then, we developed a classification system based on remote sensing data, supplemented by ancillary data, to detect the extent of mangroves at ten stages from 1987 to 2022, based on the availability of Landsat data (see Sections 2.2 and 2.3).

2.2. Landsat Images

The Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper+ (ETM+), and Landsat 8 and 9 Operational Land Imager (OLI) are the most critical data sources for detecting long-term and large-scale changes in terrestrial ecosystems [65]. In

this study, the Landsat Collection 2 atmospherically corrected surface reflectance data was obtained from the GEE from 1987 to 2022 (4670 images) (Table S2), a cloud and cloud shadow mask was applied using the CF mask algorithm [66], and a median image composite was used for three-yearly periods (1987–1989, 1993–1995, 1996–1998, 2002–2004, 2005–2007, 2008−2010, 2011−2013, 2014−2016, 2017−2019, 2020−2022)—totaling ten mangrove stages. The three years were chosen due to the availability of Landsat data and other ancillary data to identify visible changes in the mangroves, as well as to minimize the influence of extreme events. The mangrove layer is named based on the last year of the three years considered. For example, if 1987, 1988, and 1989 are used, the mangrove layer is named 1989. This approach is applied to all other stages.

Figure 3 shows the Landsat images used in this study. Landsat 8 provided the highest number of images (1843), followed by Landsat 5 (1503) and Landsat 7 (1104), while Landsat 9 contributed the least (220 images) due to its recent launch (Table S2). The historical contributions of Landsat 5 and Landsat 7 play a significant role in the long-term monitoring of mangroves in this study. The three most frequently used years were 2022 (810 images), 2004 (538 images), and 2009 (494 images), emphasizing the critical years of data availability for mangrove detection in Sri Lanka. The 141/55, 142/54, 141/54, 142/53, and 140/55 paths and rows of Landsat made the highest contribution to mangrove mapping in Sri Lanka.

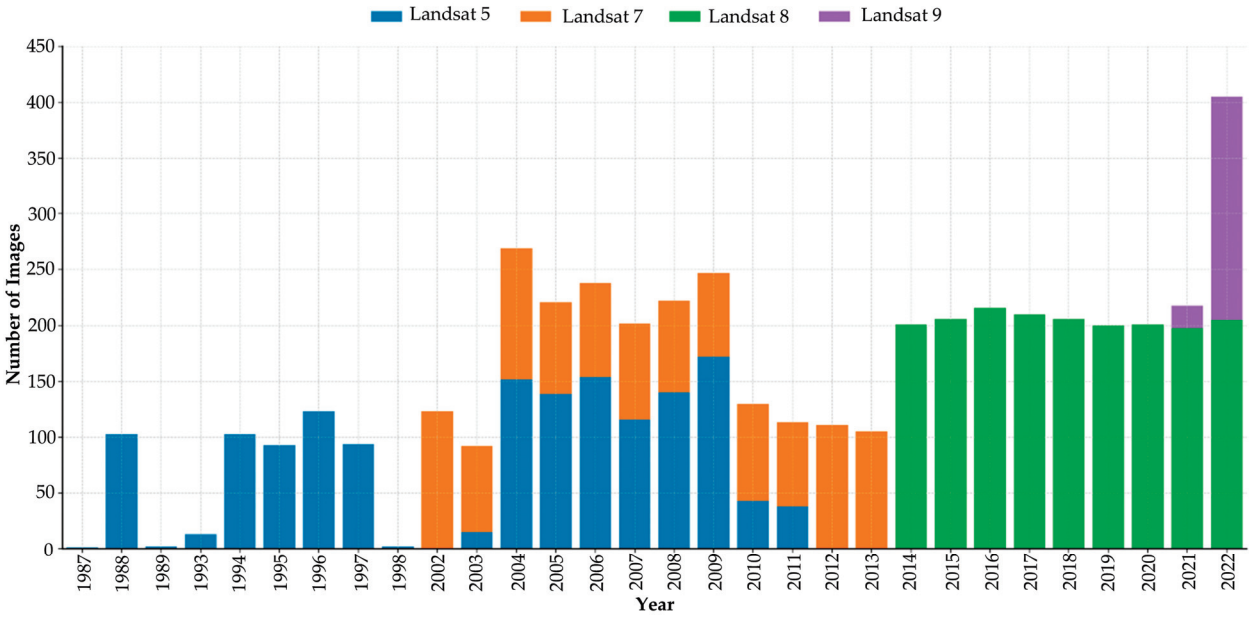


Figure 3. Number of Landsat images used for this study.

2.3. Mangrove Detection in Sri Lanka

Figure 4 shows the schematic diagram of the study. The methodology encompasses mangrove detection, spatial analysis—including loss and gain change detection—assessment of mangrove occurrence from 1989 to 2022, and evaluation of mangrove structural and stability patterns. Before delineating the study area, we visually interpreted the coastal area using other ancillary data and applied a 10 km buffer along the country's coastline. Ultimately, the study presents evidence-based recommendations to ensure the long-term sustainability of mangroves in Sri Lanka.

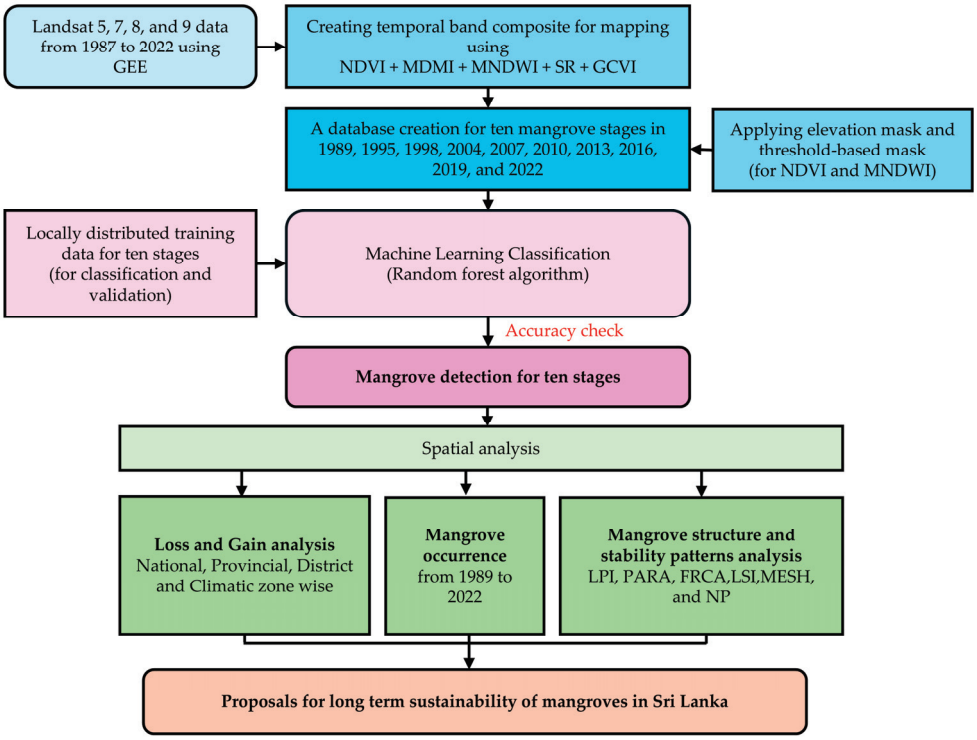


Figure 4. The schematic diagram of the study.

2.3.1. Spectral Indices Used in This Study for the Temporal Band Composite

Spectral indices are mathematical combinations of different band wavelengths that enhance specific image features. Several indices were used for mangrove detection to emphasize vegetation health, moisture content, and water presence. We used the following indices to create the temporal band composite for mangrove detection, as they are relevant to mangrove ecosystems in Sri Lanka [67,68].

Normalized Difference Vegetation Index (NDVI)

NDVI is one of the most established indices for vegetation monitoring. NDVI values range from −1 to 1. The higher values indicate dense and healthier vegetation. In the context of mangroves, their dense canopies generally produce high NDVI values. Therefore, the NDVI is crucial because it provides a baseline measure of vegetation health, helping to distinguish between mangroves and non-vegetated areas [69,70].

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}} \tag{1}$$

where NIR and Red are the reflectance values of NIR and Red of Landsat 5, 7, 8, and 9, respectively.

Normalized Difference Mangrove Index (NDMI)

The NDMI is a spectral index used to identify mangroves. The NDMI is particularly sensitive to moisture content in mangroves. Mangroves are often located in waterlogged areas and intertidal zones. The NDMI can isolate mangroves from other vegetation. The enhanced sensitivity of the NDMI to moisture conditions makes it a valuable spectral index for identifying mangroves [71,72].

$$\mathrm{NDMI} = \frac{\mathrm{SWIR} - \mathrm{Green}}{\mathrm{SWIR} + \mathrm{Green}} \tag{2}$$

where SWIR and Green are the reflectance values of SWIR and green bands of Landsat 5, 7, 8, and 9, respectively.

Modified Normalized Difference Water Index (MNDWI)

The MNDWI is used to identify water areas, a critical characteristic for mangrove detection (due to the close relationship between mangroves and aquatic environments). Mangroves thrive in intertidal zones and near estuaries and lagoons—a strong water signal (evident from the MNDWI) supports the separation of vegetation (e.g., mangroves) from water areas. The index is calculated as follows [73,74].

$$\text{MNDWI} = \frac{\text{Green} - \text{SWIR 1}}{\text{Green} + \text{SWIR 1}} \tag{3}$$

where Green and SWIR1 are the reflectance values of Green and SWIR1 of Landsat 5, 7, 8, and 9, respectively.

Simple Ratio (SR)

The simple ratio can distinguish mangroves, emphasizing NIR reflectance and low red absorption. The simple ratio is ideal for detecting health assessment and monitoring mangrove changes in the target areas [65].

$$\text{SR} = \frac{\text{NIR}}{\text{RED}} \tag{4}$$

where NIR and RED are the reflectance values of NIR and RED of Landsat 5, 7, 8, and 9, respectively.

Green Chlorophyll Vegetation Index (GCVI)

The GCVI is crucial for monitoring mangrove health, biomass, and photosynthetic activity because it can evaluate stress conditions, degradation, and restoration of mangroves. The GCVI is sensitive to the chlorophyll content in vegetation and is an effective indicator of plant health. High GCVI values indicate vigorous photosynthetic activity, suggesting the mangroves' good health condition. The GCVI can complement the NDVI. However, the GCVI provides additional insight into the biochemical properties of vegetation. The GCVI is calculated as follows [75,76]:

$$\text{GCVI} = \frac{\text{NIR}}{\text{Green}} + 1 \tag{5}$$

where NIR and Green are the reflectance values of NIR and Green wavelengths of Landsat 5, 7, 8, and 9, respectively.

2.3.2. Elevation Masking

Generally, mangroves are found in low-elevation coastal areas. Therefore, we used the elevation mask on the band composite using the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM) [77]. This analysis used elevation values of less than 60 m (<60 m). The elevation masking is critical because it removes highland areas from the band composite [78]. Hence, we modified the mangrove detection criteria by applying an elevation mask, focusing on low-lying areas, such as coastal areas, lagoons, and estuaries.

2.3.3. Threshold-Based Masking

Our study utilized threshold-based masks in the created band composite to further isolate potential mangrove areas. We used NDVI values (NDVI > 0.20) and MNDWI

values (MNDWI > −0.50) based on previous studies to help eliminate background noise and non-mangrove and water areas [79,80]. Before applying the threshold values, we visually interpreted them in the context of mangrove detection in Sri Lanka. We verified the precision by applying several threshold values and adjusting them (for the NDVI from 0.0 and the MNDWI from −1.0) at each stage. The identified and used values were ideal for creating a temporal band composite to detect mangroves in Sri Lanka (see Figure 5).

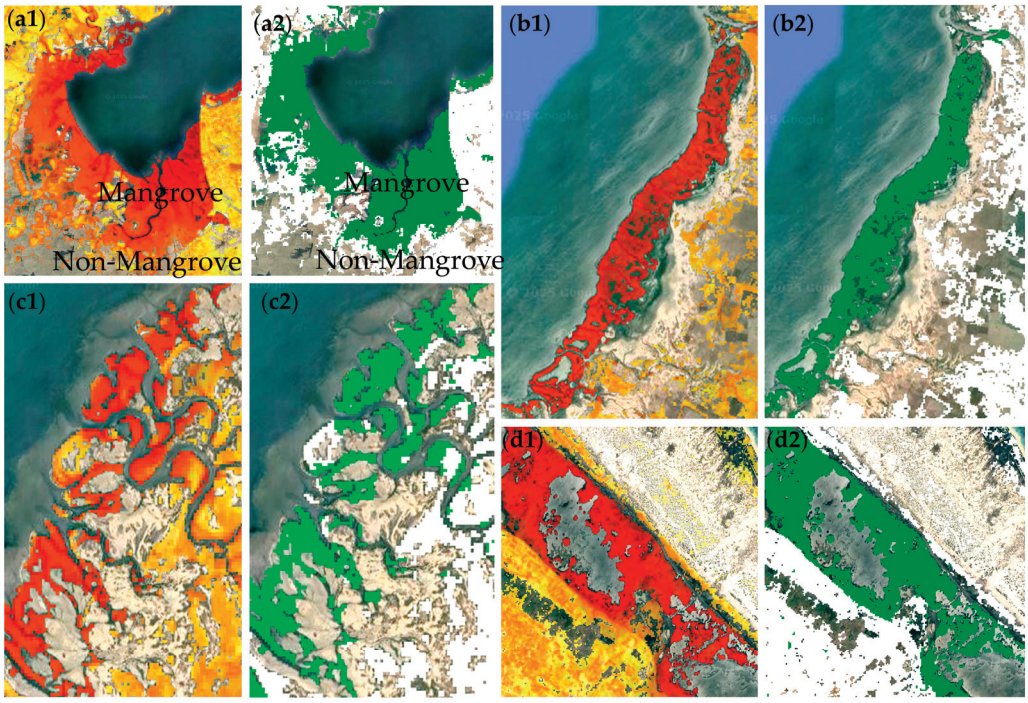


Figure 5. The selected locations of the band composites used in this study and the results after mangrove detection in 2022 using GEE. (**a1**,**a2**) band composite of Trincomalee Bay and the result of mangrove detection; (**b1**,**b2**) band composite of Vidaththalthivu and the result of mangrove detection; (**c1**,**c2**) band composite of Iluppaikadavai and the result of mangrove detection; and (**d1**,**d2**) band composite of Vadamarachchi and the result of mangrove detection.

2.3.4. Classification

Reference Samples

The accurate classification of mangrove versus non-mangrove areas depends on representative reference samples. In this study, a 'sample' refers to a single georeferenced point that has been labeled as either mangrove or non-mangrove based on reference data. Using locally distributed samples, we created a sample dataset (for training and validation) to detect mangroves in Sri Lanka. We used 70% of the samples for testing and 30% for validation (each class includes 2000 mangrove and 2000 non-mangrove samples, totaling 4000 samples per stage). The reference samples were obtained from topographic maps and other published maps from the Survey Department of Sri Lanka [52], published studies [51], Google Earth Pro [81], and visual interpretation of high-resolution satellite images (Table S3). During the mapping process, we also leveraged our expertise and understanding of the visual characteristics of mangroves, as well as our knowledge of the Sri Lankan coastal belt, to compile a reference dataset. Each classified map was cross-checked using ancillary data and visual interpretation. Misclassified areas were removed from the initial classified maps by comparing them with related data and knowledge of the Sri Lankan coastal belt.

Machine Learning Classification Using the Random Forest Method

The Random Forest (RF) classifier is selected because of its robustness and ability to handle multi-dimensional datasets [81,82]. The RF classification operates by combining multiple decision trees for classification tasks. RF can handle large datasets with multiple input variables, such as spectral bands and remote sensing indices, without overfitting. Moreover, RF can manage complex, non-linear relationships in land use and land cover, making it highly suitable for diverse and heterogeneous landscapes [81–83]. In our study, the number of trees and the number of random samples per node were 500 and 5, respectively. In this study, the RF classifier utilizes variables that include computed indices such as NDVI, NDMI, MNDWI, SR, and GCVI, as well as additional masks, including elevation, NDVI, and MNDWI (see Sections 2.3.2 and 2.3.3). Figure 5 shows the selected locations of the band composites used in this study, along with the results after mangrove detection in 2022.

2.4. Mangrove Extent and Its Changes

This study estimated the extent of mangroves and their changes over time. The results were summarized at several spatial scales, including national, provincial, and district levels, as well as climatic zone levels, to support sustainable management practices and enable robust monitoring of mangroves in Sri Lanka (Figures 2 and S1). Moreover, we investigated ten ecologically significant mangrove areas in Sri Lanka to identify changes in these areas over the study period (1987–2022). In the context of mangrove occurrence in Sri Lanka, we averaged the generated mangrove layers and categorized them into occurrence levels ranging from 100% to 10%. The results were discussed in relation to the conservation and long-term sustainability of mangroves in Sri Lanka.

2.5. Mangrove Structure and Stability Pattern Analysis

Analysis of mangrove structure and stability patterns is crucial for understanding landscape fragmentation, structural patterns, spatial configuration, and ecological functionality in Sri Lanka. Our study used six parameters, such as the Largest Patch Index (LPI), Perimeter-Area Ratio (PARA), Fractal Dimension Index (FRAC), Landscape Shape Index (LSI), Effective Mesh Size (MESH), and Number of Patches (NP) (Table S4), to provide a scientific basis for assessing mangrove fragmentation, connectivity, and habitat quality changes from 1989 to 2022. These indices are widely used in the ecological field and are particularly important for evaluating the health, structure, and functions of mangrove ecosystems [84,85]. In addition, the parameters collectively assist in exploring the impact of anthropogenic activities (urbanization, aquaculture, and deforestation) and natural pressures (sea-level rise and coastal erosion) [84,86].

We conducted a district-wise assessment to support government efforts and facilitate administrative monitoring of mangrove-related projects. Furthermore, district-level evaluations enable close tracking of mangrove sites, facilitate the initiation of restoration projects, and promote easier community engagement. This local approach also helps promote community-based mangrove management by using the knowledge of local people who are familiar with the mangrove ecosystems in their own districts. Including this local knowledge in conservation planning will strengthen mangrove protection efforts in Sri Lanka.

3. Results

3.1. Mangrove Changes in Sri Lanka

The mangroves in Sri Lanka exhibited significant changes from 1987 to 2022 due to anthropogenic activities, natural disturbances, and ecological regeneration processes

(Table 1). The average overall accuracy of the maps was over 96.19% (Table S5). The results show a sharp decline in mangrove extent from 14,627 ha in 1989 to 8596 ha in 1998 (Figure 6), which can be attributed to the civil war period, urban development, and the expansion of aquaculture. The Indian Ocean tsunami in 2004 further intensified mangrove loss (8279 ha by 2004), showing the lowest mangrove year during the study period. The massive tsunami waves uprooted mangrove trees, altered coastal geomorphology, and increased soil salinity, thereby affecting the mangrove regeneration process. However, some post-tsunami conservation practices and natural propagation helped to increase mangroves in the subsequent years (from 9820 ha in 2007 to 16,615 ha in 2022, increasing 6795 ha).

Table 1. Total area of mangroves in Sri Lanka: National scale.

Year	1989	1995	1998	2004	2007	2010	2013	2016	2019	2022
Area (ha)	14,627	10,749	8596	8279	9820	10,640	11,837	13,755	15,490	16,615

3.2. Provincial-Wise Mangrove Distribution in Sri Lanka

Provincial-wise mangrove distribution in Sri Lanka revealed significant spatial and temporal variations from 1989 to 2022 (Table 2). The Eastern Province showed the highest extent of mangroves (Table 2). Despite their decline from 1989 to 2004, the Eastern Province gradually recovered from 2007, reaching 6294 ha in 2022 (Table 2). The Northern Province has experienced a considerable increase in mangrove coverage since 2004, with a significant net gain of 1616 ha by 2022, compared to the other provinces (Table 2). The Northwestern Province exhibited a continuous decline in mangrove extent until 2007, followed by a modest recovery from 2007 (Table 2). However, the net mangrove change remained negative from 1989 to 2022. The Southern Province indicates a fluctuating trend in mangroves (initial declines until 2004; a recovery stage showed an overall gain of 1009 ha by 2022) (Table 2), indicating enhanced mangrove conservation efforts and natural propagation processes. The Western Province showed a continuous decline (net loss of 559 ha from 1987 to 2022) due to the high urban expansion and intensive coastal infrastructure development. The persistent loss in the Western Province further indicates the significant pressure of human activities on mangroves.

Table 2. Total area of mangroves in Sri Lanka (ha): Provincial-wise.

Year	Eastern	Northwestern	Northern	Southern	Western	Total
1989	6247.31	2268.2	2941.21	1737.33	1433.42	14,627.47
1995	5044.52	1685.03	1834.03	1204.65	980.73	10,748.96
1998	3622.01	1198.25	1338.68	1664.35	773.12	8596.41
2004	3244.97	1175.81	2016.32	1132.56	709.71	8279.37
2007	3861.75	1175.7	2325.56	1637.01	820.44	9820.46
2010	4585.96	1401.46	2120.09	1629.94	902.47	10,639.92
2013	4437.17	1679.46	3421.07	1561.15	738.01	11,836.86
2016	5347.86	1925.55	3933.97	1744.39	803.21	13,754.98
2019	5790.4	2079.73	4376.06	2381.79	861.95	15,489.93
2022	6293.6	2144.57	4556.73	2745.9	874.43	16,615.23
Average	4847.56	1673.38	2886.37	1743.91	889.75	
Change	46.29	−123.63	1615.52	1008.57	−558.99	

3.3. District-Wise Mangrove Distribution in Sri Lanka

District-wise mangroves in Sri Lanka reveal substantial spatio-temporal changes from 1989 to 2022 (Table S6). Trincomalee, Puttalam, Ampara, Mannar, and Batticaloa districts showed the top five districts with the highest mangrove extent, respectively (Table S7).

The districts with the lowest mangrove coverage were Kalutara, Colombo, and Matara (Table S7). Hambantota, Jaffna, and Trincomalee were the top three districts with the highest average mangrove gain (Table S8). At the same time, Gampaha, Ampara, and Batticaloa were the districts with the highest average loss from 1989 to 2022 in Sri Lanka (Table S9). The districts with historically larger mangrove extents, such as the Gampaha and Galle, showed significant declines over the study period. The Gampaha district had the most considerable mangrove extent (1252 ha in 1989), showing a sharp decline by 2004 and a partial recovery (820 ha in 2022). The Galle district showed considerable interannual variability (initial reduction in the 1990s and partial regeneration after 2010) (Table S6). In contrast, smaller mangrove extent districts, such as Colombo and Kalutara, showed a consistent mangrove loss. For example, the Colombo district dramatically decreased from 109 ha in 1989 to 29 ha in 2022 (an overall decline of 73%) (Table S6). Kalutara district showed a similar trajectory, with mangroves decreasing from 73 ha in 1987 to 26 ha in 2022 (an overall decline of 64%) (Table S6).

3.4. Mangrove Occurrence in Sri Lanka

Figure 6 shows the occurrence of mangroves in Sri Lanka from 1989 to 2022.

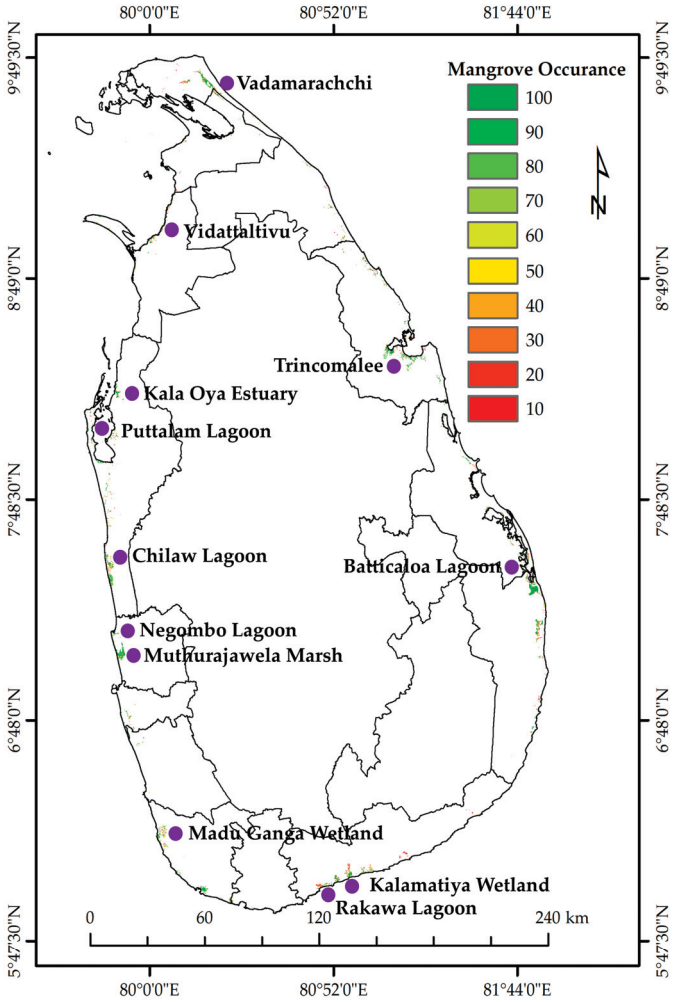


Figure 6. The occurrence of mangroves in Sri Lanka from 1989 to 2022. The labels in the figure indicate the most significant mangrove-bearing areas in Sri Lanka.

3.4.1. Mangrove Occurrence in Climatic Zone-Wise

Mangroves in Sri Lanka show distinct variations by climatic zone (Figure S1 and Table S10). The dry zone has the highest mangrove extent (11,383 ha in 100% occurrence),

significantly surpassing the wet zone (2285 ha) and intermediate zone (760 ha), respectively (Table S10). Even at lower occurrence levels (60–90%), the dry zone consistently bears the largest extent of mangroves along the northeastern, eastern, southeastern, and northwestern coasts (the areas that belong to the dry zone in Sri Lanka)—particularly, Puttalam, Trincomalee, Vedithaththe, and Batticaloa. The Jaffna Peninsula also plays a vital role in the mangroves of the dry zone. The wet zone has a limited extent of mangrove (Table S10), particularly along the Kelani River estuary, Muthurajawela, and Negombo Lagoon. The intermediate zone has the least mangrove extent across all occurrence levels (Table S10).

3.4.2. Mangrove Occurrence in Eight Selected Areas

The eight locations were selected due to their high ecological significance, extensive mangrove coverage, unique geographical background, and contribution to coastal stability in Sri Lanka (Figure 7 and Table S11). Trincomalee, Puttalam, Batticaloa, Vidattativu, Chilaw Lagoon, Kalametiya, Vadamarachchi, and Muthurajawela Marsh are the most critical mangrove-bearing areas in Sri Lanka. The above locations vary in mangrove density due to coastal geomorphology, hydrodynamic conditions, and anthropogenic activities.

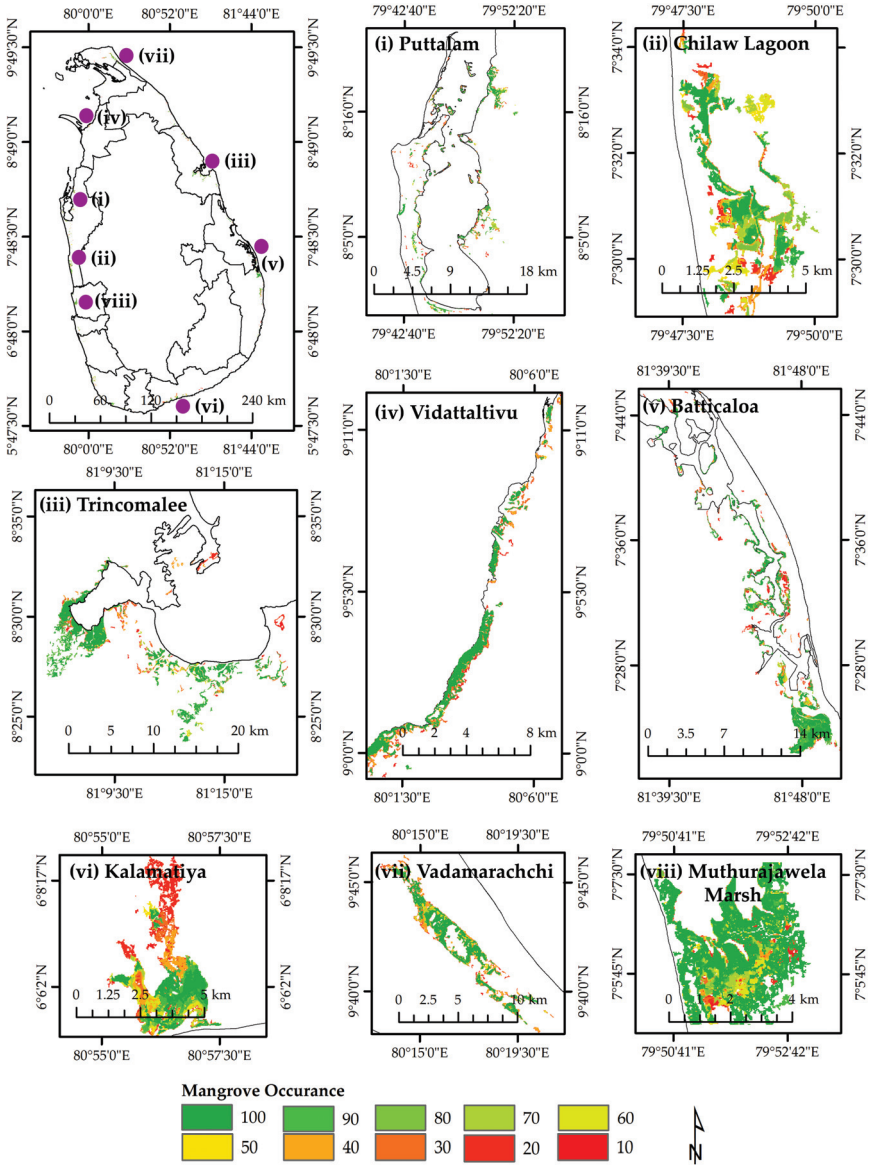


Figure 7. Mangrove occurrence in eight selected areas in Sri Lanka from 1987 to 2022.

The results showed that Trincomalee Bay has the largest mangrove occurrence area (1724 ha at 100% occurrence) (Table S11). Batticaloa Lagoon has a considerable mangrove occurrence (3037 ha at 100% occurrence), Puttalam Lagoon (1365 ha at 100% occurrence), and Vedithalathive (989 ha, 100% occurrence) (Table S11). Furthermore, Vedithalative Nature Reserve is a mangrove protected area in Sri Lanka, declared under the Fauna and Flora Protection Ordinance (Gazette No. 1956/13) due to the significance of marine and coastal ecosystems under the Department of Wildlife Conservation. Chilaw Lagoon, Muthurajawela Marsh, and Kalamatiya have a lower mangrove occurrence (Table S11). Vadamarachchi, located on the Jaffna Peninsula, has relatively higher mangrove coverage. Therefore, policy plans and governance bodies should consider integrated coastal management strategies and site-specific assessments to protect these locations and control future mangrove degradation.

3.5. Spatio-Temporal Variations in Mangrove Structure and Stability Across the Districts

Mangroves in Sri Lanka are distributed across fourteen districts (Figure 1 and Table S1). The structural characteristics and stability patterns of mangroves in each district, with a focus on significant changes, are described below.

3.5.1. Ampara District

Primary fragmentation (decrease in LPI and MESH and an increase in NP) (Figure 8a) shows deforestation from 1997 to 2004. However, from 2016 to 2022, rapid fragmentation was evident (NP increased, while LPI and MESH decreased).

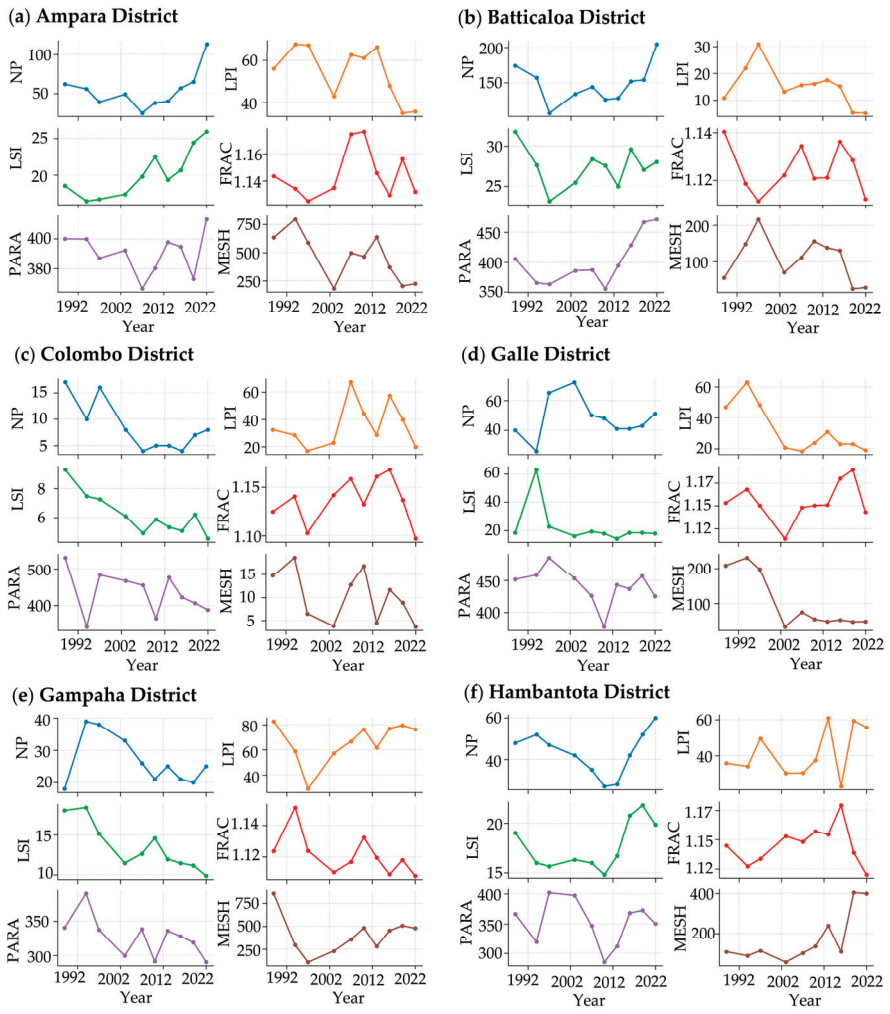


Figure 8. *Cont.*

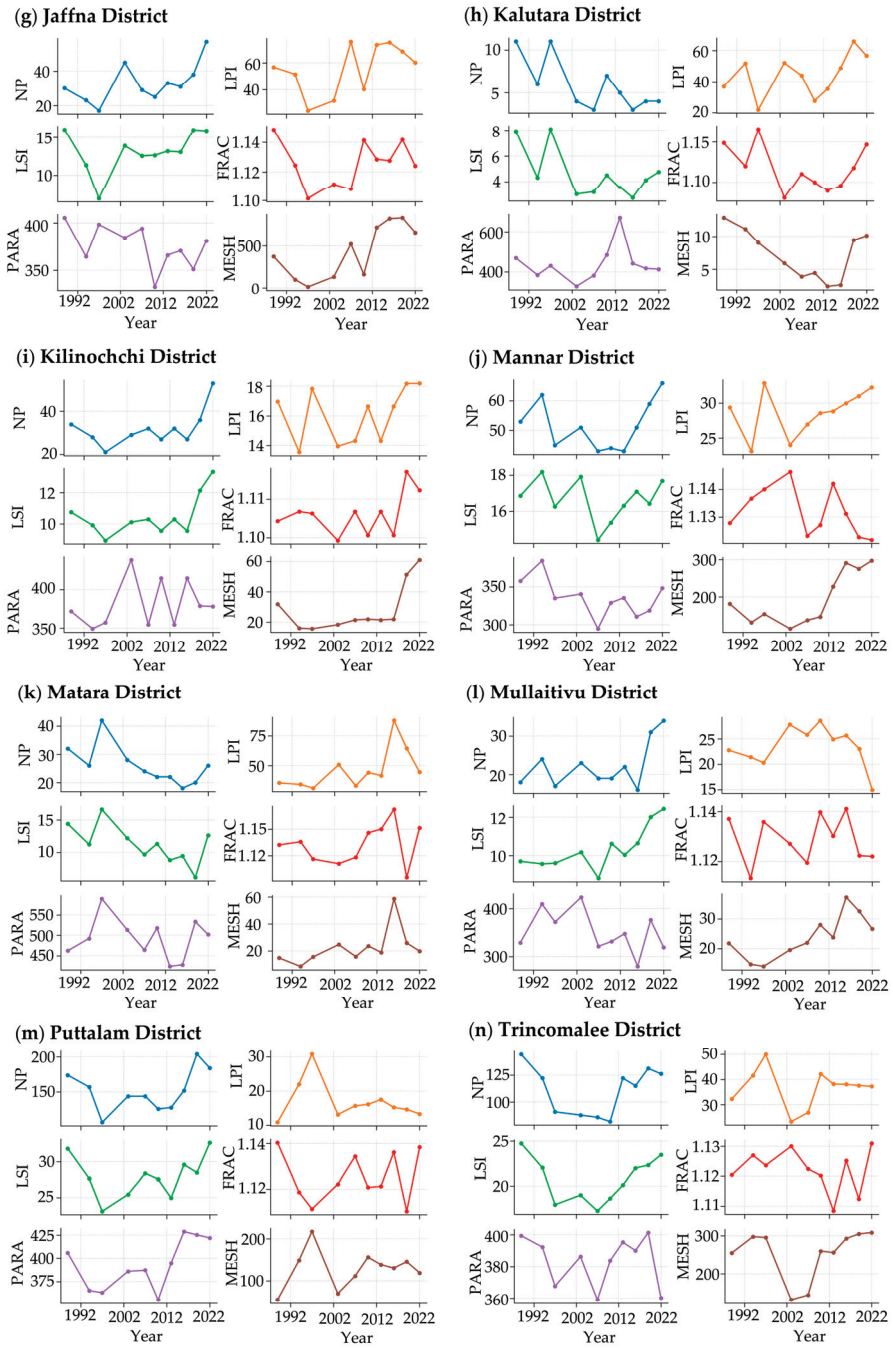


Figure 8. District-wise mangrove structure and stability patterns in Sri Lanka.

3.5.2. Batticaloa District

Severe fragmentation shows from 2016 to 2022 (decreasing in LPI and MESH, increasing in NP and PARA) (Figure 8b). The results further indicate that Batticaloa's mangroves are more severely degraded, with significant fragmentation, loss of core patches, and increased isolation.

3.5.3. Colombo District

The Colombo district exhibits a decline in LPI and MESH, accompanied by increases in NP, indicating signals of fragmentation and shrinking core mangrove areas from 2007 to 2022 (Figure 8c). Recent years have shown a decline in the spatial integrity of Colombo's mangrove, particularly in terms of connectivity (MESH) and dominance (LPI) (Figure 8c).

3.5.4. Galle District

Moderate stabilization was observed from 2007 to 2022 (Figure 8d).

3.5.5. Gampaha District

Mangroves in the Gampaha districts exhibit high spatial coherence and ecological resilience, as indicated by fewer patches, a strong LPI, compact shapes, and high MESH values from 2010 to 2022 (Figure 8e). Despite past fragmentation (the 1990s), Gampaha exhibits strong recovery, with high dominance, low edge stress, and increased connectivity.

3.5.6. Hambantota District

The increase in connectivity (MESH) and fragmentation (increasing NP) suggests clustered small patches coexisting with a strong core patch from 2013 to 2022 (Figure 8f). Overall, the mangroves in the Hambantota district indicate moderately fragmented but resilient ecosystems, exhibiting both signs of degradation and spatial adaptation by 2022.

3.5.7. Jaffna District

From 2004 to 2016, LPI, MESH, and LSI increased, indicating a clear restoration stage (Figure 8g). However, increased NP, combined with high MESH retention, suggests clustered patch growth rather than separated fragmentation from 2019 to 2022 (Figure 8g). However, the increase in NP is an early indicator of potential fragmentation (Figure 8g), necessitating continuous monitoring and protective strategies.

3.5.8. Kalutara District

Kalutara district mangrove trends show that LSI, FRAC, and MESH decreased (Figure 8h), indicating a simplification of shape, a decrease in connectivity, and an increased LPI from 2003 to 2016 (Figure 8h). Although LPI shows a high value, the mangroves are ecologically fragile, with low MESH and high PARA (Figure 8h), indicating urgent conservation and restoration strategies.

3.5.9. Kilinochchi District

The mangrove trends in the Kilinochchi district show reduced connectivity and remain as small, disconnected, and structurally weak patches due to degradation from 1989 to 2016 (Figure 8i). Spatial cohesion also improved (MESH increase) (Figure 8i), showing a modest but favorable ecological recovery, especially after 2019 (Figure 8i).

3.5.10. Manner District

The mangrove district in Manner shows increased fragmentation but stable core areas (LPI) and strong connectivity (MESH) from 2010 to 2022 (Figure 8j). Overall, the Mannar district has a resilient mangrove ecosystem, despite fragmentation, with functional connectivity and structural stability, indicating a stable and well-connected mangrove system.

3.5.11. Matara District

Mangroves in the Matara district exhibit an increased LPI, LSI, and MESH, indicating a recovery period from 2010 to 2016 (Figure 8k). However, from 2019 to 2022, the number of new mangrove patches increased while LPI and MESH decreased, suggesting a new fragmentation trend in the Matara district.

3.5.12. Mullaitivu District

The Mullaitivu district mangrove shows the best ecological structure from 2010 to 2016 (high LPI and MESH, low PARA) (Figure 8l). However, mangrove fragmentation increased from 2019 to 2022, resulting in a decrease in LPI.

3.5.13. Puttalam District

Puttalam district mangroves show stable but gradually decreasing connectivity and patch dominance from 1997 to 2016 (Figure 8m). Notably, extensive fragmentation (NP) and declining LPI, MESH, and LSI show clear signs of ecosystem degradation from 2019 to 2022 (Figure 8m), indicating that the Puttalam district is one of the most fragmented mangrove systems in Sri Lanka.

3.5.14. Trincomalee District

Mangroves in the Trincomalee district show LPI > 35% (stable core areas), steadily increasing MESH (connectivity), and relatively higher fragmentation (Figure 8n), showing Trincomalee's mangroves are a healthy but vulnerable ecosystem from 2010 to 2022 (Figure 8n), indicating strong ecological performance, with a dominant core patch.

3.6. Comparisons with Global Mangrove Watch (GMW) Data

Our study demonstrates higher accuracy (average overall accuracy = 96.29%) in mangrove mapping compared to the GMW dataset (87.4%) [39], primarily due to the use of national-level training and validation data. Conversely, the GMW dataset relies on global-scale training data. Using national-scale data enhances mapping accuracy by incorporating region-specific spectral and ecological characteristics. Our mangrove detection methodology primarily focuses on coastal mangrove mapping, which may lead to a more precise delineation of mangroves (see Section 2.2). In contrast, the GMW dataset employs a more generalized global mangrove detection typology, which may lead to overestimation or underestimation in certain areas due to the diverse ecological environments of mangroves worldwide. The GMW dataset consistently shows higher mangrove area values compared to our study due to differences in classification methodologies and the resolution of input data (Figure 9).

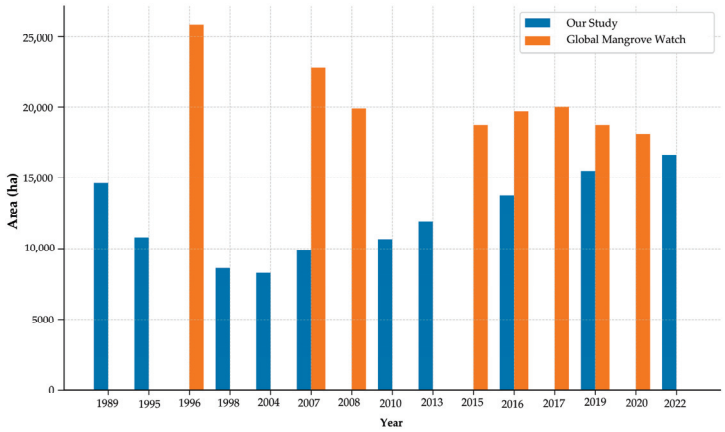


Figure 9. Comparisons of Global Mangrove Watch (GMW) data.

4. Discussion

4.1. Ecological Significance of Mangroves in Sri Lanka

Mangroves in Sri Lanka are one of the critical ecosystems, providing valuable ecosystem services such as carbon storage, biodiversity, habitat quality, nutrient cycling, and the stability of coastal areas. Our study revealed that mangroves are distributed along lagoons, estuaries, bays, and riverbanks near the coastal zone, covering approximately 16,615 ha by 2022, representing 0.25% of Sri Lanka's land area. The most extensive and continuous mangrove patches are observed in Puttalam, Kalpitiya, the Jaffna Peninsula, Trincomalee, Batticaloa, Muthurajawela, and Vedithalative coastal regions (Figures 7 and 8). However, ecologically and biologically diverse small mangrove patches are found in Negombo

Lagoon, Madu Ganga, Rekawa Lagoon, Kalamatiya, Koggala, and Anivilundawa. The Colombo, Kalutara, Galle, Matara, Ampara, and Puttalam districts exhibit a fragmented mangrove distribution, largely due to high levels of anthropogenic activities, including urban and coastal development. These areas undoubtedly provide valuable ecosystem services and carbon sequestration benefits, thereby contributing to Sri Lanka's environmental sustainability and global efforts to mitigate climate change.

The geographical context of Sri Lanka influences the distribution pattern of mangroves. For example, the eastern coastal belt of Sri Lanka is primarily affected by cyclonic activities from the Bay of Bengal, seasonal flooding and sediment deposition, and limited human activities, which provide suitable ecological conditions for propagating pristine mangrove patches. Conversely, the western coast faces the Bay of Arabia and has a more stable hydrological condition; however, significant urban expansion and coastal development pressures impact the sustainability of these areas [57]. The Intertropical Convergence Zone (ITCZ) modulates rainfall and river discharge patterns in Sri Lanka, affecting nutrient dynamics in estuaries and the lagoon system [87,88]. In Sri Lanka, 103 rivers from the central and southern highlands transport numerous sediments and organic matter to estuaries and lagoon systems (particularly in the dry zone), which is crucial for the mangrove ecosystem in Sri Lanka. Sri Lanka is home to numerous true mangroves, including critically endangered species such as *Lumnitzera littorea*, *Ceriops decandra*, and *Xylocarpus rumphii*, emphasizing the conservation importance of these mangroves [51]. Our study provides the initial assessment of mangrove occurrence in Sri Lanka, which will undoubtedly provide primary information to identify distribution patterns and changes, as well as monitor endangered and other mangrove species.

Our findings contribute to the international context by providing the first national-scale mangrove extent in Sri Lanka, using remote sensing techniques, which enables cross-country comparisons with other countries in the tropical regions. The results of fragmented versus intact mangrove patches provide new empirical findings to global discussions on the balance between anthropogenic activities and natural climatic drivers related to mangrove resilience. Our study further provides valuable insights for regional integration into blue carbon estimation and international conservation strategies.

4.2. Multi-Dimensional Challenges to Mangrove Sustainability in Sri Lanka

Mangroves in Sri Lanka face numerous threats that jeopardize their long-term sustainability due to anthropogenic activities, land-use change, coastal development, deforestation, and illegal exploitation, as well as climate change, pollution, water quality degradation, biodiversity loss, weak policy implementation, and governance issues [51]. Among them, anthropogenic activities have historically been and continue to be a primary influencing factor of mangrove loss and degradation in Sri Lanka [57]. Deforestation and changes to mangrove habitats for land use have significantly impacted ecological interactions and ecosystem functions [89]. Studies have reported that many mangroves have been cleared for aquaculture development [90]. In Sri Lanka, aquaculture is particularly prominent in the Northwestern, Southern, and Eastern Provinces—especially around areas such as Batticaloa in the Eastern Province and Chilaw Lagoon in the Northwestern Province—significantly affecting mangrove cover [51].

Mangrove transformation for agriculture (paddy cultivation and coconut plantations), mangrove land reclamation for many large projects, and illegal encroachments [51], as well as urbanization and infrastructure development (the construction of hotels, settlements, roads, and ports (air and harbor)) [59], have also contributed significantly to mangrove loss in many districts in Sri Lanka (Table S6). The results of the national-scale assessment indicate a continuous increase in mangrove cover, particularly from 2007 to 2022 (Table 1).

Many districts where human activities are predominant, such as Colombo, Gampaha, Kalutara, Matara, and Galle, exhibit a rapid loss and degradation of mangroves (Table S6). Undisturbed and low-activity districts, such as Hambantota, Jaffna, Trincomalee, Mannar, Kilinochchi, and Mullaitivu, exhibit an increasing trend (Table S6). Therefore, the national-scale mangrove assessment shows an increasing trend (Table 1). In contrast, district-wise assessments show significant spatio-temporal variations in gaining and losing (Table S6), as well as greater variations in structure and stability patterns (Figure 8). Athukorala et al. (2021) revealed that Muthurajawela Marsh and Negombo Lagoon, located in the Colombo Metropolitan Region, showed dramatic degradation of mangroves due to urban development [59]. Socioeconomic factors and unsustainable resource exploitation pose significant challenges to the conservation of mangroves in Sri Lanka. For example, local communities living near lagoons and estuaries (where extensive mangrove areas exist) depend on resources such as fuelwood, timber, and fisheries, indicating sustainable management of overexploitation and degradation of mangrove habitats [51]. Sea level rise (a direct impact of global warming) creates a considerable risk to mangrove habitats via permanent inundation and fluctuation in salinity levels [91].

The spatio-temporal changes in mangroves in Sri Lanka exhibit general global trends, where localized mangrove loss persists alongside regional gains resulting from restoration and natural regeneration processes. Similar patterns have been observed in Southeast Asia, West Africa, and Latin America [39]. Our results further emphasize the urgent need for integrated, multi-scale monitoring and management strategies. Mangrove assessment in Sri Lanka provides robust evidence to inform global initiatives, such as the Paris Agreement and the UN Decade on Ecosystem Restoration, as well as blue carbon strategies that focus on blue carbon estimation, identifying priority restoration sites, and mangrove conservation for a cost-effective climate mitigation strategy.

4.3. Integrating Spatial Structure and Conservation Strategies

Numerous studies such as the Mekong Delta in Vietnam [92], the tallest mangrove forest on Earth (Pongara National Park, Gabon) [46], the largest continuous mangrove forest in the world (Sundarbans, Bangladesh) [93], mangroves in the southwestern coast of China [94], Pongok Island [95], the Pearl River Estuary, China [96], the coastal regions of Hong Kong, China [97], the Arabian Peninsula [90], Bonaire (a small island in the Caribbean Sea) [49], and mangroves in Indonesia [98], have demonstrated that integrating mangrove change detection and the spatial distribution and stability patterns significantly enhances mangrove conservation efforts globally and supports identifying site-specific restoration priorities. The following recommendations suggest strategies for sustaining Sri Lanka's mangroves. We discuss district-wide strategies for enhancing mangrove ecosystems and their combined ecosystem services (see Figure 8 and Table S6).

- Ampara and Batticaloa districts require urgent conservation and policy actions, including reforestation, zoning restrictions, and community-based conservation.
- Colombo, Gampaha, and Kalutara districts should integrate ecological corridors and buffer zones into urban planning, incorporate mangrove protection into climate-resilient city planning, and prioritize the restoration of fragmented patches.
- Galle, Matara, Hambantota, and Mullaitivu districts require restoring connectivity between patches, with a focus on preserving the remaining core areas and monitoring coastal development. Buffer zone enforcement and the protection of emerging small patches are also required.
- Jaffna exhibits remarkable mangrove recovery; however, the rise in NP is an early signal of potential fragmentation, necessitating continuous monitoring and mangrove landscape conservation.

– Puttalam, Trincomalee, Kilinochchi, and Mannar districts require urgent habitat consolidation, the development of restoration corridors, protection from aquaculture expansion, and the implementation of ecosystem-based management and blue carbon conservation strategies, the establishment of habitat corridors, and the protection of emerging clusters.

Furthermore, ensuring the long-term sustainability of mangroves in Sri Lanka requires comprehensive strategies that combine policy enforcement, scientifically based ecological restoration, community engagement, and pollution control. In this regard, the National Guidelines for the Restoration of Mangrove Ecosystems in Central Environmental Authority [99], the Fauna and Flora Protection Ordinance of 1937 under the Wildlife Department [100], the World Conservation Union (IUCN), Sri Lanka [101], the National Aquatic Resources Agency (NARA) [102], the National Science Foundation (NSF) [103], and the Coast Conservation Department (CCD) [64] all play a unique role in mangrove protection and conservation in Sri Lanka. They should closely relate to all departments that are important for mangroves in Sri Lanka. Moreover, the Urban Development Authority [104] and the Road Development Authority [105] should closely monitor recent rapid urbanization and infrastructure development in coastal cities where mangrove forests are located.

In recent years, Sri Lanka has demonstrated a strong national obligation to mangrove conservation and ecosystem restoration. In 2015, it became the first country to legally protect all mangroves, establishing the National Expert Committee on Mangrove Conservation and Sustainable Use [106]. This was followed by the adoption of the National Policy on Conservation and Sustainable Utilization of Mangrove Ecosystems in 2020 and the development of a National Strategic Action Plan and accompanying guidelines in 2022 aimed at ensuring long-term sustainability. Recognizing these sustained efforts, Sri Lanka's mangrove restoration initiative was designated as a World Restoration Flagship and recognized as one of the most outstanding examples of large-scale and long-term mangrove ecosystem restoration globally [107]. Integrating the national climate change mitigation plan into mangrove conservation is crucial for mitigating threats to the mangrove ecosystem.

4.4. Uncertainties and Future Directions

Landsat data is useful for long-term and large-scale mapping (30 m resolution). However, Landsat data may not be sufficient to identify small or fragmented mangroves or to distinguish between different mangrove species. Future studies should utilize higher-resolution imagery (e.g., Sentinel-2 (10 m), WorldView, PlanetScope) for future mangrove mapping. Landsat data support long-term temporal analysis of mangrove distribution, while higher-resolution satellite data offer more detailed spatial data in the study areas. Furthermore, using high-resolution images will improve the accuracy of mangrove detection and may enable the separation of mangrove species. Future studies should utilize extensive field data, including species composition, biomass, soil characteristics, and other relevant parameters, to inform mangrove change detection.

5. Conclusions

This study presents the first national-scale assessment of mangrove dynamics using remote sensing in Sri Lanka. We developed a mangrove detection approach using Landsat images in the GEE to analyze mangrove changes in Sri Lanka from 1987 to 2022. Mangrove gain, loss, occurrence, structure, and stability patterns were analyzed at the national scale, provincial, district-wise, and climatic zone-wise. The results showed that mangroves in Sri Lanka have experienced significant spatio-temporal changes. The results indicated that mangrove extent increased from 1989 to 2022 (13.6%), suggesting a continuous recovery of mangroves at the national scale. Mangrove structure and stability pattern analysis

revealed that the districts that perform best ecologically are Jaffna, Trincomalee, and Gampaha. Moderate to transitional districts include Ampara, Mannar, Matara, Galle, Mullaitivu, Hambantota, and Kilinochchi. The most degraded districts are Batticaloa, Puttalam, Colombo, and Kalutara.

The dry zone of the country bears the largest extent of mangrove (11,383 ha). In contrast, the wet (2285 ha) and intermediate zones (760 ha) have lower coverage. Therefore, integrating mangrove management into national climate strategies, sustainable land use, science-based restoration, and community engagement will undoubtedly support the balance between the ecological and socioeconomic benefits of mangroves in Sri Lanka.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/land14091820/s1, Table S1. Mangrove distribution in fourteen coastal districts of Sri Lanka; Table S2. Number of Landsat images: Sensor-wise; Table S3. Sample data collection of this study; Figure S1. (a) Mangrove-bearing Provinces in Sri Lanka, and (b) Climatic zones in Sri Lanka; Table S4. Landscape parameters for mangroves in Sri Lanka; Table S5. Mangrove classification accuracy; Table S6. Total area of mangroves in Sri Lanka: District-wise (ha); Table S7. District-wise average mangrove areas in Sri Lanka (1987–2022): High to Low; Table S8. District-wise net mangrove gains areas in Sri Lanka (1987–2022): High to Low; Table S9. District-wise net mangrove loss areas in Sri Lanka (1987–2022): High to Low; Table S10. Mangrove occurrence in climatic zone-wise (ha); Table S11. Mangrove occurrence in eight selected areas

Author Contributions: Conceptualization, D.A.; methodology, D.A.; software, D.A.; validation, D.A. and Y.M.; formal analysis, D.A., S.K. and R.W.; investigation, D.A., Y.M., T.M., S.K., R.W., S.L.J.F. and N.S.K.H.; resources, D.A.; data curation, D.A.; writing—original draft preparation, D.A.; writing—review and editing, D.A., Y.M., T.M., S.K., R.W., S.L.J.F. and N.S.K.H.; visualization, D.A.; supervision, Y.M.; funding acquisition, D.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Japan Society for the Promotion of Science (JSPS) through a Postdoctoral Fellowship (Grant No. 24KF0178) awarded to Darshana Athukorala and JSPS Grant No. 24K04416.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NDMI	Normalized Difference Mangrove Index	NDVI	Normalized Difference Vegetation Index
GCVI	Green Chlorophyll Vegetation Index	MNDWI	Modified Normalized Difference Water Index
SDGs	Sustainable Development Goals	GEE	Google Earth Engine
IUCN	International Union for Conservation of Nature	CEA	Central Environmental Authority
TM	Thematic Mapper	ETM+	Enhanced Thematic Mapper+
OLI	Operational Land Imager	NIR	Near-Infrared
SWIR	Shortwave Infrared	SRTM	Shuttle Radar Topography Mission
DEM	Digital Elevation Model	SR	Simple Ratio
RF	Random Forest	LPI	Largest Patch Index
PARA	Perimeter-Area Ratio	FRAC	Fractal Dimension Index
LSI	Landscape Shape Index	MESH	Effective Mesh Size
NP	Number of Patches	ITCZ	Intertropical Convergence Zone
NARA	National Aquatic Resources Agency	CCD	Coast Conservation Department

References

1. Kuenzer, C.; Bluemel, A.; Gebhardt, S.; Quoc, T.V.; Dech, S. Remote Sensing of Mangrove Ecosystems: A Review. *Remote Sens.* **2011**, *3*, 878–928. [CrossRef]
2. Murray, N.J.; Phinn, S.R.; DeWitt, M.; Ferrari, R.; Johnston, R.; Lyons, M.B.; Clinton, N.; Thau, D.; Fuller, R.A. The Global Distribution and Trajectory of Tidal Flats. *Nature* **2019**, *565*, 222–225. [CrossRef]
3. Macreadie, P.I.; Anton, A.; Raven, J.A.; Beaumont, N.; Connolly, R.M.; Friess, D.A.; Kelleway, J.J.; Kennedy, H.; Kuwae, T.; Lavery, P.S.; et al. The Future of Blue Carbon Science. *Nat. Commun.* **2019**, *10*, 3998. [CrossRef]
4. Alongi, D.M. Present State and Future of the World's Mangrove Forests. *Environ. Conserv.* **2002**, *29*, 331–349. [CrossRef]
5. Barbier, E.B.; Hacker, S.D.; Kennedy, C.; Koch, E.W.; Stier, A.C.; Silliman, B.R. The Value of Estuarine and Coastal Ecosystem Services. *Ecol. Monogr.* **2011**, *81*, 169–193. [CrossRef]
6. Murillo-Sandoval, P.J.; Perea-Ardila, M.; Simard, M. Examining the Status of Mangrove Forest Dynamics and Community-Led Management on the Colombian Pacific Coast. *Ocean Coast. Manag.* **2025**, *266*, 107672. [CrossRef]
7. United Nations. The 17 GOALS Sustainable Development. Available online: https://sdgs.un.org/goals (accessed on 15 June 2025).
8. Choudhary, B.; Dhar, V.; Pawase, A.S. Blue Carbon and the Role of Mangroves in Carbon Sequestration: Its Mechanisms, Estimation, Human Impacts and Conservation Strategies for Economic Incentives. *J. Sea Res.* **2024**, *199*, 102504. [CrossRef]
9. Huber, L.C.; Sainge, M.N.; Feka, Z.N.; Kamara, R.A.; Kamara, A.; Sullivan, M.; Cuni-Sanchez, A. Human-Driven Degradation Impacts on Mangroves in Southern Sierra Leone. *Trees For. People* **2023**, *14*, 100445. [CrossRef]
10. Quang, N.H.; Van An, N. Four Decades of Mangrove Shoreline Changes Driven by Human Activities in the Red River Delta. *Estuar. Coast. Shelf Sci.* **2025**, *314*, 109145. [CrossRef]
11. Alongi, D.M. Carbon Cycling and Storage in Mangrove Forests. *Ann. Rev. Mar. Sci.* **2014**, *6*, 195–219. [CrossRef]
12. Chatting, M.; Al-Maslamani, I.; Walton, M.; Skov, M.W.; Kennedy, H.; Husrevoglu, Y.S.; Le Vay, L. Future Mangrove Carbon Storage Under Climate Change and Deforestation. *Front. Mar. Sci.* **2022**, *9*, 781876. [CrossRef]
13. Hilmi, N.; Arruda, G.; Broussard, D.; Maria Benitez, B.; Sauron, L.; Lamaud, T.; Jahan, N.; Hall Spencer, J.M. Blue Carbon as a Nature-Based Climate Mitigation Strategy for Mangrove Conservation in Bangladesh. *J. Nat. Conserv.* **2025**, *86*, 126885. [CrossRef]
14. Swinea, S.H.; Randall Hughes, A.; Osland, M.J.; Shepard, C.C.; Thorne, K.B.; Alemu, J.B.; Bardou, R.; Scyphers, S.B. Marshes to Mangroves: Residential Surveys Reveal Perceived Wetland Trade-Offs for Ecosystem Services. *Landsc. Urban Plan.* **2025**, *253*, 105203. [CrossRef]
15. Saria, A.E.; Li, X.; Dubi, A.; Mgalula, M.E. Spatial-Temporal Changes in Mangrove Forest in Coastal Cities of Tanzania and the Implications on Provisioning of Ecosystem Services. *J. Mar. Res.* **2020**, *1*, 107604. [CrossRef]
16. Hickey, S.M.; Phinn, S.R.; Callow, N.J.; Van Niel, K.P.; Hansen, J.E.; Duarte, C.M. Is Climate Change Shifting the Poleward Limit of Mangroves? *Estuaries Coasts* **2017**, *40*, 1215–1226. [CrossRef]
17. Malik, A.; Fensholt, R.; Mertz, O. Mangrove Exploitation Effects on Biodiversity and Ecosystem Services. *Biodivers. Conserv.* **2015**, *24*, 3543–3557. [CrossRef]
18. He, T.; Fu, Y.; Ding, H.; Zheng, W.; Huang, X.; Li, R.; Wu, S. Evaluation of Mangrove Wetlands Protection Patterns in the Guangdong–Hong Kong–Macao Greater Bay Area Using Time-Series Landsat Imageries. *Remote Sens.* **2022**, *14*, 6026. [CrossRef]
19. Xie, Z.; Li, H.; Yuan, Y.; Hu, W.; Luo, G.; Huang, L.; Chen, M.; Wu, W.; Yan, G.; Sun, X. The Spatial Patterns and Driving Mechanisms of Blue Carbon 'Loss' and 'Gain' in a Typical Mangrove Ecosystem: A Case Study of Beihai, Guangxi Province of China. *Sci. Total Environ.* **2023**, *905*, 167241. [CrossRef]
20. Turschwell, M.P.; Tulloch, V.J.D.; Sievers, M.; Pearson, R.M.; Andradi-Brown, D.A.; Ahmadia, G.N.; Connolly, R.M.; Bryan-Brown, D.; Lopez-Marcano, S.; Adame, M.F.; et al. Multi-Scale Estimation of the Effects of Pressures and Drivers on Mangrove Forest Loss Globally. *Biol. Conserv.* **2020**, *247*, 108637. [CrossRef]
21. Sudhir, S.; Arunprasath, A.; Sankara Vel, V. A Critical Review on Adaptations, and Biological Activities of the Mangroves. *J. Nat. Pestic. Res.* **2022**, *1*, 100006. [CrossRef]
22. Sierra-Correa, P.C.; Cantera Kintz, J.R. Ecosystem-Based Adaptation for Improving Coastal Planning for Sea-Level Rise: A Systematic Review for Mangrove Coasts. *Mar. Policy* **2015**, *51*, 385–393. [CrossRef]
23. Qiao, P.; Meng, Y.; Liu, C.; Zhang, T.; Li, M.; Feng, L.; Chen, Q.; Zhang, J.; Lin, C.; Gu, X.; et al. Blue Carbon Potential from Rehabilitating Urban Mangrove Forests in Coastal City. *Estuar. Coast. Shelf Sci.* **2025**, *316*, 109200. [CrossRef]
24. Pimple, U.; Leadprathom, K.; Simonetti, D.; Sitthi, A.; Peters, R.; Pungkul, S.; Pravinvongvuthi, T.; Dessard, H.; Berger, U.; Siri-on, K.; et al. Assessing Mangrove Species Diversity, Zonation and Functional Indicators in Response to Natural, Regenerated, and Rehabilitated Succession. *J. Environ. Manag.* **2022**, *318*, 115507. [CrossRef]
25. Ponmani, M.; Padmavathy, P.; Manimekalai, D.; Shalini, R.; Ravikumar, T.; Hariharan, G.; Manickavasagam, S. Vulnerability of Mangrove Ecosystems to Anthropogenic Marine Litter along the Southeast Coast of India. *Sci. Total Environ.* **2024**, *956*, 177224. [CrossRef] [PubMed]

26. Josephraj Selvaraj, J.; Victoriano Portilla-Cabrera, C. Climate Change Drives Spatiotemporal Mismatches between Bivalves and Mangrove in the Colombian Pacific Coast. *J. Nat. Conserv.* **2024**, *81*, 126679. [CrossRef]
27. Murray, N.J.; Worthington, T.A.; Bunting, P.; Duce, S.; Hagger, V.; Lovelock, C.E.; Lucas, R.; Saunders, M.I.; Sheaves, M.; Spalding, M.; et al. High-Resolution Mapping of Losses and Gains of Earth's Tidal Wetlands. *Science* **2022**, *376*, 744–749. [CrossRef] [PubMed]
28. Murray, N.J.; Phinn, S.P.; Fuller, R.A.; DeWitt, M.; Ferrari, R.; Johnston, R.; Clinton, N.; Lyons, M.B. High-Resolution Global Maps of Tidal Flat Ecosystems from 1984 to 2019. *Sci. Data* **2022**, *9*, 542. [CrossRef] [PubMed]
29. Carugati, L.; Gatto, B.; Rastelli, E.; Lo Martire, M.; Coral, C.; Greco, S.; Danovaro, R. Impact of Mangrove Forests Degradation on Biodiversity and Ecosystem Functioning. *Sci. Rep.* **2018**, *8*, 13298. [CrossRef]
30. Ruslan, N.F.N.; Goh, H.C.; Hattam, C.; Edwards-Jones, A.; Moh, H.H. Mangrove Ecosystem Services: Contribution to the Well-Being of the Coastal Communities in Klang Islands. *Mar. Policy* **2022**, *144*, 105222. [CrossRef]
31. Chmura, G.L.; Anisfeld, S.C.; Cahoon, D.R.; Lynch, J.C. Global Carbon Sequestration in Tidal, Saline Wetland Soils. *Glob. Biogeochem. Cycles* **2003**, *17*, 1111. [CrossRef]
32. Goldberg, L.; Lagomasino, D.; Thomas, N.; Fatoyinbo, T. Global Declines in Human-Driven Mangrove Loss. *Glob. Change Biol.* **2020**, *26*, 5844–5855. [CrossRef]
33. Maung, W.S.; Sasaki, J. Assessing the Natural Recovery of Mangroves after Human Disturbance Using Neural Network Classification and Sentinel-2 Imagery in Wunbaik Mangrove Forest, Myanmar. *Remote Sens.* **2021**, *13*, 52. [CrossRef]
34. Kanjin, K.; Alam, B.M. Assessing Changes in Land Cover, NDVI, and LST in the Sundarbans Mangrove Forest in Bangladesh and India: A GIS and Remote Sensing Approach. *Remote Sens. Appl. Soc. Environ.* **2024**, *36*, 101289. [CrossRef]
35. Navarro, A.; Young, M.; Macreadie, P.I.; Nicholson, E.; Ierodiaconou, D. Mangrove and Saltmarsh Distribution Mapping and Land Cover Change Assessment for South-Eastern Australia from 1991 to 2015. *Remote Sens.* **2021**, *13*, 1450. [CrossRef]
36. Dong, D.; Huang, H.; Gao, Q.; Li, K.; Zhang, S.; Yan, R. Temporal and Spatial Patterns of Blue Carbon Storage in Mangrove and Salt Marsh Ecosystems in Guangdong, China. *Land* **2025**, *14*, 1130. [CrossRef]
37. Kanniah, K.D.; Kang, C.S.; Sharma, S.; Aldrie Amir, A. Remote Sensing to Study Mangrove Fragmentation and Its Impacts on Leaf Area Index and Gross Primary Productivity in the South of Peninsular Malaysia. *Remote Sens.* **2021**, *13*, 1427. [CrossRef]
38. Nascimento, W.R.; Souza-Filho, P.W.M.; Proisy, C.; Lucas, R.M.; Rosenqvist, A. Mapping Changes in the Largest Continuous Amazonian Mangrove Belt Using Object-Based Classification of Multisensor Satellite Imagery. *Estuar. Coast. Shelf Sci.* **2013**, *117*, 83–93. [CrossRef]
39. Bunting, P.; Rosenqvist, A.; Hilarides, L.; Lucas, R.M.; Thomas, N.; Tadono, T.; Worthington, T.A.; Spalding, M.; Murray, N.J. Global Mangrove Extent Change 1996–2020: Global Mangrove. *Remote Sens.* **2022**, *14*, 3657. [CrossRef]
40. Global Mangrove Watch. Available online: https://www.globalmangrovewatch.org/ (accessed on 15 June 2025).
41. Google Earth Engine. Available online: https://earthengine.google.com/ (accessed on 15 June 2025).
42. Estoque, R.C.; Myint, S.W.; Wang, C.; Ishtiaque, A.; Aung, T.T.; Emerton, L.; Ooba, M.; Hijioka, Y.; Mon, M.S.; Wang, Z.; et al. Assessing environmental impacts and change in Myanmar's mangrove ecosystem service value due to deforestation (2000–2014). *Glob. Change Biol.* **2018**, *24*, 5391–5410. [CrossRef]
43. Sun, Y.; Ye, M.; Ai, B.; Lai, Z.; Zhao, J.; Jian, Z.; Qi, X. Annual Change in the Distribution and Landscape Health of Mangrove Ecosystems in China from 2016 to 2023 with Sentinel Imagery. *Glob. Ecol. Conserv.* **2025**, *57*, e03355. [CrossRef]
44. Giri, C.; Long, J.; Poudel, P. Mangrove Forest Cover Change in the Conterminous United States from 1980–2020. *Remote Sens.* **2023**, *15*, 5018. [CrossRef] [PubMed]
45. Hidayah, Z.; As-syakur, A.R.; Rachman, H.A. Sustainability Assessment of Mangrove Management in Madura Strait, Indonesia: A Combined Use of the Rapid Appraisal for Mangroves (RAPMangroves) and the Remote Sensing Approach. *Mar. Policy* **2024**, *163*, 106128. [CrossRef]
46. Stovall, A.E.L.; Fatoyinbo, T.; Thomas, N.M.; Armston, J.; Ebanega, M.O.; Simard, M.; Trettin, C.; Obiang Zogo, R.V.; Aken, I.A.; Debina, M.; et al. Comprehensive Comparison of Airborne and Spaceborne SAR and LiDAR Estimates of Forest Structure in the Tallest Mangrove Forest on Earth. *Sci. Remote Sens.* **2021**, *4*, 100034. [CrossRef]
47. Cortés, I.M.; Lorenzo-Trueba, J.; Rovai, A.S.; Twilley, R.R.; Chopping, M.; Fatoyinbo, T. Net Evaporation-Induced Mangrove Area Loss across Low-Lying Caribbean Islands. *Environ. Res. Clim.* **2024**, *3*, 045004. [CrossRef]
48. Tinh, P.H.; MacKenzie, R.A.; Hung, T.D.; Van Vinh, T.; Ha, H.T.; Lam, M.H.; Hanh, N.T.H.; Tung, N.X.; Hai, P.M.; Huyen, B.T. Mangrove Restoration in Vietnamese Mekong Delta during 2015-2020: Achievements and Challenges. *Front. Mar. Sci.* **2022**, *9*, 1043943. [CrossRef]
49. Casal, G.; Trégarot, E.; Cornet, C.C.; McCarthy, T.; van der Geest, M. A Cost-Effective Method to Map Mangrove Forest Extent, Composition, and Condition in Small Islands Based on Sentinel-2 Data: Implications for Management. *Ecol. Indic.* **2024**, *159*, 111696. [CrossRef]

50. Liu, X.; Fatoyinbo, T.E.; Thomas, N.M.; Guan, W.W.; Zhan, Y.; Mondal, P.; Lagomasino, D.; Simard, M.; Trettin, C.C.; Deo, R.; et al. Large-Scale High-Resolution Coastal Mangrove Forests Mapping Across West Africa with Machine Learning Ensemble and Satellite Big Data. *Front. Earth Sci.* **2021**, *8*, 560933. [CrossRef]
51. The Central Environmental Authority (CEA); The World Conservation Union (IUCN); The International Water Management Institute (IWMI). *National Wetland Directory of Sri Lanka*; The Central Environmental Authority (CEA); The World Conservation Union (IUCN); The International Water Management Institute (IWMI): Colombo, Sri Lanka, 2006; ISBN 9558177547.
52. Survey Department of Sri Lanka. Available online: https://www.survey.gov.lk/ (accessed on 15 June 2025).
53. Department of Meteorology. Available online: https://meteo.gov.lk/ (accessed on 15 June 2025).
54. Ramsar. The Convention on Wetlands. Available online: https://www.ramsar.org/ (accessed on 15 June 2025).
55. Athukorala, D.; Murayama, Y.; Herath, N.S.K.; Bandara, C.M.M.; Singh, R.K.; Fernando, S.L.J. Exploring the Cooling Effects of Urban Wetlands in Colombo City, Sri Lanka. *Remote Sens.* **2025**, *17*, 1919. [CrossRef]
56. Earth Versions—Google Earth. Available online: https://www.google.com/earth/about/versions/ (accessed on 15 June 2025).
57. Athukorala, D.; Estoque, R.C.; Murayama, Y.; Matsushita, B. Impacts of Urbanization on the Muthurajawela Marsh and Negombo Lagoon, Sri Lanka: Implications for Landscape Planning towards a Sustainable Urban Wetland Ecosystem. *Remote Sens.* **2021**, *13*, 316. [CrossRef]
58. Perera, K.A.R.S.; Amarasinghe, M.D. Carbon Sequestration Capacity of Mangrove Soils in Micro Tidal Estuaries and Lagoons: A Case Study from Sri Lanka. *Geoderma* **2019**, *347*, 80–89. [CrossRef]
59. Athukorala, D.; Estoque, R.C.; Murayama, Y.; Matsushita, B. Ecosystem Services Monitoring in the Muthurajawela Marsh and Negombo Lagoon, Sri Lanka, for Sustainable Landscape Planning. *Sustainability* **2021**, *13*, 11463. [CrossRef]
60. Satyanarayana, B.; Mulder, S.; Jayatissa, L.P.; Dahdouh-Guebas, F. Are the Mangroves in the Galle-Unawatuna Area (Sri Lanka) at Risk? A Social-Ecological Approach Involving Local Stakeholders for a Better Conservation Policy. *Ocean Coast. Manag.* **2013**, *71*, 225–237. [CrossRef]
61. Roy, S.K.; Mojumder, P.; Chowdhury, M.A.A.; Hasan, M.M. Evaluating Mangrove Forest Dynamics and Fragmentation in Sundarbans, Bangladesh Using High-Resolution Sentinel-2 Satellite Images. *Glob. Ecol. Conserv.* **2025**, *58*, e03493. [CrossRef]
62. Esri. GIS Software for Mapping and Spatial Analytics. Available online: https://www.esri.com/en-us/home (accessed on 18 June 2025).
63. IUCN. Available online: https://iucn.org/ (accessed on 18 June 2025).
64. Department of Coast Conservation and Coastal Resource Management. Available online: https://env.gov.lk/web/index.php/en/department-of-coast-conservation-and-coastal-resource-management (accessed on 18 June 2025).
65. NASA. Available online: https://www.nasa.gov/ (accessed on 18 June 2025).
66. Foga, S.; Scaramuzza, P.L.; Guo, S.; Zhu, Z.; Dilley, R.D.; Beckmann, T.; Schmidt, G.L.; Dwyer, J.L.; Joseph Hughes, M.; Laue, B. Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products. *Remote Sens. Environ.* **2017**, *194*, 379–390. [CrossRef]
67. Gopalakrishnan, L.; Satyanarayana, B.; Chen, D.; Wolswijk, G.; Amir, A.A.; Vandegehuchte, M.B.; Muslim, A.B.; Koedam, N.; Dahdouh-Guebas, F. Using Historical Archives and Landsat Imagery to Explore Changes in the Mangrove Cover of Peninsular Malaysia between 1853 and 2018. *Remote Sens.* **2021**, *13*, 3403. [CrossRef]
68. Earth Observations of Blue Carbon Ecosystems, NASA Applied Sciences (ARSET). Available online: https://appliedsciences.nasa.gov/get-involved/training/english/arset-earth-observations-blue-carbon-ecosystems (accessed on 18 June 2025).
69. Ruan, L.; Yan, M.; Zhang, L.; Fan, X.S.; Yang, H. Spatial-Temporal NDVI Pattern of Global Mangroves: A Growing Trend during 2000–2018. *Sci. Total Environ.* **2022**, *844*, 157075. [CrossRef] [PubMed]
70. Davis, Z.; Nesbitt, L.; Guhn, M.; van den Bosch, M. Assessing Changes in Urban Vegetation Using Normalised Difference Vegetation Index (NDVI) for Epidemiological Studies. *Urban For. Urban Green.* **2023**, *88*, 128080. [CrossRef]
71. Vizcaya-Martínez, D.A.; Flores-de-Santiago, F.; Valderrama-Landeros, L.; Serrano, D.; Rodríguez-Sobreyra, R.; Álvarez-Sánchez, L.F.; Flores-Verdugo, F. Monitoring Detailed Mangrove Hurricane Damage and Early Recovery Using Multisource Remote Sensing Data. *J. Environ. Manag.* **2022**, *320*, 115830. [CrossRef]
72. Shi, T.; Jue, L.; Zhongwen, H.; Huizeng, L.; Junjie, W.; Wu, G. New Spectral Metrics for Mangrove Forest Identification. *Remote Sens. Lett.* **2016**, *7*, 885–894. [CrossRef]
73. Fei, J.; Zhang, X.; Li, C.; Hao, F.; Guo, Y.; Fu, Y. A Deep Data Fusion-Based Reconstruction of Water Index Time Series for Intermittent Rivers and Ephemeral Streams Monitoring. *ISPRS J. Photogramm. Remote Sens.* **2025**, *220*, 339–353. [CrossRef]
74. Che, L.; Li, S.; Liu, X. Improved Surface Water Mapping Using Satellite Remote Sensing Imagery Based on Optimization of the Otsu Threshold and Effective Selection of Remote-Sensing Water Index. *J. Hydrol.* **2025**, *654*, 132771. [CrossRef]
75. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Xie, R.; Li, S. Integrating Satellite-Derived Climatic and Vegetation Indices to Predict Smallholder Maize Yield Using Deep Learning. *Agric. For. Meteorol.* **2021**, *311*, 108666. [CrossRef]
76. Lou, P.; Wu, T.; Yang, S.; Wu, X.; Chen, J.; Zhu, X.; Chen, J.; Lin, X.; Li, R.; Shang, C.; et al. Deep Learning Reveals Rapid Vegetation Greening in Changing Climate from 1988 to 2018 on the Qinghai-Tibet Plateau. *Ecol. Indic.* **2023**, *148*, 110020. [CrossRef]

77. Shuttle Radar Topography Mission, NASA's Earth Observing System. Available online: https://eospso.gsfc.nasa.gov/missions/shuttle-radar-topography-mission (accessed on 18 June 2025).
78. Fauzi, A.I.; Immitzer, M.; Atzberger, C. Exploring New Mangrove Horizons: A Scalable Remote Sensing Approach with Planet-NICFI and Sentinel-2 Images. *Ecol. Inform.* **2025**, *88*, 103152. [CrossRef]
79. Kotikot, S.M.; Spencer, O.; Cissell, J.R.; Connette, G.; Smithwick, E.A.H.; Durdall, A.; Grimes, K.W.; Stewart, H.A.; Tzadik, O.; Canty, S.W.J. Building a Mangrove Ecosystem Monitoring Tool for Managers Using Sentinel-2 Imagery in Google Earth Engine. *Ocean Coast. Manag.* **2024**, *256*, 107307. [CrossRef]
80. Yang, G.; Huang, K.; Sun, W.; Meng, X.; Mao, D.; Ge, Y. Enhanced Mangrove Vegetation Index Based on Hyperspectral Images for Mapping Mangrove. *ISPRS J. Photogramm. Remote Sens.* **2022**, *189*, 236–254. [CrossRef]
81. Simarmata, N.; Wikantika, K.; Tarigan, T.A.; Aldyansyah, M.; Tohir, R.K.; Fauzi, A.I.; Fauzia, A.R. Comparison of Random Forest, Gradient Tree Boosting, and Classification and Regression Trees for Mangrove Cover Change Monitoring Using Landsat Imagery. *Egypt. J. Remote Sens. Space Sci.* **2025**, *28*, 138–150. [CrossRef]
82. Zhao, C.; Jia, M.; Wang, Z.; Mao, D.; Wang, Y. Identifying Mangroves through Knowledge Extracted from Trained Random Forest Models: An Interpretable Mangrove Mapping Approach (IMMA). *ISPRS J. Photogramm. Remote Sens.* **2023**, *201*, 209–225. [CrossRef]
83. Wu, R.; Dai, Z.; Mei, X.; Luo, J.; Fagherazzi, S. Profiling of Mangrove Forest Dynamics in the Fly River Delta, Papua New Guinea. *Mar. Pollut. Bull.* **2025**, *217*, 118119. [CrossRef] [PubMed]
84. Bryan-Brown, D.N.; Connolly, R.M.; Richards, D.R.; Adame, F.; Friess, D.A.; Brown, C.J. Global Trends in Mangrove Forest Fragmentation. *Sci. Rep.* **2020**, *10*, 7117. [CrossRef] [PubMed]
85. Young, C.H.; Jarvis, P.J. Measuring Urban Habitat Fragmentation: An Example from the Black Country, UK. *Landsc. Ecol.* **2001**, *16*, 643–658. [CrossRef]
86. Reyes, G.J.; Smyth, A.R.; Qiu, J.; Reynolds, L.K. Highly Urbanized Mangrove Areas Are Small in Size, Fragmented, and Missed by Large-Scale Mapping Efforts. *Wetlands* **2024**, *44*, 103. [CrossRef]
87. Ranaweera, K.R.K.D.N.; Kamae, Y. Impact of El Niño Southern Oscillation on the First Inter-Monsoon Rainfall over Sri Lanka in the Post-El Niño Years. *Front. Clim.* **2024**, *6*, 1361322. [CrossRef]
88. Ranasinghe, P.N.; Hughen, K.A.; Wijewardhana, T.U.T.; Ortiz, J.D.; Kapuge, A.K.I.U.; Ratnayake, K.M.; Nanayakkara, N.U.; Hewawasam, A.L.T.; Siriwardana, Y.P.S.; Kodithuwakku, S. Interplay of the Indian Summer Monsoon and Intermonsoon Precipitation in Sri Lanka Due to ITCZ Migration during the Last 80000 Years. *Sci. Rep.* **2024**, *14*, 22178. [CrossRef]
89. Dutta, A.; Mohan, M.; Hendy, I.; Almealla, R.; Watt, M.S.; Burt, J.A.; Torres-florez, J.P.; Almansoori, A.; Alzahlawi, N.; Abdullah, M.; et al. Optimizing Mangrove Afforestation Site Selection in Gulf Cooperation Council Nations Using Remote Sensing and Machine Learning. *Sci. Total Environ.* **2025**, *988*, 179805. [CrossRef]
90. Taher, T.B.; Althaus, C.E.; Tranter, P.J.; Evans, M.C. Impacts of Shrimp Aquaculture on the Local Communities and Conservation of the World's Largest Protected Mangrove Forest. *Environ. Sci. Policy* **2023**, *147*, 351–360. [CrossRef]
91. Lovelock, C.E.; Cahoon, D.R.; Friess, D.A.; Guntenspergen, G.R.; Krauss, K.W.; Reef, R.; Rogers, K.; Saunders, M.L.; Sidik, F.; Swales, A.; et al. The Vulnerability of Indo-Pacific Mangrove Forests to Sea-Level Rise. *Nature* **2015**, *526*, 559–563. [CrossRef]
92. Thao, N.D.; Kiet, N.; Anh, L.T.; Huynh, P.T.D.; Hiep, L.D.; Trang, D.D.T. Wave Attenuation in Coastal Mangroves: The Case of Mekong Delta Vietnam. *IOP Conf. Ser. Earth Environ. Sci.* **2023**, *1226*, 012026. [CrossRef]
93. Saoum, M.R.; Sarkar, S.K. Monitoring Mangrove Forest Change and Its Impacts on the Environment. *Ecol. Indic.* **2024**, *159*, 111666. [CrossRef]
94. Wang, X.; Zhang, R.; Jia, M.; Wang, Z.; Zhao, C.; Guo, X.; Li, H.; Xue, Z.; Zhou, C.; Xiao, X. Significant Greening of China's Mangrove Forests during the Last Three Decades. *Ecol. Indic.* **2025**, *176*, 113632. [CrossRef]
95. Umroh; Adi, W.; Sari, S.P. Detection of Mangrove Distribution in Pongok Island. *Procedia Environ. Sci.* **2016**, *33*, 253–257. [CrossRef]
96. He, Z.; Zhao, L.; Zhuang, X.; Zeng, Y.; Hu, B.; Yu, X.; Liang, J.; Peng, Y.; Chen, Z. Tidally Driven Porewater Exchange and the Key Affecting Factors in Mangroves in the Pearl River Estuary, China. *J. Sea Res.* **2025**, *204*, 102572. [CrossRef]
97. Zhang, H.; Wei, S.; Liang, X.; Chen, Y.; Zhang, H. Scale Effects in Mangrove Mapping from Ultra-High-Resolution Remote Sensing Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2025**, *136*, 104310. [CrossRef]
98. Sasmito, S.D.; Basyuni, M.; Kridalaksana, A.; Saragi-Sasmito, M.F.; Lovelock, C.E.; Murdiyarso, D. Challenges and Opportunities for Achieving Sustainable Development Goals through Restoration of Indonesia's Mangroves. *Nat. Ecol. Evol.* **2023**, *7*, 62–70. [CrossRef]
99. Central Environmental Authority (CEA). Available online: https://www.cea.lk/ (accessed on 18 June 2025).
100. Department of Wildlife Conservation. Available online: https://www.dwc.gov.lk/ (accessed on 18 June 2025).
101. IUCN, Sri Lanka. Available online: https://iucn.org/our-work/region/asia/countries/sri-lanka (accessed on 18 June 2025).
102. National Aquatic Resources Agency (NARA). Available online: http://www.nara.ac.lk/ (accessed on 18 June 2025).
103. National Science Foundation (NSF). Available online: https://www.nsf.gov.lk/ (accessed on 18 June 2025).

104. The Urban Development Authority. Available online: https://www.uda.gov.lk/ (accessed on 18 June 2025).
105. The Road Development Authority. Available online: https://rda.gov.lk/index.php?lang=en (accessed on 18 June 2025).
106. Ministry of Environment. *National Guidelines for the Restoration of Mangrove Ecosystems of Sri Lanka*; Biodiversity Secretariat Ministry of Environment No. 416/C/1, Robert Gunawardana Mawatha Battaramulla; Ministry of Environment: Battaramulla, Sri Lanka, 2021; ISBN 9788578110796.
107. Mangrove Regeneration in Sri Lanka Recognized as One of Seven UN World Restoration Flagships, United Nations in Sri Lanka. Available online: https://srilanka.un.org/en/262045-mangrove-regeneration-sri-lanka-recognized-one-seven-un-world-restoration-flagships (accessed on 18 June 2025).

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.

land

MDPI

Article

Assessment of Landslide Susceptibility and Risk in Tengchong City, Southwestern China Using Machine Learning and the Analytic Hierarchy Process

Changwei Linghu [1], Zhipeng Qian [2], Weizhe Chen [2,*], Jiaren Li [2], Ke Yang [2], Shilin Zou [1,3], Langlang Yang [1], Yao Gao [1], Zhiping Zhu [1] and Qiankai Gao [1]

[1] Kunming General Survey of Natural Resources Center, China Geological Survey, Kunming 650100, China; linghuchangwei@mail.cgs.gov.cn (C.L.)
[2] Hubei Key Laboratory of Critical Zone Evolution, School of Earth Sciences, China University of Geosciences, Wuhan 430074, China; zpqian@cug.edu.cn (Z.Q.)
[3] Technology Innovation Center for Natural Ecosystem Carbon Sink, Ministry of Natural Resources, Kunming 650100, China
* Correspondence: wzchen@cug.edu.cn

Abstract

Southwestern China, characterized by highly undulating terrain and mountainous areas, faces frequent landslide disasters. However, previous studies in this region mostly neglected the role of extreme rainfall in landslide susceptibility assessment and the socio-economic risks threatened by landslides. To address these gaps, this study integrated 688 recorded landslides for Tengchong City in the southwest of China and 10 influencing factors (topography, lithology, climate, vegetation, and human activities), particularly two extreme precipitation indices of maximum consecutive 5 day precipitation (Rx5day) and maximum length of wet spell (CWD). These influencing factors were selected after ensuring variable independence via multicollinearity analysis. Four machine learning models were then built for landslide susceptibility assessment. The Random Forest model performed the best with an Area Under Curve (AUC) of 0.88 and identified elevation, normalized difference vegetation index (NDVI), lithology, and CWD as the four most important influencing factors. Landslides in Tengchong are concentrated in areas with low NDVI (<0.57), indicating increased vegetation cover might reduce landslide frequency. Landslide risk was further quantified via the Analytic Hierarchy Process (AHP) by integrating multiple socio-economic factors. High-risk zones were pinpointed in central-southern Tengchong (e.g., Heshun and Tuantian townships) due to their high social exposure and vulnerability. Overall, this study highlights extreme rainfall and vegetation as key modifiers of landslide susceptibility and identifies the regions with high landslide risk, which provides targeted scientific support for regional early-warning systems and risk management.

Keywords: Random Forest; landslide susceptibility; landslide risk; Analytic Hierarchy Process; extreme rainfall

1. Introduction

Landslides refer to natural processes in which a mass of rock or soil on an inclined surface moves downward, either partially or completely, induced by either natural or human-induced causes [1]. As a common geological disaster, a landslide is sudden and destructive. It not only directly affects construction, transportation, and the ecological environment but also threatens the safety of people's lives and property, causing far-reaching

social and economic impacts [2,3]. According to statistical data from the China Natural Resources Bulletin (2024) [4], landslides accounted for 57.98% (3316 out of 5719) of all recorded geological disasters in China in 2024. Southwestern China, with its complex terrain and extensive mountain ranges, is a region that has always been at high risk of landslides and other geological disasters. The Department of Emergency Management of Yunnan Province (2025) reported that geological hazard events in 2024 impacted an estimated 80,000 individuals, spanning 48 county-level administrative units across 15 prefecture-level divisions of the province [5]. Therefore, it is crucial to undertake a comprehensive landslide susceptibility assessment, geohazard risk analysis, and identification of landslide-prone areas in Yunnan province of southwest China.

Landslide susceptibility is defined as the probability of landslides occurring in an area under basic geographic conditions. Traditional models, such as the Weights of Evidence (WOE) method [6–8] and the Coefficient of Determination approach [9,10], have been widely employed in landslide susceptibility assessments. These models can reveal the dominant influencing factors of regional landslides. For example, the results obtained from the WOE method demonstrated that the distribution of landslides in Badong is predominantly concentrated along both sides of rivers and is influenced by faults [11]. Similarly, the results obtained from the Coefficient of Determination approach demonstrate that rivers, faults, and rocks are the primary factors of landslides in the Qin-Ba Mountain region [12]. Although traditional models are straightforward, easy to comprehend, and highly interpretable, these assessment techniques are limited by their incapacity to fully account for all possible influencing factors in the analysis, e.g., vegetation and extreme rainfall.

In recent years, machine learning has provided a new framework for landslide susceptibility research, leveraging its strong capabilities in data mining and pattern recognition, as well as providing more accurate predictive results. Among them, the most representative models are Random Forest (RF) [13,14], Back Propagation Neural Networks (BPNNs) [10,15], Support Vector Machines (SVMs) [9,16], Logistic Regression (LR), and Naive Bayes (NB) [17]. Researchers have conducted a lot of studies to evaluate different areas using different machine learning models. For example, the multivariate LR method is employed in northeast Kansas, identifying slope as the most significant predictor of landslide risk [18]. Two machine learning methods called RF and RF-SVM are developed to evaluate landslide susceptibility in the Rangit River watershed, and the results indicate both high regional risk and the superior predictive capability of the hybrid model [19]. The evaluation conducted in the Nandakini River basin compares three machine learning models, including RF, Deep Learning Neural Network (DLNN), and Artificial Neural Network (ANN), and demonstrates that the DLNN achieves the highest predictive accuracy in landslide susceptibility assessment [20]. Similarly, the comparative evaluation between LR and other machine learning approaches reveals that the RF model exhibits the highest predictive accuracy in the Sihjhong watershed of Taiwan [21]. The consistently better performance of machine learning models across various regions confirms their technical superiority in landslide susceptibility assessment.

Landslide risk refers to the impact of landslides on human activities and economic factors. The Analytic Hierarchy Process (AHP) is a quantitative and multi-criteria method designed for the hierarchical representation of a decision-making problem. This approach has been extensively employed in disaster risk assessment. For example, in ground subsidence risk assessment, AHP can quantitatively assign weights to hydrological, geological, anthropogenic factors and other factors [22]. Similarly, AHP can be used in flood, landslide and earthquake hazard assessment by assigning a weight value to each selected factor to provide risk maps for multi-hazard assessment modelling [23–25]. It has previously

been mathematically validated and used to identify landslide risk in India [26]. As a semi-quantitative technique, AHP combines qualitative and quantitative methods. Machine learning training and simulation rely heavily on inventory data. However, landslide risk is a further assessment by considering socio-economic factors. Therefore, machine learning-based susceptibility maps serve as a key input in the AHP framework, creating a hybrid model. This approach weights outputs of machine learning and other layers to ensure the final risk maps are statistically reliable and informed by socio-economic judgment. Consequently, the hybrid model and the resulting map are valuable for decision-making.

Rainfall, different from terrain, geology, and other factors, is greatly affected by climate change. The relationship between rainfall and landslides has always been a research hotspot. Previous work identified a close relation between rainfall and landslide [27,28]. It is well accepted that shallow landslides and debris flows are triggered by high-intensity, short rainfalls, and deep-seated landslides occur as a result of less intense rain over a long time [29–31]. However, previous studies have mostly considered the annual average precipitation [32–34]. Extreme rainfall is often a key factor in triggering landslides, especially in areas with existing potential landslides. Heavy rainfall exerts an influence on the land surface, thereby affecting the movement of land surface materials. Long-term continuous rainfall leads to severe water infiltration, affecting deep soil layers. Therefore, the integration of extreme rainfall indices (e.g., rainfall intensity and duration) is expected to enhance the predictive precision of landslide susceptibility mapping through capturing rainfall intensity extremes beyond traditional mean-based metrics.

Therefore, this study employed four commonly used machine learning models and combined them with AHP for landslide susceptibility and risk assessment in Tengchong city. Compared with previous studies, this study has two major innovations. First, in the selection of influencing factors, the important influence of extreme rainfall on landslide occurrence is fully considered. The extreme rainfall indicators are introduced, and two representative factors are screened by correlation analysis for inclusion in landslide susceptibility assessment. It improves the accuracy of landslide prediction under extreme rainfall conditions. Second, a landslide risk zoning framework is developed through the integration of AHP and machine learning modeling. The results provide valuable support to assist local government in developing targeted disaster mitigation measures and strengthening landslide risk management capabilities.

2. Data and Methods

2.1. Study Area

Tengchong is located in the western Yunnan Province of southwest China, covering an area of 5845 km^2 (Figure 1). Situated within the Indian Ocean monsoon climate zone, Tengchong exhibits a humid subtropical climate characterized by copious annual precipitation, marked seasonal contrasts between dry and wet phases, and pronounced altitudinal climatic gradients due to its complex topography, posing significant challenges for landslides and flood hazard mitigation. It lies between 24°37′–25°51′ N and 98°05′–98°46′ E, at the southern end of the Hengduan Mountains. Located on the western side of the Indian–Eurasian plate collision zone and within the Sanjiang fold system, Tengchong is tectonically active, rich in geothermal resources, and frequently affected by fault and magma activity [35]. As a result, earthquakes and geological disasters occur frequently.

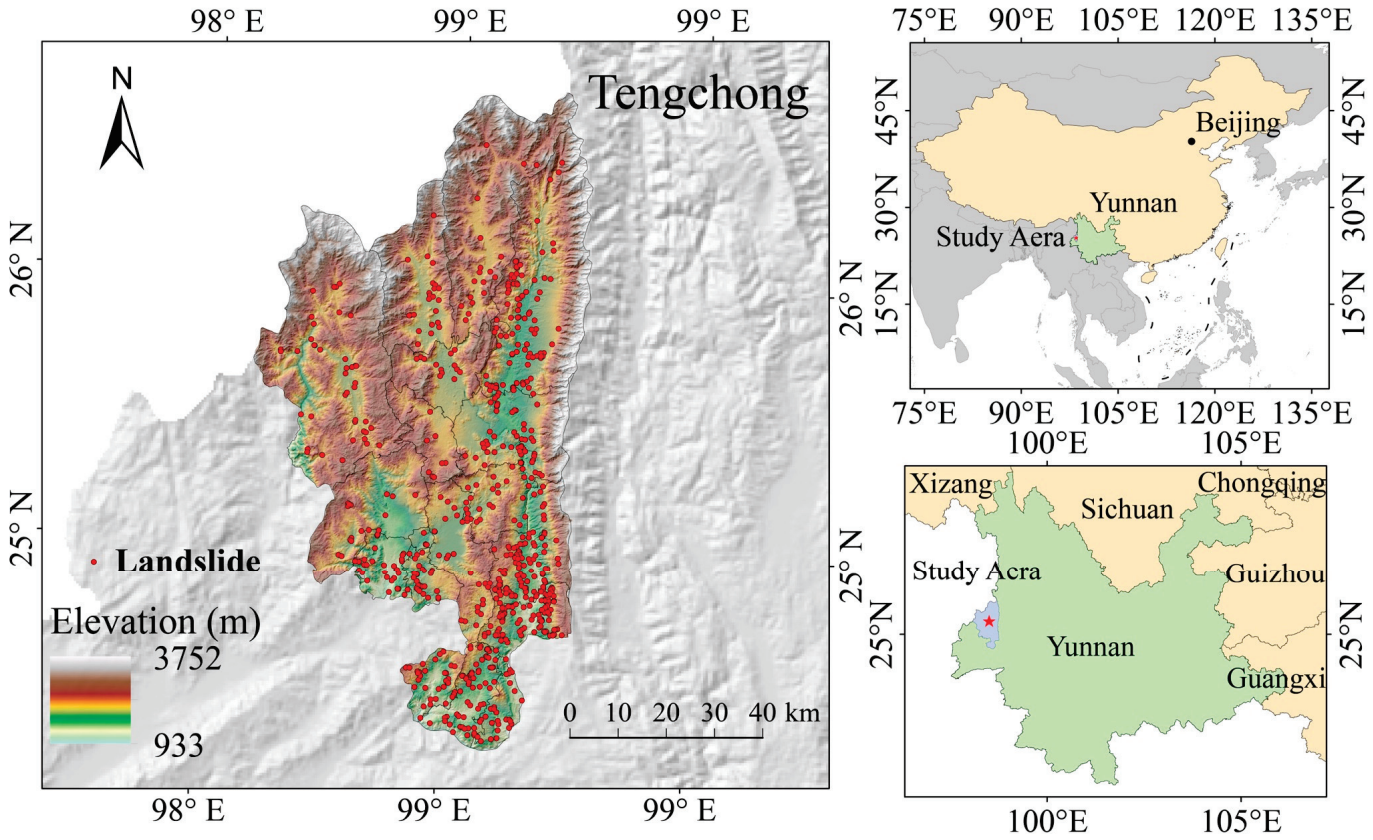


Figure 1. Location of Tengchong city of Yunnan province, southwestern China. The red dots in the left panel represent the 688 recorded landslides. The blue area with the red star in the right panels denotes the study area.

2.2. Data and Factors of Assessment

A total of 688 landslide sites, spatially dispersed across the study area, were mapped and verified by the China Geological Survey. The data for the 688 landslides are mainly derived from detailed field surveys conducted between 2000 and 2020. The dataset includes each landslide with information on landslide types, slope structures, and possible causal factors. In addition, the dataset provides precise latitudes and longitudes for specific landslide occurrence points to facilitate their easy use in subsequent analyses. This study incorporates landslide-influencing indicators, including terrain, vegetation indices, rainfall indices, geology, rivers and roads, as well as population and build-up (Table 1). For vegetation indices, rainfall data, population density and building data, average values from 2000 to 2020 are used.

Given the inconsistent resolutions of the datasets and Tengchong's spatial scope, all datasets were resampled to a uniform 30 m resolution. In detail, distance variables to target features (e.g., roads, rivers, and buildings) were first calculated at the raster scale using the Euclidean Distance tool in ArcGIS10.8. For this computation, vector-to-raster conversion was performed in two key steps: first, original vector features (e.g., linear road features) were converted to a binary source raster via the Feature to Raster tool—pixels containing the target features were assigned a value of 1, while all other pixels were set to NoData. Following this, the source raster was used as input for the Euclidean Distance tool to generate a final continuous distance raster. For each pixel in this raster, the value represents the shortest straight-line (Euclidean) distance from the pixel's center to the nearest target feature. Additionally, the rainfall raster dataset was resampled to a higher spatial resolution using the nearest neighbor method. We assumed that the average climate

state over 20 years within a 4 km grid cell is consistent. The total number of raster cells in the study area is approximately 6.32×10^6.

Table 1. List of datasets with their resolutions and data sources.

Datasets	Resolution/Scale	Data Sources	Variables
DEM	30 m	Geospatial Data Cloud Platform (http://www.gscloud.cn (accessed on 1 October 2024))	Elevation, slope, and curvature are calculated from DEM
NDVI	30 m	Surface reflection data from the Landsat 8 provided by the United States Geological Survey (https://www.usgs.gov (accessed on 1 October 2024))	Vegetation indices
Rainfall	4 km	China Daily Meteorological Dataset [36] (https://zenodo.org/records/10963932 (accessed on 1 October 2024))	Rainfall indices are calculated from daily rainfall data
Geological map	1:1,000,000	Geoscientific Data and Discovery Publishing System (https://www.ngac.cn (accessed on 1 October 2024))	Faults and stratigraphic lithologies are extracted from geological map data
River and road	1:1,000,000	National Catalogue Service for Geographic Information (www.webmap.cn (accessed on 1 October 2024))	Distances from the river and road are calculated
Population density and build-up	30 m	Global Human Settlement Layer (https://human-settlement.emergency.copernicus.eu (accessed on 1 October 2024))	Assessing landslide risk

Topographic factors include elevation, slope, and curvature. Elevation affects rainfall, vegetation, and biological activities, and consequently can cause landslides to spread in a certain range of elevations [37]. Slope also has a direct effect on landslide development [9]. The number of landslides occurring increases as the slope increases and then decreases after a certain peak [12]. Curvature is an expression of the degree of change in slope distortion [38]. Positive values indicate convex slopes, while negative values indicate concave slopes, and curvature closer to 0 suggests a gentler slope.

The land surface factors affecting landslide susceptibility are normalized difference vegetation index (NDVI) and distance from rivers. Higher NDVI values indicate healthier and denser vegetation [39]. Research shows adding NDVI to the landslide susceptibility mapping model is better than the model without NDVI [40]. The river system exhibits varying degrees of scouring and infiltration on both sides of the river, resulting in the formation of steep bank slopes, which induces landslides.

Rock and soil bodies, as well as faults, are the geologic basis for the occurrence of landslides. Different rock types have different weathering resistance, which affects the degree of development of landslides. Faults reflect the physical properties of the geologic body, affecting the stability of slopes. In fault fracture zones, the slope is prone to sliding to a certain extent. When rainfall occurs, water infiltration softens the soil in the sliding zone and influences saturation of the sliding mass. This leads to an increase in the downward sliding force of the slope mass, resulting in a landslide [41].

The main human activity factor considered in this study is the distance from roads. Excavation and unloading at the foot of slopes and accumulation at the top of slopes are the main causes of landslides induced by engineering activities. Geotechnical excavation and unloading will concentrate the stress on slopes. This leads to deterioration of slope stability conditions, thus causing landslides [42].

Atmospheric precipitation-induced elevation of pore water pressure reduces the effective stress within the soil matrix, which in turn decreases the shear resistance of potential sliding zones. This hydrological process softens the rock-soil mass through water infiltra-

tion, leading to a decline in mechanical properties and subsequent acceleration of slope failure mechanisms [43]. The expert team on Climate Change Detection and Indices (ETCCDI) and its predecessor have developed and calculated a suite of extreme indices about rainfall, which could measure the intensity and time of rainfall. Therefore, we adopt the ten extreme rainfall indicators proposed by ETCCDI, as shown in Table 2.

Table 2. Definition of extreme precipitation indices.

Indicators	Name	Definition	Unit
CDD	Maximum length of dry spell	Annual maximum number of consecutive days with daily precipitation <1 mm	day
CWD	Maximum length of wet spell	Annual maximum number of consecutive days with daily precipitation of at least 1 mm	day
Rx1day	Maximum 1 day precipitation	Annual maximum daily precipitation	mm
Rx5day	Maximum consecutive 5 day precipitation	Annual maximum continuous precipitation for 5 days	mm
R95p	Annual total PRCP when daily precipitation > 95 p	Annual accumulated precipitation greater than the 95th percentile of daily precipitation	mm
R99p	Annual total PRCP when daily precipitation > 99 p	Annual accumulated precipitation greater than the 99th percentile of daily precipitation	mm
R10mm	Annual count of days when PRCPTOT $\geq$ 10 mm	Total number of days with annual precipitation $\geq$ 10 mm	day
R20mm	Annual count of days when PRCPTOT $\geq$ 20 mm	Total number of days with annual precipitation $\geq$ 20 mm	day
PRCPTOT	Annual total precipitation in wet days	Annual accumulated precipitation greater than 1 mm per day	mm
SDII	Simple precipitation intensity index	The ratio of the total amount of precipitation of 1 mm or more to the number of days	mm/day

2.3. Machine-Learning-Based Landslide Susceptibility Assessment

The overall methodological framework (Figure 2) of this study consists of the following key steps: (1) data processing involving preparation of landslide and socio-economic factor layers; (2) random sample selection for model training and validation. This set consists of 600 samples, comprising 300 landslide and 300 non-landslide samples; (3) five-fold cross-validation is performed to evaluate more than 10 machine learning methods. The four best machine learning methods (RF, QDA, LR, and KNB) are used to produce landslide susceptibility maps. The best-performing model is selected for further analysis; (4) utilization of the Analytic Hierarchy Process (AHP) to determine factor weights. A consistency check is performed (CR < 0.1). Should the result fail the check, the judgment matrix is reconstructed until consistency is achieved; (5) integration of the optimal susceptibility map and additional factor layers via weighted overlay analysis based on AHP-derived weights; (6) generation of a final landslide risk map.

Four machine learning techniques for classification are employed to construct landslide susceptibility models. First, RF is an ensemble learning method that builds multiple decorrelated decision trees through bootstrap aggregation (bagging) with random feature selection. Each base learner trains on independently bootstrapped data samples using a randomly chosen subset of features, with final predictions determined by majority voting across the ensemble [44]. Second, Quadratic Discriminant Analysis (QDA) is a probabilistic classification method. It models each class using multivariate Gaussian distributions with class-specific parameters. By estimating distinct covariance matrices for each class, QDA generates quadratic decision boundaries that effectively handle heterogeneous class distributions. Classification is performed by comparing posterior probabilities computed through discriminant functions derived from these class-conditional density estimates [45]. Third, Logistic Regression (LR) is a statistical method for binary classification. It models the probability of class membership using a logistic function. It estimates coefficients from training data to predict binary outcomes (0 or 1) based on a decision threshold [46]. Fourth, Kernel Naive Bayes (KNB) extends conventional Naive Bayes (NB) classifiers by replacing parametric distribution assumptions with kernel density estimation. This nonparametric approach enables improved modeling of continuous features with nonlinear

distributions [47]. KNB preserves classification efficiency through posterior probability maximization while demonstrating superior capability in modeling complex data patterns compared to standard approaches.

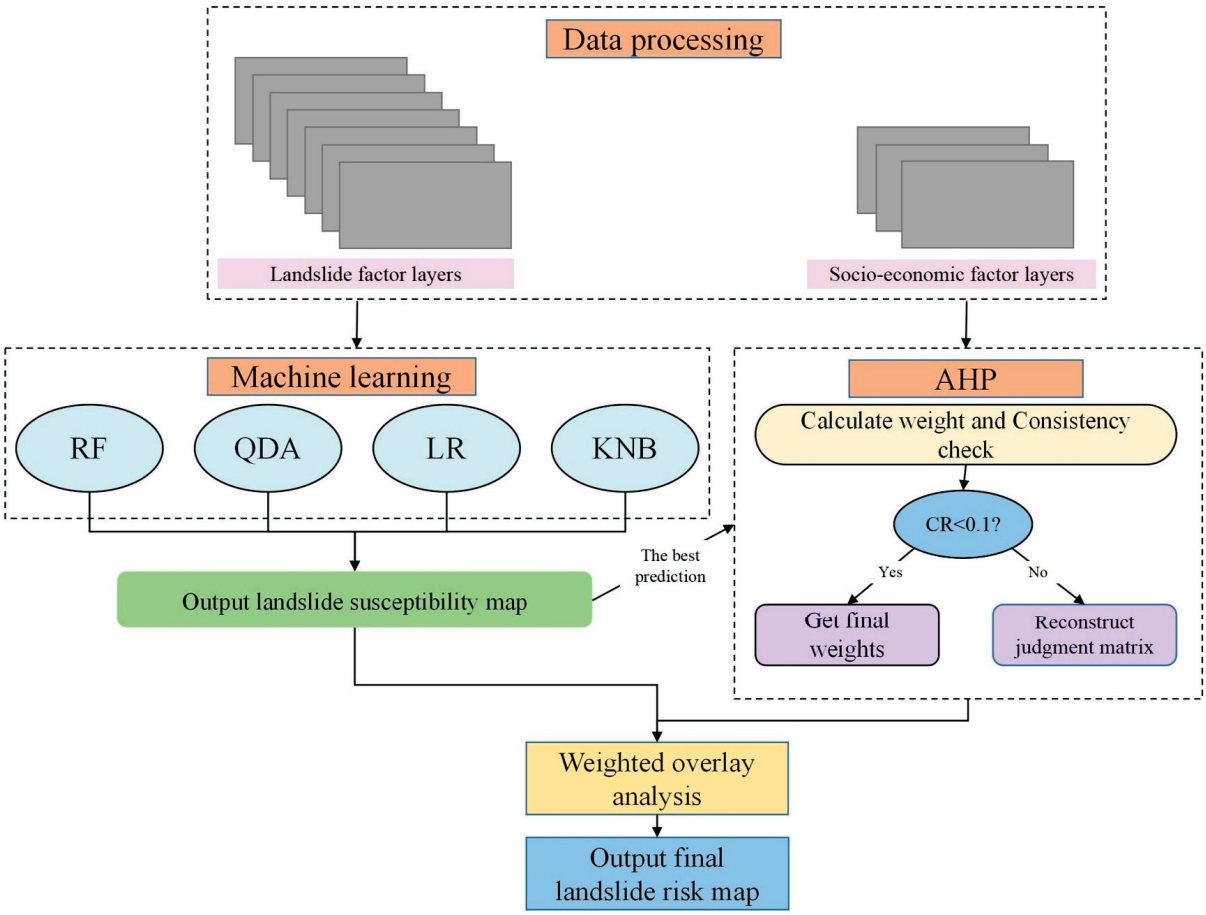


Figure 2. Framework diagram for assessing landslide susceptibility and risk.

2.4. AHP-Based Landslide Risk Assessment

The AHP is a decision-making tool that assists in solving complex problems by dividing them into simpler criteria, which can derive ratio scales from discrete and continuous pairwise comparisons [48]. AHP can effectively construct measurement systems in instances where physical or statistical measures are unavailable. This method works by quantifying subjective evaluations into relative weights for each element based on their importance [49]. Numerous studies worldwide use this approach to rank the elements depending on their importance [50–53]. AHP involves the following key steps [54]:

Step 1. Clarify the decision-making process and establish objectives.

Step 2. Creation of hierarchies and construction of relative importance matrices. The relative importance is usually measured on a scale of 1 to 9, where 1 indicates equal importance and higher values indicate greater importance [55].

Step 3. Determine the criterion weights. When determining the eigenvector, first, each element of the obtained matrix is divided by the sum of its column. Then, the arithmetic mean of each row of the calculated new matrix is taken to determine the criterion weights.

Step 4. Calculate consistency. The reliability assessment involves calculating the principal eigenvalue to derive the consistency index (CI, Equation (1)), followed by computing the consistency ratio (CR, Equation (2)) using the randomness index (RI). If the obtained

consistency ratio does not exceed the value of 0.1, it is interpreted that the criterion weights are reliable, and the decision matrix is formed.

$$CI = \frac{\lambda_{max} - n}{n - 1} (n = 1 \sim 9) \quad (1)$$

where λ_{max} is the maximum eigenvalue and n is the order of the matrix.

$$CR = \frac{CI}{RI} < 0.1 \quad (2)$$

where CI is the consistency index and RI is the random consistency index. The CI value is calculated using Equation (1). RI varies according to the matrix order n consequently (Table 3), referring to Saaty [56].

Table 3. Random consistency index (RI).

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

3. Results

3.1. Correlation Analyses of Multiple Factors

The subtropical monsoon climate in Tengchong exhibits a weakly defined seasonal cycle, with clear delineation between dry and wet periods. The rainy season extends from May to October, while the remaining months constitute the dry season. The region receives substantial precipitation, which classifies it among the southern high-rainfall zones of China. The average annual rainfall in Tengchong exceeds 1000 mm, and the city is relatively humid overall—with the western region being more humid than the eastern region (Figure 3i). The maximum consecutive dry days (CDDs) can persist for up to one month (Figure 3a), likely influenced by the monsoon climate, while the maximum consecutive wet days (CWDs) are shorter, with approximately 20 days (Figure 3b). Notably, extreme values of maximum 1 day precipitation (Rx1day) and maximum consecutive 5 day precipitation (Rx5day) occur in the northeast region, indicating the higher possibility of extreme rainfall events in these areas (Figure 3f–j). Theoretically, both short-duration, high-intensity rainfall and long-duration, low-intensity rainfall events can trigger landslides [57]. This suggests that CWDs, which reflects long-term wetting trends, and Rx5day, which represents short-term heavy rainfall, should be considered as landslide-predicting variables. Additionally, these two indices show a low correlation coefficient of -0.24 and a variance inflation factor (VIF) less than 3, making them suitable as independent factors for landslide susceptibility assessment (Figure 4a).

Correlation analyses of all factors were performed after selecting the two extreme rainfall indicators, CWDs and Rx5day (Figure 4a). Human engineering activities exhibit substantial altitude dependence, leading to a characteristically high correlation coefficient between human engineering activities and altitude (Figure 4b). Nevertheless, we cannot ignore the impact of human engineering activities on landslides based solely on the correlation coefficient. Notably, all correlation coefficients satisfy the medium multicollinearity threshold ($|r| < 0.5$) [58,59], and their VIF values are all less than 2, confirming the variables' statistical independence as predictive features for machine learning model development.

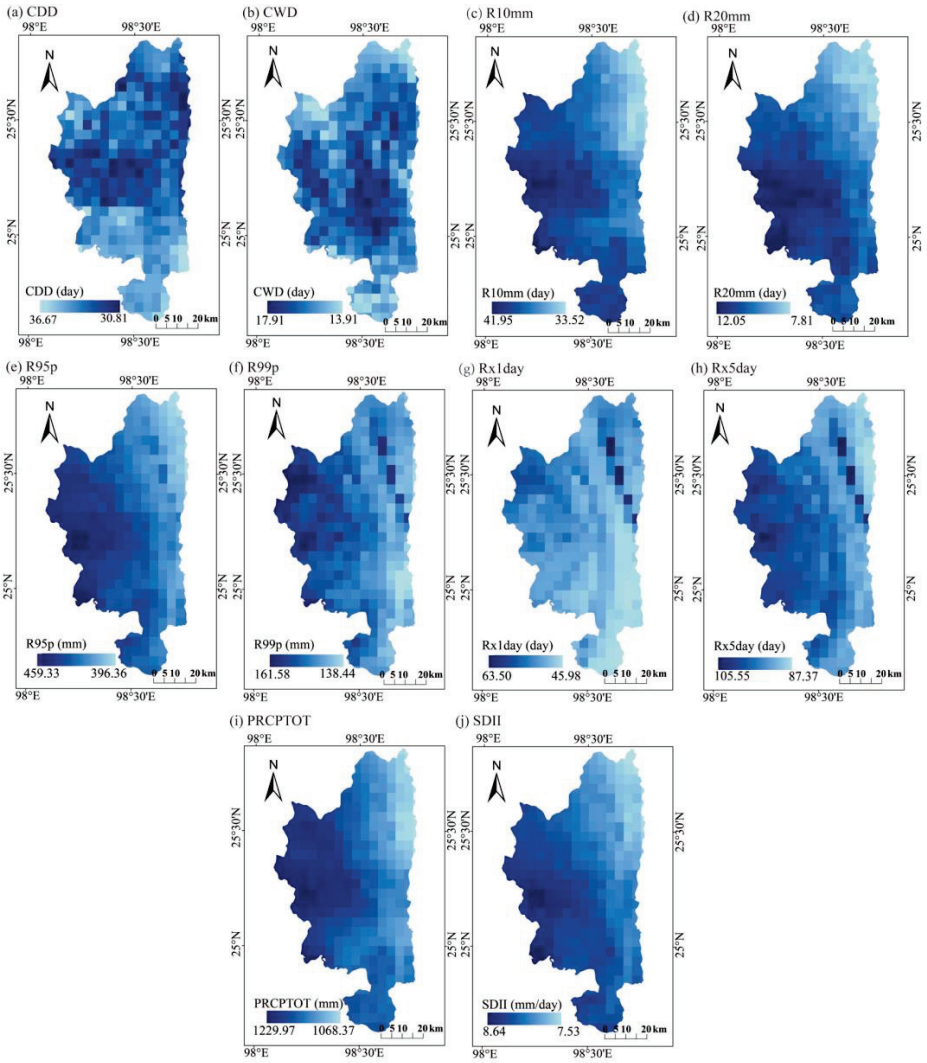


Figure 3. Spatial distribution of extreme rainfall indicators averaged from 2000 to 2020 for Tengchong. These rainfall indicators are derived from further calculations of daily rainfall. Subfigure labels (**a**–**j**) indicate the 10 extreme rainfall indicator metrics as defined in Table 2.

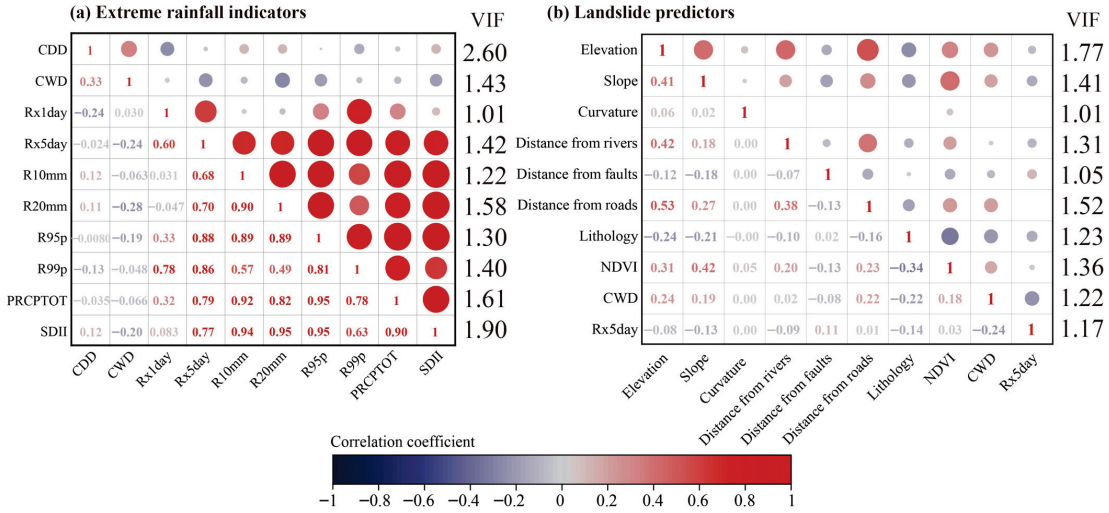


Figure 4. Correlation analyses and variance inflation factor (VIF) between the ten extreme rainfall indicators (**a**) and the ten landslide predictors (**b**). Correlation coefficient values range from −1 to 1,

with larger absolute values indicating higher correlation. Positive correlation coefficient values indicate a positive correlation between two factors, and negative values indicate a negative correlation. The magnitude of the correlation can be represented by r. A larger absolute value of r indicates a stronger positive or negative correlation. VIF is a measure of the severity of multicollinearity in multiple linear regression models. It is indicative of the ratio of the variance of the estimated regression coefficient to the variance when assuming nonlinear correlation between the independent variables. A value of VIF greater than one indicates the presence of multicollinearity, and a value closer to one suggests a lesser degree of multicollinearity. In instances of severe multicollinearity, it is imperative that appropriate methods are implemented to address the issue. The 10 landslide predictors are used for training a machine-learning-based landslide susceptibility model.

The spatial distribution of landslide disasters in Tengchong demonstrates distinct regional clustering, with pronounced concentrations in the eastern and southern sectors. Notably, a high-density cluster emerges along the 25° N latitudinal zone in the eastern region. Visually, landslide disasters are clearly concentrated in areas near rivers and roads, with weak spatial correlations with other environmental factors (Figure 5). Statistical analysis was conducted to investigate the correlations between landslides and the susceptibility factors. It should be noted that slope, curvature, NDVI, CWDs, and Rx5day do not have explicit integer threshold breakpoints. Therefore, the use of quintile processing by the number of effective grids was considered. Recorded landslides are mainly concentrated in the range of 1500 m to 1800 m above sea level (Figure 6a) and in areas with gentle slopes (Figure 6b), likely due to generally gentle topography in the region. They also tend to occur where curvature is near zero, with a higher likelihood on convex (positive curvature) slopes (Figure 6c). Landslides are concentrated near roads and rivers, while the influence of fault distance is less pronounced (Figure 6d–f). It suggests a more important role of hydrological condition and human activities on landslides. In addition, landslides are concentrated in the area with Quaternary alluvial deposits and Neogene sandstones. Tengchong has a high vegetation cover, while the landslides are concentrated in the middle and low NDVI areas (Figure 6h). In particular, both the number and relative density of landslides decrease as NDVI increases. Lastly, landslides are more likely to occur during extreme rainfall when the CWDs are longer than 15.5 days and the Rx5day ranges from 91 mm to 102 mm (Figure 6i,j).

3.2. Assessment of Landslide Susceptibility

Four machine-learning models were trained to assess landslide susceptibility for Tengchong with recorded landslides and ten predictors. The receiver operating characteristic (ROC) curve is a graphical representation of the diagnostic ability of a binary classifier system as its discrimination threshold is varied, plotting the true positive rate (TPR) against the false positive rate (FPR). The area under the ROC curve (AUC) quantifies the model's discriminative power across all possible classification thresholds, with values ranging from 0.5 (random guessing) to 1.0 (perfect discrimination). For the four models, RF, LR, QDA, and KNB, the AUC values are 0.877, 0.813, 0.807, and 0.819, with the RF model demonstrating superior performance (Figure 7a). Confusion matrix calculations for the four models indicate that the Random Forest model exhibits the highest number of correctly predicted grid cells for both landslide and non-landslide categories (Table 4). This model is the most effective. The RF model further ranks the importance of each factor and shows that elevation and NDVI have the greatest influence on landslide occurrence, followed by lithology and the extreme rainfall indicator CWDs (Figure 7b). This indicates that landslides are not only strongly correlated with terrain and geological background but also closely linked to vegetation and climatic conditions.

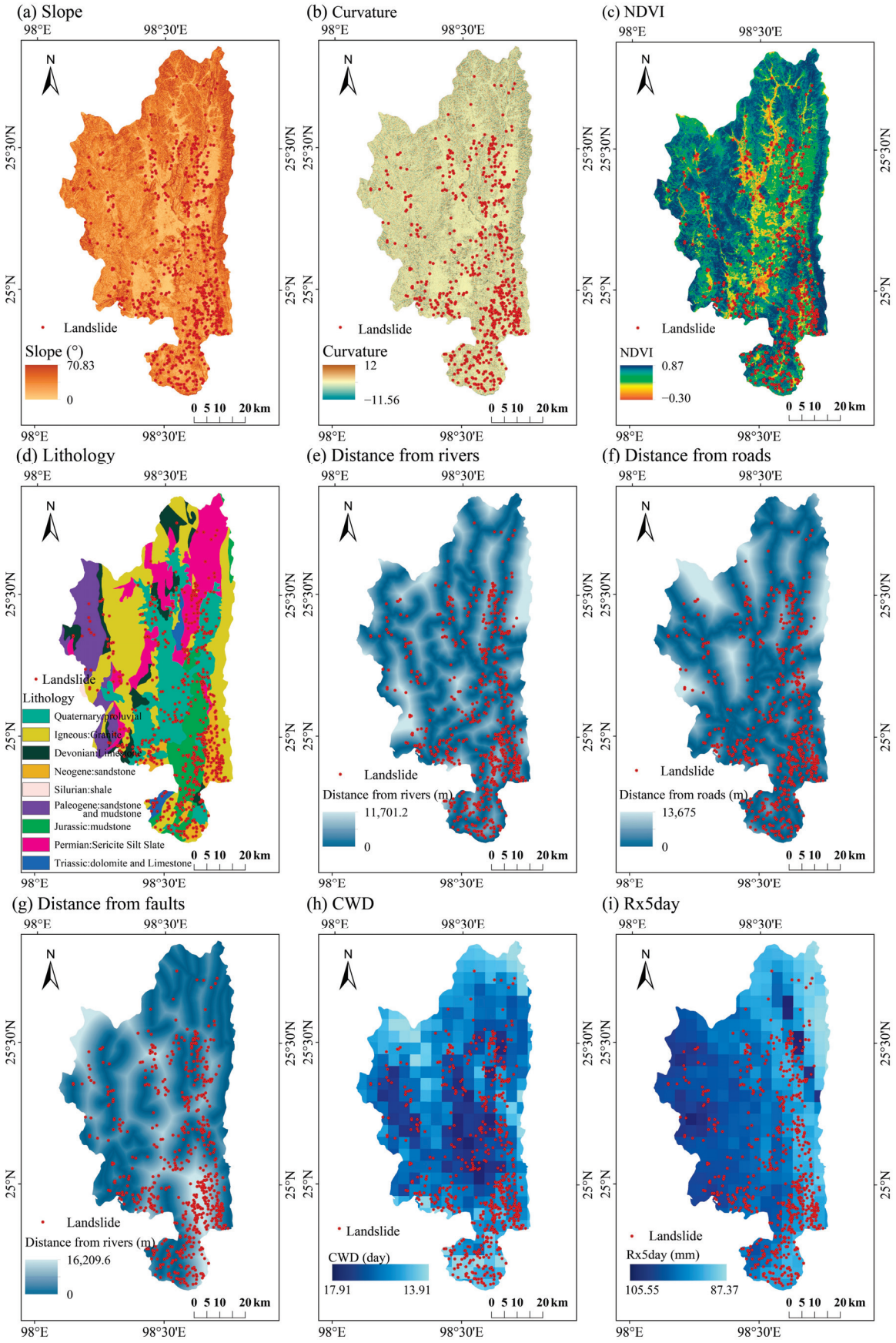


Figure 5. The spatial distribution of influencing factors and the recorded landslides.

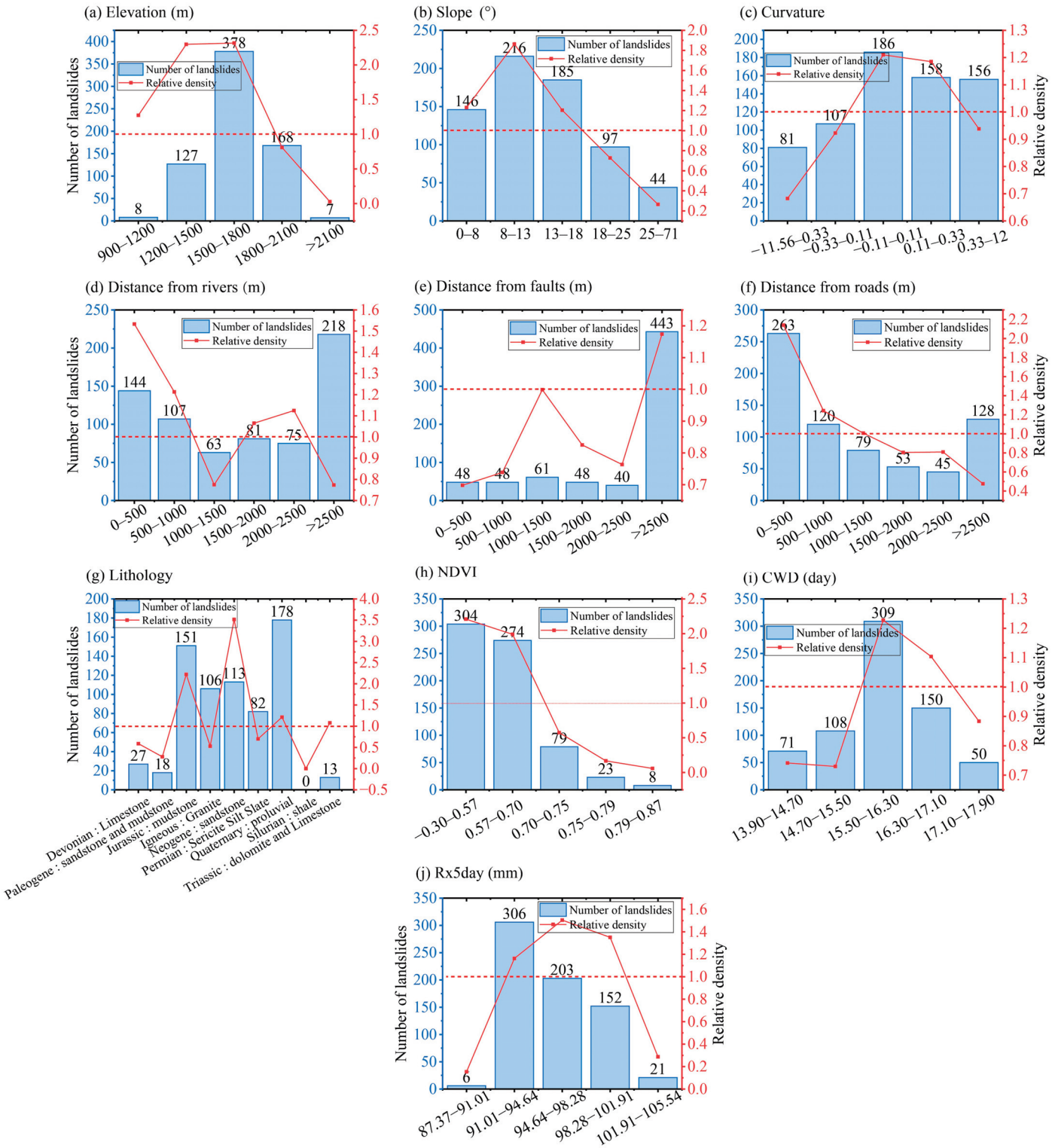


Figure 6. Number and relative density of landslides in each predictor bin. The bar graph indicates the number of landslides in each classification, and the line graph denotes the relative density. The relative density is defined as the ratio of the landslide point proportion in a classification to the area proportion of that classification within the study region. A relative density greater than one, denoted by the dotted lines in the figures, indicates that landslides are concentrated in that classification.

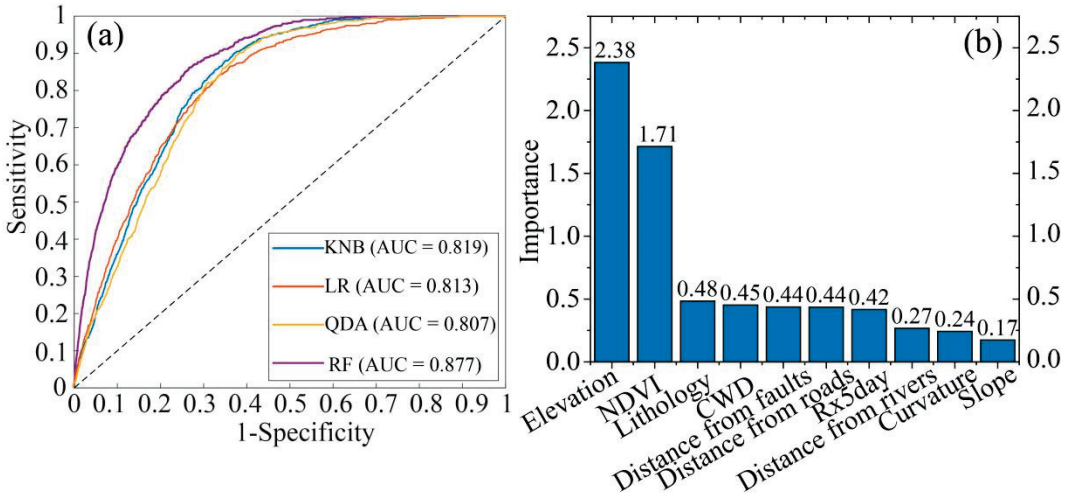


Figure 7. The ROC curves of four models and the importance score of the ten predictors from the RF model. (**a**) ROC curves of all models. The AUC (Area Under the Curve) value reflects model performance, with higher values indicating better predictive accuracy. The ROC curve shows the trade-off between the rate of correctly predicted landslides (Sensitivity) and the rate of falsely predicting landslides where none occurred (1-Specificity). (**b**) The importance score assigned by the Random Forest (RF) model represents the relative contribution of each factor to landslide susceptibility assessment. A higher value indicates a greater influence of the corresponding factor on landslide occurrence.

Table 4. Confusion matrix of the four models. 0 indicates non-landslide points, 1 indicates landslide points, 0-0 indicates correct prediction of non-landslide points, 0-1 indicates incorrect prediction of non-landslide points, 1-0 indicates incorrect prediction of landslide points, and 1-1 indicates correct prediction of landslide points.

RF	0	1
0	4,589,321	1,725,534
1	93	595
QDA	0	1
0	4,227,111	2,087,744
1	117	571
LR	0	1
0	4,546,675	1,768,180
1	168	520
KNB	0	1
0	4,212,191	2,102,664
1	98	590

The landslide susceptibility maps obtained from the four models are classified into five categories using Natural Breaks Classification, and the area of each susceptibility zone is determined (Figure 8). The RF model has a smaller percentage of very high susceptibility zones, indicating a higher accuracy in identified high-susceptibility areas. For regions with NDVI higher than 0.57, the areas with high or very high landslide susceptibility account for 62%. Most regions close to rivers or roads (<500 m) are also threatened by high or very high susceptibility.

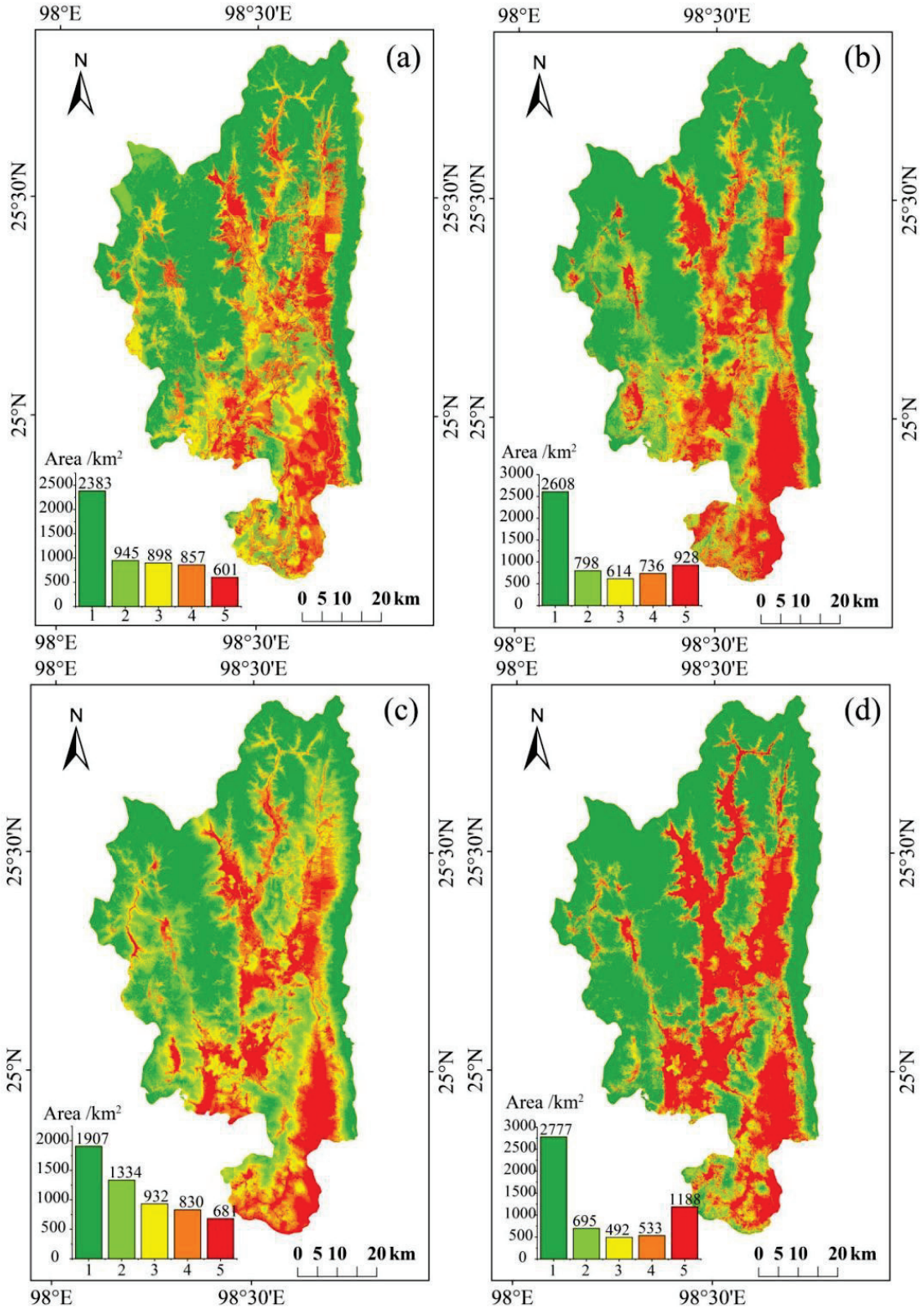


Figure 8. Landslide susceptibility maps generated by: (**a**) RF; (**b**) QDA; (**c**) LR; and (**d**) KNB. Classification thresholds were determined using the natural breaks method. The core of the natural breakpoint method involves grouping data. This is completed in order to minimize differences within and maximize differences between groups. The horizontal axis indicates susceptibility levels: 1 (very low), 2 (low), 3 (medium), 4 (high), and 5 (very high). The larger the value, the greater the susceptibility. The bar chart shows the total area of the partition. The smaller the high-risk area, the more advantageous the model is at identifying it.

The lithology statistics reveal that the Quaternary stratigraphy in Tengchong has a wide distribution and is the primary zone for high or very high landslide susceptibility. The high susceptibility area accounts for 51.3% of the total area in the region. In addition, although the Neogene stratigraphy covers a small area, the distribution of landslide susceptibility shows a clear centralization trend, with high and very high susceptibility occupying 74.9% of the area (Table 5). Collectively, we identified that areas with high landslide susceptibility are primarily associated with zones of low vegetation cover, alluvial deposits, or Neogene sandstone and are also in close proximity to rivers or roads.

Table 5. Statistics of landslide susceptibility for different ranges of NDVI, river distance, road distance, and lithology.

Factors	Range	Land Area/km^2	Very Low Susceptibility	Low Susceptibility	Moderate Susceptibility	High Susceptibility	Very High Susceptibility
NDVI	>0.57	4521.8	52.6%	18.0%	13.2%	10.4%	5.8%
	≤0.57	1162.2	0.3%	11.4%	25.8%	36.3%	26.2%
Distance from rivers	>500 m	4908.1	45.8%	16.9%	13.7%	14.0%	9.5%
	≤500 m	775.9	17.3%	15.0%	28.8%	26.0%	12.9%
Distance from roads	>500 m	4668.7	48.1%	17.4%	13.8%	12.8%	7.9%
	≤500 m	1015.3	13.2%	13.0%	25.0%	29.0%	19.7%
Lithology	Quaternary: alluvial deposits	1205.6	6.9%	15.2%	26.7%	31.2%	20.1%
	Neogene: sandstone	265.7	0.5%	7.0%	17.6%	37.5%	37.4%
	Others	4212.7	69.5%	13.7%	9.2%	4.8%	2.8%

3.3. Assessment of Landslide Risk

By integrating Tengchong's landslide susceptibility with socio-economic factors (e.g., distance to buildings, population density, distance to roads) along with the judgment matrix in Table 6, we computed the CR to be 0.019. Since this value is less than 0.1, it indicates that the landslide risk distribution map for Tengchong is successfully derived based on the calculated weights after feasibility testing. The population density is highly consistent with the construction area (Figure 9a,b). The weights for each layer obtained from the relative importance matrix and the AHP are listed in Table 6. With the weights obtained from Table 6, the landslide susceptibility, population density, distance to roads, and distance to buildings layers were multiplied by their corresponding weights, thereby producing the landslide risk layer.

Table 6. Relative importance matrix and weight of AHP for landslide risk assessment.

	Distance from the Roads	Distance from Buildings	Population Density	Landslide Susceptibility	Weight
Distance from the roads	1	1/2	1/4	1/5	0.08
Distance from buildings	2	1	1/3	1/4	0.14
Population density	4	3	1	1/2	0.23
Landslide susceptibility	5	4	2	1	0.55

The spatial distribution of landslide risk shows a negative correlation with area size, such that higher risk levels concentrate in progressively smaller zones (Figure 9c). Area, population, landslide count and density, and simulated landslide susceptibility and risk are averaged for every township (Table 7). The results indicate that the mean values of landslide susceptibility and risk are highest in Heshun, Tuantian, Wuhe, and Qingshui townships, whereas Houqiao, Mingguang, Zhonghe, and Mazhan townships exhibit relatively lower susceptibility and risk. Our simulation is in line with the number of recorded landslides per unit area. In detail, ~0.32 landslide per km^2 is observed for the four cities with the highest susceptibility, while ~0.04 landslide per km^2 is observed for the four cities with the lowest susceptibility. The Heshun township has a very high population density and landslide susceptibility, which collectively make it have the highest risk. Therefore, it is

concluded that Heshun and Tuantian townships face a more severe threat from landslides, necessitating the enhancement of landslide early-warning mechanisms in these areas.

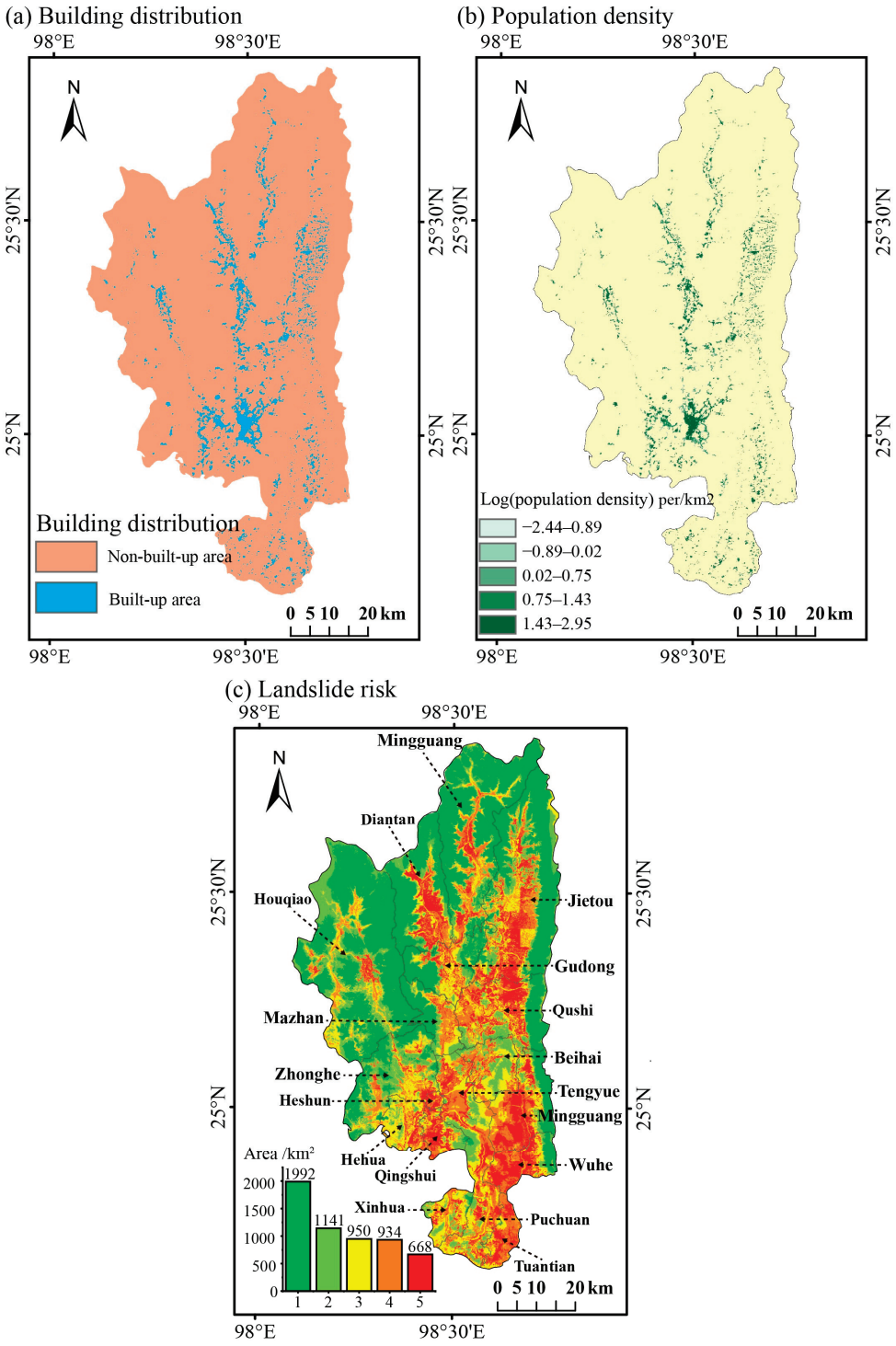


Figure 9. Distribution of population density and building areas resulted in landslide risk. (**a**) Building distribution shows the built-up and the non-built-up areas. (**b**) the logarithm of the population density. (**c**) Landslide risk based on RF and AHP assessment. Landslide risk classification methodology is consistent with landslide susceptibility classification methodology. The horizontal axis indicates risk levels: 1 (very low), 2 (low), 3 (medium), 4 (high), and 5 (very high). The larger the value, the greater the risk.

Table 7. Area, population density, landslide count and density, and simulated landslide susceptibility and risk for every township.

Townships	Land Area (km^2)	Population Density (Count/km^2)	Landslides and Density (Count/km^2)	Mean Landslide Susceptibility	Mean Landslide Risk
Heshun	17.88	2656	3 (0.17)	0.66	0.58
Tuantian	100.88	2005	27 (0.27)	0.62	0.55
Wuhe	166.39	1988	90 (0.54)	0.57	0.50
Qingshui	94.99	1217	28 (0.29)	0.56	0.51
Mangbang	298.23	1061	100 (0.34)	0.51	0.47
Puchuan	167.94	1262	61 (0.36)	0.48	0.47
Tengyue	268.34	4525	30 (0.11)	0.47	0.46
Hehua	129.80	1790	31 (0.24)	0.46	0.45
Qushi	371.79	915	51 (0.14)	0.44	0.42
Xinhua	117.11	1295	33 (0.28)	0.41	0.43
Beihai	178.91	807	22 (0.12)	0.40	0.41
Jietou	817.77	880	92 (0.11)	0.31	0.33
Diantan	398.58	566	23 (0.06)	0.29	0.32
Gudong	239.15	1979	15 (0.06)	0.28	0.32
Mazhan	164.83	1020	8 (0.05)	0.27	0.31
Zhonghe	394.72	930	15 (0.04)	0.24	0.32
Mingguang	705.54	536	18 (0.03)	0.19	0.27
Houqiao	1072.08	245	41 (0.04)	0.15	0.24

4. Discussion

Geospatial analysis using machine learning models (e.g., Random Forest) demonstrates that the relative importance of key drivers (slope angle, curvature, land cover) exhibits spatially differentiated patterns across study regions, as revealed by permutation feature importance metrics [11,12,18,60]. For Tengchong, elevation, NDVI, lithology, and CWDs are the most critical factors for landslide assessment (Figure 7a). It is evident that these factors, which represent topographic, land surface, lithological, and climatic dimensions, respectively, underscore the multifaceted nature of landslide susceptibility. This highlights that such geological hazards are governed by complex interactions among diverse environmental variables.

Topography is usually the primary factor controlling landslides. Landslide occurrence declines with increasing altitude as colluvial deposits become thinner in mountainous areas [61–63]. Our results support this idea that landslides are concentrated in lowland regions of Tengchong with elevation less than 1800 m (Figure 6a). Theoretically, slope is also among the most critical factors, with a strong correlation existing between slope angle and landslide susceptibility [18]. However, this characteristic seems less pronounced in Tengchong, where landslides are more concentrated in low-slope areas (Figure 6b). Moreover, in the RF model, the slope factor demonstrates less prominence in susceptibility assessments, ranking last among all evaluated factors. The generally low slope gradient (<25°) in Tengchong might lead to this weaker impact on landslides.

Areas with denser vegetation are usually less susceptible to landslides and vice versa [64]. This aligns with our findings in this study that landslides in Tengchong are more concentrated in vegetation sparse areas with NDVI values below 0.57 (Figure 5h). As vegetation cover increases, the relative density of landslides decreases. Vegetation has a unique soil-stabilizing effect. Therefore, when vegetation is sparse, looser soils can provide essential land surface material conditions for landslides. Vegetation in southwestern China has been increasing due to the CO_2 fertilization effect and land-use management [65]. As such, we expect that vegetation restoration may reduce landslide risk to some extent in Tengchong.

Human activities in this assessment are mainly reflected in the distance from roads. Landslides in Tengchong are relatively concentrated within 500 m of roads (Figure 6f; Table 5). On the one hand, owing to Tengchong's average elevation at approximately

2 km and its rugged hilly-mountainous terrain, the accumulation of rock and soil materials during road construction, coupled with excavation activities, compromises the stability of the original rock or soil mass. This process is also seen in the Loess Plateau [42]. On the other hand, during the operational phase, dynamic loads and vibrations exerted by road traffic further induce stress on the surrounding geotechnical medium.

Landslides in Tengchong are concentrated in areas where the rock strata consist of Neogene sandstone and Quaternary alluvial deposits. Whether it is sandstone or alluvial deposits, the mechanism for landslide occurrence in these regions mainly lies in the instability of the rock-soil structure and high permeability [7]. These characteristics make the areas with sandstone and alluvial deposits highly susceptible to disturbance and prone to landslides. The distance from faults is not significantly correlated with landslides in Tengchong. First, we infer most of the landslides in Tengchong may not have been triggered by fault activity, or that the existing faults have not shown significant movement in recent years. Second, the fault structure data are not detailed enough, or other variables reflecting tectonic activity should be proposed. Third, the current landslide survey may exhibit spatial sampling bias in that more landslides threatening human populations have been documented, whereas those in remote mountainous regions have been overlooked. It may partially explain why recorded landslides are closer to roads/rivers than to faults (Figure 6e).

Globally, regional studies are necessary because landslides are sensitive to different factors in different regions. In southern Colombia, where the climate is similar to that of Tengchong, researchers used RF to evaluate landslide susceptibility. They found that elevation and soil silt content had the most significant impact, consistent with our results [66]. In this study, we found that landslides are concentrated in low-elevation areas, whereas in Sikkim Himalaya, India, at the same latitude, landslides are more likely to occur in high-elevation regions [67]. This may be because, in India, the monsoon winds encounter the Himalayas and are forced upwards by the terrain, generating significant rainfall and triggering landslides. In another region of Yunnan, researchers found that among their 12 evaluation factors, slope has the highest importance, making it the most significant controlling factor for landslides in Yulong County [68]. Their result is different from that of our study; highly undulating terrain (more than 4000 m) may be the reason, according to the elevation map they showed. These comparisons reveal that although certain controlling factors (e.g., elevation) are universally important, the landslide-environment relationship is highly region-specific. This relationship is primarily shaped by unique local geological and climatic conditions.

There are some uncertainties and implications that should be noted for future studies. First, the current landslide survey may be subject to spatial sampling bias. The dataset primarily documents landslides that pose a threat to human populations, which may lead to the omission of events in remote mountainous regions. This oversight may explain why distance from faults—a key geological factor—exhibits a weak correlation with landslide locations (Figure 6e), especially given Tengchong's high frequency of geological activity. Additionally, the observed tendency of landslides to occur in low-slope areas (Figure 6b) is likely attributed to this same sampling bias. Second, as the study area is relatively small, variables may have spatial autocorrelation. It may therefore inflate accuracy metrics because nearby pixels share similar environmental conditions. For the extreme rainfall indicator, although the results showed their importance in landslide susceptibility assessment, their minimal variation may affect the accuracy of their impact. Third, future research could integrate field survey data with InSAR data to achieve more precise landslide identification, thereby addressing the temporal constraints of remote sensing data and the spatial coverage limitations of field surveys. Finally, the occurrence of landslides is influenced by a variety

of factors that interact with each other in a complex way. Machine learning techniques may capture these complex interactions but are difficult to explain. Sensitivity analysis of factors should be implemented to diagnose the landslide susceptibility.

5. Conclusions

This study presents a comprehensive assessment of landslide susceptibility and risk in Tengchong of southwest China by integrating multi-source influencing factors and using four machine learning algorithms and the AHP. The RF model demonstrates the highest predictive reliability (AUC = 0.88), enabling the identification of highly susceptible zones. Notably, the analysis identifies elevation as the primary controlling factor, followed by NDVI, lithology, and CWDs, underscoring the critical roles of vegetation cover and precipitation patterns in modulating landslide susceptibility. Empirical results show that landslides predominantly occur in areas with low NDVI values (<0.57), suggesting that enhanced vegetation coverage may mitigate landslide hazards. Spatial distribution analysis reveals that high-susceptibility zones are concentrated within 500 m of river networks and road infrastructures, as well as in regions underlain by alluvial deposits or Neogene sandstone formations. The integrated risk map, developed by fusing machine learning predictions with AHP-based socio-economic analysis, pinpoints Heshun and Tuantian townships as high-risk hotspots. These findings provide a scientific basis for hierarchical disaster management and emergency resource optimization in Tengchong and also offer a typical case for the risk assessment framework in global mountainous areas prone to landslides.

Author Contributions: Conceptualization, W.C.; Data curation, C.L., Z.Q., K.Y., S.Z., L.Y., Y.G., Z.Z. and Q.G.; Formal analysis, C.L.; Funding acquisition, C.L. and W.C.; Investigation, W.C.; Methodology, W.C. and J.L.; Visualization, Z.Q. and K.Y.; Writing—original draft, C.L., Z.Q. and W.C.; Writing—review and editing, C.L., Z.Q., W.C., J.L., S.Z., L.Y., Y.G., Z.Z. and Q.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (No. 42307556) and the China Geological Survey (DD20230101205).

Data Availability Statement: We thank the Geospatial Data Cloud Platform (http://www.gscloud.cn), the United States Geological Survey (USGS, https://www.usgs.gov), the Geoscientific Data and Discovery Publishing System (https://www.ngac.cn), the National Catalogue Service for Geographic Information (www.webmap.cn), GHSL (Global Human Settlement Layer, https://human-settlement.emergency.copernicus.eu), and Jielin Zhang for making their data freely accessible. The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Das, I.; Stein, A.; Kerle, N.; Dadhwal, V.K. Landslide susceptibility mapping along road corridors in the Indian Himalayas using Bayesian logistic regression models. *Geomorphology* **2012**, *179*, 116–125. [CrossRef]
2. Drid, I.; Achour, Y.; Zighmi, K.; Nouioua, T.; Khababa, O.; Rahmani, S.; Ouadfel, O.; Atoui, B. Designing gully erosion susceptibility maps (GESM) in the Algerian Eastern Tell: A case study of the K'sob River watershed. *Arab. J. Geosci.* **2022**, *15*, 1264. [CrossRef]
3. Niu, C.; Zhang, H.; Liu, W.; Li, R.; Hu, T. Using a fully polarimetric SAR to detect landslide in complex surroundings: Case study of 2015 Shenzhen landslide. *ISPRS J. Photogramm. Remote Sens.* **2021**, *174*, 56–67. [CrossRef]
4. Ministry of Natural Resources of the People's Republic of China. 2024 China Natural Resources Bulletin. Available online: https://gi.mnr.gov.cn/202503/t20250314_2881937.html (accessed on 19 March 2025).

5. Yunnan Provincial Department of Emergency Management. Disaster Situation in Yunnan Province in 2024. Available online: https://yjglt.yn.gov.cn/html/2025/tjfx_0110/4029810.html (accessed on 21 March 2025).
6. Gui, Y.; Wang, J.; Li, X. A Novel Deep Learning Model for Landslide Susceptibility Mapping. *Remote Sens.* **2024**, *16*, 1234.
7. Song, M.; Zhang, W.; Zou, Y.; Chen, J. Numerical Simulation of Uniaxial Compressive Strength and Failure Characteristics of Non-Uniform Water-Bearing Sandstone. *Materials* **2023**, *16*, 6396. [CrossRef] [PubMed]
8. Wang, M.; Zhou, X.; Wu, Y.; Zhuo, L.; Xu, H.; Liu, Y. Hydrogeochemistry of trace elements in geothermal waters from the Rehai geothermal system in Yunnan of China, a magma-indirectly related geothermal system. *Geothermics* **2024**, *124*, 103134. [CrossRef]
9. Li, C.; Liu, Y.; Lai, S.; Wang, D.; Ho, X.; Liu, Q. Landslide susceptibility analysis based on the coupling model of logistic regression and support vector machine. *J. Nat. Disasters* **2024**, *33*, 75–86.
10. Xu, W.; Yang, H.; Zhang, J.; Zhou, H.; Lan, L.; Li, X.; Chen, G.; Chen, J. Evaluation of Landslide Geological Hazard Susceptibility Based on GIS and BP Neural Network Algorithm. *Land Resour. Herald* **2025**, *22*, 98–107.
11. Yan, H.; Deying, L.; Songsong, M.; Yiqing, S. Landslide susceptibility evaluation in Badong County based on weights of evidence method. *Bull. Geol. Sci. Technol.* **2020**, *39*, 187–194.
12. Fan, W.; Wei, X.; Cao, Y.; Zheng, B. Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. *J. Mt. Sci.* **2017**, *14*, 906–925. [CrossRef]
13. Li, T.; Tian, Y.; Wu, L.; Liu, L. Landslide Susceptibility Mapping Using Random Forest. *Geogr. Geo-Inf. Sci.* **2014**, *30*, 25–30+2.
14. Yuan, X.; Liu, J.; Huang, L.; Li, X. Assessment of Landslide Susceptibility in Xinyuan County Based on Machine Learning Models. *Sci. Technol. Eng.* **2025**, *25*, 1815–1826.
15. Zhang, Y.; Song, W. Spatial Susceptibility Evaluation of Landslide in Dongchuan District of Kunming Based on BP Neural Network and Decision Tree. *Territ. Nat. Resour. Study* **2023**, 67–70. [CrossRef]
16. Sun, C.; Tie, Y.; Ning, Z.; Xu, W.; Xiong, X. Landslide susceptibility mapping in Xide County, Sichuan Province based on frequency ratio-support vector machine coupling model. *Sediment. Geol. Tethyan Geol.* **2024**, *44*, 547–559.
17. Chen, W.; Yang, Z. Landslide susceptibility modeling using bivariate statistical-based logistic regression, naïve Bayes, and alternating decision tree models. *Bull. Eng. Geol. Environ.* **2023**, *82*, 190. [CrossRef]
18. Ohlmacher, G.C.; Davis, J.C. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Eng. Geol.* **2003**, *69*, 331–343. [CrossRef]
19. Ali, S.A.; Parvin, F.; Pham, Q.B.; Khedher, K.M.; Dehbozorgi, M.; Rabby, Y.W.; Anh, D.T.; Nguyen, D.H. An ensemble random forest tree with SVM, ANN, NBT, and LMT for landslide susceptibility mapping in the Rangit River watershed, India. *Nat. Hazards* **2022**, *113*, 1601–1633. [CrossRef]
20. Bhattacharya, S.; Ali, T.; Chakravortti, S.; Pal, T.; Majee, B.K.; Mondal, A.; Pande, C.B.; Bilal, M.; Rahman, M.T.; Chakrabortty, R. Application of Machine Learning and Deep Learning Algorithms for Landslide Susceptibility Assessment in Landslide Prone Himalayan Region. *Earth Syst. Environ.* **2024**, *9*, 1427–1445. [CrossRef]
21. Chang, K.-T.; Merghadi, A.; Yunus, A.P.; Pham, B.T.; Dou, J. Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Sci. Rep.* **2019**, *9*, 12296. [CrossRef]
22. Zhan, Y.; Zhang, Y.; Zhang, J.; Xu, J.; Chen, H.; Liu, G.; Wan, Z. Risk assessment of land subsidence in Shanghai municipality based on AHP and EWM. *Sci. Rep.* **2025**, *15*, 7339. [CrossRef]
23. Burayu, D.G.; Karuppannan, S.; Shuniye, G. Identifying flood vulnerable and risk areas using the integration of analytical hierarchy process (AHP), GIS, and remote sensing: A case study of southern Oromia region. *Urban Clim.* **2023**, *51*, 101640. [CrossRef]
24. Skilodimou, H.D.; Bathrellos, G.D.; Chousianitis, K.; Youssef, A.M.; Pradhan, B. Multi-hazard assessment modeling via multi-criteria analysis and GIS: A case study. *Environ. Earth Sci.* **2019**, *78*, 47. [CrossRef]
25. Yalcin Kavus, B.; Taskin, A. Assessment of landslides induced by earthquake risk of Istanbul: A comprehensive study utilizing an integrated DFS-AHP and DFS-EDAS approach. *Soil Dyn. Earthq. Eng.* **2025**, *191*, 109285. [CrossRef]
26. Panchal, S.; Shrivastava, A.K. Landslide Hazard Assessment Using Analytic Hierarchy Process (AHP): A Case Study of National Highway 5 in India. *Ain Shams Eng. J.* **2022**, *13*, 101626. [CrossRef]
27. Cao, Y.; Guo, W.; Wu, Y.; Li, L.; Zhang, Y.; Lan, H. An hourly shallow landslide warning model developed by combining automatic landslide spatial susceptibility and temporal rainfall threshold predictions. *J. Mt. Sci.* **2022**, *19*, 3370–3387. [CrossRef]
28. Ciccarese, G.; Mulas, M.; Corsini, A. Combining spatial modelling and regionalization of rainfall thresholds for debris flows hazard mapping in the Emilia-Romagna Apennines (Italy). *Landslides* **2021**, *18*, 3513–3529. [CrossRef]
29. Abraham, M.T.; Satyam, N.; Kushal, S.; Rosi, A.; Pradhan, B.; Segoni, S. Rainfall Threshold Estimation and Landslide Forecasting for Kalimpong, India Using SIGMA Model. *Water* **2020**, *12*, 1195. [CrossRef]
30. Martelloni, G.; Segoni, S.; Fanti, R.; Catani, F. Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* **2012**, *9*, 485–495. [CrossRef]
31. Melillo, M.; Brunetti, M.T.; Peruccacci, S.; Gariano, S.L.; Roccati, A.; Guzzetti, F. A tool for the automatic calculation of rainfall thresholds for landslide occurrence. *Environ. Model. Softw.* **2018**, *105*, 230–243. [CrossRef]

32. Abraham, M.T.; Satyam, N.; Lokesh, R.; Pradhan, B.; Alamri, A. Factors Affecting Landslide Susceptibility Mapping: Assessing the Influence of Different Machine Learning Approaches, Sampling Strategies and Data Splitting. *Land* **2021**, *10*, 989. [CrossRef]
33. Huang, F.; Yang, Y.; Jiang, B.; Chang, Z.; Zhou, C.; Jiang, S.-H.; Huang, J.; Catani, F.; Yu, C. Effects of different division methods of landslide susceptibility levels on regional landslide susceptibility mapping. *Bull. Eng. Geol. Environ.* **2025**, *84*, 276. [CrossRef]
34. Ke, C.; Sun, P.; Zhang, S.; Li, R.; Sang, K. Influences of non-landslide sampling strategies on landslide susceptibility mapping: A case of Tianshui city, Northwest of China. *Bull. Eng. Geol. Environ.* **2025**, *84*, 123. [CrossRef]
35. Qin, S.; Wang, J.; Li, X. Analysis of Rainfall Infiltration Law and Stability of Unsaturated Loess Slope. *Rock Soil Mech.* **2011**, *32*, 1902–1906.
36. Zhang, J.; Liu, B.; Ren, S.; Han, W.; Ding, Y.; Peng, S. A 4 Km Daily Gridded Meteorological Dataset for China from 2000 to 2020. *Sci. Data* **2024**, *11*, 1230. [CrossRef]
37. Chen, W.; Chen, X.; Peng, J.; Panahi, M.; Lee, S. Landslide susceptibility modeling based on ANFIS with teaching-learning-based optimization and Satin bowerbird optimizer. *Geosci. Front.* **2021**, *12*, 93–107. [CrossRef]
38. Chen, C.-Y.; Chang, J.-M. Landslide dam formation susceptibility analysis based on geomorphic features. *Landslides* **2016**, *13*, 1019–1033. [CrossRef]
39. Durán Zuazo, V.H.; Rodríguez Pleguezuelo, C.R.; Francia Martínez, J.R.; Cárceles Rodríguez, B.; Martínez Raya, A.; Pérez Galindo, P. Harvest intensity of aromatic shrubs vs. soil erosion: An equilibrium for sustainable agriculture (SE Spain). *CATENA* **2008**, *73*, 107–116. [CrossRef]
40. Niraj, K.C.; Gupta, S.K.; Shukla, D.P. A Comparison of Image-Based and Physics-Based Atmospheric Correction Methods for Extracting Snow and Vegetation Cover in Nepal Himalayas Using Landsat 8 OLI Images. *J. Indian Soc. Remote Sens.* **2022**, *50*, 2503–2521. [CrossRef]
41. Zhang, Y.; Su, S.; Li, P. Numerical Simulation on Mechanism of Fault-Controlled Liujiapo Landslide. *J. Eng. Geol.* **2015**, *23*, 1127–1137.
42. Zhang, M.; Li, T. Triggering Factors and Forming Mechanism of Loess Landslides. *J. Eng. Geol.* **2011**, *19*, 530–540.
43. Yang, B.; Miao, H.; Ma, C. Landslide susceptibility assessment of Yongjia County based on multi-model coupling. *J. Henan Univ. Urban Constr.* **2024**, *33*, 91–99+127. [CrossRef]
44. Fang, K.; Wu, J.; Zhu, J.; Xie, B. A Review of Technologies on Random Forests. *J. Stat. Inf.* **2011**, *26*, 32–38.
45. Hu, Y.; Du, H. Research on color recognition based on quadratic discriminant analysis. *KEJIFENG* **2016**, 17–18. [CrossRef]
46. Wang, D.; Han, L. Research on Intelligent Operation and Maintenance of Subway Fan and Water Pump Based on Logistic Regression Classification Algorithm. *Intell. Build. Smart City* **2024**, 157–159. [CrossRef]
47. Zuo, J. Research on Automatic Identification of Abnormal Nodes in Wireless Sensor Networks Based on Naive Bayesian Networks. *Autom. Instrum.* **2025**, *40*, 124–128.
48. Al-Kordi, H.; Al-Amri, A.; Raju, G. Landslide susceptibility mapping using geospatial, analytical hierarchy process (AHP), and binary logistic regression (BLR) techniques—A study of Wadi Habban Basin, Shabwah, Yemen. *Results Earth Sci.* **2025**, *3*, 100103. [CrossRef]
49. Olson, D.L. (Ed.) The Analytic Hierarchy Process. In *Decision Aids for Selection Problems*; Springer: New York, NY, USA, 1996; pp. 49–68.
50. Aksoy, H. Determination of landslide susceptibility with Analytic Hierarchy Process (AHP) and the role of forest ecosystem services on landslide susceptibility. *Environ. Monit. Assess.* **2023**, *195*, 1525. [CrossRef]
51. Chattaraj, D.; Paul, B.; Sarkar, S. Integrated Multi-parametric Analytic Hierarchy Process (AHP) and Geographic Information System (GIS) based Spatial modelling for Flood and Water logging Susceptibility Mapping: A case study of English Bazar Municipality of Malda, West Bengal, India. *Nat. Hazards Earth Syst. Sci. Discuss.* **2021**, *2021*, 1–20. [CrossRef]
52. Demirel, B.; Yildirim, E.; Can, E. GIS-based landslide susceptibility mapping using AHP, FMEA, and Pareto systematic analysis in central Yalova, Türkiye. *Eng. Sci. Technol. Int. J.* **2025**, *64*, 102013. [CrossRef]
53. Swain, K.C.; Singha, C.; Nayak, L. Flood Susceptibility Mapping through the GIS-AHP Technique Using the Cloud. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 720. [CrossRef]
54. Kılıç, G.; Altıntaş, O. A Comparative Analysis of Classical AHP, Fuzzy AHP, and Z-Fuzzy AHP Methods for Flood Risk Assessment in Büyükçekmece District. In *Intelligent and Fuzzy Systems*; Kahraman, C., Cevik Onar, S., Cebi, S., Oztaysi, B., Tolga, A.C., Ucal Sari, I., Eds.; Springer: Cham, Switzerland, 2024; pp. 120–127.
55. Batur, M.; Karaman, H.; Alkan, R.M.; Ozener, H. Mapping Social Vulnerability to Nuclear Accidents: A Comparative Framework Using BWM and AHP Methods. *Appl. Spat. Anal. Policy* **2025**, *18*, 45. [CrossRef]
56. Saaty, T.L. How to make a decision: The analytic hierarchy process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [CrossRef]
57. Perera, E.N.C.; Jayawardana, D.T.; Jayasinghe, P. A Rainfall Intensity-Duration Threshold for Mass Movement in Badulla, Sri Lanka. *GEP* **2017**, *5*, 135–152. [CrossRef]
58. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988; pp. 79–81.

59. Schober, P.; Boer, C.; Schwarte, L.A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [CrossRef]
60. Althuwaynee, O.F.; Pradhan, B.; Park, H.-J.; Lee, J.H. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *CATENA* **2014**, *114*, 21–36. [CrossRef]
61. Nakileza, B.R.; Nedala, S. Topographic influence on landslides characteristics and implication for risk management in upper Manafwa catchment, Mt Elgon Uganda. *Geoenviron. Disasters* **2020**, *7*, 27. [CrossRef]
62. Yi, Y.; Zhang, Z.; Zhang, W.; Xu, Q.; Deng, C.; Li, Q. GIS-based earthquake-triggered-landslide susceptibility mapping with an integrated weighted index model in Jiuzhaigou region of Sichuan Province, China. *Nat. Hazards Earth Syst. Sci.* **2019**, *19*, 1973–1988. [CrossRef]
63. Herrera-Coy, M.C.; Calderón, L.P.; Herrera-Pérez, I.L.; Bravo-López, P.E.; Conoscenti, C.; Delgado, J.; Sánchez-Gómez, M.; Fernández, T. Landslide Susceptibility Analysis on the Vicinity of Bogotá-Villavicencio Road (Eastern Cordillera of the Colombian Andes). *Remote Sens.* **2023**, *15*, 3870. [CrossRef]
64. Niraj, K.C.; Singh, A.; Shukla, D.P. Effect of the Normalized Difference Vegetation Index (NDVI) on GIS-Enabled Bivariate and Multivariate Statistical Models for Landslide Susceptibility Mapping. *J. Indian Soc. Remote Sens.* **2023**, *51*, 1739–1756. [CrossRef]
65. Chen, C.; Park, T.; Wang, X.; Piao, S.; Xu, B.; Chaturvedi, R.K.; Fuchs, R.; Brovkin, V.; Ciais, P.; Fensholt, R.; et al. China and India lead in greening of the world through land-use management. *Nat. Sustain.* **2019**, *2*, 122–129. [CrossRef] [PubMed]
66. Correa-Muñoz, N.A.; Martinez-Martinez, L.J.; Murillo-Feo, C.A. Landslide Susceptibility Mapping Using Supervised Learning Methods—Case Study: Southwestern Colombia. In *Geopedology: An Integration of Geomorphology and Pedology for Soil and Landscape Studies*; Zinck, J.A., Metternicht, G., del Valle, H.F., Angelini, M., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 315–335.
67. Mandal, S.P.; Chakrabarty, A.; Maity, P. Comparative Evaluation of Information Value and Frequency Ratio in Landslide Susceptibility Analysis along National Highways of Sikkim Himalaya. *Spat. Inf. Res.* **2018**, *26*, 127–141. [CrossRef]
68. Qin, Y.; Zhao, Z.; Zhou, D.; Chang, K.; Mou, Q.; Yang, Y.; Hu, Y. Landslide Susceptibility Assessment in Yulong County Using Contribution Degree Clustering Method and Stacking Ensemble Coupled Model Based on Certainty Factor. *Remote Sens.* **2024**, *16*, 3582. [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.

land

MDPI

Article

Bridging Heritage Systems: Multi-Scale Spatial Coupling Between Tangible and Intangible Cultural Heritage in China Using Hierarchical Bayesian Model and Causal Inference

Yuxi Liu, Xinyu Du, Yu Bai, Qibing Chen * and Shiliang Liu

College of Landscape Architecture, Sichuan Agricultural University, Chengdu 611130, China
* Correspondence: cqb@sicau.edu.cn

Abstract

Cultural heritage systems play a crucial role in decoding human–environment interactions and social evolution. This study aims to reveal the spatial coupling characteristics of tangible and intangible cultural heritage in China, as well as the heterogeneity of their driving mechanisms. After quantifying heritage coupling at three geographic scales, we integrated a hierarchical Bayesian model with a hybrid causal inference framework to identify the correlations, causal effects, and heterogeneity of the driving factors. The empirical results indicate the following: (1) The coupling patterns exhibit scale dependence. The proportion of strongly coupled areas decreases from the prefecture-level scale to the provincial scale but increases at the cultural–geographical unit scale. This suggests that China's cultural system has a cohesive effect that transcends administrative boundaries. (2) The hierarchical Bayesian model identifies the significant effects of mean annual temperature, population density, GDP–population interaction, transport–hydrological network interaction, and industrial structure. Effect strengths generally peak at the prefecture-level scale and decrease at the provincial scale. (3) Causal inference estimates the causal effects of mean annual temperature, transport–hydrological network interaction, mean annual precipitation, and water network density on coupling. (4) Heterogeneity tests reveal that the positive causal effect of transport–hydrological network interaction and the negative causal effect of mean annual precipitation are significant only in low-temperature regions. By integrating hierarchical modeling with causal verification, this study elucidates the mechanisms underlying heritage coupling. This provides a scientific basis for understanding the spatial patterns of cultural heritage systems and formulating differentiated conservation policies.

Keywords: human–land relations; heritage systems; spatial coupling; machine learning; heterogeneity analysis

1. Introduction

As carriers of human history and civilization, cultural heritage embodies the collective memory of nations and serves as a bridge connecting the past and present. China's symbiotic system of tangible (TCH) and intangible (ICH) cultural heritage provides valuable insights into cultural evolution on a global scale [1]. TCH represents the material imprints that document historical processes, while ICH perpetuates cultural traditions in living practice. The Convention Concerning the Protection of the World Cultural and Natural Heritage and the Convention for the Safeguarding of the Intangible Cultural Heritage have, respectively, established international normative frameworks for TCH and ICH. However,

their separate structures have led to persistent integration challenges in policy-making and management practices [2]. Scholars have increasingly employed perspectives such as the assemblage theory [3] to critically reflect on this binary framework, conceptualizing heritage as a dynamically evolving system composed of material spaces, social actors, institutions, and cultural practices [4]. By examining the spatial interactions and structural logics among elements and incorporating comprehensive systems of culture and nature, this research field has developed the theory of heritage systems, which is characterized by fuzzy boundaries and value associations [5,6].

In this study, heritage systems are defined as multi-level evolutionary systems encompassing material remains, living practices, and spatial environments, which are embedded within broader natural and socio-economic systems. Their systemic boundaries are generated across regional and local scales. TCH and ICH interact through spatial proximity and cultural processes to form interconnected subsystems that constitute complementary relational networks. Despite its limitations, the binary TCH–ICH framework is still employed as an analytical tool due to the methodological and normative advantages of official inventory systems. These inventories provide heritage datasets with clearly defined spatial boundaries, ensuring regional comparability, value assessment criteria, and institutional frameworks, which facilitate the identification of factors influencing coupling patterns within heritage systems. In other words, using the official system serves to test, rather than perpetuate, the binary classification model and analyze the value judgments and cultural processes embedded within this institutional component. The research objective is to reveal the dynamic processes and synergistic relationships obscured by this framework through empirical methods [7], thereby transcending the static classification logic that treats the two heritage types as mutually independent entities. This contributes to understanding the spatial association patterns and driving mechanisms between them, providing a scientific basis for formulating adaptive conservation strategies.

From the perspective of cultural geography, cultural landscapes represent the manifestation of cultural relationships within the material environment [8], and long-term interactions between space and culture provide a theoretical basis for understanding heritage distribution. Within this framework, heritage is conceptualized as cultural elements that aggregate, diffuse and co-locate within specific territories. Their formation processes reflect the ongoing evolution of social practices, local identities and social networks [9]. Previous studies have indicated that, although TCH and ICH are institutionally categorized as distinct types, they often exhibit functionally complementary and spatially intertwined relationships in practice [10]. This interrelation provides a critical theoretical foundation for understanding the coupling patterns between the two heritage types based on heritage pattern analysis. Recent research on the spatial aspects of cultural heritage has become increasingly interdisciplinary [11,12], primarily focusing on distribution patterns and influencing factors. Scholars have commonly employed methods such as kernel density estimation [13], standard deviational ellipses [14], and nearest-neighbor indices [15] to analyze clustering characteristics and spatiotemporal patterns across multiple geographic scales, including the national [16], watershed [17], provincial [18], and municipal levels [19]. Within frameworks that encompass natural [20], social [21], and cultural factors [22], quantitative approaches—such as decision trees [23], geographic detectors [24], and geographically weighted regression [25]—have also become popular for elucidating underlying mechanisms. However, two systematic gaps remain in the field. Firstly, TCH and ICH are frequently considered separately [1], often lacking analysis of their spatial coupling. Only a few studies have examined the coupled interactions between heritage and external systems, such as tourism [26], urbanization, or cultural transmission [27], while the internal coupling between TCH and ICH has received limited attention. Secondly, from

a methodological perspective, traditional statistical approaches struggle to address the high dimensionality, nonlinearity, and heterogeneity inherent [28] to heritage data. While some machine learning models can capture complex patterns, their limited interpretability [29] restricts their ability to provide causal evidence to inform heritage policy. In particular, the driving mechanisms of coupling may vary across spatial scales, posing challenges for cross-scale heterogeneity analyses.

To address these research gaps, this study focuses on the spatial coupling patterns between TCH and ICH in China, as well as the heterogeneity of their driving mechanisms. The research aims to identify the characteristics of heritage association networks in different regions, and to examine how social and natural factors influence coupling patterns. First, an improved co-location quotient (CLQ) theory is used to identify heritage coupling patterns across three geographical scales. Subsequently, three categories of variables are delineated based on the results of a hierarchical Bayesian model (HBM) and causal identification tools. Finally, causal inference methods are applied to obtain estimates of the causal effects of factors and evidence of heterogeneity under optimal grouping. This multi-level framework combines data-driven HBM with logic-driven causal inference, enabling the identification of correlation patterns and model-based causal relationships. It provides rigorous scientific explanations for the operational mechanisms of heritage coupling systems, promoting the transition of heritage revitalization from experience-based to evidence-based decision-making.

2. Materials and Methods

2.1. Research Area

With a history spanning thousands of years, China is located in the eastern part of the Eurasian continent and covers approximately 9.6 million km^2 (Figure 1). Its three-tiered topographical pattern extending from west to east, combined with complex climatic gradients, has shaped diverse geographical units that constitute a natural arena for cultural differentiation and integration among the country's 56 ethnic groups [30]. A total of 5060 national-level cultural relic protection units, ranging from the Neolithic period to the present day, form a comprehensive spectrum of TCH, encompassing ancient architecture, archeological sites, and other typological categories. Additionally, 3610 ICH items are distributed across ten categories, including folk literature, traditional music, and customs. The historical information and human–land relations embedded within these two types of heritage reflect the transnational value of Chinese culture in engineering technology, ceremonial institutions, and ideological frameworks. The spatial distribution of this heritage, characterized by distinct geographical differentiation and clustering patterns, reflects the evolutionary logic of cultural patterns [31] and provides crucial empirical samples for global cultural heritage research. China's cultural heritage protection system represents a significant experiential supplement to the World Heritage Convention, holding substantial practical significance for investigating the coupling characteristics and influencing mechanisms of heritage spatial distribution.

2.2. Data Sources

TCH was determined based on a list of national key cultural relics protection units published by the National Cultural Heritage Administration, and their spatial coordinates were obtained using the Baidu coordinate retrieval system. ICH data were derived from the standardized dataset compiled by Guo Y. et al., organized by the geographical location of the applying regions or institutions [32]. Data on the proportion of the tertiary industry, per capita GDP, and the population share of different ethnic groups were obtained from statistics publicly available on government websites. Ethnic diversity was calcu-

lated using the Shannon–Wiener index [33], which measures social group richness based on the population share of each ethnic group. Provincial-scale digital elevation model (DEM) data (https://www.resdc.cn/data.aspx?DATAID=337, accessed on 11 March 2025) and geological hazard points (https://www.resdc.cn/data.aspx?DATAID=290, accessed on 10 March 2025) were obtained from the Resource and Environmental Science Data Platform. Elevation, aspect, and slope parameters were then extracted from these using ArcGIS Pro 3.1.4. The 1 km mean annual temperature raster was obtained from the China extreme and average temperature index dataset (1961–2012), provided by the National Tibetan Plateau Data Center [34]. The 1 km mean annual precipitation data were derived from the monthly precipitation dataset developed by Peng S. et al. [35] and converted into the appropriate units after being aggregated to annual values. The 30 m resolution Normalized Difference Vegetation Index (NDVI) raster was obtained from the research outputs of Dong J. et al. [36] and Yang J. et al. [37]. The 100 m population density raster was sourced from the Lake–Watershed Science Data Center, National Earth System Science Data Center, National Science & Technology Infrastructure of China [38] (http://lake.geodata.cn, accessed on 7 March 2025). Road and hydrographic data were obtained from the Geospatial Data Cloud (http://www.gscloud.cn, accessed on 12 March 2025) and OpenStreetMap (https://www.openstreetmap.org/, accessed on 12 March 2025), respectively. All variables corresponding to each geographical unit were processed using spatial statistics in ArcGIS Pro to ensure the spatial consistency of the attribute data, thereby supporting subsequent modeling and analysis.

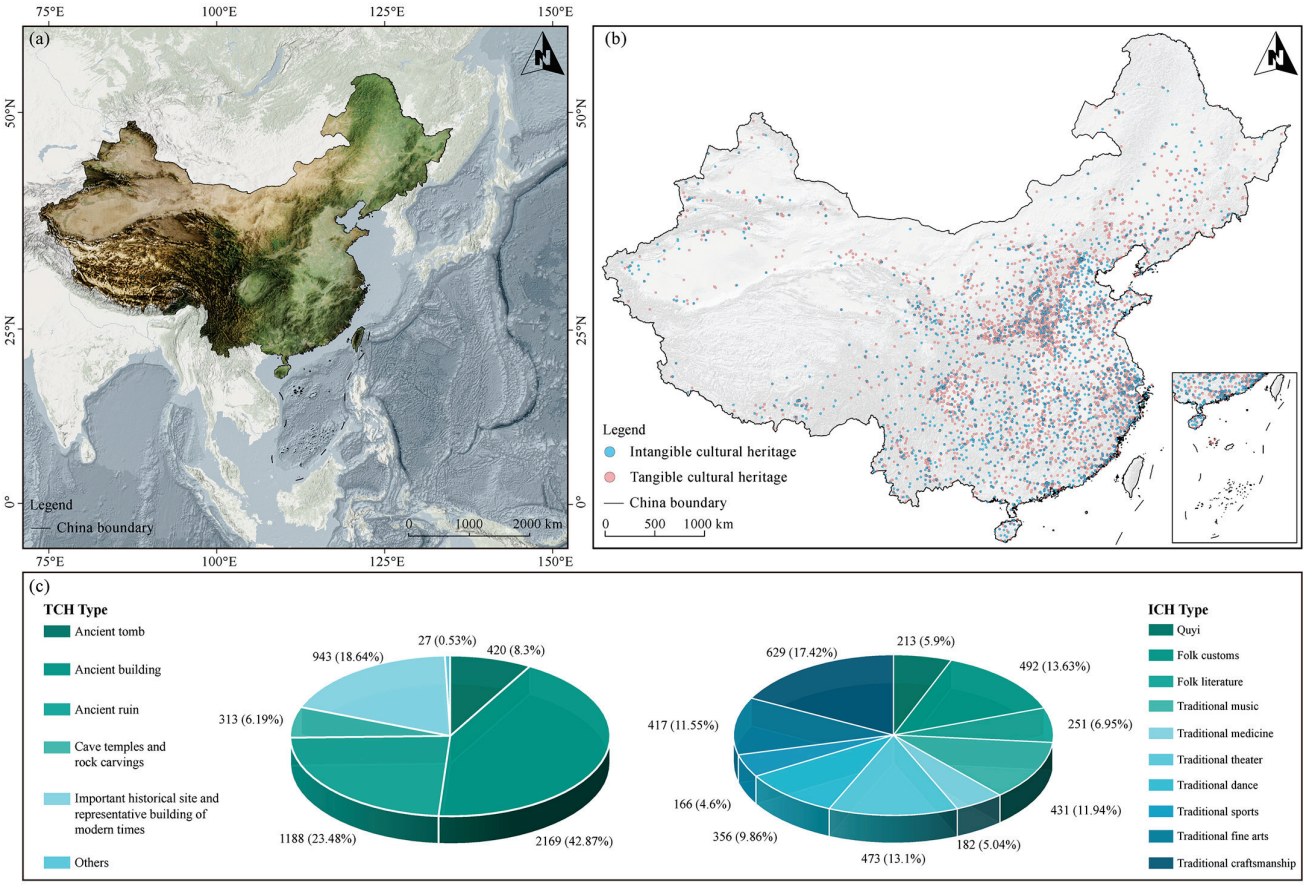


Figure 1. Spatial distribution and TCH/ICH proportions of cultural heritage in China. (**a**) Geographic location of China; (**b**) spatial distribution of tangible and intangible cultural heritage sites; (**c**) proportional distribution of TCH and ICH by category.

2.3. Data Processing and Variable Selection

To ensure data quality and model robustness, this study employed a standardized data processing workflow. First, samples with missing coupling degree values were removed. Median imputation was then used to handle missing values in feature variables, thereby minimizing distributional assumptions and reducing sensitivity to outliers [39]. Subsequently, the interquartile range (IQR) method was used to identify and remove outliers [40]. To improve skewed data distributions and enhance model convergence within the Bayesian framework, the Yeo–Johnson transformation was applied to all continuous variables:

$$Y^{(\lambda)} = \begin{cases} \frac{(Y+1)^{\lambda}-1}{\lambda}, & \text{if } Y \geq 0,\ \lambda \neq 0 \\ \ln(Y+1), & \text{if } Y \geq 0,\ \lambda = 0 \\ -\frac{(-Y+1)^{2-\lambda}-1}{2-\lambda}, & \text{if } Y < 0,\ \lambda \neq 2 \\ -\ln(-Y+1), & \text{if } Y < 0,\ \lambda = 2 \end{cases} \quad (1)$$

In Equation (1), Y represents the original data; the transformation parameter, λ, is determined using the maximum likelihood estimation method to correct the transformation shape [41].

To eliminate scale effects and ensure comparability between different features, Z-score standardization was applied to the transformed data [42].

The aforementioned workflow optimized data characteristics while preserving the underlying relational structures required for causal inference methods. In this study, drivers are formally defined as covariates that, after controlling for confounding variables and spatial scale effects, exhibit a directionally clear and statistically significant non-zero effect on the coupling index. This definition aims to link mechanisms such as economic structure and environmental exposure to the TCH–ICH coupling process and to examine their cross-regional heterogeneity. Regarding indicator selection, we conducted correlation analysis, multicollinearity testing, and significance testing on 13 baseline features and their constructed bivariate interaction terms, with the final variable set determined based on posterior inference. The construction of interaction variables was guided by both empirical verification of indicators and consideration of established theories, such as infrastructure complementarity and economic agglomeration effects [43].

To further assess the potential causal relevance and statistical robustness of the variables, the instrumental variables method [44] was employed to validate all potential factors. When an instrumental variable passes the weak instrument test (F-statistic > 10) and the second-stage regression yields $p < 0.1$, it receives a score of 2. Variables that fail the test but have second-stage regression ($p < 0.1$) receive a score of 1. Those that are significant only in conventional regression ($p < 0.1$) but lack valid instrumental variables receive a score of 0.5. This scoring scheme was designed to quantify the strength of causal evidence, rather than make definitive causal claims. Combined with the significance levels obtained from the HBM, the primary treatment variables were designated as those that were significant and had a score of 2. These variables serve as core explanatory variables for causal effects and their heterogeneity. Those that were significant and had a score of 1 were designated robustness variables for sensitivity analysis. Non-significant variables with a score of 0.5 were designated as exploratory variables to capture potential specificities. It should be noted that some variables exhibited relatively weak instrumental strength, indicating uncertainty in the causal effect estimates. Therefore, a dual-validation strategy was adopted to enhance the robustness of the results.

Theoretically, factors such as climate and topography are linked to heritage through their influence on population distribution. To statistically isolate these baseline effects, an analytical framework was constructed to support causal inference. Its primary task was

to control for confounding factors in order to estimate the net effects of key variables on heritage coupling, thereby addressing the scientific question of why and how such coupling occurs. Thus, by disentangling complex interdependencies, this approach mitigated the risk of tautological reasoning regarding human–heritage correlations and helped to reveal the differentiated mechanisms driving the spatial coupling of heritage.

2.4. Research Methods

We developed a progressive analytical framework integrating multi-scale identification, hierarchical modeling, and causal validation. First, the degree of heritage coupling was measured based on the processed data, and correlations for the initially screened variables were identified using a Bayesian model. Subsequently, three types of variables were re-screened based on significance levels and instrumental variables, enabling the identification of model-based causal effect estimates and heterogeneity under optimal grouping strategies (Figure 2).

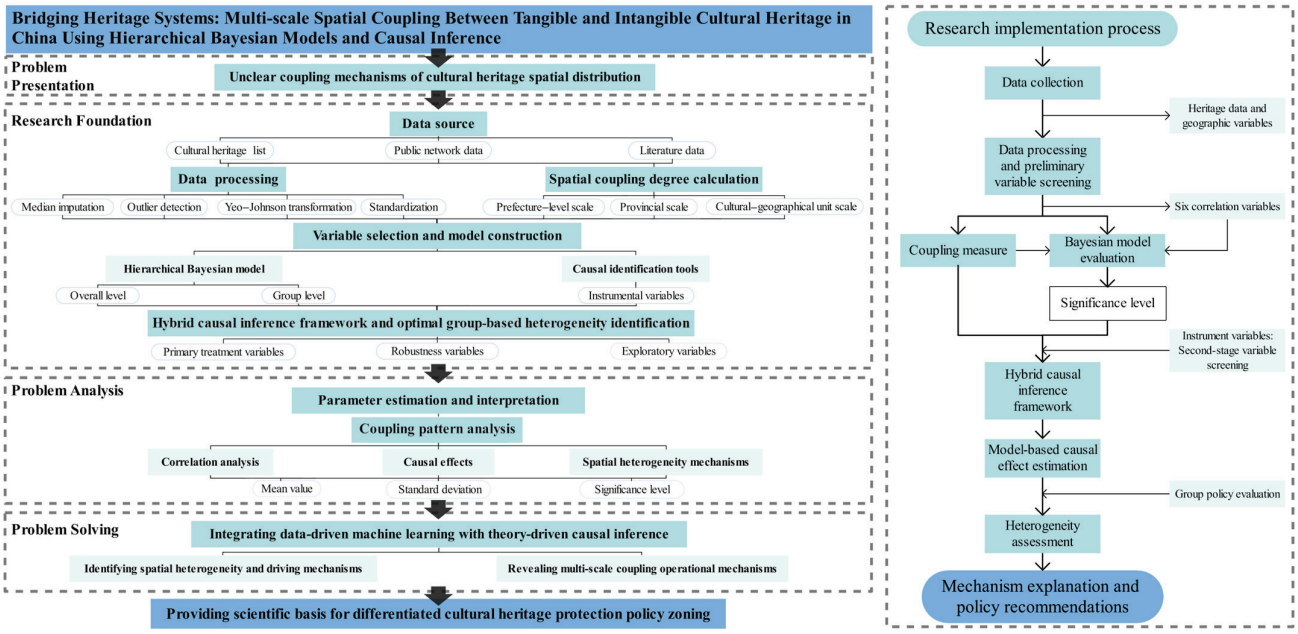


Figure 2. Research framework and implementation workflow diagram. Dark blue (R126, G175, B220): research title and significance; teal (R177, G219, B223): core analytical frameworks and processes; light green (R233, G248, B245): methods and supporting information.

(1) Spatial Coupling Degree Calculation Method

Coupling describes the dynamic relational mechanisms through which system components interact and co-evolve, and it has been assigned multiple meanings across disciplines, including correlation and interdependence. The term 'coupling' is used in heritage studies to refer to the cultural and social associations reflected through spatial co-location. Due to potential shared historical trajectories, social transformations and environmental constraints, spatial co-location constitutes a core condition for heritage coupling [9]. For example, traditional villages and historic buildings often form cultural ensembles together with folk customs, festivals and ritual practices [16]. Such symbiotic relationships between TCH and ICH represent a marker of cultural system integrity and have given rise to holistic conservation approaches that emphasize their interconnections [45].

Drawing on system coupling theory [46] and research on place attachment in cultural geography [47], this study integrated spatial analysis with cultural interpretation. The term 'coupling' refers to the association strength at the level of spatial co-location, measured through operational indicators, that is, the systematic co-existence patterns in which TCH

and ICH deviate from a random spatial distribution. Spatial coupling analysis does not equate 'coupling' with mere 'proximity', but instead conceptualizes spatial co-location as a spatial representation of cultural associations [48]. A high degree of coupling indicates that the two have formed a closely interconnected cultural ecosystem whose spatial clustering patterns reflect functional complementarity and cultural symbiosis. This theoretical lens shifts the meaning of heritage from isolated elements to the relational networks formed among them [49]. Accordingly, the measurable mathematical indicator captures statistical proximity and spatial evidence of interactive relationships between TCH and ICH.

This study used the modified co-location quotient (MCLQ) index to quantify the coupling intensity between TCH and ICH. Based on the CLQ theoretical framework, this method assesses the degree of association of point features by comparing the actual number of neighboring point pairs observed with the expected value under theoretical random distribution conditions [50]. The CLQ method is often adapted in spatial co-location analysis to suit the characteristics of different datasets [51]. A correction coefficient was introduced to adjust for the impact of geographical unit boundaries on the effective search area of heritage sites. Additionally, a geographical unit-based analytical framework was adopted, with differentiated search radii designed accordingly.

The number of neighboring point pairs satisfying the distance threshold conditions was first calculated:

$$C_i = \sum_{j=1}^{m} \sum_{k=1}^{n} I(D_{jk} \leq r) \tag{2}$$

Here, m and n denote the number of TCH and ICH points within a geographic unit, respectively; D_{jk} represents the Euclidean distance between TCH point j and ICH point k; and $I(\cdot)$ is an indicator function that takes the value 1 when the distance condition is satisfied, and 0 otherwise.

The expected number of nearest-neighbor point pairs was estimated under the assumption that heritage points were randomly distributed within the unit, determined by the ratio of the search area to the unit area. A correction factor was applied to account for the reduction in the effective search area for heritage points located near the unit boundaries.

$$E_i = \frac{m \times n \times \pi r^2}{A_i} \times \lambda \tag{3}$$

$$\lambda = \min\left(1, \frac{A_i}{\pi r^2}\right) \tag{4}$$

In the above equations, A_i represents the area of the geographical unit (km^2); λ represents the boundary effect correction coefficient; and r represents the search radius.

Finally, standardized MCLQ values were obtained by comparison with the values expected under a random distribution.

$$\mathrm{MCLQ}_i = \frac{C_i}{E_i} \tag{5}$$

Geographical units were classified into four coupling types based on MCLQ values: strongly coupled areas (MCLQ > 2.0), characterized by high spatial clustering and strong co-distribution of both heritage types; weakly coupled areas (1.0 < MCLQ $\leq$ 2.0), exhibiting moderate spatial association; independently distributed areas (0.5 $\leq$ MCLQ $\leq$ 1.0), exhibiting relatively random distribution patterns; and repulsively distributed areas (MCLQ < 0.5), exhibiting spatial separation between the two heritage types.

Considering China's administrative system and geographical characteristics, the analysis units were defined at three geographical scales: prefecture-level scale, provincial scale, and cultural–geographical unit scale. The search radii were set at 20 km, 150 km, and 250 km, respectively. This multi-scale framework follows place formation theory [52], recognizing that heritage associations are products of multi-level social processes [53]. Smaller radii are suited for identifying community-level symbiotic relationships, while larger radii can capture association patterns within cultural networks. Different radii correspond to different natural or social scales, reflecting interactive processes at various levels [54]. Historically, cultural practices within prefecture-level administrative units or settlements primarily unfolded within daily life circles and transportation networks [55]. Based on this theory, the 20 km radius was set to correspond to local interaction ranges, capturing the functional symbiotic relationships between TCH and ICH. Provincial administrative divisions often follow natural geographical boundaries and have important cultural and institutional integration functions. The 150 km radius was set to correspond to the spatial scale of regional cultural networks, capturing co-distribution patterns [56]. China's cultural geography is divided into seven regions. These units share common historical origins and cultural concepts, forming cultural ecological zones that transcend administrative boundaries [57]. The 250 km radius was set to capture spatial evidence of cultural diffusion and heritage evolution at the macro scale, identifying coupling patterns driven by geo-historical processes. This differentiated configuration considered the average spatial extent of geographical units and the distribution density of heritage sites at each scale, ensuring the effective identification of cross-scale spatial associations [58].

From the perspective of conservation practice, a strong coupling state signals active functional linkages, whereas a weak coupling state may indicate breaks in cultural transmission. This observation is consistent with the phenomena of cultural continuity and functional discontinuity documented in studies of traditional villages [57]. The coupling types described above are regarded as external manifestations of latent cultural interactions, but they do not directly reveal underlying social or cultural relations. Relying solely on this diagnostic indicator cannot quantify the social or environmental driving mechanisms, so we constructed interaction terms and implemented machine-learning models. This approach enabled us to empirically test the actual effects of factors and mechanism pathways, link spatial attributes to cultural or social explanatory mechanisms, and thereby strengthen the scientific validity of the coupling-system indicators.

(2) Hierarchical Bayesian Model

The core of HBM lies in nesting effects at both the overall level and group level, achieving information sharing and bias correction through hyperparameter constraints [59]. Compared with traditional regression methods, HBM not only avoids underestimation of standard errors caused by ignoring intra-group correlations but also maintains robustness under conditions of unbalanced sample sizes across groups or sparse local data [60]. This characteristic effectively integrates the regional heterogeneity and multi-level data structures of the three types of geographical units, thereby preventing the information loss that would result from single-level models.

The three geographical grouping patterns were used as grouping criteria, enabling separate modeling at each geographical level and revealing the similarities and differences between the overall and local effects [61]. The sample was randomly divided, with 80% allocated to the training set and the remainder to the test set, thus balancing the requirements for model training and independent validation. The basic model form is as follows:

Group level: Data from each group are directly processed for fitting, and the Markov Chain Monte Carlo (MCMC) method is used to estimate parameters for each group [62]. This method preserves inter-group differences through independent sampling. Group-

specific parameters are subject to structural constraints due to shrinkage effects, which prevent overfitting by exhibiting a tendency to shrink toward the overall mean [63].

$$y_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma^2),\ \mu_{ij} = \alpha_j + \sum_{k=1}^{K} \beta_{jk} X_{ijk} \tag{6}$$

In Equation (6), y_{ij} represents the MCLQ value of the i-th observation in the j-th group; X_{ijk} represents the k-th feature variable; β_{jk} is the corresponding regression coefficient, revealing the variable effects within different groups; α_j represents the group-specific intercept term; and σ^2 represents the variance parameter.

Overall level: Imposing weakly informative hyperpriors on group-level intercepts and slopes stabilizes estimates for small-sample groups while preserving between-group heterogeneity. When sample sizes are small or residual variance is large, posterior estimates become more sensitive to hyperpriors [64]. Tighter variance hyperpriors strengthen partial pooling, draw small-sample parameters toward the overall mean, and suppress extreme estimates, thereby reducing overfitting. Looser variance hyperpriors widen uncertainty and more fully reveal cross-group differences. The mean hyperprior determines the direction and magnitude of shrinkage. Accordingly, a zero-centered, weakly informative specification with an appropriate scale helps avoid systematic bias in small-sample groups [65]. The hyperprior on the residual term also indirectly modulates shrinkage in small groups by altering the effective noise level.

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \tau_\alpha^2),\ \beta_{jk} \sim \mathcal{N}(\mu_{\beta_k}, \tau_{\beta_k}^2) \tag{7}$$

Here, μ_α and $\mu_{\beta k}$ serve as hyperpriors representing overall-level mean effects; τ_α^2 and $\tau_{\beta k}^2$ are variance parameters reflecting the degree of effect variation across different geographical units.

The model captures overall patterns and inter-group variations through a multi-layered parameter structure, enabling cross-scale effect decomposition. Uncertainty quantification provides complete posterior probability distributions for the estimated parameters. MCMC sampling convergence was diagnosed using the Gelman–Rubin statistic (R^) and Monte Carlo standard error (MCSE) [66]. For model evaluation, the coefficient of determination (R^2) was employed to quantify the model's fitting capability [67]. Root mean square error (RMSE) was used to measure the accuracy of the model's predictions [68]. The Bayesian p-value (BPV) was introduced to test the consistency between the observed and generated data [69]. Together, these three metrics provided comprehensive evaluation criteria for HBM performance from the perspectives of goodness of fit, prediction accuracy, and statistical validity. In addition, five-fold cross-validation was used to assess out-of-sample predictive stability. One of the five randomly partitioned subsets was designated as the test set and the remaining subsets as the training set, with fitting and prediction repeated across folds. The RMSE and mean absolute error (MAE) on each test fold were then averaged to quantify generalization performance [70].

To further assess whether the hierarchical structure of the HBM had adequately captured spatial dependence, multi-strategy spatial autocorrelation diagnostics were applied to the model residuals. Given the sensitivity to the specification of spatial weight matrices, three robustness schemes [71] were implemented: (a) queen contiguity matrices; (b) adaptive distance–band matrices with thresholds set at the 75th and 90th percentiles of the k-nearest-neighbor (KNN) distance distribution (with k = 8, 6, and 3 for the three geographical scales); and (c) KNN matrices chosen to balance statistical power in small samples against over-connection in large samples (with k = 6, 3, and 2 for the three scales). All weight matrices were row-standardized to ensure comparability across heterogeneous neighborhood structures [72]. For each specification, the global Moran's I of the residu-

als was computed, and significance was assessed using a 999-permutation Monte Carlo test [71]. If the residuals exhibited no significant spatial autocorrelation across all strategies ($p > 0.05$), we concluded that the HBM had sufficiently captured spatial dependence, obviating the need for additional spatially structured terms.

(3) Causal Effect Estimation and Heterogeneity Analysis

A hybrid causal inference framework based on double machine learning (DML) and ordinary least squares (OLS) was employed to investigate the causal mechanisms underlying the cultural heritage coupling. Within this framework, the DML approach was used to control for high-dimensional confounders and reduce model specification bias [73], while the OLS approach was used to retain interpretability. This design ensured accurate estimation and robust, reliable identification of causal effects. The basic mathematical structure is expressed as follows:

$$Y_i = \alpha + \beta T_i + \gamma X_i + \epsilon_i \tag{8}$$

$$Y_i = \theta T_i + g(X_i) + \epsilon_i \tag{9}$$

$$T_i = m(X_i) + \nu_i \tag{10}$$

$$\hat{g}(x) = E[Y|\ X = x],\ \hat{m}(x) = E[T|\ X = x] \tag{11}$$

$$\hat{\theta}_{DML} = \left(\sum_{i=1}^{n} (T_i - \hat{m}(X_i))^2 \right)^{-1} \sum_{i=1}^{n} (T_i - \hat{m}(X_i))(Y_i - \hat{g}(X_i)) \tag{12}$$

Equation (8) specifies the OLS regression model used as the baseline estimation. Y_i denotes the MCLQ value of the i-th geographical unit; T_i represents the treatment variable; X_i is the vector of the control variables; β is the target causal effect parameter; γ is the coefficient vector for the control variables; α is the intercept term; and ϵ_i is the random error term.

Equations (9)–(12) present the core structure of the DML method [74]. θ denotes the target causal effect parameter (average treatment effect, ATE); $\hat{g}(X_i)$ and $\hat{m}(X_i)$ denote the estimated conditional expectation functions of the outcome and treatment variables with respect to the covariates; ϵ_i and ν_i are the random error terms; $\hat{\theta}_{DML}$ is the causal effect estimated by DML; $(T_i - \hat{m}(X_i))$ is the residualized treatment variable that removes the influence of control variables; $(Y_i - \hat{g}(X_i))$ is the residualized outcome variable that eliminates the effect of control variables; and n is the sample size.

The DML estimation required randomly splitting the sample into K folds to implement a cross-fitting strategy [75]. Using data from the other folds, $\hat{g}(X_i)$ and $\hat{m}(X_i)$ were estimated separately based on the random forest algorithm and then used to predict the target fold. The number of trees was adaptively adjusted according to the sample size to balance estimation accuracy and overfitting risk. The maximum tree depth was constrained using a logarithmic function based on the sample size, and a feature subsampling strategy was employed to reduce variance. Residuals were then computed to ensure the unbiasedness of the causal effect estimates. Finally, ordinary least squares regression was performed on the residualized variables, and the causal effect estimates from each fold were weighted by sample size to obtain $\hat{\theta}_{DML}$.

We employed Conley heteroscedasticity and autocorrelation consistent (HAC) standard errors to assess the robustness of our causal estimates to potential spatial autocorrela-

tion [76]. Given the asymptotic nature of Conley–HAC and the small sample sizes at the other two scales, the main analysis was conducted only for the prefecture-level sample. Bandwidths were chosen in a data-driven manner: we used the 25th and 50th percentiles of the nearest-neighbor distances between city centroids as the lower and main bandwidths and determined an adaptive bandwidth by requiring that $\geq$70% of nodes had at least one neighbor. Distances were computed as projected Euclidean distances, and spatial weights were constructed using the Bartlett (triangular) kernel [77]. Variable selection focused on pre-specified main treatment variables and robustness-check variables to mitigate post-selection inference bias.

For the best grouping identified through subsequent multi-group strategies, the significance and degree of heterogeneity were assessed using Cochran's Q test and the I^2 statistic, respectively [78]. The Q statistic follows a chi-square distribution with $k-1$ degrees of freedom. Typically, $I^2 > 75\%$ indicates high heterogeneity, $25\% < I^2 \leq 75\%$ indicates moderate heterogeneity, and $I^2 \leq 25\%$ indicates low heterogeneity.

To ensure reliable causal inference, differentiated robustness checks were constructed based on the theoretical importance of variables and the strength of causal evidence [79].

$$\Delta_{\text{robust}} = \frac{\left|\hat{\theta}_{\text{extended}} - \hat{\theta}_{\text{base}}\right|}{\left|\hat{\theta}_{\text{base}}\right|} \times 100\% \tag{13}$$

$$\Delta_{\text{norm}} = \frac{\left|\hat{\theta}_{\text{OLS}} - \hat{\theta}_{\text{DML}}\right|}{\sqrt{\left(\text{SE}_{\text{OLS}}^2 + \text{SE}_{\text{DML}}^2\right)}} \tag{14}$$

Equation (13) tests for the influence of potential confounding variables on the main effect. If the change in effect is less than 20%, the results are considered robust. $\hat{\theta}_{\text{base}}$ and $\hat{\theta}_{\text{extended}}$ denote the estimated causal effects of the main treatment variable in the baseline model and the extended model that incorporates robustness variables, respectively. Equation (14) compares the estimation differences between the two methods. If $\Delta_{\text{norm}} < 1.96$, the results are considered statistically consistent across the two methods. SE_{OLS} and SE_{DML} represent the standard errors of the two estimates, respectively.

A rigorous, comprehensive test was conducted for the primary treatment variables. Potential confounding variables were added sequentially to examine the stability of their coefficients, and the consistency of the method was verified across the baseline and each extended model. For robustness variables, the focus was on assessing methodological consistency, supplemented by a simplified version of Equation (13) to test the effects of key control variables and avoid excessive testing. For exploratory variables, the emphasis was placed on evaluating the stability of the coefficients across different model specifications. Consistency checks between methods were only performed for significant variables in the baseline model. Furthermore, to clearly illustrate the complex structure of the constructed HBM and causal inference models, Figure 3 displays the hierarchical structure, prior distribution specifications, and core parameters.

If longitudinal data become available, the framework can be extended temporally. The HBM can be extended to a cross-classified spatiotemporal hierarchy, where time-varying random intercepts and key slopes with weak temporal-dependence priors are introduced within each geographic scale. Spatially structured components can also be combined with the temporal process through a separable space–time covariance [80], preserving cross-scale analysis while providing spatiotemporal trajectories of key driver effects. Concurrently, the causal module can be extended via a two-way fixed-effects difference-in-differences design [81]. By aligning units to the onset of treatment and incorporating lead and lag indicators, dynamic effect paths can be estimated in an event-study form, using cluster-robust standard errors to ensure robustness against serial correlation [82].

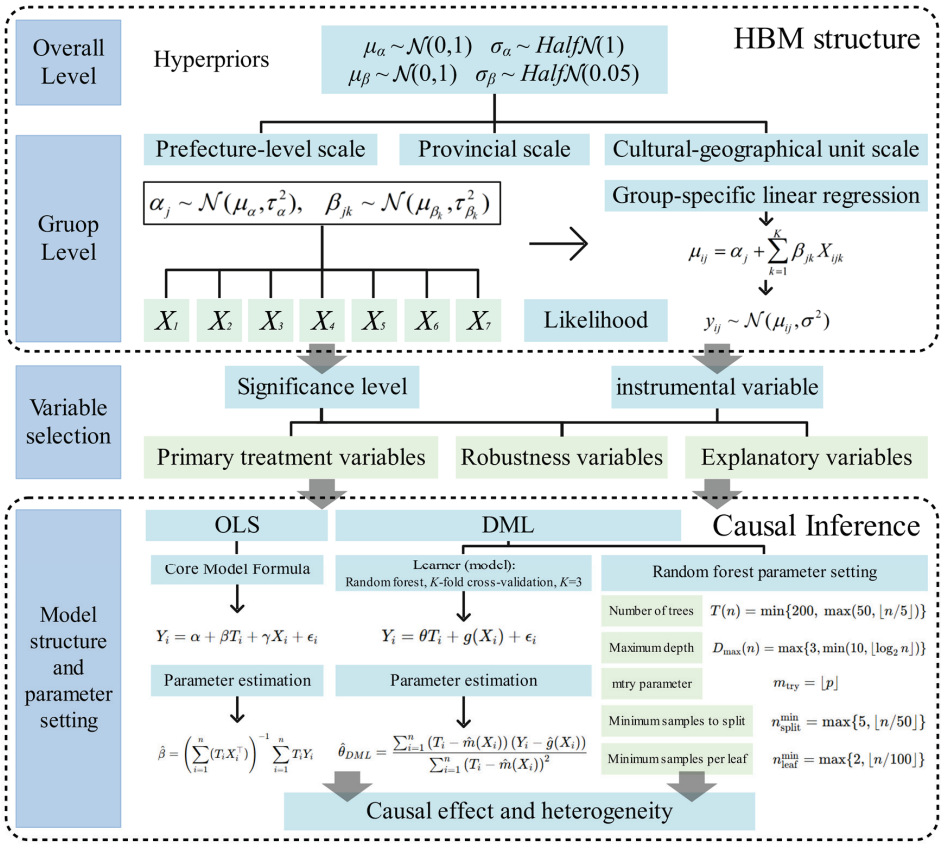


Figure 3. Schematic diagram of the HBM and causal inference model structure.

(4) Optimal Grouping Strategy for Heterogeneity Analysis

We used a multidimensional evaluation framework to identify the optimal grouping strategy for heterogeneity. Sample units were grouped according to key variables, enabling causal effects to be analyzed comparatively within each group. Four grouping methods were applied: median split, tertile split, quartile extreme, and K-means clustering. The composite scoring function encompassed three aspects: statistical significance testing, effect size measurement, and the magnitude of intergroup differences.

To quantify the differences between the groups, Cohen's *d* and Partial Eta Squared were used as effect size indices for two-group and multiple-group comparisons, respectively [83]. The corresponding calculation formulas are as follows:

$$d = \frac{\overline{X}_1 - \overline{X}_2}{s_{\text{pooled}}} \tag{15}$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{16}$$

$$\text{CV}_{\text{between}} = \frac{\sigma(\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k)}{\overline{X}} \tag{17}$$

Cohen's *d* effect size is typically classified as small (d ≈ 0.2), medium (d ≈ 0.5), or large (d ≈ 0.8). The coefficient of variation between groups ($\text{CV}_{\text{between}}$) measures the relative dispersion among group means, with values greater than 0.4 indicating a large effect size.

Based on these parameters, the magnitude of intergroup differences was introduced to evaluate the performance disparities across different classification combinations. The optimal combination was required to reach statistical significance ($p < 0.05$) and to exhibit a large effect size [84]. Additionally, each group had to meet the sample size requirement

for statistical inference ($n \geq 30$), and the grouping results had to be consistent with geographical and social logic. This weighting scheme emphasized the central role of effect size while assigning equal importance to statistical significance and the magnitude of group differences.

$$\text{Score} = 0.2 \times (1 - p) + 0.6 \times \text{Effect Size} + 0.2 \times \frac{\Delta\mu}{\max(\Delta\mu)} \quad (18)$$

where p represents the significance level obtained from statistical testing; Effect Size represents the standardized effect size; $\Delta\mu$ represents the mean difference between groups; and $\max(\Delta\mu)$ represents the maximum mean difference.

3. Results

3.1. Multi-Scale Spatial Coupling Patterns of Cultural Heritage

According to the MCLQ index results, the spatial coupling of cultural heritage had a pronounced scale effect across the three analyzed scales (Figure 4). At the prefecture-level scale, strongly coupled areas dominated, accounting for 55.5% of the total, indicating a high degree of co-existence and spatial overlap between the two types of heritage at a fine spatial scale. At the provincial scale, both weakly coupled and independently distributed areas reached their peak proportions of 40.6%, reflecting a weakening and differentiation of heritage associations at intermediate scales. At the cultural–geographical unit scale, the proportions of weakly and strongly coupled areas increased to 42.8% and 28.6%, respectively. These results revealed a reorganization pattern at larger scales, indicating a deviation from the overall trend of continuous weakening in the coupling relationships.

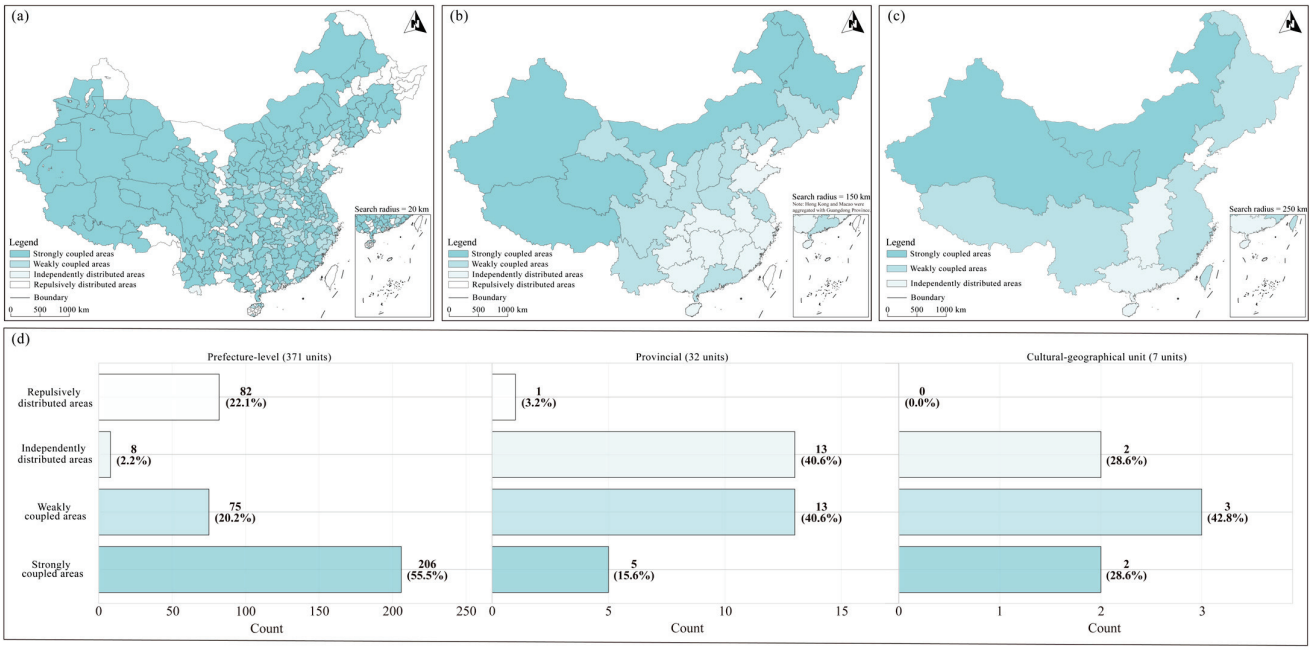


Figure 4. Spatial coupling patterns between TCH and ICH based on MCLQ analysis across three scales. (**a**) Prefecture-level scale; (**b**) provincial scale; (**c**) cultural–geographical unit scale; (**d**) number of units by coupling type across three scales.

It should be emphasized that the spatial coupling of cultural heritage did not decline monotonically with increasing scale. From the prefecture-level to provincial scale, the proportion of strongly coupled areas decreased markedly while the proportions of weakly coupled and independently distributed areas increased substantially, in line with the general pattern of spatial attenuation. Theoretically, expanding the search range at larger

scales would substantially increase the expected value, leading to lower MCLQ indices. However, the proportion of strongly coupled areas rebounded at the cultural–geographical unit scale. This manifested spatially in specific regions in the northwest, north, and northeast, which consistently maintained high coupling—an unexpected response pattern. The re-aggregation trend beyond administrative boundaries at the geographical level may reflect intrinsic cross-regional cohesion and cultural identity [85]. This cross-scale stability suggests that the Chinese cultural system possesses a binding force that maintains connections between TCH and ICH at the macro-spatial level. Consequently, the resulting integrated cultural ecosystem endows the two types of heritage with sufficiently strong associations to overcome the dilution effect at larger scales.

Overall, the prefecture-level scale better reveals local symbiosis between heritage types, while the provincial scale highlights spatial differentiation and independence. The cultural–geographical unit scale reflects the integration and continuity of cultural traditions over a broader area. This comparison underscores the scale sensitivity and regional variability of heritage coupling, the driving mechanisms of which require further quantitative investigation.

3.2. Performance and Diagnostic Assessment of HBM

The three basic variables selected through the aforementioned mechanism were as follows: mean annual temperature (X_1), reflecting regional thermal conditions; the Shannon–Wiener index (X_2), capturing the diversity of ethnic composition; and population density (X_3), representing the density of cultural carriers and the social foundation for heritage transmission. The four interaction variables were as follows: GDP–population interaction (X_4), constructed from per capita GDP and population density, capturing the combined effects of economic development and population size; transport–hydrological network interaction (X_5), comprising road network density and hydrological connectivity, reflecting the synergistic influence of transportation accessibility and water network connectivity; industrial structure (X_6), defined as the interaction between GDP and the proportion of tertiary industry, representing the effect of tertiary sector upgrading on heritage; and topographic–climate interaction (X_7), integrating elevation and mean annual temperature to account for the interplay between topography and climate.

Quantitative validation showed that the selected features had a mean correlation of $|r| = 0.309$, which was significantly higher than that of the unselected features ($|r| = 0.227$). Compared with the full-variable model ($R^2 = 0.633$), this seven-variable system achieved good explanatory power within a linear approximation framework ($R^2 = 0.612$). Additionally, the mean variance inflation factor (VIF) of the selected variables decreased to 1.71, facilitating the maintenance of independence among hierarchical effects and stabilizing parameter estimation in the HBM. Overall, this feature selection strategy achieved an optimized balance between statistical performance and model stability.

The HBM demonstrated strong predictive performance and model fit (Table 1). The R^2 values for the training and test sets were 0.612 and 0.553, respectively, indicating that the selected covariates and their interaction terms explained 61.2% and 55.3% of the variance in coupling degree, respectively. The RMSE values were 0.456 and 0.504, respectively, demonstrating good predictive accuracy and generalization capability. Cross-validation yielded a mean RMSE of 0.472 ± 0.038 and a mean MAE of 0.294 ± 0.039. The close agreement between the cross-validated RMSE and the test-set RMSE corroborated the reliability of the model evaluation. Moreover, the small standard deviations indicated low sensitivity to the particular data partitioning. The BPV, close to the ideal value of 0.5, suggested adequate model fit without evidence of systematic bias or misspecification. The

slight reduction in explanatory power and the controlled increase in prediction error on the test set were both within acceptable thresholds.

Table 1. HBM evaluation metrics.

Evaluation Metrics	Training Dataset	Testing Dataset	Cross-Validation (Mean ± SD)
R^2	0.612	0.553	—
RMSE	0.456	0.504	0.472 ± 0.038
MAE	—	—	0.294 ± 0.039
BPV	0.513	—	—

To further assess the fitting quality and the validity of assumptions in the HBM, this study conducted multiple diagnostics based on residual characteristics and predictive performance (Figure 5).

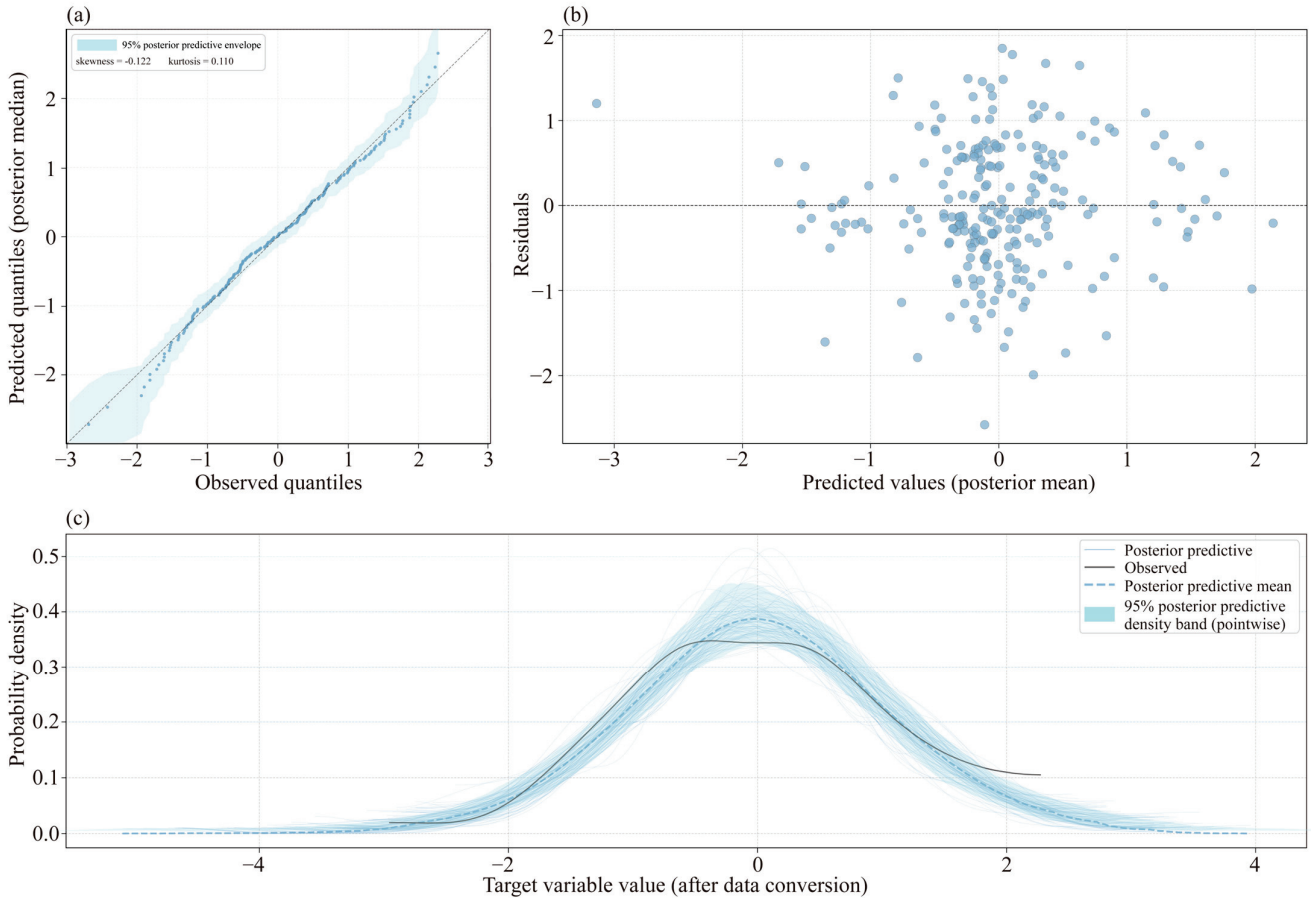


Figure 5. HBM performance diagnostics. (**a**) Posterior predictive QQ plot; (**b**) residuals and predicted values; (**c**) posterior predictive check.

As Figure 5a shows, the data points are closely aligned with the theoretical quantile lines, and the residuals largely conform to the normality assumption. The skewness of −0.122 and the kurtosis of 0.110 are both close to the theoretical value of 0. Only slight deviations occur in the distribution tails, with no evident heavy-tailed or peaked behavior, indicating an absence of significant non-normality [86]. Figure 5b shows no obvious funnel-shaped or systematic patterns, and the residuals exhibit homoscedastic characteristics. The residuals remain relatively uniformly dispersed across the full range, confirming that the model maintains stable predictive accuracy for geographical units at different coupling levels. In Figure 5c, the posterior predictive means largely overlap with the observed

data in most regions. The 100 posterior sample curves, reflecting quantified uncertainty, display approximately unimodal normal distributions similar to the observed data. The peak positions largely coincide, and the distribution widths and tail decay patterns are comparable, indicating that the model successfully captures the underlying probabilistic generative process of the coupling degree.

Multi-strategy spatial autocorrelation diagnostics indicated that the HBM residuals showed no significant spatial dependence at any scale (Table 2), suggesting that the hierarchical structure had appropriately modeled spatial association. At the prefecture-level scale, Moran's I ranged from 0.0029 to 0.0391 (p = 0.06–0.45). At the provincial scale, residual spatial structuring was even weaker, with Moran's I between −0.0692 and 0.0017 (p = 0.36–0.47). Despite the very small sample size at the cultural–geographical unit scale, the results remained non-significant (p = 0.25–0.49). All diagnostic strategies supported residual spatial independence. Accordingly, no additional spatially structured terms were introduced, thereby avoiding the risk of over-parameterization.

Table 2. Residual spatial autocorrelation diagnostics for the HBM across scales.

Geographic Scale	Weight Matrix	Moran's *I*	*z*-Score	*p*-Value
Prefecture-level	Queen contiguity	0.0029	0.15	0.45
	Distance band (309.7 km)	0.0391	1.38	0.08
	Distance band (393.3 km)	0.0364	1.59	0.06
	KNN (k = 6)	0.0273	0.94	0.17
Provincial	Queen contiguity	−0.0692	−0.21	0.42
	Distance band (1011.4 km)	−0.0381	0.02	0.47
	Distance band (1307.6 km)	−0.0581	−0.25	0.43
	KNN (k = 3)	0.0017	0.31	0.36
Cultural–geographical unit	Queen contiguity	−0.0499	0.53	0.25
	KNN (k = 2)	−0.1838	−0.07	0.49

Overall, the HBM achieved stable predictive accuracy and residual spatial independence across scales, supporting its use as a reliable basis for subsequent causal inference without additional spatial terms.

3.3. Posterior Inference Results of HBM

Based on the diagnostics of the MCMC chains, the sampling trajectories of the β parameters at the overall level exhibit good mixing (Figure 6). The trajectories display stable random-walk patterns within the target range, with no apparent drift or stagnation, indicating that the model converged to a stable posterior distribution. Convergence diagnostics show that the $\hat{R}$ values are all below the convergence threshold of 1.1 and fall within the [1.015, 1.037] range. As a proportion of the posterior standard deviation, the MCSE is below 5% for all parameters, generally satisfying the required sampling precision. The dashed lines on the left indicate that the posterior distributions of the three scale groups differ significantly, confirming that the HBM successfully captures the statistical characteristics of between-group differences.

The posterior distributions of the regression coefficients for the seven feature variables at the overall level showed clear patterns of statistical significance. The mean values of β_{X1}, β_{X2}, β_{X3}, β_{X4}, β_{X5}, β_{X6}, and β_{X7} were −0.240, 0.035, −0.233, −0.121, 0.161, 0.224, and −0.002, respectively, with corresponding standard deviations of 0.068, 0.061, 0.074, 0.060, 0.057, 0.050, and 0.061, respectively; the 95% highest density intervals (HDIs) were (−0.371, −0.111), (−0.088, 0.153), (−0.369, −0.080), (−0.235, −0.002), (0.049, 0.276), (0.129, 0.322), and (−0.115, 0.124), respectively. The relatively low standard deviations for all feature variables indicated that the uncertainty in parameter estimation was well

controlled. X_1, X_3, and X_4 exhibited statistically significant negative effects on coupling, whereas X_5 and X_6 exhibited statistically significant positive effects. No evidence was found for a significant effect of X_2 or X_7, suggesting that their direct associations with coupling are weak and may instead operate indirectly through mediating variables.

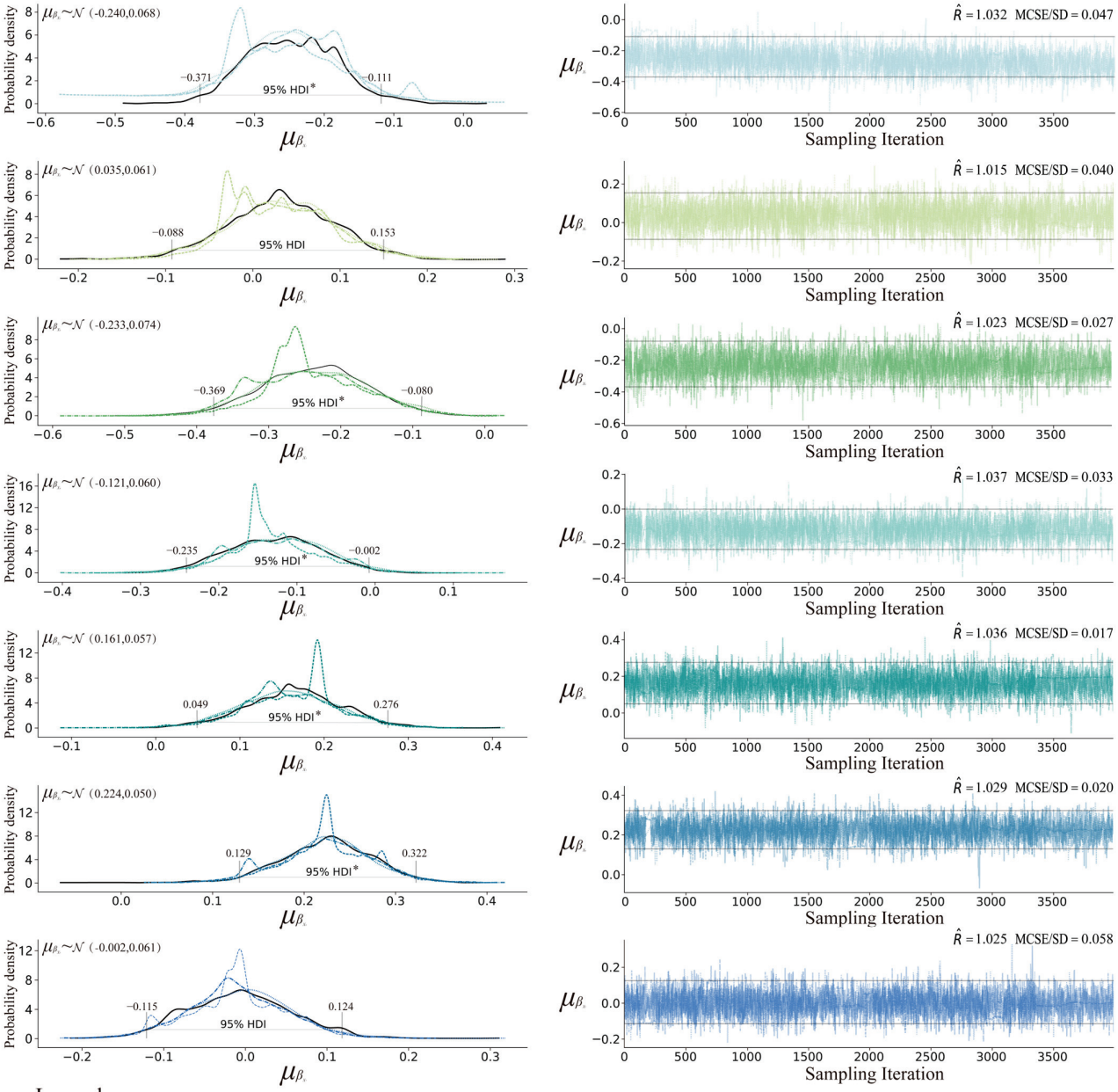


Figure 6. Posterior distributions and MCMC samples of overall level regression coefficients. Colors indicate individual variables, with each variable's posterior distribution and corresponding MCMC samples shown in the same color.

Even under hyperprior constraints, differences between parameter groups were maintained through independent sampling (Figure 7). The five significant factors at the overall level remained significant across all three scales, reflecting consistency in their influence mechanisms. At the group level, the effect strengths of the variables displayed some differentiation. Specifically, the mean values of β_{X1} were -0.2447, -0.2386, and -0.2440; β_{X3} were -0.2324, -0.2298, and -0.2307; β_{X4} were -0.1169, -0.1198, and -0.1185; β_{X5} were 0.1666, 0.1648, and 0.1644; and β_{X6} were 0.2331, 0.2198, and 0.2263. Notably, the interaction

between GDP and population density had a negative effect, whereas the interaction with the share of the tertiary sector was positive. This suggests that different economic development trajectories may influence heritage coupling through either spatial restructuring or service-oriented pathways.

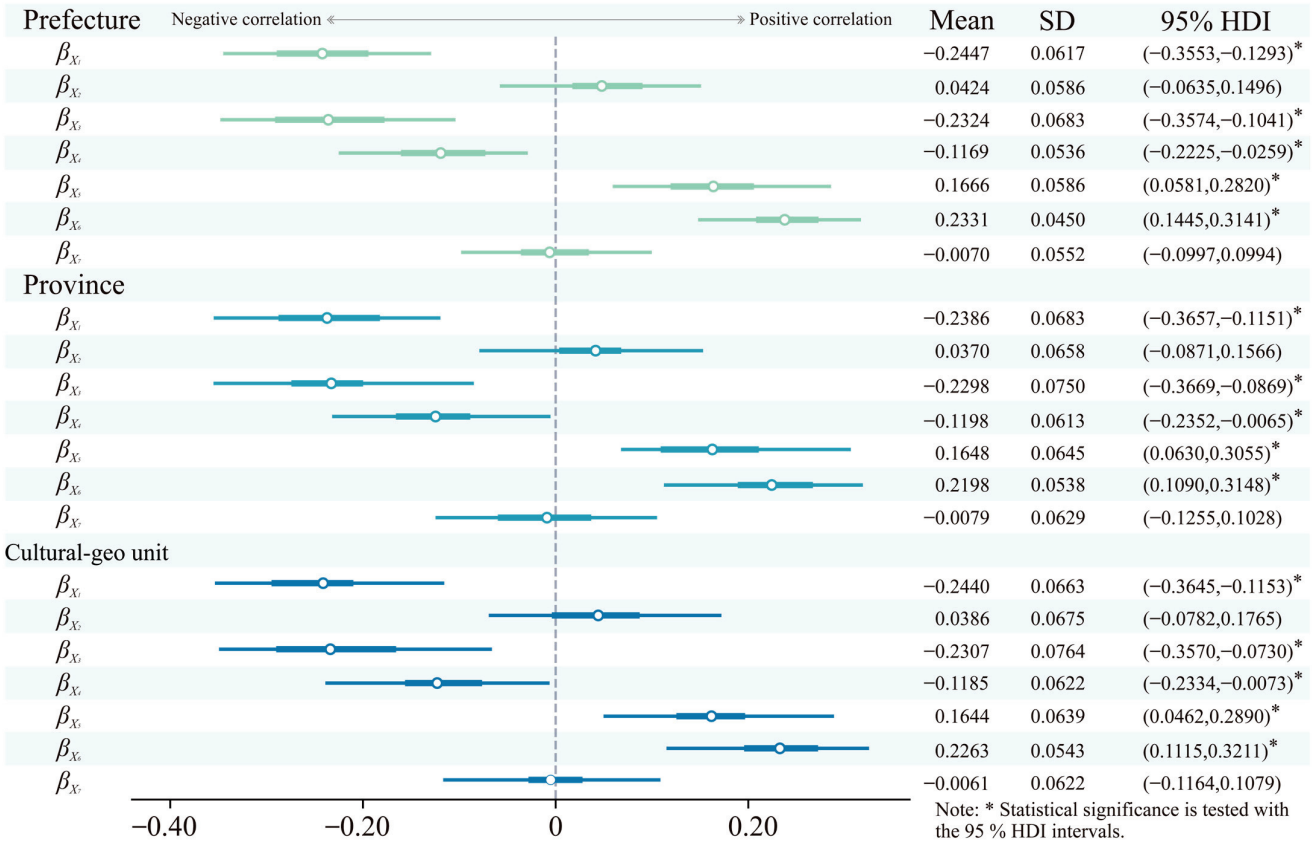


Figure 7. Posterior distributions of group-level regression coefficients. Different colors represent the three geographic levels, with all seven coefficients for each level shown in the same color.

Except for X_4, the remaining four significant variables reached their maximum effect strength at the prefecture-level scale. This pattern reflected the amplification of local factors at fine spatial units and the regionalized characteristics of heritage transmission. X_1, X_3, and X_6 exhibited the lowest effect strength at the provincial scale, forming a U-shaped distribution with relatively strong effects at the other two scales. By contrast, X_4 reached its maximum effect strength at the provincial scale, while X_5 showed a monotonically decreasing pattern. These differences indicated that the coupling characteristics of multi-scale geographic processes were influenced by distinct driving mechanisms. The provincial scale did not simply inherit the hierarchical changes in scale; specific inhibitory mechanisms may have systematically weakened or diluted certain coupling processes. The high effects observed at the prefecture-level scale reflected the direct influence of localized processes, whereas the rebound of effects at the cultural–geographical unit scale reflected the regulatory role of the overall cultural system on coupling processes. These findings deviated from the traditional assumption of diminishing scale effects [87] and revealed a nonlinear interaction pattern between administrative boundary effects and the continuity of geographic processes.

3.4. Causal Effect Estimation and Comparison

The mixed causal inference framework further refined observed correlations into model-based causal effect estimates, which necessarily rely on core assumptions such as

model specification and instrumental-variable validity. Accordingly, the resulting causal effects should be interpreted as estimates conditional on these assumptions. Based on the combined validation of HBM significance and the strength of causal evidence, X_1 and X_5 were selected as the primary treatment variables, and X_3 and X_4 were used as robustness variables. The exploratory variables included mean annual precipitation (X_8), elevation (X_9), river network density (X_{10}), road network density (X_{11}), tertiary sector share (X_{12}), and slope gradient (X_{13}). At the prefecture-level scale, Conley–HAC standard errors were estimated for alternative bandwidths (Table 3). The significance levels of X_1, X_5, X_3, and X_4 remained unchanged, but under the baseline bandwidth, their standard errors changed by −2.7%, +5.6%, −1.9%, and −4.8%, respectively. Across the full bandwidth range, the maximum variation ranged from −5.1% to +9.5%. These patterns indicated that the core inferences were robust to plausible spatial correlation. The adaptive bandwidth achieved 75.2% node coverage, suggesting that this setting adequately captured the principal scale of inter-city spatial dependence. Overall, the causal inference model exhibited no material spatial dependence and, so, introducing an additional spatially structured component was unnecessary. Subsequently, we compared causal effect estimates from OLS and DML (Figure 8).

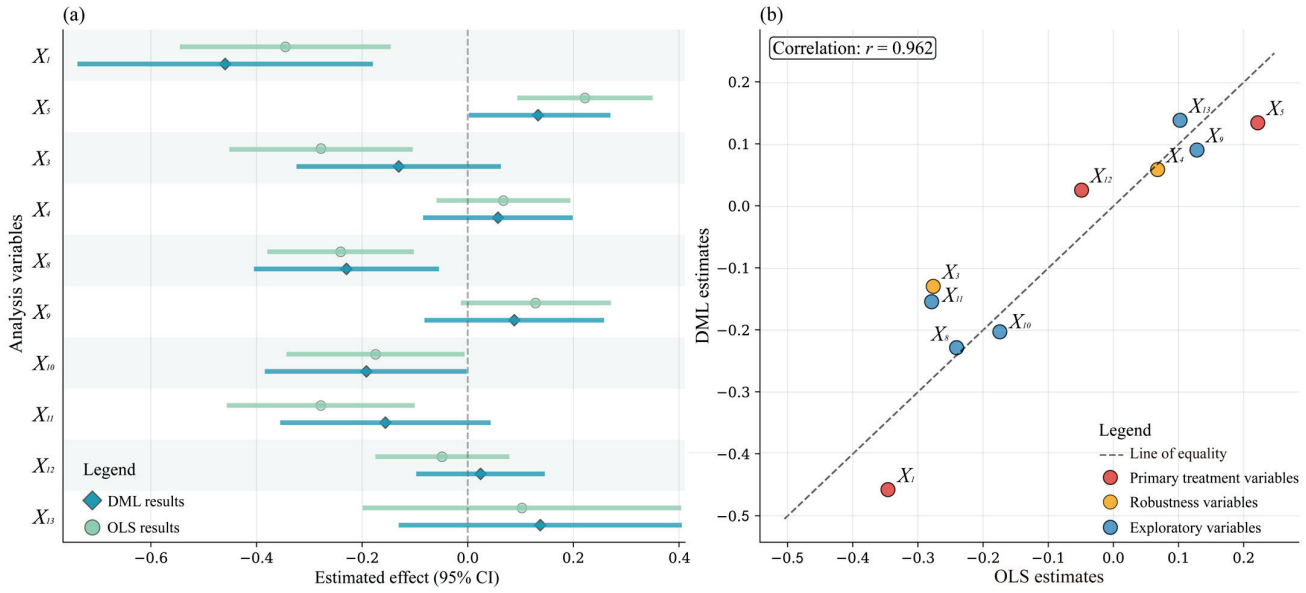
Table 3. Spatial robustness checks for causal estimates at the prefecture level.

Variable	SE/*p*	Conley 90.7 km (SE/*p*/ΔSE)	Conley 112.8 km (SE/*p*)	Conley 138.6 km (SE/*p*)
X_1	0.0994/0.0001	0.0970/0.0001/−2.4%	0.0967/0.0001/−2.7%	0.0968/0.0001/−2.6%
X_5	0.0563/0.0045	0.0583/0.0062/3.7%	0.0594/0.0071/5.6%	0.0616/0.0095/9.5%
X_3	0.0875/0.0033	0.0859/0.0027/−1.8%	0.0858/0.0027/−1.9%	0.0854/0.0026/−2.4%
X_4	0.0652/0.4955	0.0619/0.4728/−5.1%	0.0621/0.4742/−4.8%	0.0630/0.4800/−3.5%
Node Coverage	—	25.6%	50.0%	75.2%

X_1, X_5, X_8, and X_{10} showed significant and robust causal effect estimates under both estimation strategies. X_1 exhibited a strong negative effect estimate, indicating that low-temperature regions were more conducive to maintaining heritage associations and traditional lifestyles. The ATEs were −0.345 and −0.459, with the latter representing a 33% increase in effect magnitude. This suggested that, after controlling for selection bias and confounding variables, temperature had a more pronounced negative impact on coupling intensity. X_5 showed a significant positive effect estimate, with ATEs of 0.222 and 0.133. The approximately 40% reduction in the latter effect may have resulted from DML's control of the independent effects of water systems and roads, as well as their interactions with other infrastructure components. X_8 exhibited a robust negative effect estimate, with ATEs of −0.240 and −0.229, indicating that high precipitation negatively impacted heritage coupling. X_{10} exhibited a significant negative effect estimate, with ATEs of −0.174 and −0.192. Its positive interaction with road networks suggested that the negative effect of water system density was reduced by road conditions. X_3 and X_{11} only achieved significant effect estimates under OLS estimation, with ATEs of −0.278, potentially due to mediating variables [88]. X_4, X_9, X_{12}, and X_{13} were not significant under either estimation strategy, indicating limited evidence for their independent causal effects.

The strong correlation between the OLS and DML estimates (r = 0.962) indicated high concordance between the two methods in identifying the direction and magnitude of effects. DML was particularly effective at controlling high-dimensional confounders and capturing nonlinear features, typically providing more conservative yet robust estimates [75]. For instance, DML amplified the negative effect of the primary treatment variable, X_1, while moderately attenuating the positive effect of X_5. As a parametric baseline method, OLS emphasized interpretability and computational efficiency, providing an intuitive linear

approximation. The complementarity of the two methods offered an empirical basis for understanding the potential causal mechanisms underlying heritage coupling systems.



(c)

Variable	ATE	SE	95%CI	*p*-value	Variable	ATE	SE	95%CI	*p*-value
$X_{1(OLS)}$	−0.345	0.100	(−0.542,−0.148)*	0.001	$X_{1(DML)}$	−0.459	0.141	(−0.735,−0.183)*	0.001
$X_{3(OLS)}$	0.222	0.064	(0.096,0.347)*	0.001	$X_{3(DML)}$	0.133	0.067	(0.001,0.267)*	0.047
$X_{8(OLS)}$	−0.278	0.087	(−0.448,−0.107)*	0.001	$X_{8(DML)}$	−0.131	0.097	(−0.320,0.059)	0.178
$X_{4(OLS)}$	0.067	0.063	(−0.057,0.191)	0.287	$X_{4(DML)}$	0.057	0.071	(−0.081,0.195)	0.418
$X_{5(OLS)}$	−0.240	0.069	(−0.376,−0.104)*	0.001	$X_{5(DML)}$	−0.229	0.088	(−0.401,−0.058)*	0.009
$X_{9(OLS)}$	0.129	0.071	(−0.011,0.268)	0.070	$X_{9(DML)}$	0.088	0.085	(−0.078,0.254)	0.300
$X_{10(OLS)}$	−0.174	0.085	(−0.340,−0.009)*	0.039	$X_{10(DML)}$	−0.192	0.097	(−0.381,−0.002)*	0.047
$X_{11(OLS)}$	−0.278	0.089	(−0.453,−0.103)*	0.002	$X_{11(DML)}$	−0.156	0.100	(−0.351,0.040)	0.118
$X_{12(OLS)}$	−0.047	0.064	(−0.173,0.076)	0.445	$X_{12(DML)}$	0.024	0.060	(−0.094,0.142)	0.688
$X_{13(OLS)}$	0.102	0.152	(−0.197,0.401)	0.503	$X_{13(DML)}$	0.137	0.135	(−0.127,0.401)	0.309

Figure 8. Causal effects of selected variables: comparison of OLS and DML estimates. (**a**) Causal effects of selected variables: OLS and DML estimates; (**b**) comparison of causal effect estimates: OLS and DML methods; (**c**) statistical summary of causal effect estimates. * Indicates statistical significance at the 5% level ($p < 0.05$).

3.5. Heterogeneity Grouping and Differentiated Effects

The top five heterogeneity grouping strategies are presented in Table 4. Among them, the mean annual temperature grouped using K-means clustering ($k = 2$) achieved the highest overall score of 0.9810. This grouping strategy performed well across all three dimensions: statistical significance, effect size, and inter-group differences. The K-means binary grouping effectively identified critical thresholds [89] in temperature variation. This strategy aligned with the climatic differentiation observed in China and was suitable for revealing the differentiated mechanisms of influencing factors across distinct climate zones.

The differentiated robustness testing strategies for the three types of variables provided a credible basis for the heterogeneity analysis. For the primary treatment variables X_1 and X_5, the Δ_{robust} pass rates were 87.5% and 100%, respectively, indicating that their coefficients remained stable after most or all confounding variables were included. The Δ_{norm} values were 0.658 and 0.949, respectively, demonstrating substantial consistency between the two estimation methods. Notably, only X_5 passed the heterogeneity test, confirming its differentiated effect across groups (Table 5).

Table 4. Optimal heterogeneity grouping results.

Grouping Variable	Optimal Grouping Method	*p*-Value	Effect Size	Δμ	max(Δμ)	Composite Score
Annual temperature	K-means (k = 2)	1.71×10^{-13}	0.9789	3.1728	3.2767	0.9810
NDVI	K-means (k = 2)	1.83×10^{-7}	0.9671	3.2767	3.2767	0.9803
Annual temperature	Quartile extreme	1.26×10^{-7}	0.9131	3.5932	3.5932	0.9479
Heritage density	K-means (k = 2)	4.93×10^{-9}	0.9186	3.0819	3.2767	0.9393
Elevation	K-means (k = 2)	4.71×10^{-9}	0.9054	3.0348	3.2767	0.9285

Table 5. Primary treatment variable robustness testing results.

Variable	Δ_{robust} Pass Rate	Δ_{norm}	Heterogeneity
X_1	87.5%	0.658	×
X_5	100%	0.949	✓

Note: ✓indicates heterogeneity between groups; × indicates no significant heterogeneity.

The robustness variables X_3 and X_4 had Δ_{robust} pass rates of 100%, demonstrating high stability after controlling for confounding factors. Their Δ_{norm} values were 1.073 and 0.456, respectively, indicating statistical consistency between the two methods. However, neither variable passed the heterogeneity test (Table 6).

Table 6. Robustness variables robustness testing results.

Variable	Δ_{robust} Pass Rate	Δ_{norm}	Heterogeneity
X_3	100%	1.073	×
X_4	100%	0.456	×

Note: × indicates no significant heterogeneity.

The six exploratory variables exhibited pronounced differences in robustness characteristics (Table 7). According to the differentiated testing strategy, only the variables X_8, X_{10}, and X_{11}, which exhibited significant effects in the baseline model, were subjected to method consistency testing. X_8 exhibited the best overall performance, with a Δ_{norm} value of 0.104 and a Δ_{robust} pass rate of 75%, reflecting strong method consistency and stability. The fact that it passed the heterogeneity test confirmed significant effect differences across groups. For X_{10} and X_{11}, the Δ_{robust} pass rates were 50% and 70%, respectively, and the Δ_{norm} values were 0.149 and 0.938, respectively. However, neither variable passed the heterogeneity test. For the baseline model variables X_9, X_{12}, and X_{13}, the Δ_{robust} values did not exceed 50%, with no heterogeneity observed.

Table 7. Exploratory variable robustness testing results.

Variable	Δ_{robust} Pass Rate	Δ_{norm}	Heterogeneity
X_8	75%	0.104	✓
X_9	50%	—	×
X_{10}	50%	0.149	×
X_{11}	75%	0.938	×
X_{12}	0%	—	×
X_{13}	50%	—	×

Note: Δ_{norm} is calculated only for significant variables. ✓indicates heterogeneity between groups; × indicates no significant heterogeneity.

These results indicated that X_5 and X_8 exhibited significant heterogeneity. Although X_1 did not pass the heterogeneity test, its significant effect and Δ_{robust} pass rate of 87.5% suggested the possibility of a more complex pattern. Based on K-means clustering (k = 2) of the mean annual temperature, we focused on the heterogeneity mechanisms of X_5 and X_8 while retaining an exploratory analysis of X_1 (Figure 9).

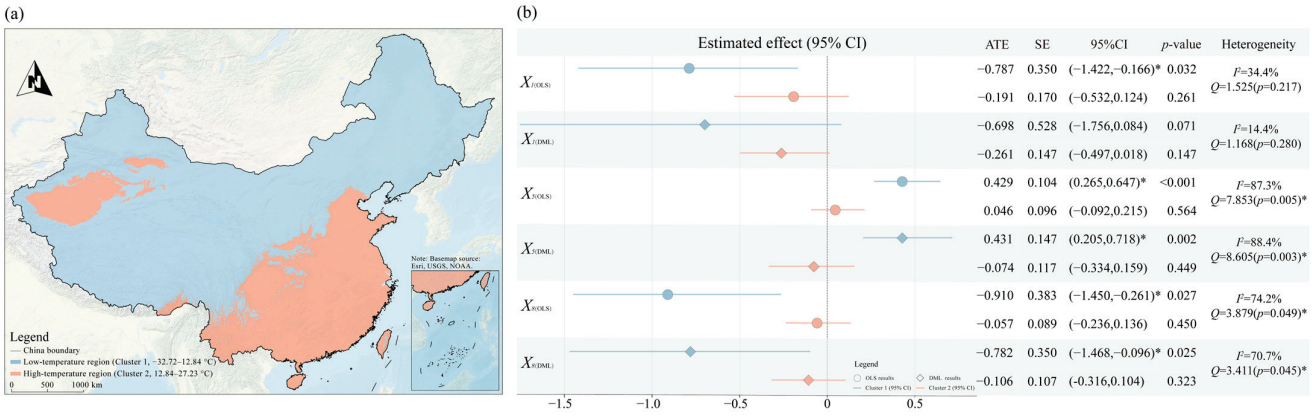


Figure 9. Heterogeneity estimation under mean annual temperature grouping. (**a**) Grouping based on mean annual temperature; (**b**) forest plot comparing OLS and DML heterogeneous effects. * Indicates statistical significance at the 5% level ($p < 0.05$).

X_1 not only served as the basis for grouping but also exhibited differences in its effect on coupling. The OLS estimates indicated a significant negative effect estimate in low-temperature areas (ATE = −0.787), whereas the effect in high-temperature areas was nonsignificant. By contrast, the DML approach showed that neither region reached conventional significance levels. The I^2 values were 34.4% and 14.4% for the two methods, respectively, and the Q test did not detect any significant differences between the groups. This result suggested that temperature may exert a nonlinear influence on heritage coupling, although the evidence remains insufficient under the current dataset.

X_5 exhibited the strongest heterogeneity. Both methods showed a significant positive effect estimate in the low-temperature region (ATE = 0.429 and ATE = 0.431), while the effect was not significant in the high-temperature region. The I^2 values were 87.3% and 88.4%, and the Q test indicated very high between-group heterogeneity. These results suggested that well-developed transport and waterway networks may have been key to overcoming climatic constraints in low-temperature regions [90], facilitating the flow and integration of cultural elements. By contrast, in high-temperature regions, where climatic conditions were relatively favorable, the marginal contribution of transportation and waterways was markedly reduced.

X_8 also exhibited significant heterogeneity. Both methods showed a significant negative effect estimate in the low-temperature region (ATE = −0.910 and ATE = −0.782), whereas the effect was not significant in the high-temperature region. The I^2 values were 74.2% and 70.7%, and the Q test was significant. These findings indicated that the impact of precipitation on coupling was temperature-dependent. Precipitation heterogeneity reflected the role of climate in shaping the spatial patterns of agricultural civilizations and cultural forms [91]. In low-temperature regions, excessive precipitation may have suppressed heritage linkages through mechanisms such as hindering transportation and constraining social activities, whereas this suppressive effect was no longer significant in high-temperature regions.

It should be noted that heterogeneity identification depends on the grouping method and testing criteria. Inter-group inconsistencies represent statistical evidence obtained under specific strategies and do not completely negate underlying mechanisms. These results revealed the complex regulatory role of geographical factors in the cultural heritage coupling system under different temperature conditions. Temperature was found to directly affect heritage coupling and constrain the influence of other factors. Both X_5 and X_8 exerted significant effects in low-temperature regions, while their impacts weakened in high-temperature regions. This temperature-centered environmental regulation mechanism

demonstrated that the influence of geographical factors on cultural heritage coupling did not operate independently but instead formed a differentiated network of interactive effects under specific environmental contexts.

4. Discussion

4.1. Interpretation of Driving Mechanisms from HBM and Causal Evidence

This study reveals the multi-scale driving mechanisms of spatial coupling between TCH and ICH in China by integrating HBM and causal inference methods. The progressive evidence chain, from correlation to model-based causal effect estimates and heterogeneity, provides multiple empirical foundations for understanding the complexity of heritage systems. Figure 10 illustrates the spatial distribution of the five core variables, facilitating an understanding of their geographical associations with degrees of coupling.

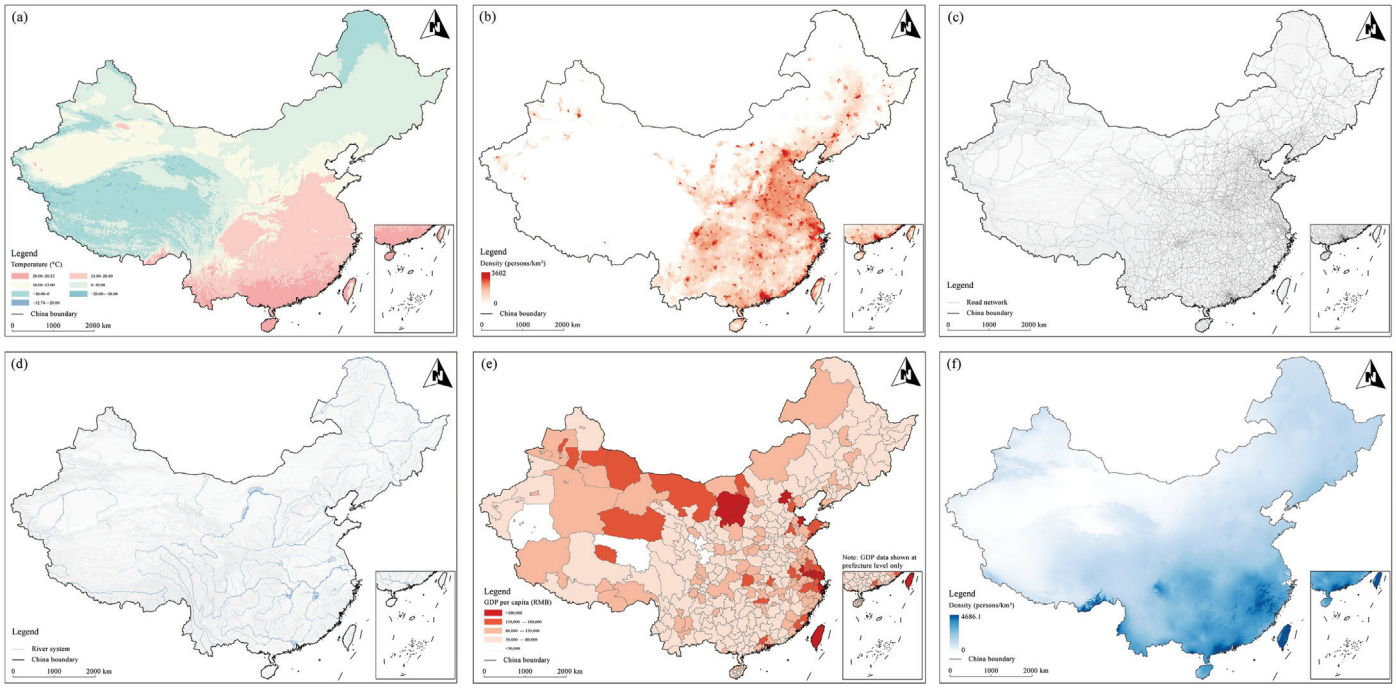


Figure 10. Spatial distribution of driving variables. (**a**) Mean annual temperature; (**b**) population density; (**c**) road network distribution; (**d**) water system distribution; (**e**) per capita GDP; (**f**) mean annual precipitation.

The significant negative effect estimate of X_1 indicates higher coupling intensity in low-temperature regions. Although the heterogeneity does not reach a statistically significant level, this result still reflects the nonlinear role of temperature. In China, temperature exhibits clear latitudinal and altitudinal gradients, with southern regions generally warmer than vast areas in the west and northeast. This spatial correspondence suggests that temperature, as a fundamental environmental variable, shapes heritage patterns by influencing human activity and cultural practices [92]. The underlying mechanism may lie in the tendency of low-temperature environments to form relatively enclosed geographic units, which consolidate community structures and cultural identities [93], thereby facilitating the co-transmission of both types of heritage. In the context of global warming, high temperatures may weaken the spatial linkage between heritage types by reshaping traditional lifestyles and increasing environmental pressures [94]. Notably, mean annual temperature also emerges as the optimal variable for heterogeneity grouping, supporting the foundational role of climatic conditions in heritage coupling systems.

X_3 and its interaction with per capita GDP (X_4) both exert significant negative effect estimates. Conventionally, areas with high population density often host rich heritage resources. For instance, migration and urban expansion in Jordan have shaped the multi-layered evolution of As-Salt's architecture and cultural landscape [95]. However, our results indicate that in regions of high population concentration and economic prosperity in China, heritage coupling intensity tends to be lower. Using the 'Hu Huanyong Line' to delineate population density patterns, we observe that the eastern and southeastern regions display lower overall coupling than the western and northwestern regions. This decoupling phenomenon suggests that cultural transmission in modern societies is challenging, highlighting the need for integrated protection. Rapid urbanization may accelerate the spatial separation and cultural decoupling of the two heritage types by dismantling local cultural contexts [27]. TCH may relocate or disappear due to urban renewal, while ICH may lose its continuity as the linkage between traditional communities and heritage spaces weakens. In contrast to the inhibitory effect of X_4 on coupling, the interaction between per capita GDP and tertiary industry share (X_6) exhibits a significant positive effect estimate. Studies on cultural tourism coupling in the Beijing–Tianjin–Hebei urban agglomeration indicate that industrial restructuring and digital infrastructure are key drivers of integration [96]. This suggests that when economic development is accompanied by optimized heritage tourism and creative cultural industries, the heritage value chain can be restructured to actively connect both types of heritage. Such a shift implies that heritage protection can transform from a mere cost burden into a value-creating process, providing market-based pathways to support the coordinated development of heritage.

X_5 exhibits a robust positive effect estimate, showing significant heterogeneity between low- and high-temperature regions. Transportation networks and river systems form a collaborative network that serves as critical infrastructure for cultural exchange in relatively isolated cold environments, acting as traditional channels of cultural dissemination. This is essential for maintaining cultural connections between dispersed communities and heritage sites. By contrast, this effect is not significant in high-temperature regions with greater openness, suggesting that cultural transmission in these areas relies more on other mechanisms, such as tourism or institutional protection. Notably, river density (X_{10}) alone demonstrates a negative effect estimate in causal tests, indicating that overly dense river networks, in the absence of adequate transportation infrastructure, may isolate communities from heritage sites [97], thereby weakening the coupling degree. However, when interacting with road networks, it exhibits a significant positive effect estimate. Studies on the coupling between village cultural landscapes and socio-economic systems demonstrate that improving transportation accessibility facilitates the dual objectives of heritage protection and economic development [98]. The pattern of individual inhibition but interactive promotion highlights that road networks can mitigate the spatial fragmentation caused by river systems, thereby supporting the maintenance and dissemination of cultural activities and ultimately reinforcing the integration of TCH and ICH. The construction of interaction variables followed the collaborative logic of scientific elements based on indicator validation, allowing the estimation of synergistic effects that exceed individual contributions. This ensures that the causal evidence represents not only statistical signals but also underlying environmental linkages.

X_8 has a significant negative effect estimate, with significant heterogeneity. The negative impact of precipitation is significant in low-temperature regions, likely because the combination of cold and high precipitation creates a harsh natural environment that amplifies geographical isolation. Concentrated rainfall and spring floods can cause structural degradation of heritage in cold areas and prolong restoration cycles [99], disrupting the heritage transmission chain. Frequent precipitation and flooding make heritage protection

more difficult and hinder human mobility and cultural exchange, ultimately reducing cross-regional cultural integration. Conversely, this effect is not significant in high-temperature regions, which may benefit from long-established adaptive mechanisms such as moisture-resistant architecture and water transportation systems in southern areas. These differentiated effects reflect how environmental constraints translate into cultural outcomes [100]. Distinct cultural systems, such as those of Jiangnan and Lingnan, have evolved in warm and humid conditions, exemplifying human societies' adaptive transformation of nature.

4.2. Differentiated Governance Strategies for Heritage Protection

Multi-level evidence provides valuable insights into heritage conservation and regional governance (Table 8). Given the central role of climate and its moderating effects, differentiated, climate-adaptive conservation strategies should be developed based on coupling relationships. Heritage impact mechanisms exhibit significant differentiation across varying temperature contexts, as stone and timber materials demonstrate different material-specific thresholds to climatic stress [101]. Combining the positive effects of road–water interaction with the negative effects of water system density, water management should proceed in parallel with accessibility infrastructure development. In low-temperature regions, the synergistic effects of road–water interconnection should be fully utilized, with integrated transport projects promoting the clustered development of heritage sites within cultural corridors [102]. In areas with dense water systems but limited land transport, priority should be given to improving terrestrial infrastructure or enhancing waterway services. This helps to prevent spatial fragmentation from reducing coupling degrees. Key improvements should focus on enhancing the landscapes of vernacular communities and making heritage more accessible, as well as on strengthening the interaction between heritage and daily life. In high-temperature or climate-vulnerable areas, priority should be given to heritage risk assessment and protective restoration. At the same time, cultural transmission mechanisms or digital platforms should be established to reduce the impacts of climatic stress on coupling degrees.

Regarding the negative effects in densely populated and economically developed areas, excessive urbanization leading to cultural decoupling should be prevented. Heritage conservation buffer zones should be incorporated into urban renewal and land planning to maintain connections between heritage and daily life [103]. Cultural activity circles or subsidy mechanisms should be established to encourage local culture and traditional crafts. Simultaneously, heritage impact assessments and cultural space protection clauses can be incorporated into urban construction to prevent heritage spatial separation caused by development. The positive effects of economic structural transition toward tertiary industries indicate that heritage conservation can be transformed from cost consumption into value generation through the development of cultural industries and heritage tourism [104]. Cultural industrial parks can be established to promote spatial clustering and reproduction while mitigating the risks of homogenization caused by over-commercialization.

Differentiated governance frameworks are crucial for formulating and implementing heritage conservation policy [105]. Prefecture-level scales are suitable for undertaking specific infrastructure construction and community cultural conservation projects. Provincial scales can play coordinating roles across administrative regions and integrate resources. Cultural–geographical unit scales should respect the authenticity of cultural ecosystems by conducting conservation and transmission planning centered on cultural spaces. Thus, a three-tier indicator system encompassing 'risk assessment–accessibility–socioeconomic structure' is recommended, with performance indicators such as coupling degree changes and cultural activity participation rates established for process management and evaluation. Through multi-scale collaborative governance mechanisms, regional cultural balanced

development and factor sharing can be promoted more effectively, facilitating heritage activation and socioeconomic synergy.

Table 8. Key findings and policy recommendations.

Variable and Effect Direction	Theoretical Mechanism	Policy Recommendations	Implementation Scale
X_1: Negative	Low temperature enhances geographic enclosure and cultural cohesion	Integrate climate adaptation into heritage planning; prioritize protection in cold regions	Provincial/cultural–geographical unit
X_3: Negative	Population agglomeration accelerates urbanization and deconstructs vernacular culture	Embed heritage buffer zones in urban areas; conduct heritage impact assessments	Prefecture level
X_4: Negative	Compounded effects of economy and population exacerbate cultural decoupling risks	Prevent cultural decoupling from aggressive urbanization; establish holistic protection mechanisms	Prefecture–provincial combined
X_5: Positive	Infrastructure complementarity generates synergistic effects	Develop cultural–tourism transport corridors; prioritize road connectivity improvements	Provincial/cultural–geographical unit
X_6: Positive	Industrial restructuring enhances heritage value chains	Promote tourism and creative industries; balance commercialization with authenticity protection	Prefecture/provincial
X_8: Negative	Low temperature and precipitation compound geographic isolation and structural degradation	Strengthen moisture and flood prevention facilities in cold-humid regions; establish adaptive protection systems	Provincial/cultural–geographical unit
X_{10}: Negative	Road networks can mitigate spatial fragmentation caused by dense river networks	Coordinate waterway management with accessibility improvements	Prefecture/provincial

5. Conclusions

This study established an integrated analytical framework for coupling degree assessment that combines multi-scale geographic units, Bayesian modeling, and causal inference. A coupling degree measurement system for heritage was constructed at the prefecture-level, provincial, and cultural–geographical unit scales, based on the MCLQ index. The key driving variables were then identified using the HBM. Bayesian statistics and causal identification tools were employed together to screen potential variables and categorize them hierarchically. Subsequently, the model-based causal effects were estimated using a differentiated hybrid framework. Following robustness checks, optimal grouping methods were applied to capture heterogeneity in factor effects across groups, and the spatial distribution patterns of heritage coupling and the heterogeneity of its driving mechanisms in China were systematically examined. The main empirical findings are as follows:

(1) Heritage coupling patterns showed scale dependence and cultural cohesion. The proportion of strongly coupled areas decreased from the prefecture-level to the provincial scale, consistent with typical spatial attenuation patterns, but rose noticeably at the cultural–geographical unit scale. This anomaly revealed a cohesive effect for China's cultural system that transcended administrative boundaries.

(2) The HBM identified five critical drivers and their differentiated impacts. X_1, X_3, and X_4 exhibited significant negative effects, whereas X_5 and X_6 showed significant positive effects. Except for X_4, the effect intensity of the remaining four variables peaked at the prefecture-level scale and generally diminished at the provincial scale. This suggested the presence of a suppressive mechanism at the provincial level, which attenuated the effects of these factors.

(3) The hybrid causal inference framework provided model-based causal estimation evidence for four key variables. Through dual testing with OLS and DML verification, X_1,

X_8, and X_{10} displayed significant negative causal effect estimates, while X_5 exhibited a significant positive causal effect estimate.

(4) Heterogeneity analysis detected differentiated causal effects for two variables. In K-means clustering (k = 2) based on annual average temperature, the positive effect estimate of X_5 and the negative effect estimate of X_8 were significant in the low-temperature group but not in the high-temperature group. This indicated that temperature conditions moderated the driving mechanisms of heritage coupling.

By organically integrating data-driven HBM with logic-driven causal inference and incorporating a wide range of geographic and social variables, this study uncovers multiple driving mechanisms of heritage coupling and their heterogeneity. However, several empirical limitations exist regarding causal interpretation. Firstly, some factors involving potential mediating pathways could not be incorporated into the hybrid framework. Secondly, it was not possible to continuously explore the temporal evolution of the heritage coupling system due to constraints in historical data. Spatial units partially reflect institutional and social processes rather than purely neutral natural variables, and thus cannot be completely distinguished from cultural dynamics. Furthermore, some instrumental variables with limited strength suggest potential weak instrument bias. This study presents optimal estimates based on the current model assumptions and data conditions rather than definitive causal relationships obtained under experimental conditions. Future research could employ quasi-natural experimental designs, spatiotemporal panels and multi-scenario grouping strategies to improve identification capabilities. Structural equation modeling could also be applied to clarify theoretical causal pathways and mediation testing. The analytical framework could incorporate qualitative elements such as participatory geographic information systems, policies, and community behaviors, thereby enabling a deeper exploration of the differences and integration pathways between official recognition and local needs. From a practical perspective, this study primarily aimed to validate and deepen causal understanding of the intrinsic drivers of heritage coupling, thus providing a methodological reference for the conservation of cultural heritage at multiple scales.

Author Contributions: Conceptualization, Y.L.; data curation, Y.L., X.D. and Y.B.; methodology, Y.L. and X.D.; software, Y.L.; formal analysis, Y.L. and X.D.; investigation, Y.L., X.D. and Y.B.; validation, Y.L., X.D. and S.L.; visualization, Y.L.; supervision, Q.C. and X.D.; writing—original draft preparation, Y.L., Y.B. and S.L. writing—review and editing, Y.L., X.D. and Y.B.; project administration, Q.C.; funding acquisition, Q.C.; Y.L. and X.D. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China [32271944]. and the Sichuan Philosophy and Social Science Foundation Program [SCJJ24ND261].

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

Acknowledgments: We gratefully acknowledge the data support from the National Tibetan Plateau/Third Pole Environment Data Center for providing the temperature index dataset (http://data.tpdc.ac.cn, accessed on 6 March 2025) and from the Lake–Watershed Science Data Center, National Earth System Science Data Center, National Science & Technology Infrastructure of China, for providing the population dataset (http://lake.geodata.cn, accessed on 7 March 2025).

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

Abbreviations

The following abbreviations are used in this manuscript:

TCH	Tangible Cultural Heritage
ICH	Intangible Cultural Heritage
CLQ	Co-location Quotient
HBM	Hierarchical Bayesian Model
DEM	Digital Elevation Model
NDVI	Normalized Difference Vegetation Index
IQR	Interquartile Range
MCLQ	Modified Co-Location Quotient
MCMC	Markov Chain Monte Carlo
$\hat{R}$	Gelman–Rubin Statistic
MCSE	Monte Carlo Standard Error
R^2	Coefficient of Determination
RMSE	Root Mean Square Error
BPV	Bayesian *p*-Value
MAE	Mean Absolute Error
KNN	K-Nearest Neighbor
DML	Double Machine Learning
OLS	Ordinary Least Squares
HAC	Heteroscedasticity and Autocorrelation Consistent
ATE	Average Treatment Effect
HDI	Highest Density Interval

References

1. Shen, W.; Chen, Y.L.; Cao, W.W.; Yu, R.Y.; Rong, P.J.; Cheng, J.L. Spatial pattern and its influencing factors of national-level cultural heritage in China. *npj Herit. Sci.* **2024**, *12*, 384. [CrossRef]
2. Malik, U.S.; Micoli, L.L.; Caruso, G.; Gabriele, G. Integrating quantitative and qualitative analysis to evaluate digital applications in museums. *J. Cult. Herit.* **2023**, *62*, 304–313. [CrossRef]
3. Guttormsen, T.S.; Skrede, J.; Guzman, P.; Fouseki, K.; Bonacchi, C.; Pastor Pérez, A. Assemblage urbanism: The role of heritage in urban placemaking. *J. Cult. Herit. Manag. Sustain. Dev.* **2023**, *ahead-of-print*. [CrossRef]
4. Taylor, K. The Historic Urban Landscape paradigm and cities as cultural landscapes. Challenging orthodoxy in urban conservation. *Landsc. Res.* **2016**, *41*, 471–480. [CrossRef]
5. Monteiro, V.; Painho, M.; Vaz, E. Is the heritage really important? A theoretical framework for heritage reputation using citizen sensing. *Habitat Int.* **2015**, *45*, 156–162. [CrossRef]
6. Ferretti, V.; Comino, E. An integrated framework to assess complex cultural and natural heritage systems with Multi-Attribute Value Theory. *J. Cult. Herit.* **2015**, *16*, 688–697. [CrossRef]
7. Ripp, M.; Gustafsson, C.; Jones, Z.; Majumdar, S.; Ginzarly, M. Historic Urban Landscapes and Heritage Systems as the Basis for Sustainable Urban Development. *Land* **2025**, *14*, 1783. [CrossRef]
8. Taylor, K.; Lennon, J. Cultural landscapes: A bridge between culture and nature? *Int. J. Herit. Stud.* **2011**, *17*, 537–554. [CrossRef]
9. Zhang, Z.W.; Cui, Z.; Fan, T.S.; Ruan, S.Y.; Wu, J.M. Spatial distribution of intangible cultural heritage resources in China and its influencing factors. *Sci. Rep.* **2024**, *14*, 4960. [CrossRef] [PubMed]
10. Xue, Q.M.; Huang, Y.H. The spatial relationship and influence mechanism of traditional villages and intangible cultural heritage: A case study of the upper reaches of the Yellow River Basin. *Hum. Soc. Sci. Commun.* **2025**, *12*, 142. [CrossRef]
11. Li, C.; Qian, Y.Y.; Li, Z.K.; Tong, T. Identifying factors influencing the spatial distribution of minority cultural heritage in Southwest China. *npj Herit. Sci.* **2024**, *12*, 117. [CrossRef]
12. Ma, Q.B.; Liu, T.T.; Yang, Z.; Zhang, S.; Zhang, X.Y.; Sun, M.N.; Hou, X.X. The spatio-temporal distribution and influencing factors of cave temples and stone carvings in China. *npj Herit. Sci.* **2025**, *13*, 311. [CrossRef]
13. Feng, Y.; Wei, H.; Huang, Y.; Li, J.W.; Mu, Z.Q.; Kong, D.Z. Spatiotemporal evolution characteristics and influencing factors of traditional villages: The Yellow River Basin in Henan Province, China. *npj Herit. Sci.* **2023**, *11*, 97. [CrossRef]
14. Liu, E.L.; Zhang, J.; Chi, H.N. Spatial distribution and driving factors of Chinese dragon and lion dance athletes based on GIS. *Heliyon* **2024**, *10*, e23409. [CrossRef] [PubMed]
15. Dong, B.L.; Bai, K.; Sun, X.L.; Wang, M.T.; Liu, Y. Spatial distribution and tourism competition of intangible cultural heritage: Take Guizhou, China as an example. *npj Herit. Sci.* **2023**, *11*, 64. [CrossRef]

16. Chen, W.X.; Yang, Z.; Yang, L.Y.; Wu, J.H.; Bian, J.J.; Zeng, J.; Liu, Z.L. Identifying the spatial differentiation factors of traditional villages in China. *npj Herit. Sci.* **2023**, *11*, 149. [CrossRef]
17. Jiang, L.W.; Cai, J.; Fan, Y.L. Spatiotemporal Distribution and Heritage Corridor Construction of Vernacular Architectural Heritage in the Cao'e River, Jiaojiang River, and Oujiang River Basin. *Land* **2025**, *14*, 1484. [CrossRef]
18. Zhang, W.Y.; Ma, L.H.; Xu, J.Y.; Peng, Y.J. Spatial–temporal distribution characteristics and influencing factors of immovable cultural heritage in Hubei, China. *Sci. Rep.* **2025**, *15*, 24570. [CrossRef]
19. Zhang, J.H.; Jiang, L.; Wang, X.Q.; Chen, Z.L.; Xu, S.B. A Study on the Spatiotemporal Aggregation and Corridor Distribution Characteristics of Cultural Heritage: The Case of Fuzhou, China. *Buildings* **2024**, *14*, 121. [CrossRef]
20. Zhu, L.T.; Li, Z.W.; Su, H.M.; Wang, X. Temporal and spatial distribution of ancient sites in Shaanxi Province using geographic information systems (GIS). *npj Herit. Sci.* **2021**, *9*, 121. [CrossRef]
21. Laguna-Palma, D.; Palacios, O.; Mokránová, K. Assessing the risks of northeastern African archaeological heritage and their relationship to human–environmental processes: A Bayesian network approach. *World Archaeol.* **2023**, *55*, 366–381. [CrossRef]
22. Zhang, S.R.; Chi, L.; Zhang, T.Y.; Ju, H.R. Spatial pattern and influencing factors of land border cultural heritage in China. *npj Herit. Sci.* **2023**, *11*, 187. [CrossRef]
23. Liu, W.X.; Xue, Y.; Shang, C. Spatial distribution analysis and driving factors of traditional villages in Henan province: A comprehensive approach via geospatial techniques and statistical models. *npj Herit. Sci.* **2023**, *11*, 185. [CrossRef]
24. Pang, L.; Wu, L.N. Distribution characteristics and influencing factors of Intangible Cultural Heritage in Beijing-Tianjin-Hebei. *npj Herit. Sci.* **2023**, *11*, 19. [CrossRef]
25. Liu, S.; Ge, J.; Bai, M.; Yao, M.; Zhu, Z.N. Uncovering the factors influencing the vitality of traditional villages using POI (point of interest) data: A study of 148 villages in Lishui, China. *npj Herit. Sci.* **2023**, *11*, 123. [CrossRef]
26. Du, Y.X.; Chen, L.; Xu, J.Y. Interactive effects of intangible cultural heritage and tourism development: A study based on the data panel PVAR model and coupled coordination model. *npj Herit. Sci.* **2024**, *12*, 401. [CrossRef]
27. Yan, L.; Shen, X.; Liu, P.L. Spatiotemporal evolution of the coupling coordination between inheritance of local culture and new-style urbanization in China. *npj Herit. Sci.* **2022**, *10*, 73. [CrossRef]
28. Huser, R.; Opitz, T.; Wadsworth, J.L. Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes. *Environ. Data Sci.* **2025**, *4*, e3. [CrossRef]
29. Fiorucci, M.; Khoroshiltseva, M.; Pontil, M.; Traviglia, A.; Del Bue, A.; James, S. Machine Learning for Cultural Heritage: A Survey. *Pattern Recognit. Lett.* **2020**, *133*, 102–108. [CrossRef]
30. Zhuang, Y. *Environmental Foundations to the Rise of Early Civilisations in China*; Cambridge University Press: Cambridge, UK, 2024; pp. 23–79.
31. Wang, F.L.; Zhang, T.Y.; Zhang, S.R. MGWR reveals scale heterogeneity shaping intangible cultural heritage distribution in China. *npj Herit. Sci.* **2025**, *13*, 367. [CrossRef]
32. Guo, Y.; Yan, J.Y.; Wang, Z.B.; Li, J.X. Spatial Dataset of 3610 Items of China's National Intangible Cultural Heritage in Five Packages. *J. Glob. Change Data Discov.* **2022**, *6*, 257–264. [CrossRef]
33. Dong, L.B.; Bettinger, P.; Liu, Z.G. Stand spatial structural diversity: Developing and validating a novel index. *For. Ecol. Manag.* **2024**, *569*, 122157. [CrossRef]
34. Ma, W.; Zhao, C. *China Extreme and Average Temperature Index Dataset (1961–2012)*; National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2024. [CrossRef]
35. Peng, S.Z.; Ding, Y.X.; Liu, W.Z.; Li, Z. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [CrossRef]
36. Dong, J.W.; Zhou, Y.; You, N.S.; Cheng, C.; Yang, J.L. *China 30 m Annual Maximum NDVI Dataset from 2000 to 2022*; National Ecosystem Science Data Center: Beijing, China, 2021. (In Chinese) [CrossRef]
37. Yang, J.; Dong, J.; Xiao, X.; Dai, J.; Wu, C.; Xia, J.; Zhao, G.; Zhao, M.; Li, Z.; Zhang, Y.; et al. Divergent shifts in peak photosynthesis timing of temperate and alpine grasslands in China. *Remote Sens. Environ.* **2019**, *233*, 111395. [CrossRef]
38. WorldPop; Columbia University. *Global High Resolution Population Denominators Project—Funded by The Bill and Melinda Gates Foundation (OPP1134076)*; University of Southampton: Southampton, UK, 2018. [CrossRef]
39. Tawakuli, A.; Havers, B.; Gulisano, V.; Kaiser, D.; Engel, T. Survey: Time-series data preprocessing: A survey and an empirical analysis. *J. Eng. Res.* **2025**, *13*, 674–711. [CrossRef]
40. Goswami, T.; Sinha, G.R. *Statistical Modeling in Machine Learning: Concepts and Applications*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 287–306. [CrossRef]
41. Atkinson, A.C.; Riani, M.; Corbellini, A. The Box–Cox Transformation: Review and Extensions. *Stat. Sci.* **2021**, *36*, 239–255. [CrossRef]
42. Demircioğlu, A. The effect of feature normalization methods in radiomics. *Insights Imaging* **2024**, *15*, 2. [CrossRef]
43. Rosik, P.; Wójcik, J. Transport infrastructure and regional development: A survey of literature on wider economic and spatial impacts. *Sustainability* **2022**, *15*, 548. [CrossRef]

44. Crits-Christoph, P.; Gallop, R.; Gaines, A.; Rieger, A.; Connolly Gibbons, M.B. Instrumental variable analyses for causal inference: Application to multilevel analyses of the alliance–outcome relation. *Psychother. Res.* **2018**, *30*, 53–67. [CrossRef]
45. Ginzarly, M.; Houbart, C.; Teller, J. The Historic Urban Landscape approach to urban management: A systematic review. *Int. J. Herit. Stud.* **2018**, *25*, 999–1019. [CrossRef]
46. Li, Y.F.; Li, Y.; Zhou, Y.; Shi, Y.L.; Zhu, X.D. Investigation of a coupling model of coordination between urbanization and the environment. *J. Environ. Manag.* **2012**, *98*, 127–133. [CrossRef] [PubMed]
47. Diener, A.C.; Hagen, J. Geographies of place attachment: A place-based model of materiality, performance, and narration. *Geogr. Rev.* **2020**, *112*, 171–186. [CrossRef]
48. Pan, Z.; Chen, Y.L.; Huang, Y.H.; Zheng, L. The vernacular cultural landscape in traditional villages: Global hotspots, emerging trends, and a case study of China's Qilu cultural district. *Front. Earth Sci.* **2024**, *12*, 1511292. [CrossRef]
49. Liu, B.B.; Wu, C.; Xu, W.X.; Shen, Y.N.; Tang, F.L. Emerging trends in GIS application on cultural heritage conservation: A review. *npj Herit. Sci.* **2024**, *12*, 139. [CrossRef]
50. Leslie, T.F.; Kronenfeld, B.J. The Colocation Quotient: A New Measure of Spatial Association Between Categorical Subsets of Points. *Geogr. Anal.* **2011**, *43*, 306–326. [CrossRef]
51. Xie, J.F.; Zhang, J.; Yang, Z.N.; Yu, T.F.; Xie, Y. Characterizing the spatial organization of urban functions using a distance-based colocation quotient approach: With special reference to Beijing. *Habitat Int.* **2025**, *163*, 103474. [CrossRef]
52. Pierce, J.; Martin, D.G.; Murphy, J.T. Relational place-making: The networked politics of place. *Trans. Inst. Br. Geogr.* **2010**, *36*, 54–70. [CrossRef]
53. Yu, Y.; Wang, J. Crouching the tiger, or hiding the dragon? scale in China's heritage production. *Built Herit.* **2024**, *8*, 11. [CrossRef]
54. Tian, Y.S.; Kong, X.S.; Liu, Y.L. Combining weighted daily life circles and land suitability for rural settlement reconstruction. *Habitat Int.* **2018**, *76*, 1–9. [CrossRef]
55. Li, B.L.; Lu, Y.F.; Li, Y.D.; Zuo, H.Y.; Ding, Z.Q. Research on the spatiotemporal distribution characteristics and accessibility of traditional villages based on geographic information systems—A case study of shandong province, China. *Land* **2024**, *13*, 1049. [CrossRef]
56. Wang, L.; Zhang, F.F.; Zang, Y.Z.; Duan, J. Understanding the Regional Integration Process from the Perspective of Agglomeration and Urban Networks: Case Study in Central China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12834. [CrossRef] [PubMed]
57. Su, H.R.; Wang, Y.W.; Zhang, Z.; Dong, W. Characteristics and Influencing Factors of Traditional Village Distribution in China. *Land* **2022**, *11*, 1631. [CrossRef]
58. Kaplan, N.; Omer, I. Multiscale Accessibility—A New Perspective of Space Structuration. *Sustainability* **2022**, *14*, 5119. [CrossRef]
59. Wang, M.Y.; Pan, S.J.X.; Tao, Y.Q.; Sun, H.L.; Li, X.Y. Hierarchical Bayesian modelling of quasi-region-specific soil porosity. *Ocean Eng.* **2022**, *266*, 113052. [CrossRef]
60. Jia, X.Y.; Sedehi, O.; Papadimitriou, C.; Katafygiotis, L.S.; Moaveni, B. Nonlinear model updating through a hierarchical Bayesian modeling framework. *Comput. Meth. Appl. Mech. Eng.* **2022**, *392*, 114646. [CrossRef]
61. Yu, D.L.; Fang, C.L. How neighborhood characteristics influence neighborhood crimes: A Bayesian hierarchical spatial analysis. *Int. J. Environ. Res. Public Health.* **2022**, *19*, 11416. [CrossRef]
62. Rittirong, J.; Bryant, J.; Aekplakorn, W.; Prohmmo, A.; Sunpuwan, M. Obesity and occupation in Thailand: Using a Bayesian hierarchical model to obtain prevalence estimates from the National Health Examination Survey. *BMC Public Health* **2021**, *21*, 914. [CrossRef]
63. Veenman, M.; Stefan, A.M.; Haaf, J.M. Bayesian hierarchical modeling: An introduction and reassessment. *Behav. Res. Methods* **2023**, *56*, 4600–4631. [CrossRef]
64. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [CrossRef]
65. Polson, N.G.; Scott, J.G. On the half-Cauchy prior for a global scale parameter. *Bayesian Anal.* **2012**, *7*, 887–902. [CrossRef]
66. Roy, V. Convergence Diagnostics for Markov Chain Monte Carlo. *Annu. Rev. Stat. Appl.* **2020**, *7*, 387–412. [CrossRef]
67. Dias, F.; Suhadolnik, N.; Camargo, H.; Da Silva, S. Predicting the pulse of the Amazon: Machine learning insights into deforestation dynamics. *J. Environ. Manag.* **2024**, *362*, 121359. [CrossRef] [PubMed]
68. Karunasingha, D.S.K. Root mean square error or mean absolute error? Use their ratio as well. *Inf. Sci.* **2022**, *585*, 609–629. [CrossRef]
69. Martin, O. *Bayesian Analysis with Python: A Practical Guide to Probabilistic Modeling*; Packt Publishing Ltd.: Birmingham, UK, 2024; pp. 152–154.
70. Ahmad, S.; Batool, A.; Ali, Z. Spatial predictive analysis of drought duration in relation to climate change using interpolation techniques. *Stoch. Environ. Res. Risk Assess.* **2025**, *39*, 639–656. [CrossRef]
71. Anselin, L.; Syabri, I.; Kho, Y. *GeoDa: An introduction to Spatial Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 73–89. [CrossRef]

72. Anselin, L. *An Introduction to Spatial Data Science with GeoDa: Volume 1: Exploring Spatial Data*; Chapman and Hall/CRC: New York, NY, USA, 2024; pp. 176–388. [CrossRef]
73. Baiardi, A.; Naghi, A.A. The value added of machine learning to causal inference: Evidence from revisited studies. *Econom. J.* **2024**, *27*, 213–234. [CrossRef]
74. Hünermund, P.; Louw, B.; Caspi, I. Double machine learning and automated confounder selection: A cautionary tale. *J. Causal Inference* **2023**, *11*, 20220078. [CrossRef]
75. Moccia, C.; Moirano, G.; Popovic, M.; Pizzi, C.; Fariselli, P.; Richiardi, L.; Ekstrøm, C.T.; Maule, M. Machine learning in causal inference for epidemiology. *Eur. J. Epidemiol.* **2024**, *39*, 1097–1108. [CrossRef]
76. Kelejian, H.H.; Prucha, I.R. HAC estimation in a spatial framework. *J. Econom.* **2007**, *140*, 131–154. [CrossRef]
77. Fitzenberger, B.; Mena, G.; Nimczik, J.; Sunde, U. Personality traits across the life cycle: Disentangling age, period and cohort effects. *Econ. J.* **2022**, *132*, 2141–2172. [CrossRef]
78. Borenstein, M. Avoiding common mistakes in meta-analysis: Understanding the distinct roles of Q, I-squared, tau-squared, and the prediction interval in reporting heterogeneity. *Res. Synth. Methods* **2023**, *15*, 354–368. [CrossRef] [PubMed]
79. Lu, X.; White, H. Robustness checks and robustness tests in applied economics. *J. Econom.* **2014**, *178*, 194–206. [CrossRef]
80. Wang, F.F.; Chen, X.; Xue, F. A review of Bayesian spatiotemporal models in spatial epidemiology. *ISPRS Int. J. Geo Inf.* **2024**, *13*, 97. [CrossRef]
81. De, C.C.; d'Haultfoeuille, X. Two-way fixed effects estimators with heterogeneous treatment effects. *Am. Econ. Rev.* **2020**, *110*, 2964–2996. [CrossRef]
82. MacKinno, J.G.; Nielsen, M.Ø.; Webb, M.D. Cluster-robust inference: A guide to empirical practice. *J. Econom.* **2023**, *232*, 272–299. [CrossRef]
83. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Front. Psychol.* **2013**, *4*, 863. [CrossRef] [PubMed]
84. Maher, J.M.; Markey, J.C.; Ebert-May, D. The Other Half of the Story: Effect Size Analysis in Quantitative Research. *CBE Life Sci. Educ.* **2013**, *12*, 345–351. [CrossRef]
85. Qian, X.Y.; Yu, Y.; Liu, R.J. The Spatial Relationship Characteristics and Differentiation Causes Between Traditional Villages and Intangible Cultural Heritage in China. *Buildings* **2025**, *15*, 2094. [CrossRef]
86. Kamath, A.; Poojari, S.; Varsha, K. Assessing the robustness of normality tests under varying skewness and kurtosis: A practical checklist for public health researchers. *BMC Med. Res. Methodol.* **2025**, *25*, 206. [CrossRef]
87. Wang, W.L.; Wang, R.; Lu, Y.Y.; Meng, Y.Y.; Zhou, Z. A novel spatial heterogeneity decay model for depicting scale-effects in urban land. *GISci. Remote Sens.* **2025**, *62*, 2547122. [CrossRef]
88. Zuo, S.Z.; Ghosh, D.; Ding, P.; Yang, F. Mediation Analysis with the Mediator and Outcome Missing Not at Random. *J. Am. Stat. Assoc.* **2024**, *120*, 794–804. [CrossRef]
89. Ikotun, A.M.; Ezugwu, A.E.; Abualigah, L.; Abuhaija, B.; Heming, J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Inf. Sci.* **2023**, *622*, 178–210. [CrossRef]
90. Antonson, H.; Buckland, P.; Nyqvist, R. A society ill-equipped to deal with the effects of climate change on cultural heritage and landscape: A qualitative assessment of planning practices in transport infrastructure. *Clim. Change* **2021**, *166*, 18. [CrossRef]
91. Abebaw, S.E. A Global Review of the Impacts of Climate Change and Variability on Agricultural Productivity and Farmers' Adaptation Strategies. *Food Sci. Nutr.* **2025**, *13*, e70260. [CrossRef]
92. Li, J.C.; Han, L.Y.; Zhang, G.M.; Su, Z.Z.; Zhao, Y.F. Temporal-spatial variations of human settlements in relation to environment change during the Longshan culture and Xia-Shang periods in Shanxi Province, China. *Quat. Int.* **2017**, *436*, 129–137. [CrossRef]
93. Rubin, S.E.; Mørch, I.; Olsen, N.; Nørtoft, K. Important intergenerational transmission of knowledge in promotion of well-being and cultural identity in Greenland. *Int. J. Circumpolar Health* **2024**, *83*, 2420479. [CrossRef] [PubMed]
94. Maldonado-Erazo, C.P.; Álvarez-García, J.; Río-Rama, M.d.l.C.d.; Durán-Sánchez, A. Scientific Mapping on the Impact of Climate Change on Cultural and Natural Heritage: A Systematic Scientometric Analysis. *Land* **2021**, *10*, 76. [CrossRef]
95. Cotella, V.A.; Aburamadan, R.; Di Perna, E. HBIM applications in the world heritage city of as-salt, Jordan: Architecture as a reflection of cultural diversity shaped by migratory flows. *Digit. Appl. Archaeol. Cult. Herit.* **2025**, *37*, e00418. [CrossRef]
96. Du, H.F.; Liu, J.G. Exploring Spatial–Temporal Coupling and Its Driving Factors of Cultural and Tourism Industry in the Beijing–Tianjin–Hebei Urban Agglomeration, China. *Sustainability* **2025**, *17*, 890. [CrossRef]
97. Fuller, M.R.; Doyle, M.W.; Strayer, D.L. Causes and consequences of habitat fragmentation in river networks. *Ann. N. Y. Acad. Sci.* **2015**, *1355*, 31–51. [CrossRef] [PubMed]
98. Yang, M.L.; Wu, C.; Gong, L.; Tan, G.W. Coupling Coordination Relationship between Cultural Landscape Conservation and Socio-Economic System in Ethnic Villages of Southeast Guizhou. *Land* **2024**, *13*, 1223. [CrossRef]
99. Wang, X.M.; Li, H.M.; Wang, Y.F.; Zhao, X.F. Assessing climate risk related to precipitation on cultural heritage at the provincial level in China. *Sci. Total Environ.* **2022**, *835*, 155489. [CrossRef]

100. Yang, H.; He, J.; Su, Y.F.; Xu, J.C. Adaptation to climate change: Ethnic groups in Southwest China. *Environ. Hazards* **2021**, *21*, 117–136. [CrossRef]
101. Chen, Z.H.; Gao, Q.; Wu, Y.J.; Li, J.X.; Li, X.W.; Li, X.; Wang, Z.B.; Cui, H.Y. World Cultural Heritage sites are under climate stress and no emissions mitigation pathways can uniformly protect them. *Commun. Earth Environ.* **2025**, *6*, 628. [CrossRef]
102. Li, X.B.; Zhu, R.; Shi, C.Y.; Yang, X.K.; Chen, J.Z.; Wei, K.X. Research on the construction of intangible cultural heritage corridors in the Yellow River Basin based on geographic information system (GIS) technology and the minimum cumulative resistance (MCR) model. *npj Herit. Sci.* **2024**, *12*, 271. [CrossRef]
103. Lv, R.B.; Liu, Y.P.; Zhang, L.; Kong, D.Z. Urban historic heritage buffer zone delineation: The case of Shedian. *npj Herit. Sci.* **2022**, *10*, 64. [CrossRef]
104. Min, W.Q. A scientometric review of cultural heritage management and sustainable development through evolutionary perspectives. *npj Herit. Sci.* **2025**, *13*, 215. [CrossRef]
105. Li, J.; Wu, X.; Du, Y. Reframing Cultural Heritage Policy Through Place-Based Perspectives: The Evolution of China's ICH Governance Amid Historical Continuity and Global Convergence. *Land* **2025**, *14*, 1425. [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.

land

MDPI

Article

Towards Trustworthy Urban Land Use Classification: A Synergistic Fusion of Deep Learning and Explainable Machine Learning with a Nanning Case Study

Yusheng Zheng [1,2], Xinying Huang [1,2] and Huanmei Yao [1,2,*]

[1] School of Resources, Environment and Materials, Guangxi University, Nanning 530004, China; 2315393091@st.gxu.edu.cn (Y.Z.); 2515390341@st.gxu.edu.cn (X.H.)
[2] Key Laboratory of Environmental Protection (Guangxi University), Education Department of Guangxi Zhuang Autonomous Region, Nanning 530004, China
* Correspondence: yaohuanmei@gxu.edu.cn

Abstract

While artificial intelligence (AI) has advanced urban land use classification, its application in high-stakes decision making, such as urban planning, demands not only high accuracy but also transparency and interpretability. This study evaluates the potential of Google Satellite Embeddings (GSE), a ready-to-use dataset of AI-generated numerical features that capture deep land cover characteristics, for land use classification in the central urban area of Nanning in 2022. A synergistic analytical framework was constructed by integrating the 64 high-dimensional features of GSE data with the feature attribution of Shapley Additive Explanations (SHAP), merging deep learning features with explainable machine learning. The results demonstrate that the XGBoost model (OA = 85.00% ± 2.24%) significantly outperformed the Random Forest (RF) model (OA = 81.87% ± 1.72%) overall. Key abstract features were successfully interpreted as comprehensible geographic semantics, with A51 and A36 corresponding to built-up intensity and vegetation cover, respectively. Moreover, XGBoost enabled more refined decisions than Random Forest (RF) due to its superior ability to distinguish between functionally distinct classes that have similar physical appearances. This framework provides a scalable and transferable analytical solution for the challenges of feature limitations and insufficient model transparency in urban land use classification.

Keywords: urban land use; machine learning; satellite embeddings; explainable AI; semantic analysis

1. Introduction

Driven by global urbanization and the rise in smart cities, the practice of governing cities, which encompasses urban planning and resource management, is transitioning from traditional, static approaches to data-driven, dynamic, and fine-grained paradigms [1–3]. In this context, accurate and continuously updated urban land use information has become the bedrock of scientific decision making and resource allocation [4]. This foundational data underpins critical applications ranging from environmental management, such as modeling urban heat islands, to strategic planning for public infrastructure and services. While artificial intelligence (AI) has significantly enhanced the accuracy of land use classification, the increasing complexity accompanying these performance gains has rendered model

decision making opaque, introducing trust deficits in practical applications [5,6]. For high-stakes domains such as urban planning, a model must not only deliver precise results but also articulate the logic and evidence behind its decisions [7,8]. Therefore, enhancing the interpretability of high-performance AI models, especially complex deep learning and ensemble algorithms, is a critical prerequisite for their trustworthy application in planning.

To accurately map complex urban landscapes, current research has increasingly integrated multi-source data and advanced algorithms [6]. On one hand, the fusion of traditional remote sensing features like spectral and texture information with social sensing data from sources such as OpenStreetMap (OSM) and Points of Interest (POI) has shifted the unit of analysis from pixels to urban parcels, significantly improving semantic representation [9–11]. This shift from pixels to parcels is a crucial methodological decision grounded in the distinction between land cover and land use. While pixel-based and object-based image analysis excel at classifying physical land cover (surfaces such as asphalt, grass, and concrete), they are often less suited for identifying the socio-economic land use, which involves distinguishing between functions like commercial and residential activities [10,12]. The urban parcel, defined by roads and property boundaries, represents the fundamental semantic unit at which urban functions are organized and managed [13]. Therefore, adopting a parcel-based approach aligns the unit of analysis with the actual phenomenon being studied. The consensus in the literature is that aggregating image features within these parcels creates a more holistic and stable representation of function, which mitigates the salt-and-pepper noise inherent in pixel-based methods and yields spatially coherent outputs directly usable for urban planners [9,11,14]. On the other hand, AI models have been widely adopted to boost the accuracy and automation of land use classification [15,16]. Traditional machine learning models like Random Forest (RF) offer a balance between computational cost and performance [17]. Deep learning, despite its computational demands, has proven effective and robust in identifying urban land use patterns by automatically learning deep semantic features from multi-source data [18,19]. However, despite the advances in fine-grained urban understanding driven by the synergy of multi-source data and AI, existing methods still rely heavily on manual feature engineering and expert knowledge, and struggle to comprehensively capture the fine-grained spatial structures of complex urban scenes [20,21]. Moreover, relying entirely on deep learning models trained from scratch often leads to the dual problems of high computational costs and poor generalizability [22,23].

While deep learning methods excel at feature extraction, their internal representations are highly abstract, creating an opacity that limits trustworthy application in high-stakes domains [24]. In response, Explainable Artificial Intelligence (XAI) has emerged to demystify these black boxes, with tools like SHAP enabling the quantification of each feature's contribution to a prediction [25–27]. However, a critical limitation persists in current XAI applications for urban science: their outputs often remain confined to the level of machine logic. They excel at pinpointing what data the model prioritizes but fall short of articulating why in a geographically meaningful way. Pioneering studies exemplify this challenge. Methods like Grad-CAM++ and explainable boosting machines have been successfully used to reveal which image regions and POI types were most influential in a land use model [28]. In a similar vein, feature attribution and counterfactual analysis have been leveraged on graph-based models to show how mobility data contributed to predictions [29]. Despite their value, these explanations stop short of providing a true geographic narrative. They tell a planner that certain pixels or POI counts were weighted heavily, but not that the decision was based on recognizing a pattern of high-density built form, complex vertical structure, or a specific texture of mixed land cover—which is the language of urban planning. Thus, a crucial gap exists in bridging this interpretive divide:

https://doi.org/10.3390/land15010158

moving from identifying important abstract features to decoding the tangible geographic concepts they represent.

The recent emergence of the GSE data offers a novel approach for addressing the shallow feature problem in remote sensing. Generated by Google's AlphaEarth Foundation (AEF), it transforms global, multi-source satellite imagery into a high-dimensional feature vector through deep learning [30]. These pre-encoded features automatically capture the deep spectral-spatial structures and semantic patterns of land cover, holding the potential to fundamentally bypass the reliance on manually engineered features and provide an unprecedented feature foundation for fine-grained, parcel-level classification [31].

This presents an opportunity to forge a third way that navigates the trade-off between traditional methods and end-to-end deep learning. While traditional approaches using manually engineered features are interpretable but often limited by a shallow feature bottleneck, end-to-end deep learning models offer high accuracy but are computationally expensive black boxes. The proposed framework seeks to synergize these approaches by leveraging the pre-encoded deep features from GSE to achieve high performance, while utilizing explainable machine learning to maintain efficiency and ensure the transparency required for trustworthy application in urban planning.

Therefore, this study applies this synergistic framework to the land use classification of the central urban area of Nanning. It aims to achieve two core objectives.

- To evaluate the performance of machine learning models using GSE data for urban land use classification.
- To enhance model interpretability by decoding the geographic semantics of the features and dissecting the model's decision-making mechanisms.

This study provides a synergistic analytical framework that integrates deep learning features with explainable machine learning, thereby paving the way for more trustworthy and interpretable approaches in smart city planning.

2. Materials and Methods

2.1. Study Area

This study focuses on the central urban area of Nanning, the capital city of the Guangxi Zhuang Autonomous Region in Southern China. It consists of the contiguous urban built-up region inside the Nanning Ring Expressway and the Nanning International Railway Port area, with a total area of approximately 665 km^2 (Figure 1). This area spans several major administrative districts, including Xixiangtang, Qingxiu, Xingning, Jiangnan, Yongning, and Liangqing, which represent diverse stages of urban development and functional concentrations.

As a frontier city for China's open cooperation with the Association of Southeast Asian Nations (ASEAN), Nanning's development serves as a representative microcosm of the rapid urbanization processes occurring throughout southern China [32]. The central urban district, which hosts the majority of the city's population and economic activities, has evolved into an urban environment characterized by high functional complexity and significant landscape heterogeneity. A significant challenge arises from this complexity: land parcels with distinct functions often exhibit similar physical characteristics in remote sensing imagery [33]. This phenomenon poses a significant challenge to both the accuracy of land use classification and the transparency of model decision-making [34,35].

2.2. Data Source and Preprocessing

2.2.1. World Imagery

Esri World Imagery, a global service provided through the ArcGIS 10.8 platform, integrates multi-source satellite and aerial imagery with resolutions up to 0.3 m and offers

https://doi.org/10.3390/land15010158

multiple historical versions. This imagery service can be directly loaded and utilized within the ArcGIS environment. For this study, the 2022 Esri World Imagery was employed to support the visual interpretation of land parcel functions.

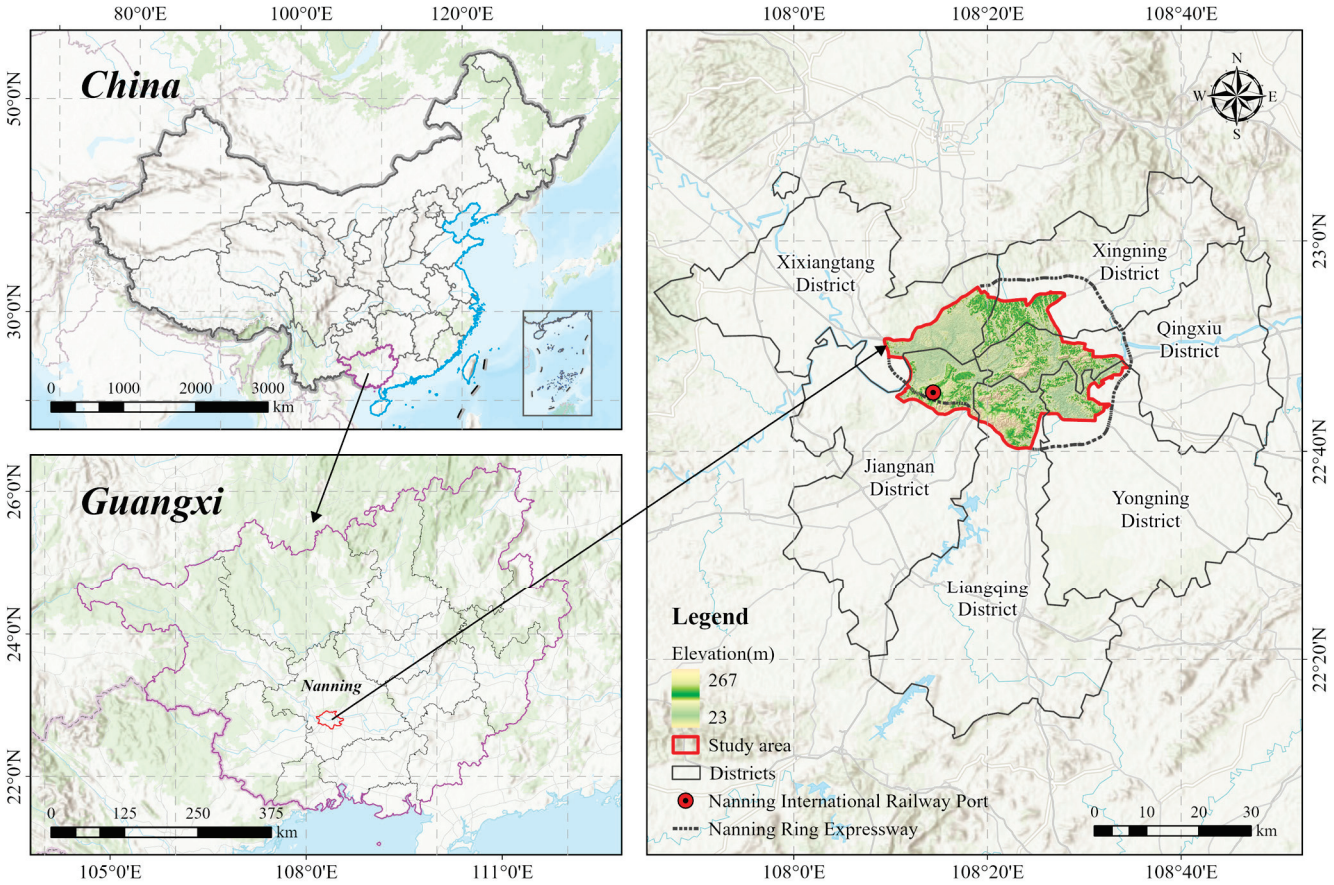


Figure 1. Overview of the study area.

2.2.2. OpenStreetMap (OSM) Network Data

As a collaborative project providing free, editable global maps, OSM is a well-established and effective data source for capturing urban functional spatial patterns [36,37]. The 2022 dataset for the central urban area of Nanning was utilized in this study. First, the data was manually verified to eliminate redundant information. Subsequently, the cleaned roads were classified into a hierarchical system (Table 1), which was developed by integrating categories from the Ministry of Housing and Urban-Rural Development (MoHURD) and methodologies from previous research [38,39], and adapting them to reflect the specific local context of the study area. This classified hierarchical network is crucial as it serves as the primary structural framework for delineating urban parcels, which are the fundamental units of our analysis. The detailed methodology for how these road levels are utilized in the parcel generation process is described in Section 2.3.1.

Table 1. Road Classification.

Classes	Road Descriptions	Road Widths (m)
Level 1	Motorway, motorway_link, primary, primary_link, trunk, trunk_link, rail	40
Level 2	Secondary, secondary_link	20
Level 3	Tertiary, tertiary_link, residential, service, unclassified, unknown, other	10

To obtain an accurate river network, which serves as a crucial natural barrier for parcel delineation, a hybrid approach was adopted. The river network data from OSM was used as a baseline. Then, to ensure the highest level of positional accuracy and alignment with the current landscape, a manual visual correction of the OSM data was performed against the 2022 Esri World Imagery.

2.2.3. Points of Interest (POI) Data

POI effectively captures the spatial distribution of urban functions, offering crucial semantic information for land use classification [40,41]. For this study, 2022 POI data for the central urban area of Nanning were acquired from the Amap Web Service Application Programming Interface (API) using Python scripts.

Drawing upon the urban land use classification standards from the Ministry of Housing and Urban-Rural Development (MoHURD) and methodologies from previous research [9,42,43], a simplified, four-category functional area system was developed. Within this framework, commercial and industrial lands, which share similar economic functions, were consolidated into a single class.

This consolidation is methodologically justified by two key factors specific to this study. First, our focus on the central urban area means that the observed industrial parcels are predominantly light industry, logistics centers, and wholesale markets, which are functionally and physically distinct from the heavy industry typically located on the urban periphery. From a remote sensing perspective, these central urban industrial areas share similar physical signatures with old city commercial areas, such as low, metal-roofed buildings. Second, this simplification allows the model to learn a more robust, unified signature for areas of intense economic activity based on their shared physical attributes captured by GSE, which is crucial for the primary goal of this study: evaluating the interpretability of this new data source.

The final system comprises four categories: institution, open space, business, and residence (Table 2). The detailed classification of POI types into these categories is provided in Supplementary Table S1.

Table 2. Urban Land Use Classification.

Classification	Main Subcategories	Description
Institution	Governmental organization; Medical service; Culture and Education.	This class primarily includes land for public administration, public services, and education.
Open space	Tourist attraction	This category encompasses public green spaces and other open areas intended for recreation, ecological functions, or public access.
Business	Enterprises; Shopping; Accommodation service	This class represents areas dedicated to commercial activities, financial services, corporate functions, and some light industry.
Residence	Commercial house; Daily life service	This category includes land primarily used for housing.

2.2.4. Google Satellite Embedding (GSE) Data

GSE data served as the primary feature source for our machine learning models. It is essential to clarify that GSE is not a classification product but a high-dimensional geospatial embedding vector instead of traditional spectral bands. It is generated by Google's

https://doi.org/10.3390/land15010158

AlphaEarth Foundations (AEF) [30], an embedding field model based on a Transformer architecture developed by Google DeepMind.

The AEF model is trained using self-supervised learning, specifically employing contrastive learning techniques. This allows it to learn universal geospatial representations from massive, unlabeled Earth observation data. The process learns and encodes a year's worth of multi-modal, time-series satellite imagery into a highly condensed semantic space. The result is a 64-dimensional (A00–A63) vector for each 10 m pixel that represents deep semantic features rather than direct physical signals. This encoding process aims to maximize discriminative information useful for downstream tasks while filtering out noise and redundancy [31]. This vector comprehensively reflects the spectral, structural, and contextual properties of land cover, capturing deep patterns that transcend singular physical responses and thereby overcoming the shallow feature bottleneck of traditional remote sensing.

A key strength of the AEF model is its ability to process heterogeneous data. Its architecture includes source-specific encoders for different data types (e.g., optical, radar). Through joint training on a global scale, the model learns to intelligently fuse information, meaning each dimension of a GSE vector can encapsulate complex patterns from multiple sensors. While trained globally, these embeddings provide fine-grained features applicable at local scales, making them suitable for our Nanning study area.

The 2022 annual, 10 m resolution product was acquired from the GOOGLE/SATELLITE _EMBEDDING/V1/ANNUAL dataset on the Google Earth Engine (GEE) platform. These 64 dimensions constitute an inseparable feature vector and were therefore all utilized for model training. As a ready-to-use dataset, it has been internally processed to handle cloud cover and data gaps. Furthermore, its vectors are of unit length, obviating the need for normalization. This significantly simplifies the data preprocessing workflow and makes the dataset inherently compatible with tree-based classifiers such as XGBoost and RF.

2.2.5. Sentinel-2 Multispectral Imagery

Sentinel-2 multispectral imagery was selected for its high spatial resolution (10 m) and rich spectral information. Its bands include the essential red, near-infrared (NIR), and short-wave infrared (SWIR) wavelengths required for calculating classic remote sensing indices such as the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Built-up Index (NDBI) [44,45].

Within the GEE platform, the COPERNICUS/S2_SR_HARMONIZED dataset was accessed. An annual mean composite image for the central urban area of Nanning was synthesized to create a data source for subsequent spectral and texture feature extraction. This composite was generated using all available 2022 images with a cloud cover of less than 40%.

2.3. Methods

This study proposes a synergistic analytical framework that integrates deep learning features with explainable machine learning (Figure 2). This framework consists of three core steps.

1. First, a high-precision sample library was constructed by fusing multi-source geospatial data with expert visual interpretation within ArcMap 10.8.
2. Second, two parallel feature sets were formulated on the GEE platform: the high-dimensional deep embedding features for model training, and a set of physical reference features for attribution.
3. Finally, the modeling and analysis pipeline was implemented in Python 3.9. We trained and evaluated machine learning models using the scikit-learn (for RF) and xg-

https://doi.org/10.3390/land15010158

boost libraries, which were then deconstructed using the shap library to analyze their internal decision-making mechanisms. This involved quantifying the contributions of key abstract features and systematically correlating them with the physical reference features using the pandas and scipy libraries.

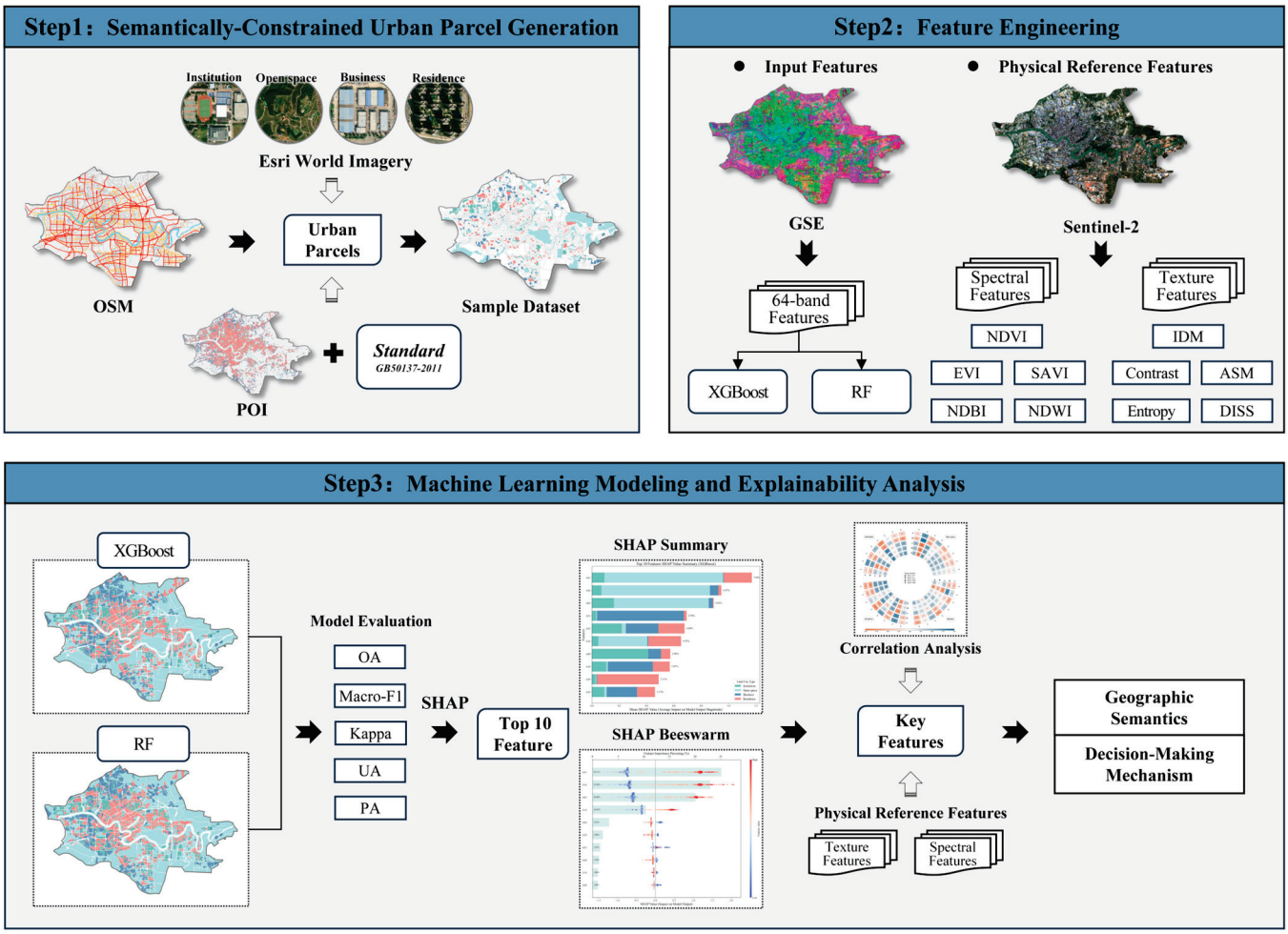


Figure 2. Methodology workflow. (1) Step1: The sample dataset is established through expert annotation based on Esri World imagery, POI, and Standard, after delineating land parcels using OSM. (2) Step2: Input features are extracted from GSE data, and a set of physical reference features is constructed using Sentinel-2. (3) Step3: XGBoost and RF models are evaluated, followed by a SHAP analysis. SHAP summary and beeswarm plots are generated for the top 10 features to identify key features. Finally, a correlation analysis is conducted between key features and physical reference features to interpret their geographical semantics and elucidate the model's decision-making mechanism.

2.3.1. Semantically Constrained Urban Parcel Generation

OSM-guided physical parcel delineation. The central urban area of Nanning was first segmented into morphologically intact and non-overlapping initial parcels adopting an established method for automated parcel characterization [37]. This topological segmentation approach aligns with the framework of Essential Urban Land Use Categories (EULUC), where road networks serve as the fundamental skeleton of urban functional organization [9,11]. The process was executed within the ArcGIS platform by applying multi-level buffer analysis to the pre-processed 2022 OSM network and then overlaying natural barriers such as rivers.

POI-based semantic labeling and creation of a candidate sample pool. Next, the initial physical parcels were spatially associated with four major categories of POI data. The functional composition of each parcel was quantified by calculating the density of each of the four major POI categories. A preliminary label was assigned only to parcels where a single function was clearly dominant, defined as constituting over 70% of the total POI count within that parcel. This quantitative filtering strategy is consistent with recent frequency-density methodologies, which advocate for high thresholds to mitigate the inherent noise in social sensing data and ensure the functional purity of training samples [39,46]. This process was designed to focus the subsequent manual verification effort precisely on those parcels where POI signals were strongest, enabling efficient and targeted expert correction.

Visual Interpretation with Multi-Source Information and Final Sample Verification. To ensure definitive label accuracy and mitigate any remaining POI-induced semantic bias, a rigorous, multi-stage verification process was implemented on the candidate sample pool. First, a multi-source interpretation protocol was established. For each candidate parcel, interpreters did not rely solely on the high-resolution imagery but synergistically used it with the specific POI name labels within that parcel. This synthesis directly addresses the challenge of spectral confusion in complex urban environments, where functionally distinct zones often exhibit nearly identical physical morphologies [47]. For example, while both business and institution parcels might appear as high-rise buildings on imagery, reviewing the POI names (e.g., "Nanning No. 2 High School," "First Affiliated Hospital of Guangxi Medical University") allowed for unambiguous classification as institution, in strict accordance with the Standard for Urban Land Use Classification and Planning of Construction Land [42]. This approach effectively leveraged POI data as a semantic clue rather than a mere counter, directly addressing the risk of misclassification based on visual morphology alone. By integrating semantic labels with physical textures, this protocol adheres to the multi-view fusion paradigm recommended for fine-grained urban classification [33]. Second, to guarantee label consistency and quality in the absence of multiple annotators, a systematic iterative review was employed. The expert first conducted an initial round of labeling for the entire candidate set. After a deliberate time interval to minimize cognitive bias, a second, independent review round was performed. During this second pass, the expert re-evaluated each label against the established protocol, paying special attention to potentially ambiguous cases. Any parcels where the initial label seemed questionable upon re-inspection were subjected to a final, more intensive investigation before a definitive label was assigned. Parcels that remained ambiguous even after this multi-step scrutiny were discarded from the final sample set.

Ultimately, a final dataset of 800 high-precision samples—200 for each of the four functional categories—was curated that were unambiguously labeled (Figure 3). The aim was for spatial representativeness rather than absolute spatial uniformity in designing the sampling strategy. Urban functional areas are inherently clustered, and a strictly uniform sampling might fail to capture the diversity of these real-world patterns. Therefore, this approach focused on selecting samples for each class from multiple, geographically distinct locations across the study area. For example, samples for the business class were drawn from the old city center, the new CBD, and several secondary commercial hubs. This strategy ensures the model captures the high intra-class variability typical of heterogeneous urban landscapes [39]. This carefully constructed dataset serves as the foundational ground truth for all subsequent analyses.

https://doi.org/10.3390/land15010158

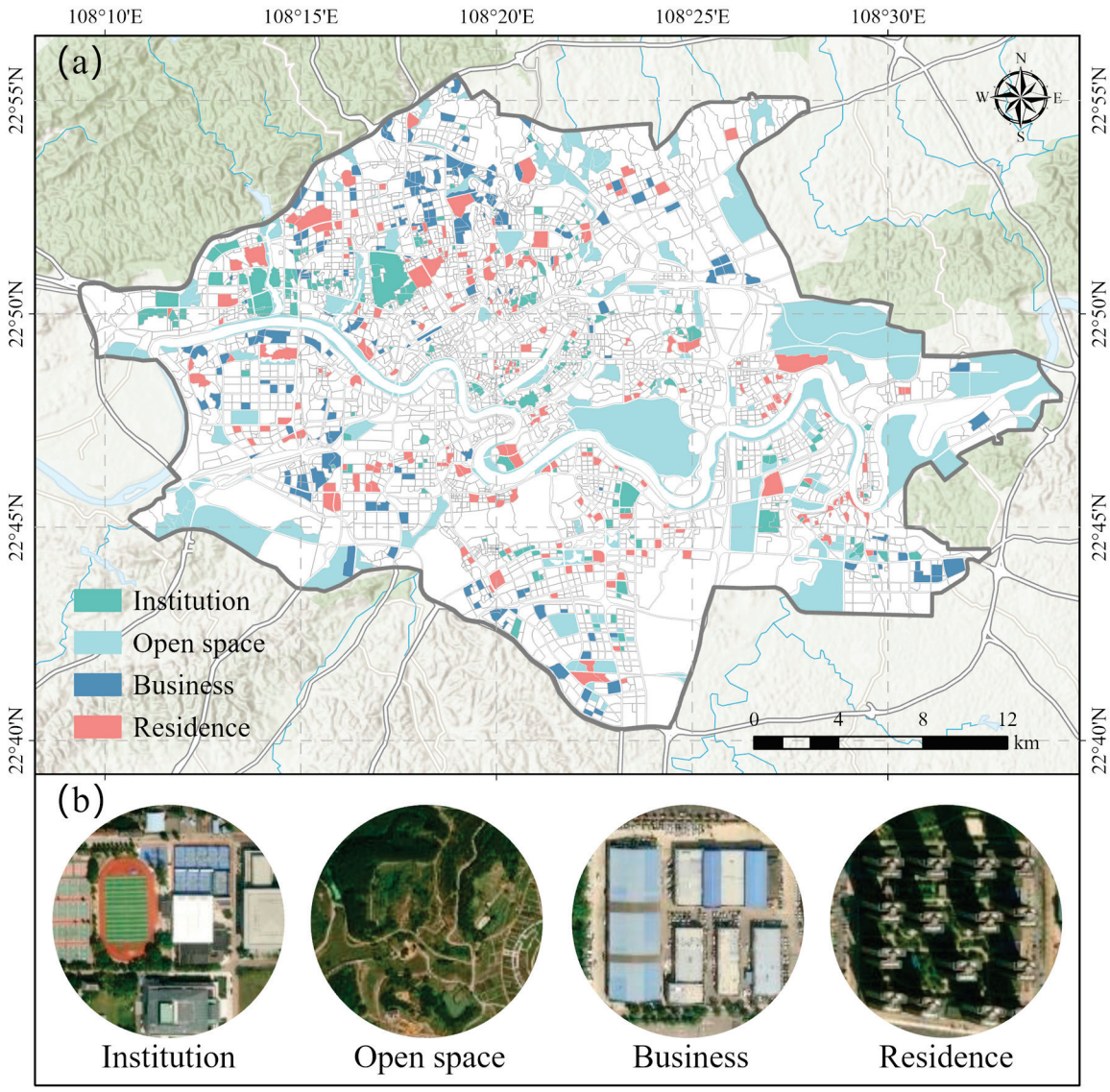


Figure 3. (**a**) Spatial distribution of the samples; (**b**) The representative samples.

2.3.2. Feature Engineering

Generation of high-dimensional input features. The core input features for the models were generated by aggregating the pixel-level GSE data. Specifically, the mean value of each of the 64 GSE feature bands was calculated on the GEE platform for each parcel. Aggregating these high-resolution raster features at the parcel level is not a loss of detail; rather, it creates a robust, composite feature signature that holistically represents the typical mixture of surfaces (buildings, vegetation, pavement) associated with a specific function. This process resulted in a single 64-dimensional feature vector per parcel, serving as a comprehensive, high-level representation of its unique deep spectral-spatial patterns.

Extraction of physical reference features. To facilitate model interpretation, a white-box feature set composed of traditional remote sensing characteristics was extracted from the annual Sentinel-2 composite imagery (Table 3). This set is designed to serve as a bridge, connecting the abstract deep learning features to established geographical knowledge. Specifically, it comprises two components:

- First, a series of spectral features were calculated to quantitatively assess the vegetation, built-up, and water status of the land surface [48,49].

https://doi.org/10.3390/land15010158

- Second, texture features, derived from the Gray-Level Co-occurrence Matrix (GLCM), were computed to describe the spatial structure and heterogeneity of the land surface, complementing the spectral signatures which are often insufficient in heterogeneous urban environments [50]. However, utilizing the full set of GLCM features often leads to high data redundancy and computational inefficiency due to strong inter-correlations [51]. Therefore, drawing upon established feature selection strategies for remote sensing classification, a specific subset of metrics was chosen to form a representative and minimally redundant set [39,52]. These metrics targeted three complementary textural dimensions critical for distinguishing urban functions: (1) Spatial Order/Disorder (ASM, Entropy) to characterize the complexity of building arrangements; (2) Local Contrast (Contrast, Dissimilarity) to capture the edge intensity typical of built-up areas; and (3) Homogeneity (IDM) to identify uniform surfaces such as open spaces and water bodies.

Table 3. Physical Reference Features.

Feature Types	Indices	Description
Spectral features	Normalized Difference Vegetation Index (NDVI)	Indicates the presence and density of green vegetation
	Enhanced Vegetation Index (EVI)	Quantifies vegetation with reduced sensitivity to soil and atmospheric noise
	Soil-Adjusted Vegetation Index (SAVI)	Minimizes soil brightness effects to accurately map sparse vegetation
	Normalized Difference Built-up Index (NDBI)	Highlights impervious surfaces and built-up areas
	Normalized Difference Water Index (NBVI)	Delineates open water bodies and aids in identifying building shadows
Textural features	Entropy	Measures the disorder or complexity of the spatial texture
	Contrast	Captures the intensity of local variations and edges
	Angular Second Moment (ASM)	Measures textural uniformity and order
	Inverse Difference Moment (IDM)	Quantifies the local homogeneity of the image texture
	Dissimilarity (DISS)	Describes the distinctiveness of spatial patterns and local contrast

2.3.3. Machine Learning Modeling and Explainability Analysis

Model training and performance evaluation. Two machine learning algorithms, RF and XGBoost, were selected for comparison. These algorithms represent the two dominant paradigms in ensemble learning—Bagging and Boosting, respectively—and are widely utilized for their robustness in modeling complex, non-linear relationships [53,54].

To mitigate the potential risk of overfitting given the high-dimensional feature space (64 features) relative to the sample size (800 parcels), several strategies were implemented to ensure model robustness. First, a 5-fold cross-validation strategy was used for both model training and evaluation, providing a more reliable estimate of generalization performance [55]. Second, hyperparameter tuning was performed using a grid search within the cross-validation loop to identify the optimal parameters that control model complexity (e.g., tree depth, learning rate). Third, an early stopping mechanism was employed for the XGBoost model specifically. This technique halts the training process if performance on a validation set does not improve for a specified number of iterations, effectively preventing the model from fitting to noise in the training data. Model performance was assessed using

https://doi.org/10.3390/land15010158

Overall Accuracy (OA), the Kappa Coefficient, Producer's Accuracy (PA), User's Accuracy (UA), and the Macro-F1.

Quantifying feature contributions to model decisions. The TreeSHAP algorithm, an efficient variant optimized for tree-based models, was utilized to quantify feature contributions at both global and local levels.

- At the global level, the most influential features were identified by calculating the mean absolute SHAP value for each feature across all samples. To ensure stable results, these importance values were averaged across the five folds of the cross-validation, and the final ranking is presented as a percentage contribution.
- At the local level, stratified SHAP beeswarm plots were employed for visual analysis. These plots intuitively display how a single feature contributes to each individual sample's prediction, while simultaneously using color to reveal the patterns between the feature's original value (high or low) and the direction and magnitude of its contribution [56,57].

Attributing Physical Meaning to Deep Features. A stratified Spearman correlation analysis was performed to investigate the intrinsic meaning of the abstract deep learning features. This analysis was conducted between the key features and our white-box set of physical reference features. The stratified approach—calculating correlations separately within each land use category—was designed to uncover how the association patterns between a single deep feature and the physical metrics vary across different geographical contexts.

3. Results

3.1. Performance Evaluation of Embedding-Based Models

3.1.1. Quantitative Evaluation

The performance of the XGBoost and RF models, trained on the high-dimensional GSE features, was evaluated using a 5-fold cross-validation strategy. The evaluation revealed that the XGBoost model demonstrated superior overall performance compared to the RF model (Table 4). The OA of XGBoost reached 85.00% ± 2.24%, marking a 3.13 percentage point improvement over RF's accuracy of 81.87% ± 1.72%. Furthermore, its superiority in maintaining a balance between precision and recall was evidenced by a higher Macro-F1 of 84.81% ± 2.32%, compared to RF's 81.72% ± 1.73%. Similarly, XGBoost achieved a higher Kappa Coefficient of 0.8000 ± 0.0298, compared to RF's 0.7583 ± 0.0230, indicating a stronger agreement between its classification results and the ground truth, as well as superior model stability.

Table 4. Comparison of Model Performance.

Evaluation Metrics		RF	XGBoost
OA		81.87% ± 1.72%	85.00% ± 2.24%
Macro F1		81.72% ± 1.73%	84.81% ± 2.32%
Kappa		0.7583 ± 0.0230	0.8000 ± 0.0298
UA	Institution	69.00% ± 4.36%	74.50% ± 5.10%
	Open space	98.00% ± 1.87%	99.00% ± 1.22%
	Business	84.50% ± 6.60%	87.00% ± 7.81%
	Residence	76.00% ± 7.84%	79.50% ± 8.72%
PA	Institution	72.95% ± 6.08%	80.02% ± 4.21%
	Open space	96.15% ± 2.84%	96.18% ± 2.42%
	Business	81.97% ± 5.68%	84.11% ± 4.50%
	Residence	76.89% ± 1.20%	80.32% ± 5.65%

https://doi.org/10.3390/land15010158

Performance disparities were particularly pronounced for specific land use categories. For institutions, a category characterized by complex physical morphology, XGBoost achieved a PA of 80.02% and a UA of 74.50%. These figures were substantially higher than those of RF, which were 72.95% and 69.00%, respectively. For the often-confused business and residence categories, XGBoost also demonstrated an accuracy improvement of 2 to 4 percentage points. Both models performed exceptionally well in classifying open space, a category with distinct features, consistently achieving accuracies above 96%.

3.1.2. Qualitative Evaluation

The qualitative assessment visually corroborated and provided context for the quantitative findings by visualizing the classification results (Figure 4).

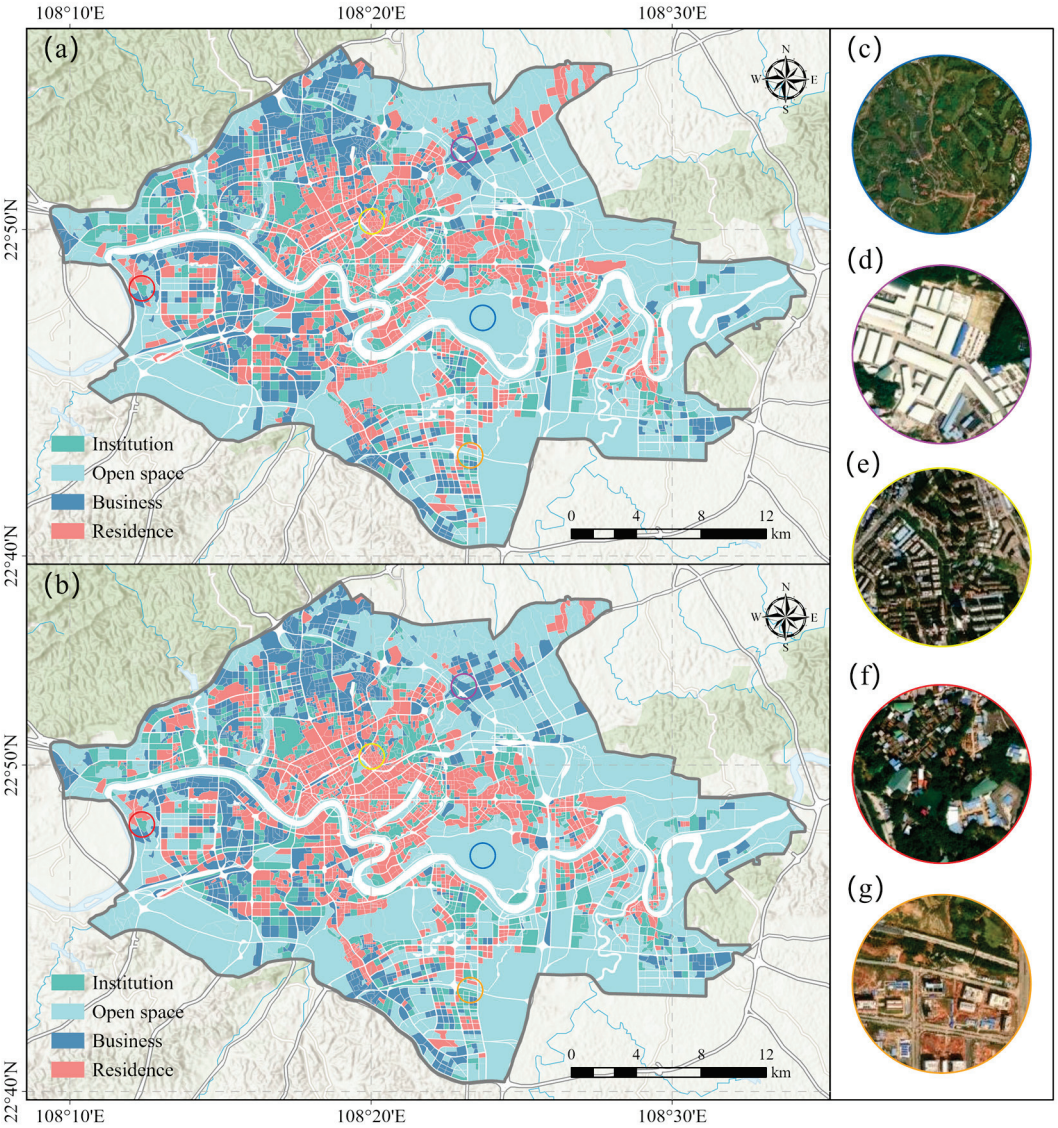


Figure 4. Visualization of Classification Results. (**a**) Classification results of XGBoost; (**b**) Classification results of RF; (**c**) A green park, identified by its large, contiguous green texture, was correctly classified as open space; (**d**) The area of low-rise, metal-roofed buildings was correctly classified as business; (**e**) The mixture of high-rise buildings and green space was correctly classified as residence; (**f**) A village, characterized by a dense cluster of small buildings and significant green space, was misclassified as institution; (**g**) Land under development, characterized by bare earth and uncompleted buildings, was misclassified as an institution.

https://doi.org/10.3390/land15010158

First, the areas of high model consensus aligned perfectly with the high-accuracy land use categories. Both models demonstrated precise identification of categories with relatively distinct physical characteristics, such as open space, business and residence (Figure 4c–e). This observation is consistent with the high OA and PA scores these categories achieved in the quantitative evaluation.

Furthermore, the visualizations offered a clear explanation for the performance disparity observed for the institution category, where the quantitative difference was most significant. This category is challenging because its physical features—such as building densities and vegetation cover that often fall somewhere between those of business and residence—are inherently ambiguous. This ambiguity led to varying degrees of confusion and misclassification by both models (Figure 4f,g).

3.2. Identification and Physical Interpretation of Key Features

An analysis of feature importance for both models, conducted using the SHAP, revealed a strong consensus in identifying key discriminative features (Figure 5). A significant overlap was observed, with 8 of the top 10 most important features being common to both models: A36, A41, A51, A05, A52, A14, A18, and A40. This high degree of agreement indicated that these features are the core drivers for distinguishing land parcel types in the study area. More importantly, it suggested they possess algorithm-agnostic discriminative value, making them robust indicators regardless of the specific model used.

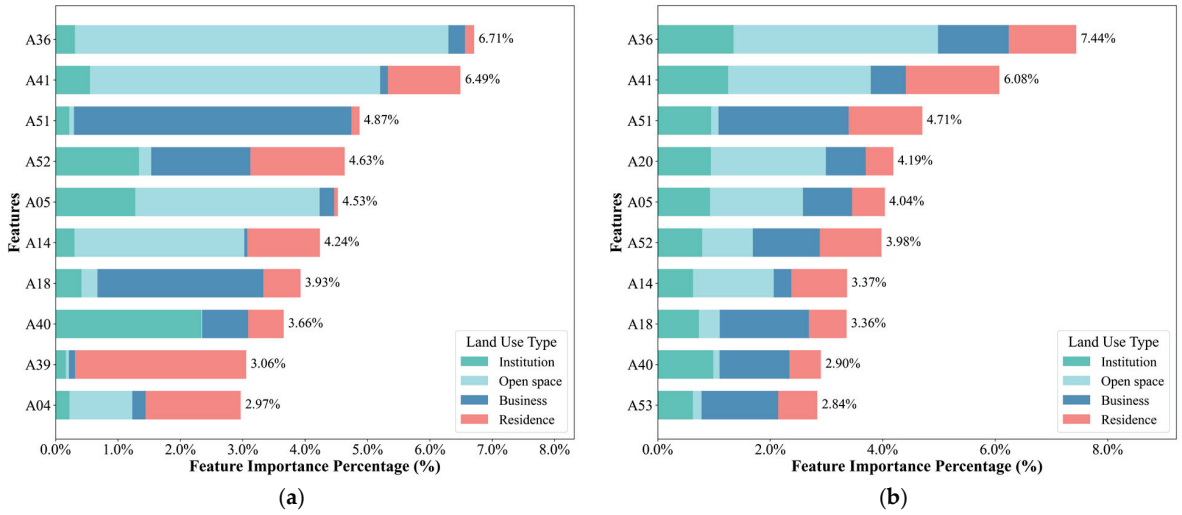


Figure 5. SHAP Summary Plot for the Top 10 Features. (**a**) XGBoost; (**b**) RF.

This consensus was further corroborated when examining the contributions of individual features to specific classes. For instance, in both models, A36 and A41 are primary contributors to the identification of open space. This suggested a strong correlation with surface cover information, such as vegetation or water bodies. Likewise, A51 consistently emerged as a key contributor for identifying the business across both models, implying that it likely captures spectral-spatial information associated with high-density built-up environments.

A stratified Spearman correlation analysis between the key features and a set of traditional spectral features and GLCM texture features revealed complex patterns of association (Figure 6).

- Features associated with the built-up environment.

Across the institution, residence, and open space classes, A51 functions as a universal indicator of built-up intensity. It exhibits significant positive correlations with both NDBI

and NDWI (r > 0.5) and strong negative correlations with vegetation indices like NDVI (r < −0.5). However, a key anomaly revealed its deeper properties. Within the Business class, the correlation between A51 and NDBI weakened dramatically (r = 0.18). In contrast, its correlation with NDWI became highly positive (r = 0.65), suggesting a functional shift from a simple spectral response to capturing the more complex structure of the urban canopy.

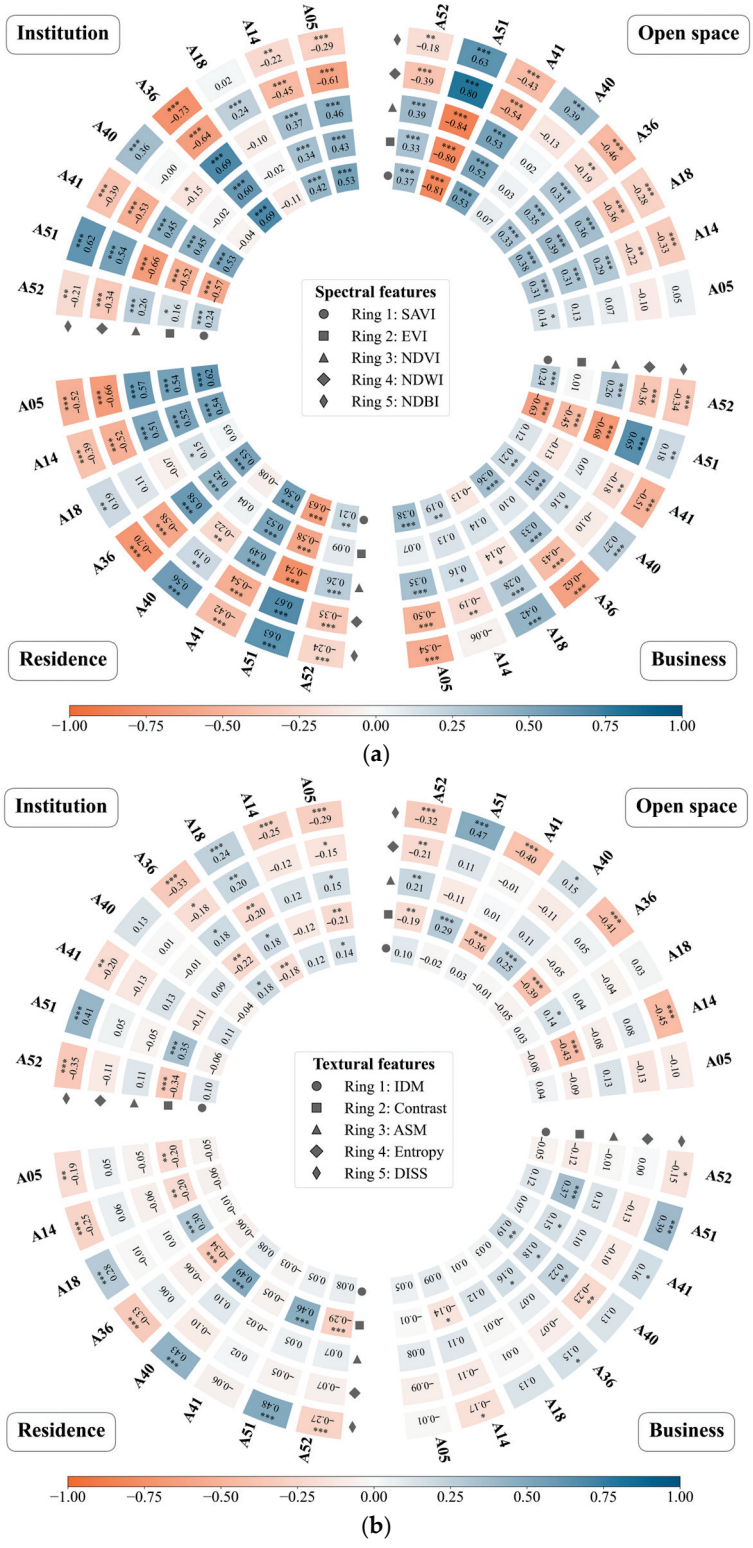


Figure 6. Correlation analysis of key features with spectral and texture features. (**a**) Spectral features; (**b**) Texture features. Significance levels are denoted as follows: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

https://doi.org/10.3390/land15010158

A40 behaved differently. Its positive correlations with NDBI (r = 0.56) and the DISS texture metric (r = 0.43) were most pronounced within the residence class. Feature A18, meanwhile, demonstrated a more multifaceted behavior, correlating positively with NDBI (r = 0.42) and NDWI (r = 0.28) in business, but with the DISS (r = 0.28) and Contrast (r = 0.30) texture metrics in Residence.

- Features associated with vegetation.

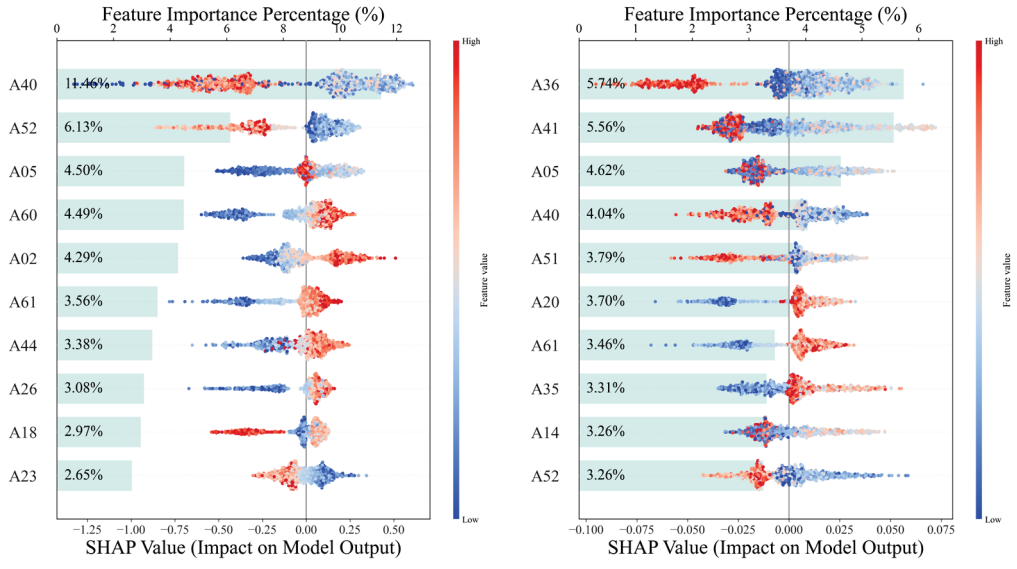
The correlation analysis revealed that A36 exhibited a uniform yet critical contrast pattern. On one hand, it showed strong positive correlations with various vegetation indices in built-up areas (r = 0.42 to 0.69). On the other hand, it demonstrated strong negative correlations with NDBI and NDWI, which represent built-up and shadow properties, respectively (r = −0.43 to −0.73). Critically, both the positive and negative correlations consistently and substantially weakened specifically within the open space class.

A41 and A14 not only correlated positively with vegetation indices but also negatively with texture metrics such as DISS and Contrast in open space. This suggested they are sensitive to the spatial configuration of vegetation rather than just its presence. A05 displayed a highly context-dependent association pattern. Its correlation with vegetation indices peaked within the residence class, while it also showed a strong negative correlation with NDWI in both institution and residence.

3.3. SHAP-Based Insights into Classification Decision

The SHAP beeswarm plots provide an intuitive visualization of the model's decision-making logic (Figure 7). They reveal how the model leverages the identified key features, with strategies that display both consensus and divergence across different classes.

A strong consensus in logic emerged between the two models for spectrally distinct classes. Specifically, both models identified business by leveraging high values of feature A51, a strong built-up signal, as the core positive driver. This single feature constituted 16.57% of XGBoost's and 8.41% of RF's decision weight. Conversely, high values in vegetation features like A52 and A36 acted as powerful negative predictors. The reverse held true for open space, where high values in vegetation features (e.g., A36, A41, and A05) provided the decisive positive push for classification.



(**a**) (**b**)

Figure 7. *Cont.*

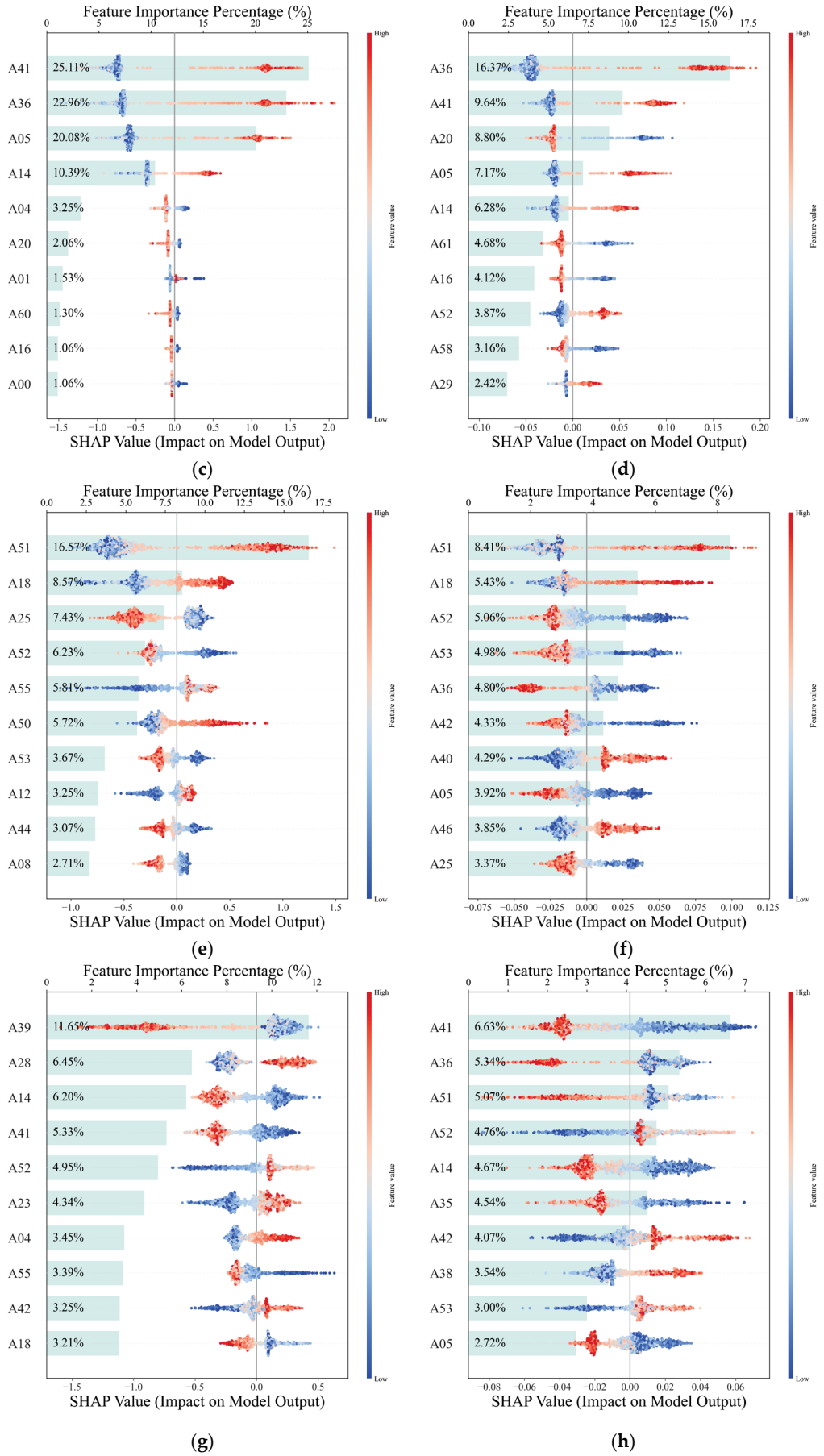


Figure 7. SHAP Beeswarm Plot. Each pair of plots compares the XGBoost (**left**) and the RF (**right**) for a specific category. (**a**,**b**) Institution; (**c**,**d**) Open space; (**e**,**f**) Business; (**g**,**h**) Residence.

https://doi.org/10.3390/land15010158

For spectrally ambiguous classes, a significant divergence in strategy emerged between the two models. In identifying institutions, the RF model exhibits a diffuse decision pattern with no single dominant feature, relying primarily on negative contributions from high values in various vegetation and built-up features. Its logic for residence was similarly indistinct, mainly driven by negative contributions from multiple core features. In stark contrast, XGBoost developed a more sophisticated strategy. It pinpointed A39 as the most crucial discriminator for residential areas; high values of this feature contribute strong negative SHAP values, effectively forming a powerful exclusion rule. Simultaneously, the model leveraged high values of features like A52 and A28 as positive contributors to construct a more nuanced profile of the residence.

4. Discussion

A core finding of this study, as presented in the results, is the remarkable consensus between the Random Forest and XGBoost models in identifying the most important features. This convergence, where two algorithmically distinct paradigms—Bagging and Boosting—independently prioritize the same set of GSEs (e.g., A36, A51, A41), is not a trivial observation. According to the principles of explainable machine learning, such cross-model agreement serves as a powerful indicator of feature robustness, suggesting that the identified predictors capture genuine underlying phenomena rather than model-specific artifacts or noise [58]. It serves as a powerful, algorithm-agnostic validation of the Google Satellite Embeddings themselves. This high degree of agreement strongly suggests that the discriminative power resides within the intrinsic structure of the GSE data, rather than being an artifact of a specific modeling choice. It demonstrates that GSE features successfully encode stable and meaningful geospatial patterns that are universally recognizable across different learning architectures, a characteristic often cited as a hallmark of high-quality representation learning [59]. Consequently, this algorithm-agnostic validation provides the impetus for a deeper inquiry into the nature of these embeddings, moving from simply identifying which features are important to understanding what they represent geographically and how they inform the models' decision-making logic.

4.1. Decoding the Geospatial Semantics of Key Features

4.1.1. Features for Built-Up Area Characterization

As direct literature on the geographic meaning of these specific features is not yet available, the interpretations were grounded in established remote sensing principles. The analysis revealed a clear functional differentiation among the most important features, which were then validated against classic physical indices. A51, A40, and A18 were identified as crucial features for characterizing built-up areas. Among these, A51 functions as a comprehensive indicator that comprehensively quantifies the intensity of the built environment. In contrast, A40 and A18 are more specialized, exhibiting strong class dependency in identifying distinct urban patterns.

It is established that NDBI effectively indicates impervious surfaces, NDWI captures structural information via building shadows [49], and texture metrics quantify the spatial heterogeneity of built-up areas [60–62]. The correlation analysis revealed that A51 is strongly correlated with indices from all three categories. It confirms that A51 is not a single-dimensional metric but a composite feature simultaneously encoding spectral, structural, and textural information, thereby providing a holistic quantification of built-up intensity. A critical insight emerges from the behavior of A51 in complex commercial districts: while its correlation with NDBI weakened, its strong association with NDWI persisted. This indicates that A51 is not merely a proxy for NDBI; rather, it is a versatile feature whose representational focus can shift from surface spectral properties to urban

https://doi.org/10.3390/land15010158

canopy structure. Consequently, A51 can be conceptualized as an adaptive composite indicator capable of adaptively adjusting its focus. The inherent flexibility enables a comprehensive and dynamic quantification of built-up intensity, explaining why XGBoost can extract rich contextual information from this single feature to achieve high accuracy. This interpretation is corroborated by recent research applying GSE to urban air quality; A51 was identified as a dominant predictor for SO2 concentrations, explicitly linking this feature to the capture of industrial land-use characteristics [63]. The view that A51 effectively encodes high-intensity urban morphologies is supported by this external validation.

In contrast to the broad applicability of A51, A40 and A18 exhibit pronounced class dependency. A40 precisely characterizes the spatial patterns unique to residential areas through its positive correlations with both NDBI and texture metrics. Meanwhile, A18 plays a more multifaceted role; it is sensitive to both the building-shadow complexes in business districts and the spatial textures of residential areas. This suggests that A18 is instrumental in identifying transitional built-up patterns between these two distinct types.

4.1.2. Features for Vegetation Characterization

A36, A41, A14, A05, and A52 were identified as key indicators for vegetation characterization. Clear functional differentiation exists within this group: A36 serves as a fundamental metric for overall surface greenness, while the remaining features specialize in identifying specific spatial patterns of green spaces.

Vegetation indices such as NDVI, EVI, and SAVI are the established standards for quantifying surface vegetation cover [48,64], whereas NDBI and NDWI characterize impervious surfaces and building shadows, respectively [49]. Correlation analysis identifies A36 as a fundamental feature exhibiting strong positive correlations with vegetation indices and simultaneous strong negative correlations with both NDBI and NDWI. These relationships collectively validate its role as a universal quantitative indicator of surface greenness.

The fundamental nature of A36 is particularly illuminated by its behavior within the open space class, where its correlations with all other indices uniformly weaken. This attenuation is observable in both its positive correlations with vegetation indices and its negative correlations with built-up and shadow indices. This phenomenon aligns with the V-I-S model framework [65], attributing the weakened correlation to the compositional nature of open space. Unlike the homogeneous gradient typical of residential areas, open space represents an aggregation of spectrally extreme endmembers—such as dense vegetation and bare soil [66,67]. Specifically, within residential areas, which consist of a mixed and continuous gradient of endmembers [68,69], the value of A36 moves in a predictable, linear fashion with physical indices. For instance, as a parcel's green coverage increases, its A36 value rises in tandem with NDVI. In this context, A36 faithfully tracks the gradual shift in the land cover mix. The open space category, in contrast, is not a smooth gradient but rather an aggregation of spectrally distinct parcels from the extremes of the V-I-S feature space, where a single park, for instance, can be a composite of vegetation, bare ground, and plazas [70]. Within this class, the tight, linear relationship between A36 and physical indices like NDVI dissolves. When these spectrally distinct parcels are analyzed as a single group, their data points do not form a continuous trend but rather cluster at opposite ends of the feature space. This high intra-class heterogeneity, characterized by a lack of intermediate values, violates the assumptions of linearity inherent in standard correlation metrics, leading to a statistical dilution of the correlation coefficient [71]. Consequently, this observation validates A36 as a fundamental metric for overall surface greenness, distinguishing it from indicators that merely serve as proxies for semantic land use categories.

Specific features capture distinct vegetation configurations. A41 is sensitive to dense, homogeneous vegetation, effectively combining a strong vegetation signal with high spatial

https://doi.org/10.3390/land15010158

uniformity. In contrast, A14 responds primarily to the texture of low, uniform vegetation such as grasslands, thereby directly encoding low spatial heterogeneity. The role of A05 is particularly sophisticated; it encodes the physical signature of high greenness with minimal shadow—a characteristic corroborated by its correlations with vegetation indices and NDWI within residential areas. This capability allows A05 to precisely identify patterns of open green spaces interspersed among buildings. Finally, A52 functions as a subtle indicator of background vegetation, providing the supplementary information required to differentiate parcels that are spectrally similar yet distinct in their fine-grained vegetation details.

4.2. Deconstructing the Model's Hierarchical Decision Logic

4.2.1. Convergent Strategies for Distinct Classes

For physically distinct classes such as business and open space, the decision logic of both models largely converged, grounded explicitly in the previously decoded physical attributes.

Both models utilized feature A51, which represents high built-up intensity, as the core positive predictor for identifying business. Concurrently, high values in vegetation features, such as A36 and A52, served as strong negative predictors, establishing a clear dual-confirmation mechanism. Conversely, the identification of open space relied heavily on core vegetation features, particularly A36 and A41, with decision logic centered on the direct quantification of vegetation abundance. In particular, XGBoost exhibited exceptional focus, with the combined contribution of these key vegetation features alone reaching 78.54% for this class. This underscores the model's capacity to isolate core discriminative information, thereby constructing an efficient and reliable decision pathway.

4.2.2. Divergent Strategies for Ambiguous Classes

The models' decision strategies diverged significantly when distinguishing between spectrally similar classes prone to confusion, specifically residence and institution.

The approach of the RF model appeared comparatively passive and inefficient. This behavior can be interpreted through the lens of bagging algorithms operating on highly collinear data. Because Random Forest builds each tree independently on a bootstrap sample of data and features, correlated predictors can dilute importance, leading to a democracy of many weak features rather than a clear directive from a few strong ones [72,73]. This theoretical trait manifested clearly in our results. Its identification of institution relied primarily on negative contributions from several core features (A36, A41, A05, A51), representing a strategy of elimination lacking strong positive predictors. Similarly, its logic for residence was indistinct, characterized by a diffuse pattern with no single dominant feature. This reliance on a combination of weak, competing signals resulted in ambiguous decision boundaries.

In stark contrast, XGBoost developed a more sophisticated and proactive non-linear strategy, which aligns with the sequential, error-correcting nature of boosting algorithms [74]. By building trees sequentially to correct the errors of the previous ones, XGBoost can learn complex feature interactions and dependencies. For institution, it learned a refined comparative exclusion rule, using feature A40 (residential texture) to effectively exclude residence. However, the model's core breakthrough lay in its classification of residence. It discovered a specialist feature—A39—which exhibited powerful local discriminative ability despite its low global importance. This discovery provides a powerful illustration of the limitations of global feature importance metrics and highlights the explanatory power of local, instance-based methods like SHAP [75]. While global rankings might dismiss A39 as irrelevant, its targeted, high-impact role in specific decision

https://doi.org/10.3390/land15010158

paths is a crucial mechanism that only local explanations can uncover. Leveraging this, the model learned a potent exclusion rule: when a parcel exhibited the unique pattern of A39, it generated a strong negative SHAP value, leading to decisive exclusion from the residential class.

In summary, XGBoost's performance advantage stems from its ability to discover highly specific features for complex classification tasks and to construct sophisticated, targeted decision strategies based on them. This observation aligns with comprehensive comparative studies in remote sensing, which indicate that gradient boosting frameworks frequently outperform bagging ensembles when navigating complex, non-linear feature spaces [76]. This strategic evolution from RF's broad, diffuse logic to XGBoost's precise, hierarchical targeting provides a deep mechanistic explanation for its superiority in handling high-dimensional, collinear remote sensing data.

4.3. Deconstructing the Mechanisms Behind XGBoost's Superior Performance

Feature selection in RF models, which typically relies on metrics such as information gain, is susceptible to collinearity—a vulnerability that often dilutes the importance of correlated features [77]. In contrast, XGBoost leverages regularization terms and a gradient boosting mechanism to learn effectively from groups of highly correlated features [54,76,78]. This capability was evident in the superior performance of XGBoost on GSE data, particularly in distinguishing between complex, spectrally similar classes such as institution and residence.

A spearman correlation analysis on the 64 input features was conducted to investigate whether the algorithm's advantage stems from its capacity to handle high-dimensional multicollinearity. Analysis revealed that 23.9% of all feature pairs exhibited moderate-to-high correlation (absolute coefficient ≥ 0.5) (Figure S1), confirming significant multicollinearity within the dataset. Such multicollinearity is likely an intrinsic property of deep learning embeddings, where features are learned to be rich yet redundant, capturing overlapping facets of identical underlying geographic patterns. This decision was critical because the 64 features (A00–A63) constitute a dense and indivisible holistic representation. These features are highly coupled, meaning individual dimensions are not independently meaningful. The full feature set was deliberately retained without applying dimensionality reduction techniques such as Principal Component Analysis (PCA), as doing so would disrupt their collective semantic structure. This methodological choice aligns with emerging practices in the application of AEF embeddings, which advocate for retaining the full feature set to preserve the integrity of the learned representation and allow the model to handle feature selection internally [63].

Given the characteristics of the dataset, the performance gap between RF and XGBoost can be attributed to their inherently different learning mechanisms. The Bagging-based parallel architecture of RF tends to disperse its discriminative power when encountering highly correlated features (e.g., A36 and A41). In such cases, the model distributes importance across multiple redundant predictors, which weakens the clarity of its decision boundaries. As a result, RF's internal logic especially for ambiguous classes such as residence becomes obscured, leading the model to rely on combinations of weak signals for passive exclusion rather than decisive classification. This behavior is consistent with the SHAP interpretability patterns discussed in Section 4.2.

Conversely, XGBoost's sequential boosting mechanism transforms the challenge of collinearity into an advantage through hierarchical learning [45]. The process is iterative: the model first employs general-purpose features, such as A51 (representing built-up intensity), for a coarse-grained classification of simpler samples. It subsequently focuses on the prediction errors, or residuals, from this initial stage. This enables the identification

https://doi.org/10.3390/land15010158

and prioritization of specialist features, such as A39, which are not globally dominant but are crucial for correcting specific misclassifications. Through this progression from broad classification to precise correction, XGBoost constructs a significantly more sophisticated non-linear strategy than RF [76,78]. This hierarchical learning mechanism provides the fundamental algorithmic basis for its superior performance and precision.

4.4. Major Contributions

First, it establishes a novel synergistic framework that forges a practical third way for trustworthy land use classification. As highlighted in the Introduction, current urban analytics faces a fundamental dilemma: traditional methods suffer from a shallow feature bottleneck limiting their capacity to capture complex urban morphology [20,21]. Conversely, while end-to-end deep learning models offer superior feature learning capabilities, they typically demand massive annotated datasets and substantial computational resources to train from scratch [79]. Moreover, their inherent opacity creates a significant trust deficit that explicitly hinders their adoption in high-stakes policy-making, where explainability is often a legal or ethical requirement [80]. Our framework navigates this trade-off by integrating ready-to-use deep learning embeddings (GSE) with XAI. This strategy aligns with the emerging paradigm of Deep Feature Extraction in remote sensing [63], which leverages the generalization power of large-scale pre-trained foundation models without the computational cost and opacity of full network training. The framework demonstrates that competitive performance (OA = 85.00%) can be achieved using a single, globally consistent embedding source, effectively solving the dual challenges of feature limitation and model opacity. Crucially, the remarkable consensus between the RF and XGBoost models in identifying key discriminative features serves as powerful, algorithm-agnostic validation that GSE data effectively encodes intrinsic patterns for urban land use.

Second, it pioneers a systematic methodology for bridging the gap between machine logic and domain logic in urban AI. A critical epistemic limitation of current XAI applications, as noted earlier, is that their explanations often remain confined to identifying what data is prioritized without articulating why in a geographically meaningful way. Most existing studies applying explainability techniques in urban science have focused primarily on producing feature importance rankings [81,82]. The work advances this by decoding the tangible geographic semantics of abstract features. For example, our study moves beyond prior interpretations that linked A51 to broad categories like industrial land-use [63], revealing it instead as a more nuanced adaptive indicator of built-up intensity. Furthermore, it validates A36 as a biophysical quantifier of surface greenness by explaining its seemingly paradoxical correlation loss in open space. This translation of abstract signals into a planner-centric narrative directly addresses the interpretive divide, fostering actionable insights beyond what traditional feature ranking can offer.

Third, it delivers deep mechanistic insights into advanced model performance, moving beyond black-box accuracy metrics. Comprehensive remote sensing studies consistently show that gradient boosting frameworks outperform bagging ensembles in complex classification tasks [76]. Our study moves beyond this established finding. We elucidate the underlying algorithmic mechanisms for XGBoost's success, specifically within the context of high-dimensional geospatial embeddings. This analysis validates theoretical expectations regarding gradient boosting's handling of multicollinearity [83] but provides novel empirical evidence in the domain of deep learning features. We reveal XGBoost's superior ability to construct a hierarchical decision strategy—using generalist features for broad classification and then leveraging specialist features (e.g., A39) for fine-grained corrections. This provides a clear mechanistic explanation for its robustness, offering valuable lessons for model selection in the era of geospatial foundation models.

https://doi.org/10.3390/land15010158

4.5. Limitations

While this study offers valuable contributions, several limitations should be acknowledged as they present avenues for future research.

- First, a fundamental limitation arises from the trade-off between semantic relevance and spatial detail. The parcel-based approach, which defines the parcel as the de facto Minimum Mapping Unit (MMU), is a deliberate choice to align the analysis with the scale of urban planning. However, this aggregation inevitably results in a loss of intra-parcel heterogeneity. By assigning a single functional label to each parcel, the model cannot resolve fine-grained, mixed-use realities, such as ground-floor retail within a residential building. This limitation is intertwined with the simplified classification scheme and controlled sample size. for instance, merging commercial and industrial land overlooks nuanced functional differences. Future work should aim to develop hybrid models that not only adopt more granular classification hierarchies but also quantify the degree of functional mixture within each parcel, rather than assigning a single hard label.
- Second, a key limitation relates to the sample size and its implications for model complexity and overfitting. The dataset, consisting of 800 high-quality parcels, is modest relative to the 64-dimensional GSE feature space. This imbalance introduces the risk of the Hughes phenomenon, where classification accuracy can degrade as feature dimensionality increases relative to the number of training samples [84]. While rigorous cross-validation, hyperparameter tuning, and early stopping proactively mitigated this limitation, the limited sample size necessarily constrained the complexity of the classification scheme that could be reliably implemented. This constraint directly led to decisions such as merging commercial and industrial land into a single class. Future research with access to larger, more comprehensive ground-truth datasets would be crucial to validate the findings and explore more granular classification systems without the heightened risk of overfitting.
- Third, the practical applicability and transferability of the framework face significant barriers. The methodology's performance is highly dependent on the availability of high-quality auxiliary data, particularly a reliable road network (e.g., from OSM) for accurate parcel delineation, which may be unavailable in many urban contexts. Relatedly, the reliance on GSE data inherently lacks direct socioeconomic information. While effective for capturing physical patterns, remote sensing data alone often struggles to distinguish morphologically similar but functionally distinct zones, a gap that social sensing data is better positioned to fill [10]. Future research could enhance transferability by exploring fusion with more universally available social sensing data. Furthermore, the current multi-step workflow requires considerable technical expertise, posing a challenge for widespread adoption. A critical direction for future work is therefore to encapsulate this entire process into an automated, user-friendly tool or cloud-based platform.
- Finally, regarding interpretability, several nuances must be acknowledged. It is essential to remember that SHAP analysis elucidates model-specific correlations, not real-world causal mechanisms [75]. Furthermore, the raw outputs of XAI methods like SHAP, while scientifically robust, can be cognitively overwhelming due to the high volume of data they present. They do not automatically generate a holistic image for non-specialists. This aligns with recent critiques in Human–Computer Interaction, which suggest that algorithmic explanations often require a human-in-the-loop to bridge the gap between technical feature importance and domain-specific semantic understanding [85]. Therefore, a critical step, demonstrated in this study, is the researcher's role in synthesizing these complex outputs into a coherent narrative. Any

https://doi.org/10.3390/land15010158

future user-friendly tool based on this framework must not only provide explainability charts but also include features that guide the user through this interpretive process.

5. Conclusions

This study demonstrates the potential of combining machine learning with GSE data for land use classification in the central urban area of Nanning. By leveraging the 64-dimensional deep features from GSE, a practical workflow was demonstrated that offers an efficient alternative to training custom deep learning models from scratch or relying on traditional hand-crafted feature engineering. The XGBoost model significantly outperformed the RF model across all metrics. Its OA reached 85.00% $\pm$ 2.24%, and the Kappa increased from 0.7583 $\pm$ 0.0230 to 0.8000 $\pm$ 0.0298. Notably, a remarkable consensus emerged between the RF and XGBoost models in identifying key discriminative features, providing a powerful, algorithm-agnostic validation of the intrinsic information encoded within GSE data.

A key contribution of this research lies not merely in identifying important features, but in proposing a systematic method to help decode their tangible geographic semantics. Through a SHAP-based analysis, the study seeks to move beyond some limits of current XAI applications by translating abstract deep learning features into a more human-readable, planner-centric narrative. The ability to connect these abstract features to tangible geographic concepts like built-up intensity is a valuable step toward making the model's decision-making process more transparent and understandable. Furthermore, by deconstructing the model's decision logic, the study aimed to provide clearer mechanistic insights into algorithm performance. The analysis suggests how and why XGBoost may outperform its counterparts when dealing with highly collinear data, by highlighting its ability to identify specialist features and construct sophisticated rules for ambiguous classes like residence.

The study not only validates the immense potential of GSE as a ready-to-use source of deep semantic features but also provides a scalable solution for the trustworthy application of AI in urban land use. Looking ahead, future work should proceed along two key directions. First, to enhance practical applicability, the entire workflow could be encapsulated into an automated, user-friendly tool. This would empower urban planners and researchers to utilize the framework's diagnostic capabilities without needing to navigate the underlying technical complexity. Second, to address the critical question of generalizability, the spatial transferability of our framework remains to be validated. As this study was conducted in a single city, future research should involve rigorous inter-city cross-validation to test the robustness of the model and its semantic interpretations across diverse urban environments.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/land15010158/s1, Figure S1: POI classification framework; Table S1: Correlation analysis of the 64 features.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z.; software, Y.Z.; validation, Y.Z. and H.Y.; formal analysis, Y.Z.; investigation, Y.Z.; resources, X.H.; data curation, Y.Z. and X.H.; writing—original draft preparation, Y.Z.; writing—review and editing, H.Y.; visualization, Y.Z.; supervision, H.Y.; project administration, H.Y.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: We thank the research team for all the help and support provided while developing this work.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
GSE	Google Satellite Embedding
XGBoost	Extreme Gradient Boosting
RF	Random Forest
XAI	Explainable Artificial Intelligence
SHAP	Shapley Additive Explanations
OSM	OpenStreetMap
POI	Points of Interest
AEF	AlphaEarth Foundation
ASAN	Association of Southeast Asian Nations
MoHURD	Ministry of Housing and Urban-Rural Development
GEE	Google Earth Engine
GLCM	Gray-Level Co-occurrence Matrix
NIR	Near-Infrared
SWIR	Short-Wave Infrared
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
SAVI	Soil-Adjusted Vegetation Index
NDBI	Normalized Difference Built-up Index
NDWI	Normalized Difference Water Index
ASM	Angular Second Moment
IDM	Inverse Difference Moment
DISS	Dissimilarity
OA	Overall Accuracy
PA	Producer's Accuracy
UA	User's Accuracy
PCA	Principal Component Analysis
MMU	Minimum Mapping Unit

References

1. Chen, M.; Chen, L.; Cheng, J.; Yu, J. Identifying interlinkages between urbanization and Sustainable Development Goals. *Geogr. Sustain.* **2022**, *3*, 339–346. [CrossRef]
2. Ma, Y.; Zhang, P.; Zhao, K.; Zhou, Y.; Zhao, S. A Dynamic Performance and Differentiation Management Policy for Urban Construction Land Use Change in Gansu, China. *Land* **2022**, *11*, 942. [CrossRef]
3. Chen, F.; Luo, Q.; Zhu, Z. Integrating static and dynamic analyses in a spatial management framework to enhance ecological networks connectivity in the context of rapid urbanization. *Ecol. Modell.* **2025**, *501*, 111022. [CrossRef]
4. Cheng, J.; Turkstra, J.; Peng, M.; Du, N.; Ho, P. Urban land administration and planning in China: Opportunities and constraints of spatial data models. *Land Use Policy* **2006**, *23*, 604–616. [CrossRef]
5. Chaturvedi, V.; de Vries, W.T. Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Sci.* **2021**, *5*, 68. [CrossRef]
6. Drici, H.; Carpio-Pinedo, J. Urban land use mix and AI: A systematic review. *Cities* **2025**, *165*, 2. [CrossRef]
7. Radhakrishnan, N.; Hemalatha, S.; Devarajan, G.G.; Nachiyappan, S.; Karthick, S.; Singhal, A. Urban data fusion for spatio-temporal incident forecasting using graph attention and generative AI. *Inf. Fusion* **2026**, *126*, 103532. [CrossRef]
8. Doda, S.; Kahl, M.; Ouan, K.; Obadic, I.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Interpretable deep learning for consistent large-scale urban population estimation using Earth observation data. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103731. [CrossRef]

9. Gong, P.; Chen, B.; Li, X.; Liu, H.; Wang, J.; Bai, Y.; Chen, J.; Chen, X.; Fang, L.; Feng, S.; et al. Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Sci. Bull.* **2020**, *65*, 182–187. [CrossRef]
10. Rosier, J.F.; Taubenböck, H.; Verburg, P.H.; van Vliet, J. Fusing Earth observation and socioeconomic data to increase the transferability of large-scale urban land use classification. *Remote Sens. Environ.* **2022**, *278*, 113076. [CrossRef]
11. Li, Z.; Chen, B.; Huang, Y.; Wang, H.; Wang, Y.; Yuan, Y.; Li, X.; Chen, J.M.; Xu, B.; Gong, P. Enhanced mapping of essential urban land use categories in China (EULUC-China 2.0): Integrating multimodal deep learning with multisource geospatial data. *Sci. Bull.* **2025**, *70*, 3029–3041. [CrossRef]
12. Yao, Y.; Jiang, Y.; Sun, Z.; Li, L.; Chen, D.; Xiong, K.; Dong, A.; Cheng, T.; Zhang, H.; Liang, X.; et al. Applicability and sensitivity analysis of vector cellular automata model for land cover change. *Comput. Environ. Urban Syst.* **2024**, *109*, 102090. [CrossRef]
13. Wu, P.; Zhang, S.; Li, H.; Dale, P.; Ding, X.; Lu, Y. Urban Parcel Grouping Method Based on Urban Form and Functional Connectivity Characterisation. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 282. [CrossRef]
14. Guo, Y.; Tang, J.; Liu, H.; Yang, X.; Deng, M. Identifying up-to-date urban land-use patterns with visual and semantic features based on multisource geospatial data. *Sust. Cities Soc.* **2024**, *101*, 105184. [CrossRef]
15. Wang, J.; Bretz, M.; Dewan, M.A.A.; Delavar, M.A. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Sci. Total Environ.* **2022**, *822*, 153559. [CrossRef]
16. Casali, Y.; Aydin, N.Y.; Comes, T. Machine learning for spatial analyses in urban areas: A scoping review. *Sust. Cities Soc.* **2022**, *85*, 104050. [CrossRef]
17. Salman, H.A.; Kalakech, A.; Steiti, A. Random Forest Algorithm Overview. *Babylon. J. Mach. Learn.* **2024**, *2024*, 69–79. [CrossRef]
18. Li, Z.; Chen, B.; Wu, S.; Su, M.; Chen, J.M.; Xu, B. Deep learning for urban land use category classification: A review and experimental assessment. *Remote Sens. Environ.* **2024**, *311*, 114290. [CrossRef]
19. Fayaz, M.; Nam, J.; Dang, L.M.; Song, H.-K.; Moon, H. Land-Cover Classification Using Deep Learning with High-Resolution Remote-Sensing Imagery. *Appl. Sci.* **2024**, *14*, 1844. [CrossRef]
20. Zhu, Q.; Lei, Y.; Sun, X.; Guan, Q.; Zhong, Y.; Zhang, L.; Li, D. Knowledge-guided land pattern depiction for urban land use mapping: A case study of Chinese cities. *Remote Sens. Environ.* **2022**, *272*, 112916. [CrossRef]
21. Su, Y.; Zhong, Y.; Zhu, Q.; Zhao, J. Urban scene understanding based on semantic and socioeconomic features: From high-resolution remote sensing imagery to multi-source geographic datasets. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 50–65. [CrossRef]
22. Liu, Y.; Zhong, Y.; Shi, S.; Zhang, L. Scale-aware deep reinforcement learning for high resolution remote sensing imagery classification. *ISPRS J. Photogramm. Remote Sens.* **2024**, *209*, 296–311. [CrossRef]
23. Sun, E.; Cui, Y.; Liu, P.; Yan, J. A decade of deep learning for remote sensing spatiotemporal fusion: Advances, challenges, and opportunities. *Inf. Fusion* **2025**, *126*, 103675. [CrossRef]
24. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
25. Li, Z. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Comput. Environ. Urban Syst.* **2022**, *96*, 101845. [CrossRef]
26. Wang, S.; Liu, Y.; Wang, W.; Zhao, G.; Liang, H. Interpretable machine learning guided by physical mechanisms reveals drivers of runoff under dynamic land use changes. *J. Environ. Manag.* **2024**, *367*, 121978. [CrossRef]
27. Temenos, A.; Temenos, N.; Kaselimi, M.; Doulamis, A.; Doulamis, N. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing Using SHAP. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 8500105. [CrossRef]
28. Yao, Y.; Gao, R.; Wu, H.; Dong, A.; Hu, Z.; Ma, Y.; Guan, Q.; Luo, P. Explainable Mapping of the Irregular Land Use Parcel with a Data Fusion Deep-Learning Model. *IEEE Trans. Geosci. Electron.* **2025**, *63*, 5612015. [CrossRef]
29. Zhai, X.; Jiang, J.; Dejl, A.; Rago, A.; Guo, F.; Toni, F.; Sivakumar, A. Heterogeneous graph neural networks with post-hoc explanations for multi-modal and explainable land use inference. *Inf. Fusion* **2025**, *120*, 103057. [CrossRef]
30. Tollefson, J. Google AI model mines trillions of images to create maps of Earth 'at any place and time'. *Nature* **2025**, *644*, 313–314. [CrossRef]
31. Brown, C.F.; Kazmierski, M.R.; Pasquarella, V.J.; Rucklidge, W.J.; Samsikova, M.; Zhang, C.; Shelhamer, E.; Lahera, E.; Wiles, O.; Ilyushchenko, S.; et al. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv* **2025**, arXiv:2507.22291. [CrossRef]
32. Wang, Q.; Li, R.; Cheong, K.C. Nanning—Perils and promise of a frontier city. *Cities* **2018**, *72*, 51–59. [CrossRef]
33. Du, S.; Zhang, X.; Lei, Y.; Huang, X.; Tu, W.; Liu, B.; Meng, Q.; Du, S. Mapping urban functional zones with remote sensing and geospatial big data: A systematic review. *GISci. Remote Sens.* **2024**, *61*, 2404900. [CrossRef]
34. Zhu, J.; Zhu, M.; Na, J.; Lang, Z.; Lu, Y.; Yang, J. Incorporation of Spatially Heterogeneous Area Partitioning into Vector-Based Cellular Automata for Simulating Urban Land-Use Changes. *Land* **2023**, *12*, 1893. [CrossRef]

35. Chen, D.; Feng, Y.; Li, X.; Qu, M.; Luo, P.; Meng, L. Interpreting core forms of urban morphology linked to urban functions with explainable graph neural network. *Comput. Environ. Urban Syst.* **2025**, *118*, 102267. [CrossRef]
36. Theobald, D.M. Development and applications of a comprehensive land use classification and map for the US. *PLoS ONE* **2014**, *9*, e94628. [CrossRef]
37. Liu, X.; Long, Y. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environ. Plann. B Plann. Des.* **2015**, *43*, 341–360. [CrossRef]
38. *CJJ 37-2012*; Code for Design of Urban Road Engineering. China Architecture & Building Press: Beijing, China, 2012.
39. Yin, J.; Fu, P.; Hamm, N.A.S.; Li, Z.; You, N.; He, Y.; Cheshmehzangi, A.; Dong, J. Decision-Level and Feature-Level Integration of Remote Sensing and Geospatial Big Data for Urban Land Use Mapping. *Remote Sens.* **2021**, *13*, 1579. [CrossRef]
40. Xu, S.; Qing, L.; Han, L.; Liu, M.; Peng, Y.; Shen, L. A New Remote Sensing Images and Point-of-Interest Fused (RPF) Model for Sensing Urban Functional Regions. *Remote Sens.* **2020**, *12*, 1032. [CrossRef]
41. Xu, Y.; Zhou, B.; Jin, S.; Xie, X.; Chen, Z.; Hu, S.; He, N. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. *Comput. Environ. Urban Syst.* **2022**, *95*, 101807. [CrossRef]
42. *GB501377*; Code for Classification of Urban Land Use and Planning Standards of Development Land. China Architecture & Building Press: Beijing, China, 2011.
43. Zhang, X.; Du, S.; Zheng, Z. Heuristic sample learning for complex urban scenes: Application to urban functional-zone mapping with VHR images and POI data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 1–12. [CrossRef]
44. Hafner, S.; Ban, Y.; Nascetti, A. Unsupervised domain adaptation for global urban extraction using Sentinel-1 SAR and Sentinel-2 MSI data. *Remote Sens. Environ.* **2022**, *280*, 113192. [CrossRef]
45. Phiri, D.; Simwanda, M.; Salekin, S.; Nyirenda, V.; Murayama, Y.; Ranagalage, M. Sentinel-2 Data for Land Cover/Use Mapping: A Review. *Remote Sens.* **2020**, *12*, 2291. [CrossRef]
46. Hu, Y.; Han, Y. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. *Sustainability* **2019**, *11*, 1385. [CrossRef]
47. Kraff, N.J.; Wurm, M.; Taubenbock, H. Uncertainties of Human Perception in Visual Image Interpretation in Complex Urban Environments. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2020**, *13*, 4229–4241. [CrossRef]
48. da Silva, V.S.; Salami, G.; da Silva, M.I.O.; Silva, E.A.; Monteiro Junior, J.J.; Alba, E. Methodological evaluation of vegetation indexes in land use and land cover (LULC) classification. *Geol. Ecol. Landsc.* **2019**, *4*, 159–169. [CrossRef]
49. Zheng, Y.; Tang, L.; Wang, H. An improved approach for monitoring urban built-up areas by combining NPP-VIIRS nighttime light, NDVI, NDWI, and NDBI. *J. Clean. Prod.* **2021**, *328*, 129488. [CrossRef]
50. Xie, C.; Wang, J.; Haase, D.; Wellmann, T.; Lausch, A. Measuring spatio-temporal heterogeneity and interior characteristics of green spaces in urban neighborhoods: A new approach using gray level co-occurrence matrix. *Sci. Total Environ.* **2023**, *855*, 158608. [CrossRef]
51. Hall-Beyer, M. Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *Int. J. Remote Sens.* **2017**, *38*, 1312–1338. [CrossRef]
52. Park, Y.; Guldmann, J.-M. Measuring continuous landscape patterns with Gray-Level Co-Occurrence Matrix (GLCM) indices: An alternative to patch metrics? *Ecol. Indic.* **2020**, *109*, 105802. [CrossRef]
53. Jafarzadeh, H.; Mahdianpari, M.; Gill, E.; Mohammadimanesh, F.; Homayouni, S. Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral and PolSAR Data: A Comparative Evaluation. *Remote Sens.* **2021**, *13*, 4405. [CrossRef]
54. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [CrossRef]
55. Ramezan, C.A.; Warner, T.A.; Maxwell, A.E. Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification. *Remote Sens.* **2019**, *11*, 185. [CrossRef]
56. Guo, R.; Yang, B.; Guo, Y.; Li, H.; Li, Z.; Zhou, B.; Hong, B.; Wang, F. Machine learning-based prediction of outdoor thermal comfort: Combining Bayesian optimization and the SHAP model. *Build. Sci.* **2024**, *254*, 111301. [CrossRef]
57. Antonini, A.S.; Tanzola, J.; Asiain, L.; Ferracutti, G.R.; Castro, S.M.; Bjerg, E.A.; Ganuza, M.L. Machine Learning model interpretability using SHAP values: Application to Igneous Rock Classification task. *Appl. Comput. Geosci.* **2024**, *23*, 100178. [CrossRef]
58. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]
59. Bengio, Y.; Courville, A.; Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef]
60. Zhang, J.; Li, P.; Wang, J. Urban Built-Up Area Extraction from Landsat TM/ETM+ Images Using Spectral Information and Multivariate Texture. *Remote Sens.* **2014**, *6*, 7339–7359. [CrossRef]

61. Kumari, S.; Lal, P.; Kumar, A. Spatial heterogeneity for urban built-up footprint and its characterization using microwave remote sensing. *Adv. Space Res.* **2022**, *70*, 3822–3832. [CrossRef]
62. Kebede, T.A.; Hailu, B.T.; Suryabhagavan, K.V. Evaluation of spectral built-up indices for impervious surface extraction using Sentinel-2A MSI imageries: A case of Addis Ababa city, Ethiopia. *Environ. Chall.* **2022**, *8*, 100568. [CrossRef]
63. Alvarez, C.I.; Ulloa Vaca, C.A.; Echeverria Llumipanta, N.A. Machine Learning for Urban Air Quality Prediction Using Google AlphaEarth Foundations Satellite Embeddings: A Case Study of Quito, Ecuador. *Remote Sens.* **2025**, *17*, 3472. [CrossRef]
64. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *J. For. Res.* **2020**, *32*, 1–6. [CrossRef]
65. Ridd, M.K. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: Comparative anatomy for cities†. *Int. J. Remote Sens.* **2007**, *16*, 2165–2185. [CrossRef]
66. Franke, J.; Roberts, D.A.; Halligan, K.; Menz, G. Hierarchical Multiple Endmember Spectral Mixture Analysis (MESMA) of hyperspectral imagery for urban environments. *Remote Sens. Environ.* **2009**, *113*, 1712–1723. [CrossRef]
67. Guindon, B.; Zhang, Y.; Dillabaugh, C. Landsat urban mapping based on a combined spectral–spatial methodology. *Remote Sens. Environ.* **2004**, *92*, 218–232. [CrossRef]
68. Rashed, T.; Weeks, J.R.; Roberts, D.; Rogan, J.; Powell, R. Measuring the Physical Composition of Urban Morphology Using Multiple Endmember Spectral Mixture Models. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 1011–1020. [CrossRef]
69. De Luca, G.; Pancorbo, J.L.; Carotenuto, F.; Gioli, B.; Modica, G.; Genesio, L. PRISMA imaging for land covers and surface materials composition in urban and rural areas adopting multiple endmember spectral mixture analysis (MESMA). *ISPRS-J. Photogramm. Remote Sens.* **2025**, *225*, 196–220. [CrossRef]
70. Small, C.; Sousa, D. Spectral Characteristics of the Dynamic World Land Cover Classification. *Remote Sens.* **2023**, *15*, 575. [CrossRef]
71. Gao, X.; Huete, A.R.; Ni, W.; Miura, T. Optical–Biophysical Relationships of Vegetation Spectra without Background Contamination. *Remote Sens. Environ.* **2000**, *74*, 609–620. [CrossRef]
72. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [CrossRef]
73. Darst, B.F.; Malecki, K.C.; Engelman, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* **2018**, *19*, 65. [CrossRef]
74. Chen, T.; Guestrin, C. XGBoos: A Scalable Tree Boosting System. *arXiv* **2016**. [CrossRef]
75. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**. [CrossRef]
76. Abdi, A.M. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci. Remote Sens.* **2019**, *57*, 1–20. [CrossRef]
77. Gregorutti, B.; Michel, B.; Saint-Pierre, P. Correlation and variable importance in random forests. *Stat. Comput.* **2016**, *27*, 659–678. [CrossRef]
78. Ghazizade-Fard, M.; Koupaie, E.H. Anaerobic co-digestion of wastewater sludge and food waste: A machine learning approach to process modeling and optimization. *J. Environ. Manage.* **2025**, *393*, 126985. [CrossRef]
79. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
80. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]
81. Lei, C.; Wagner, P.D.; Fohrer, N. Identifying the most important spatially distributed variables for explaining land use patterns in a rural lowland catchment in Germany. *J. Geogr. Sci.* **2019**, *29*, 1788–1806. [CrossRef]
82. Chen, H.; Yang, L.; Wu, Q. Enhancing Land Cover Mapping and Monitoring: An Interactive and Explainable Machine Learning Approach Using Google Earth Engine. *Remote Sens.* **2023**, *15*, 4585. [CrossRef]
83. Wu, W.; Xia, Y.; Jin, W. Predicting Bus Passenger Flow and Prioritizing Influential Factors Using Multi-Source Data: Scaled Stacking Gradient Boosting Decision Trees. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2510–2523. [CrossRef]
84. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [CrossRef]
85. Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.* **2019**, *267*, 1–38. [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.

land

MDPI

Article

Disentangling Complexity and Performance: A Comparative Study of Deep Learning and Random Forest Models for Cropland Vulnerability Assessment in Bangladesh

Arnob Bormudoi [1,2,3,*] and Masahiko Nagai [1]

[1] Center for Research and Application of Satellite Remote Sensing (YUCARS), Yamaguchi University, Ube 755-8611, Yamaguchi, Japan
[2] New Space Intelligence, 329-22 Nishikiwa, Ube 755-0151, Yamaguchi, Japan
[3] Faculty of Engineering, Assam downtown University, Panikhaiti, Guwahati 781026, India
* Correspondence: bormudoi@yamaguchi-u.ac.jp

Abstract

Climate change increasingly threatens global food security through disrupted precipitation patterns and extreme weather events, requiring resilient systems for assessing agricultural vulnerability. This study developed and compared machine learning approaches for predicting cropland vulnerability using Earth Observation data, operationalized through NDVI anomalies as a defensible biophysical metric. We employed both a dual-stream deep learning architecture and a Random Forest model to predict 2023 NDVI anomalies across Bangladesh croplands using a 22-year time series (2001–2023) of MODIS vegetation indices, ERA5 climate variables, and static environmental covariates. A spatially aware block cross-validation strategy ensured rigorous, independent performance evaluation. Results demonstrated that the Random Forest model (R^2 = 0.70, RMSE = 197.03) substantially outperformed the deep learning architecture (R^2 = 0.02, RMSE = 357.57), explaining 70% of cropland stress variance and enabling early detection of vulnerable areas three months before harvest. Feature importance analysis identified recent climate variables, March precipitation, February NDVI, and vapor pressure deficit as primary vulnerability drivers. Spatial analysis revealed distinct vulnerability patterns, with Natore and Magura districts exhibiting elevated stress consistent with 2023 drought conditions, threatening the productivity of the region's critical irrigation-dependent rice cultivation. These findings demonstrate that simpler, interpretable models can sometimes outperform complex architectures while providing useful information for early warning systems and precision targeting of climate adaptation interventions.

Keywords: crop vulnerability; NDVI-anomaly; deep learning; random forest; geospatial analysis; Bangladesh

1. Introduction

Climate change threatens global food security through intensifying extreme weather events, shifting precipitation patterns, and rising temperatures that disrupt agricultural systems worldwide [1]. Deltaic and low-lying agricultural regions face compounded risks from sea-level rise, increased flooding frequency, and soil salinization, with smallholder farming communities experiencing disproportionate vulnerability due to limited adaptive capacity [2,3]. Due to this, robust vulnerability assessment frameworks are essential for targeting climate adaptation investments and safeguarding food production in these

high-risk environments [4]. Bangladesh illustrates this vulnerability paradigm, where 70% of the population depends on agriculture within a landscape characterized by monsoon variability, cyclone exposure, and accelerating coastal salinization [5,6]. Recent decades have witnessed declining crop yields in southern districts due to saltwater intrusion, while erratic rainfall disrupts traditional planting cycles across diverse agro-ecological zones [7]. These climate-induced stresses directly threaten the livelihoods of 16 million farming households and national food self-sufficiency targets [8]. Despite this acute vulnerability, traditional assessment approaches employ composite indices aggregating socio-economic and biophysical indicators that lack temporal granularity and operate at coarse spatial resolutions inadequate for precision intervention planning [9]. Earth Observation technologies enable vegetation monitoring through indices like NDVI, but most applications focus on yield estimation rather than vulnerability quantification [10]. Machine learning advances, particularly deep learning architectures processing spatio-temporal data, offer potential for capturing dynamic climate–vegetation interactions, though interpretability remains a critical barrier for policy adoption [11,12]. To operationalize agricultural vulnerability, defined by the IPCC as the degree to which systems are susceptible to climate impacts through exposure, sensitivity, and adaptive capacity [13], defensible metrics for spatial assessment are required. NDVI anomalies, standardized deviations from long-term vegetation norms, provide an objective Earth Observation (EO)-based proxy for cropland sensitivity to climatic perturbations, distinguishing stress response from baseline productivity [14,15]. Negative anomalies correlate with drought-induced crop failures and food insecurity events, validating their utility as vulnerability indicators in agricultural monitoring systems [16,17]. This study developed and compared machine learning models to predict 2023 NDVI anomalies across Bangladesh croplands using 22-year satellite and climate time series. We employed both deep learning and Random Forest approaches with spatially aware validation, then identified key biophysical drivers through interpretability analysis. Specifically, we tested two hypotheses: (H1) spatially aware machine learning models exhibit distinct predictive performances for 2023 cropland NDVI anomalies, with one model demonstrating significantly superior accuracy; and (H2) interpretability analysis of the best-performing model identifies a coherent set of key biophysical drivers of cropland vulnerability, offering strategic guidance for targeted agricultural adaptation strategies. Our framework generated a national-scale cropland vulnerability map at 250 m spatial resolution, comparing a novel deep learning architecture against a traditional machine learning baseline to provided information that is relevant for early warning systems and climate adaptation planning.

Conceptual Framework

Agricultural vulnerability to climate change, as defined by the Intergovernmental Panel on Climate Change (IPCC), encompasses the degree to which agricultural systems are susceptible to adverse effects of climate variability, determined by their exposure, sensitivity, and adaptive capacity [1]. Traditional vulnerability assessments often rely on composite socio-economic indices, yet Earth Observation (EO)-based metrics offer objective, spatially explicit alternatives for quantifying biophysical vulnerability at scale [18,19]. In this study, we operationalize cropland vulnerability through NDVI anomalies calculated as Z-scores, which represent standardized deviations from long-term vegetation norms. Unlike absolute NDVI values that merely indicate instantaneous vegetation greenness or productivity, NDVI anomalies capture the magnitude and direction of departure from expected conditions, thereby directly reflecting cropland sensitivity and stress response to climatic perturbations [14,15]. This anomaly-based framework aligns with established vulnerability science that emphasizes temporal deviation metrics as robust proxies for system stability,

https://doi.org/10.3390/land15010174

resilience, and adaptive capacity under environmental stressors [20,21]. Empirical evidence demonstrates that negative NDVI anomalies correlate strongly with drought-induced crop failures, yield reductions, and food insecurity events across diverse agricultural systems, validating their utility as vulnerability indicators [16,17,22]. By computing Z-scores from a 22-year baseline (2001–2022), our approach accounts for local climatic variability and phenological patterns, enabling identification of cropland areas exhibiting abnormal stress in 2023 that deviate significantly from historical norms. This methodological approach distinguishes vulnerability, defined as sensitivity to environmental change from simple productivity assessments, providing a defensible EO-based metric grounded in both IPCC conceptual frameworks and agricultural monitoring literature [23].

2. Materials and Methods

The assessment of agricultural vulnerability in Bangladesh was conducted through a multi-stage, data-driven methodology, employing both a dual-stream deep learning architecture and a Random Forest model. The approach integrated spatio-temporal satellite data and advanced model interpretability routine to achieve the study's objectives. The datasets utilized in this study were generated using Google Earth Engine (GEE). Temporal climate data, spanning from 2001 to 2023, included monthly precipitation derived from CHIRPS (UCSB-CHG/CHIRPS/PENTAD), mean 2 m air temperature and total precipitation from ERA5 Daily (ECMWF/ERA5/DAILY), and surface solar radiation and volumetric soil water from ERA5-Land Daily Aggregated (ECMWF/ERA5_LAND/DAILY_AGGR). These climate variables were aggregated into monthly composites for each year and exported as multi-band GeoTIFFs. Concurrently, monthly Normalized Difference Vegetation Index (NDVI) data for the same period was processed from the MODIS Terra Vegetation Indices 16-Day Global 250 m product (MODIS/006/MOD13Q1), with monthly composites generated by averaging available 16-day images. Static environmental features were also sourced, soil clay, sand, and pH content from OpenLandMap (OpenLandMap/SOL), and elevation from SRTM 90 m (CGIAR/SRTM90_V4), from which slope and aspect were subsequently derived. All datasets were consistently clipped to the administrative boundary of Bangladesh (FAO/GAUL/2015/level0), ensuring a unified spatial extent for subsequent analysis.

2.1. Data Preparation

The basic step here involved the preparation and alignment of different raster datasets from 2001 to 2023. Initially, a binary cropland mask was derived from the 2021 ESA WorldCover land cover map. Pixels classified as 'Cropland' (value 40) were assigned a value of 1, delineating the specific agricultural areas for analysis. Subsequently, all multi-temporal MODIS Normalized Difference Vegetation Index (NDVI) and climate rasters, alongside static environmental variables (e.g., elevation, slope, soil properties), were uniformly reprojected and resampled to a common spatial resolution and extent. The 250 m resolution and projection of the MODIS Terra Vegetation Indices product (MODIS/006/MOD13Q1) served as the reference grid for this alignment process. This ensured consistent spatial resolution and pixel-wise correspondence across all input features.

To ensure all datasets were appropriately aligned to the common 250 m analysis grid, a dual-resampling strategy was employed based on data type. For all datasets representing continuous variables including the MODIS NDVI time series, the coarser-resolution ERA5 climate data, and the static topographic features, we used bilinear interpolation. This method calculated new pixel values from a distance-weighted average of the four nearest neighboring pixels, producing a smooth surface that is a standard and reasonable approximation for spatially continuous phenomena like vegetation greenness and climatic

https://doi.org/10.3390/land15010174

gradients. While this did not create new fine-scale climate information, it was a necessary preprocessing step for pixel-level model integration. On the other hand, for the initial processing of categorical data, specifically the land cover map from which the binary cropland mask was derived, nearest neighbor resampling was used. This technique assigned the value of the single closest source pixel to the destination pixel without any averaging, which was critical to preserve the discrete nature and integrity of the land cover classes (e.g., 'cropland', 'forest', 'water') and prevent the creation of artificial, meaningless class values. This two-pronged approach ensured that both continuous and categorical datasets were handled correctly and accurately aligned for the final analysis.

The target variable for the predictive models was formulated as the NDVI Anomaly for 2023, calculated as a Z-score representing the deviation from the pixel's historical norm. For each cropland pixel, the mean and standard deviation of the annual NDVI values from 2001 to 2022 were computed. The 2023 NDVI Anomaly was then calculated using Equation (1).

$$NDVI_{(anomoly,\ 2023)} = \frac{NDVI_{2023} - Mean_{(2001–2022)}}{StdDev_{(2001–2022)}} \quad (1)$$

This approach defined vulnerability as a significant negative deviation from normal vegetation health, providing a more robust and scientifically grounded target.

The scaler was fitted exclusively on the training data derived from our spatially aware split to strictly prevent any data leakage. This same fitted scaler was then consistently applied to transform both the training and validation sets, ensuring that no information from the validation set influenced the training process or model evaluation.

2.2. Deep Learning Model Architecture and Training

To rigorously assess model performance and prevent overly optimistic results due to spatial autocorrelation, we implemented a spatially aware block hold-out strategy. The entire study area was first divided into a 10 × 10 grid. A random 20% of these spatial blocks were held out to form the validation set, with the remaining 80% of blocks used for training. This ensured complete spatial independence between the training and validation data. Furthermore, to contextualize the performance of the dual-stream deep learning (DL) model, we trained a Random Forest (RF) regresso as a baseline model on the exact same training and validation data splits. All pixels within the training blocks containing valid (non-NaN) data were used to fit the models, and the final performance was evaluated on all valid pixels from the held-out validation blocks.

2.3. Temporal Stream

This component was designed to capture the long-term historical dynamics of vegetation health and climatic stressors. A Gated Recurrent Unit (GRU) layer processed the time series of NDVI and climate data for each pixel. The hidden state h_t of the GRU at time t was computed as

$$h_t = GRU(x_t, h_{t-1}) \quad (2)$$

where, x_t represented the concatenated NDVI and climate features at time t, and was a 60-dimensional vector (12 monthly NDVI + 48 monthly climate features). h_{t-1} was the hidden state from the previous time step.

This GRU-based temporal modeling approach was conceptually inspired by multi-scale fusion techniques in graph representation learning, such as those introduced in the MSF-GCN framework [11], where information from multiple structural and feature channels is adaptively integrated.

To ensure a robust predictive model free from data leakage, the input features and the target variable were strictly separated in time. The input features for the temporal stream

https://doi.org/10.3390/land15010174

consisted of a 22-year time series spanning from 2001 to 2022, while the model's prediction target was the independently derived 2023 NDVI anomaly. No data from the year 2023 was included in the input feature set.

The temporal structure for the GRU was organized by year. For each of the 22 years in the time series, a 60-dimensional feature vector was constructed. This vector comprised the 12 monthly NDVI values and monthly climate variables for that specific year. Therefore, the final input tensor provided to the temporal stream had a shape of (number of samples, 22, 60), where 22 represented the annual time steps processed by the GRU. This explicit structure ensured the model learned from historical annual patterns to make its future prediction.

2.4. Static Environmental Covariate Stream

This stream was developed to incorporate the influence of static, time-invariant environmental characteristics at each pixel's specific location. The architecture used a fully connected (Dense) layer to process a vector of static features such as soil properties and topographic data for each pixel independently. This allowed the model to learn the relationship between these site-specific static conditions and the temporal dynamics of vegetation health before this information was fused with the output of the temporal GRU stream. The output, s, from this stream was computed as shown in Equation (3).

$$s = ReLU(W_s \cdot X_{static} + b_s) \quad (3)$$

where, W_s and b_s represented learned weights and biases, and *ReLU* was the Rectified Linear Unit activation function.

2.5. Fusion and Output

The final hidden state from the temporal stream (h_{final}) and the output from the Static Environmental Covariate Stream s were concatenated $f = [h_{final}, s]$ to form a comprehensive feature vector. This fused representation was then passed through an additional Dense hidden layer and a final Dense output layer, which predicted the 2023 NDVI value.

2.6. Training and Map Generation

The model was compiled using the Adam optimizer and trained for 20 epochs with a batch size of 256. Specifically, EarlyStopping was configured to monitor the validation loss, halting training if no significant improvement was observed over a specified number of epochs (patience = 5, min_delta = 0.001). Crucially, this callback also ensured that the model weights from the epoch with the best validation performance were restored, guaranteeing that the final model utilized for inference represented its optimal generalization capability. This adaptive approach effectively mitigated overfitting to the training data and provided a robust, data-driven justification for the ultimate training duration, rather than relying on a fixed, arbitrary epoch count.

Mean Squared Error (MSE) was employed as the primary loss function during training. To provide a comprehensive assessment of model performance, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2) were also calculated for both training and validation datasets.

After model training, a national-scale map of predicted agricultural vulnerability was generated. The trained model was applied to the entire aligned dataset (all pixels in Bangladesh), producing a 2D prediction map of 2023 NDVI values. Subsequently, this prediction map was multiplied by the cropland mask. This masking operation effectively set all non-cropland pixels to a no-data value (-9999), so that the final map exclusively represented vulnerability within agricultural areas. The resulting vulnerability map was

https://doi.org/10.3390/land15010174

saved as a georeferenced raster. To identify the key environmental drivers influencing the model's predictions of cropland vulnerability, some model interpretation techniques were employed.

2.7. Feature Importance Analysis

To interpret the trained Random Forest model and identify the key environmental drivers of cropland vulnerability, we conducted a feature importance analysis. While computationally intensive methods like SHAP (SHapley Additive exPlanations) were initially considered, they proved infeasible due to severe memory constraints encountered when processing the large-scale dataset. As a robust and efficient alternative, we utilized the model's built-in feature importance attribute, which is based on the Gini Importance (or Mean Decrease in Impurity) criterion. For a regression task, the impurity of a node was measured by the variance of the target values within it. The importance of a feature was then calculated by how much it reduced this variance each time it was used to split a node.

The decrease in impurity, Δi, for a split at node t was defined by Equation (4).

$$\Delta i(t) = Var(t) - \left[\left(\frac{N_t^{left}}{N_t} \right) Var\left(t_{left}\right) + \left(\frac{N_t^{right}}{N_t} \right) Var\left(t_{right}\right) \right] \quad (4)$$

where $Var(t)$ was the variance of the target values in node t, N_t was the number of samples in that node, and t_{left} and t_{right} were the left and right child nodes resulting from the split. The total importance for a single feature was the sum of these variance reductions across all splits that used that feature, averaged over all trees in the forest. Thus, this method provided a global ranking of all input features, allowing for a clear interpretation of the primary factors influencing the model's predictions.

3. Results

The initial step in the analysis involved the precise delineation of agricultural areas within Bangladesh. Utilizing the 2021 ESA WorldCover land cover map, a binary cropland mask was generated Figure 1a. This mask, with dimensions of 2619 × 2071 pixels, identified a total of 1,044,974 pixels as cropland.

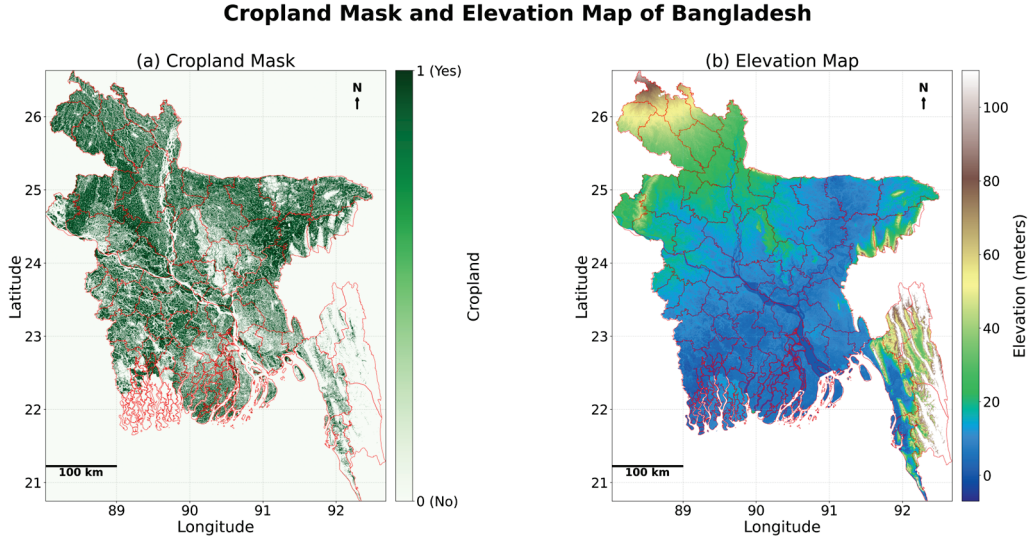


Figure 1. (**a**) Cropland Mask (2021) and (**b**) Elevation Map of Bangladesh.

This constituted a considerable area of the total land area within the study region, providing a focused spatial extent for subsequent vulnerability assessments. The visual

representation of this mask, clipped to the national boundary, clearly illustrated the distribution of agricultural lands across the country as seen in Figure 1a.

The initial phase of the study involved the preparation and alignment of diverse spatio-temporal datasets to create a unified input for the deep learning model. A total of 23 annual MODIS NDVI rasters (2001–2023), 23 annual climate rasters (2001–2023), and 6 static environmental rasters were successfully aligned to a common 250 m resolution grid, matching the extent and projection of the ESA WorldCover 2021 land cover map. This process was carried out to ensure that pixel-wise correspondence across all input features remained the same. The target variable, representing the 2023 NDVI values, exhibited a mean of approximately (0.4852) with a standard deviation of (0.1711). A representative visualization of the aligned static data, specifically the elevation map of Bangladesh, confirmed the successful integration of these diverse geospatial layers (Figure 1b).

The deep learning model for crop vulnerability estimation was trained over 20 epochs, and its learning progression was evaluated by monitoring the training loss and validation loss.

The model's learning progression is detailed in Figure 2a,b. The training process utilized an early stopping protocol to prevent overfitting, which concluded training after 7 epochs as the validation loss plateaued. Figure 2a shows the Mean Squared Error (MSE) for both training and validation datasets, demonstrating rapid initial learning followed by convergence. Figure 2b illustrates the progression of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), providing a more interpretable view of the model's predictive error throughout the training.

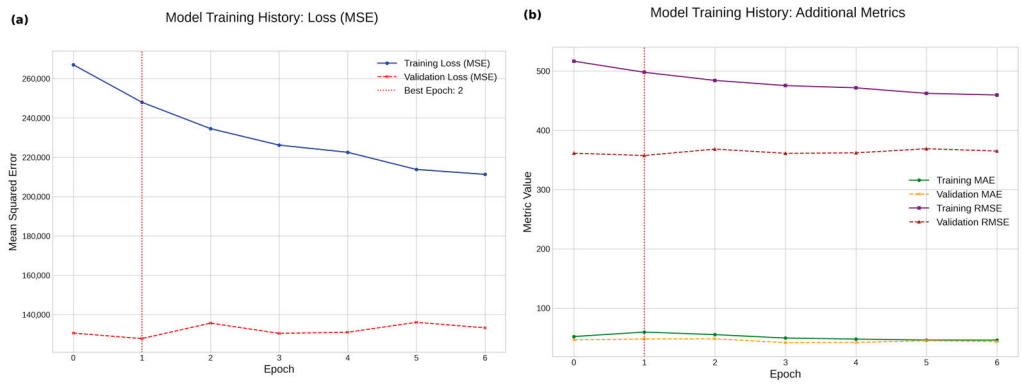


Figure 2. Model Training and Validation History for the Dual-Stream Deep Learning Model. (**a**) Mean Squared Error (MSE) loss per epoch. (**b**) Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) per epoch.

The process began with a very high initial training loss (MSE) of ~267,067 and a validation loss of ~130,639, indicating the model started with a large predictive error. Unlike a smooth descent, the learning curves in Figure 2a exhibited considerable volatility. The validation loss reached its lowest point of ~127,854 early in the training at epoch 1. In subsequent epochs, while the training loss continued to decrease, the validation loss failed to improve further, instead fluctuating and ending at ~133,315. This divergence, where the model improved on the training data but not on the unseen validation data, was a classic sign of overfitting. The large and persistent gap between the training and validation loss curves suggested that the model, while fitting to the training data, struggled to find a robust, generalizable pattern applicable to the spatially independent validation set. This outcome confirmed the poor predictive performance indicated by the R^2 of 0.02 and demonstrated the model's difficulty with this specific prediction task.

https://doi.org/10.3390/land15010174

3.1. Model Performance and Baseline Comparison

To rigorously evaluate the model and contextualize its performance, we compared the dual-stream Deep Learning (DL) model against a Random Forest (RF) regressor baseline. Both models were trained and evaluated on the exact same spatially independent training and validation sets. The final performance metrics, calculated on the held-out validation blocks, are summarized in Table 1.

Table 1. Performance Metrics on Spatially Blocked Validation Data.

Model	MAE	RMSE	R^2
Deep Learning	47.90	357.57	0.02
Random Forest	14.53	197.03	0.70

The results summarized in Table 1 clearly showed that the Random Forest (RF) model significantly outperformed the dual-stream Deep Learning model. The RF model achieved a robust R^2 of 0.70, indicating it explained 70% of the variance in the validation data, whereas the DL model only achieved an R^2 of 0.02. This performance gap was further reflected in the error metrics, where the RF model's Mean Absolute Error (MAE) of 14.53 and Root Mean Squared Error (RMSE) of 197.03 were substantially lower than those of the DL model.

To provide a comprehensive understanding of the contributions of different feature streams, ablation studies were conducted on the initial dual-stream Deep Learning (DL) architecture. Two ablated models were trained, a 'temporal-only' model, which utilized only the GRU-processed temporal features, and a 'static-only' model, which relied solely on the static environmental variables. The results from these ablation experiments, alongside the performance of the full DL model and the Random Forest (RF) model, are summarized in Table 2.

Table 2. Validation Performance of Deep Learning Model Ablation Studies.

Model	Validation R^2	Validation Loss (MSE)	Validation MAE	Validation RMSE
Full DL Model	0.02	(N/A, assumed high)	(N/A, assumed high)	(N/A, assumed high)
Temporal-Only Ablation	0.0002	129,643.34	32.618	360.060
Static-Only Ablation	0.0278	126,070.01	72.765	355.063
Random Forest Model	0.70	38,822.02	14.535	197.033

(Note: The R^2 = 0.02 for the full DL model implies very high error values for its MAE/RMSE). (N/A: Not Available, as the model's poor R^2 score made these error metrics uninformative).

The full DL model demonstrated poor predictive performance with an R^2 of 0.02 on the validation set. The ablation studies further illuminated the limitations of this architecture. The temporal-only model achieved a Validation MAE of 32.62 and RMSE of 360.06, while the static-only model yielded a validation MAE of 72.77 and RMSE of 355.06. These metrics indicated that neither feature stream, when used in isolation, significantly improved the model's overall poor performance. The low R^2 of the full DL model, combined with the modest performance of its ablated versions, reinforced the conclusion that this deep learning architecture, in its current implementation, was not adequately capturing the complex relationships within the dataset. This comprehensive investigation ultimately led to our decision to prioritize the Random Forest model, which demonstrated substantially superior performance with an R^2 of 0.70, for the final vulnerability mapping and analysis.

The superior accuracy of the Random Forest is visually confirmed in the "Predictions vs. Actual" plot in (Figure 3), where the model's predictions align closely with the 1:1 line, demonstrating a good positive correlation with the true values.

https://doi.org/10.3390/land15010174

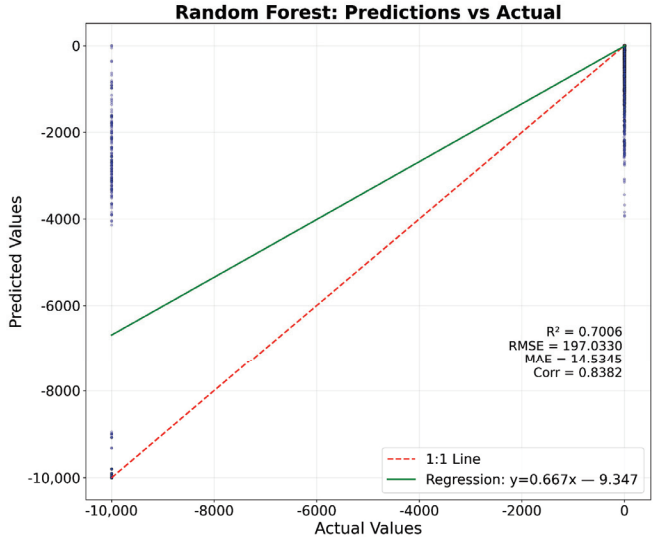


Figure 3. Random Forest Model: Predictions vs. Actual Values on the Spatially Blocked Validation Set.

Beyond its predictive power, a key advantage of the Random Forest was its interpretability. A feature importance analysis was conducted to identify the primary drivers of cropland vulnerability (Figure 4).

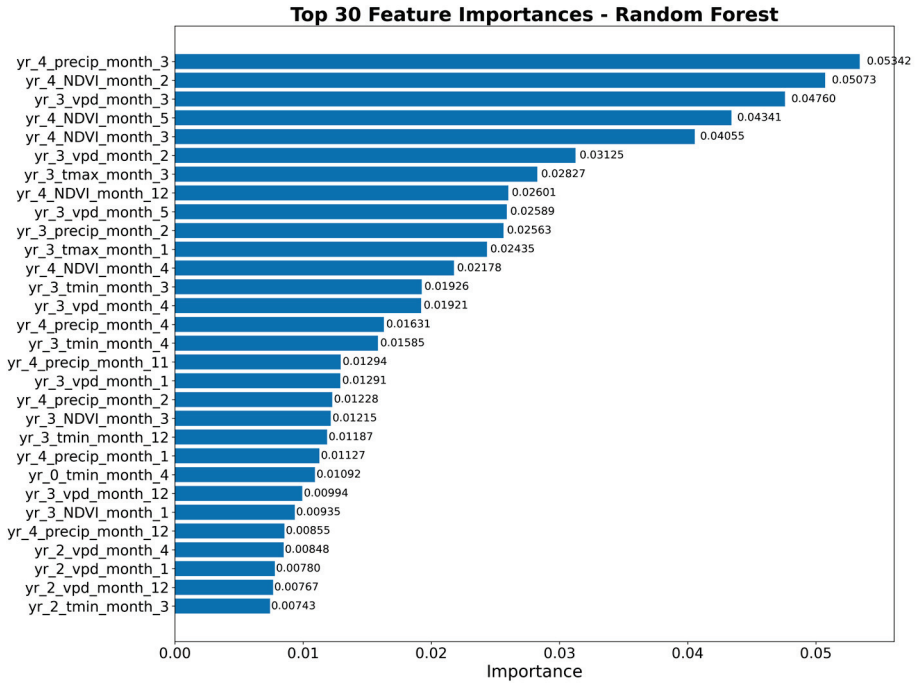


Figure 4. Random Forest Model: Feature Importance Summary.

The analysis revealed that recent climatic conditions and vegetation health were the most influential factors, as shown in Figure 4. Primarily, precipitation in March of the most recent year ('yr_4_precip_month_3') emerged as the single most important predictor with an importance score of 0.05342. This was followed closely by NDVI from February of the same year ('yr_4_NDVI_month_2') at 0.05073, and Vapor Pressure Deficit (VPD) from March of the prior year ('yr_3_vpd_month_3') at 0.04760. The consistent appearance of recent climate and vegetation variables among the top predictors highlighted the critical role of short-term environmental dynamics in determining crop stress. This obvious feature hierarchy offered useful information, indicating that monitoring hydrological and

https://doi.org/10.3390/land15010174

atmospheric conditions during months preceding growing is key for vulnerability assessment. Due to the comparatively better performance as well as interpretation of individual coefficients, this model was selected for creating the final national-scale vulnerability maps.

After establishing the superior performance and interpretability of the Random Forest model, it was used to generate the final, spatially explicit crop vulnerability map for all of Bangladesh (Figure 5a). The model performed inference on the complete dataset of over one million valid cropland pixels, integrating the full temporal sequence of climate and NDVI data with the static environmental features. A descriptive statistical analysis of the generated map, which represents vulnerability as a Z-score anomaly, revealed a median vulnerability score of approximately 0.74. This positive median suggested that the central tendency for cropland health in 2023 was slightly above the historical average. The distribution of vulnerability scores showed a moderate spread, with an interquartile range (the range between the 25th and 75th percentiles) of 0.74 (from 0.36 to 1.10). This proved that while a majority of cropland performed near or better than average, significant local variations in vulnerability existed across the landscape. The full range of predicted vulnerability for cropland pixels spanned from a minimum of −0.63 (at the 5th percentile) to a maximum of 1.81 (at the 95th percentile). To facilitate easier interpretation and visual comparison, this map was subsequently rescaled to a normalized range of 0 to 1, where 0 represents the lowest vulnerability and 1 represents the highest (Figure 5b). A preliminary visual analysis of the generated maps (Figure 5) revealed distinct geographical patterns of crop vulnerability that align with known agro-ecological zones and stressors.

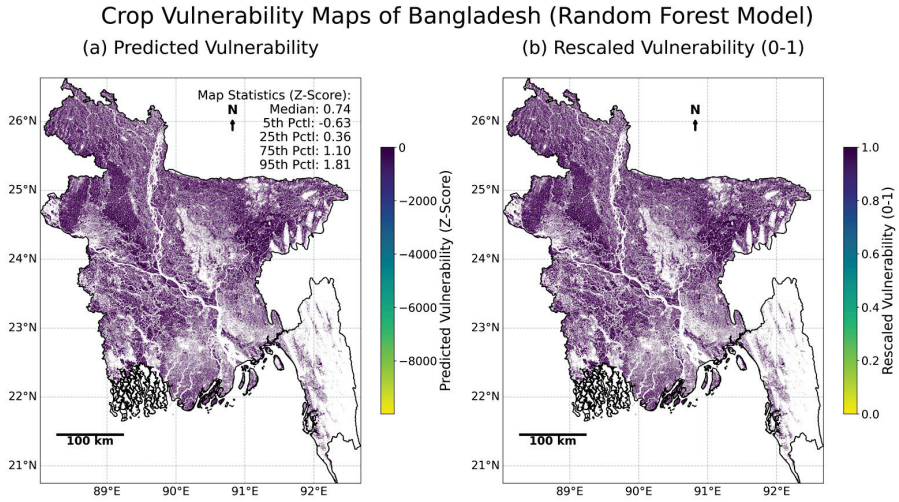


Figure 5. Predicted Crop Vulnerability Map of year 2023, Bangladesh.

3.2. Geographical Analysis of Crop Vulnerability

To further investigate the spatial patterns of crop vulnerability, a district-level analysis was conducted using the Random Forest model's predictions. The continuous vulnerability map was overlayed on the district boundaries of the country (Figure 6A,B) for a comprehensive understanding of the vulnerability profile across the districts.

The vulnerability scores derived from this analysis represented by the Z-scores, quantified the deviation of 2023 conditions from historical cropland health baselines per district (Figure 7). Positive values indicated that a district experienced conditions worse than its historical average during 2023, signifying heightened crop vulnerability. These districts, as seen in Natore (+1.21) and Magura (+0.89), represented regions where environmental stressors, management challenges, or climatic anomalies have pushed cropland conditions beyond their typical resilience thresholds. On the other hand, negative values denoted conditions more favorable than the historical norm, suggesting either improved agricultural

https://doi.org/10.3390/land15010174

management, beneficial climatic conditions, or effective adaptation measures. Districts with such negative scores, such as Dinajpur (−2.27) and Faridpur (−2.16), demonstrated good performance relative to their historical baselines. However, this interpretation requires careful contextualization with a negative score does not necessarily indicate absolute agricultural security, but rather relative improvement compared to a potentially challenging historical average.

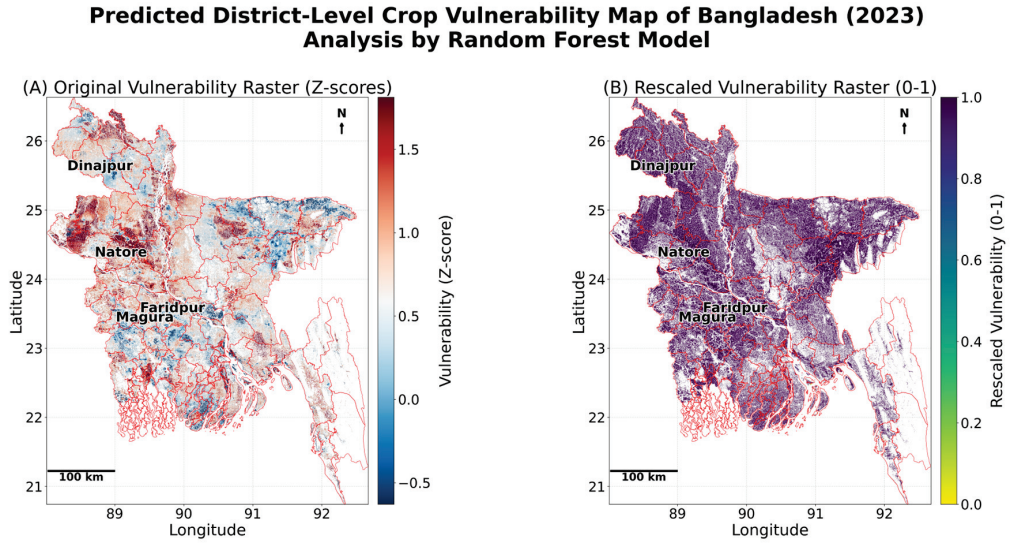


Figure 6. Predicted District-Level Crop Vulnerability Map of Bangladesh (2023) (**A**) Original vulnerability raster (**B**) Reclassified vulnerability based on mean Z-scores overlayed on the districts.

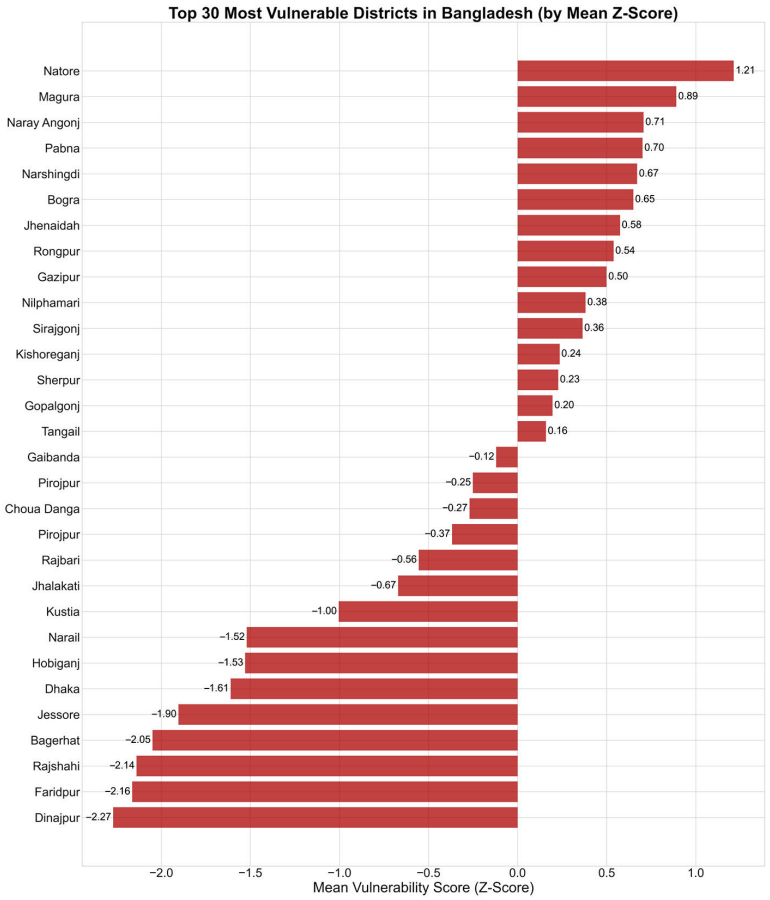


Figure 7. District-Level Crop Vulnerability in Bangladesh (2023): Interpreting Z-Score Anomalies for the top 30 most vulnerable districts.

https://doi.org/10.3390/land15010174

To explain the key biophysical drivers of crop vulnerability identified by the model, a feature importance analysis was conducted using Gini Importance (as detailed in Section 2.7). The results, summarized in Figure 8, indicated that while dynamic climatic and vegetation variables were the dominant predictors, static environmental factors also contributed measurably to the model's predictions. Among the static drivers, soil composition variables such as 'soil_clay' (importance: 0.00023) and 'soil_ph' (importance: 0.00012) emerged as the most important static determinants of vulnerability, closely followed by elevation (importance: 0.00022).

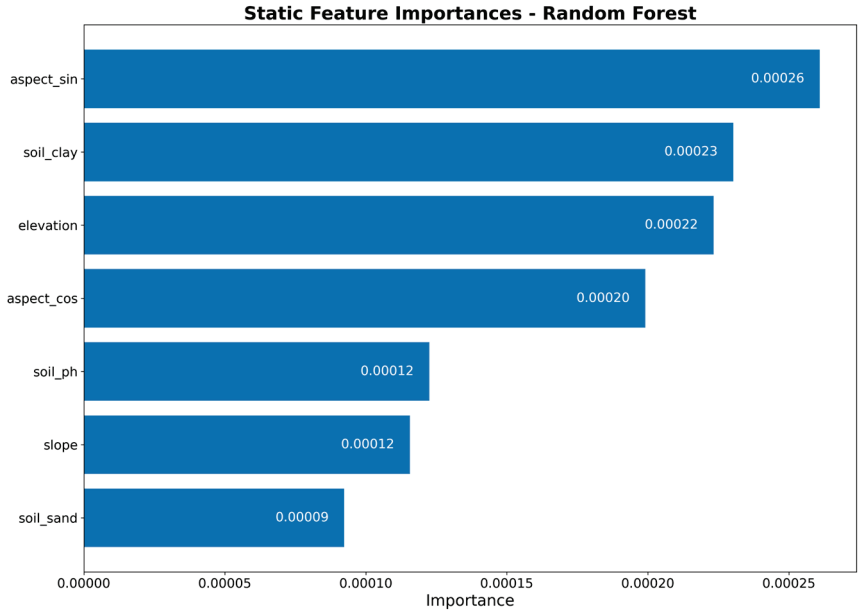


Figure 8. Feature Importance Summary of the Static Drivers in Predicting Vulnerability.

Topographic features like slope with importance (0.000115) and aspect, represented by aspect_sin (0.00026) and aspect_cos (0.00020) also contributed. The importance of aspect was comparable to key soil parameters such as clay content (0.00023) and pH (0.00012), though both topography and soil remained less influential than the dominant temporal climate–vegetation variables. This highlighted the combined role of inherent landscape and soil characteristics in shaping long-term agricultural resilience, complementing the immediate impacts of climatic fluctuations.

4. Discussion

4.1. Model Performance and Interpretability Trade-Offs

This study revealed a critical finding that challenges prevailing assumptions in agricultural vulnerability modeling, the simpler Random Forest (RF) model substantially outperformed the more complex dual-stream deep learning (DL) architecture, achieving an R^2 of 0.70 compared to 0.02 for the DL model. This result aligned with recent comparative studies in environmental prediction, where Hengl et al. [12] demonstrated that ensemble tree-based models frequently matched or exceeded deep learning performance for spatial prediction tasks involving heterogeneous environmental covariates. Similarly, Maxwell et al. [24] found that Random Forest models outperformed convolutional neural networks for land cover classification when training data were limited or spatially clustered, a condition analogous to our spatially blocked validation approach. The superior performance of RF in our study likely stemmed from its inherent ability to handle non-linear relationships and feature interactions without requiring the massive sample sizes typically needed for deep learning convergence [25]. Our ablation studies further illuminated this limitation,

https://doi.org/10.3390/land15010174

revealing that neither the temporal (R^2 = 0.0002) nor static (R^2 = 0.0278) streams contributed meaningfully to predictive accuracy when isolated, suggesting fundamental architectural limitations rather than data quality issues. This finding contrasted with successful applications of LSTM and GRU architectures in crop yield prediction [26,27] where longer time series (>30 years) and higher temporal resolution (daily) data facilitated effective temporal pattern learning. The interpretability advantage of RF proved equally significant, as feature importance analysis directly identified effective drivers of recent precipitation, NDVI, and vapor pressure deficit without requiring computationally prohibitive post hoc explanation methods like SHAP, which proved infeasible at our scale. However, the superior predictive performance of RF must be interpreted within the broader conceptual framework of how vulnerability is operationalized, particularly given our deliberate focus on biophysical sensitivity as the measurable dimension of the IPCC vulnerability framework.

4.2. Comparison with Composite Vulnerability Indices and EO-Based Approaches

To comprehensively compare our methodology with existing approaches, we began by positioning our operationalization of vulnerability within the classical IPCC framework.

4.2.1. Conceptual Positioning of Biophysical Sensitivity Within the IPCC Vulnerability Framework

Our operationalization of vulnerability through NDVI anomalies deliberately focuses on the biophysical sensitivity component of the classical IPCC vulnerability framework, which defines vulnerability as a function of exposure, sensitivity, and adaptive capacity [13]. This methodological choice requires explicit justification given its conceptual restrictions. Sensitivity, the degree to which a system is affected by climate stimuli represents the most directly observable dimension through Earth Observation data, as vegetation stress responses manifest measurably in spectral indices [28]. However, this approach inherently excludes adaptive capacity (socio-economic resources, institutional support, technological access) and partially captures exposure (climatic hazard intensity) without accounting for differential human exposure pathways [29,30]. This limitation is not merely technical but conceptual as by isolating biophysical sensitivity, we measure potential vulnerability rather than realized vulnerability, which emerges from the interaction of all three dimensions [31]. Turner et al. [19] demonstrated that communities with identical biophysical sensitivity exhibit vastly different outcomes based on adaptive capacity, while Eakin and Luers [32] showed that focusing solely on sensitivity risks misidentifying vulnerable populations where high adaptive capacity buffers climate impacts. This inherent conceptual restriction means our map highlights areas of potential biophysical stress but does not fully account for existing coping mechanisms or adaptive capacities at the local level. Consequently, a region appearing "vulnerable" on our map might in reality be highly resilient due to robust irrigation infrastructure, well-functioning credit markets, or diversified livelihood portfolios [33], while another appearing "resilient" might be highly susceptible to minor stress due to lack of adaptive resources, high debt burdens, or limited market access [34]. This distinction is critical for avoiding misinterpretations in policy and resource allocation strategies. For instance, our identification of Natore district as highly vulnerable (vulnerability score +1.21) reflects biophysical cropland stress but does not indicate whether local farmers possess drought-resistant seed varieties, access to supplemental irrigation, or off-farm income sources that might substantially mitigate actual livelihood impacts [35]. Conversely, districts with moderate biophysical vulnerability scores might harbor highly vulnerable sub-populations such as landless laborers, female-headed households, or marginal farmers whose socio-economic marginalization amplifies their susceptibility to even modest crop stress [36,37]. Therefore, our framework provides a necessary but partly insufficient vulnerability assessment a biophysical foundation that identifies where climate stress is occurring

https://doi.org/10.3390/land15010174

but requires integration with socio-economic data to determine who is most affected and why, as advocated by the vulnerability science community [38,39]. This conceptual positioning shapes how we must interpret the spatial patterns discussed in Section 4.4 and highlights the importance of the uncertainties examined in Section 4.3.

4.2.2. Methodological Comparison with Traditional Approaches

Traditional composite vulnerability indices, such as the Livelihood Vulnerability Index (LVI) employed by Hahn et al. [40] and Alam et al. [41] in Bangladesh, aggregated socio-economic indicators (income diversity, social networks, institutional access) with biophysical factors (water availability, land degradation) using equal or expert-assigned weights. While these frameworks captured multidimensional vulnerability, they suffered from three critical limitations our approach addressed, (1) reliance on infrequent household surveys that are temporally static, (2) coarse spatial resolution (typically district or sub-district level) that masked within-region heterogeneity, and (3) subjective weighting schemes that reduced reproducibility [39]. Our NDVI anomaly-based approach provided objective, annually updatable, moderate-resolution (250 m) vulnerability estimates, though admittedly at the cost of excluding socio-economic dimensions. This trade-off mirrored the evolution in drought monitoring, where standardized indices like the Standardized Precipitation Evapotranspiration Index (SPEI) [42] and satellite-based Vegetation Health Index (VHI) [43] increasingly complemented rather than replaced socio-economic assessments. Recent integrated frameworks, such as the spatial drought risk model by Meza et al. [44], demonstrated the potential of combining EO-derived hazard layers with socio-economic exposure data, an approach our biophysical vulnerability map could readily support. In particular, our findings corroborated regional studies identifying precipitation variability and vapor pressure deficit as primary cropland stressors in South Asian deltaic systems [44,45], while extending spatial coverage beyond previous sub-national assessments [46]. Having established the conceptual positioning and methodological advantages of our approach, we now turn to examining the specific uncertainties inherent in our data sources and modeling choices, and critically, how these uncertainties might quantitatively affect the spatial vulnerability patterns we observed.

4.3. Model Uncertainties and Spatial Scale Considerations

Several sources of uncertainty warranted acknowledgment in interpreting our results, with each potentially exerting quantifiable effects on the observed spatial vulnerability patterns. First, the ERA5 climate reanalysis data, while providing consistent spatial coverage, exhibited documented biases in representing localized precipitation extremes and monsoonal dynamics in complex topography [47]. These systematic biases could lead to a substantial underestimation of actual precipitation deficits in flood-prone northeastern districts during extreme monsoon events [47], potentially causing our model to misclassify genuinely stressed areas as moderately vulnerable or even resilient. Conversely, ERA5's smoothing of localized convective rainfall could overestimate drought severity in districts that experienced isolated but intense precipitation events not captured at the 0.25° reanalysis resolution [48]. Second, the bilinear resampling of coarse-resolution climate data (0.25° ERA5, approximately 27.5 km) to the 250 m MODIS grid, while necessary for model integration, did not create genuine fine-scale climate information and may have introduced spatial autocorrelation artifacts [49]. This resampling assigns identical climate values to all 250 m pixels within each 27.5 km ERA5 cell, potentially masking the fine-scale microclimatic gradients in temperature and precipitation that exist across topographic features, water bodies, or urban-rural transitions, a known limitation when using coarse-resolution climate data in high-resolution ecological applications [50,51]. Such homogenization could cause

https://doi.org/10.3390/land15010174

our model to miss localized vulnerability hotspots where terrain-induced climate variations create stress conditions not represented in the coarser input data. Third, the 250 m MODIS resolution, though adequate for national assessment, likely obscured field-level heterogeneity critical in smallholder systems where plot sizes averaged 0.5 hectares [52,53]. Each 250 m pixel (6.25 hectares) can encompass over a dozen typical smallholder plots in Bangladesh. Consequently, our vulnerability scores represent spatial averages that could mask severe stress affecting a substantial minority of farms within a pixel while others remain unaffected, a fundamental constraint of moderate-resolution data for heterogeneous systems [54]. This averaging effect may substantially underestimate peak vulnerability in heterogeneous landscapes where stress is spatially clustered at sub-pixel scales, a known consequence of scale mismatch in agricultural remote sensing [55]. The spatially blocked validation strategy, while mitigating spatial autocorrelation bias, may have inadvertently penalized the DL model if vulnerability patterns exhibited strong spatial structure that RF could exploit through localized feature interactions [56]. While this methodological choice could have inflated the apparent gap [57], preliminary analysis with random cross-validation suggested the core finding of RF superiority was robust. Finally, our 2001–2022 baseline period encompassed significant land use transitions, including expansion of dry-season irrigation and crop intensification [58], potentially shifting the reference conditions against which 2023 anomalies were calculated. This temporal non-stationarity could systematically bias vulnerability estimates, as the baseline 'normal' vegetation condition no longer reflects current management intensity. In districts experiencing rapid agricultural transformation, this could theoretically shift anomaly scores by a notable margin. Future iterations may incorporate higher-resolution climate products (e.g., CHIRPS at 5 km) and Sentinel-2 imagery (10 m) to address scale mismatches, alongside ground-truth validation datasets to quantify prediction uncertainties through ensemble approaches [59] and other advanced methodologies outlined for Earth science applications [60].

With these uncertainty sources and their potential magnitudes explicitly identified, we can now interpret the observed spatial vulnerability patterns with appropriate caution, recognizing where our estimates may be most reliable and where they require validation.

4.4. Spatial Patterns and Agro-Ecological Contextualization

The spatial distribution of cropland vulnerability revealed by our Random Forest model exhibited patterns consistent with known agro-ecological stressors in Bangladesh, though these patterns must be interpreted through the lens of the uncertainties and conceptual limitations discussed in Sections 4.2.1 and 4.3. Districts with elevated vulnerability scores during the 2023 dry season, such as Natore (+1.21) in the northwest and Magura (+0.89) in the southwest, correspond to regions historically characterized by intensified drought conditions and groundwater depletion. This spatial alignment is corroborated by previous regional assessments; importantly, Shahid and Behrawan [61] identified the northwestern Barind Tract including Natore as a critically drought-prone zone with declining groundwater tables that directly threaten irrigation-dependent Boro rice cultivation. However, the magnitude of vulnerability in these districts should be interpreted with awareness of the spatial resolution's impact discussed in Section 4.3. The 250 m resolution means our map may not capture intra-district variability, potentially homogenizing conditions where specific smallholder farms might be experiencing substantially higher stress than the pixel average, particularly in areas with heterogeneous soil drainage or uneven irrigation access. Furthermore, as highlighted in Section 4.2.1, these biophysical vulnerability scores do not account for adaptive capacity; Natore's actual livelihood vulnerability may be substantially lower if farmers have adopted stress-tolerant varieties or possess diversified income sources [32,33]. On the other hand, districts exhibiting pronounced

https://doi.org/10.3390/land15010174

negative vulnerability scores, such as Dinajpur (−2.27) and Faridpur (−2.16), demonstrated high biophysical resilience. This resilience was achieved despite seasonal forecasts for the 2023 monsoon indicating below-normal to near-normal rainfall and a heightened risk of drought conditions for these regions [62], suggesting the probable presence of effective buffering factors such as irrigation infrastructure or adaptive agricultural practices. Yet, as discussed in Section 4.3, the apparent resilience of these districts could be influenced by ERA5 biases, which might smooth over localized extreme events not fully captured by our reanalysis input data [63,64]. Additionally, the conceptual limitation emphasized in Section 4.2.1 applies here because even if biophysical conditions were favorable, vulnerable sub-populations within these 'resilient' districts such as landless laborers or marginal farmers might still face severe livelihood impacts due to socio-economic marginalization and gendered inequalities, which our vegetation-based metric cannot detect [36,37]. The strong predictive importance of March precipitation and February NDVI in our model reflected the critical role of pre-monsoon moisture availability for dry-season crop establishment, a relationship extensively documented for Boro rice and other Rabi crops in South Asian agricultural systems [45,65]. However, the temporal resampling of climate data to monthly aggregates (Section 4.3) may have masked sub-monthly variability critical for specific phenological stages [66], meaning our model might miss vulnerability driven by brief but intense dry spells occurring within otherwise "normal" monthly precipitation totals. Interestingly, coastal districts did not uniformly exhibit high vulnerability scores despite well-documented salinity intrusion concerns [67]. This pattern requires careful interpretation as NDVI anomalies primarily capture moisture-driven vegetation stress rather than salinity stress, which may manifest differently in spectral signatures [68]. Consequently, our map might underestimate vulnerability in coastal areas where salinity degrades soil fertility and crop quality without necessarily reducing canopy greenness to the same degree as drought stress [69]. This suggests that the 2023 growing season may have experienced sufficient freshwater availability, that farmers successfully adopted salt-tolerant varieties [70], or that our biophysical metric is insufficiently sensitive to salinity-specific stress a hypothesis requiring ground-truth validation through household surveys, soil salinity measurements, and yield data. The 250 m MODIS resolution (Section 4.3) further complicates coastal interpretation, as salinity intrusion often creates sharp spatial gradients at sub-pixel scales. Severely affected areas may represent only a fraction of a pixel and thus be masked in our averaged vulnerability scores, a known limitation of moderate-resolution data for detecting fine-scale environmental stressors [54,55]. These spatial patterns, interpreted through the explicit consideration of uncertainty sources and conceptual limitations, provide a foundation for designing operational early warning systems, though such systems must integrate our biophysical vulnerability layer with socio-economic data to achieve comprehensive risk assessment.

4.5. Implications for Early Warning Systems and Adaptive Management

The demonstrated capability of the Random Forest model to predict cropland vulnerability with 70% explained variance using readily available satellite and reanalysis data presented significant opportunities for operational early warning systems in Bangladesh, though the conceptual and methodological considerations discussed in Sections 4.2–4.4 shape how such systems should be designed and interpreted. Unlike traditional composite indices requiring extensive field surveys, our approach enabled near-real-time vulnerability assessment within weeks of data availability, a temporal advantage critical for timely intervention deployment [71]. The identification of recent climate variables (particularly March precipitation and February-March vapor pressure deficit) as primary drivers suggested that vulnerability forecasts could be generated as early as mid-growing season, providing

https://doi.org/10.3390/land15010174

a 2–3 month lead time for targeted support measures including supplemental irrigation allocation, crop insurance payouts, or input subsidy distribution [72,73]. This predictive window aligned with operational drought early warning systems implemented in East Africa, where satellite-based vegetation monitoring triggered anticipatory humanitarian response [74,75]. However, operationalizing our findings requires explicit acknowledgment of the uncertainties quantified in Section 4.3 and the conceptual limitations outlined in Section 4.2.1. Early warning systems based solely on our biophysical vulnerability layer risk misallocating resources if they do not account for adaptive capacity heterogeneity, directing irrigation support to districts with high biophysical stress scores but robust existing infrastructure would be less effective than targeting areas with moderate stress but limited coping resources [33,35]. To address this, operational systems should integrate our 250 m vulnerability maps with spatially explicit socio-economic data such as poverty rates, credit access, irrigation coverage, and market connectivity to generate composite risk indices that capture both biophysical exposure and social vulnerability [76,77]. Such integration would address the "necessary but insufficient" nature of our assessment emphasized in Section 4.2.1. Furthermore, the 250 m spatial resolution of our vulnerability maps exceeded the administrative granularity of most existing agricultural support programs in Bangladesh, which typically operated at upazila (sub-district) level, enabling more precise targeting of vulnerable farming communities. However, as discussed in Section 4.4, the 250 m resolution still obscures field-level heterogeneity critical for smallholder systems, potentially missing a significant portion of peak vulnerability occurring at sub-pixel scales, a known limitation when applying moderate-resolution data to heterogeneous smallholder landscapes [53]. Operational systems should therefore treat our maps as first-stage screening tools that identify priority upazilas or unions for ground-truthed assessments, rather than as definitive vulnerability assignments at the pixel level, following best practices for integrating remote sensing into decision-support framework [10,78]. Combining our moderate-resolution national coverage with targeted high-resolution (10 m Sentinel-2) assessments in flagged priority areas would balance scalability with precision [78]. The model's interpretability clearly attributing vulnerability to specific climatic and environmental factors facilitates communication with non-technical stakeholders including extension officers and farmer organizations, addressing a critical barrier to machine learning adoption in agricultural policy [79]. The explicit feature importance rankings (March precipitation, February NDVI, vapor pressure deficit) provide functional entry points for adaptation planning enhancing water storage capacity for pre-monsoon irrigation, promoting early-maturing varieties to avoid late-season moisture stress, or expanding weather-indexed insurance tied to these specific predictors [80,81]. However, communicating the uncertainty bounds discussed in Section 4.3 such as the potential for significant underestimation of precipitation deficits and systematic bias from temporal non-stationarity is equally critical to prevent over-confidence in specific vulnerability estimates. Probabilistic vulnerability forecasts incorporating these uncertainty ranges, rather than deterministic maps, would better support risk-informed decision-making [82]. Finally, the strong performance of the simpler Random Forest model over complex deep learning (Section 4.1) has important implications for system sustainability and scalability. RF models require minimal computational resources for retraining and inference, enabling deployment on standard computing infrastructure accessible to national agricultural agencies, whereas deep learning architectures would necessitate the specialized hardware and expertise outlined in comprehensive reviews of the field [60]. This computational accessibility, combined with the model's interpretability and the ready availability of MODIS and ERA5 data, positions our approach as a viable operational tool for resource-constrained contexts, provided that users understand its scope as a biophysi-

https://doi.org/10.3390/land15010174

cal sensitivity assessment requiring socio-economic complementation for comprehensive vulnerability mapping.

5. Conclusions

This study addressed a critical gap in agricultural vulnerability assessment by developing and comparing machine learning approaches for predicting cropland stress in Bangladesh, a nation facing acute climate risks to food security. In direct response to our primary objective, we successfully generated a national-scale, moderate-resolution (250 m) cropland vulnerability map for 2023 using 22-year satellite and climate time series data. Our comparative analysis definitively demonstrated that the Random Forest model substantially outperformed the dual-stream deep learning architecture, achieving an R^2 of 0.70 versus 0.02 on spatially independent validation data. This finding challenged prevailing assumptions favoring deep learning complexity and revealed the continued relevance of interpretable ensemble methods for heterogeneous environmental prediction tasks, particularly when training data exhibited spatial structure and sample sizes were constrained relative to feature dimensionality. Addressing our second objective, feature importance analysis of the superior Random Forest model identified recent climatic variables, specifically March precipitation, February NDVI, and March vapor pressure deficit as the primary biophysical drivers of cropland vulnerability. This finding validated our NDVI anomaly-based operationalization of vulnerability as a defensible Earth Observation metric sensitive to short-term climate perturbations, while also revealing that static environmental factors (soil properties, topography) played secondary but measurable roles. The spatial patterns of predicted vulnerability aligned with documented agro-ecological stressors, with northwestern districts (Natore, Magura) exhibiting elevated stress consistent with 2023 drought conditions, while northern and central districts (Dinajpur, Faridpur) demonstrated resilience linked to favorable precipitation patterns. The practical implications of these results extend across multiple domains of agricultural policy and climate adaptation planning in Bangladesh. First, the demonstrated accuracy and efficiency of the Random Forest approach established a viable framework for operational early warning systems, enabling near-real-time vulnerability assessment within weeks of satellite data availability, a temporal advantage over traditional survey-based indices requiring months to compile. The 2–3 month lead time provided by identifying March precipitation and February-March vegetation conditions as key predictors created feasible windows for deploying individualized approaches, including supplemental irrigation allocation in high-vulnerability districts, expedited crop insurance claim processing, and strategic distribution of drought-resistant seed varieties for subsequent planting seasons. Second, the 250 m spatial resolution of our vulnerability maps exceeded the administrative granularity of existing agricultural support programs, enabling precision targeting of vulnerable farming communities within districts rather than uniform interventions across heterogeneous landscapes. This spatial precision could substantially improve the cost-effectiveness of adaptation investments by concentrating resources in genuinely stressed areas while avoiding unnecessary expenditures in resilient zones. Third, the model's interpretability, transparently linking vulnerability to specific, measurable climatic and environmental factors facilitated communication with extension services, farmer organizations, and policymakers who require understandable justifications for resource allocation decisions rather than opaque algorithmic predictions. From a methodological perspective, this study contributed a reproducible framework integrating freely available Earth Observation data (MODIS, ERA5, CHIRPS) with open-source machine learning tools, enabling replication across other climate-vulnerable agricultural regions in South Asia and beyond. The rigorous spatially aware validation strategy employed here should serve as a standard for future agricultural prediction studies to avoid

https://doi.org/10.3390/land15010174

inflated performance claims arising from spatial autocorrelation. However, several important limitations warrant emphasis. Our NDVI anomaly approach captured biophysical vulnerability dimensions while excluding socio-economic factors (adaptive capacity, market access, institutional support) that critically modulate actual food security outcomes, the 250 m resolution, though improved over district-level indices, remained coarse relative to smallholder plot sizes and the reliance on reanalysis climate data introduced documented biases in representing localized precipitation extremes. Future research should prioritize integration of household-level socio-economic surveys with higher-resolution satellite imagery (Sentinel-2 at 10 m) and ground-truth yield measurements to develop truly comprehensive vulnerability assessments bridging biophysical and human dimensions. In conclusion, this study demonstrated that simpler, interpretable machine learning models can outperform complex deep learning architectures for agricultural vulnerability assessment when appropriately designed and validated, while simultaneously providing the transparency necessary for operational deployment in policy contexts. The resulting vulnerability maps and identified climatic drivers offer immediately workable information for Bangladesh's agricultural adaptation planning, with broader methodological lessons for climate-vulnerable food systems globally.

6. Limitations

While NDVI anomalies provide a spatially explicit and objective proxy for biophysical cropland vulnerability, this approach inherently focuses on vegetation health as an indicator and does not directly capture the socio-economic dimensions of vulnerability, including farmer adaptive capacity, market access, credit availability, or institutional support system [19,33]. NDVI-based assessments represent a proxy for crop stress rather than direct measurements of yield loss or food insecurity, as the relationship between vegetation indices and actual harvest outcomes can be modulated by crop type, management practices, and post-harvest factors [54,78]. Additionally, NDVI saturation in densely vegetated areas and sensitivity to atmospheric conditions, soil background, and sensor characteristics may introduce uncertainties in anomaly detection, particularly in heterogeneous agricultural landscapes [4]. The 250 m spatial resolution of MODIS data, while suitable for national-scale assessments, may not capture field-level variability critical for smallholder farming systems that dominate Bangladesh's agricultural landscape [53]. Future research should integrate household-level socio-economic data, higher-resolution imagery, and ground-truth yield measurements to develop comprehensive vulnerability frameworks that bridge biophysical and human dimensions of climate risk [23].

A further limitation concerns the absence of formal uncertainty quantification and sensitivity analysis in our modeling framework, which restricts assessment of result robustness. Our reliance on ERA5 climate reanalysis products introduces inherent uncertainties from model assimilation processes, spatial interpolation, and parameterization schemes that may not fully capture localized extreme events or microclimatic variations [48,83]. The temporal resampling strategy employed to align multi-source datasets, while methodologically necessary, potentially propagates uncertainties through aggregation and may mask sub-monthly climate variability critical for crop phenological stages [66,84]. Furthermore, the Random Forest model's sensitivity to hyperparameter configurations and the spatial block cross-validation strategy's dependence on block size selection warrant systematic sensitivity analyses to establish confidence intervals around predictions [56,85]. Future implementations may incorporate ensemble modeling and data assimilation approaches [86], along with systematic sensitivity testing of input data quality, temporal aggregation methods, and model parameterization [87], to provide probabilistic vulnerability estimates with quantified confidence bounds.

https://doi.org/10.3390/land15010174

Author Contributions: Conceptualization, A.B. and M.N.; methodology, A.B. and M.N.; software, A.B.; validation, A.B.; formal analysis, A.B.; investigation, A.B.; resources, A.B. and M.N.; data curation, A.B.; writing—original draft preparation, A.B.; writing—review and editing, A.B. and M.N.; visualization, A.B.; supervision, M.N.; project administration, M.N.; funding acquisition, M.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Satellite-derived datasets used in this study are large in size and therefore not publicly hosted. The data are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the author used ChatGPT (GPT-5, OpenAI, 2025) for language editing and refinement of text clarity. The author reviewed and edited all AI-generated content and takes full responsibility for the final version of the manuscript.

Conflicts of Interest: Author Arnob Bormudoi was employed by New Space Intelligence. The remaining author 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

1. Field, C.B.; Barros, V.R. *Climate Change 2014–Impacts, Adaptation and Vulnerability: Global and Sectoral Aspects*; Cambridge University Press: Cambridge, UK, 2014.
2. Frischen, J.; Meza, I.; Rupp, D.; Wietler, K.; Hagenlocher, M. Drought Risk to Agricultural Systems in Zimbabwe: A Spatial Analysis of Hazard, Exposure, and Vulnerability. *Sustainability* **2020**, *12*, 752. [CrossRef]
3. West, H.; Quinn, N.; Horswell, M. Remote Sensing for Drought Monitoring & Impact Assessment: Progress, Past Challenges and Future Opportunities. *Remote Sens. Environ.* **2019**, *232*, 111291.
4. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A Commentary Review on the Use of Normalized Difference Vegetation Index (NDVI) in the Era of Popular Remote Sensing. *J. For. Res.* **2021**, *32*, 1–6. [CrossRef]
5. Aryal, J.P.; Sapkota, T.B.; Rahut, D.B.; Krupnik, T.J.; Shahrin, S.; Jat, M.L.; Stirling, C.M. Major Climate Risks and Adaptation Strategies of Smallholder Farmers in Coastal Bangladesh. *Environ. Manag.* **2020**, *66*, 105–120. [CrossRef]
6. Lázár, A.N.; Clarke, D.; Adams, H.; Akanda, A.R.; Szabo, S.; Nicholls, R.J.; Matthews, Z.; Begum, D.; Saleh, A.F.M.; Abedin, M.A. Agricultural Livelihoods in Coastal Bangladesh Under Climate and Environmental Change–A Model Framework. *Environ. Sci. Process. Impacts* **2015**, *17*, 1018–1031. [CrossRef] [PubMed]
7. Kabir, H.; Golder, J. Rainfall Variability and Its Impact on Crop Agriculture in Southwest Region of Bangladesh. *J. Climatol. Weather Forecast.* **2017**, *5*, 1–20.
8. SAC; FAO; AFA; ICA-AP. *United Nations Decade of Family Farming 2019–2028, Regional Action Plan to Implement the UNDFF and Achieve the SDGs in South Asia*; SAC: Dhaka, Bangladesh; FAO: Rome, Italy; AFA: Quezon City, Philippines; ICA-AP: New Delhi, India, 2021; ISBN 978-92-5-134526-9.
9. Rojas, O.; Vrieling, A.; Rembold, F. Assessing Drought Probability for Agricultural Areas in Africa with Coarse Resolution Remote Sensing Imagery. *Remote Sens. Environ.* **2011**, *115*, 343–352. [CrossRef]
10. Enenkel, M.; See, L.; Bonifacio, R.; Boken, V.; Chaney, N.; Vinck, P.; You, L.; Dutra, E.; Anderson, M. Drought and Food Security–Improving Decision-Support via New Technologies and Innovative Collaboration. *Glob. Food Secur.* **2015**, *4*, 51–55. [CrossRef]
11. Kong, Z.; Ren, J.; Wang, L.; Guo, G. Multi-Scale Fusion Graph Convolutional Networks. *Expert Syst. Appl.* **2025**, *298*, 129815. [CrossRef]
12. Hengl, T.; Nussbaum, M.; Wright, M.N.; Heuvelink, G.B.; Gräler, B. Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables. *PeerJ* **2018**, *6*, e5518. [CrossRef]
13. Füssel, H.-M.; Klein, R.J.T. Climate Change Vulnerability Assessments: An Evolution of Conceptual Thinking. *Clim. Change* **2006**, *75*, 301–329. [CrossRef]
14. Anyamba, A.; Tucker, C.J. Historical Perspective of AVHRR NDVI and Vegetation Drought Monitoring. *Remote Sens. Drought Innov. Monit. Approaches* **2012**, *23*, 20.
15. Lloyd, D. A Phenological Classification of Terrestrial Vegetation Cover Using Shortwave Vegetation Index Imagery. *Int. J. Remote Sens.* **1990**, *11*, 2269–2279. [CrossRef]
16. Funk, C.; Budde, M.E. Phenologically-Tuned MODIS NDVI-Based Production Anomaly Estimates for Zimbabwe. *Remote Sens. Environ.* **2009**, *113*, 115–125. [CrossRef]
17. Tucker, C.J.; Choudhury, B.J. Satellite Remote Sensing of Drought Conditions. *Remote Sens. Environ.* **1987**, *23*, 243–251. [CrossRef]

18. Kuenzer, C.; Ottinger, M.; Wegmann, M.; Guo, H.; Wang, C.; Zhang, J.; Dech, S.; Wikelski, M. Earth Observation Satellite Sensors for Biodiversity Monitoring: Potentials and Bottlenecks. *Int. J. Remote Sens.* **2014**, *35*, 6599–6647. [CrossRef]
19. Turner, B.L.; Kasperson, R.E.; Matson, P.A.; McCarthy, J.J.; Corell, R.W.; Christensen, L.; Eckley, N.; Kasperson, J.X.; Luers, A.; Martello, M.L.; et al. A Framework for Vulnerability Analysis in Sustainability Science. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8074–8079. [CrossRef]
20. Folke, C. Resilience: The Emergence of a Perspective for Social–Ecological Systems Analyses. *Glob. Environ. Change* **2006**, *16*, 253–267. [CrossRef]
21. Holling, C.S. *Resilience and Stability of Ecological Systems*; Annual Reviews: San Mateo, CA, USA, 1973.
22. Prasad, A.K.; Chai, L.; Singh, R.P.; Kafatos, M. Crop Yield Estimation Model for Iowa Using Remote Sensing and Surface Parameters. *Int. J. Appl. Earth Obs. Geoinf.* **2006**, *8*, 26–33. [CrossRef]
23. Simelton, E.; Fraser, E.D.; Termansen, M.; Forster, P.M.; Dougill, A.J. Typologies of Crop-Drought Vulnerability: An Empirical Analysis of the Socio-Economic Factors That Influence the Sensitivity and Resilience to Drought of Three Major Food Crops in China (1961–2001). *Environ. Sci. Policy* **2009**, *12*, 438–452. [CrossRef]
24. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
25. Hastie, T.; Tibshirani, R.; Friedman, J.; Franklin, J. The Elements of Statistical Learning: Data Mining, Inference and Prediction. *Math. Intell.* **2005**, *27*, 83–85. [CrossRef]
26. Kuwata, K.; Shibasaki, R. Estimating Crop Yields with Deep Learning and Remotely Sensed Data. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; IEEE: New York, NY, USA, 2015; pp. 858–861.
27. Khaki, S.; Wang, L.; Archontoulis, S.V. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [CrossRef]
28. Atzberger, C. Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sens.* **2013**, *5*, 949–981. [CrossRef]
29. O'Brien, K.; Eriksen, S.; Nygaard, L.P.; Schjolden, A. Why Different Interpretations of Vulnerability Matter in Climate Change Discourses. *Clim. Policy* **2007**, *7*, 73–88. [CrossRef]
30. Adger, W.N. Vulnerability. *Glob. Environ. Change* **2006**, *16*, 268–281. [CrossRef]
31. Hufschmidt, G. A Comparative Analysis of Several Vulnerability Concepts. *Nat. Hazards* **2011**, *58*, 621–643. [CrossRef]
32. Eakin, H.; Luers, A.L. Assessing the Vulnerability of Social-Environmental Systems. *Annu. Rev. Environ. Resour.* **2006**, *31*, 365–394. [CrossRef]
33. Agrawal, A.; Perrin, N. Climate Adaptation, Local Institutions and Rural Livelihoods. In *Adapting to Climate Change: Thresholds, Values, Governance*; Cambridge University Press: Cambridge, UK, 2009; pp. 350–367.
34. Eriksen, S.H.; Brown, K.; Kelly, P.M. The Dynamics of Vulnerability: Locating Coping Strategies in Kenya and Tanzania. *Geogr. J.* **2005**, *171*, 287–305. [CrossRef]
35. Pelling, M.; High, C. Understanding Adaptation: What Can Social Capital Offer Assessments of Adaptive Capacity? *Glob. Environ. Change* **2005**, *15*, 308–319. [CrossRef]
36. Neumayer, E.; Plümper, T. The Gendered Nature of Natural Disasters: The Impact of Catastrophic Events on the Gender Gap in Life Expectancy, 1981–2002. *Ann. Assoc. Am. Geogr.* **2007**, *97*, 551–566. [CrossRef]
37. Demetriades, J.; Esplen, E. The Gender Dimensions of Poverty and Climate Change Adaptation. *Soc. Dimens. Clim. Change Equity Vulnerability Warm. World* **2010**, *4*, 133–143. [CrossRef]
38. Birkmann, J.; Cardona, O.D.; Carreño, M.L.; Barbat, A.H.; Pelling, M.; Schneiderbauer, S.; Kienberger, S.; Keiler, M.; Alexander, D.; Zeil, P.; et al. Framing Vulnerability, Risk and Societal Responses: The MOVE Framework. *Nat. Hazards* **2013**, *67*, 193–211. [CrossRef]
39. Hinkel, J. "Indicators of Vulnerability and Adaptive Capacity": Towards a Clarification of the Science–Policy Interface. *Glob. Environ. Change* **2011**, *21*, 198–208. [CrossRef]
40. Hahn, M.B.; Riederer, A.M.; Foster, S.O. The Livelihood Vulnerability Index: A Pragmatic Approach to Assessing Risks from Climate Variability and Change—A Case Study in Mozambique. *Glob. Environ. Change* **2009**, *19*, 74–88. [CrossRef]
41. Alam, G.M.; Alam, K.; Mushtaq, S. Climate Change Perceptions and Local Adaptation Strategies of Hazard-Prone Rural Households in Bangladesh. *Clim. Risk Manag.* **2017**, *17*, 52–63. [CrossRef]
42. Beguería, S.; Vicente-Serrano, S.M.; Reig, F.; Latorre, B. Standardized Precipitation Evapotranspiration Index (SPEI) Revisited: Parameter Fitting, Evapotranspiration Models, Tools, Datasets and Drought Monitoring. *Int. J. Climatol.* **2014**, *34*, 3001–3023. [CrossRef]
43. Kogan, F.N. Application of Vegetation Index and Brightness Temperature for Drought Detection. *Adv. Space Res.* **1995**, *15*, 91–100. [CrossRef]

44. Meza, I.; Siebert, S.; Döll, P.; Kusche, J.; Herbert, C.; Eyshi Rezaei, E.; Nouri, H.; Gerdener, H.; Popat, E.; Frischen, J. Global-Scale Drought Risk Assessment for Agricultural Systems. *Nat. Hazards Earth Syst. Sci.* **2020**, *20*, 695–712. [CrossRef]
45. Mondal, M.S.; Rahman, M.A.; Mukherjee, N.; Huq, H.; Rahman, R. Hydro-Climatic Hazards for Crops and Cropping System in the Chars of the Jamuna River and Potential Adaptation Options. *Nat. Hazards* **2015**, *76*, 1431–1455. [CrossRef]
46. Alamgir, M.; Shahid, S.; Hazarika, M.K.; Nashrrullah, S.; Harun, S.B.; Shamsudin, S. Analysis of Meteorological Drought Pattern During Different Climatic and Cropping Seasons in Bangladesh. *JAWRA J. Am. Water Resour. Assoc.* **2015**, *51*, 794–806. [CrossRef]
47. Prakash, S.; Mitra, A.K.; Momin, I.M.; Rajagopal, E.N.; Basu, S.; Collins, M.; Turner, A.G.; Achuto Rao, K.; Ashok, K. Seasonal Intercomparison of Observational Rainfall Datasets over India During the Southwest Monsoon Season. *Int. J. Climatol.* **2015**, *35*, 2326–2338. [CrossRef]
48. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]
49. Goodchild, M.F. Geographical Data Modeling. *Comput. Geosci.* **1992**, *18*, 401–408. [CrossRef]
50. Ashouri, H.; Hsu, K.-L.; Sorooshian, S.; Braithwaite, D.K.; Knapp, K.R.; Cecil, L.D.; Nelson, B.R.; Prat, O.P. PERSIANN-CDR: Daily Precipitation Climate Data Record from Multisatellite Observations for Hydrological and Climate Studies. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 69–83. [CrossRef]
51. Daly, C.; Halbleib, M.; Smith, J.I.; Gibson, W.P.; Doggett, M.K.; Taylor, G.H.; Curtis, J.; Pasteris, P.P. Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States. *Int. J. Climatol.* **2008**, *28*, 2031–2064. [CrossRef]
52. Tittonell, P.; Muriuki, A.; Shepherd, K.D.; Mugendi, D.; Kaizzi, K.C.; Okeyo, J.; Verchot, L.; Coe, R.; Vanlauwe, B. The Diversity of Rural Livelihoods and Their Influence on Soil Fertility in Agricultural Systems of East Africa–A Typology of Smallholder Farms. *Agric. Syst.* **2010**, *103*, 83–97. [CrossRef]
53. Jain, M.; Mondal, P.; DeFries, R.S.; Small, C.; Galford, G.L. Mapping Cropping Intensity of Smallholder Farms: A Comparison of Methods Using Multiple Sensors. *Remote Sens. Environ.* **2013**, *134*, 210–223. [CrossRef]
54. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A Scalable Satellite-Based Crop Yield Mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [CrossRef]
55. Jin, Z.; Azzari, G.; Lobell, D.B. Improving the Accuracy of Satellite-Based High-Resolution Yield Estimation: A Test of Multiple Scalable Approaches. *Agric. For. Meteorol.* **2017**, *247*, 207–220. [CrossRef]
56. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
57. Wadoux, A.M.-C.; Heuvelink, G.B.; De Bruin, S.; Brus, D.J. Spatial Cross-Validation Is Not the Right Way to Evaluate Map Accuracy. *Ecol. Model.* **2021**, *457*, 109692. [CrossRef]
58. Rahman, S. Six Decades of Agricultural Land Use Change in Bangladesh: Effects on Crop Diversity, Productivity, Food Availability and the Environment, 1948-2006: Land Use Change in Bangladesh, 1948–2006. *Singap. J. Trop. Geogr.* **2010**, *31*, 254–269. [CrossRef]
59. Dietterich, T.G. Ensemble Methods in Machine Learning. In *Multiple Classifier Systems*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1857, pp. 1–15, ISBN 978-3-540-67704-8.
60. Camps-Valls, G.; Tuia, D.; Zhu, X.X.; Reichstein, M. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*; John Wiley & Sons: Hoboken, NJ, USA, 2021.
61. Shahid, S.; Behrawan, H. Drought Risk Assessment in the Western Part of Bangladesh. *Nat. Hazards* **2008**, *46*, 391–413. [CrossRef]
62. Bangladesh Meteorological Department Seasonal Outlook 2023: Guidance for Decision Makers, July and JAS 2023. 2023. Available online: https://www.bmd.gov.bd/file/2024/04/22/pdf/165422.pdf (accessed on 13 January 2026).
63. Gleixner, S.; Demissie, T.; Diro, G.T. Did ERA5 Improve Temperature and Precipitation Reanalysis over East Africa? *Atmosphere* **2020**, *11*, 996. [CrossRef]
64. Nogueira, M. Inter-Comparison of ERA-5, ERA-Interim and GPCP Rainfall over the Last 40 Years: Process-Based Analysis of Systematic and Random Differences. *J. Hydrol.* **2020**, *583*, 124632. [CrossRef]
65. Islam, M.A.; Warwick, N.; Koech, R.; Amin, M.N.; de Bruyn, L.L. The Importance of Farmers' Perceptions of Salinity and Adaptation Strategies for Ensuring Food Security: Evidence from the Coastal Rice Growing Areas of Bangladesh. *Sci. Total Environ.* **2020**, *727*, 138674. [CrossRef] [PubMed]
66. White, J.W.; Hoogenboom, G.; Kimball, B.A.; Wall, G.W. Methodologies for Simulating Impacts of Climate Change on Crop Production. *Field Crops Res.* **2011**, *124*, 357–368. [CrossRef]
67. Dasgupta, S.; Hossain, M.M.; Huq, M.; Wheeler, D. Climate Change and Soil Salinity: The Case of Coastal Bangladesh. *Ambio* **2015**, *44*, 815–826. [CrossRef]
68. Ivushkin, K.; Bartholomeus, H.; Bregt, A.K.; Pulatov, A.; Kempen, B.; De Sousa, L. Global Mapping of Soil Salinity Change. *Remote Sens. Environ.* **2019**, *231*, 111260. [CrossRef]

69. Scudiero, E.; Skaggs, T.H.; Corwin, D.L. Regional-Scale Soil Salinity Assessment Using Landsat ETM+ Canopy Reflectance. *Remote Sens. Environ.* **2015**, *169*, 335–343. [CrossRef]
70. Tuong, T.P.; Bouman, B.A.M. Rice Production in Water-Scarce Environments. In *Water Productivity in Agriculture: Limits and Opportunities for Improvement*; Kijne, J.W., Barker, R., Molden, D., Eds.; CABI Publishing: Wallingford, UK, 2003; pp. 53–67, ISBN 978-0-85199-669-1.
71. Verdin, J.; Funk, C.; Senay, G.; Choularton, R. Climate Science and Famine Early Warning. *Philos. Trans. R. Soc. B Biol. Sci.* **2005**, *360*, 2155–2168. [CrossRef]
72. Brown, M.E. Remote Sensing Technology and Land Use Analysis in Food Security Assessment. *J. Land Use Sci.* **2016**, *11*, 623–641. [CrossRef]
73. Binswanger-Mkhize, H.P. Is There Too Much Hype About Index-Based Agricultural Insurance? *J. Dev. Stud.* **2012**, *48*, 187–200. [CrossRef]
74. Funk, C.; Shukla, S.; Thiaw, W.M.; Rowland, J.; Hoell, A.; McNally, A.; Husak, G.; Novella, N.; Budde, M.; Peters-Lidard, C. Recognizing the Famine Early Warning Systems Network: Over 30 Years of Drought Early Warning Science Advances and Partnerships Promoting Global Food Security. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 1011–1027. [CrossRef]
75. Coughlan de Perez, E.; van den Hurk, B.; Van Aalst, M.K.; Jongman, B.; Klose, T.; Suarez, P. Forecast-Based Financing: An Approach for Catalyzing Humanitarian Action Based on Extreme Weather and Climate Forecasts. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 895–904. [CrossRef]
76. Tate, E. Social Vulnerability Indices: A Comparative Assessment Using Uncertainty and Sensitivity Analysis. *Nat. Hazards* **2012**, *63*, 325–347. [CrossRef]
77. Cutter, S.L.; Boruff, B.J.; Shirley, W.L. Social Vulnerability to Environmental Hazards. In *Hazards Vulnerability and Environmental Justice*; Routledge: Abingdon-on-Thames, UK, 2012; pp. 115–132.
78. Becker-Reshef, I.; Justice, C.; Barker, B.; Humber, M.; Rembold, F.; Bonifacio, R.; Zappacosta, M.; Budde, M.; Magadzire, T.; Shitote, C. Strengthening Agricultural Decisions in Countries at Risk of Food Insecurity: The GEOGLAM Crop Monitor for Early Warning. *Remote Sens. Environ.* **2020**, *237*, 111553. [CrossRef]
79. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef]
80. Jensen, N.D.; Barrett, C.B.; Mude, A.G. Index Insurance Quality and Basis Risk: Evidence from Northern Kenya. *Am. J. Agric. Econ.* **2016**, *98*, 1450–1469. [CrossRef]
81. Barnett, B.J.; Barrett, C.B.; Skees, J.R. Poverty Traps and Index-Based Risk Transfer Products. *World Dev.* **2008**, *36*, 1766–1785. [CrossRef]
82. Rosenzweig, C.; Elliott, J.; Deryng, D.; Ruane, A.C.; Müller, C.; Arneth, A.; Boote, K.J.; Folberth, C.; Glotter, M.; Khabarov, N.; et al. Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3268–3273. [CrossRef]
83. Betts, A.K.; Chan, D.Z.; Desjardins, R.L. Near-Surface Biases in ERA5 over the Canadian Prairies. *Front. Environ. Sci.* **2019**, *7*, 129. [CrossRef]
84. Bolton, D.K.; Friedl, M.A. Forecasting Crop Yield Using Remotely Sensed Vegetation Indices and Crop Phenology Metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84. [CrossRef]
85. Probst, P.; Wright, M.N.; Boulesteix, A. Hyperparameters and Tuning Strategies for Random Forest. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
86. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.E.; Liang, S.; Chen, Z.; Xue, J.-H.; Wu, Y. Assimilation of Remote Sensing into Crop Growth Models: Current Status and Perspectives. *Agric. For. Meteorol.* **2019**, *276*, 107609. [CrossRef]
87. Wallach, D.; Palosuo, T.; Thorburn, P.; Gourdain, E.; Asseng, S.; Basso, B.; Buis, S.; Crout, N.; Dibari, C.; Dumont, B. How Well Do Crop Modeling Groups Predict Wheat Phenology, given Calibration Data from the Target Population? *Eur. J. Agron.* **2021**, *124*, 126195. [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.

land

MDPI

Article

Landslide Susceptibility Assessment Based on a TSPF-BiLSTM Model: A Case Study of Sangzhi County, Hunan Province

Kangcheng Zhu [1,2], Yuzhong Kong [3], Xiangyun Kong [3], Sen Hu [1,2], Junmeng Zhao [2,4,*], Ciren Pu [4], Junzhe Teng [1,2], Weiyan Luo [1,2], Yang Pu [1,2], Taijin Su [2,3], Xingwang Chen [3] and Zhen Jiang [1]

[1] School of Ecology and Environment, Xizang University, Lhasa 850000, China; zkc505@163.com (K.Z.)
[2] Geodynamics Laboratory, Xizang University, Lhasa 850000, China
[3] School of Engineering, Xizang University, Lhasa 850000, China
[4] School of Science, Xizang University, Lhasa 850000, China
* Correspondence: zhaojm@itpcas.ac.cn

Abstract

In karst mountainous areas where high-dimensional features coexist with extremely limited sample sizes, accurate landslide susceptibility mapping remains challenging. To address this issue, we propose an ensemble framework termed the Triple-Source Probabilistic Fusion Bidirectional Long Short-Term Memory network (TSPF-BiLSTM). The approach was tested in Sangzhi County, Hunan Province, by integrating three base learners—Random Forest (RF), LightGBM, and AdaBoost. Their raw outputs were first calibrated using five-fold Platt scaling to generate posterior probabilities on a unified scale. A bidirectional LSTM was then employed to perform deep nonlinear fusion of these cross-model probability features. Using a total of 618 landslide and 618 non-landslide samples (split into training and testing sets), the TSPF-BiLSTM model achieved a mean AUC of 0.9525 ($\pm$0.0115) under ten-fold cross-validation, outperforming not only the individual base learners but also standalone deep learning models (CNN and Transformer). The frequency ratio in the very high susceptibility zone reached 3.97, significantly exceeding all benchmark models and confirming its superior capability in high-risk area identification. Multi-model importance analysis identified NDVI, elevation, and annual rainfall as the dominant regional landslide predisposing factors. Within the specific ranges of NDVI 0–0.686, elevation 155–462 m, and annual rainfall 1273.6–1301 mm, landslide frequency ratios consistently exceeded 1.96. The proposed framework, with its probability-level fusion and embedded regularization mechanisms, effectively mitigated overfitting despite the small sample size, providing a robust technical solution for geological hazard risk identification and prevention in the data-scarce karst terrain of the Wuling Mountains.

Keywords: Sangzhi County; landslide susceptibility; probabilistic fusion; bidirectional LSTM; karst mountains

1. Introduction

Landslides are among the most widespread and destructive natural hazards in mountainous regions worldwide, posing persistent and severe threats to human life, infrastructure, and sustainable development [1]. To facilitate proactive risk management and mitigation, Landslide Susceptibility Assessment (LSA) has emerged as a fundamental tool that predicts the spatial probability of landslide occurrence by analyzing relationships between historical landslides and various conditioning factors such as geology, topography, and land use [2]. The evolution of LSA methodologies illustrates a clear trajectory

from conventional statistical analysis toward advanced computational intelligence and, increasingly, multidisciplinary integration.

Early LSA research extensively employed statistically based models. Reichenbach et al. provided a systematic review of techniques, including logistic regression and frequency ratio, establishing a solid foundation for the field [1]. However, the performance of statistical models is highly sensitive to analysis scale and data quality. For instance, Polykretis et al. demonstrated that a more detailed analysis scale (e.g., 20 m grids) yields significantly higher accuracy and predictive capability for logistic regression models compared to regional scales [3]. Yang et al. further quantified the impacts of landslide spatial representation (point vs. polygon) and raster resolution, finding that finer resolutions and polygon representation generally lead to improved model performance [4]. These studies collectively underscore the critical challenges of scale effects and data uncertainty in model construction.

To overcome the limitations of traditional statistical methods, Machine Learning (ML) and Deep Learning (DL) algorithms have been widely adopted. Initial research focused on comparing model performance in specific case studies; for example, Lee et al. showed that an Artificial Neural Network outperformed logistic regression in Inje, Korea [5]. Subsequent research has advanced toward more complex architectures and integration with physical principles. Liu et al. proposed an innovative physics-informed data-driven framework for the Three Gorges Reservoir Area, using the Scoops3D model to guide non-landslide sample selection before applying Random Forest, which improved the AUC by 20.1% [6]. Wang et al. introduced a landslide density-based Bidirectional Long Short-Term Memory model with multichannel input and an optimized sampling strategy for seismically active mountainous regions, emphasizing the importance of extracting high-dimensional geospatial features and ensuring model generalization [7]. Xia et al. developed a KNN-GCN model for slope-unit-based mapping that incorporates spatial correlations between units [8].

Despite the proliferation of models, significant challenges remain unresolved. First, many advanced ML/DL models operate as "black boxes" lacking interpretability, which limits their credibility for critical engineering decisions. In contrast, physics-informed methods such as limit equilibrium are theoretically more acceptable [9]. Beyond the landslide domain, interpretable modeling frameworks have also been developed in related geohazard fields. For instance, in deformation monitoring, Mohamadi et al. proposed an interpretable causality model for complex subsidence based on PS-InSAR cross-heading orbit analysis, which successfully distinguishes contributions of different factors to surface deformation [10]. Such cross-domain methodological insights offer valuable references for enhancing the transparency of LSA models. Second, models often exhibit strong regional dependency; their performance depends heavily on the specific study area, data quality, and feature selection methods, hindering application across diverse geological settings [11]. Third, most LSA models are static, unable to effectively incorporate dynamic triggering factors such as rainfall or seismic activity [12]. Recent advances in spatiotemporal data mining provide new opportunities for dynamic modeling. Mohamadi et al. integrated PS-InSAR time series with space-time cubes and emerging hot spot analysis to identify anomalous spatiotemporal patterns (new cold spots and new hot spots) preceding building collapses, achieving a 97% detection rate [13]. This methodology can be adapted to LSA for capturing precursor deformation patterns related to landslide initiation. Fourth, a particularly pressing issue is the "few-shot" predicament: acquiring large-scale, high-quality landslide inventories is costly and challenging, especially in remote and hazardous mountainous terrain, often resulting in a scarcity of labeled samples [14]. Conventional data-driven methods are prone to overfitting under such conditions, compromising model

https://doi.org/10.3390/land15040579

robustness and generalization. Recent research has explored the meta-learning paradigm to address this challenge. The core idea is to enable models to learn how to learn, allowing rapid adaptation to new tasks with only a handful of samples. By learning generalizable knowledge from related tasks, this approach offers a promising solution to the few-shot learning problem in LSA [14,15].

In exploring the integration of dynamic factors, recent studies have analyzed the role of soil hydrodynamic parameters, which are governed by profound physical mechanisms. When soil water content exceeds field capacity and approaches saturation, water replaces air in soil pores, generating positive pore water pressure. This effectively "lifts" soil particles, reducing frictional resistance on slopes. Furthermore, vegetation not only provides mechanical reinforcement through root systems but also actively extracts soil moisture via transpiration. The hydrological benefit of vegetation is mainly attributed to the suction generated by transpiration, which lowers soil water content and pore water pressure, thereby increasing soil shear strength. Research indicates that transpiration-induced suction maintained by vegetation can significantly improve slope stability under extreme rainfall; this hydrological effect is crucial under different rainfall patterns, and vegetation removal would make soil more susceptible to reaching a saturated state [16]. Integrating these physical mechanisms with data-driven models is essential for enhancing model interpretability.

Although hybrid models combining Convolutional Neural Networks and Bidirectional Long Short-Term Memory networks (e.g., CNN-BiLSTM-AM) are recognized as high-performing deep learning architectures due to their ability to extract spatial features and mine temporal dependencies [17,18], no single "best" method exists for LSA. Model performance depends critically on the specific study area's geo-environmental context, data quality, feature selection methods, and strategies for handling few-shot conditions [17]. Therefore, further elucidating model generalizability, parameter sensitivity, and adaptability under few-shot scenarios is essential for enhancing the scientific contribution of any study.

These universal challenges are acutely relevant for Sangzhi County, Zhangjiajie—a region characterized by complex geology, concentrated rainfall, and significant human activity, leading to high landslide hazard risk. However, its landslide inventory is relatively limited, making it a typical few-shot study area. Tailored susceptibility research addressing its specific geo-environmental context while balancing accuracy, interpretability, and few-shot adaptability is notably lacking. Therefore, this study aims to develop an interpretable landslide susceptibility assessment model for Sangzhi County that integrates the strengths of physical mechanism understanding and machine learning. The model seeks to address the black-box limitation and regional dependency of current approaches, explore the influence mechanisms of key dynamic triggering factors, and enhance model robustness and generalization in few-shot scenarios by exploring meta-learning or similar strategies. The ultimate goal is to provide a scientific, accurate, and practical decision-support tool for landslide risk management in this area.

2. Study Area and Data

2.1. Study Area

Sangzhi County is located on the north-western margin of Hunan Province and administratively belongs to Zhangjiajie City (as shown in Figure 1). It lies between 109°41′–110°46′ E and 29°17′–29°84′ N and forms an elongated east-west strip roughly 104 km long and 51 km wide, covering 3474 km^2. Situated in the core of the Wuling Mountains, the terrain tilts from north-west to south-east and is classified as medium- to high-relief mountains. The highest point is the summit of Cape Mountain in the Badagongshan massif

https://doi.org/10.3390/land15040579

(1898.4 m a.s.l.), whereas the lowest point is the Liuyangxi valley floor (154 m a.s.l.), giving a maximum relief of 1744 m. The bedrock consists mainly of slate and sandstone of the Triassic and Silurian systems, with relics of the older Sinian sequence. The climate is humid subtropical monsoon, with a mean annual temperature of 16.6 °C and mean annual precipitation of 1440.9 mm, placing the county among the three rainstorm centers of Hunan. Forest resources are abundant: woodland occupies 79.2% of the total area, the standing timber volume is about 18.86 million m^3, and the forest cover reaches 76.2%. A dense drainage network is formed by the Lishui and Loushui Rivers together with more than 410 tributaries. The Badagongshan National Nature Reserve harbors large tracts of pristine evergreen broad-leaved forest and rich flora and fauna, serving as a key ecological barrier in the middle Yangtze region.

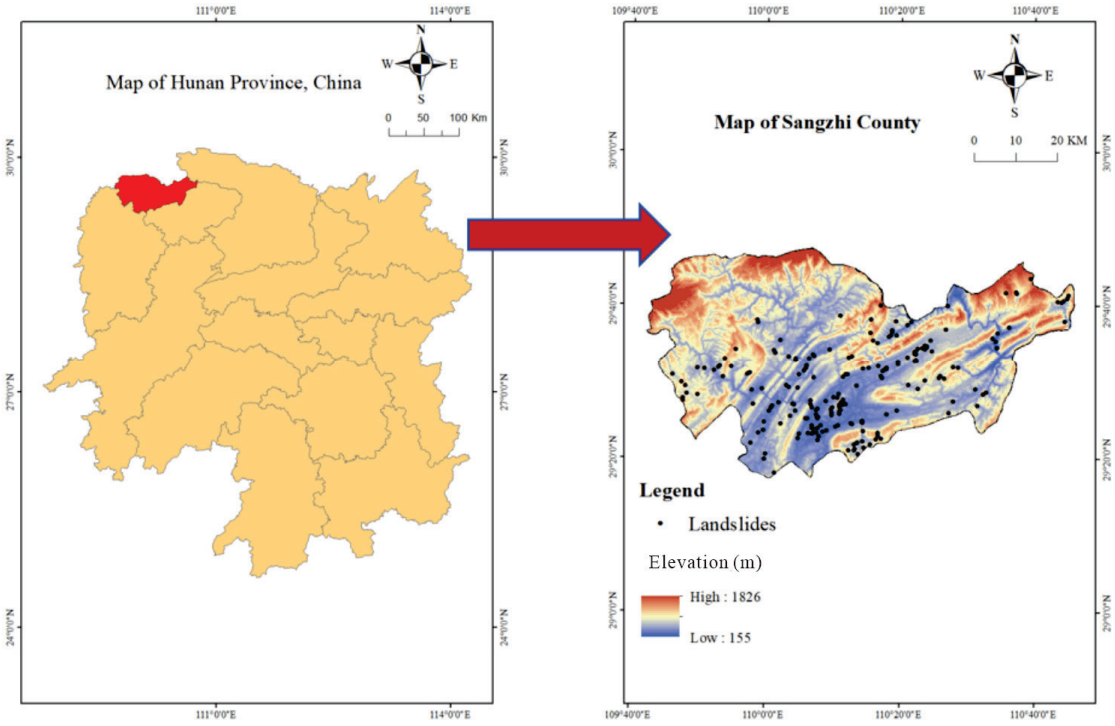


Figure 1. Overview map of the study area.

2.2. Data

2.2.1. Landslide and Non-Landslide Sample Preparation

Landslide locations (n = 206) were interpreted from historical archives and high-resolution Google Earth imagery and are concentrated in the central and southern parts of the county. To create a balanced dataset, a 30 m buffer was centered on each landslide centroid, and two additional points were randomly placed inside the buffer; together with the centroid, this yielded three positive samples per landslide. The 30 m buffer radius was selected based on the typical size of landslide initiation zones reported in the literature (e.g., diameters ranging from 10 to 50 m), ensuring that sampled points remain within the landslide-affected area while capturing intra-zone variability in topographic and environmental conditions. Negative samples were generated by randomly distributing 618 points at least 1 km away from any landslide polygon, a distance chosen to exclude latent or unmapped failures and to reduce spatial autocorrelation, thereby decreasing noise and improving model discrimination. The 206 landslides were then randomly split into 144 for training (~70%) and 62 for testing (~30%). Accordingly, the positive samples from the training landslides (432 samples) and an equal number of randomly selected negative

https://doi.org/10.3390/land15040579

samples formed the training set. In comparison, the remaining positive (186) and negative (186) samples constituted the testing set, approximately representing a 7:3 split ratio. The final inventory contains 618 positive and 618 negative samples in total.

2.2.2. Environmental Factors

Thirteen environmental predictors were selected for landslide-susceptibility modeling: elevation, slope, aspect, distance to roads, distance to faults, distance to rivers, mean annual rainfall, lithology, land-use type, topographic wetness index (TWI), NDVI, soil type, and vegetation type. Descriptions of evaluation-factor datasets are shown in Table 1. All covariates were resampled to 30 m resolution, yielding 3,799,219 grid cells across the county (Figure 2).

Table 1. Description of evaluation-factor datasets.

ID	Factor Name (Unit)	Abbrev.	Native Resolution	Resampled Resolution	Data Source	Period/Year
1	Normalized Difference Vegetation Index (–)	NDVI	30 m	30 m	Chinese Ecosystem Research Network	2019
2	Vegetation Functional Group (–)	VFG	30 m	30 m	CASEarth	2020
3	Elevation (m)	DEM	12.5 m	30 m	NASA ALOS 12.5 m DEM	2006–2011
4	Slope angle (°)	Slope	12.5 m	30 m	NASA ALOS 12.5 m DEM	2006–2011
5	TWI	TWI	12.5 m	30 m	NASA ALOS 12.5 m DEM	2006–2011
6	Slope aspect (–)	Aspect	12.5 m	30 m	NASA ALOS 12.5 m DEM	2006–2011
7	Soil type	Soil	shp	30 m	the 1:1,000,000 Soil Map of China	1995
8	Mean annual rainfall (mm)	Rain	1 km	30 m	Qinghai-Tibet Plateau Data Centre	2018–2022 mean
9	Distance to river (m)	DistRiv	Shp	30 m	OpenStreetMap (updated 2024)	2024
10	Distance to fault (m)	DistFau	1:200,000-scale mapping	30 m	1:200 k geological map	2022
11	Lithology (–)	Litho	1:200,000-scale mapping	30 m	1:200 k geological map	2022
12	Land-use type (–)	LU	30 m	30 m	Wuhan University dataset	2021
13	Distance to road (m)	DistRoad	Shp	30 m	OpenStreetMap (updated 2024)	2024

https://doi.org/10.3390/land15040579

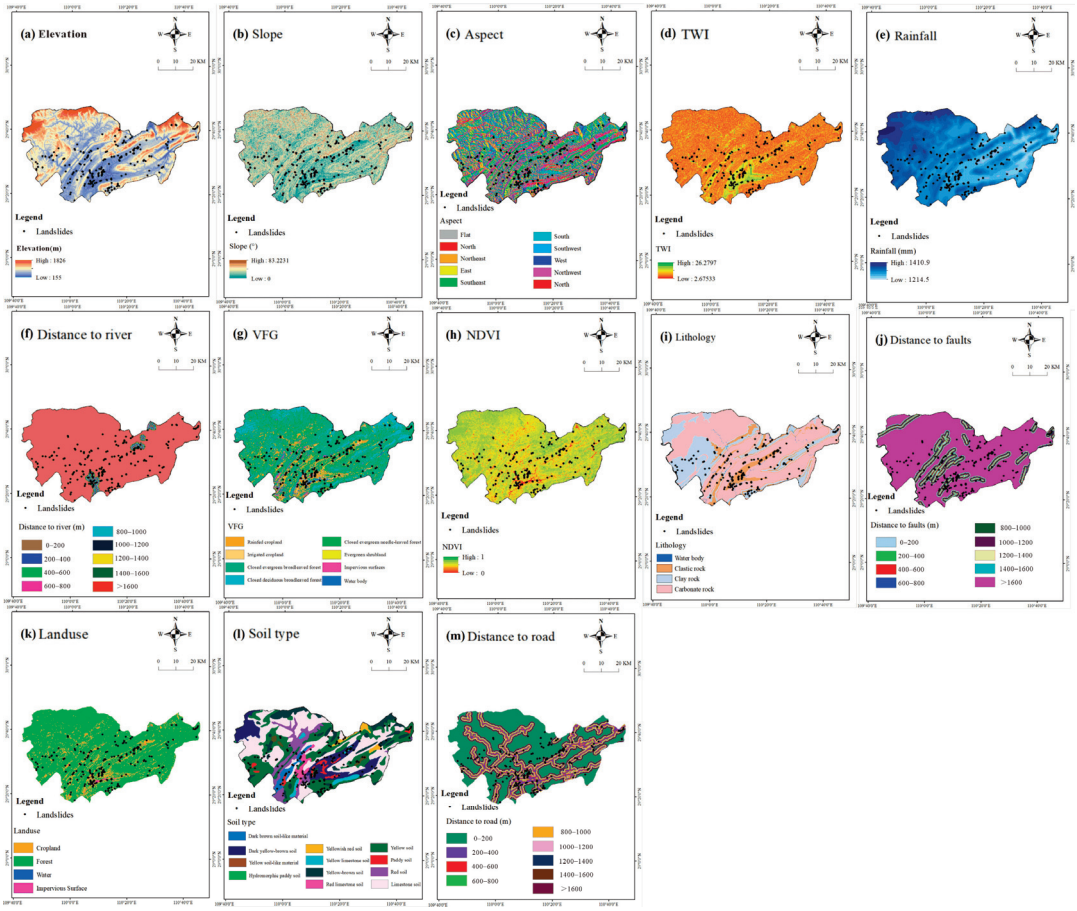


Figure 2. Grading chart of evaluation factors.

2.2.3. Topographic and Topographic Wetness Factors

Topography directly governs slope stability: elevation modulates lithostatic stress and groundwater flow; steeper slopes increase gravitational driving forces; aspect controls solar irradiance and rainfall redistribution, thereby altering rock/soil mechanical properties; TWI quantifies potential surface saturation, with mid-range values frequently coinciding with high landslide densities. In this study, elevation, slope, and aspect data were derived from NASA's 12.5 m-resolution DEM product provided by Japan's ALOS satellite, collected between 2006 and 2011 (Figure 2a–d).

2.2.4. Climatic and Hydrological Factors

Rainfall is the main external trigger, raising pore-water pressure and reducing shear strength. River undercutting at the toe of slopes further decreases stability; therefore, distance to rivers was included. Mean annual rainfall for 2018–2022 (native 1 km) was downscaled to 30 m by kriging [19], and river networks were extracted from OpenStreetMap (updated 2024) (Figure 2e,f).

2.2.5. Vegetation Characteristic Factors

Vegetation enhances slope stability through root reinforcement and evapotranspiration-induced suction. Vegetation type ranks rooting strength (trees > shrubs > herbs), whereas NDVI indicates canopy density; higher NDVI generally implies denser cover and greater mechanical reinforcement. The vegetation functional group (vegetation type) data (30 m) were obtained from the CASEarth land cover product [20], and the NDVI layer (30 m) was acquired from the Chinese Ecosystem Research Network [21].

https://doi.org/10.3390/land15040579

2.2.6. Geological and Soil Factors

Lithology, distance to faults, land-use type, and soil type control shear strength, permeability, and surface loading. Alternating hard/soft beds and intensely fractured fault zones promote failure, whereas land-use changes alter infiltration and loading. Lithology and faults were digitized from the 1:200,000 geological map; land use was taken from the 30 m CLCD product [22]; soil type was vectorized from the 1:1,000,000 Soil Map of China.

2.2.7. Human Activity Factor

Road construction alters original topography, redistributes in-situ stress, and induces dynamic loading; inadequate drainage further raises groundwater levels. Distance to roads was therefore adopted as the anthropogenic predictor, with road networks extracted from OpenStreetMap (updated 2024) (Figure 2m).

3. Materials and Methods

3.1. General Workflow

3.1.1. Data Acquisition and Preprocessing

Historical landslide inventories, a 12.5 m resolution DEM, 1 km rainfall grids, 1:200,000 geological maps, and OpenStreetMap road and river layers were systematically compiled for Sangzhi County. All layers were projected to a common coordinate system and resampled to 30 m to generate a standardized raster stack.

3.1.2. Figures, Tables, and Schemes

High-resolution Google Earth imagery was visually interpreted to delineate 206 landslide scarps. A balanced sampling strategy was then applied: two additional points were randomly generated within a 30 m buffer around each scarp centroid, yielding three positive samples per slide; 618 negative points were randomly placed >1 km away from any landslide polygon, resulting in 618 positive and 618 negative samples (n = 1236).

3.1.3. Factor Screening and Importance Analysis

Multicollinearity was first diagnosed using Pearson correlation and variance inflation factor (VIF); thirteen factors with VIF < 5 were retained. Their relative importance was subsequently assessed through information gain, entropy weight, Random Forest, LightGBM, and AdaBoost to identify the dominant drivers of landsliding.

3.1.4. Model Development and Performance Evaluation

A Triple-Source Probabilistic Fusion Bidirectional LSTM (TSPF-BiLSTM) framework was proposed. Posterior probabilities from Random Forest, LightGBM, and AdaBoost were calibrated by five-fold Platt scaling and concatenated into a six-dimensional vector. This vector was fed into a 4-layer bidirectional LSTM with attention and residual connections for deep nonlinear fusion. The proposed TSPF-BiLSTM model was compared with five baseline models, including Random Forest (RF), LightGBM (LGB), AdaBoost (ADA), Convolutional Neural Network (CNN), and Transformer. Performance was quantified with AUC, accuracy, F1-score, and recall under ten-fold cross-validation to ensure robustness.

3.1.5. Susceptibility Mapping and Validation

The optimal model was applied to each 30 m pixel to produce a landslide-susceptibility index. Natural-break classification divided the county into five classes (very low, low, moderate, high, and very high). Frequency-ratio analysis and spatial overlay with historical landslides were used to validate the model's ability to identify high-hazard zones, providing a scientific basis for risk reduction and land-use planning in Sangzhi County.

https://doi.org/10.3390/land15040579

3.2. Machine-Learning Algorithms

3.2.1. Random Forest (RF)

RF uses Bootstrap resampling to generate multiple training subsets; each subset grows a CART decision tree, and node splitting is performed only within a randomly selected feature subset. The final prediction is obtained by majority vote (or averaging), while the out-of-bag (OOB) error provides an unbiased estimate of generalization performance. The ensemble is inherently robust to high-dimensional, small-sample, and missing-data scenarios [23].

3.2.2. LightGBM (LGB)

Within the GBDT framework, LGB employs histogram-based gradient boosting: continuous features are first discretized into k bins, and optimal splits are searched on the histogram. A depth-limited leaf-wise growth strategy, combined with exclusive feature bundling (EFB) and one-sided gradient sampling (GOSS), markedly reduces memory footprint and computation time. The algorithm handles categorical features directly without one-hot encoding while maintaining competitive accuracy [24].

3.2.3. AdaBoost

AdaBoost sequentially trains a series of weak classifiers; at each iteration, the weights of misclassified samples are increased, and the weighted error of the base learner is computed. The final strong classifier is formed by weighted majority voting. The method is simple to implement, requires minimal parameter tuning, resists over-fitting as iterations increase, and can be combined with any weak learner, offering high versatility [25].

3.2.4. CNN

Convolutional Neural Networks (CNNs) employ convolutional layers that apply sliding filters to extract local patterns through hierarchical feature learning, making them particularly effective for spatial feature extraction [26]. In this study, we implement a CNN architecture comprising two convolutional layers (kernel size 1×3, with batch normalization and ReLU activation), an adaptive max-pooling layer, and two fully connected layers for landslide susceptibility classification. Key parameters include: hidden size = 64, dropout rate = 0.3, and Adam optimizer (learning rate = 0.001, weight decay = 1×10^{-4}).

3.2.5. Transformer

The Transformer architecture, built upon self-attention mechanisms, processes input sequences in parallel and captures long-range dependencies without recurrence, revolutionizing sequence modeling [27]. Our implementation utilizes a single Transformer encoder layer with multi-head attention (8 heads), a feed-forward network (dimension = 256), and two fully connected layers. Critical parameters include: embedding dimension = 128, dropout rate = 0.1, and Adam optimizer (learning rate = 0.001, weight decay = 1×10^{-4}).

3.3. Triple-Source Probabilistic Fusion Bidirectional LSTM (TSPF-BiLSTM)

TSPF-BiLSTM is a probability-level, end-to-end ensemble learning framework that performs second-stage nonlinear deep fusion of posterior probabilities generated by heterogeneous base learners for landslide susceptibility prediction tasks. The network architecture comprises three sequentially connected modules, as illustrated in Figure 3.

https://doi.org/10.3390/land15040579

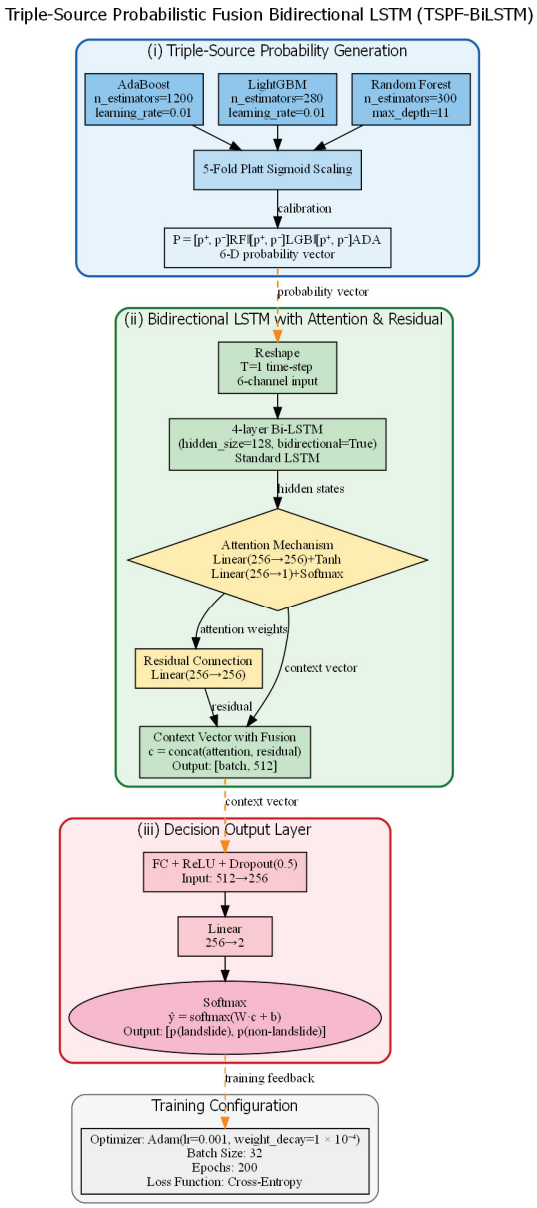


Figure 3. Architecture of the proposed model.

3.3.1. Triple-Source Probability Generation

To capture multi-dimensional characteristics of landslide susceptibility, three heterogeneous base learners were independently trained with the following core parameter configurations: Random Forest with 300 decision trees, a maximum depth of 11, a fixed number of features randomly selected as candidates at each split, and the Gini index as the split criterion; LightGBM with 280 boosting rounds, a learning rate of 0.01, a maximum depth of 5, gradient-based one-side sampling to improve training efficiency, and information gain as the split criterion; AdaBoost with 1200 weak classifiers, a learning rate of 0.01, decision stumps (maximum depth of 1) as base learners, and iterative sample weight adjustments to focus on difficult-to-classify instances. Posterior probabilities from each base learner were calibrated using 5-fold Platt Scaling to mitigate probability distribution biases arising from structural differences among models. The three sets of calibrated probabilities (for each class) were concatenated into a 6-dimensional feature vector, serving as input for the subsequent fusion model.

https://doi.org/10.3390/land15040579

$$P = \left[{p_{RF}}^{+}, {p_{RF}}^{-}, {p_{LGB}}^{+}, {p_{LGB}}^{-}, {p_{Ada}}^{+}, {p_{Ada}}^{-} \right] \in R^{n \times 6} \quad (1)$$

where superscripts pRF+, pRF− denote positive/negative class probabilities. Calibration guarantees that the three base learners share a common probability scale and supplies an unbiased input for subsequent deep fusion.

3.3.2. Bidirectional LSTM with Attention and Residual Fusion

The 6-D vector is reshaped into a sequence of length 1 with 6 channels, then fed into a 4-layer bidirectional LSTM. The forward and backward hidden states of the bidirectional LSTM are concatenated to form the final hidden state representation:

$$h_t = \text{Bi-LSTM}(P, \theta_{lstm}), \; H = [{h^{\rightarrow}}_t; {h^{\leftarrow}}_t] \in R^{256} \quad (2)$$

where ${h^{\rightarrow}}_t$ and ${h^{\leftarrow}}_t$ represent the forward and backward LSTM hidden states at time step t, respectively, and [;] denotes vector concatenation. With time steps T = 1, we obtain hidden states $H \in R^{B \times 1 \times 256}$, where B is the batch size.

The LSTM output passes through an additive attention mechanism ($256 \rightarrow 256 \rightarrow 1$, Tanh + Softmax) to generate attention weights, and the weighted sum produces the context vector. A linear transformation ($256 \rightarrow 256$) is applied to the attention output as a residual, and both are concatenated into a 512-D fusion vector.

3.3.3. Decision Output Layer

After passing through a fully connected layer with ReLU activation, the vector c is regularized by Dropout (p = 0.5) to mitigate overfitting under small-sample conditions. This dropout rate has been widely adopted in landslide-susceptibility assessments and consistently shown across diverse small-scale datasets to offer a robust compromise between preserving model capacity and enhancing generalization, before the softmax layer outputs the final class probabilities:

$$y = \text{softmax}(W_{out} \cdot \text{ReLU}(W_{fc} \cdot c + b_{fc}) + b_{out}) \quad (3)$$

A dropout rate of 0.5 has been widely validated in landslide susceptibility assessments, achieving an optimal balance between preserving model capacity and enhancing generalization, particularly suitable for small-scale dataset scenarios.

Training configuration: Adam optimizer (learning rate 0.001, weight decay 0.0001), batch size 32, 200 epochs, cross-entropy loss, and ReduceLROnPlateau scheduler.

3.4. Entropy-Weighting Model

Entropy weighting uses information entropy to quantify indicator dispersion. Entropies are computed column-wise on the normalized indicator matrix; a higher entropy indicates greater disorder and a smaller weight. Difference coefficients are derived as (1 − entropy), normalized to objective weights, and multiplied by the standardized data to produce composite scores. The procedure is entirely data-driven, eliminates subjective judgment, and is often coupled with TOPSIS for multi-criteria evaluation [28].

3.5. Information-Gain Model

Information gain employs Shannon entropy to measure dataset purity. For feature A, gain is defined as $\Delta H = H(S) - H(S \mid A)$; a larger gain indicates a stronger ability to reduce uncertainty, so the criterion is used for decision-tree splitting. The metric is simple, interpretable, and can guide tree growth or serve as a filter-based feature selector to improve subsequent modeling efficiency [29].

https://doi.org/10.3390/land15040579

4. Results

4.1. Multicollinearity Diagnosis and Feature Selection

4.1.1. Pearson Correlation Matrix Screening

In the present study, pairwise Pearson correlation coefficients were computed for all thirteen environmental predictors. As illustrated in Figure 4, no absolute correlation exceeded 0.8, indicating the absence of severe linear multicollinearity; consequently, all variables were retained for subsequent analysis.

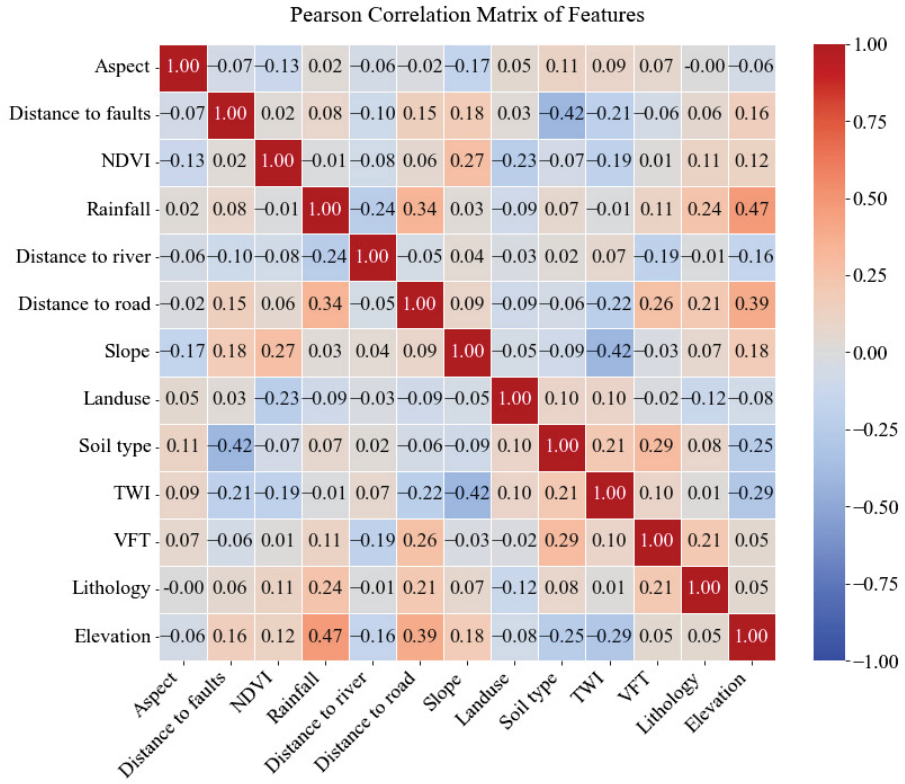


Figure 4. Pearson covariance index plot.

4.1.2. Variance-Inflation-Factor (VIF) Refinement

To further exclude potential multicollinearity, VIF was calculated for each predictor (Figure 5). All VIF values were <5, well below the customary warning threshold of 10 and the stricter cut-off of 10; the maximum VIF was only 1.7, the minimum 1.1, and the mean 1.42 ± 0.17. The low VIFs indicate strong independence among the predictors; therefore, all 13 environmental factors were retained for subsequent contribution-rate analysis.

4.1.3. Analysis of the Contribution Rates of the Five Models

Figure 6a–e compares the importance ranks returned by five algorithms. Cross-model consensus shows that rainfall and elevation are the top two drivers of landsliding in Sangzhi County (Figure 6). Random Forest assigns the highest weights to elevation (0.186), rainfall (0.164), NDVI (0.111) and lithology (0.095); AdaBoost emphasizes rainfall (0.210) and elevation (0.150), followed by distance to faults (0.110) and slope (0.130); LightGBM ranks elevation (0.149), NDVI (0.142), rainfall (0.144) and aspect (0.110) as most influential; Entropy identifies NDVI (0.154), rainfall (0.151), and elevation (0.149) as the key contributors, with aspect (0.137) following closely; Information Gain highlights TWI (0.173) and NDVI (0.153), with rainfall (0.149) and aspect (0.143) next.

https://doi.org/10.3390/land15040579

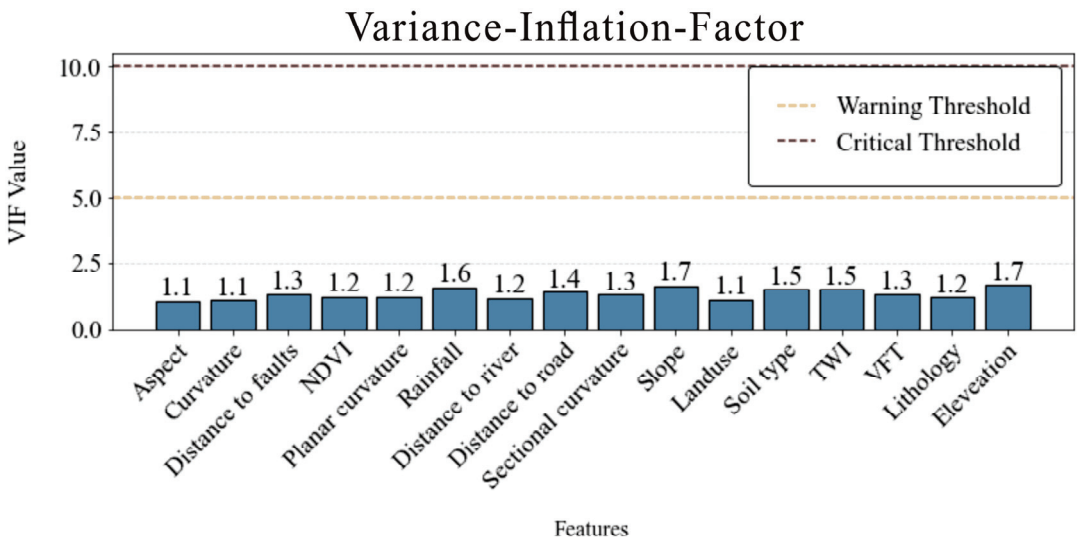


Figure 5. VIF values of the conditioning factors.

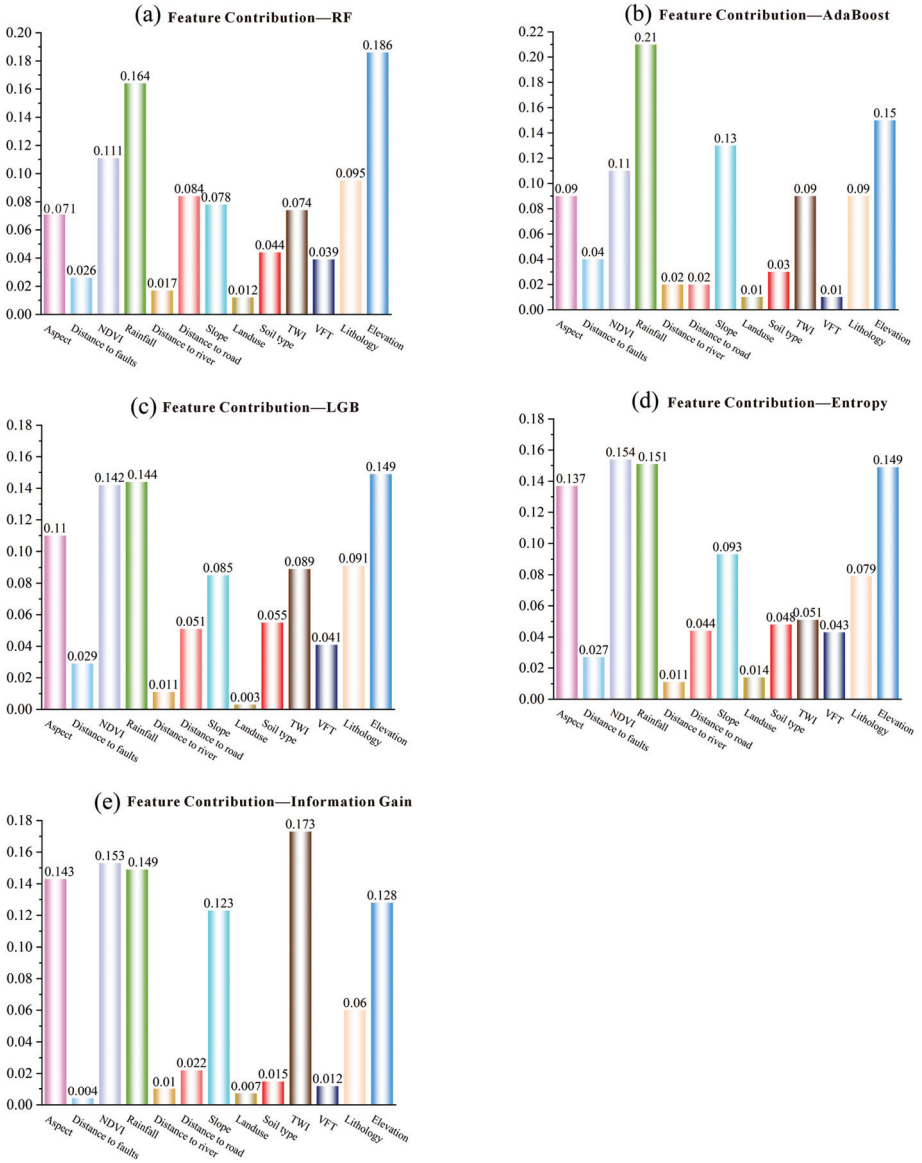


Figure 6. Contribution rate of evaluation factors of different models.

https://doi.org/10.3390/land15040579

Despite minor differences, the consistent primacy of rainfall and elevation underlines their key roles: rainfall directly controls soil saturation, pore-water pressure, and rock/soil strength, while elevation integrates relief, hydrology, and human activity intensity, indirectly reflecting slope stability. High ranks for slope and NDVI across algorithms further confirm the importance of topography and vegetation cover.

4.2. ROC-Curve-Based Accuracy Assessment

4.2.1. Comparative Analysis of Model Performance and Stability

The ROC curve synthetically evaluates the discriminative performance of a binary classifier by plotting the true-positive rate (TPR) against the false-positive rate (FPR) at consecutive decision thresholds [30]. In this study, ten-fold cross-validation (10-Fold CV) was adopted to assess the performance of six models—Random Forest (RF), LightGBM (LGB), AdaBoost, Convolutional Neural Network (CNN), Transformer, and TSPF-BiLSTM—and the corresponding ROC curves are shown in Figure 7. The area under the curve (AUC) was computed, yielding mean values of 0.9403 ± 0.0129, 0.9003 ± 0.0237, 0.8394 ± 0.0323, 0.8693 ± 0.0237, 0.8563 ± 0.0231, and 0.9525 ± 0.0115 for RF, LGB, AdaBoost, CNN, Transformer, and TSPF-BiLSTM, respectively. During the 10-fold CV, the original dataset was randomly divided into ten mutually exclusive subsets; nine subsets were used for training, and the remaining one for validation. This procedure was repeated ten times, and the mean and standard deviation of the performance metrics were calculated to estimate model stability and generalization capability. Examination of 95% confidence intervals (derived from the mean AUC and standard deviation) shows that TSPF-BiLSTM not only achieves the highest mean AUC (0.9525) but also possesses the narrowest interval, indicating excellent stability and generalization. The RF model exhibits high discriminative performance (AUC = 0.9403) with a narrow confidence interval, reflecting robust and consistent outcomes. The LGB model delivers a mean AUC of 0.9003 with a moderate confidence interval, denoting acceptable stability. The CNN and Transformer models yield mean AUCs of 0.8693 and 0.8563, respectively, with comparable interval widths, suggesting moderate performance and generalization. AdaBoost yields the lowest mean AUC (0.8394) and the widest confidence interval, signifying substantial performance fluctuations and limited generalization. Overall, within the landslide-susceptibility models constructed in this study, the Triple-Source Probabilistic Fusion BiLSTM framework demonstrates the best comprehensive performance, combining the highest discriminative accuracy with the most robust estimates; RF also shows exceptionally strong and stable performance, while LGB, CNN, and Transformer offer acceptable to moderate results, and AdaBoost performs relatively poorly under the present experimental setting.

4.2.2. Base Learner Contribution Analysis Based on Attention Weights

To explicitly characterize the contributions of the three base learners within the TSPF-BiLSTM fusion framework, we analyze the self-attention weights generated during forward propagation. The additive attention mechanism, applied after the bidirectional LSTM layer, dynamically assigns importance weights to the hidden representations derived from the probability outputs of the base models. Similar interpretable attention mechanisms have been successfully applied in landslide susceptibility mapping studies, demonstrating their capability to quantify input contributions and enhance model interpretability [31]. Across ten-fold cross-validation, the mean attention weights—interpreted as the relative focus of the model on each base learner's output during the fusion process—are quantified as follows: Random Forest: 0.4025 ± 0.2189 (approximately 40.2%), LightGBM: 0.2359 ± 0.0830 (23.6%), AdaBoost: 0.3617 ± 0.2004 (36.2%). These results suggest that Random Forest receives relatively higher average attention weights in the fusion process, with AdaBoost

https://doi.org/10.3390/land15040579

also receiving substantial attention. At the same time, LightGBM contributes stably but to a lesser extent. The observed standard deviations, particularly for Random Forest and AdaBoost, reflect the adaptive nature of the attention mechanism, which modulates focus according to input-specific characteristics. It should be noted that attention weights indicate the model's focus on different base learner outputs during dynamic fusion, and their representation of feature importance is not entirely equivalent to direct measures such as Gini impurity. This dynamic weighting effectively harnesses the complementary strengths of each base learner—Random Forest's robustness with high-dimensional features, AdaBoost's focus on challenging samples, and LightGBM's efficiency in gradient-based optimization—thereby enhancing the overall discriminative performance and generalization ability of the integrated framework. The explicit quantification of attention weights elucidates the internal fusion mechanism and offers interpretable evidence of how probability-level integration improves model stability and prediction accuracy in landslide susceptibility mapping under data-scarce conditions.

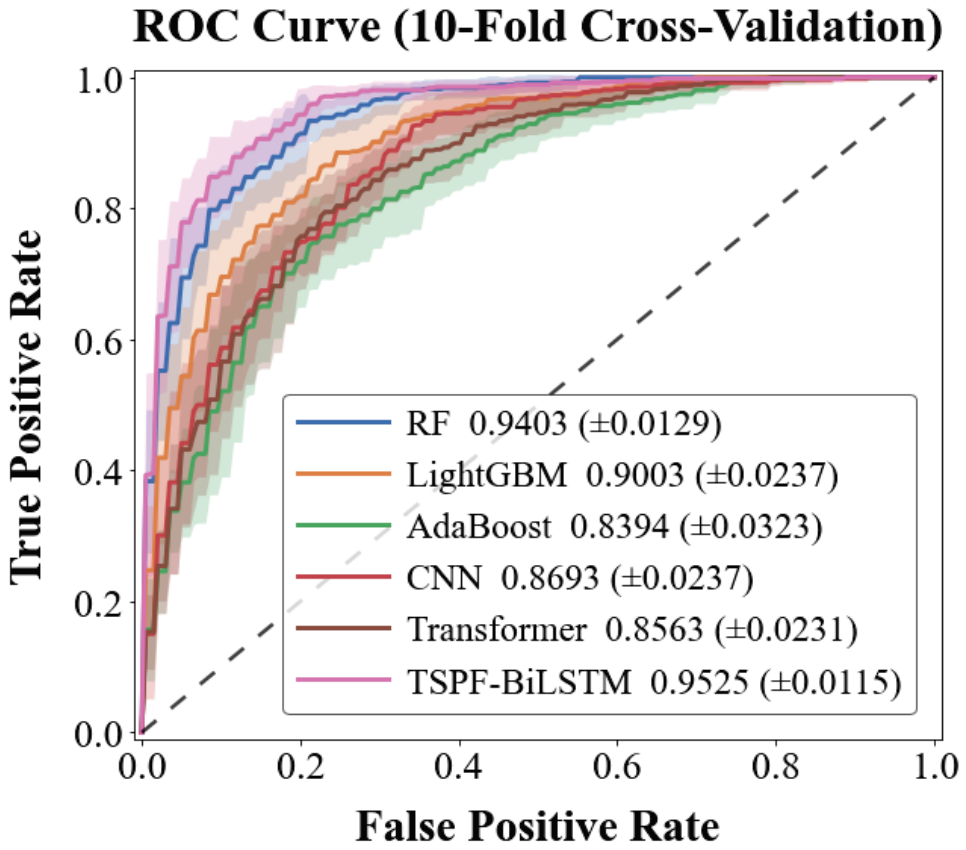


Figure 7. ROC curves of the six models under ten-fold cross-validation.

4.3. Multi-Metric Comparison of Recall, F1-Score, ACC, and Precision Across Six Models

Although the ROC-based analysis indicates that the TSPF-BiLSTM model achieves a significantly higher AUC than its competitors, relying solely on AUC may provide an incomplete assessment—particularly under the imbalanced landslide-present versus landslide-absent conditions typical in regional susceptibility studies [32]. The AUC reflects the overall discriminative capacity of a classifier; however, when class prevalence is skewed, it can overestimate the true ability to detect the minority class and may conceal deficiencies in identifying critical high-risk locations [33,34]. To obtain a more holistic appraisal, we therefore supplement AUC with accuracy (ACC), recall, precision, and the F1-score, all derived from the ten-fold cross-validation averages (Figure 8).

https://doi.org/10.3390/land15040579

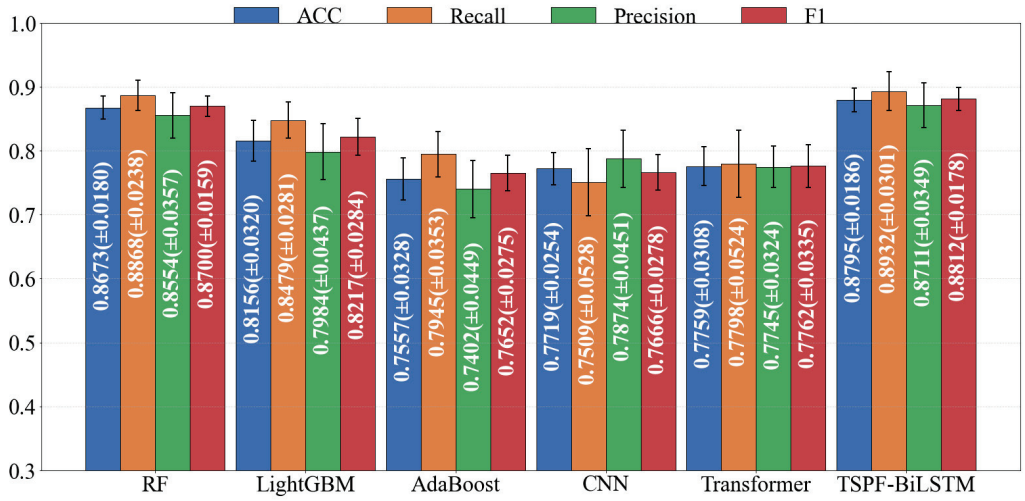


Figure 8. Performance metrics comparison of the six models.

ACC can be disproportionately influenced by correct classifications of the majority class in imbalanced datasets, potentially creating a "high-accuracy illusion" that masks poor performance on the minority class of interest [35]. Among the six models, TSPF-BiLSTM attains the highest ACC (0.8795 ± 0.0186), followed by RF (0.8673 ± 0.0180) and LightGBM (0.8156 ± 0.0320). CNN (0.7719 ± 0.0254) and Transformer (0.7759 ± 0.0308) perform moderately, whereas AdaBoost records the lowest ACC (0.7557 ± 0.0328), revealing a marked shortfall in overall classification performance.

Recall (sensitivity) measures the proportion of true landslide pixels correctly identified. It was introduced as one of the core evaluation metrics in this study because it directly reflects the model's coverage of actual landslide areas; a higher recall can effectively reduce blind spots in geological hazard assessment caused by missed detections [36]. TSPF-BiLSTM again achieves the highest recall (0.8932 ± 0.0301), demonstrating enhanced detection of high-risk locations. RF shows a competitive recall of 0.8868 ± 0.0238, while LightGBM follows with 0.8479 ± 0.0281. Transformer (0.7798 ± 0.0524), CNN (0.7509 ± 0.0528), and AdaBoost (0.7945 ± 0.0353) exhibit lower recall values, implying a less robust capability to capture complex landslide signatures, particularly where false negatives are critical.

Precision indicates the reliability of positive predictions. TSPF-BiLSTM yields the highest precision (0.8711 ± 0.0349), followed by RF (0.8554 ± 0.0357), suggesting both models make fewer false-positive errors. LightGBM (0.7984 ± 0.0437), Transformer (0.7745 ± 0.0324), CNN (0.7874 ± 0.0451), and AdaBoost (0.7402 ± 0.0449) show progressively lower precision, reflecting a higher tendency to incorrectly label stable areas as susceptible.

The F1-score harmonizes precision and recall, offering a single metric for performance under class imbalance [37]. TSPF-BiLSTM yields the best F1-score (0.8812 ± 0.0178), indicating a well-balanced trade-off between omission and commission errors. RF achieves a close second with 0.8700 ± 0.0159. LightGBM (0.8217 ± 0.0284), Transformer (0.7762 ± 0.0335), CNN (0.7666 ± 0.0278), and AdaBoost (0.7652 ± 0.0275) lag substantially behind.

Collectively, TSPF-BiLSTM delivers the best results on all four complementary metrics, achieving the highest ACC, Recall, Precision, and F1-score. RF performs robustly across all metrics, closely following TSPF-BiLSTM in multiple categories. LightGBM follows with acceptable but lower performance, while CNN, Transformer, and AdaBoost yield the weakest overall scores, with AdaBoost consistently at the bottom. These findings confirm that the proposed fusion strategy, which synergizes dynamic weight adjustment with sequence-aware learning, can effectively enhance generalization in complex geological environments.

https://doi.org/10.3390/land15040579

In the following section, frequency-ratio analysis over the very-high susceptibility zone is used to further validate the model's ability to localize historical landslide pixels.

4.4. Landslide-Susceptibility Mapping of the Six Models

The trained RF, LGB, AdaBoost, CNN, Transformer, and TSPF-BiLSTM models were applied to every 30 m pixel in Sangzhi County (3,799,219 cells) to estimate landslide occurrence probability. The resulting probability surfaces were imported into ArcGIS 10.7 and classified into five susceptibility levels—very low, low, moderate, high, and very high—using the natural breaks method. The susceptibility maps produced by the four algorithms are presented in Figure 9.

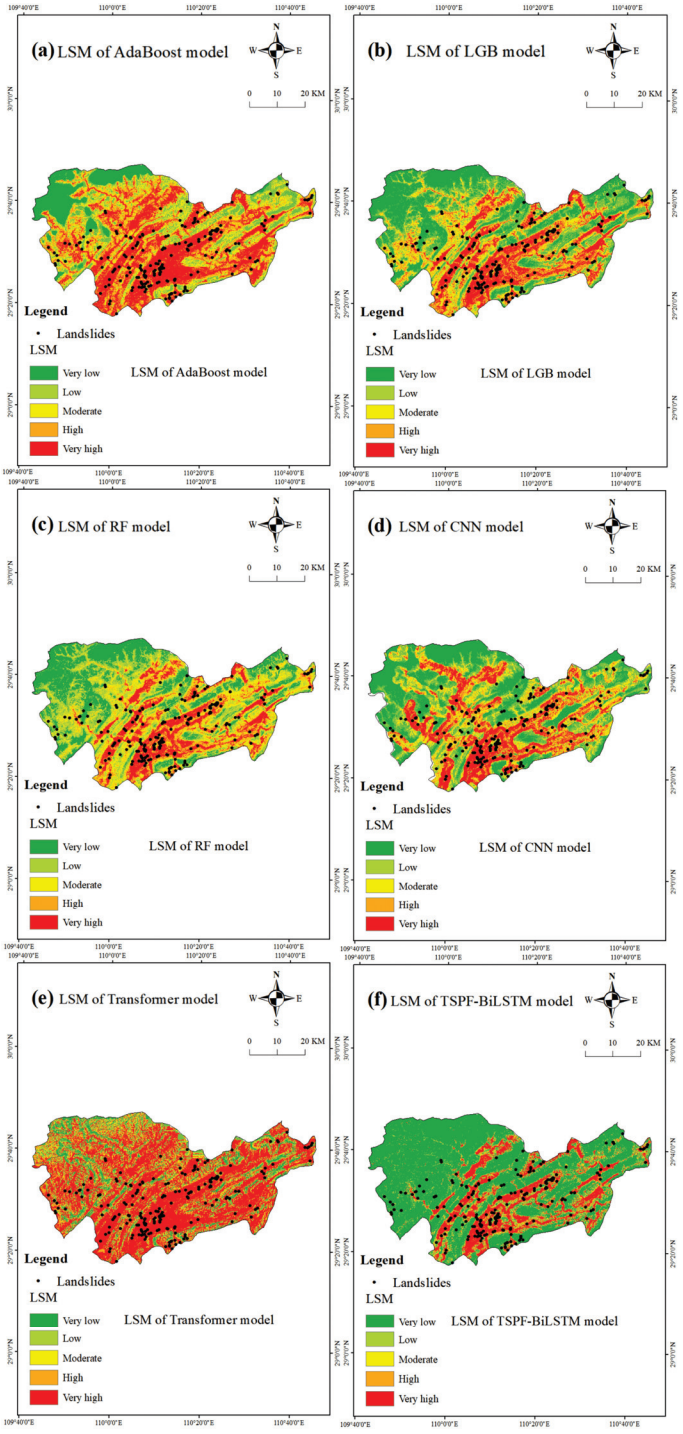


Figure 9. Evaluation plot of susceptibility for the 6 models.

Figure 10(a) and (b) illustrate the predicted distribution of susceptibility classes across the entire study area and within the known landslide inventory for each model, respectively. Figure 10c presents the key evaluation metric, the Landslide Frequency Ratio (FR), where a higher value indicates a greater ability of the model to concentrate known landslides into high-susceptibility zones.

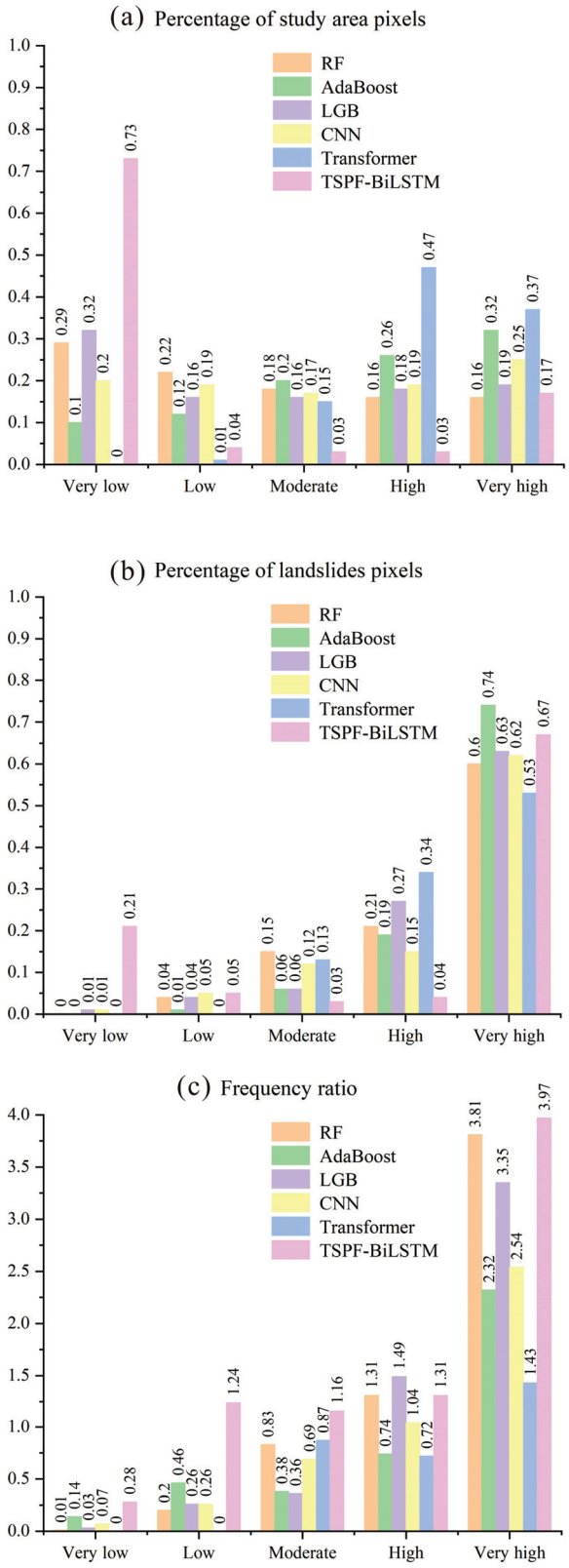


Figure 10. Overview of the classification of susceptibility areas for different models.

Analysis of the figures reveals distinct performance differences among the models. The TSPF-BiLSTM model shows a unique pattern in Figure 10a, with its predictions heavily concentrated in the "Very Low" susceptibility class. However, within the landslide inventory depicted in Figure 10b, the distribution shifts decisively, concentrating in the "High" and "Very High" classes, similar to other models. This contrast suggests strong discriminatory power. This is confirmed in Figure 10c, where TSPF-BiLSTM achieves the highest Frequency Ratio (3.97) in the "Very High" susceptibility class, significantly outperforming all others, demonstrating its superior efficacy in identifying the highest-risk zones where landslides actually occur.

In contrast, the Transformer model shows a strong tendency to classify the vast majority of pixels as "Very High" susceptibility in both Figure 10a,b, yet it yields the lowest FR (1.43) for this class in Figure 10c, indicating a significant prediction bias. The RF and LGB models demonstrate robust, similar FR performance in Figure 10c, particularly in the "Very High" class (3.81 and 3.35, respectively), making them strong benchmarks. The CNN model exhibits the most distinct FR peak in the "High" susceptibility class. The AdaBoost model shows intermediate performance across all figures.

In conclusion, the TSPF-BiLSTM framework demonstrates a significant advantage in accurately delineating areas at very high risk of landslides, offering a more reliable tool for landslide susceptibility mapping.

5. Discussion

5.1. Superiority of the TSPF-BiLSTM Probabilistic-Fusion Framework in Karst Mountains

The TSPF-BiLSTM model achieved a mean AUC of 0.9525 with a standard deviation of 0.0115 under ten-fold cross-validation, substantially outperforming all baseline models (Random Forest: 0.9403 ± 0.0129; LightGBM: 0.9003 ± 0.0237; AdaBoost: 0.8394 ± 0.0323; CNN: 0.8693 ± 0.0237; Transformer: 0.8563 ± 0.0231). Its narrow confidence interval reflects excellent stability across different data splits—a particularly valuable characteristic given the limited sample size (618 samples per class) in the study area.

This superior performance stems from the framework's multi-level architecture. First, Platt–Sigmoid calibration transforms the raw outputs of three heterogeneous base learners into posterior probabilities on a unified scale, effectively mitigating the confidence-distribution shifts commonly encountered in ensemble methods. Second, the bidirectional LSTM layer with attention mechanism and residual connections performs deep nonlinear fusion of the calibrated probability vectors, capturing complex probabilistic interactions that simple linear weighting cannot represent. Attention-weight analysis (Section 4.2.2) revealed that the model dynamically leverages the complementary strengths of each base learner: Random Forest received the highest average attention (40.2%), reflecting its robustness to high-dimensional features; AdaBoost received substantial attention (36.2%) for its ability to focus on difficult samples; and LightGBM contributed stably (23.6%), providing efficient gradient-based optimization. Third, a dropout rate of 0.5 balances model capacity with the limited sample size, effectively preventing overfitting.

The practical significance of this improved performance is evident in the frequency-ratio analysis. Within the very-high susceptibility zone, the TSPF-BiLSTM model achieved a frequency ratio of 3.97, substantially higher than Random Forest (3.81) and LightGBM (3.35). This indicates that the model not only classifies accurately but also concentrates predicted high-risk areas more precisely where historical landslides have actually occurred—a crucial characteristic for practical hazard management.

In contrast, the Transformer model classified 76% of the study area as very-high susceptibility yet yielded the lowest frequency ratio (1.43) for that class, revealing clear over-prediction bias likely resulting from the mismatch between its high model capacity

https://doi.org/10.3390/land15040579

and insufficient regularization for small datasets. The CNN model showed a distinct performance peak in the high-susceptibility class rather than the very-high class, suggesting that its local feature extraction may not fully capture regional-scale landslide drivers. AdaBoost alone performed the poorest overall (AUC = 0.8394, ACC = 0.7557), consistent with its known sensitivity to noise and outliers in imbalanced datasets.

5.2. Relationship Between Model Differences and Landslide Probability

The six models exhibit distinct prediction patterns (Figure 9), but these differences are not arbitrary—they reflect varying sensitivities to the environmental factors controlling landslide occurrence. The TSPF-BiLSTM framework, through probabilistic fusion, effectively captures consensus among base learners while suppressing individual biases. Its highest frequency ratio in the very-high class demonstrates that the fused output aligns more closely with the spatial distribution of historical landslides.

The over-prediction bias of the Transformer model warrants further analysis. Its tendency to classify most areas as very-high susceptibility likely stems from the self-attention mechanism overfitting to noise patterns in the small dataset when processing high-dimensional features, resulting in overly generous decision boundaries. The CNN model's peak in the high-susceptibility class rather than the very-high class may indicate that its convolutional kernels primarily extract local topographic features while failing to fully integrate regional-scale climatic and geological controls. These differences further underscore the necessity of ensemble strategies that integrate multiple perspectives; the TSPF-BiLSTM framework achieves this integration at the probability level, generating a more balanced and physically plausible susceptibility map.

5.3. Identification of Dominant Predisposing Factors and Landslide Response Mechanisms

Multi-model contribution analysis (Section 4.1.3) consistently identified NDVI (Table 2), elevation (Table 3), and annual rainfall (Table 4) as the three most important predisposing factors, with mean contribution rates of 0.152, 0.147, and 0.136, respectively. Frequency-ratio validation further quantified their influence:

Table 2. Landslide distribution and frequency ratio for different NDVI classes.

NDVI Class	Pixel Count	Area Ratio (%)	Landslide Pixel Count	Landslide Ratio (%)	Frequency Ratio
0–0.494	30,438	0.008	4	1.9	2.424
0.494–0.686	91,226	0.024	15	7.3	3.032
0.686–0.8	282,766	0.074	44	21.4	2.870
0.8–0.862	1,356,789	0.357	88	42.7	1.196
0.862–1	2,038,000	0.536	55	26.7	0.498

Table 3. Landslide distribution and frequency ratio for different elevation classes.

Elevation Class (m)	Pixel Count	Area Ratio (%)	Landslide Pixel Count	Landslide Ratio (%)	Frequency Ratio
155–462	968,843	0.255	124	60.2	2.361
462–658	1,111,989	0.293	62	30.1	1.028
658–871	890,423	0.234	15	7.3	0.311
871–1143	594,633	0.157	4	1.9	0.124
1143–1826	233,416	0.061	1	0.5	0.079

https://doi.org/10.3390/land15040579

Table 4. Landslide distribution and frequency ratio for different rainfall classes.

Rainfall Class (mm)	Pixel Count	Area Ratio (%)	Landslide Pixel Count	Landslide Ratio (%)	Frequency Ratio
1214.5–1273.6	637,332	0.168	59	28.6	1.707
1273.6–1301	922,259	0.243	98	47.6	1.960
1301–1328.8	1,130,027	0.297	37	18	0.604
1328.8–1361.4	827,306	0.218	11	5.3	0.245
1361.4–1410.9	282,295	0.074	1	0.5	0.065

NDVI 0–0.686: Frequency ratios of 2.424–3.032, indicating that areas with sparse or low vegetation cover are substantially more susceptible to landsliding.

Elevation 155–462 m: Containing 60.2% of historical landslides with a frequency ratio of 2.361, reflecting the concentration of slope instability at moderate elevations where human activity and natural processes intersect.

Annual rainfall 1273.6–1301 mm: Frequency ratio of 1.960, confirming that areas receiving intermediate-to-high rainfall are most prone to failure.

These intervals align well with records from geologically similar karst areas in northern Guizhou and southeastern Chongqing, providing a quantitative basis for cross-regional landslide zoning within the Wuling Mountains karst belt and supporting the regional applicability of the identified drivers.

It is important to clarify that elevation is a static topographic attribute and should not be misinterpreted as a triggering factor. The primary triggering mechanism in this region is rainfall, which directly controls soil saturation and pore-water pressure dynamics. However, the current study employed mean annual rainfall—a long-term climatic average—rather than event-based rainfall data. While this captures spatial climatic patterns, it does not represent the short-term, transient processes that actually initiate slope failure. Similarly, the analysis did not include direct measurements of soil pore-water pressure or key soil hydraulic parameters such as field capacity and permanent wilting point, which govern the transient hydrological response of slopes to rainfall.

5.4. Methodological Considerations and Limitations

Several methodological issues warrant discussion. First, all environmental predictors were resampled to a uniform 30 m resolution before analysis. The original resolutions varied considerably: DEM (12.5 m), NDVI and land use (30 m), and rainfall (1 km). Resampling to 30 m preserved the detail of fine-resolution layers while downscaling coarse layers through kriging (for continuous variables) or nearest-neighbor assignment (for categorical variables). This standardization ensured spatial consistency and avoided scale-induced artefacts, as confirmed by multicollinearity diagnosis (all VIF < 5). Nevertheless, downscaling rainfall through kriging introduces some uncertainty that future work could address through multi-scale feature extraction techniques.

Second, the excellent performance metrics (AUC = 0.9525, F1 = 0.8812) were obtained through within-area training and testing (70/30 split). While this demonstrates strong internal consistency within the study region, it does not guarantee transferability to other areas. The regional dependency of machine learning models—discussed in the Introduction—remains a fundamental challenge requiring external validation in geologically similar areas.

Third, despite its strong performance, the current framework does not incorporate direct measurements of soil pore-water pressure or key hydraulic parameters such as field capacity and permanent wilting point. These parameters govern the transient hydrological

https://doi.org/10.3390/land15040579

response of slopes to rainfall and are essential for representing the physical processes that lead to slope failure.

5.5. Future Research Directions

The limitations identified above point toward several important directions for future research:

Integration of dynamic triggering factors: Moving beyond static and averaged factors to incorporate event-based rainfall intensity and duration data would capture the transient processes that actually initiate landslides. Ideally, this would be complemented by real-time or satellite-derived soil moisture information (e.g., from Sentinel-1) to model pore-water pressure dynamics directly. Furthermore, fusing dynamic rainfall data with time-series NDVI data and leveraging the powerful temporal prediction capability of deep learning can realize the upgrade from static susceptibility assessment to dynamic hazard prediction, which requires the systematic collection of landslide point datasets with temporal attributes as the core verification data for model training and optimization.

Incorporation of soil hydraulic parameters: Including field capacity, permanent wilting point, and other soil hydraulic properties would substantially enhance the physical basis of susceptibility assessments, enabling the model to better represent infiltration and saturation processes.

Cross-regional validation and transfer learning: Systematic external validation in neighboring catchments with similar geology is needed to assess model generalizability. Techniques such as transfer learning or domain adaptation—and potentially the meta-learning paradigm mentioned in the Introduction—could enable few-shot adaptation to new areas with limited data.

Multi-scale feature fusion: Developing architectures that can simultaneously process predictors at their native resolutions would better preserve scale-specific information and reduce uncertainty introduced by resampling.

Enhanced interpretability: While attention-weight analysis provided insight into the fusion process, applying techniques such as SHAP or LIME to generate local explanations for individual predictions would further bridge the "black-box" gap and build trust in model outputs for practical applications.

Introduction of vegetation-related characteristic factors: Combined with the regional characteristics of abundant vegetation resources and extensive bamboo distribution in Sangzhi County, the core karst area of the Wuling Mountains, plant-related evaluation factors such as Forest Canopy Height (FCH) and Above-Ground Biomass (AGB) can be introduced into the model in the future. Bamboo has a strong slope-stabilizing effect due to its developed root system and light-plant mass, which is an important vegetation type for landslide prevention and control in the study area. Quantitative research on the root distribution and slope-stabilizing mechanism of bamboo and other typical vegetation will become a key research direction, and the development of relevant research needs to be based on the systematic collection of measured vegetation data.

Addressing these directions would progressively transform static susceptibility maps into more dynamic, physically grounded hazard-forecasting tools, ultimately supporting real-time early-warning systems and adaptive land management in vulnerable karst mountain regions.

6. Conclusions

This study developed and validated a novel landslide-susceptibility assessment framework—the Triple-Source Probabilistic Fusion Bidirectional LSTM (TSPF-BiLSTM)—specifically designed for data-scarce environments. The framework integrates

https://doi.org/10.3390/land15040579

posterior probabilities from three heterogeneous base learners (Random Forest, LightGBM, AdaBoost) through Platt–Sigmoid calibration and a bidirectional LSTM with attention and residual connections for deep nonlinear fusion.

Validated in the karst terrain of Sangzhi County, Hunan Province, using a total of 618 landslide and 618 non-landslide samples (split into training and testing sets), the TSPF-BiLSTM model achieved superior performance across all evaluation metrics: the highest AUC (0.9525 ± 0.0115), the highest F1-score (0.8812 ± 0.0178), and the highest frequency ratio in the very-high susceptibility class (3.97). These results demonstrate that a probability-level fusion framework with appropriate regularization can effectively enhance model stability and generalization even under severe sample-size constraints.

Multi-model feature importance analysis, supported by five independent techniques, consistently identified elevation, annual rainfall, and NDVI as the dominant predisposing factors controlling landslide occurrence in the study area. Frequency-ratio validation further pinpointed the highest-risk intervals: NDVI 0–0.686 (FR = 2.424–3.032), elevation 155–462 m (FR = 2.361), and annual rainfall 1273.6–1301 mm (FR = 1.960). These intervals align well with records from geologically similar karst areas, providing a quantitative basis for cross-regional landslide zoning within the Wuling Mountains karst belt. Notably, NDVI, as a key vegetation factor, also reflects the close correlation between vegetation coverage and landslide susceptibility in the study area, laying the foundation for subsequent in-depth research on the slope-stabilizing effects of typical vegetation such as bamboo.

It is important to distinguish between predisposing factors and triggering mechanisms. Elevation is a static topographic attribute and should not be considered a triggering factor. The primary triggering factor in this region is rainfall, which governs pore-water pressure dynamics and slope saturation. The current study used mean annual rainfall as a spatial proxy, which captures climatic patterns but does not represent event-scale processes. This distinction frames the most important direction for future work: the integration of dynamic triggering factors, particularly event-based rainfall intensity and duration, along with direct or indirect measurements of soil pore-water pressure and key hydraulic parameters such as field capacity and permanent wilting point. Fusing dynamic hydrometeorological data with time-series vegetation-monitoring data and combining deep learning's temporal prediction capability will further improve the model's dynamic hazard prediction ability, for which the collection of time-attributed landslide inventory data is an essential prerequisite.

The TSPF-BiLSTM framework offers a robust modelling paradigm for landslide susceptibility mapping in sample-limited karst mountainous areas. Its outputs can be applied to guide practical mitigation measures in Sangzhi County, including resident relocation planning, terraced farmland optimization, and low-risk land development. Given the abundant vegetation resources and extensive bamboo distribution in the study area, the research results also provide a scientific basis for the selection of slope-stabilizing vegetation and the formulation of ecological control measures for landslides. By simultaneously delineating high-risk zones requiring intervention and low-susceptibility areas suitable for urban expansion and agricultural allocation, this approach effectively bridges disaster risk reduction with regional development planning—offering tangible value for formulating sustainable development strategies in vulnerable mountainous regions.

Author Contributions: Conceptualization, K.Z.; Methodology, K.Z. and J.Z.; Software, K.Z. and Y.K.; Validation, K.Z., Y.K., X.K., J.T., Y.P. and X.C.; Formal analysis, X.K.; Investigation, S.H., C.P., W.L., T.S. and Z.J.; Writing—review & editing, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was made possible through the generous support of several funding sources. The projects that provided financial backing include: (1) National Natural Science Foundation of China Key Project "Formation Mechanism and Subsequent Earthquake Early Warning of the Jash Strong Earthquake Group" (42230307), spanning from 1 January 2023 to 31 December 2027. (2) Tibet University Talent Introduction & Cultivation Project [00061166] "Deep Crustal Structure and Seismic Activity in Western China", 1 January 2023–31 December 2027.

Data Availability Statement: The code and test data supporting the findings of this study are publicly available in the Mendeley Data repository: Zhu, Kangcheng (2026), "Code for the TSPF-BiLSTM model and test data for landslide susceptibility", Mendeley Data, V1, doi: 10.17632/7zhyhm82tf.1.

Acknowledgments: This study was supported by the research project led by Junmeng Zhao from Xizang University. We also extend our sincere gratitude to Hua Wu from Xizang University for his insightful guidance and hardware support throughout this research.

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

References

1. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]
2. Hearn, G.J.; Hart, A.B. Landslide susceptibility mapping: A practitioner's view. *Bull. Eng. Geol. Environ.* **2019**, *78*, 5811–5826. [CrossRef]
3. Polykretis, C.; Faka, A.; Chalkias, C. Exploring the Impact of Analysis Scale on Landslide Susceptibility Modeling: Empirical Assessment in Northern Peloponnese, Greece. *Geosciences* **2018**, *8*, 261. [CrossRef]
4. Yang, S.; Li, D.; Sun, Y.; She, X. Effect of landslide spatial representation and raster resolution on the landslide susceptibility assessment. *Environ. Earth Sci.* **2024**, *83*, 132. [CrossRef]
5. Saro, L.; Woo, J.S.; Kwan-Young, O.; Moung-Jin, L. The spatial prediction of landslide susceptibility applying artificial neural network and logistic regression models: A case study of Inje, Korea. *Open Geosci.* **2016**, *8*, 382–392. [CrossRef]
6. Liu, S.; Wang, L.; Zhang, W.; Sun, W.; Fu, J.; Xiao, T.; Dai, Z. A physics-informed data-driven model for landslide susceptibility assessment in the Three Gorges Reservoir area. *Geosci. Front.* **2023**, *14*, 101621. [CrossRef]
7. Wang, J.; Jaboyedoff, M.; Chen, G.; Luo, X.; Derron, M.-H.; Hu, Q.; Fei, L.; Prajapati, G.; Choanji, T.; Luo, S.; et al. Landslide susceptibility prediction and mapping using the LD-BiLSTM model in seismically active mountainous regions. *Landslides* **2024**, *21*, 17–34. [CrossRef]
8. Xia, D.; Tang, H.; Glade, T.; Tang, C.; Wang, Q. KNN-GCN: A Deep Learning Approach for Slope-Unit-Based Landslide Susceptibility Mapping Incorporating Spatial Correlations. *Math. Geosci.* **2024**, *56*, 1011–1039. [CrossRef]
9. Bednarik, M.; Yilmaz, I.; Kralovičová, L. Deterministic approach to assess landslide susceptibility and landslide activity in the Central-Western Region of Slovakia. *Bull. Eng. Geol. Environ.* **2024**, *83*, 327. [CrossRef]
10. Mohamadi, B.; Balz, T.; Younes, A. A Model for Complex Subsidence Causality Interpretation Based on PS-InSAR Cross-Heading Orbits Analysis. *Remote Sens.* **2019**, *11*, 2014. [CrossRef]
11. Zhang, Z.; Sun, J. Regional Landslide Susceptibility Assessment and Model Adaptability Research. *Remote Sens.* **2024**, *16*, 2305. [CrossRef]
12. Oliveira, S.C.; Zêzere, J.L.; Garcia, R.A.C.; Pereira, S.; Vaz, T.; Melo, R. Landslide susceptibility assessment using different rainfall event-based landslide inventories: Advantages and limitations. *Nat. Hazards* **2024**, *120*, 9361–9399. [CrossRef]
13. Mohamadi, B.; Balz, T.; Younes, A. Towards a PS-InSAR Based Prediction Model for Building Collapse: Spatiotemporal Patterns of Vertical Surface Motion in Collapsed Building Areas—Case Study of Alexandria, Egypt. *Remote Sens.* **2020**, *12*, 3307. [CrossRef]
14. Chen, L.; Ding, Y.; Zhu, Q.; Zeng, H.; Zhang, L.; Liu, F. Few-Shot Prediction of Landslide Susceptibility Based on Meta-Learning Paradigm. *Geomat. Inf. Sci. Wuhan Univ.* **2024**, *49*, 1367–1376. [CrossRef]
15. Hospedales, T.M.; Antoniou, A.; Micaelli, P.; Storkey, A.J. Meta-Learning in Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 5149–5169. [CrossRef]
16. Guo, H.; Ng, C.W.W.; Zhang, Q. Three-dimensional numerical analysis of plant-soil hydraulic interactions on pore water pressure of vegetated slope under different rainfall patterns. *J. Rock Mech. Geotech. Eng.* **2024**, *16*, 3696–3706. [CrossRef]
17. Ju, X.; Li, J.; Sun, C.; Li, B. Landslide Susceptibility Assessment Using a CNN–BiLSTM-AM Model. *Sustainability* **2024**, *16*, 9476. [CrossRef]
18. Zhang, X.; Zhu, Y.; Deng, L.; Guo, J.; Wu, J.; Jia, H. Design of a Landslide Prediction Model Based on Dual Attention and Dual LSTM Convolutional Neural Network. *CEUR Workshop Proc.* **2023**, *3344*, 206–216. Available online: https://ceur-ws.org/Vol-3344/paper30.pdf (accessed on 20 February 2026).

19. Peng, S. *1-km Monthly Precipitation Dataset for China (1901–2024)*; National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2020. [CrossRef]
20. Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [CrossRef]
21. Yang, J.; Dong, J.; Xiao, X.; Dai, J.; Wu, C.; Xia, J.; Zhao, G.; Zhao, M.; Li, Z.; Zhang, Y.; et al. Divergent shifts in peak photosynthesis timing of temperate and alpine grasslands in China. *Remote Sens. Environ.* **2019**, *233*, 111395. [CrossRef]
22. Yang, J.; Huang, X. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. *Earth Syst. Sci. Data* **2021**, *13*, 3907–3925. [CrossRef]
23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
24. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 3149–3157.
25. Freund, Y.; Schapire, R.E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
26. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*; Neural Information Processing Systems Foundation: Long Beach, CA, USA, 2017; pp. 5998–6008.
28. Chen, W.; Peng, J.; Hong, H.; Shahabi, H.; Pradhan, B.; Liu, J.; Zhu, A.-X.; Pei, X.; Duan, Z. Landslide susceptibility modelling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. *Sci. Total Environ.* **2018**, *626*, 1121–1135. [CrossRef]
29. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [CrossRef]
30. Richardson, E.; Trevizani, R.; Greenbaum, J.A.; Carter, H.; Nielsen, M.; Peters, B. The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns* **2024**, *5*, 100994. [CrossRef]
31. Wang, J.; Zhu, H.-H.; Zhang, W.; Tan, D.-Y.; Pasuto, A. Enhancing Landslide Displacement Prediction Using a Spatio-Temporal Deep Learning Model with Interpretable Features. *J. Geophys. Res. Mach. Learn. Comput.* **2025**, *2*, e2025JH000592. [CrossRef]
32. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
33. Branco, P.; Torgo, L.; Ribeiro, R.P. A survey of predictive modeling under imbalanced distributions. *ACM Comput. Surv.* **2016**, *49*, 1–50. [CrossRef]
34. López, V.; Fernández, A.; García, S.; Palade, V.; Herrera, F. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Inf. Sci.* **2013**, *250*, 113–141. [CrossRef]
35. Farhadpour, S.; Warner, T.A.; Maxwell, A.E. Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sens.* **2024**, *16*, 533. [CrossRef]
36. Kumar, R.P.; Bharathi Mohan, G.; Rohith, N.D.; Mukesh, K.; Harish, B.; Kakarla, Y. Identification of Landslides from Remote Sensing Satellite Imagery Using YOLOv7. In Proceedings of the 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 6–8 July 2023; pp. 1–5. [CrossRef]
37. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [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.

land

MDPI

Article

Beyond Sprawl: How Urban Morphology Shapes Carbon Emission Intensity Categories via SHAP-PDP Framework

Yingkai Tang [1], Wangping Liu [1], Xi Yao [2], Liangzhao Chen [3] and Min Li [3,*]

[1] School of Business, Sichuan University, Chengdu 610065, China; tang@scu.edu.cn (Y.T.); liuwangping@stu.scu.edu.cn (W.L.)
[2] School of Accountancy, Xinjiang University of Finance & Economics, Urumqi 830012, China; yx@xjufe.edu.cn
[3] School of Public Administration, Sichuan University, Chengdu 610065, China; 2024325010019@stu.scu.edu.cn
* Correspondence: limingg@scu.edu.cn

Abstract

Aligning urban morphology with carbon emission intensity categories is essential for advancing sustainable urban development and achieving dual carbon objectives. This study utilizes data from 336 Chinese cities across 2010, 2015, and 2020 to construct multi-dimensional morphological indicators. Spectral clustering categorizes cities into four distinct classes: high-emission intensity, medium-emission ecological, medium-emission developmental, and low-emission. An integrated gradient boosting framework, combined with SHAP and PDP interpretability tools, identifies key morphological drivers and their nonlinear contributions to class assignments. Results demonstrate that morphological features exert nonlinear and threshold-dependent effects on carbon emission intensity category assignments, exhibiting substantial spatial heterogeneity across urban clusters. Core drivers, such as economic density and the landscape shape index, follow distinctly different decision pathways in each category. Furthermore, morphological factors produce non-additive interactive effects that generate region-specific shifts in classification probability. Through this classification-oriented approach, the study provides policymakers with a systematic and readily interpretable reference to inform the formulation of context-specific low-carbon spatial planning strategies.

Keywords: urban morphology; carbon emission intensity; interpretable machine learning; nonlinear effects; spectral clustering

1. Introduction

Driven by the dual imperatives of the global climate crisis and the sustainable development agenda, the carbon reduction performance of cities, as key sources of emissions, has become a focal point of international attention. Cities serve as engines of economic growth, contributing over 60% of global greenhouse gas emissions and consuming more than 78% of energy. By 2050, the global urban population is projected to increase by 2.5 billion people, with nearly 90% of this growth concentrated in Asia and Africa. As one of the countries experiencing the fastest urbanization globally, China's dramatic evolution of urban spatial structures has driven economic development while simultaneously creating severe challenges including ecological space compression, intensified resource consumption, and rising carbon emissions [1,2]. Traditional urban development models, whether characterized by uncontrolled sprawl or a simplistic pursuit of compactness, struggle to adequately address the complex challenges of climate change and carbon reduction targets [3,4]. Therefore, moving beyond binary qualitative judgments of urban morphology as

'good' or 'bad' to deeply explore the complex relationships between its multi-dimensional characteristics and carbon emission intensity holds critical theoretical and practical urgency for formulating precise, efficient, and spatially adaptive low-carbon urban planning and management policies.

Urban morphology, referring to the structure, pattern, and three-dimensional characteristics of urban physical spaces, is widely recognized as a key adjustable variable influencing urban carbon metabolism processes [5–7]. Existing research has accumulated rich findings from multiple dimensions. At the theoretical framework level, scholars have systematically elaborated on how urban morphology affects urban thermal environments, air quality, and carbon metabolism by altering surface energy balance, material flows, and human activity patterns [8–10]. Sharifi (2019), from the perspective of resilient cities, demonstrated that macro-scale morphological characteristics (such as polycentricity and connectivity) may indirectly improve environmental performance by enhancing system adaptability and resilience [11]. In terms of empirical research, substantial work has focused on revealing associations between specific morphological indicators and individual environmental elements. For instance, the positive correlation between building density, floor area ratio, and urban heat island intensity (UHI) has been confirmed by multiple studies [12,13], while green space ratio and blue–green space connectivity have been shown to play positive roles in mitigating heat island effects and improving air quality [14–16]. With advances in remote sensing and geographic information system (GIS) technologies, research on the environmental effects of urban three-dimensional morphology, such as building height, sky view factor, and surface roughness, has become increasingly sophisticated, revealing the important role of vertical dimensions in regulating wind environments and solar radiation reception [17,18].

Despite these fruitful achievements, the existing literature still contains several gaps that require deepening and integration when explaining the complex geographical relationship between urban morphology and carbon emission intensity. First, most studies rely on traditional econometric models such as linear regression and correlation analysis, which assume stable, homogeneous linear relationships between variables [19]. However, urban systems, as typical complex adaptive systems, likely exhibit nonlinear, threshold, and saturation effects in the interactions among their internal elements [20]. For example, the ecological benefits of green space areas may only become significant after reaching a certain threshold, while increased building density may reduce per capita energy consumption through intensive effects in the initial stage but produce negative effects beyond a critical point due to obstructed ventilation and aggravated heat accumulation [21,22]. Such nonlinear mechanisms have not been adequately revealed and quantified in the existing research dominated by linear thinking. Second, existing studies often focus on single or several morphological indicators, lacking a comprehensive analytical framework that can systematically integrate multi-dimensional characteristics of urban development scale, spatial pattern, and socio-economic density. Urban ecological environments result from multi-dimensional interactions between natural substrates and artificial built environments, and fragmented analyses may overlook key interactive responses and synergistic effects [23].

Third, conventional urban classification schemes frequently rely on administrative boundaries, economic aggregates, or demographic thresholds, which often obscure intrinsic spatial configurations and their distinct associations with emission patterns [24]. Such predefined categorization methods may introduce arbitrary boundaries and overlook latent morphological heterogeneity within broadly defined city groups. To address this limitation, this study adopts a data-driven clustering approach grounded in multi-dimensional morphological indicators. Unlike traditional typologies, this methodology allows urban spatial categories to emerge empirically from observed physical configurations, thereby

https://doi.org/10.3390/land15050738

establishing a more coherent alignment between spatial form and carbon emission intensity category assignment. The application of spectral clustering further accommodates complex feature relationships and yields balanced category distributions, providing a structured foundation for subsequent classification modeling. This morphology-driven framework offers a complementary perspective to existing urban typologies, facilitating differentiated spatial governance strategies tailored to distinct form–emission trajectories.

Fourth, while spatial form constitutes a critical adjustable variable, carbon emission intensity is concurrently shaped by socio-economic, technological, and institutional factors. Economic structure, particularly industrial composition and regional energy intensity, remains a primary determinant of emission trajectories [25]. Technological advancement and energy efficiency improvements consistently demonstrate mitigating associations with carbon output [26]. Demographic characteristics, including population density and urbanization levels, influence baseline energy demands and transportation networks [27]. Furthermore, regional climate conditions and policy interventions introduce contextual variability in urban emission performance [28]. Acknowledging these multi-dimensional drivers is necessary for isolating the specific contribution of urban morphology. By incorporating these factors as control variables in the analytical framework, the study can more systematically attribute variations in emission category assignment to spatial form rather than confounding socio-economic or technological shifts. Finally, methodologically traditional models have limited capabilities in handling high dimensionality, collinearity, and complex interactions, making it difficult to accurately capture and explain nonlinear and non-additive relationship structures among variables. Although some scholars have begun applying Geodetector to detect spatial differentiation [29–31], or using machine learning methods such as Random Forest for prediction [32], how to deeply integrate powerful machine learning predictive capabilities with result interpretability to clearly elucidate the independent contributions, nonlinear marginal effects, and interaction mechanisms of various morphological factors on carbon emission intensity remains a cutting-edge challenge facing current research.

To systematically address these challenges, this study employs data from 336 prefecture-level cities (excluding Taiwan Province, Hong Kong SAR, and Macao SAR) in China for 2010, 2015, and 2020, yielding a total of 1008 samples. Based on this dataset, the study aims to construct an analytical system integrating multi-dimensional urban landscape pattern (urban morphology) indicators and adopt an innovative dual-interpretability machine learning framework to reveal their nonlinear influence mechanisms on carbon emission intensity. Specifically, the study will first construct a comprehensive morphological indicator system covering multi-dimensional characteristics. Subsequently, spectral clustering will be employed to classify national cities into different types based on morphological characteristics to identify their underlying heterogeneity. On this basis, this study will construct and optimize a gradient boosting model, combined with a SHAP (Shapley Additive exPlanations) and partial dependence plot (PDP) interpretability framework [33]. While SHAP quantitatively decomposes the global contribution and importance ranking of each morphological feature, PDPs are explicitly utilized to visualize the average marginal effects and nonlinear response trajectories of these features on carbon emission intensity. This dual approach addresses the interpretability challenges of traditional 'black-box' models: it maintains robust predictive performance while enabling the identification of critical thresholds and variations in influence patterns. This offers a detailed perspective on how specific morphological changes relate to carbon dynamics, extending beyond conventional feature ranking.

The potential contributions of this study lie in the following three aspects:

https://doi.org/10.3390/land15050738

First, at the theoretical level, this study extends existing understanding by challenging the assumption that simplifies the relationship between urban morphology and carbon emission intensity into linear connections. By employing a classified modeling perspective, it empirically characterizes the prevalent nonlinear responses and context-dependent threshold effects between the two. This perspective shifts the focus from mere prediction of continuous values to identifying key morphological pathways associated with carbon emission intensity variations, offering deeper insights into the complex nonlinear dynamics in urban systems.

Second, at the methodological level, this study addresses the trade-off between predictive performance and interpretability in continuous variable regression modeling. By integrating the gradient boosting tree model with the SHAP-PDP interpretability framework, a robust analytical framework is constructed. This method enhances model generalization while quantifying the differentiated contributions of various morphological features to carbon emission intensity levels and their nonlinear influence patterns, providing a valuable methodological reference for related research.

Third, at the practical level, based on the above theory and methods, the study conducts empirical analysis on national urban samples. The research conclusions offer evidence to identify key morphological factors and their potential threshold ranges that could be considered for regulation to support low-carbon goals in different types of cities. This provides scientific evidence for national and local governments to implement typified, differentiated territorial spatial planning and targeted low-carbon urban management.

2. Study Area and Data Sources

2.1. Study Area

China was selected as the research subject primarily based on its typicality and global reference significance. China is undergoing the largest and fastest urbanization process in world history, encompassing various urban types from high-density megacities to numerous small- and medium-sized cities at different development stages, forming a highly representative 'natural experimental field' in the global urban development spectrum. This extreme spatial heterogeneity provides a rich sample foundation for systematically exploring the complex nonlinear relationship between urban morphology and carbon emission intensity. Compared with Western cities that have completed urbanization, Chinese cities exhibit greater uniqueness and complexity, as spatial morphological evolution and carbon reduction pressures coexist and interact during rapid industrialization and urbanization [34]. This makes the influence mechanisms identified in the Chinese context not only capable of providing scientific evidence for promoting ecological civilization construction and precise spatial planning under the 'dual carbon' goals but also offering important empirical experience and strategic insights for other developing countries currently in rapid urbanization, with certain international promotion value.

To ensure the completeness and comparability of the research data, this study excluded cities with severe missing data during sample selection, ultimately identifying 336 cities as the study area. Geographically, China exhibits pronounced altimetry and topographical heterogeneity, characterized by a three-step descending terrain gradient. The western region is dominated by high-altitude plateaus and mountainous systems, the central zone comprises transitional basins and hills, and the eastern region features low-lying plains and coastal belts. This complex altimetry fundamentally constrains urban spatial expansion, infrastructure layout, and land-use continuity, thereby generating a highly varied morphological landscape that serves as the critical geographical foundation for this study.

https://doi.org/10.3390/land15050738

2.2. Data Sources and Preprocessing

2.2.1. Data Sources

This study integrated spatial and statistical data from multiple sources. Land use data (with a resolution of 1 km) were obtained from the Resource and Environment Data Center of the Chinese Academy of Sciences (https://www.resdc.cn/, accessed on 1 December 2023). Urban landscape pattern metrics were calculated using Fragstats 4.2 software. Population and economic output data were derived from the China City Statistical Yearbook (https://data.cnki.net/, accessed on 1 January 2022). Carbon emissions data were from the China City Greenhouse Gas Working Group (https://www.cityghg.com/, accessed on 1 January 2024). This dataset provides detailed carbon emissions data for various Chinese cities and has been widely used in previous studies. Given that the dataset is currently only updated to 2020, and to ensure consistency across all data sources, the study period is limited to 2010, 2015, and 2020, yielding a total of 1008 samples. All indicators were standardized to ensure comparability in both statistical definition and spatial coverage across years, forming a consistent panel dataset that supports the subsequent analysis of nonlinear relationships between urban form and carbon emission intensity.

2.2.2. Urban Morphological Calculations

The characteristic variable system constructs a multi-dimensional urban morphological indicator framework, mainly including the following three categories. The first category is scale characteristics, with the core indicator being the proportion of built-up area, used for the basic measurement of urban land development intensity and spatial expansion, which has been widely applied to characterize urbanization levels and their ecological environmental effects [35]. The second category is spatial morphological characteristics, drawing on mature landscape ecology principles and index systems [36] which aim to quantify the pattern attributes of urban two-dimensional space. Among them, patch density (PD) and mean patch area (AREA_MN) jointly reflect the degree of spatial fragmentation and average patch scale. Largest patch index (LPI) represents the dominance of dominant patches. Landscape shape index (LSI) and mean shape index (SHAPE_MN) describe the complexity and irregularity of patch outlines. Edge density (ED) reflects boundary length per unit area. Aggregation index (AI), clumpiness (CLUMPY), cohesion index (COHESION), and effective mesh size (MESH) measure spatial aggregation states and landscape connectivity from different perspectives. Split index (SPLIT) reflects the degree of landscape division. Indicators in the spatial morphological characteristics dimension provide standardized analytical tools for understanding the impact of spatial configuration on ecological processes [37]. The third category is density characteristics, including people density and economic density, calculated respectively as the ratio of urban population and secondary and tertiary industry added value to urban area. These two indicators are key variables characterizing the spatial agglomeration intensity of socio-economic activities, crucial for understanding the impact of human activities on resources and the environment [38,39].

Given the high dimensionality and potential multicollinearity among the initial landscape metrics, we conducted a systematic variable screening process to reduce redundancy prior to modeling. Pairwise correlation analysis was first performed, and variables with correlation coefficients exceeding 0.7 were flagged. Subsequently, a variance inflation factor (VIF) test was applied, and indicators with VIF > 10 were sequentially excluded to ensure statistical independence. This screening process reduced the initial 18 morphological indicators to 14 core variables (listed in Table 1), which adequately capture the multi-dimensional attributes of urban spatial configuration while strictly controlling for multicollinearity. These selected indicators were then standardized and incorporated into subsequent analy-

https://doi.org/10.3390/land15050738

ses. Variable definitions are shown in Table 1, and the descriptive statistics are provided in Table A1 of Appendix A.

Table 1. Urban morphological indicators.

Category	Metric Name	Definition
Scale	Proportion of Built-Up Area	Urban land proportion
Shape	PD	Urban patch density
	LPI	Largest urban patch index
	ED	Urban edge density
	LSI	Urban landscape shape index
	AREA_MN	Mean urban patch area
	SHAPE_MN	Mean urban patch shape index
	CLUMPY	Urban clumpiness index
	COHESION	Urban patch cohesion index
	MESH	Urban effective mesh size
	SPLIT	Urban splitting index
	AI	Urban aggregation index
Density	People Density	Ratio of the urban population to the urban area size
	Economic Density	Ratio of the total output value of the secondary and tertiary industries to the urban area size

2.2.3. Carbon Emission Intensity Assessment

The core dependent variable of this study is carbon emission intensity, quantified using carbon dioxide emissions per unit of gross domestic product (GDP). This indicator is a key comprehensive measure for assessing the carbon production efficiency and low-carbon development level of regional or urban economic systems [40]. Its core value lies in combining the scale effect of economic activities with the absolute level of carbon emissions, directly revealing the coupling relationship between economic growth and carbon emission pressure, and thus being widely used in international comparisons and policy evaluations [41].

Conceptually, this indicator relates closely to the idea of decoupling in sustainability research. It measures the extent to which economic growth can be separated from resource use and environmental impact, with a specific focus on carbon emissions [42,43]. A decline in carbon emission intensity signals improvements in economic structure, production technology, or management practices that enable greater economic output with lower carbon inputs [44]. Compared to total emissions, this intensity measure offers a more equitable basis for comparing cities at different stages of development, as it controls for differences in economic scale [45]. It is also more directly linked to the efficiency and pattern of economic development than per capita emissions. For these reasons, carbon emission intensity has become a widely used benchmark for evaluating whether urban development pathways are consistent with low-carbon and sustainability goals.

For predictive modeling, the continuous emission intensity was operationalized as a binary classification (high- vs. low-carbon categories). This transformation is methodologically grounded in three considerations. First, urban carbon intensity typically exhibits a pronounced right-skewed distribution with extreme high-value outliers, which can destabilize continuous regression and exacerbate overfitting. Second, in the absence of universally mandated policy thresholds, the sample median provides a statistically neutral benchmark to objectively differentiate relatively efficient from inefficient urban systems. Third, median-based dichotomization yields a balanced class distribution, which enhances machine learning robustness by mitigating majority-class bias and ensuring consistent sensitivity in detecting morphological drivers across both categories. This approach preserves

https://doi.org/10.3390/land15050738

essential distributional characteristics while aligning the analytical framework with the study's comparative typology objectives.

3. Methodology

3.1. Study Framework

This study adopts a systematic machine learning analysis framework to explore the impact of urban morphological indicators on carbon emission intensity (shown in Figure 1). First, an urban morphological indicator system including landscape patterns and socio-economic attributes is constructed, and carbon emission intensity indicators are calculated. Second, variance inflation factor detection is used to remove highly collinear indicators, retaining representative indicators as input for spectral clustering analysis to classify research samples into different types. Third, six machine learning algorithms, including gradient boosting, Random Forest, Multi-Layer Perceptron (MLP), eXtreme Gradient Boosting (XGBoost), Logistic Regression, and Support Vector Machine (SVM), are compared, and the model with optimal predictive performance is selected as the baseline model. Subsequently, the Bayesian optimization algorithm combined with an early stopping strategy is employed to train and optimize the model. To comprehensively interpret the results, this research integrated SHAP values for evaluating global feature importance and interaction effects with partial dependence plots (PDPs) to visualize average marginal effects and nonlinear response trajectories. While SHAP quantifies contribution magnitudes, PDPs specifically pinpoint critical threshold levels and direction shifts in the influence mechanisms. Overall, this synergistic framework effectively characterizes the complex nonlinear responses, precise thresholds, and interaction directions of urban morphological indicators on carbon emission intensity.

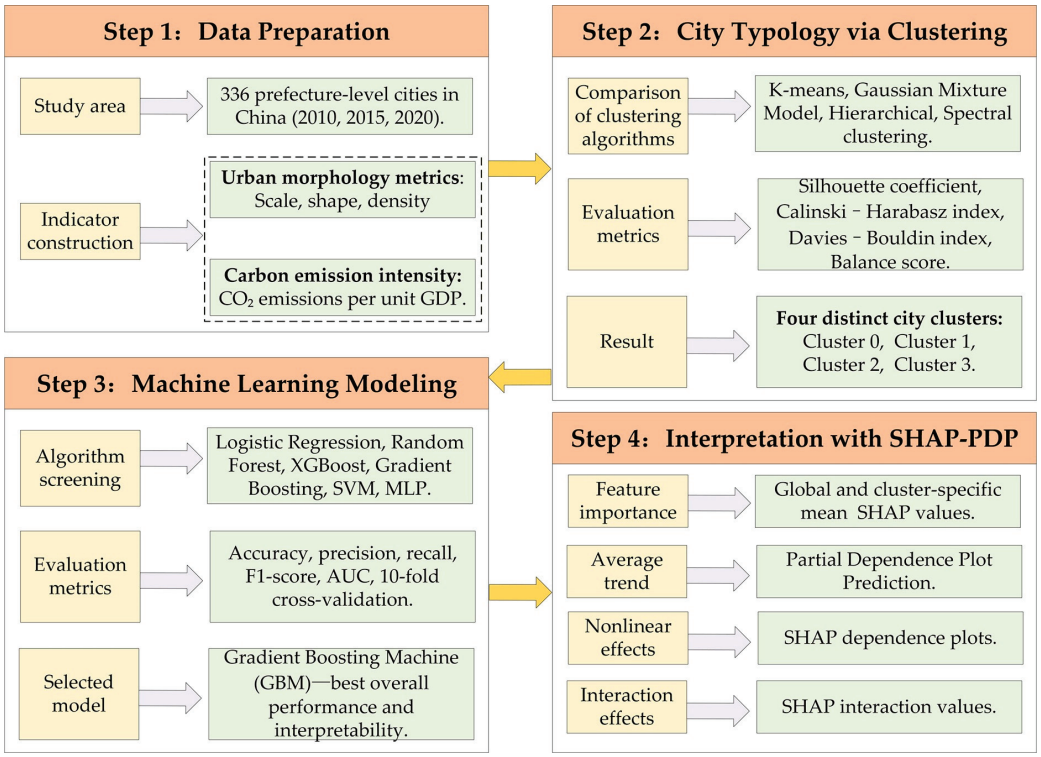


Figure 1. Research flowchart.

3.2. Overview of Methods

3.2.1. Clustering Method Selection

To identify urban groups with similar morphological characteristics, this study employs a comprehensive clustering analysis framework. Given the relative stability of urban morphological traits over the observation period, three-year observational data are aggregated into representative city-level feature values to capture typical spatial patterns.

In terms of algorithm selection, this study systematically compares four classical methods: K-means clustering, Gaussian mixture model, hierarchical clustering, and spectral clustering. Spectral clustering is a graph theory-based algorithm that constructs a similarity graph from data points and utilizes eigenvalues and eigenvectors for dimensionality reduction and partitioning, making it highly effective for nonlinear and non-convex data distributions. In geospatial analysis, where data frequently exhibit spatial autocorrelation and heterogeneous distributions, this method constructs spatial weight matrices based on distance or topological relationships to effectively capture both local similarities and global structural dependencies among geographic units. Unlike traditional distance-based approaches, it is better suited for delineating irregular urban functional patterns and complex landscape configurations. Given the inherent nonlinear characteristics and potential manifold structures in urban morphological data, spectral clustering is selected as the primary analytical tool. Its specific advantages include the capacity to model nonlinear relationships, optimize graph partitioning to maximize intra-cluster similarity while minimizing inter-cluster dissimilarity, and produce spatially coherent results that align well with visual validation. All algorithms are evaluated using identical standardized feature sets and assessed through silhouette coefficients, Calinski–Harabasz indices, Davies–Bouldin indices, and clustering balance metrics. Specifically, the silhouette coefficient quantifies intra-cluster cohesion relative to inter-cluster separation, with values approaching 1 indicating well-defined and distinct partitions. The Calinski–Harabasz (CH) index calculates the ratio of between-cluster dispersion to within-cluster dispersion, where higher values denote more compact and well-separated groupings. Conversely, the Davies–Bouldin (DB) index measures the average similarity between each cluster and its most similar counterpart based on centroid distances and scatter; lower values reflect superior cluster separation and minimal overlap. By jointly applying these complementary indicators, this study ensures a robust, multi-dimensional assessment of cluster validity to determine the optimal partition. A composite scoring mechanism ensures an optimal balance between structural clarity and interpretability, ultimately confirming spectral clustering as the most robust approach.

Furthermore, this study adopts a two-step analytical strategy—first grouping cities by morphological similarity, then estimating regression models within each cluster—rather than directly modeling spatial heterogeneity through continuous local regression or hierarchical techniques. This framework is selected for three primary reasons. First, urban morphological evolution often exhibits discrete structural regimes rather than smooth spatial gradients; clustering explicitly identifies these mechanistic breakpoints, whereas continuous heterogeneity models may obscure distinct pathway variations. Second, by isolating cities with comparable spatial configurations, the within-cluster regression significantly reduces multicollinearity among landscape metrics and minimizes unobserved confounding, thereby yielding more stable and interpretable coefficient estimates. Third, the resulting typology-based framework directly aligns with urban planning practices, enabling targeted policy interventions tailored to specific morphological categories. While direct heterogeneity modeling remains valuable for capturing localized variations, the proposed two-step strategy offers superior clarity in mechanism identification, enhanced statistical robustness within homogeneous subgroups, and stronger practical relevance for differentiated urban governance.

https://doi.org/10.3390/land15050738

To determine the optimal number of clusters, a comprehensive evaluation system is established. The trends of multiple clustering quality indicators are analyzed for cluster counts ranging from 2 to 10. By identifying inflection points and stable regions within these trends and integrating practical research requirements, the optimal cluster count is selected to ensure both statistical significance and clear interpretability.

To ensure the reliability of the clustering outcomes, a systematic robustness testing framework is implemented. Pattern stability is evaluated across multiple dimensions, including consistency checks under varying data aggregation schemes, comparative performance analyses of alternative algorithms, and sensitivity tests across different cluster count settings. This multi-level validation ensures that the results are not dependent on specific methods or arbitrary parameter choices, thereby strengthening the scientific rigor and reliability of the findings and providing a solid foundation for subsequent heterogeneity analysis.

3.2.2. Machine Learning Model

This study adopts a systematic machine learning analysis framework to explore the influencing mechanisms of urban morphological indicators on carbon emission intensity. The model construction and optimization process includes four key stages: data division, model comparison, hyperparameter optimization, and model interpretation.

(1) Data Division and Preprocessing

First, the complete dataset is randomly divided into training and testing sets at a 7:3 ratio to ensure independence of model evaluation. The training set is used for model training and parameter tuning, while the testing set is used for final model performance verification. To address class imbalance issues, Synthetic Minority Oversampling Technique is applied to the training set to improve the model's recognition ability for minority class samples.

(2) Model Screening and Comparison

To determine the most suitable model for predicting carbon emission intensity, this study systematically compares six mainstream machine learning algorithms, including Logistic Regression, Random Forest, XGBoost, gradient boosting, Support Vector Machine, and Multi-Layer Perceptron. This multi-model comparison strategy helps avoid model bias that may arise from relying on a single algorithm and is a standard practice for ensuring the robustness of analysis conclusions [46]. Model performance evaluation adopts multi-dimensional indicators, including accuracy, precision, recall, F1 score, AUC-ROC curve area, and five-fold cross-validation accuracy. The basis for selecting this comprehensive indicator system is that classification task evaluation needs to simultaneously consider the model's overall correctness, discrimination ability for specific categories, and ranking quality. Single indicators may be misleading, while multi-dimensional evaluation can provide a more comprehensive performance profile [47]. Among them, cross-validation is a key technique for evaluating model generalization capability and reducing overfitting [48]. By comprehensively comparing evaluation results of various models, the model with optimal performance across multiple indicators is selected as the baseline model.

(3) Hyperparameter Optimization

For the selected baseline model, Bayesian optimization algorithm is used for fine-tuning of its key hyperparameters. The optimization process is based on tree-structured Parzen estimators, evaluating performance of different parameter combinations through five-fold cross-validation. To prevent overfitting, 15% of training data is reserved as early stopping validation set, automatically terminating training when model performance no longer improves. The optimization objective is to maximize performance indicators on the validation set, ultimately obtaining the optimal hyperparameter combination and best iteration rounds.

https://doi.org/10.3390/land15050738

(4) Model Ensemble and Interpretation

Based on selecting gradient boosting model as the optimal model, to further improve prediction stability and accuracy, this study simultaneously attempts two strategies: voting ensemble and stacking ensemble. Voting ensemble combines prediction results of multiple base learners through weighting, while stacking ensemble uses outputs of base learners as inputs for meta-learners for secondary training. Ensemble model evaluation results show that voting ensemble and stacking ensemble have improvements over single models on some indicators, but considering model complexity and computational efficiency, gradient boosting model has better interpretability while maintaining excellent prediction performance. Therefore, after comprehensive consideration of performance, complexity, and interpretability, this study ultimately selects gradient boosting model as the optimal model for subsequent in-depth feature importance and mechanism analysis.

Model interpretability analysis adopts a dual-framework approach combining SHAP and PDPs. While SHAP values, grounded in cooperative game theory, quantify each feature's contribution to identify key influencing factors, PDPs are explicitly employed to visualize the average marginal effects and nonlinear response trajectories of individual indicators on carbon emission intensity. This synergy overcomes the limitations of single-method interpretation—SHAP reveals global importance rankings, whereas PDP precisely delineates critical threshold levels and direction shifts—providing a robust basis for understanding the complex action mechanisms of urban morphological indicators.

4. Results

4.1. City Types Based on Urban Morphology

To accurately identify urban groups with similar morphological characteristics, this study did not pre-set a single clustering algorithm but systematically evaluated the comprehensive performance of four mainstream clustering methods under different data integration strategies. First, to obtain more representative city-level characteristics and smooth annual fluctuations, an integrated dataset based on long-term trends of panel data was constructed. To ensure robustness of analysis conclusions, primary and secondary different data integration strategies were adopted to capture the central tendency of urban morphology, forming two sets of comparable but logically different foundational data. Subsequently, for each dataset, four algorithms including K-means clustering, Gaussian mixture model, agglomerative hierarchical clustering, and spectral clustering were applied. For each method, the optimal number of clusters on corresponding data was automatically determined based on multiple internal validity indicators, including silhouette coefficient, Calinski–Harabasz index, and Davies–Bouldin index. Evaluation not only focused on clustering structure compactness and separation but also specifically introduced balance scores to measure the balance of sample sizes across categories, avoiding invalid categories with too few samples.

Comprehensive evaluation results show significant differences in performance among different clustering methods. As shown in Figure 2, spectral clustering demonstrates the highest comprehensive scores under both primary and secondary integration strategies. Specifically, under the primary integration strategy, spectral clustering has a silhouette coefficient of 0.191, slightly lower than K-means (0.296), but its balance score reaches 0.737, far superior to other methods, indicating that its four categories have the most balanced sample size distribution (as shown in Figure 3a, category sample sizes are 74, 114, 61, and 87). Under the alternative integration strategy, spectral clustering also leads with a balance score of 0.823, ensuring statistical validity of classification (as shown in Figure 3b, category sample sizes are 69, 107, 73, and 87). In contrast, K-means and agglomerative hierarchical clustering produce extremely unbalanced sample sizes among categories in

https://doi.org/10.3390/land15050738

most cases (Figure 3c), with balance scores often approaching zero. Gaussian mixture model performance shows greater volatility across different data foundations (Figure 3d,e).

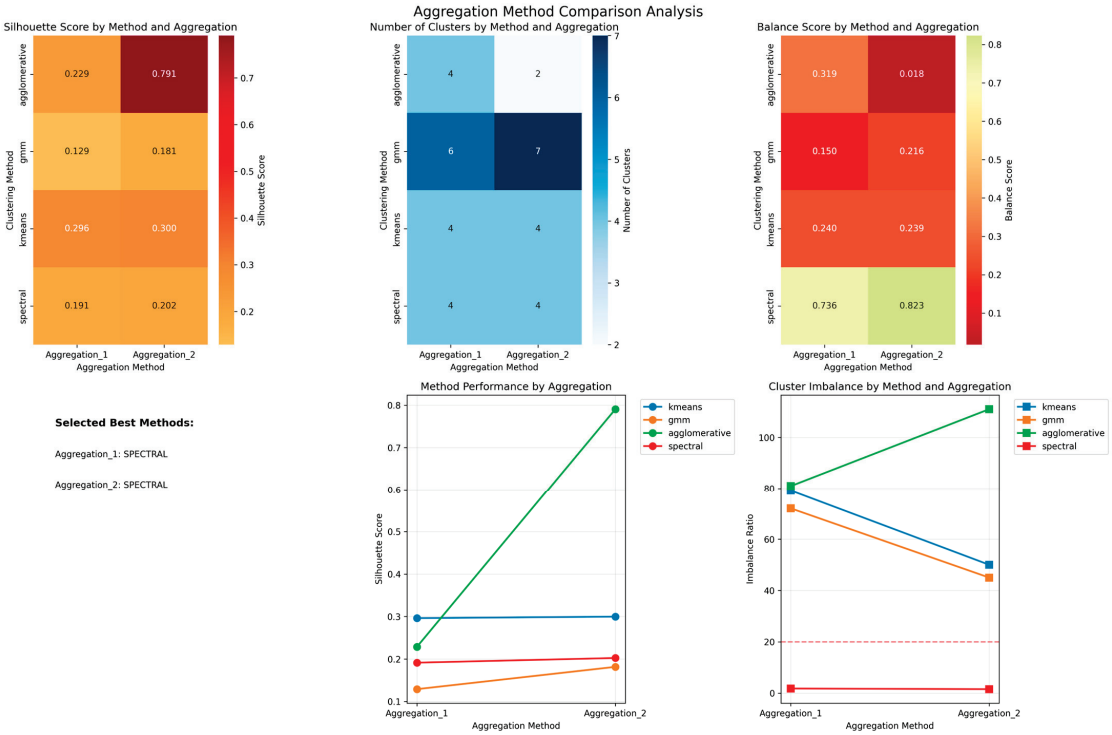


Figure 2. Cluster method performance comparison. **Note:** "kmeans" is the abbreviation of K-means.

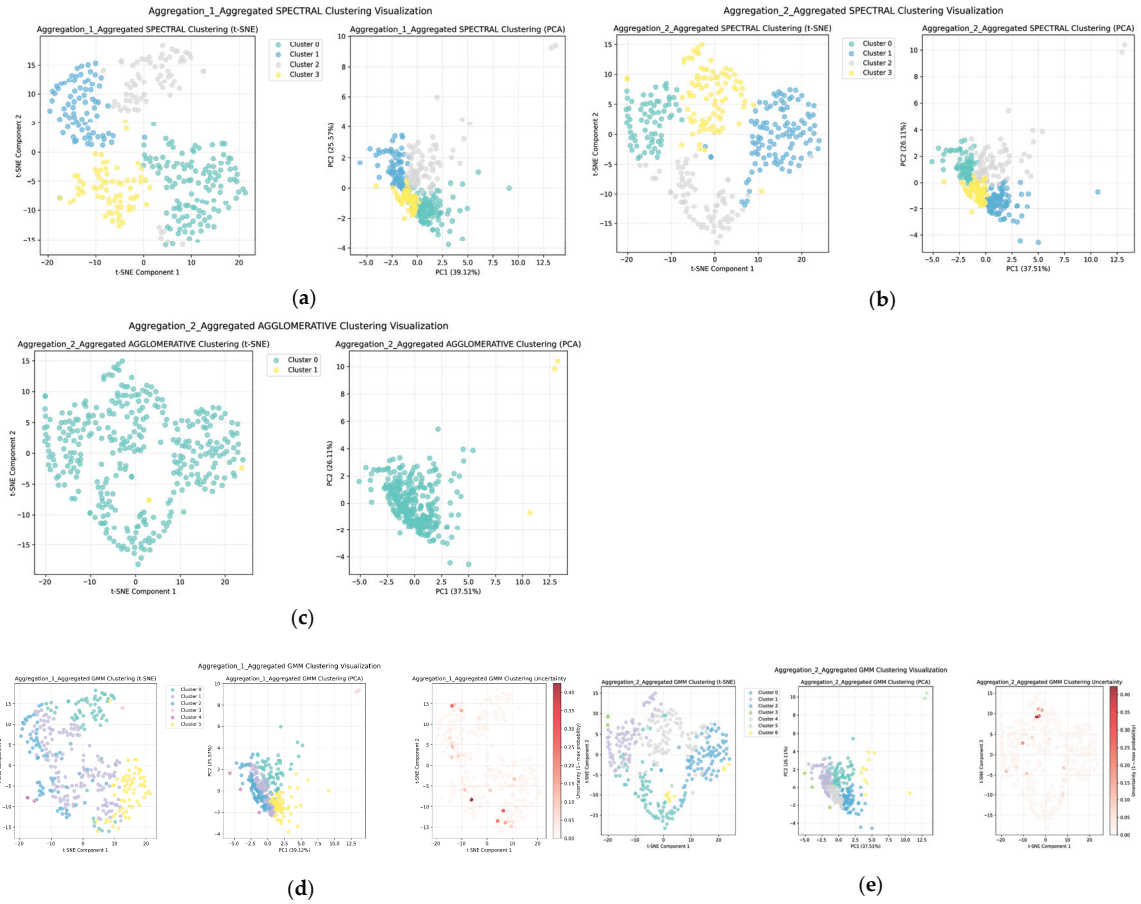


Figure 3. Spatial distribution map of spectral clustering results.

https://doi.org/10.3390/land15050738

Therefore, this study ultimately selects spectral clustering as the optimal method for urban morphological classification. This method performs partitioning based on similarity graphs of data points, excelling at identifying complex, non-spherical data distribution structures. Through t-SNE and PCA dimensionality reduction techniques for visualizing spectral clustering results under different integration strategies, the different urban categories formed show good separation trends in low-dimensional feature space, and the core category structures remain consistent between primary and secondary strategies. This systematic method comparison and sensitivity testing process, comprehensively considering clustering quality, sample balance, and conclusion robustness, ultimately ensuring the scientificity and reliability of urban type classification.

Based on a clustering analysis of urban morphological characteristics, Chinese cities were classified into four distinct groups (Figure 4). By grouping cities with similar morphological features, this approach effectively controls for within-group variation in urban forms while highlighting between-group differences in their relationship with carbon emission intensity. This classification avoids the confounding effects that would arise from analyzing cities with divergent morphological types together, thereby providing a robust basis for examining how urban form influences carbon emission intensity across different spatial development patterns.

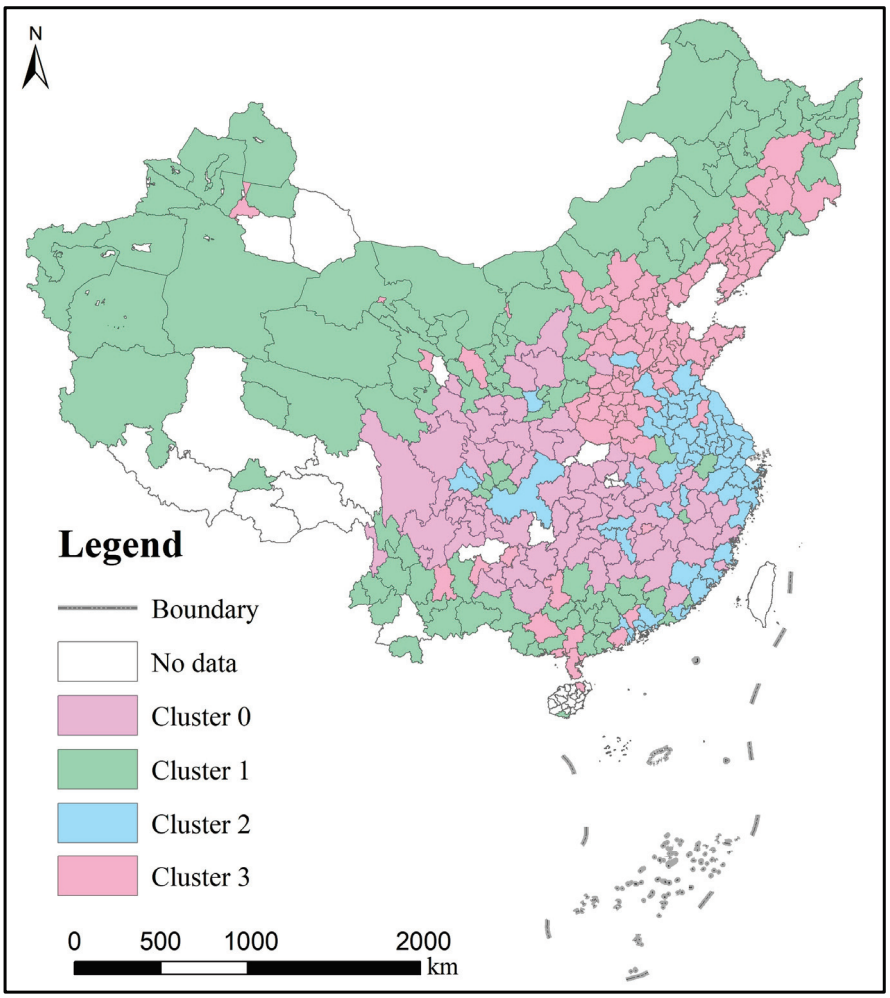


Figure 4. Study area.

The resulting groups exhibit pronounced spatial clustering that aligns with the regional altimetry and topographical gradients outlined in Section 2.1. Cluster 0 (C0) is predominantly distributed across central and Western China, as well as the southern inland

regions, covering Sichuan, Chongqing, Guizhou, Hubei, Hunan, Jiangxi, and Anhui. This is the most spatially extensive cluster. Cities in this group share similar morphological characteristics, shaped largely by the topographical constraints of mountainous and hilly terrain in inland areas, with consistency observed in built-up area structure and density patterns. Cluster 1 (C1) is primarily located along the northern frontiers, the northwest, and the coastal margins of South China, including Inner Mongolia, Xinjiang, Qinghai, Gansu, Heilongjiang, Jilin, Northern Liaoning, Southern Guangdong and Guangxi, and Western Fujian. These cities are generally situated in sparsely populated or ecologically sensitive regions. Morphologically, these areas display low-density and dispersed spatial patterns with limited continuity in built-up zones, a characteristic that aligns with the development trajectories of resource-based and frontier cities. Cluster 2 (C2) is highly concentrated in the eastern coastal region and the core areas of the middle and lower Yangtze River, encompassing Shanghai, Jiangsu, Zhejiang, Eastern Anhui, and Eastern Fujian. This cluster corresponds to China's most economically advanced coastal areas. Morphologically, these cities are characterized by high-density compact urban forms with highly contiguous built-up areas, reflecting the mature stage of urbanization typical of the eastern seaboard. Cluster 3 (C3) is mainly distributed across the North China Plain and the Bohai Rim region, covering Beijing, Tianjin, Hebei, Shandong, Henan, Shanxi, and Southern Liaoning. Cities in this group exhibit medium-to-high-density urban forms with monocentric ring-like expansion patterns. The consistency in morphological metrics across this cluster aligns with the spatial characteristics typical of traditional industrial bases and densely populated agglomerations in Northern China.

4.2. Gradient Boosting Model Fitting Results

Guided by the research objectives and insights from the existing literature, this study adopts a classification modeling framework in the machine learning stage. We acknowledge the valuable contributions of prior studies, such as Huang et al. (2026) [21], Cheng et al. (2026) [49], and Ru et al. (2026) [50], which effectively employed regression-based predictive models to capture continuous variations in urban efficiency. While regression is well-suited for modeling continuous variables, policy formulation often requires clear, qualitative distinctions to enable targeted interventions. Specifically, decision-makers benefit from directly identifying whether a city operates in a low-carbon or high-carbon efficiency state, rather than interpreting abstract continuous scores that require additional thresholding. To address this practical need, this study shifts to a classification framework that directly outputs categorical judgments, thereby providing an end-to-end analytical tool that enhances policy relevance and operational clarity. To further ensure the robustness and generalizability of the classification model, we integrate the Synthetic Minority Oversampling Technique (SMOTE) to address sample imbalance, Bayesian optimization for systematic hyperparameter tuning, and an early stopping strategy to mitigate overfitting risk. These methodological enhancements collectively advance the reliability of the model while aligning closely with the differentiated analytical requirements of urban carbon governance.

To determine the optimal modeling method, this study systematically compares six mainstream machine learning algorithms, including gradient boosting, Random Forest, Multi-Layer Perceptron, eXtreme Gradient Boosting, Logistic Regression, and Support Vector Machine. Model performance is comprehensively evaluated through mean accuracy of ten-fold cross-validation and multiple indicators, including accuracy, precision, recall, F1 score, and AUC. Model performance comparison results are shown in Figure 5a. gradient boosting model performs optimally across multiple indicators, with test set accuracy reaching 64.4%, mean cross-validation accuracy of 60.6%, and AUC value of 0.67, significantly outperforming other models in comprehensive performance. Benefiting from

https://doi.org/10.3390/land15050738

the naturally balanced class distribution (504 low-carbon vs. 504 high-carbon samples), the gradient boosting model demonstrated effective discriminative capability, yielding balanced recall and precision across both categories without the need for synthetic over-sampling. Its performance evaluation confusion matrix, ROC curve, and precision–recall curve are detailed in Figure 5b. In comparison, the Random Forest model has the highest AUC value of 0.689, with performance evaluation shown in Figure 5c; while Support Vector Machine and other models show relatively weaker comprehensive performance, with specific evaluations shown in Figure 5d.

To further improve prediction stability and accuracy, this study explores ensemble learning strategies based on single models, constructing voting ensemble and stacking ensemble models. Ensemble model evaluation results show that stacking ensemble slightly outperforms voting ensemble in accuracy, F1 score, and AUC value, with AUC reaching 0.689 and accuracy of 64% (Figure 5e,f). However, considering that stacking ensemble model structure is more complex with higher computational costs while performance improvement is relatively limited, gradient boosting model maintains high prediction performance while having advantages of model simplicity, strong interpretability, and high computational efficiency. Therefore, after comprehensive weighing of performance, complexity, and interpretability, this study ultimately selects gradient boosting model as the optimal model for subsequent in-depth feature importance and mechanism analysis.

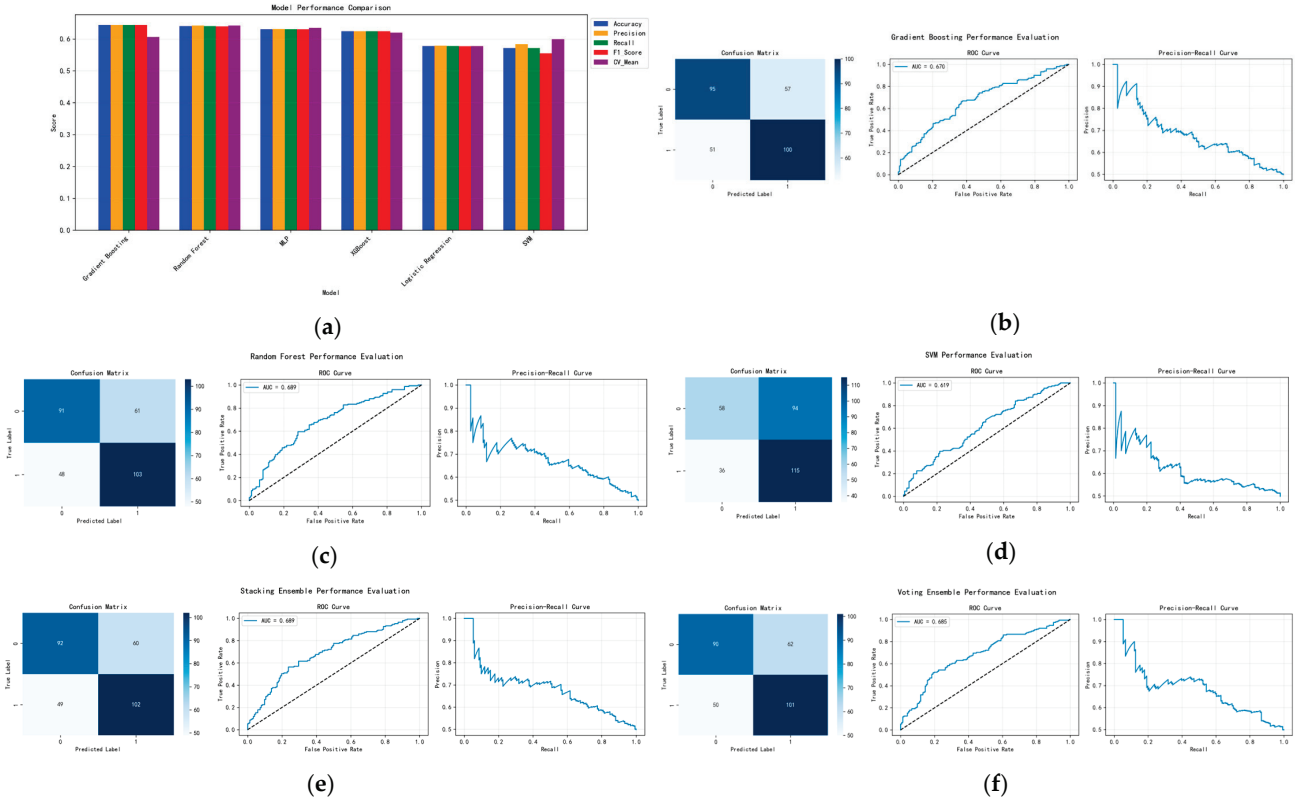


Figure 5. Different machine learning models prediction performance comparison.

4.3. SHAP-PDP Analysis of Driving Factors

In terms of methodological design, this study references the previous work of Huang et al. (2026) but does not follow their adopted OPGD method for feature screening and interaction detection [21]. The main consideration is that OPGD, as a classical statistical method based on spatial differentiation, has analytical logic not entirely consistent with the internal mechanism of subsequent machine learning models, and its discretization process may introduce subjective bias. To ensure overall logical consistency from feature analysis

to model interpretation, this study chooses to directly use the optimized gradient boosting model and complete all feature importance ranking, influence direction determination, and nonlinear relationship characterization using the SHAP framework. This approach not only fully leverages modern advantages of machine learning methods in interpretability but also avoids redundancy and inconsistency that may arise from mixed methods.

To further elucidate the specific response patterns of carbon emission intensity to urban morphological changes, we employed partial dependence plots (PDPs) to visualize the average marginal effects of key features (Figure 6b–i). The PDP results reveal predominantly nonlinear relationships, with clear threshold effects that complement the SHAP-based feature importance rankings. SHAP and partial dependence plots serve as model interpretability tools but differ in calculation logic and interpretive meaning. The PDP method marginalizes other features to reveal the global average effect of a single variable on classification probability. This approach focuses on macro trends and critical thresholds. SHAP relies on cooperative game theory to quantify the local contribution of each feature to individual predictions. It naturally captures interactions between features. These two methods produce complementary rather than contradictory results. They connect macro average trends with micro individual variations. The SHAP distribution does not conflict with the PDP curve. It supplements the PDP findings by displaying the actual dispersion of samples around the average trend.

For proportion of built-up area, ED, and PD, the global average trend shows a monotonic increase in the predicted probability of high carbon emission intensity as these variables rise, indicating that higher urbanization intensity and fragmentation consistently elevate emission risk. In contrast, CLUMPY and SHAPE_MN exhibit decreasing trends, suggesting that greater spatial aggregation and more regular patch shapes are associated with a lower likelihood of high emissions. Several variables—including LSI, LPI, and AI—display complex nonlinear responses with multiple inflection points, where the direction or magnitude of their effect on emission intensity shifts at specific threshold values. These PDP-derived patterns confirm that the influence of urban morphology is not uniform but is highly context-dependent, reinforcing the need for nuanced, threshold-aware policy interventions.

Complementing these trend observations, the SHAP summary plot based on the full sample (Figure 6a) quantifies the relative magnitude of each factor's contribution. The analysis identifies proportion of built-up area as the most critical determinant, possessing the highest mean absolute SHAP value of 0.0200, which underscores its dominant global impact on carbon emission intensity. Following closely are SHAPE_MN (0.0193) and LSI (0.0188), highlighting the substantial role of landscape configuration. Economic density and LPI round out the top five influential characteristics with mean absolute SHAP values of 0.0173 and 0.0135, respectively. Together, the PDP trends and SHAP magnitudes offer a comprehensive view, where built-up proportion is the primary driver in terms of strength, while the configuration of the landscape (shape and density) plays an equally critical role in shaping the trajectory of urban carbon emissions.

C0, as the high-emission intensity urban cluster (Figure 7a), has AREA_MN as its most important characteristic, with a mean absolute SHAP value reaching 0.0311, ranking first among all important cluster characteristics. CLUMPY ranks second with a mean absolute SHAP value of 0.0174, and LSI follows closely with 0.0113. Proportion of built-up area and SPLIT enter the top five important characteristics with mean absolute SHAP values of 0.0102 and 0.0096, respectively.

https://doi.org/10.3390/land15050738

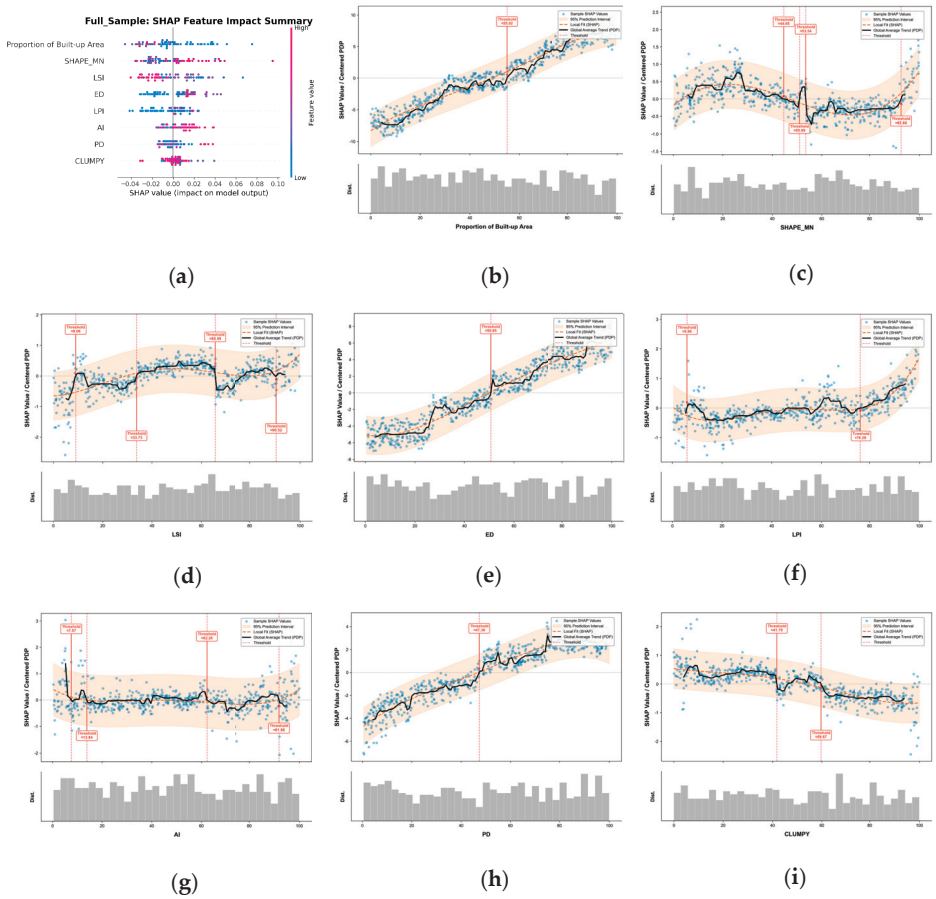


Figure 6. SHAP-based feature importance and PDP-derived marginal effects of urban morphological indicators on carbon emissions intensity (full sample). **Note**: (**a**) Represents the SHAP importance ranking in the full sample; (**b**–**i**) respectively represent the PDP and SHAP composite images of the Proportion of Built-up Area, SHAPE_MN, LSI, ED, LPI, AI, PD and CLUMPY.

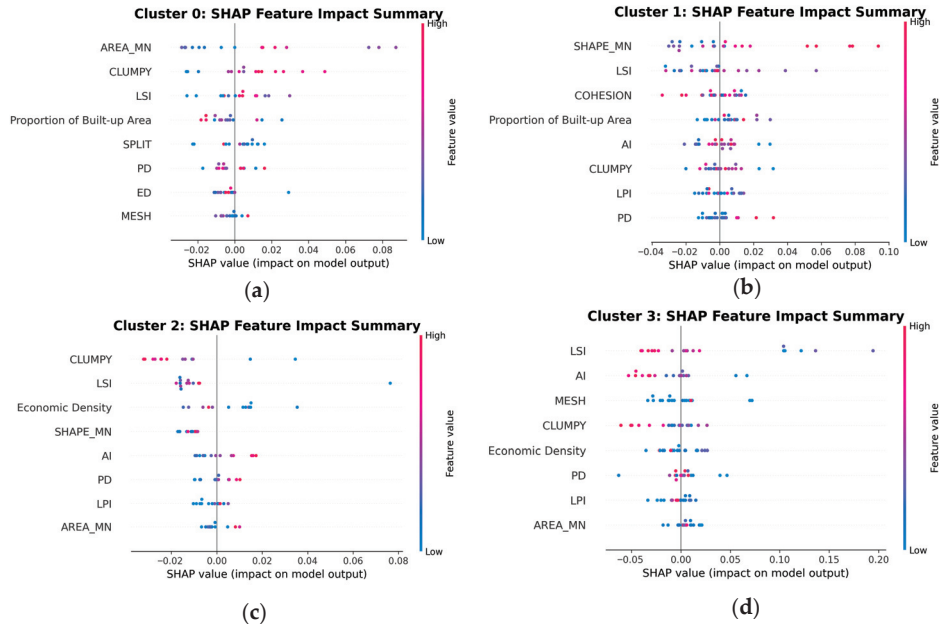


Figure 7. Feature importance of urban morphology indicators based on mean absolute SHAP values for the entire sample and four city zones. **Note**: (**a**–**d**) respectively represent the feature value rankings based on SHAP importance in Cluster 0, Cluster 1, Cluster 2, and Cluster 3.

https://doi.org/10.3390/land15050738

C1, as the medium-emission ecological urban cluster (Figure 7b), has SHAPE_MN as its most important morphological characteristic, with a mean absolute SHAP value of 0.0266. LSI ranks second with a mean absolute SHAP value of 0.0160, and COHESION ranks third with 0.0094. Proportion of built-up area and AI enter the top five important characteristics with mean absolute SHAP values of 0.0090 and 0.0089, respectively.

In C2, representing the medium-emission urban cluster (Figure 7c), CLUMPY becomes the most important characteristic with a mean absolute SHAP value of 0.0223. LSI ranks second with a mean absolute SHAP value of 0.0183, and economic density ranks third with 0.0125. SHAPE_MN and AI enter the top five important characteristics with mean absolute SHAP values of 0.0124 and 0.0080, respectively.

C3, as the low-emission urban cluster (Figure 7d), has LSI as its most important characteristic, with a mean absolute SHAP value as high as 0.0562, far higher than the importance of this characteristic in other clusters. AI ranks second with a mean absolute SHAP value of 0.0244, and MESH and CLUMPY tie for third with 0.0202. Economic density enters the top five important characteristics with a mean absolute SHAP value of 0.0145.

Cross-cluster comparative analysis shows that LSI enters the top five important characteristics in all four clusters, indicating that this characteristic has universal importance across different city types, particularly showing the highest importance in C3 low-emission cluster (0.0562), which reflects the significant impact of landscape shape complexity on carbon emissions in all city types. Proportion of built-up area ranks first at the full sample level and enters the top five in C0 and C1 clusters, indicating its most significant overall impact on urban carbon emissions. CLUMPY appears in three clusters—C0, C2, and C3—indicating that this characteristic has important influence on cities with different emission levels. However, dominant characteristics vary across different clusters: the high-emission urban cluster (C0) is dominated by AREA_MN, reflecting the impact of large continuous patches on high emissions; the ecological urban cluster (C1) is dominated by SHAPE_MN, indicating the importance of patch shape regularity for medium-emission ecological cities; the medium-emission urban cluster (C2) is dominated by CLUMPY, reflecting the influence of patch aggregation degree; and the low-emission urban cluster (C3) is dominated by LSI, highlighting the prominent role of landscape boundary complexity in emission reduction.

Comparison with full sample analysis results reveals that proportion of built-up area has the highest importance at the full sample level (0.0200), though importance rankings vary significantly across clusters, only entering the top five in C0 and C1 but not in C2 and C3, indicating that the global importance of this characteristic may be dominated by some high-weight cities. LSI ranks third at the full sample level (0.0188), but enters the top five in all clusters, even reaching 0.0562 in C3, demonstrating its universal importance across city types. These findings indicate that SHAP analysis based on the full sample can identify globally important characteristics, while clustering analysis can reveal heterogeneity of characteristic influence under different urban types, providing basis for differentiated emission reduction policy formulation.

4.4. SHAP Results Based on the Bayesian-Optimized Gradient Boosting Model

4.4.1. Nonlinear and Threshold Effects of Urban Morphology Indicators on Carbon Emission Intensity

In C0, AREA_MN, CLUMPY, and LSI dominate the classification outcome. Figure 8a demonstrates that all three features consistently increase the probability of assigning a city to the high carbon intensity category. AREA_MN exerts the strongest marginal contribution toward high carbon classification. Higher CLUMPY values similarly elevate this probability. LSI also shifts predictions toward the high carbon category, although its probabilistic impact increases more gradually. This pattern suggests that cities with larger, more aggregated,

https://doi.org/10.3390/land15050738

and more complex landscape patches face a higher likelihood of high carbon intensity classification. From a physical and socio-economic perspective, larger patch sizes and higher aggregation often correspond to lower land-use mixing and relatively dispersed spatial structures, which may extend commuting distances and limit public transport network coverage, thereby reducing overall transport efficiency. Concurrently, complex landscape boundary configurations (LSI) may imply there are increased exterior surface areas of building envelopes, elevating heating and cooling energy demands and collectively steering the model toward high carbon intensity classification.

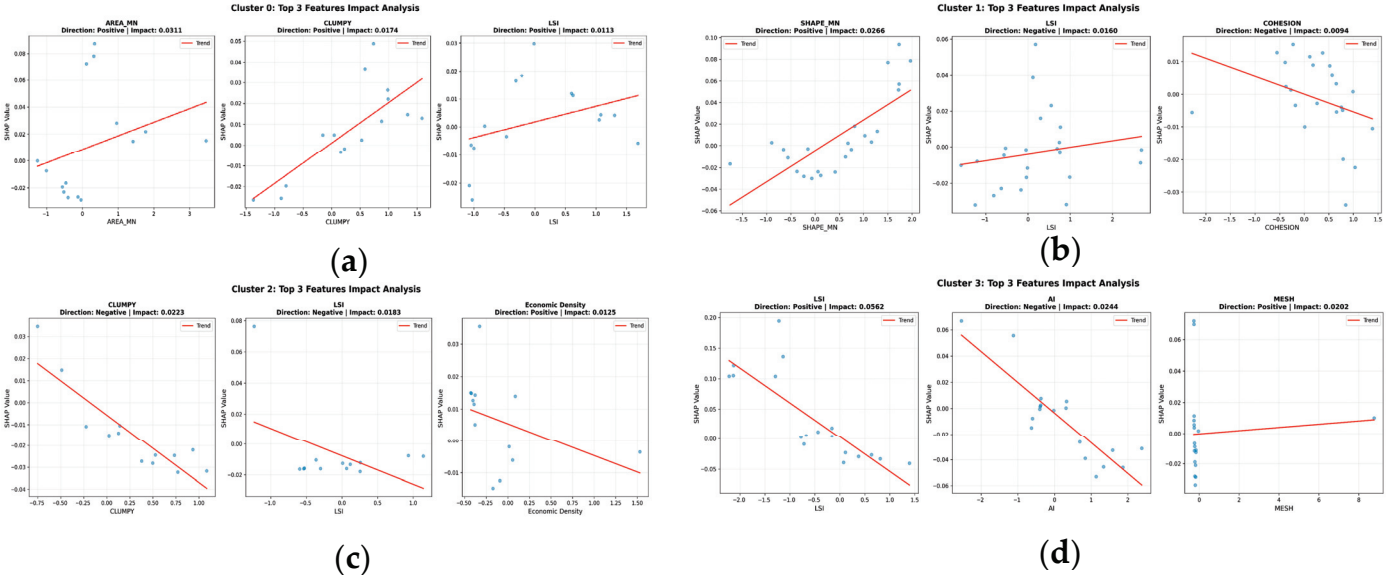


Figure 8. Nonlinear effects of key urban morphology indicators by urban zone. **Note**: (**a**–**d**) respectively represent the top three feature values in terms of importance in Cluster 0, Cluster 1, Cluster 2, and Cluster 3 (from left to right). In Cluster 0, they are in the fol-lowing order: AREA_MN, CLUMPY and LSI; In Cluster 1, they are in the following order: SHAPE_MN, LSI and COHESION; In Cluster 2, they are in the following order: CLUMPY, LSI and Economic Density; In Cluster 3, they are LSI, AI and MESH in se-quence.

In C1, SHAPE_MN, LSI, and COHESION exert the greatest influence on classification decisions. Figure 8b reveals complex nonlinear relationships for these indicators. SHAPE_MN substantially increases the probability of high carbon classification. This indicates that more irregular patch shapes elevate the likelihood of this outcome. Conversely, LSI and COHESION decrease the probability of high carbon classification and instead promote assignment to the low carbon category. The opposing directional effects between SHAPE_MN and LSI imply that individual patch shape irregularity and overall landscape pattern complexity alter classification results through distinct pathways. Specifically, high irregularity in individual patch shapes (SHAPE_MN) typically reflects fragmented land use or disordered development, which may compromise the connectivity of road and municipal utility networks and increase energy consumption for infrastructure deployment and maintenance. In contrast, increased overall landscape complexity (LSI) and enhanced inter-patch connectivity (COHESION) are often associated with mature mixed-use zones or highly integrated urban networks. These configurations facilitate shorter travel distances and improve accessibility for public transit and active mobility, thereby promoting low carbon category assignment at the system level.

In C2, CLUMPY, LSI, and economic density emerge as the primary classification drivers. Figure 8c shows that CLUMPY and LSI both reduce the probability of high carbon classification. Higher landscape aggregation and greater shape complexity therefore favor

https://doi.org/10.3390/land15050738

the low carbon category. Economic density produces the opposite effect by substantially increasing the likelihood of high carbon classification as economic activity density rises. This divergence demonstrates that landscape morphology and economic density exert opposing forces on the model decision boundary within this cluster. This phenomenon reflects a dynamic balance between spatial form optimization and economic activity intensity. Higher landscape aggregation and morphological complexity may indicate that urban areas have entered a compact development phase, where agglomeration effects promote the intensive utilization of infrastructure and improve energy system efficiency. However, increased economic density is directly associated with the concentration of commercial and industrial activities, which entails higher electricity consumption, logistics turnover, and production energy use. The resulting direct and embodied carbon emissions may partially offset the decarbonization potential derived from morphological optimization, leading to opposing effects within the classification decision.

In C3, LSI, AI, and MESH rank as the most influential indicators. Figure 8d displays a distinct classification pattern. LSI generates the strongest positive contribution to high carbon classification probability among all features across every cluster. AI exerts an equally strong but opposing effect by markedly decreasing this probability and reinforcing low carbon classification. MESH shows a modest tendency to increase the high carbon classification likelihood. These results indicate that extreme landscape shape complexity acts as the primary driver elevating the probability of high carbon intensity. Higher landscape aggregation conversely serves as a key factor promoting low carbon classification. The positive influence of extreme morphological complexity (LSI) may correlate with historical irregular street networks or inefficient spatial layouts, which increase travel detour probabilities and energy transmission losses. Highly aggregated spatial structures (AI) typically correspond to contiguous and compact urban fabrics that support centralized heating, regional energy planning, and efficient public service coverage, thereby reinforcing low carbon category assignment. The modest positive contribution of MESH may reflect the presence of large single-function zones, such as industrial parks or core commercial districts, which are characterized by concentrated energy consumption patterns and specific emission profiles.

Examination of the SHAP dependence plots (Figures 7 and 8) indicates that most morphological indicators do not exhibit linear relationships with carbon intensity classification. Several curves display consistent directional trends. The rate of probability shift varies across the feature range. This variation suggests that predictive contribution changes as indicator values fluctuate. Certain intervals produce sharper changes in class probability. Other intervals yield gradual shifts. This pattern reflects varying marginal contributions to the model decision boundary. These nonlinear probability responses indicate that urban morphology likely involves complex threshold effects on category assignment. When a morphological indicator reaches specific levels, its incremental impact on classification likelihood shifts substantially or stabilizes. Such nonlinear responses may correspond to critical states of urban infrastructure capacity, inflection points in transport network efficiency, or transitions in building energy consumption patterns. The variation in feature importance and directional effects across clusters further reflects heterogeneity in developmental stages, industrial structures, geographic constraints, and spatial planning policies. For instance, regions undergoing spatial expansion are more susceptible to patch scale effects, whereas economically dense areas are predominantly driven by the intensity of industrial and commercial activities. These mechanistic hypotheses provide a physical and socio-economic interpretive framework for understanding the linkage between morphological characteristics and carbon emission categories. The subsequent interaction analysis will

https://doi.org/10.3390/land15050738

further clarify whether these morphological indicators produce synergistic or antagonistic effects on classification outcomes.

4.4.2. Interactive Effects of Urban Morphology Indicators on Carbon Emission Intensity

Analysis of the complete dataset identifies three indicator pairs with the strongest interaction effects. Figure 9a presents the interaction between proportion of built-up area and LPI with an interaction value of 0.0057. Figure 9b displays the interaction between proportion of built-up area and SHAPE_MN with a value of 0.0056. Figure 9c illustrates the interaction between AI and CLUMPY with a value of 0.0043. High interaction intensity indicates that these feature pairs jointly shape the model decision boundary and substantially influence classification outcomes. The direction of the interaction values further clarifies how these combinations alter category assignments.

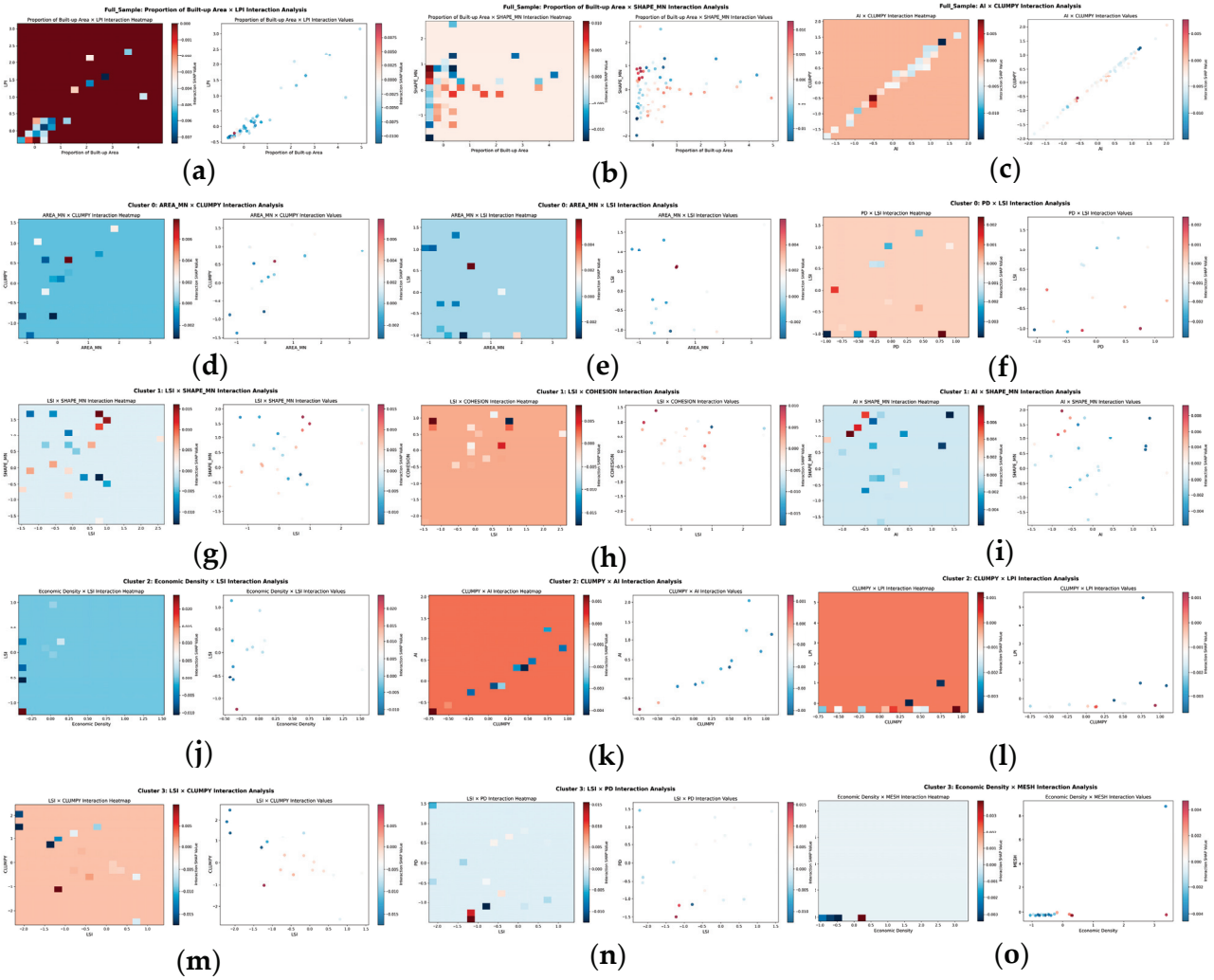


Figure 9. SHAP interaction of key urban morphology indicators by urban zone. **Note**: In the figure, it shows the interaction effects between the top three most important features in the entire sample and the sample subset, as well as the most interacting feature among them. (**a**–**c**) represent the respective interaction relationships between the Proportion of Built-up Area and LPI, SHAPE_MN, as well as AI and CLUMPY within the entire sample. (**d**–**f**) successively represent the interaction relationships of AREA_MN with CLUMPY, LSI, as well as PD and LSI in Cluster0. (**g**–**i**) respectively represent the interaction relationships between LSI and SHAPE_MN, COHE-SION, as well as between AI and SHAPE_MN in Cluster1. (**j**–**l**) show the relationship between Economic Density and LSI, CLUMPY, and AI and LPI in Cluster2, respectively. (**m**–**o**) represent the interaction relationships between LSI and CLUMPY, PD, Economic Density, and MESH respectively in Cluster 3.

https://doi.org/10.3390/land15050738

Examining the strongest pair of proportion of built-up area and LPI reveals predominantly negative SHAP interaction values across most value ranges. Negative values indicate that simultaneous changes in these two features jointly decrease the probability of assigning a city to the high carbon intensity category. This combined effect consistently promotes classification into the low carbon intensity category rather than increasing the likelihood of high carbon outcomes. This suggests that a high proportion of built-up areas, when coupled with a dominant largest patch, signifies a compact urban structure. Such compactness likely reduces transportation energy consumption and promotes economies of scale in infrastructure, thereby enhancing carbon emission intensity.

The interaction between proportion of built-up area and SHAPE_MN follows a comparable directional pattern. Conversely, the interaction between AI and CLUMPY shifts between positive and negative contributions across different value intervals. This variation demonstrates that their combined impact on classification probability is highly context dependent and reflects complex joint decision pathways within the model. Specifically, the transition from positive to negative values implies that while moderate aggregation might initially support development, excessive clumping without proper functional organization could lead to inefficiencies, highlighting the need for balanced spatial layouts.

Analysis of distinct urban development clusters reveals substantial heterogeneity in dominant interaction patterns, confirming the context-dependent nature of morphological impacts on carbon classification.

In C0, the strongest joint effects occur between AREA_MN and CLUMPY (Figure 9d, interaction value 0.0021) and between AREA_MN and LSI (Figure 9e, interaction value 0.0021). Within specific value ranges, the AREA_MN and CLUMPY combination yields strongly positive SHAP interaction values. This indicates that large patch sizes combined with high aggregation synergistically increase the probability of high carbon classification. From a planning perspective, this reveals a "scale inefficiency" in this cluster. It implies that simply merging small patches into larger aggregated blocks without optimizing internal functions may increase carbon intensity, suggesting that urban renewal in these areas should focus on internal structural optimization rather than mere expansion. The interaction between patch density and LSI similarly exhibits a tendency to elevate this classification likelihood (Figure 9f).

In C1, the interaction network centers on LSI and SHAPE_MN, which produce the strongest joint effect (Figure 9g, interaction value 0.0065). Predominantly positive interaction values across most ranges demonstrate that combining overall landscape complexity with irregular patch shapes synergistically elevates the probability of high carbon classification. This suggests that in these cities, the morphological penalty of having both complex landscapes and irregular shapes is severe. Notable interactions also emerge between LSI and COHESION (Figure 9h) and between the aggregation index and SHAPE_MN (Figure 9i), further confirming that spatial configuration and shape regularity are the dominant morphological determinants for carbon intensity in this cluster.

In C2, strong interactions are observed among key morphological indicators, specifically involving aggregation and shape metrics. The economic density and LSI pairing dominates (Figure 9j, interaction value 0.0064). Strongly positive interaction values in specific intervals indicate that high economic density combined with landscape complexity substantially increases the likelihood of high carbon classification. Regarding pure morphological interactions, significant joint effects also characterize the CLUMPY and AI combination (Figure 9k) and the CLUMPY and LPI combination (Figure 9l). These results highlight that in C2, the spatial aggregation of patches (CLUMPY and AI) interacts intensely with landscape composition, playing a critical role alongside economic factors.

https://doi.org/10.3390/land15050738

In C3, LSI anchors the interaction network, exhibiting the strongest joint effect with CLUMPY (Figure 9m, interaction value 0.0066). High concurrent values generate strongly positive interactions that markedly increase the probability of high carbon classification. Conversely, the LSI and PD interaction yields negative values at low indicator levels, indicating a synergistic reduction in high carbon classification likelihood. This suppressive effect diminishes as either of the metrics rise (Figure 9n). This implies that while low-density, simple landscapes are carbon efficient, uncontrolled increases in complexity and aggregation can rapidly degrade environmental performance. Therefore, maintaining low landscape complexity is critical for sustaining low carbon intensity in C3 cities. The economic density and MESH interaction remains comparatively weak (Figure 9o). These patterns underscore that managing landscape shape complexity alongside spatial aggregation patterns is critical for accurate classification in this cluster.

5. Discussion

5.1. Nonlinear Effects of Urban Morphology on Carbon Emission Intensity

Prior research has predominantly examined linear relationships between urban morphology and continuous environmental metrics, with limited attention to the complex nonlinear mechanisms governing carbon intensity category assignment. Departing from this tradition, the present study integrates a gradient boosting classifier with a dual-interpretability framework. This methodology systematically quantifies the nonlinear contributions of morphological features to categorical prediction and examines their heterogeneity across distinct urban types. Results indicate that morphological features exert nonlinear effects on carbon category assignment, with the magnitude and direction of their contributions varying across city types. Partial dependence plots reveal distinct class probability trajectories, demonstrating that the relationship between morphological indicators and classification likelihood rarely follows a monotonic pattern. Instead, it frequently exhibits threshold regions where the predicted probability of category membership shifts noticeably. The detailed examination of these threshold dynamics and their quantitative benchmarks is provided in Section 5.2.

Specifically, this study identifies multiple pathways through which urban morphology influences carbon intensity category assignment. First, spatial configurations regulate the distribution and interaction of human activities and material flows. Higher spatial aggregation and landscape connectivity may enhance operational efficiency yet require alignment with economic density. When local capacity constraints are approached, these features can shift the classification probability toward higher emission categories. Second, morphological characteristics influence category assignment by modulating local microclimates. Landscape complexity and fragmentation affect ventilation and surface evaporation, thereby altering building energy demands. The nonlinear probability trajectories clarify how these microclimatic influences transition from reducing to increasing the likelihood of classification into specific intensity categories as morphological complexity rises. Finally, morphological attributes interact with the spatial distribution of socio-economic activities. Combined contributions of economic density and specific morphological metrics, such as effective mesh size, can noticeably shift classification likelihood across urban categories. Fundamentally, urban landscape patterns operate as complex systems in which spatial configuration influences category assignment through threshold-dependent nonlinear contributions. This perspective offers an alternative to linear frameworks and provides a structured reference for developing context-specific low-carbon spatial planning strategies.

https://doi.org/10.3390/land15050738

5.2. Threshold Effects of Urban Morphology on Carbon Emission Intensity

The influence of key morphological indicators on carbon emission intensity category assignment does not follow a linear trajectory. Instead, the directional impact and contribution intensity of these indicators undergo observable shifts within distinct value ranges. As demonstrated in the analytical results and visualized in Figure 6, PDPs identify critical transition zones where these directional changes occur, offering measurable reference points for spatial regulation. For example, the marginal effect of urban land proportion (proportion of built-up area) on classification probability shifts from reducing to increasing the likelihood of high emission category assignment as urban expansion progresses. Similarly, the landscape shape index (LSI) exhibits a clear inflection point, beyond which increased boundary complexity consistently correlates with a higher probability of assignment to high emission categories. The mean urban patch shape index (SHAPE_MN) demonstrates a stabilizing influence up to a certain level, after which its marginal contribution gradually diminishes. These threshold-dependent patterns align with classical urban economic theory, particularly the dynamic trade-off between agglomeration economies and congestion effects. Within the present analytical framework, initial morphological compactness generates efficiency gains through shared infrastructure and reduced spatial friction; however, once critical thresholds are surpassed, these agglomeration benefits are progressively offset by congestion externalities and capacity constraints, consistent with established expectations of diminishing returns in urban spatial systems. Recognizing these nonlinear response trajectories enables planners to anticipate critical transition points and adjust spatial configurations proactively before adverse emission outcomes materialize.

Crucially, these threshold effects exhibit substantial heterogeneity across urban clusters. In the high-emission C0 cluster, thresholds for spatial aggregation and patch size appear at lower values compared to other groups, suggesting that capacity constraints emerge earlier in cities with terrain-induced spatial fragmentation. In the medium-emission ecological C1 cluster, the threshold for shape complexity interacts differently with cohesion metrics, indicating that maintaining regular patch shapes up to a certain level supports lower-emission classification. The medium-emission developmental C2 cluster shows that economic density modifies morphological thresholds, with high economic activity lowering the tolerance for landscape complexity before classification probability shifts upward. In the low-emission C3 cluster, the landscape shape index threshold is particularly pronounced, highlighting that strict control of boundary irregularity is essential to maintain low-emission status. These divergent trajectories suggest that optimizing single morphological indicators is not universally applicable. Strategies aimed at influencing category assignment through spatial regulation should account for local baseline conditions, recognizing that exceeding context-specific thresholds may yield limited or counterproductive outcomes. Furthermore, interaction effects among morphological factors indicate that combined thresholds shift under different variable pairings, reinforcing the need to evaluate morphological regulation within an integrated, cluster-specific framework.

5.3. Implications for Urban Planning and Management

Urban spatial form functions as a direct regulatory instrument for carbon management rather than a passive physical backdrop. The identified nonlinear thresholds and interaction patterns provide measurable targets for spatial planning and development control. To systematically translate these findings into actionable governance, it is essential to categorically align spatial interventions with the distinct economic development models characterizing each cluster. The four categories reflect divergent developmental trajectories, each exhibiting a unique mechanism through which economic activities interact with urban form to either exacerbate or mitigate carbon intensity. Effective implementation requires ex-

https://doi.org/10.3390/land15050738

plicitly mapping these development models to ensure morphological adjustments directly support low-carbon economic transitions.

Cities in the high-emission cluster (i.e., C0) typically follow an extensive expansion model constrained by complex topography and reliant on resource intensive industries. Economic growth in these regions drives continuous physical land consumption. Planning authorities must enforce strict boundaries for urban expansion and mandate minimum landscape aggregation thresholds. Directing new industrial development toward existing compact nodes reduces infrastructure duplication and lowers transportation emissions. Municipal zoning codes should require regular plot shapes and integrated green corridors to counteract ventilation blockage from fragmented layouts. Assertively, transitioning from extensive land-driven growth to morphological consolidation constitutes the critical pathway to decouple economic expansion from carbon escalation in terrain-constrained regions. This morphological consolidation directly links spatial efficiency to reduced carbon intensity while sustaining industrial output.

Cities in the medium-emission ecological cluster (i.e., C1) operate under a dispersed development model driven by resource extraction and ecological conservation mandates. Spatially scattered economic activities increase transport distances and buildings' energy demands. Planning must prioritize consolidating isolated settlements into coordinated service hubs. Regulatory instruments should establish connectivity standards that link functional zones without increasing overall landscape complexity. Investment in centralized heating networks and shared logistics facilities within these hubs directly couples economic efficiency with morphological carbon reduction. Crucially, shifting from dispersed resource-dependent expansion to hub-based consolidation ensures that ecological preservation mandates are structurally reinforced by compact spatial organization, directly translating economic activity into measurable carbon reduction. Consolidating economic functions while maintaining ecological corridors ensures resource activities do not trigger disproportionate emission increases.

Cities in the medium-emission developmental cluster (i.e., C2) exhibit a dense compact growth model characteristic of mature economic centers. Substantial economic output per unit area generates localized emission pressures from concentrated commercial and industrial activities. Urban management should decouple economic intensification from spatial congestion by promoting vertical mixed-use zoning and expanded underground infrastructure. Planning regulations must cap landscape shape complexity while preserving designated ventilation corridors. Allocating high value commercial functions to transit oriented nodes and relocating heavy logistics to peripheral zones ensures economic density improvements translate into measurable emission reductions. Fundamentally, economic intensification in mature centers must be structurally decoupled from horizontal sprawl; enforcing vertical densification and strict spatial boundaries guarantees that economic agglomeration yields proportional carbon intensity reduction rather than compounding environmental loads. Maintaining strict spatial boundaries while upgrading vertical capacity allows economic growth to proceed without increasing landscape fragmentation.

Cities in the low-emission cluster (i.e., C3) follow a traditional industrial development model characterized by monocentric ring expansion. Historical industrial layouts and aging infrastructure sustain moderate baseline emissions within an already efficient spatial configuration. Economic revitalization must avoid uncontrolled boundary irregularity. Municipal authorities should adopt control standards that limit the landscape shape index at current levels while mandating energy efficient retrofits for existing industrial facilities. Development incentives should prioritize brownfield redevelopment within established urban rings to preserve ecological buffers. Specifically, preserving the existing monocentric equilibrium while channeling economic revitalization inward demonstrates

https://doi.org/10.3390/land15050738

that morphological stability, rather than outward expansion, is the most effective driver of carbon reduction in historically industrialized cities. Channeling economic investment into existing industrial cores rather than expanding outward preserves the current reduced emission morphological equilibrium and prevents carbon intensity escalation.

The successful application of these spatial governance measures depends on integrating morphological thresholds directly into municipal development approval workflows. Local planning departments should establish automated compliance checks that evaluate proposed land use changes against cluster specific emission probability curves. Regular monitoring of landscape metric shifts will allow authorities to adjust zoning parameters before threshold crossings occur. This targeted approach ensures that economic development strategies consistently align with spatial configurations that minimize carbon intensity across diverse urban contexts.

5.4. Limitations and Future Directions

This study examines the nonlinear, threshold-dependent, and interactive contributions of urban morphology to carbon emission intensity category assignment. Given the constraints of available data and methodological scope, several limitations remain, and future research may extend the analysis in the following directions.

First, the interpretability framework primarily identifies contribution patterns and relative importance, but the underlying physical and socio-economic mechanisms driving these threshold shifts require further exploration. Future studies could integrate multi-source datasets, such as high-resolution building energy consumption records, traffic mobility patterns, or microclimate observations, and combine these with urban climate or transportation models to provide more detailed mechanistic explanations for the observed threshold and interaction effects.

Second, the current analyses relies on cross-sectional panel data, which captures static associations across discrete time points. Urban morphology and carbon emission patterns are inherently dynamic, and longitudinal changes may involve cumulative effects, lagged responses, or path dependencies. Constructing continuous temporal datasets and applying spatiotemporal modeling techniques or sequential machine learning approaches could help clarify the long-term evolution of morphological influences on category assignment.

Third, this study primarily focuses on morphological indicators, while carbon emission intensity is also shaped by non-morphological factors such as energy structure, industrial composition, technological adoption, and policy interventions. These variables are not explicitly included in the current model and may operate as unobserved influences. As a result, the importance rankings of morphological features may partly reflect correlations with these unmeasured variables rather than their intrinsic causal effects. While this study establishes robust statistical associations between urban morphology and carbon emission categories, the results do not imply direct causality. Instead, urban morphology acts as a "spatial constraint framework" that shapes carbon emission patterns indirectly by influencing traffic efficiency, energy consumption modes, and built environment configurations. Future frameworks could incorporate broader socio-economic and policy indicators (where data availability permits) to more comprehensively assess the relative contributions of morphological factors and explore their synergistic pathways with technological and institutional variables.

6. Conclusions

Based on a sample of 336 Chinese cities, this study constructs an urban morphological indicator system encompassing multi-dimensional characteristics to classify urban carbon emission intensity levels. To identify urban typologies, the K-means clustering algorithm,

https://doi.org/10.3390/land15050738

Gaussian mixture models, agglomerative hierarchical clustering, and spectral clustering were systematically compared. Comparative evaluation indicates that spectral clustering yields comparatively balanced category distributions, which supports the subsequent statistical representation of distinct urban types. Consequently, this study applies spectral clustering to partition the sample into four urban clusters characterized by differing development trajectories. Integrating a Bayesian optimized gradient boosting classifier with a dual-interpretability framework combining SHAP and PDPs, the study examines the relationship between urban morphology and carbon emission intensity category assignment from three perspectives. These perspectives include feature relevance, nonlinear probability transitions, and interaction mechanisms. While SHAP quantify global feature contributions to category assignment, PDPs illustrate average marginal probability shifts, facilitating the identification of transition boundaries and nonlinear classification trajectories. The results suggest that the association between urban morphology and emission category assignment exhibits nonlinearity and threshold dependency, with dominant features, directional influences, and probability transition ranges varying across urban types.

The findings indicate that uniform morphological regulation approaches may not suit all urban contexts. Within the high-emission intensity cluster, mean urban patch area and urban clumpiness demonstrate positive associations with higher emission category assignment. PDPs indicate probability transition ranges beyond which spatial expansion and extensive agglomeration correspond with elevated classification probabilities, suggesting the need for measured development controls. In the medium-ecological cluster, complex interactions and compensatory relationships exist among shape indicators. Spatial planning may benefit from coordinating morphological complexity across landscape and patch scales within probability ranges associated with lower emission categories. Within the medium-emission developmental cluster, interaction effects between economic density and morphological characteristics are notable, suggesting that spatial adjustments during development may help moderate the association between economic intensification and higher emission category assignment. In the low emission cluster, landscape shape complexity correlates with increased classification probabilities, whereas spatial aggregation demonstrates a mitigating association. PDPs indicate a probability shift boundary for the landscape shape index, suggesting that monitoring this metric may support category stability.

Furthermore, interaction effects among morphological factors exhibit considerable complexity. Analysis using SHAP indicates that combinations of key morphological features produce non-additive effects on category assignment probabilities. The magnitude and direction of these interactions demonstrate substantial variation across urban clusters. Identical feature pairs may exhibit synergistic, compensatory, or statistically neutral associations depending on the urban context. Complementing this observation, stratified partial dependence plots demonstrate how these interaction dynamics influence probability distributions and transition boundaries across different urban settings.

Overall, this study examines the nonlinear, threshold, and interaction patterns associated with the relationship between urban morphology and carbon emission intensity categories across national and subregional scales. By utilizing partial dependence plots to identify probability transition ranges and SHAP to assess feature relevance, this research offers a reference framework for developing differentiated spatial planning strategies across varying urban development contexts. Additionally, the analytical framework may be applied to examine the relationship between urban morphology and other ecological indicators, offering methodological considerations for spatial optimization and urban resilience assessment.

Author Contributions: Conceptualization, Y.T., W.L., X.Y., L.C. and M.L.; methodology, W.L. and L.C.; software, W.L. and L.C.; validation, Y.T., W.L., X.Y., L.C. and M.L.; formal analysis, W.L. and L.C.; investigation, W.L. and L.C.; resources, L.C. and M.L.; data curation, L.C.; writing—original

draft preparation, Y.T., W.L. and L.C.; writing—review and editing, Y.T., W.L., X.Y., L.C. and M.L.; visualization, W.L. and L.C.; supervision, Y.T., X.Y. and M.L.; project administration, Y.T., L.C. and M.L.; funding acquisition, Y.T., L.C. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation (No. 2023NSFSC0521), Sichuan Provincial Philosophy and Social Science Research (No. SC23TJ039), Fundamental Research Funds for the Central Universities: Self-initiated Project of the School of Public Administration (2025), and Sichuan Provincial Philosophy and Social Science Fund Young Talents Project (No. SCJJ25QN12).

Data Availability Statement: Dataset available on request from the authors.

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

Appendix A

Table A1. Descriptive statistics of urban morphological indicators within the clustering area.

Metric Name	C0		C1		C2		C3	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Proportion of Built-Up Area	0.0090	0.0115	0.0241	0.0091	0.0414	0.0755	0.0058	0.0382
PD	0.0023	0.0032	0.0015	0.0010	0.0088	0.0111	0.0007	0.0023
LPI	0.0826	0.1516	0.9647	0.2150	1.0190	3.3013	0.0911	1.7176
ED	0.1921	0.2979	0.5051	0.1646	1.2060	1.9581	0.1215	0.7867
LSI	10.7093	12.0361	9.8622	7.8374	14.9860	16.0590	7.1264	10.4533
AREA_MN	153.5667	172.5931	1086.5476	523.8283	353.6216	612.8499	397.6154	1576.7277
SHAPE_MN	1.8720	1.8954	2.281	2.0088	1.9020	1.9320	1.9884	2.3927
CLUMPY	0.8596	0.8593	0.9379	0.9162	0.9038	0.9051	0.9117	0.9365
COHESION	95.1344	95.0874	98.3481	96.5445	97.6999	97.8763	96.4639	98.4111
MESH	2.5207	6.6003	115.6101	20.6720	114.1223	1695.1472	3.2395	690.1546
SPLIT	707,887.5633	12,628,614.1753	7541.8321	100,084,683.9418	7173.0992	13,276.8932	579,233.1124	28,914.1427
AI	86.0197	85.9966	93.9178	91.6565	90.8741	91.0708	91.1994	93.8399
People Density	1.1493	33.4046	3.0357	23.8877	5.3802	33.9811	0.9734	29.4489
Economic Density	5.9114	294.0912	23.9293	258.2912	57.3151	458.4860	5.2685	263.8001

References

1. Seto, K.C.; Fragkias, M.; Güneralp, B.; Reilly, M.K. A meta-analysis of global urban land expansion. *PLoS ONE* **2011**, *6*, e23777. [CrossRef]
2. Hong, S.; Hui, E.C.M.; Lin, Y. Relationship between urban spatial structure and carbon emissions: A literature review. *Ecol. Indic.* **2022**, *144*, 109456. [CrossRef]
3. Zheng, S.; Huang, Y.; Sun, Y. Effects of urban form on carbon emissions in China: Implications for low-carbon urban planning. *Land* **2022**, *11*, 1343. [CrossRef]
4. Sun, C.; Zhang, Y.; Ma, W.; Wu, R.; Wang, S. The impacts of urban form on carbon emissions: A comprehensive review. *Land* **2022**, *11*, 1430. [CrossRef]
5. Dalde, M.C.; Nitivattananon, V.; Sharma, D.; Ninsawat, S. Effects of urban form and socio-economic factors on transport-related carbon dioxide emissions: A structural equation approach. *Int. J. Transp. Sci. Technol.* **2026**, *21*, 191–207. [CrossRef]
6. Jareemit, D.; Srivanit, M.; Tanapant, S.; Limeechokchai, B. Role of urban morphology integrated building envelope materials in achieving zero emissions: A simulation-based study in complex-shaped urban blocks. *Energy Rep.* **2025**, *13*, 27–39. [CrossRef]
7. Li, L.; Sun, S.; Zhong, L.; Han, J.; Qian, X. Novel spatiotemporal nonlinear regression approach for unveiling the impact of urban spatial morphology on carbon emissions. *Sustain. Cities Soc.* **2025**, *125*, 106381. [CrossRef]
8. Oke, T.R. The energetic basis of the urban heat island. *Q. J. R. Meteorol. Soc.* **1982**, *108*, 1–24. [CrossRef]
9. Gago, E.J.; Roldan, J.; Pacheco-Torres, R.; Ordóñez, J. The city and urban heat islands: A review of strategies to mitigate adverse effects. *Renew. Sustain. Energy Rev.* **2013**, *25*, 749–758. [CrossRef]
10. Yan, H.; Zhou, M.; Wang, K.; Wu, X.; Lin, T.; Zhang, G.; Liu, Y.; Ye, H. Can urban three-dimensional compact form reduce building operational carbon emissions? New evidence from a prototypical coastal city. *Sustain. Cities Soc.* **2025**, *130*, 106664. [CrossRef]
11. Sharifi, A. Resilient urban forms: A macro-scale analysis. *Cities* **2019**, *85*, 1–14. [CrossRef]
12. Imhoff, M.L.; Zhang, P.; Wolfe, R.E.; Bounoua, L. Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sens. Environ.* **2010**, *114*, 504–513. [CrossRef]
13. Yang, J.; Yang, Y.; Sun, D.; Jin, C.; Xiao, X. Influence of urban morphological characteristics on thermal environment. *Sustain. Cities Soc.* **2021**, *72*, 103045. [CrossRef]

14. Bowler, D.E.; Buyung-Ali, L.; Knight, T.M.; Pullin, A.S. Urban greening to cool towns and cities: A systematic review of the empirical evidence. *Landsc. Urban Plan.* **2010**, *97*, 147–155. [CrossRef]
15. Kabisch, N.; Strohbach, M.; Haase, D.; Kronenberg, J. Urban green space availability in European cities. *Ecol. Indic.* **2016**, *70*, 586–596. [CrossRef]
16. Yu, B.; Pan, J. How do 2D and 3D urban morphology impact spatial patterns of thermal environment? A nested multi-scale local climate zone perspective. *Build. Environ.* **2026**, *288*, 114014. [CrossRef]
17. Chen, Y.; Ma, W.; Shao, Y.; Wang, N.; Yu, Z.; Li, H.; Hu, Q. The impacts and thresholds detection of 2D/3D urban morphology on the heat island effects at the functional zone in megacity during heatwave event. *Sustain. Cities Soc.* **2025**, *118*, 106002. [CrossRef]
18. Tang, G.; Du, X.; Wang, S. Impact mechanisms of 2D and 3D spatial morphologies on urban thermal environment in high-density urban blocks: A case study of Beijing's Core Area. *Sustain. Cities Soc.* **2025**, *123*, 106285. [CrossRef]
19. Bereitschaft, B.; Debbage, K. Urban form, air pollution, and CO_2 emissions in large US metropolitan areas. *Prof. Geogr.* **2013**, *65*, 612–635. [CrossRef]
20. Zhang, X.; Meng, Q.; Xu, J.; Li, K. Urbanization impacts on anthropogenic carbon dioxide emissions and the roles of urban morphologies: Insights from urban socioeconomic clusters and local climate zones in Shanghai, China. *Sustain. Cities Soc.* **2025**, *129*, 106494. [CrossRef]
21. Huang, Y.; Wang, Z.; Zhao, H.; You, D.; Wang, W.; Peng, Y. Quantifying the nonlinear effects of urban morphological features on ecological resilience: Evidence from cities in China. *Sustain. Cities Soc.* **2026**, *136*, 107072. [CrossRef]
22. Wang, W.; Wang, Y.; Shen, C. Quantifying the nonlinear and interactive effects of urban form on resilience to extreme precipitation: Evidence from 192 cities of Southern China. *Sustain. Cities Soc.* **2025**, *125*, 106366. [CrossRef]
23. Guo, L.; Du, S.; Sun, W.; Fan, D.; Wu, Y. Multi-scale impact of urban building function and 2D/3D morphology on urban heat island effect: A case study in Shanghai, China. *Energy Build.* **2025**, *338*, 115719. [CrossRef]
24. Shi, K.; Xu, T.; Li, Y.; Chen, Z.; Gong, W.; Wu, J.; Yu, B. Effects of urban forms on CO_2 emissions in China from a multi-perspective analysis. *J. Environ. Manag.* **2020**, *262*, 110300. [CrossRef] [PubMed]
25. Xie, B.; Liu, R.; Dwivedi, R. Digital economy, structural deviation, and regional carbon emissions. *J. Clean. Prod.* **2024**, *434*, 139890. [CrossRef]
26. Li, R.; Li, L.; Wang, Q. The impact of energy efficiency on carbon emissions: Evidence from the transportation sector in Chinese 30 provinces. *Sustain. Cities Soc.* **2022**, *82*, 103880. [CrossRef]
27. Palani, H.; Acosta-Sequeda, J.; Karatas, A.; Derrible, S. The role of socio-demographic and economic characteristics on energy-related occupant behavior. *J. Build. Eng.* **2023**, *75*, 106875. [CrossRef]
28. Xue, J.; Liu, T.; Long, F.; Ge, C.; Weng, Z. Short and long-term impact assessment of urban low-carbon pilot policies on household decarbonization. *Environ. Impact Assess. Rev.* **2026**, *117*, 108234. [CrossRef]
29. Wang, J.F.; Li, X.H.; Christakos, G.; Liao, Y.L.; Zhang, T.; Gu, X.; Zheng, X.Y. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [CrossRef]
30. Huang, Y.; Li, S.; Lin, J.; Zheng, L.; Zhuang, C.; Guan, C.; Guo, Y.; Zhuang, Y. Nonlinear and threshold effects of urban building form on carbon emissions. *Energy Build.* **2025**, *329*, 115243. [CrossRef]
31. Chen, L.; Tang, P.; Li, J.; Li, J. Assessment of Low-Carbon Utilization in Territorial Space and Identification of Its Driving Factors: A Case Study of the Yangtze River Economic Belt in China. *Land* **2025**, *14*, 738. [CrossRef]
32. Li, Y.; Liu, L.; Zhang, T.; Liang, Y.; Chen, Z.; Chu, L.; He, C. Climate change and human activities drive the warm-season rooftop solar photovoltaic potential in the Chinese Chengdu-Chongqing urban agglomeration. *Sustain. Cities Soc.* **2025**, *119*, 106110. [CrossRef]
33. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*; NIPS Foundation: San Diego, CA, USA, 2017; Volume 30.
34. Xia, Z.; Li, C.; Zhang, F.; Yan, Y.; Sun, J.; Ma, D.; Guo, Z. Spatial–temporal heterogeneous analysis of the coupling coordination between new urbanization and urban ecological resilience. *Int. J. Urban Sci.* **2025**, 1–29. [CrossRef]
35. Chen, J.; Gao, M.; Cheng, S.; Liu, X.; Hou, W.; Song, M.; Li, D.; Fan, W. China's city-level carbon emissions during 1992–2017 based on the inter-calibration of nighttime light data. *Sci. Rep.* **2021**, *11*, 3323. [CrossRef]
36. McGrane, S.J. Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: A review. *Hydrol. Sci. J.* **2016**, *61*, 2295–2311. [CrossRef]
37. McGarigal, K.; Marks, B.J. *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure*; USDA Forest Service, Pacific Northwest Research Station: Portland, OR, USA, 1995.
38. Bettencourt, L.M. The origins of scaling in cities. *Science* **2013**, *340*, 1438–1441. [CrossRef] [PubMed]
39. Seto, K.C.; Sánchez-Rodríguez, R.; Fragkias, M. The new geography of contemporary urbanization and the environment. *Annu. Rev. Environ. Resour.* **2010**, *35*, 167–194. [CrossRef]

40. Bashmakov, I.A.; Nilsson, L.J.; Acquaye, A.; Bataille, C.; Cullen, J.M.; de la Rue du Can, S.; Fischedick, M.; Geng, Y.; Tanaka, K. *Climate Change 2022: Mitigation of Climate Change*; Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2022; Chapter 11.
41. Andrew, J.; Kaidonis, M.A.; Andrew, B. Carbon tax: Challenging neoliberal solutions to climate change. *Crit. Perspect. Account.* **2010**, *21*, 611–618. [CrossRef]
42. Tapio, P. Towards a theory of decoupling: Degrees of decoupling in the EU and the case of road traffic in Finland between 1970 and 2001. *Transp. Policy* **2005**, *12*, 137–151. [CrossRef]
43. Wu, Y.; Zhu, Q.; Zhu, B. Decoupling analysis of world economic growth and CO_2 emissions: A study comparing developed and developing countries. *J. Clean. Prod.* **2018**, *190*, 94–103. [CrossRef]
44. Stern, N. The economics of climate change. *Am. Econ. Rev.* **2008**, *98*, 1–37. [CrossRef]
45. Wang, S.; Zhao, Y.; Wiedmann, T. Carbon emissions embodied in China–Australia trade: A scenario analysis based on input–output analysis and panel regression models. *J. Clean. Prod.* **2019**, *220*, 721–731. [CrossRef]
46. Caruana, R.; Niculescu-Mizil, A. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*; Association for Computing Machinery: New York, NY, USA, 2006; pp. 161–168.
47. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
48. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning*; Springer: New York, NY, USA, 2009; Volume 2.
49. Cheng, Y.; Zhang, X.; Yu, S.; Liu, Y.; Hu, J.; Jiang, Y.; Zhang, C.; Wu, X. Leveraging Explainable Machine Learning to Decipher Ecosystem Health and Nonlinear Dynamics in the Henan Yellow River Basin. *Land* **2026**, *15*, 429. [CrossRef]
50. Ru, J.; Li, J.; Gan, L.; Yusufu, G. Digital–Intelligent Transformation and Urban Carbon Efficiency in the Yellow River Basin: A Hybrid Super-Efficiency DEA and Interpretable Machine-Learning Framework. *Land* **2026**, *15*, 159. [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.

land

MDPI

Article

Analysis of Factors Influencing Fire Risk in High-Density Urban Areas Based on the CatBoost-SHAP Model

Yunlong Wei and Hu Li *

School of Human Settlements Science and Design, North China University of Water Resources and Electric Power, Zhengzhou 450046, China; x20231141019@stu.ncwu.edu.cn
* Correspondence: lihu@ncwu.edu.cn

Abstract

Urban fire risk in high-density cities is characterized by complex spatial heterogeneity and nonlinear relationships with the built environment, population distribution, and climatic conditions. However, most existing studies rely on linear assumptions and offer limited interpretability. To address this gap, we developed an interpretable analytical framework that integrates the CatBoost model with SHAP (SHapley Additive exPlanations), using Futian District in Shenzhen as a case study. We constructed a fire risk surface from historical fire incident data using kernel density estimation (KDE) and incorporated multiple urban environmental factors—including points of interest (POIs), road networks, and meteorological variables—as explanatory variables. The CatBoost model captured nonlinear relationships, while SHAP quantified feature importance and revealed interaction effects. The results show that urban fire risk is strongly associated with the spatial agglomeration of population-related facilities, especially high-density commercial and residential areas, as well as thermal conditions. Several variables exhibit clear nonlinear threshold effects, with their influence on fire risk varying markedly across different intensity ranges. Interaction analysis further indicates that combinations of built-environment characteristics and climatic factors jointly shape the spatial pattern of fire risk. These findings provide empirical insights into the spatial mechanisms underlying urban fire risk and highlight the value of interpretable machine learning in urban safety research. The proposed framework offers a practical tool for developing more targeted, evidence-based fire risk management strategies in high-density urban areas.

Keywords: urban fire; risk assessment; ArcGIS; machine learning; CatBoost algorithm; SHAP

1. Introduction

Urban fires are common urban disasters that have been increasing in frequency, posing severe threats to public safety and property. As urbanization accelerates and urban governance becomes more granular, identifying fire risk factors and developing effective prevention strategies have emerged as critical research topics in the field of urban safety [1]. Although fire incidents are characterized by their stochasticity, the integration of big data [2] and machine learning (ML) technologies [3] allows for a more profound understanding of risk factors and enables more precise urban fire risk assessments [4].

Currently, methodologies for analyzing urban fire risk factors primarily include the DEMATEL-ISM method [5], AHP-FCE method [6], entropy weight method [7], multi-criteria decision-making [8], risk matrix method [9], TOPSIS [10], and various machine learning techniques [11–13]. Zamanialaei et al. (2025) [14] used machine learning

to assess structure-level fire risk in California's wildland–urban interface (WUI), identifying building spacing and defensible space as key predictors. Complementing their WUI-focused study, the present work applies an interpretable CatBoost-SHAP framework to intra-urban fires in high-density settings, thereby extending data-driven fire risk assessment to purely urban contexts. Research scales in this domain span from individual buildings [15–17] and blocks [18,19] to urban grids [20,21]. For instance, Sun et al. (2019) developed an evaluation index system using the Analytic Hierarchy Process (AHP) based on historical fire statistics and theoretical analysis, thereby promoting the integration of smart fire protection within smart city frameworks [22]. Wei et al. (2018) proposed a rapid assessment approach combining fuzzy mathematics with Support Vector Machines (SVM); by training the model with samples derived from fuzzy comprehensive evaluations, they successfully predicted risks for specific sports facilities with high accuracy [23]. Kim et al. (2023) [24] developed a prediction model for fire occurrence and casualties using datasets from Seoul (2017–2021). By combining built environment characteristics with meteorological and demographic data, they enhanced predictive performance and identified high-risk zones in the Gangnam district. Furthermore, Oh and Jeong (2022) introduced an XGBoost-based prediction method utilizing grid-level data (e.g., population, aged buildings, and high-rise structures), which outperformed baseline models like SVM and logistic regression in localized fire prevention [25]. However, despite the predictive superiority of ML, existing studies often suffer from the "black-box" nature of these algorithms, failing to provide actionable spatial thresholds for urban planners. Alongside traditional methods' issues of high subjectivity and coarse spatial resolution, current research lacks the capacity to fully capture complex nonlinear relationships or the spatiotemporal transition of driving mechanisms during rapid urbanization. Specifically, there is a lack of in-depth analysis regarding the fine-grained spatiotemporal evolution of fire risks within urban core areas.

To address these gaps, the present study focuses on the Futian District of Shenzhen. By integrating historical fire records with multi-source built environment data, we investigate the spatiotemporal distribution of fire incidents. Specifically, this study fills the identified gaps by: (1) constructing an interpretable CatBoost [26]-SHAP [27] framework to decouple the non-linear influences of multi-source urban data; (2) identifying the spatiotemporal transition of fire risk drivers in Futian District; and (3) quantifying critical management thresholds (e.g., firefighting service radii and thermal limits) to support evidence-based urban resilience planning. This research provides a scientific basis for optimizing fire resource allocation and advancing resilient urban planning.

2. Materials and Methods

Figure 1 presents the overall methodological workflow used to quantify, model, and interpret the spatial distribution of urban fire risk in Futian District during 2008–2018. The study timeline is first segmented into two distinct urban development stages—(i) the Urban Rapid Development Period (2008–2012) and (ii) the Spatial Structure Optimization Period (2013–2018)—to capture temporal transitions in urban morphology and fire risk patterns.

The analytical framework includes three main steps. The Fire Risk Index is first derived from historical fire points of interest (POI) data (2008–2012 and 2013–2018) through kernel density estimation (KDE). Parallel to this, Fire Risk Impact Factors are generated from multiple geospatial datasets: (a) POI kernel densities of public gathering places, residential places, flammable/explosive sites, large-scale buildings, and vulnerable facilities; (b) raster-based environmental variables, including topographic factors (DEM, slope), climatic variables (temperature and precipitation), and urban/environmental indicators such as nighttime light intensity (NTL), NDVI, road density, built-up density, and distance

https://doi.org/10.3390/land15050796

to fire stations. All impact layers are standardized to the same spatial resolution and time windows.

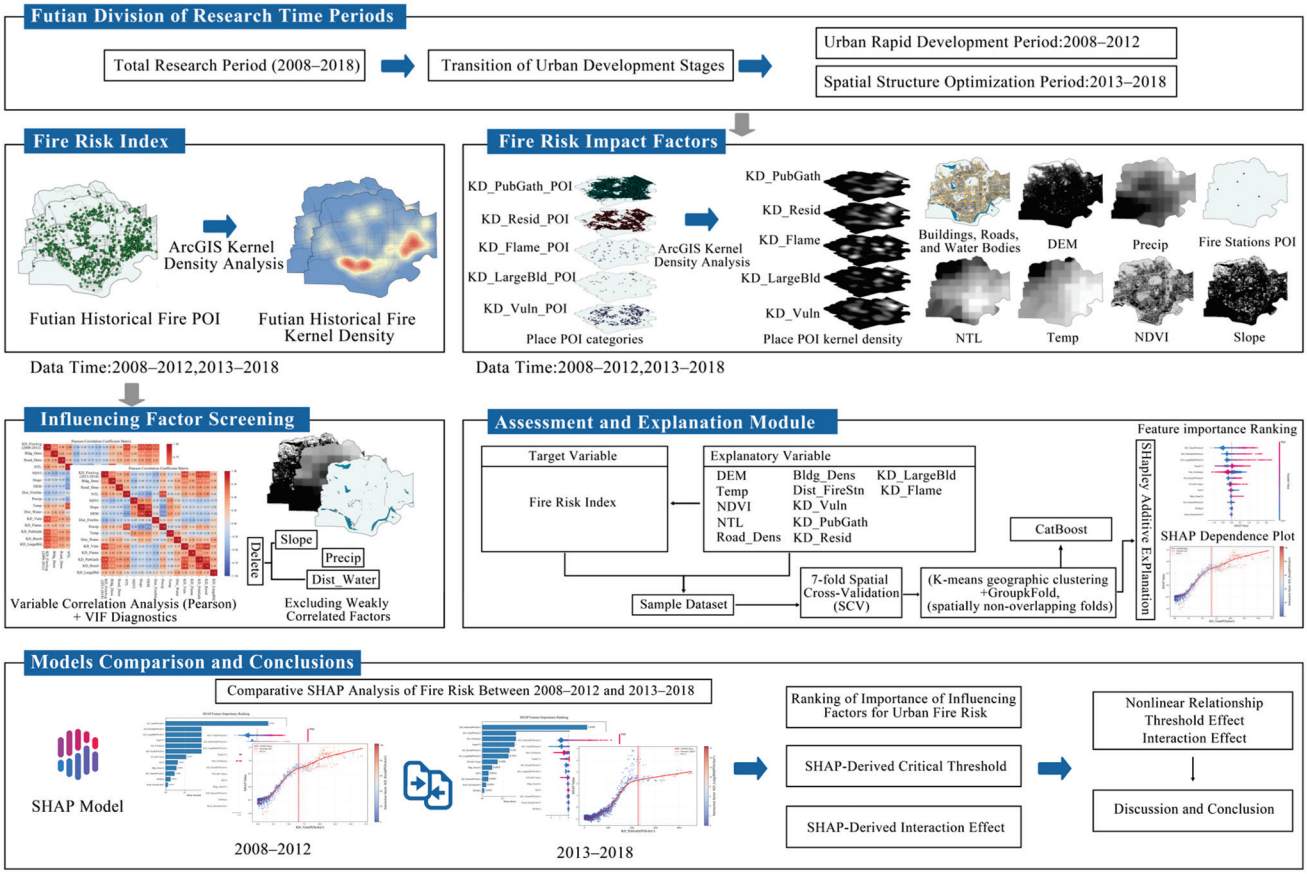


Figure 1. Research framework. The overall workflow illustrates the process of urban fire risk assessment in Futian District using multi-source data, kernel density estimation (KDE), CatBoost modeling, and SHAP (SHapley Additive exPlanations) interpretation. The study period is divided into two phases: the Urban Rapid Development Period (2008–2012) and the Spatial Structure Optimization Period (2013–2018). FRI denotes the Fire Risk Index. Key abbreviations in the figure: POI = Points of Interest; KD_ = Kernel Density; DEM = Digital Elevation Model; NDVI = Normalized Difference Vegetation Index; NTL = Nighttime Light Intensity; SCV = Spatial Cross-Validation; SHAP = SHapley Additive exPlanations.

Subsequently, Influencing Factor Screening is performed via variable correlation analysis (Pearson heatmap) and collinearity diagnostics. Variables exhibiting weak correlation or strong collinearity—specifically slope, annual precipitation, and distance to water bodies—are excluded to reduce multicollinearity and enhance model interpretability.

The core Assessment and Explanation Module employs a supervised CatBoost gradient-boosting framework. The fire risk index serves as the target (dependent) variable, while the screened explanatory variables (DEM, annual mean temperature, building density, residential density, public gathering places density, road density, nighttime light intensity, NDVI, flammable/explosive sites density, large-scale buildings density, vulnerable facilities density, and distance to fire stations) constitute the feature set. After hyperparameter tuning and model training/validation, SHAP (SHapley Additive exPlanations) values are computed for global and local interpretability. The results include (i) feature importance rankings, (ii) SHAP dependence plots showing nonlinear effects, and (iii) interaction heatmaps.

Finally, the Models Comparison and Conclusions section conducts a comparative SHAP analysis between the two development periods (2008–2012 vs. 2013–2018), high-

https://doi.org/10.3390/land15050796

lighting shifts in factor dominance, critical risk thresholds, and interaction effects. These insights feed directly into the ranking of influencing factors, derivation of SHAP-derived critical thresholds and interaction effects, and the ultimate discussion and conclusions on nonlinear threshold effects and policy implications for urban fire-risk management.

2.1. Study Area: Futian District, Shenzhen, China

Futian District is located in the central part of the Shenzhen Special Economic Zone and serves as a key central urban area for Shenzhen's development and construction (Figure 2). The district covers 78.66 km^2 and includes 10 sub-districts. It represents a typical high-density urban built-up area [28]. Futian District concentrates a substantial number of high-rise buildings, commercial complexes, aging residential communities, and "urban villages" (chengzhongcun), resulting in fire safety risks that exhibit diversified and complex characteristics. Between 2008 and 2018, a total of 2762 fire incidents were recorded in the district. After data cleaning to remove duplicate records, 2654 valid fire cases were retained for further analysis. Persistent safety hazards—such as the illegal parking of electric bicycles and the improper storage of debris obstructing fire access lanes—further underscore its status as a priority area for urban public safety governance.

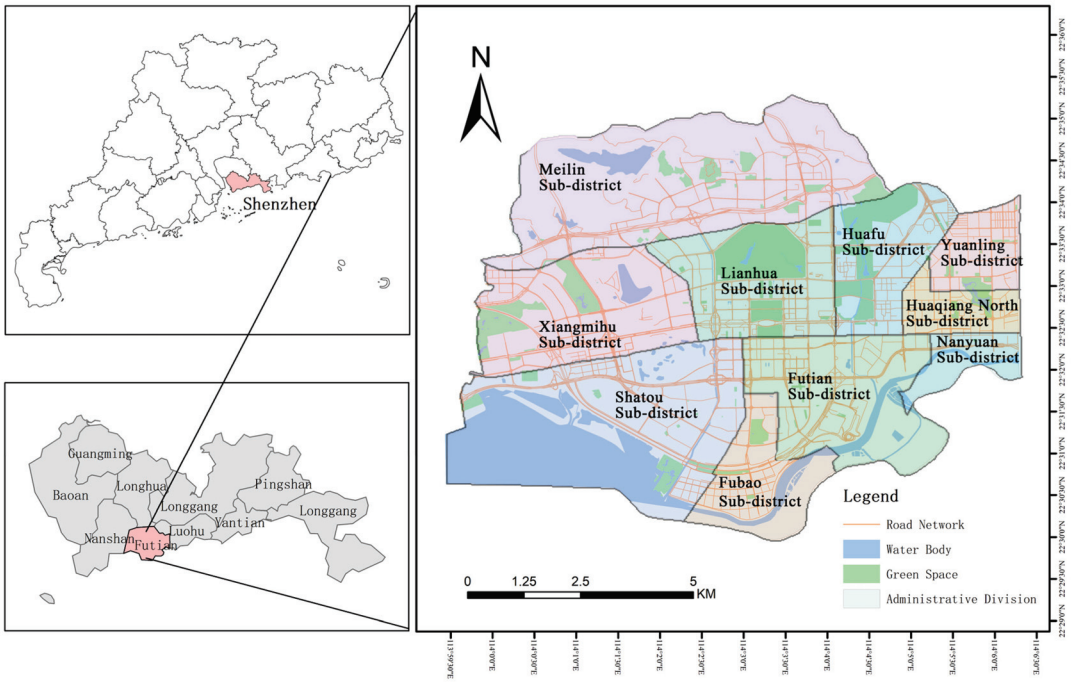


Figure 2. Study area of Futian District, Shenzhen.

2.2. Data Sources and Preprocessing

The data used in this study mainly include elevation, slope, normalized difference vegetation index (NDVI), nighttime light intensity, building density, road density, and POI data, all of which are obtained from publicly available online sources. The specific data sources are listed in Table 1.

Historical fire incident data for Futian District, Shenzhen, were obtained from the Shenzhen Municipal Government Data Open Platform [29], covering all publicly available records from January 2008 to December 2018. The initial dataset contained 2761 records. We applied a systematic multi-step cleaning protocol using Python 3.13 (pandas and geopandas) to ensure data quality and reproducibility. Duplicate records were removed using Python (pandas and geopandas). Duplicates were defined as entries with identical timestamps, overlapping coordinates within a 10 m tolerance, and consistent incident

descriptions (n = 64). Subsequently, fire locations were geocoded via the Baidu Maps API and transformed to the WGS-84 coordinate system in ArcGIS 10.8. Records with coordinates falling outside the administrative boundary of Futian District or containing ambiguous addresses that could not be reliably verified were excluded to maintain spatial consistency with other geospatial layers, representing 35 instances. An additional 8 records were discarded due to missing critical fields such as alarm dates or location descriptions. All exclusion criteria were applied uniformly and documented in the preprocessing workflow to ensure analytical transparency. Following these procedures, 2654 valid fire incidents remained for analysis and were stratified into two sub-periods: 2008–2012 and 2013–2018, to align with the district's transition from rapid urban expansion to spatial structure optimization around 2012. The kernel density of these fire incidents was calculated to construct the Fire Risk Index (FRI), which represents the continuous spatial intensity of urban fire risk. Notably, the original open-platform records lacked a standardized fire severity classification. To avoid introducing subjective categorization bias or sample imbalance, all incidents were treated as binary spatial events, and kernel density estimation was employed to capture continuous variations in overall fire risk intensity.

POI data were obtained from the Baidu Maps API. After deduplication and filtering, the raw POI data were reclassified according to fire risk characteristics into five major categories. Large-scale buildings include factories, shopping malls, and warehousing and logistics facilities, which are characterized by high fire loads and concentrated combustible materials. Residential places mainly refer to residential communities, where the use of electricity and fire-related activities is concentrated. Public gathering places include accommodation, catering, supermarkets, internet cafés, and tutoring institutions, which typically contain abundant combustible materials and high population density. Vulnerable facilities include hospitals, schools, nursing homes, and childcare facilities, where occupants often have limited emergency response capability. Flammable and explosive places include petrochemical enterprises, new energy enterprises, and gas stations, where flammable materials are highly concentrated.

Table 1. Multi-source geographic data.

Data Category	Data Name	Abbreviation	Data Source
Natural Environment	Elevation	DEM	Geospatial Data Cloud [30] (https://www.gscloud.cn)
	Slope	Slope	Geospatial Data Cloud [30] (https://www.gscloud.cn)
	Normalized Difference Vegetation Index	NDVI	National Ecosystem Science Data Center [31] (https://doi.org/10.12199/nesdc.ecodb.rs.2021.012)
	Annual Total Precipitation	Precip	National Tibetan Plateau Data Center [32] (https://doi.org/10.5281/zenodo.3114194)
	Annual Mean Temperature	Temp	National Tibetan Plateau Data Center [33] (https://doi.org/10.11888/Meteoro.tpdc.270961)
	Distance to Water Bodies	Dist_Water	OpenStreetMap [34] (https://www.openstreetmap.org/#map=11/22.3567/114.1363) (accessed on 22 October 2025)
Built Environment	Road Density	Road_Dens	OpenStreetMap [34] (https://www.openstreetmap.org/#map=11/22.3567/114.1363) (accessed on 22 October 2025)
	Building Density	Bldg_Dens	East Asia Building High-Quality Vector Dataset [35] (https://doi.org/10.5281/zenodo.8174931)
	Nighttime Light Intensity	NTL	China Long-Term Nighttime Light Dataset (2000–2020) [36] (https://doi.org/10.3974/geodb.2022.06.01.V1)
	Vulnerable Facilities Density	KD_Vuln	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)
	Public Gathering Places Density	KD_PubGath	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)
	Flammable/Explosive Sites Density	KD_Flame	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)

Table 1. *Cont.*

Data Category	Data Name	Abbreviation	Data Source
	Residential Density	KD_Resid	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)
	Large-scale Buildings Density	KD_LargeBld	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)
	Distance to Fire Stations	Dist_FireStn	Baidu Maps API POI data [37] (2012–2018) (https://lbsyun.baidu.com/products/map) (accessed on 14 October 2025)
Historical Fire Data	Fire Risk Index	FRI	Calculated using Formula (2)

Baidu POI data were period-specifically matched (2012 snapshot for 2008–2012 period and 2018 snapshot for 2013–2018 period).

Some raster-based variables (e.g., NDVI and temperature) have relatively coarse spatial resolution compared to the scale of the study area. To ensure spatial consistency, all variables were resampled to a unified resolution and aggregated to a 240 m × 240 m grid, which serves as the analytical unit. This resolution represents a balance between data availability and the need to capture intra-urban spatial variability. In addition, kernel density estimation (KDE) was applied to transform discrete fire incidents into continuous spatial surfaces, which helps reduce the sensitivity of the analysis to differences in data resolution. Therefore, the modeling framework focuses on relative spatial patterns and nonlinear relationships rather than fine-scale pixel accuracy.

To ensure temporal consistency, POI data from 2012 and 2018 are used to represent the two study periods (2008–2012 and 2013–2018), respectively. This approach captures the dominant built-environment characteristics of each phase while maintaining comparability [37]. Given the marked slowdown in large-scale development after 2012, the 2012 POI dataset effectively captures the vast majority of functional facilities added between 2008 and 2012, thereby serving as a robust representation of the built environment characteristics during that specific period. By the same token, the 2018 POI dataset reasonably reflects the overall conditions during the 2013–2018 period. This snapshot approach allows for a more accurate depiction of the actual urban functional distribution and POI density characteristics during each phase, avoiding potential temporal biases that might arise from relying on a single static dataset. Furthermore, both snapshots adhere to a consistent POI classification system, thereby ensuring methodological comparability across the two distinct time periods.

To further ensure the robustness of the CatBoost model and mitigate potential multicollinearity among explanatory variables, a Variance Inflation Factor (VIF) diagnostic was conducted for all 15 candidate predictors across both study periods (2008–2012 and 2013–2018). After Pearson correlation screening, three weakly correlated variables (Slope, Precip, Dist_Water) were excluded. VIF values for the remaining 12 variables are shown in Table 2. The VIF quantifies how much the variance of a regression coefficient is inflated due to linear dependence with other features; A commonly used threshold is VIF < 10, indicating acceptable levels of collinearity [38].

Table 2. Variance inflation factor (VIF) diagnostics for explanatory variables across two study periods.

Variable		VIF (2008–2012)	VIF (2013–2018)	Collinearity Status
Elevation	DEM	4.90	4.84	Acceptable
Normalized Difference Vegetation Index	NDVI	2.23	2.18	Acceptable
Annual Mean Temperature	Temp	3.26	3.73	Acceptable
Building Density	Bldg_Dens	2.23	2.24	Acceptable
Road Density	Road_Dens	1.53	1.52	Acceptable
Nighttime Light Intensity	NTL	4.24	3.7	Acceptable
Vulnerable Facilities Density	KD_Vuln	5.02	4.62	Moderate (Period 1) *

https://doi.org/10.3390/land15050796

Table 2. *Cont.*

Variable		VIF (2008–2012)	VIF (2013–2018)	Collinearity Status
Public Gathering Places Density	KD_PubGath	4.52	5.67	Acceptable
Flammable/Explosive Sites Density	KD_Flame	1.54	1.79	Acceptable
Residential Density	KD_Resid	9.53	4.39	Moderate (Period 1) *
Large-scale Buildings Density	KD_LargeBld	2.68	1.54	Acceptable
Distance to Fire Stations	Dist_FireStn	1.56	1.86	Acceptable

* Threshold: VIF < 10.0 indicates acceptable multicollinearity. Variables with 5.0 ≤ VIF < 10.0 are marked with *; all values decreased or remained stable in the second period (2013–2018), supporting model robustness.

As shown in Table 2, all retained variables exhibited VIF values below the threshold of 10.0 in both periods. Specifically, 13 out of 15 variables showed VIF < 5.0, indicating negligible multicollinearity. Two variables—KD_Resid (Residential Density) and KD_Vuln (Vulnerable Facilities Density)—exhibited moderately elevated VIF values in the first period (9.53 and 5.02, respectively), but both decreased substantially in the second period (4.39 and 4.62), reflecting improved feature independence following urban spatial restructuring.

Notably, although tree-based models like CatBoost are relatively robust to multicollinearity, conducting VIF screening enhances the interpretability of SHAP values by reducing feature redundancy. The diagnostic results confirm that multicollinearity does not substantially affect the stability of feature importance estimation in our modeling framework.

2.3. Indicator System Construction

This study focuses on urban fire risk assessment and influencing factor analysis in Futian District, Shenzhen. A CatBoost regression model was constructed based on multi-source data. The kernel density of urban fire incident points is selected as the Urban Fire Risk Index, which reflects the spatial clustering characteristics and potential intensity of fire risk.

Based on an extensive review of related literature and expert consultation [39–42], the following variables were selected as explanatory features: elevation (DEM), slope (Slope), normalized difference vegetation index (NDVI), annual total precipitation (Precip), annual mean temperature (Temp), distance to water bodies (Dist_Water), building density (Bldg_Dens), road density (Road_Dens), nighttime light intensity (NTL), vulnerable facilities density (KD_Vuln), public gathering places density (KD_PubGath), flammable/explosive sites density (KD_Flame), residential density (KD_Resid), large-scale buildings density (KD_LargeBld), and distance to fire stations (Dist_FireStn).

Although vegetation coverage in Futian District is relatively limited due to its high urbanization level, NDVI was retained as a proxy variable to capture micro-scale environmental heterogeneity. NDVI reflects variations in surface moisture and urban thermal conditions. These factors may influence fire ignition and spread. Moreover, in highly urbanized areas, fragmented green spaces and urban vegetation may still contribute to localized cooling effects and humidity regulation. Therefore, NDVI was not interpreted as a dominant ecological driver, but rather as a complementary environmental indicator interacting with temperature and built-environment variables.

Although building density reflects the physical intensity of the built environment, nighttime light intensity (NTL) captures the functional dimension of urban activity, particularly human activity intensity and energy consumption patterns during nighttime. In high-density urban areas, NTL has been widely used as a proxy for socioeconomic activity and urban vitality. These aspects are not fully represented by structural indicators such as building density. Therefore, the inclusion of NTL enables the model to distinguish between

https://doi.org/10.3390/land15050796

areas with similar physical density but different levels of human activity, which may lead to different fire risk profiles.

It is acknowledged that the distance to fire stations is more directly related to fire response capacity rather than the intrinsic probability of fire occurrence. However, in this study, this variable is incorporated as a contextual modifier to capture the spatial heterogeneity of fire risk outcomes. Areas with longer distances to fire stations may experience delayed response times, which can exacerbate fire consequences and indirectly influence the observed spatial pattern of fire incidents. Therefore, this variable is not interpreted as a direct causal driver of fire occurrence, but rather as a proxy for reflecting the spatial configuration of emergency service accessibility within the urban system.

2.4. Spatial Analysis Methods

For the spatial analysis, a regular grid-based spatial unit system covering the entire Futian District was established using ArcGIS 10.8. The grid cell size was set to 240 m × 240 m, generating a total of 1596 grid cells as independent analytical units. The geocoded historical fire event point data were imported into the spatial database, and kernel density estimation (KDE) was applied (with the quartic kernel function used by default in ArcGIS) to analyze the spatial point pattern.

2.4.1. Kernel Density Estimation (KDE)

The spatial intensity of urban fire incidents was quantified using KDE, which is defined as:

$$f(x) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{d(x, x_i)}{h}\right) \quad (1)$$

where n is the number of fire incidents, h is the bandwidth (determined by Silverman's rule of thumb to ensure spatial smoothness), and K is the kernel function. This method allows for the identification of high-risk "hotspots" beyond discrete points, producing continuous fire risk surfaces for both sub-periods.

This method quantifies the influence intensity of fire incident points on surrounding areas through a weighted function that decays with distance, effectively smoothing the data distribution and revealing spatial clustering characteristics. All kernel density calculations, including the Fire Risk Index (FRI) and POI indicators, adopted a unified parameter system: the cell size was set to 30 m, and the search radius and bandwidth parameter (h) were optimized according to the Silverman rule [43]:

$$h = 0.9 \times min\left(\sigma, \frac{IQR}{1.34}\right) \times n^{-1/5} \quad (2)$$

where σ denotes the standard deviation, IQR represents the interquartile range, and n indicates the sample size; these parameters were automatically computed using the "Bandwidth Calculator" tool in ArcGIS 10.8.

By calculating the kernel density value (fire events per km^2) for each grid cell, a continuous spatial risk surface was generated. This density value serves as the regional Fire Risk Index (FRI), which is quantified using the following formula:

$$FRI_i = \frac{1}{h^2}\sum_{j=1}^{n} K\left(\frac{d_{ij}}{h}\right) \quad (3)$$

where FRI_i represents the Fire Risk Index of grid cell i, K denotes the kernel density function, d_{ij} represents the Euclidean distance between the center of grid cell i and fire

https://doi.org/10.3390/land15050796

incident point j, h is the bandwidth parameter, and n denotes the total number of fire incidents within the bandwidth range.

To assess the robustness of the Fire Risk Index (FRI) to the chosen analytical parameters, a sensitivity analysis was conducted. The baseline configuration (240 m × 240 m grid and bandwidth determined by Silverman's rule) was compared against four alternative setups: (i) 120 m grid, (ii) 480 m grid, (iii) bandwidth × 0.8, and (iv) bandwidth × 1.2. Pearson correlation coefficients between the baseline FRI and the alternative FRI surfaces exceeded 0.92 in all cases ($p < 0.01$), and the spatial overlap of high-risk hotspots (Getis-Ord Gi* hotspot analysis at 99% confidence) remained above 87%. These results indicate that the identified fire risk patterns and subsequent SHAP-derived thresholds are robust to reasonable variations in grid resolution and bandwidth, although finer grids (120 m) slightly increase local variability in low-density areas while coarser grids (480 m) tend to smooth extreme hotspots.

2.4.2. Spatial Autocorrelation

Spatial autocorrelation analysis quantifies the degree of autocorrelation of geographic variables within a study area, reflecting spatial dependence and heterogeneity. It can identify spatial hotspots, cold spots, and outliers, and help interpret the spatial structure and variation patterns of geographic phenomena. This method is used to explore the spatial distribution and temporal variation characteristics of the historical fire kernel density (FRI) in Futian District, Shenzhen. The analysis was conducted using ArcGIS [44].

Among the methods employed, Global Spatial Autocorrelation (Global Moran's I) [45] focuses on testing the overall spatial clustering tendency of the study object. It is used to determine whether the fire kernel density exhibits significant spatial clustering or dispersion patterns during the study period.

High/Low clustering analysis (Getis-Ord General G), a global spatial statistical method, is used to measure the spatial clustering degree of high or low values within the study area. This statistic standardizes the weighted sum of attribute values for spatial units and their neighboring units, comparing the calculated observed value with the expected value under a random distribution. The method generates a z-score and p-value to determine clustering significance: when the z-score is positive, the observed General G index is greater than the expected index, indicating a high-value clustering pattern; when the z-score is negative, the expected index exceeds the observed index, indicating a low-value clustering pattern. We use it to identify the overall high/low clustering pattern of fire kernel density [46].

Following a "global–local" analytical framework, this study first tests the overall clustering trend using global spatial autocorrelation methods. Subsequently, the Getis-Ord General G statistic is applied to conduct high/low clustering analysis. Combined with the actual geographic context of Futian District and the characteristics of the fire kernel density dataset, multidimensional statistical significance tests and feature interpretation are performed to more comprehensively reveal the spatial distribution patterns and heterogeneity of fire kernel density in the study area.

2.5. CatBoost-SHAP Modeling Framework

2.5.1. CatBoost Algorithm

CatBoost is a high-performance gradient boosting decision tree algorithm developed by Yandex. It improves upon the traditional Gradient Boosting Decision Tree (GBDT) framework by learning and predicting based on symmetric decision trees. By addressing issues related to categorical feature processing and prediction shift, the algorithm effectively reduces overfitting [47,48].

https://doi.org/10.3390/land15050796

To address spatial autocorrelation in grid-based data, a 7-fold spatial cross-validation (SCV) framework was adopted. The study area was divided into seven spatial clusters using K-means based on geographic coordinates. Model performance was quantified using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). R^2 reflects the explanatory power of the model for the target variable, with values closer to 1 indicating better performance. RMSE and MAE quantify the magnitude of prediction errors, with smaller values indicating higher accuracy. Together, these three indicators provide a comprehensive evaluation of the goodness-of-fit and predictive accuracy of the urban fire risk model. The calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

where n represents the number of samples, y_i denotes the observed value, $\hat{y}_i$ represents the predicted value, and $\overline{y}$ denotes the mean of the observed values y_i. The coefficient of determination R^2 measures the goodness-of-fit of the model to the data, with a value range of [0, 1]. Values closer to 1 indicate a better model fit.

A comprehensive grid search strategy was employed to optimize CatBoost hyperparameters, including tree depth, learning rate, and regularization coefficients, ensuring an optimal balance between predictive accuracy and model stability. To enhance the scientific rigor of the interpretability analysis, a statistical verification framework was introduced for identifying critical thresholds. Rather than relying on subjective visual inspection of SHAP dependence plots, piecewise linear fitting combined with 1000-iteration bootstrap resampling was applied to calculate 95% confidence intervals (CIs) for each identified breakpoint. This approach quantifies predictive uncertainty and provides a statistically robust foundation for evidence-based policy recommendations. In practical application, the optimized CatBoost algorithm efficiently processes multiple influencing factors, accurately captures nonlinear associations and interaction effects among built environment and climatic variables, and suppresses overfitting through spatially constrained validation. This provides a reliable modeling foundation for subsequent SHAP-based interpretation.

To investigate the relationship between regional fire kernel density and its influencing factors, the fire risk kernel density of Futian District during 2008–2012 and 2013–2018 was used as the dependent variable. The independent variables included DEM, NDVI, Temp, Bldg_Dens, Road_Dens, NTL, KD_Vuln, KD_PubGath, KD_Flame, KD_Resid, KD_LargeBld, and Dist_FireStn.

2.5.2. SHAP Method

This study adopts the SHAP (SHapley Additive explanations) method proposed by Scott M. Lundberg et al. [27] to identify key factors influencing changes in urban fire risk and to determine the direction of their effects. SHAP quantifies the contribution of each feature at the individual sample level, enabling an intuitive interpretation of how different variables influence variations in urban fire risk. SHAP interaction values are used to analyze the interaction effects among feature variables, helping to explain complex nonlinear relationships. The calculation formula is as follows:

https://doi.org/10.3390/land15050796

$$f(x) = \phi_0 + \sum_{i=1}^{n} \phi_i \tag{7}$$

where $f(x)$ represents the predicted value of the Fire Risk Index, ϕ_0 denotes the baseline value, and ϕ_i represents the SHAP value of each feature, reflecting the marginal contribution of that feature to the prediction result. To further analyze the interaction effects among features, SHAP interaction values are employed:

$$\phi_{ij} = f(S \cup \{i,j\}) - f(S \cup \{i\}) - f(S \cup \{j\}) + f(S) \tag{8}$$

where ϕ_{ij} represents the interaction contribution value between feature i and feature j, and S denotes a subset of features.

Equation (8) measures the joint effect of two features and reveals how their combined influence affects the model's prediction results.

SHAP dependence plots were subsequently generated to visualize the marginal relationship between each feature and the predicted risk. To rigorously identify nonlinear threshold effects and quantify their uncertainty, this study applied a bootstrap-based piecewise linear fitting approach (pwlf package, 1000 bootstrap resamples) on the SHAP dependence plots. The identified breakpoints and their 95% confidence intervals (95% CI) provide statistically supported critical thresholds for each key feature.

3. Results

3.1. Spatiotemporal Variation in Fire Risk in Futian District

3.1.1. Temporal Variation Characteristics of Historical Fires in Futian District

As shown in Figure 3, a total of 2654 fire incidents occurred in Futian District, Shenzhen, between 2008 and 2018, with an annual average of 241 fires. Figure 3a indicates that the interannual variation in fire frequency exhibits significant fluctuations, showing an initial increase followed by a decline. Among these years, 2013 represents the peak year of fire occurrence, with 436 recorded fires, accounting for 16.4% of the total during the study period. In contrast, 2008 was the year with the lowest fire frequency, with only 110 incidents, representing 4.1% of the total. The decline observed after 2013 may be related to the strengthening of national and local fire safety policies.

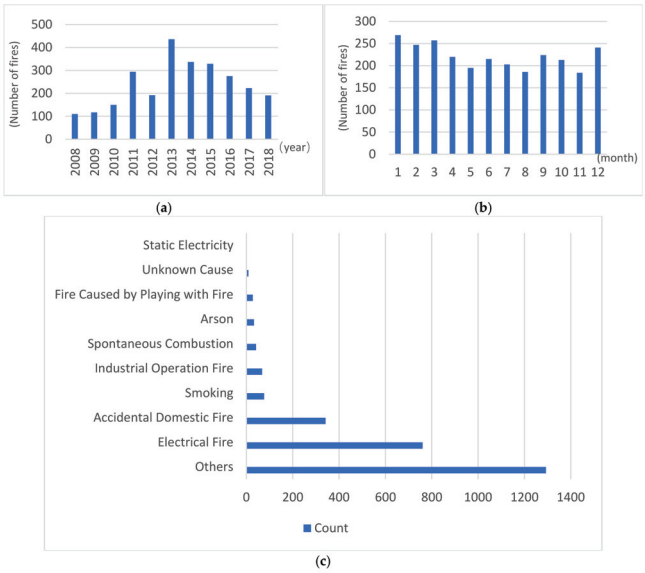


Figure 3. Historical fire statistics of Futian District. (**a**) Interannual distribution of historical fires in Futian District. (**b**) Intermonthly distribution of historical fires in Futian District. (**c**) Bar chart of fire causes in Futian.

https://doi.org/10.3390/land15050796

Figure 3b shows the monthly distribution averaged across years. At the monthly scale, fire incidents occur more frequently in January, February, and March. Among them, January records the highest number of fires, with 269 incidents, accounting for approximately 10.1% of the annual total. This pattern may be associated with dry and low-humidity winter conditions, increased use of heating equipment, and fireworks during the Chinese New Year period. In contrast, November shows the lowest fire frequency, with 184 incidents, accounting for 6.9% of the annual total. The relatively higher humidity and moderate temperatures during this month make combustible materials less likely to ignite.

Figure 3c illustrates that, among the 2654 fire incidents recorded in Futian District, Shenzhen, the distribution of fire causes exhibits marked heterogeneity. Due to incomplete classification in the original open-data records, 1293 incidents (48.72%) were categorized as "Other", reflecting a common limitation in municipal fire databases. Among the classified causes, electrical fires ranked first with 761 incidents (28.67%), followed by accidental domestic fires with 342 incidents (12.89%). This pattern underscores the prominence of electrical infrastructure overload and residential fire-use behaviors as primary preventable risk sources in high-density urban settings.

3.1.2. Global Spatial Autocorrelation Analysis

This study used ArcGIS to conduct a global spatial autocorrelation analysis of the spatial distribution and temporal variation in historical fire kernel density (FRI) in Futian District, Shenzhen. The results of the Global Moran's I analysis indicate that during both 2008–2012 and 2013–2018, the historical fire kernel density in Futian District exhibited a highly significant ($p < 0.01$) positive spatial autocorrelation (clustering) pattern. Specifically, the Global Moran's I value was 0.959 for 2008–2012 and 0.961 for 2013–2018. Before modeling, this Global Moran's I test verified the spatial autocorrelation of fire incidents. The results (Moran's I > 0.95, $p < 0.01$) confirmed a significant spatial clustering pattern, justifying the use of spatial-based machine learning models.

The results (Figure 4) show that during 2008–2012, areas with high annual average fire kernel density were mainly concentrated in the eastern and central–southern parts of Futian District, including Shatou Subdistrict, Futian Subdistrict, Huaqiangbei Subdistrict, and Yuanling Subdistrict, with sporadic distribution also observed in Meilin Subdistrict in the northern part of the district. During 2013–2018, the annual average fire kernel density increased in Shatou Subdistrict and Huaqiangbei Subdistrict, while it decreased in Meilin Subdistrict and Yuanling Subdistrict.

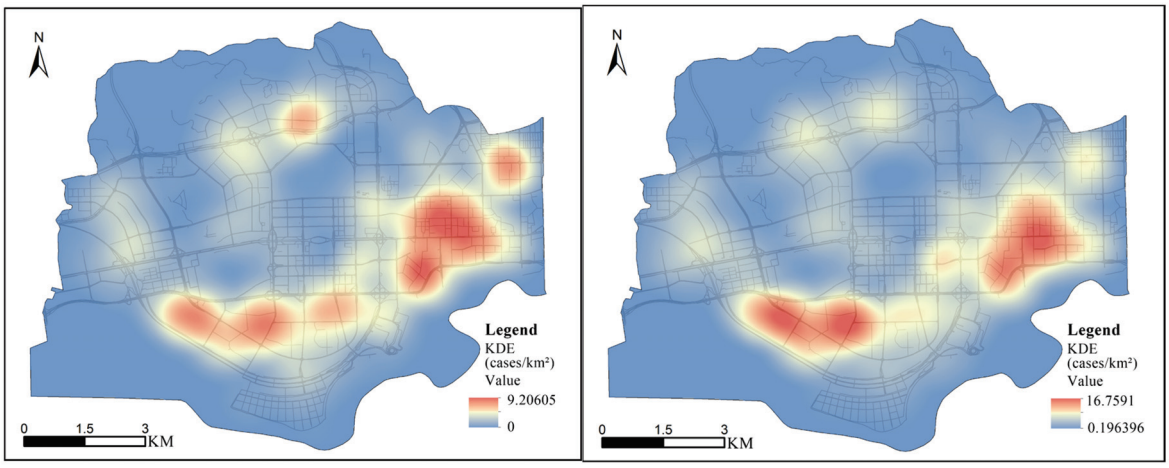


Figure 4. Annual average kernel density of fires in Futian District. (**a**) Mean annual fire kernel density, 2008–2012; (**b**) mean annual fire kernel density, 2013–2018.

https://doi.org/10.3390/land15050796

3.2. Spatial Correlation Analysis of Fire Incidents in Futian District

High/Low clustering analysis using the Getis-Ord General G statistic further indicates that the fire kernel density in the study area exhibited a significant high-value clustering pattern ($p < 0.01$) during both periods, with z-scores of 51.69 and 51.83, respectively.

Hotspot analysis (Getis-Ord Gi*) was applied to identify statistically significant hotspot and cold spot areas. Figure 5 shows that the spatial distribution patterns of hotspots remained highly consistent across the two periods. For 2008–2012, the areas of hotspots, cold spots, and non-significant regions were 19.69 km^2, 21.76 km^2, and 52.16 km^2, respectively. For 2013–2018, the corresponding areas were 17.88 km^2, 20.97 km^2, and 54.77 km^2.

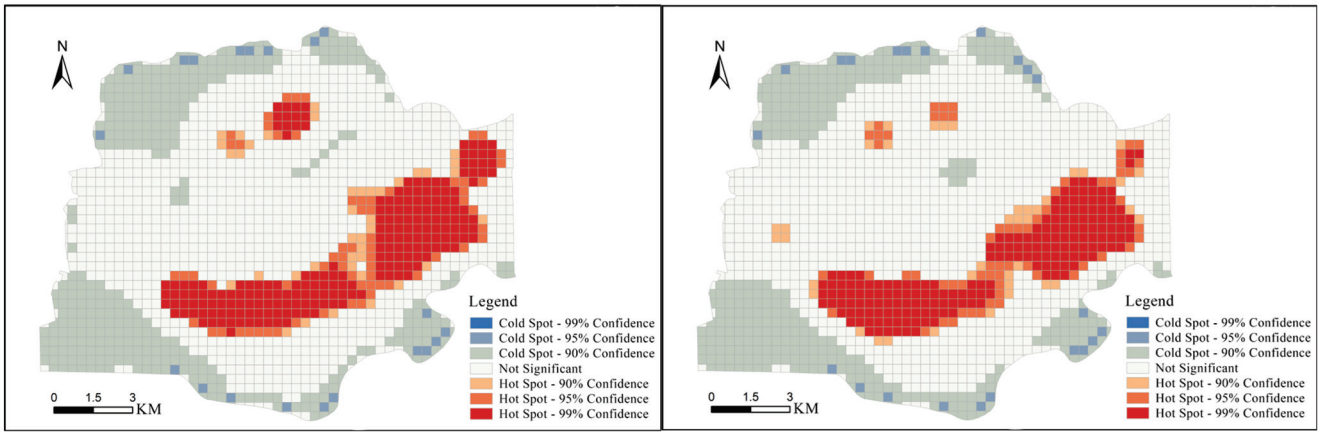


Figure 5. Spatial distribution of kernel density hotspots of fire incidents in Futian District, Shenzhen City. (**a**) Spatial distribution of fire kernel density hotspots in Futian District, 2008–2012; (**b**) spatial distribution of fire kernel density hotspots in Futian District, 2013–2018.

Comparative analysis indicates that both hotspot and cold spot areas showed a decreasing trend, while the area of non-significant regions expanded, accompanied by slight morphological changes in hotspot areas. Hotspots remained concentrated in the eastern and southern parts of the study area, whereas cold spots were mainly distributed in the central and western areas, demonstrating significant spatial heterogeneity.

3.3. Variable Correlation and Multicollinearity Diagnosis

Prior to model construction, a rigorous screening process was applied to the 15 initial explanatory variables to ensure data quality and model robustness. This process involved two steps: (1) examining pairwise correlations to identify weakly related factors, and (2) conducting multicollinearity diagnostics to prevent redundancy among retained predictors.

Figure 6 presents the results based on the Pearson correlation coefficient matrix. Considering the results from both 2008–2012 and 2013–2018, built-environment factors exhibit significant positive correlations with fire kernel density. Among them, the kernel density of vulnerable facilities, residential places, public gathering places, and large-scale buildings forms a core driving cluster.

In addition, building density, nighttime light intensity, and road density further confirm the positive influence of urbanization processes on fire risk. In contrast, distance to fire stations acts as a key inhibitory factor and shows a significant negative effect.

Natural geographic factors such as slope, elevation, annual total precipitation, and distance to water bodies show weak correlations, indicating that topographic and hydrological characteristics have limited explanatory power within the highly urbanized fire risk

https://doi.org/10.3390/land15050796

system of Futian District. Significant multicollinearity exists among variables, which needs to be addressed through the regularization mechanisms of machine learning models.

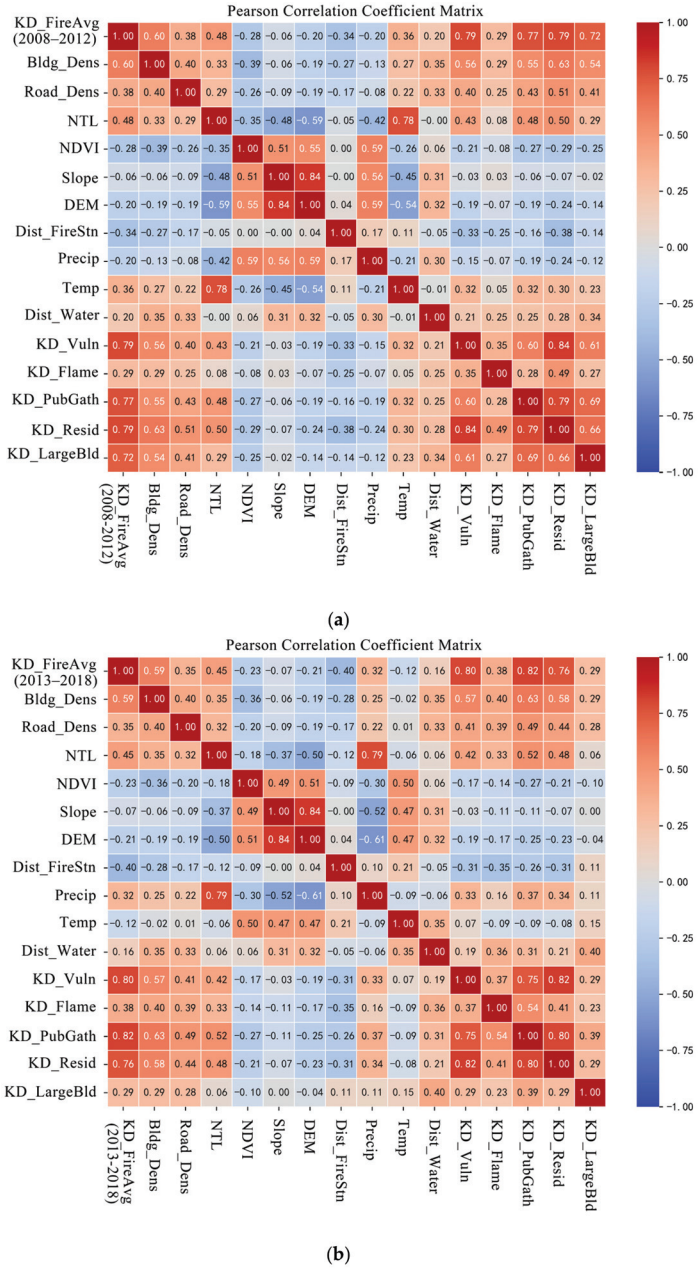


Figure 6. Pearson correlation coefficient heatmap. (**a**) Heatmap of Pearson correlation coefficients, 2008–2012; (**b**) heatmap of Pearson correlation coefficients, 2013–2018. Abbreviations: KD_FireAvg: Annual Average Fire Kernel Density (dependent variable).

Therefore, three weakly correlated factors—slope, annual total precipitation, and distance to water bodies—were excluded to improve the model's generalization ability. While correlation analysis provides an overall understanding of relationships among variables, it cannot directly reveal the actual driving effects of each variable on regional fire kernel density. Further investigation using the CatBoost algorithm and the interpretable method SHAP (SHapley Additive Explanations) is therefore required.

3.4. CatBoost Model Performance

To capture the nonlinear relationships between urban environmental factors and fire risk, a CatBoost regression model was constructed using the 12 screened predictors

(detailed in Section 2.2). To mitigate information leakage induced by strong spatial autocorrelation, conventional random splitting was replaced with a rigorous 7-fold spatial cross-validation framework. The study area was partitioned into geographically distinct clusters using K-means clustering. GroupKFold was then applied to ensure that training and testing sets were spatially non-overlapping, enabling a realistic evaluation of out-of-sample performance. Concurrently, a comprehensive grid search was conducted across hyperparameter spaces (iterations: 800–1200, depth: 4–6, learning rate: 0.02–0.08, subsample: 0.7–0.9, reg_lambda: 1–3, colsample_bylevel: 0.7–0.9) to identify the optimal configuration balancing predictive accuracy and model stability.

As shown in Table 3, the optimized CatBoost model achieved robust performance under spatial validation, with mean R^2 values of 0.7572 ± 0.0464 (2008–2012) and 0.7484 ± 0.1461 (2013–2018). The corresponding RMSE and MAE were 0.8233 ± 0.2792 and 1.3507 ± 0.6894 events/km^2 (2008–2012) and 0.5673 ± 0.2286 and 0.9683 ± 0.4654 events/km^2 (2013–2018), respectively. Although these values are lower than those from random splitting ($R^2 > 0.94$), they avoid spatial leakage and provide a more realistic estimate of model generalization. The relatively narrow standard deviations confirm consistent generalization performance, indicating stable model performance across folds.

Table 3. Seven-fold spatial cross-validation evaluation metrics.

Study Period	RMSE (Events/km^2)	MAE (Events/km^2)	R^2
	Test Set	Test Set	Test Set
2008–2012	0.8233 ± 0.2792	0.5673 ± 0.2286	0.7572 ± 0.0464
2013–2018	1.3507 ± 0.6894	0.9683 ± 0.4654	0.7484 ± 0.1461

To further evaluate the model's predictive accuracy, scatter plots of predicted versus observed fire counts were generated for both study periods (Figure 7). For 2008–2012 (Figure 7a), the points are generally aligned along the 1:1 line, with a linear fit of $y = 0.74x + 0.46$. This indicates strong overall agreement between predictions and observations, although the slope slightly below 1 suggests a modest tendency toward underprediction at higher fire counts. The color gradient (representing predicted values) shows that higher predicted values correspond well to higher observed counts, with limited scatter in the low-value range.

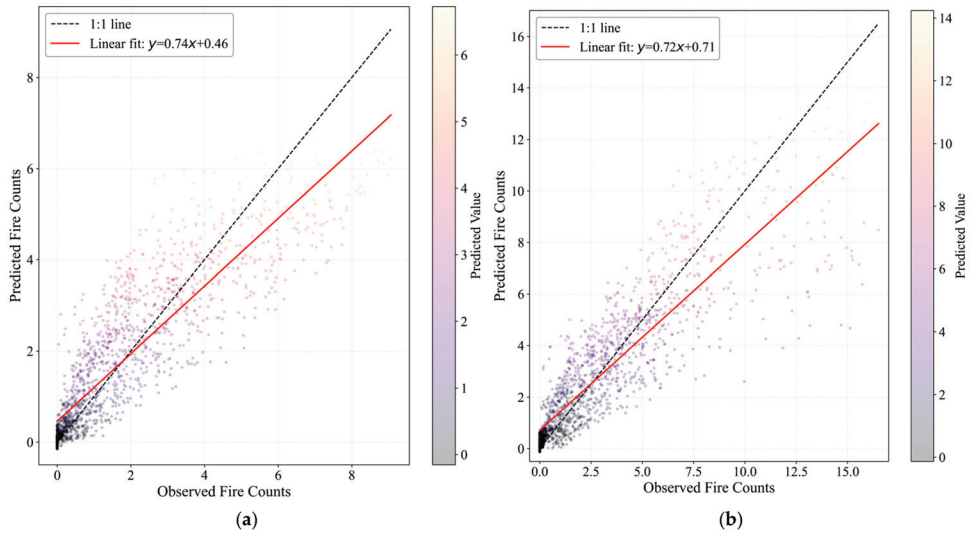


Figure 7. Scatter plots of predicted versus observed fire Counts in Futian District, Shenzhen. (**a**) Model validation results in Futian District, Shenzhen, 2008–2012; (**b**) model validation results in Futian District, Shenzhen, 2013–2018.

https://doi.org/10.3390/land15050796

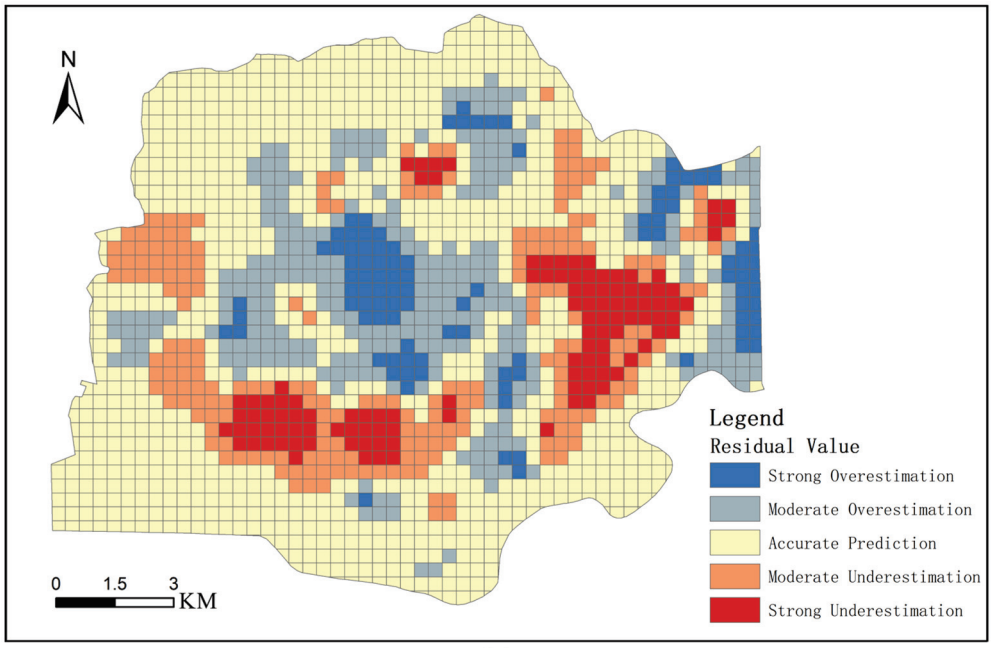
For 2013–2018 (Figure 7b), the relationship remains linear but exhibits greater dispersion, with a fitted line of $y = 0.72x + 0.71$. The model continues to capture the overall trend effectively, yet the slightly lower slope and increased scatter at mid-to-high observed values indicate a more pronounced under-prediction in areas of elevated fire activity compared with the earlier period.

To address potential zero-inflation issues arising from the grid-based structure of the dataset (1596 grid cells), a zero-inflation diagnostic was performed for both study periods (Table 4). In the raw data, 6.3% of grid cells (100/1596) exhibited zero fire risk index (FRI = 0) during 2008–2012, while this proportion increased to 12.7% (202/1596) during 2013–2018. After training the full CatBoost model, the proportion of predicted values near zero (<0.01) dropped substantially to 2.6% (42/1596) in 2008–2012 and 1.9% (31/1596) in 2013–2018. These results indicate that the model does not simply replicate the zero-inflated pattern observed in the raw data. Instead, it successfully captures meaningful spatial variations in fire risk intensity across the study area.

Table 4. Zero-inflation diagnostics for the fire risk index.

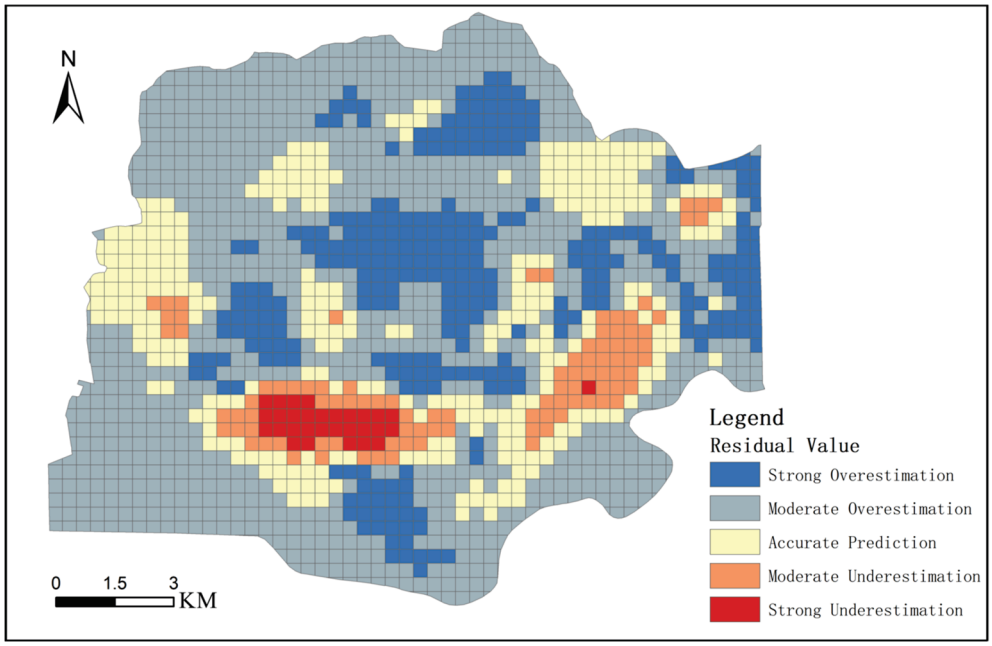
Study Period	Raw Data: Zero FRI (n, %)	Model Predictions: Near-Zero (<0.01) (n, %)
2008–2012	100 (6.27%)	42 (2.63%)
2013–2018	202 (12.66%)	31 (1.94%)

To further evaluate the spatial performance of the model, residuals (observed minus predicted values) were mapped across the 1596 grid cells for the two study periods (Figure 8). For the 2008–2012 period, the residuals are generally centered around zero, indicating a relatively good overall model fit. However, distinct spatial heterogeneity is observed. Areas of underestimation (positive residuals) are mainly concentrated in the central–southern and eastern parts of the study area, forming localized clusters that correspond to high-risk zones identified in the kernel density results. In contrast, overestimation (negative residuals) is primarily distributed in the central region, suggesting that the model tends to slightly over-predict fire risk in areas with moderate fire activity. Overall, the residual pattern indicates that while the model captures the general spatial distribution of fire risk, it exhibits limitations in accurately representing localized high-intensity clusters.



(**a**)

Figure 8. *Cont.*



(**b**)

Figure 8. Spatial distribution of model residuals (observed minus predicted values) across 1596 grid cells in Futian District, Shenzhen. (**a**) Residual patterns and spatial heterogeneity in Futian District, Shenzhen, 2008–2012; (**b**) shift in overestimation and underestimation clusters in Futian District, Shenzhen, 2013–2018.

For the 2013–2018 period, the spatial pattern of residuals exhibits a more pronounced structure. Overestimation becomes more widespread, particularly in the northern and central regions, where large contiguous areas of negative residuals are observed, indicating that the model tends to over-predict fire risk in these zones. Meanwhile, underestimation is mainly concentrated in the central–southern corridor and parts of the eastern region, forming clear linear or clustered patterns. Compared with the earlier period, the spatial distribution of residuals shows a noticeable shift, reflecting the evolving nature of fire risk drivers under changing urban conditions. These results suggest that although the model maintains good overall predictive performance, certain emerging or localized factors—such as intensified human activities or functional land-use changes—may not be fully captured, leading to systematic biases in specific areas.

3.5. Analysis of Influencing Factor Importance

Figures 9 and 10 present the relative importance of factors influencing fire risk in Futian District, Shenzhen, during 2008–2012 and 2013–2018, respectively. In Figures 9a and 10a, the contribution (SHAP value) and relative importance ranking of the 12 influencing factors are shown. Figures 9b and 10b combine factor importance with their influence on fire risk. Each point in the figure represents the predictive effect (SHAP value) of built-environment variables on individual samples, and samples with identical SHAP values are vertically stacked.

The results indicate that during 2008–2012, the key influencing factors of fire risk, ranked by importance, were: KD_Vuln, KD_PubGath, KD_LargeBld, Temp, Dist_FireStn, and KD_Resid. The KD_Vuln ranked first, mainly because such facilities (e.g., schools, hospitals) are critical zones where evacuation is difficult and the presence of vulnerable populations creates significant potential for severe consequences. The high importance of the KD_PubGath and KD_LargeBld is associated with the distribution of facilities in densely populated urban cores, where high human traffic and significant electrical loads create prominent fire hazards. The KD_Resid ranked last among the top factors, largely

because residential areas may be more spatially dispersed or have lower building densities compared to the concentrated commercial and vulnerable facility clusters, resulting in a relatively lower marginal contribution to the prediction.

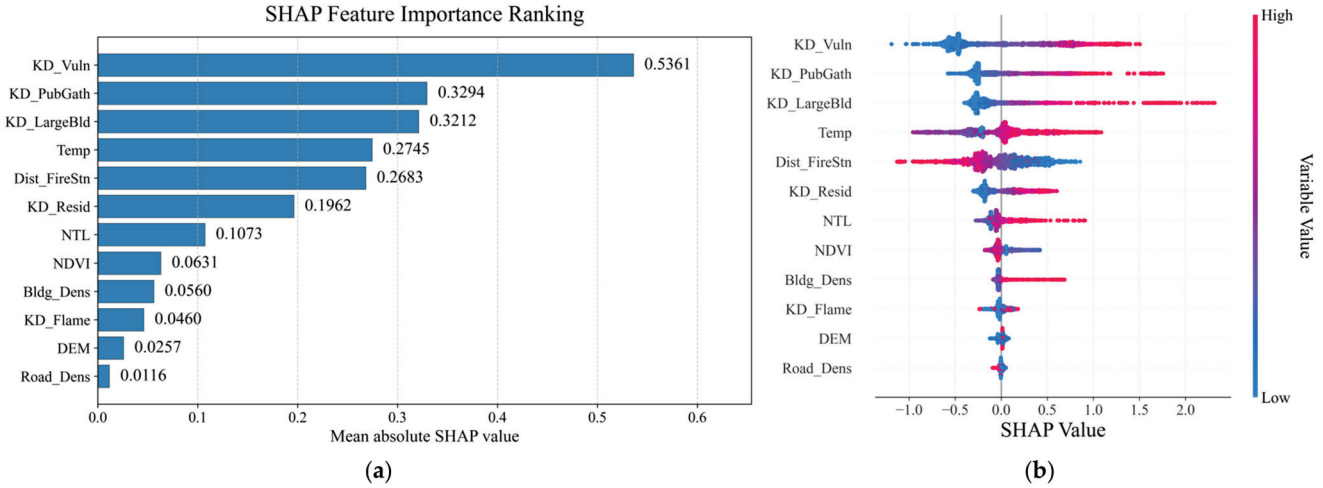


Figure 9. Relative importance of factors influencing fire risk in Futian District, Shenzhen, from 2008 to 2012. (**a**) Mean absolute SHAP values of fire risk factors in Futian District, Shenzhen, 2008–2012; (**b**) SHAP summary plot, 2008–2012.

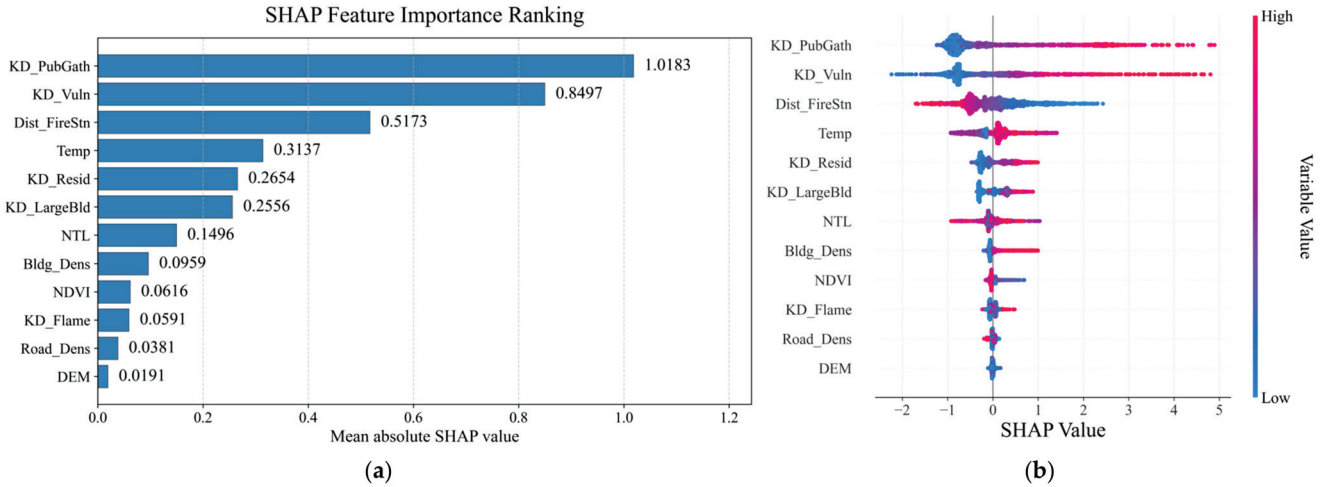


Figure 10. Relative importance of factors influencing fire risk in Futian District, Shenzhen, from 2013 to 2018. (**a**) Mean absolute SHAP values of fire risk factors in Futian District, Shenzhen, 2013–2018; (**b**) SHAP summary plot, 2013–2018.

During 2013–2018, the ranking of influencing factors shifted to: KD_PubGath, KD_Vuln, Dist_FireStn, Temp, KD_Resid, and KD_LargeBld. The KD_PubGath rose to first place, reflecting the continued growth of visitor flows in the Futian central area and the Huaqiangbei Commercial Area, where large populations gather. The KD_Resid increased to fifth place due to accelerated urban renewal processes, increasing residential density and exposing issues such as aging infrastructure in older residential communities. The NTL emerged as a notable factor (ranking seventh), reflecting the growing intensity of nighttime economic activities and the associated increase in electricity use and fire-related scenarios. In contrast, the importance of the KD_LargeBld declined, likely due to improved spatial planning of newly constructed large buildings and stricter fire safety inspection standards.

Overall, KD_PubGath, KD_Vuln, and Dist_FireStn consistently remain the core influencing factors across both periods. This indicates a shift in dominant fire risk drivers from building expansion (2008–2012) to population concentration and activity intensity

https://doi.org/10.3390/land15050796

(2013–2018). Meanwhile, NDVI, Bldg_Dens, Road_Dens, and DEM all exhibit relatively low mean absolute SHAP values during both periods, indicating limited influence on urban fire risk.

3.6. Feature Interaction Analysis

To quantitatively address the reviewers' concerns regarding feature interactions, the SHAP interaction value matrices and the top 10 dominant interaction pairs for both study periods were computed and summarized (Figure 11 and Table 5). The results indicate a substantial intensification and structural shift in synergistic effects over time. In the 2008–2012 period, the interaction between Dist_FireStn and Temp was the primary driver (0.0581). However, by 2013–2018, the interaction strength increased substantially, with the synergy between KD_PubGath and KD_Vuln emerging as the dominant factor (0.1399). Furthermore, significant interactions were observed between NTL and socio-economic densities (KD_PubGath and KD_Vuln), supporting the presence of complex nonlinear interactions between urban development intensity and fire risk vulnerability in high-density areas.

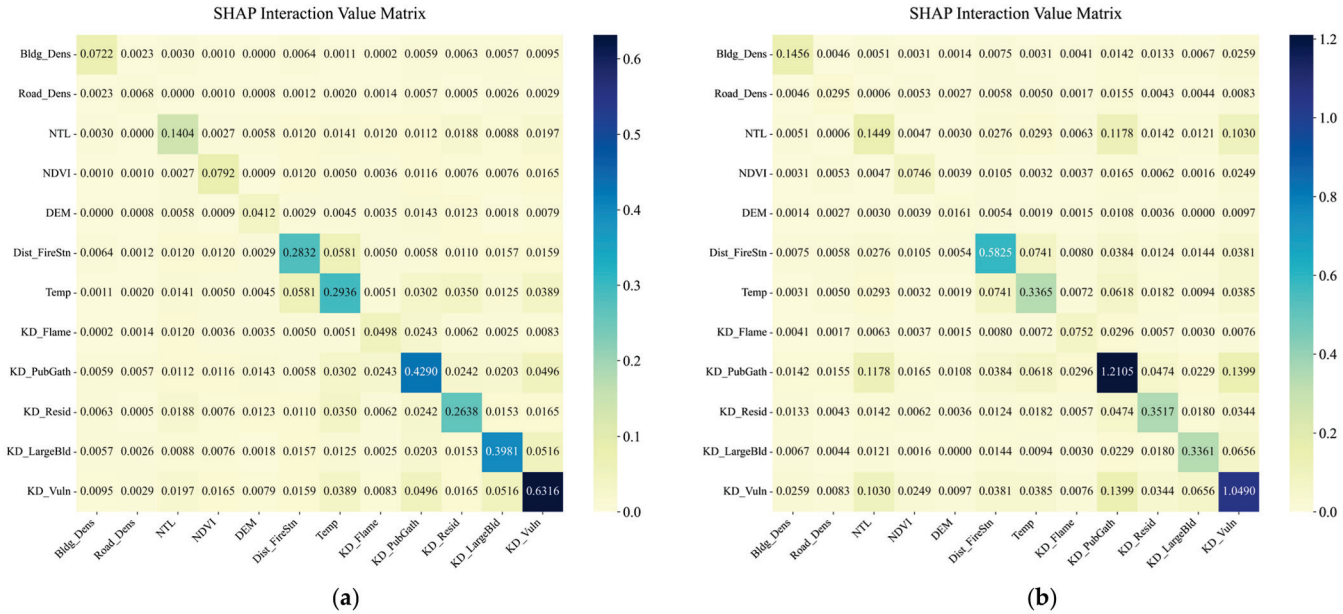


Figure 11. SHAP interaction value matrices for urban fire risk factors: (**a**) 2008–2012; (**b**) 2013–2018.

Table 5. Top 10 pairwise feature interactions based on mean absolute SHAP interaction values.

Rank	2008–2012 Interaction Pair	Value	2013–2018 Interaction Pair	Value
1	Dist_FireStn × Temp	0.0581	KD_PubGath × KD_Vuln	0.1399
2	KD_LargeBld × KD_Vuln	0.0516	NTL × KD_PubGath	0.1178
3	KD_PubGath × KD_Vuln	0.0496	NTL × KD_Vuln	0.103
4	Temp × KD_Vuln	0.0389	Dist_FireStn × Temp	0.0741
5	Temp × KD_Resid	0.035	KD_LargeBld × KD_Vuln	0.0656
6	Temp × KD_PubGath	0.0302	Temp × KD_PubGath	0.0618
7	KD_Flame × KD_PubGath	0.0243	KD_PubGath × KD_Resid	0.0474
8	KD_PubGath × KD_Resid	0.0242	Temp × KD_Vuln	0.0385
9	KD_PubGath × KD_LargeBld	0.0203	Dist_FireStn × KD_PubGath	0.0384
10	NTL × KD_Vuln	0.0197	Dist_FireStn × KD_Vuln	0.0381

3.7. Nonlinear Relationships Between Features and Regional Fire Risk

We further examined the nonlinear relationships between influencing factors and regional fire risk by analyzing scatter plots combined with locally weighted regression

(LOWESS) fitting curves. Colors represent the feature with the strongest interaction effect in each sample. To ensure statistical rigor, critical thresholds were identified using a bootstrap-based piecewise linear fitting method (pwlf package, 1000 bootstrap resamples). The identified breakpoints and 95% confidence intervals (95% CI) are reported below. Figures 12 and 13 show the nonlinear relationships between environmental factors and fire risk in Futian District, Shenzhen, during 2008–2012 and 2013–2018, respectively.

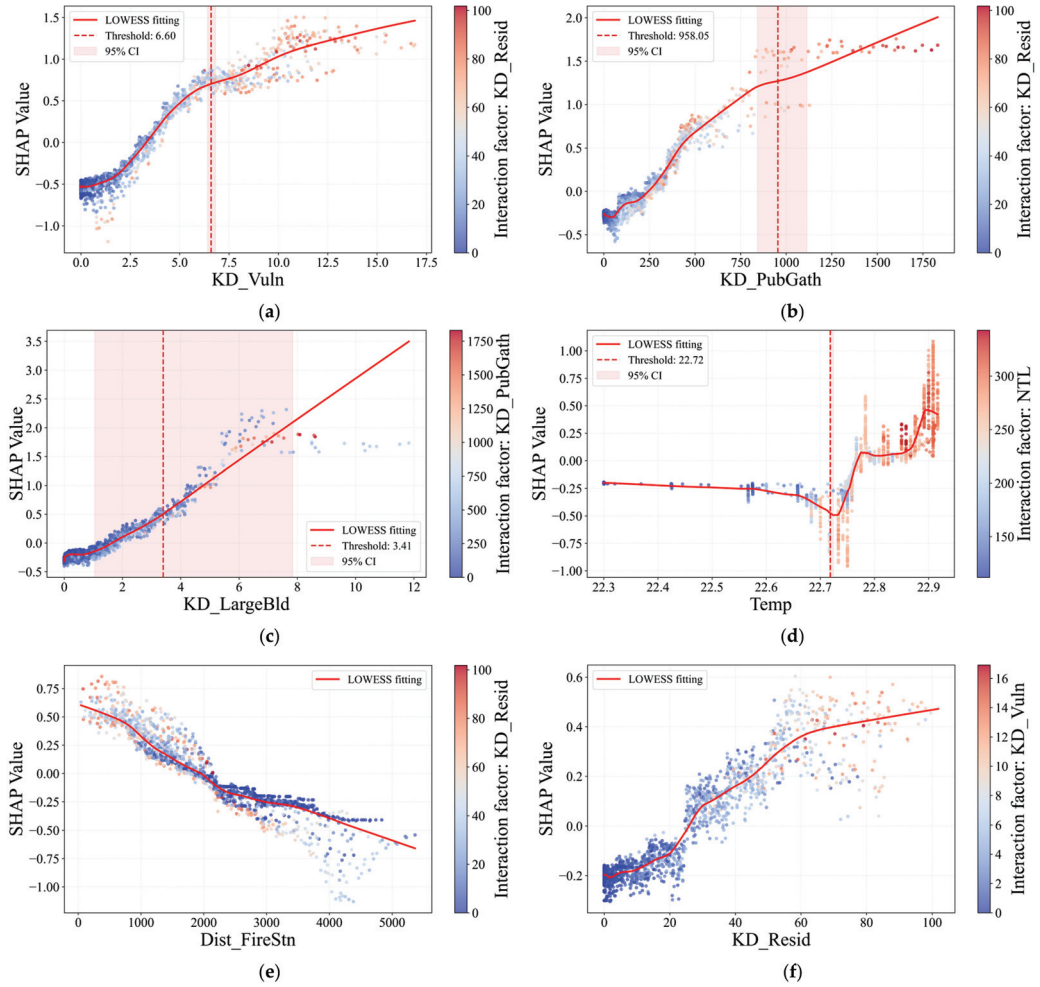


Figure 12. Nonlinear relationships between core environmental factors and fire risk in Futian District, Shenzhen, 2008–2012. (**a**) KD_Vuln; (**b**) KD_PubGath; (**c**) KD_LargeBld; (**d**) Temp; (**e**) Dist_FireStn; (**f**) KD_Resid.

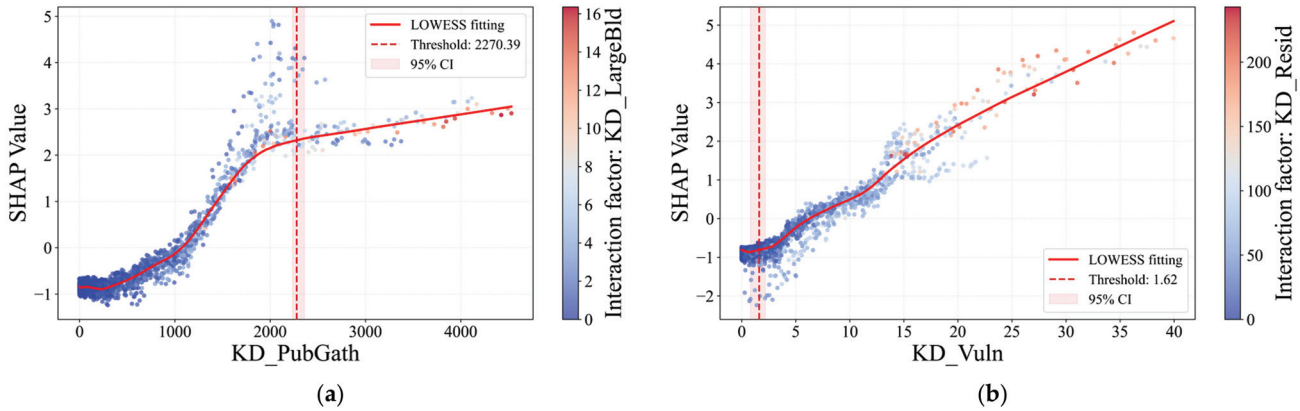


Figure 13. *Cont.*

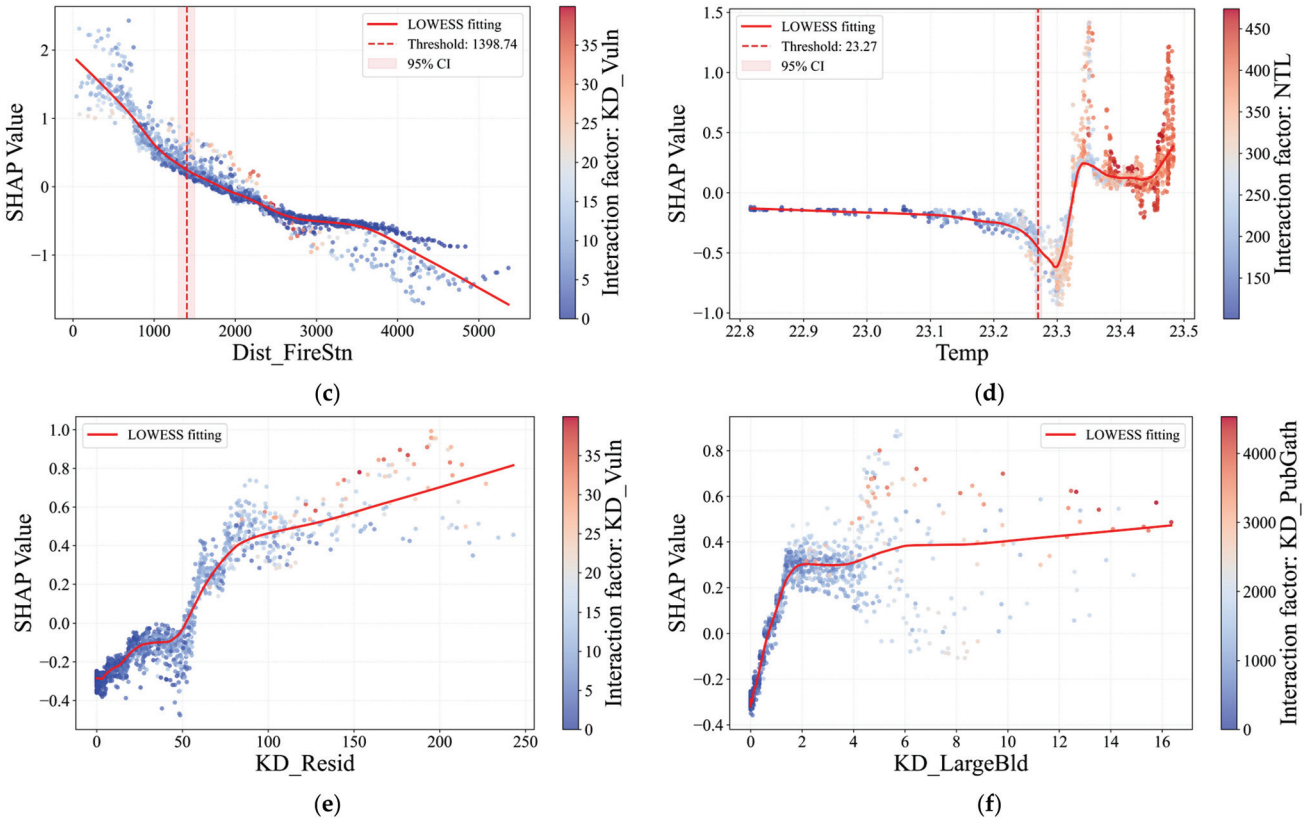


Figure 13. Nonlinear relationships between core environmental factors and fire risk in Futian District, Shenzhen, 2013–2018. (**a**) KD_PubGath; (**b**) KD_Vuln; (**c**) Dist_FireStn; (**d**) Temp; (**e**) KD_Resid; (**f**) KD_LargeBld.

In Figures 12 and 13, the threshold (red dashed line) and 95% confidence interval (pink shading) are displayed only for those variables exhibiting a distinct change in the direction or slope of the LOWESS-fitted relationship.

The results for 2008–2012 (Figure 12) show that KD_Vuln (a) is positively correlated with SHAP values: when the kernel density exceeds 6.60 POI/km^2, it begins to exert a positive effect on FRI; the effect peaks at approximately 8 POI/km^2, with SHAP values stabilizing at around 1.5. A strong positive interaction exists between KD_Vuln and KD_Resid; when the latter is high, the positive contribution of KD_Vuln to SHAP values is further amplified.

KD_PubGath (b) shows a positive association with SHAP values. When the density exceeds 958.05 POI/km^2, SHAP values turn positive, indicating a positive effect on FRI; the growth rate slows after exceeding approximately 1000 POI/km^2. This variable also exhibits a positive interaction with KD_Resid: higher residential density corresponds to higher SHAP values.

KD_LargeBld (c) shows an overall positive correlation with SHAP values. When the density is below 3.41 POI/km^2, SHAP values are negative, indicating a negative effect on FRI; once it exceeds 3.41 POI/km^2, SHAP values become positive, and the growth rate remains strong without significant slowing. This variable also shows a positive interaction with KD_PubGath, with a strong amplification effect in medium- and high-density ranges.

Temp (d) shows that when the temperature is less than 22.72 °C, SHAP values are negative, indicating a negative influence on FRI. When the temperature exceeds 22.72 °C, SHAP values turn positive, suggesting a clear threshold effect. This pattern is likely related to spatial thermal heterogeneity within the study area. Specifically, coastal zones tend to maintain relatively lower temperatures due to the thermal regulation effect of adjacent

water bodies, which contributes to reduced fire occurrence. In contrast, inland urban areas with higher temperatures are more likely to experience intensified FRI, potentially driven by the urban heat island effect, potentially increasing energy demand and reducing local moisture conditions.

In addition, Temp is positively correlated with NTL: areas with higher temperature generally correspond to higher light intensity, while lower temperature areas exhibit relatively lower nighttime light levels.

Dist_FireStn (e) shows that when the distance is less than 2000 m, SHAP values are positive, indicating a positive influence on FRI. When the distance exceeds 2000 m, SHAP values become negative. Dist_FireStn is negatively correlated with KD_Resid: areas near fire stations generally have higher KD_Resid (POI/km^2), while more distant areas show lower KD_Resid levels.

KD_Resid (f) shows a clear positive correlation with SHAP values. When the density is below approximately 25 POI/km^2, SHAP values are predominantly negative or close to zero, indicating a limited or slightly suppressing effect on FRI. Once the density exceeds this threshold, SHAP values turn positive and increase steadily, suggesting a significant enhancement of FRI. This pattern indicates a distinct threshold effect. Furthermore, higher Temp further strengthens this positive contribution, as reflected by the color gradient in the figure, where higher-temperature observations are associated with larger SHAP values at similar density levels. This interaction may be attributed to intensified human activity and elevated thermal conditions. High residential density concentrates population and daily activities. Under higher temperatures, increased energy use, reduced moisture, and more frequent ignition sources jointly contribute to higher FRI.

The results for 2013–2018 (Figure 13) show that KD_PubGath (a) is positively correlated with SHAP values: when the density is below the bootstrap-derived breakpoint of 2270.39 POI/km^2, SHAP values are negative, indicating a negative effect on FRI; once it exceeds 2270.39 POI/km^2, SHAP values become positive, and growth slows beyond ~2300 POI/km^2. A strong interaction is observed between KD_PubGath and KD_LargeBld; when the latter is high (as indicated by redder points), the positive contribution of KD_PubGath to SHAP values is further amplified.

KD_Vuln (b) is positively correlated with SHAP values: when the density is below the bootstrap-derived breakpoint of 1.62 POI/km^2, SHAP values are negative, indicating a negative effect on FRI; once it exceeds 1.62 POI/km^2, SHAP values turn positive, and SHAP values increase substantially with higher densities. A strong interaction is observed between KD_Vuln and KD_Resid; when the latter is high (as indicated by redder points), the positive contribution of KD_Vuln to SHAP values is further amplified.

Dist_FireStn (c) shows that when the distance is less than the bootstrap-derived breakpoint of 1398.74 m, SHAP values are positive, indicating a positive influence on FRI. When the distance exceeds 1398.74 m, SHAP values become negative. Dist_FireStn is negatively correlated with KD_Vuln: areas near fire stations generally have higher KD_Vuln (POI/km^2), while more distant areas show lower KD_Vuln levels.

Temp (d) shows negative SHAP values when the temperature is below the bootstrap-derived breakpoint of 23.27 °C, indicating a negative influence on FRI; once it exceeds 23.27 °C, SHAP values become positive, indicating a clear threshold effect. This phenomenon is closely related to the interaction between Temp and NTL; higher NTL values (as indicated by redder points) strengthen the positive effect of Temp on SHAP values, whereas lower NTL values reinforce its negative effect, reflecting the synergistic amplification of fire risk by elevated temperature and intense anthropogenic activity in urban environments.

KD_Resid (e) is positively correlated with SHAP values. SHAP values rise rapidly from negative to positive as the density increases, indicating a transition from negative

https://doi.org/10.3390/land15050796

to positive effect on FRI; once the density is higher, SHAP values become positive, and the growth rate slows significantly, stabilizing at around 0.8. A strong positive interaction exists between KD_Resid and KD_Vuln; when the latter is high (as indicated by redder points), the positive contribution of KD_Resid to SHAP values is further amplified.

KD_LargeBld (f) shows negative SHAP values in the range of 0–0.5 POI/km^2, indicating a negative influence on FRI; once the density exceeds 0.5 POI/km^2, SHAP values turn positive. This phenomenon is also related to the spatial concentration of high-fire-load structures and urban density variations. Higher densities of KD_PubGath strengthen the positive effect of KD_LargeBld on SHAP values, whereas lower densities reinforce its negative or neutral effect in less developed areas.

Comparing the results for 2008–2012 (Figure 12) and 2013–2018 (Figure 13) reveals both notable similarities and clear dynamic shifts in the nonlinear relationships between the core environmental factors and regional fire risk.

In terms of similarities, the six key factors—KD_PubGath, KD_Vuln, KD_Resid, Dist_FireStn, Temp, and KD_LargeBld—remain central to fire risk prediction across both periods. Their relationships with SHAP values consistently exhibit a threshold-dependent pattern: SHAP values are predominantly negative (or near zero) below the identified breakpoint and turn positive above it, indicating a shift from risk suppression to risk amplification. Significant interaction effects among variables are also evident in both periods, with color gradients highlighting how one factor modulates the influence of another (e.g., KD_Resid amplifying KD_Vuln and KD_PubGath; Temp interacting with NTL).

In terms of differences, two major changes stand out. First, the critical thresholds display substantial temporal evolution. The positive-effect breakpoint for KD_PubGath rose sharply from 958.05 POI/km^2 to 2270.39 POI/km^2 (with growth slowing after approximately 2300 POI/km^2); the threshold for KD_Vuln decreased from 6.60 POI/km^2 to 1.62 POI/km^2; the threshold for KD_Resid shifted to a substantially higher density range where the growth rate slows significantly once density is higher; the breakpoint distance for Dist_FireStn decreased from 2000 m to 1398.74 m; the temperature threshold rose modestly from 22.72 °C to 23.27 °C; and the breakpoint for KD_LargeBld fell from 3.41 POI/km^2 to 0.5 POI/km^2. These shifts reflect changing urban conditions, such as intensified population density and evolving spatial patterns.

Second, interaction patterns and the role of auxiliary factors strengthened in the later period. The positive interaction between KD_PubGath and KD_LargeBld became more pronounced, while the synergistic amplification between Temp and NTL grew more explicit: higher NTL values notably reinforced the positive effect of temperature on fire risk. In addition, the overall importance ranking shifted (with KD_PubGath rising to first place and NTL emerging as a notable seventh-ranked factor in 2013–2018), underscoring a transition from building-expansion-driven risk to population-concentration- and activity-intensity-driven risk.

These evolving feature dependence patterns demonstrate that the influence of built-environment factors on urban fire risk is inherently dynamic rather than static. Risk assessment and governance strategies should therefore incorporate temporal context and adopt period-specific adjustments to remain effective.

3.8. Local Diagnostics and Spatiotemporal Mechanism Evolution

To further unravel the non-linear driving mechanisms and their temporal evolution, this section conducts a comparative local SHAP analysis on two representative hotspots—Huaqiangbei (a premier commercial hub) and Shatou (a typical urban village/mixed-use area)—across the 2008–2012 and 2013–2018 periods. The base value, representing the average predicted risk, remains stable at 1.732 across both periods ac-

https://doi.org/10.3390/land15050796

cording to the latest model training, providing a consistent benchmark for evaluating how specific local features deviate from the global mean to determine the final Fire Risk Index (FRI). The detailed feature contributions and their spatiotemporal shifts, visualized in the 2 × 2 waterfall diagnostic matrix in Figure 14, reveal distinct evolutionary trajectories for these two urban functional zones.

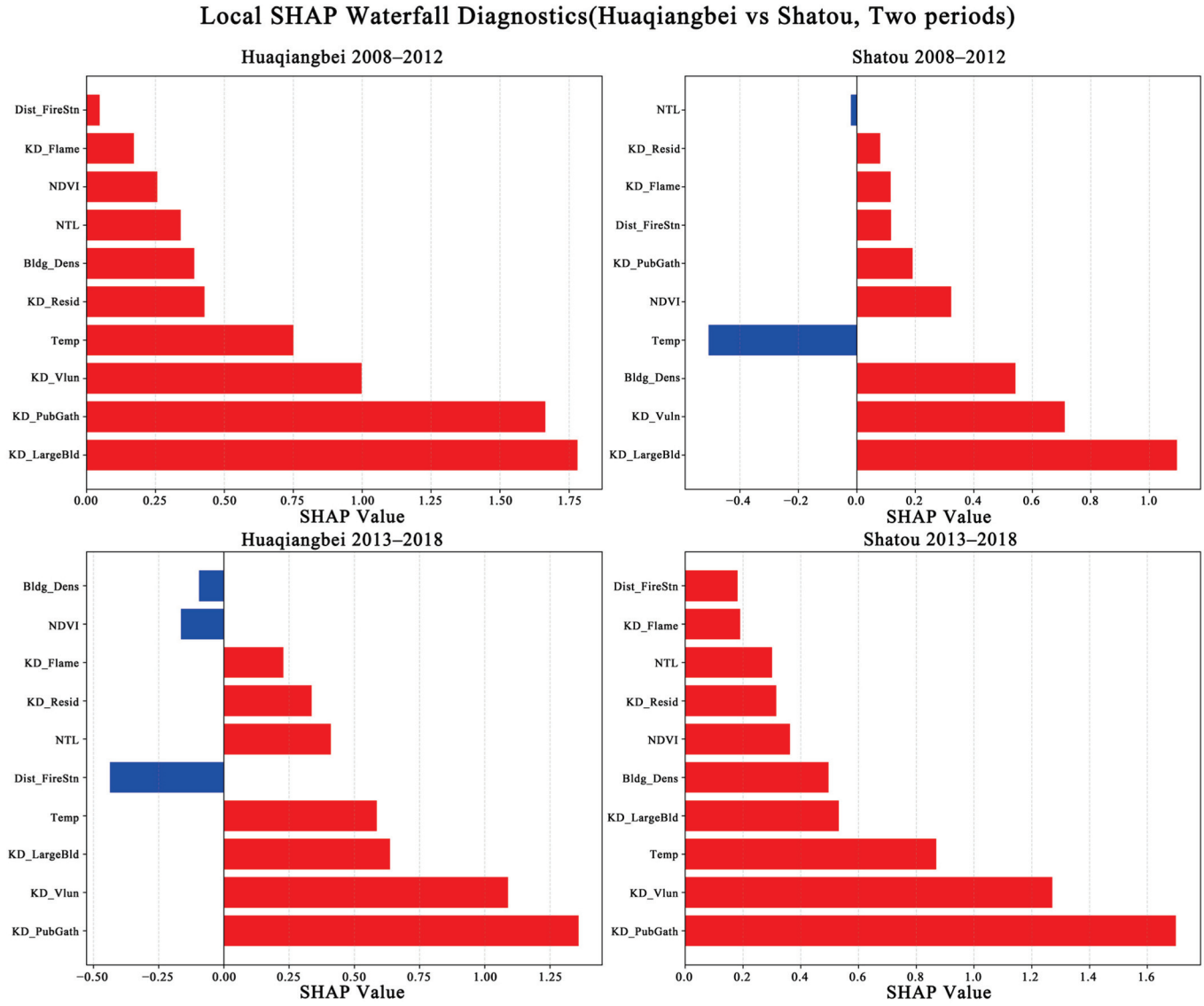


Figure 14. Local SHAP waterfall diagnostics (Huaqiangbei vs. Shatou, Two Periods).

3.8.1. Huaqiangbei: Paradigm Shift from Activity-Driven to Vulnerability-Dominated Risk

In Huaqiangbei, the local diagnostics reveal a clear mechanistic transition reflecting its evolution from a rapid wholesale expansion phase to a mature, managed commercial core.

During the 2008–2012 period, the risk was overwhelmingly driven by the sheer scale of the built environment and intense commercial pressure, where KD_LargeBld contributed the most significant positive SHAP value (approximately +1.78), followed by KD_PubGath (+1.66) and KD_Vuln (+1.00). This early phase underscores how massive building volumes and high functional density directly dictated the fire load in a burgeoning market district.

By the 2013–2018 period, while the area remained high-risk, the dominant drivers shifted toward the interplay of social vulnerability and persistent functional activity. KD_PubGath (+1.35) and KD_Vuln (+1.08) continued to be the primary risk enhancers, but the influence of physical building scale (KD_LargeBld) diminished notably to +0.63.

https://doi.org/10.3390/land15050796

Crucially, Dist_FireStn emerged as a powerful risk-reducing factor with a negative SHAP value of −0.43, indicating that the maturation of the district has been accompanied by more effective fire service integration and localized safety management. The emergence of slight negative contributions from NDVI and Bldg_Dens further suggests that modern urban renewal and density controls have begun to exert a synergistic mitigating effect on the fire risk profile.

3.8.2. Shatou: Intensification of Infrastructure Vulnerability Synergy in Urban Villages

In stark contrast to the adaptive transition observed in Huaqiangbei, Shatou exemplifies a scenario of risk intensification and structural inertia typical of informal high-density settlements.

During the 2008–2012 period, the risk mechanism was relatively fragmented; although KD_LargeBld (+1.10) and KD_Vuln (+0.71) were the leading drivers, the impact of environmental factors like Temp was actually negative (−0.50), reflecting a different socio-spatial equilibrium in the early stage.

However, by the 2013–2018 period, the diagnostic mechanism underwent a dramatic escalation, characterized by a "vulnerability–infrastructure gap" loop. KD_PubGath surged to become the dominant risk driver with a SHAP contribution of +1.72, while KD_Vuln's impact intensified to +1.27, highlighting the extreme pressure of population crowding within limited spatial confines. Most significantly, the role of Temp flipped to a strong positive contribution (+0.87), suggesting that environmental thermal stress now acts as a critical catalyst for fire risk in the poorly ventilated, high-density fabric of the urban village. Furthermore, Dist_FireStn maintained a positive contribution (+0.18), confirming that Shatou remains functionally isolated from effective fire response thresholds.

These findings emphasize that while general city governance has improved, the "structural traps" of urban villages like Shatou necessitate place-specific, infrastructure-first interventions—such as decentralized micro-fire stations—to break the cycle of escalating risk.

Overall, the local SHAP trajectories across the two periods and two functional zones reveal substantial spatiotemporal heterogeneity: Huaqiangbei demonstrates adaptive risk transformation through functional maturation, whereas Shatou underscores structural inertia requiring place-specific, infrastructure-first remedies. These findings underscore the necessity of dynamic, grid-level "precision fire governance" rather than uniform district-wide strategies in megacity redevelopment.

4. Discussion

4.1. Spatiotemporal Evolution and the Paradigm Shift in Risk Drivers

The findings show a clear spatiotemporal differentiation of fire risk within Futian District, demonstrating a gradual shift from physical–structural dominance toward activity-related patterns. During the first period (2008–2012), high-risk hotspots were primarily concentrated in the eastern and southern parts of the district, covering approximately 19.69 km^2. These clusters were mainly associated with the density of large-scale buildings (KD_LargeBld) and vulnerable populations (KD_Vuln), which corresponds to a stage in which fire risk was closely related to rapid built-up expansion and uneven infrastructure conditions.

In the second period (2013–2018), the total hotspot area decreased to 17.88 km^2, which is likely related to improvements in urban governance and the implementation of policies such as the "Refined Urban Renewal" initiative introduced around 2012. Despite the overall reduction, Nighttime Light (NTL) and the density of public gathering places (KD_PubGath) became more prominent in the model results, indicating that fire risk is

https://doi.org/10.3390/land15050796

increasingly associated with patterns of economic activity and human use of space. In high-density urban environments such as Shenzhen, fire risk is linked not only to the spatial distribution of buildings but also to the intensity and timing of urban activities, particularly in areas characterized by active nighttime economies and mixed commercial functions. This observation is consistent with previous studies emphasizing the relationship between land use intensity and urban safety resilience, while also indicating that the relative importance of risk factors varies across different stages of urban development.

4.2. Decoupling Non-Linear Thresholds Through XAI (Explainable AI)

An important aspect of this study is the use of the CatBoost–SHAP framework with 7-fold spatial cross-validation to examine nonlinear relationships between urban environmental factors and fire risk. Compared with conventional models, this approach allows for a more detailed characterization of how the effects of individual variables vary across different value ranges. For example, the analysis of temperature (Temp) shows the presence of threshold-like behavior. The estimated critical values increased from approximately 22.72 °C in 2008–2012 to 23.27 °C in 2013–2018 (with 95% CI derived from 1000 bootstrap resamples). Beyond these ranges, the marginal contribution of temperature to predicted fire risk becomes more pronounced, which is likely related to combined effects such as increased energy demand and intensified urban heat conditions.

A similar nonlinear pattern is observed for emergency response accessibility. The SHAP results indicate that the distance to fire stations (Dist_FireStn) is associated with fire risk in a nonlinear manner, with a noticeable change in marginal effects around 1398.74 m. The adoption of spatial cross-validation effectively mitigates information leakage caused by strong spatial autocorrelation, thereby enhancing the reliability of these SHAP-derived thresholds. Beyond this distance, the contribution to predicted risk increases more rapidly. At the same time, this effect appears to be moderated in areas with higher densities of public gathering places (KD_PubGath), which is likely linked to the presence of fire safety facilities and stricter building regulations in large commercial complexes. These findings show that fire risk is shaped by the combined influence of spatial configuration and functional characteristics, and that uniform planning standards may not fully capture such variability. In this sense, the results provide an empirical complement to recent studies that emphasize perceptual or qualitative assessments of urban safety, by offering a quantitative perspective on how nonlinear responses of physical land-use variables relate to fire risk patterns [49].

4.3. Comparative Analysis with Existing Literature

Compared with traditional fire risk studies, this research extends previous findings on the relationship between population density and fire occurrence. While earlier work has often emphasized a linear association between density and fire frequency, the present analysis reveals a more complex pattern in which the effects of activity-related density (e.g., KD_PubGath) vary under different environmental conditions. For instance, public gathering places do not appear to uniformly increase fire risk; instead, their contribution becomes more pronounced under higher temperature conditions, indicating an interaction between human activity intensity and climatic stress.

In addition, while prior studies have highlighted the role of proximity to fire stations as a key factor in reducing fire risk, the results in this study suggest a more heterogeneous spatial pattern. Although accessibility to firefighting services remains important, areas such as Shatou and Huaqiangbei continue to exhibit relatively high risk levels, even with improved infrastructure. This shows that, in highly dense urban environments, physical proximity alone may not fully account for variations in fire risk.

https://doi.org/10.3390/land15050796

More broadly, these findings are consistent with recent research emphasizing the importance of integrating multiple dimensions—such as socio-economic activity and environmental conditions—into urban risk assessment. At the same time, the present study highlights the value of considering nonlinear effects and feature interactions, which are often overlooked in conventional approaches. By incorporating these aspects, the analysis provides a more nuanced understanding of how different factors are associated with fire risk in complex urban systems [50,51].

5. Policy Implications

5.1. Precision Governance for "Dual-Core" High-Density Clusters

The persistence of high-risk clusters in Shatou and Huaqiangbei necessitates a transition from "one-size-fits-all" firefighting strategies to zone-specific, precision governance.

For Huaqiangbei (Commercial Core): Given that KD_PubGath and NTL are the primary risk drivers, land management departments should implement IoT-based (Internet of Things) real-time monitoring of electrical loads and population flow. Safety audits should focus on the internal management of commercial POIs rather than just building-scale structural inspections.

For Shatou (Mixed-use/Urban Village): The focus should remain on bridging the 1398.74 m service gap for emergency response. In these dense quarters where traditional fire trucks may struggle to navigate, the deployment of "Micro-Fire Stations" and decentralized water supply systems is essential to maintain the "Golden 4 min" response time within the 1398.74 m service gap identified by the model.

5.2. Integrating Urban Heat Mitigation with Fire Prevention

The identified thermal thresholds (23.27 °C, derived from SHAP dependence plots under spatial cross-validation) provide a scientific imperative for integrating fire safety into climate adaptation strategies. Urban planners should prioritize "Nature-based Solutions" (NbS) in high-risk zones to mitigate the urban heat island effect. Strategies such as increasing green roof coverage, promoting permeable pavements, and creating urban wind corridors can lower local micro-environmental temperatures. By maintaining ambient temperatures below the 23.27 °C critical limit, cities can indirectly reduce the probability of heat-induced spontaneous combustion and electrical failures, thereby enhancing the overall resilience of the urban land system.

5.3. Cause-Specific Interventions and Targeted Governance

The cause-level analysis presented in Section 3.1.1 further refines these policy priorities by linking dominant fire typologies to targeted interventions. Given that electrical fires account for 28.67% of all incidents and strongly correlate with the SHAP-identified importance of residential density, fire prevention strategies should prioritize aging residential communities and urban villages. Targeted measures should include mandatory upgrades to outdated electrical wiring, standardized deployment of safe charging facilities for electric bicycles, and community-based fire safety awareness campaigns. Similarly, the prominence of accidental domestic fires (12.89%) alongside the high-risk contribution of public gathering places underscores the necessity for stringent, routine fire safety inspections in high-occupancy venues, coupled with mandatory emergency evacuation drills and smart early-warning systems.

Furthermore, the observed interaction between ambient temperature and built-environment factors indicates that fire risk management should adopt dynamic, seasonally adaptive protocols. During dry, high-temperature periods (particularly winter months with low humidity and peak summer periods), municipal authorities should intensify public

https://doi.org/10.3390/land15050796

outreach, temporarily restrict high-risk ignition activities (e.g., fireworks), and pre-position mobile firefighting resources in historically vulnerable sub-districts. This cause-informed, temporally responsive governance framework ensures that urban fire safety strategies move beyond static spatial zoning to directly address the root ignition mechanisms and environmental triggers of fire incidents.

6. Conclusions

This study develops an interpretable machine learning framework to examine the spatial patterns and influencing factors of urban fire risk in a high-density urban context. By integrating the CatBoost model with SHAP analysis, the study captures nonlinear relationships and provides a detailed interpretation of how different factors influence fire risk distribution.

The results show that urban fire risk is closely related to the spatial concentration of population-related facilities, including commercial, residential, and vulnerable-use spaces. These factors exhibit heterogeneous effects across different density levels, with several variables showing evident threshold characteristics. In addition, climatic conditions, particularly temperature, are associated with variations in fire risk, although their effects are not strictly linear. Interaction analysis further indicates that the combined presence of built environment features and thermal conditions reinforces or moderates fire risk patterns in specific areas.

Overall, this study demonstrates the effectiveness of interpretable machine learning methods in urban risk analysis and provides a useful reference for understanding complex spatial phenomena in high-density cities.

6.1. Limitations

It should be noted that this study focuses on a single high-density urban district, and the generalizability of the findings to other cities requires further validation. In addition, the use of POI data and kernel density estimation introduces certain uncertainties related to data completeness and spatial smoothing.

Furthermore, the use of aggregated POI-based indicators (e.g., commercial facilities) rather than more fine-grained classifications may mask potential heterogeneity in fire risk associated with different subtypes of urban activities. Due to limitations in data availability, this study adopts POI kernel density as a proxy for urban functional intensity, which may not fully capture all dimensions of land-use heterogeneity.

In addition, the use of historical POI snapshots (2012 and 2018) may introduce minor uncertainties related to positional shifts in individual POIs over time, although this approach helps ensure temporal consistency in matching.

Moreover, the raw fire records provided by the Shenzhen Municipal Government Open Data Platform do not include standardized information on fire severity levels. The transition from conventional random splitting to rigorous 7-fold spatial cross-validation further improves model robustness against spatial autocorrelation, although it results in moderately lower R^2 values compared with random partitioning, reflecting more realistic out-of-sample performance in geographically heterogeneous urban environments.

6.2. Future Research Directions

Future research could incorporate multi-source and higher-resolution datasets, including more detailed land-use classifications, business-type data, and environmental variables, to further improve the granularity of fire risk analysis.

In addition, sensitivity analyses based on multi-temporal POI datasets could be conducted to better account for potential spatial and temporal uncertainties in POI data. Com-

https://doi.org/10.3390/land15050796

parative studies across different urban contexts would also help evaluate the robustness and transferability of the findings.

Furthermore, future work could integrate more detailed fire incident records, including severity levels and internal operational data from fire departments, to enable more refined analyses of fire risk mechanisms across different fire intensities.

Author Contributions: Conceptualization, H.L.; methodology, Y.W.; validation, Y.W.; formal analysis, Y.W.; investigation, Y.W.; resources, H.L.; data curation, Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, H.L. and Y.W.; visualization, Y.W.; supervision, H.L.; project administration, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Major Project of Philosophy and Social Science Research in Higher Education Institutions of Henan Province (2025-YYZD-12).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

DEM	Elevation
Slope	Slope
NDVI	Normalized Difference Vegetation Index
Precip	Annual Total Precipitation
Temp	Annual Mean Temperature
Dist_Water	Distance to Water Bodies
Bldg_Dens	Building Density
Road_Dens	Road Density
NTL	Nighttime Light Intensity
KD_Vuln	Vulnerable Facilities Density
KD_PubGath	Public Gathering Places Density
KD_Flame	Flammable/Explosive Sites Density
KD_Resid	Residential Density
KD_LargeBld	Large-scale Buildings Density
KD_Vuln_POI	Vulnerable Facilities Points of Interest
KD_PubGath_POI	Public Gathering Places Points of Interest
KD_Flame_POI	Flammable/Explosive Sites Points of Interest
KD_Resid_POI	Residential Points of Interest
KD_LargeBld_POI	Large-scale Buildings Points of Interest
Dist_FireStn	Distance to Fire Stations
FRI	Fire Risk Index
POI	Points of Interest
KDE	Kernel Density Estimation
GBDT	Gradient Boosting Decision
ML	Machine Learning
AHP	Analytic Hierarchy Process
SVM	Support Vector Machine

References

1. Bai, M.; Liu, Q. Evaluating Urban Fire Risk Based on Entropy-Cloud Model Method Considering Urban Safety Resilience. *Fire* **2023**, *6*, 62. [CrossRef]
2. Kim, J.S.; Kim, J.H.; Lee, S.K. Analysis of Fire-Accident Factors Using Big-Data Analysis Method for Construction Areas. *KSCE J. Civ. Eng.* **2018**, *22*, 1535–1543. [CrossRef]

3. Taylor, M.; Dean, E.; Fielding, J.; Lyon, R.; Reilly, D.; Francis, H.; Kwasnica, V. The Use of Machine Learning for Accidental Dwelling Fire Prevention. *Fire Saf. J.* **2025**, *157*, 104510. [CrossRef]
4. Cui, S.; Wang, N.; Zhao, E.; Zhang, J.; Zhang, C. Assessing Urban Fire Risk: An Ensemble Learning Approach Based on Scenarios and Cases. *Int. J. Disaster Risk Reduct.* **2024**, *114*, 104941. [CrossRef]
5. Kang, J.; Wang, H.; Jin, H.; Wang, Z.; Zhang, J. Risk Analysis and Management of Hydrogen Station Fire and Explosion Accidents Using DEMATEL-ISM and Complex Network Models. *Int. J. Hydrogen Energy* **2025**, *136*, 1230–1241. [CrossRef]
6. Pei, J.; Zhang, H.; Yin, S.; Wang, M.; Luo, Y. Research on Forest Fire Risk Assessment and Prevention and Control Countermeasures: A Case Study of Qinyuan County, Shanxi Province, China. *Ecol. Indic.* **2025**, *176*, 113719. [CrossRef]
7. Liu, F.; Zhao, S.; Weng, M.; Liu, Y. Fire Risk Assessment for Large-Scale Commercial Buildings Based on Structure Entropy Weight Method. *Saf. Sci.* **2017**, *94*, 26–40. [CrossRef]
8. Bostani, A.M.; Pourebrahim, S. Integrated Fire Risk Assessment Using Time-Series Satellite Data and Multi-Criteria Decision-Making Techniques. *Environ. Sustain. Indic.* **2025**, *28*, 100916. [CrossRef]
9. Xie, S.; Dong, S.; Chen, Y.; Peng, Y.; Li, X. A Novel Risk Evaluation Method for Fire and Explosion Accidents in Oil Depots Using Bow-Tie Analysis and Risk Matrix Analysis Method Based on Cloud Model Theory. *Reliab. Eng. Syst. Saf.* **2021**, *215*, 107791. [CrossRef]
10. Gulum, P.; Ayyildiz, E.; Taskin Gumus, A. A Two Level Interval Valued Neutrosophic AHP Integrated TOPSIS Methodology for Post-Earthquake Fire Risk Assessment: An Application for Istanbul. *Int. J. Disaster Risk Reduct.* **2021**, *61*, 102330. [CrossRef]
11. Seo, M.S.; Osorio, E.E.C.; Yoo, H.H. Analysis of Fire Risk Factors in Seoul, Korea, by Machine Learning. *Sens. Mater.* **2022**, *34*, 4841–4854. [CrossRef]
12. Bouzeraa, Y.; Bouchemal, N.; Djaaboub, S.; Hristov, G.; Zahariev, P. Machine Learning-Based Wildfire Susceptibility Mapping: A GIS-Integrated Predictive Framework. *Appl. Sci.* **2025**, *15*, 12188. [CrossRef]
13. Sivrikaya, F.; Demirel, D. Fire Risk Mapping Using Machine Learning Method and Remote Sensing in the Mediterranean Region. *Adv. Space Res.* **2025**, *75*, 5402–5419. [CrossRef]
14. Zamanialaei, M.; San Martin, D.; Theodori, M.; Purnomo, D.M.J.; Tohidi, A.; Lautenberger, C.; Qin, Y.; Trouvé, A.; Gollner, M. Fire Risk to Structures in California's Wildland-Urban Interface. *Nat. Commun.* **2025**, *16*, 8041. [CrossRef] [PubMed]
15. Rohde, D.; Corcoran, J.; Chhetri, P. Spatial Forecasting of Residential Urban Fires: A Bayesian Approach. *Comput. Environ. Urban Syst.* **2010**, *34*, 58–69. [CrossRef]
16. Granda, S.; Ferreira, T.M. Assessing Vulnerability and Fire Risk in Old Urban Areas: Application to the Historical Centre of Guimarães. *Fire Technol.* **2019**, *55*, 105–127. [CrossRef]
17. Chhetri, P.; Corcoran, J.; Stimson, R.J.; Inbakaran, R. Modelling Potential Socio-Economic Determinants of Building Fires in South East Queensland. *Geogr. Res.* **2010**, *48*, 75–85. [CrossRef]
18. Xia, Z.; Zhang, X.; Zhai, G.; Zhang, Y. Integrating Visual Spatial Vulnerability to Quantify Fire-Prone Neighborhoods in Cities: A Case Study of Nanjing, China. *Int. J. Disaster Risk Reduct.* **2025**, *128*, 105758. [CrossRef]
19. Palazzi, N.C.; Juliá, P.B.; Ferreira, T.M.; Rosas, J.; Monsalve, M.; de la Llera, J.C. Fire Risk Assessment of Historic Urban Aggregates: An Application to the Yungay Neighborhood in Santiago, Chile. *Int. J. Disaster Risk Reduct.* **2023**, *86*, 103550. [CrossRef]
20. Song, C.; Kwan, M.-P.; Song, W.; Zhu, J. A Comparison between Spatial Econometric Models and Random Forest for Modeling Fire Occurrence. *Sustainability* **2017**, *9*, 819. [CrossRef]
21. Pei, Z.; Li, J.; Guo, J.; Li, Q.; Chen, J. Using Local Co-Location Quotient and Niche-Based Model to Assess Fire Risk in Urban Environments: A Case Study of Beijing, China. *Sustain. Cities Soc.* **2023**, *99*, 104989. [CrossRef]
22. Sun, J.; Lin, M.; Wu, J.; Liu, X.; Sun, T.; Fang, H. Study on Urban Fire Risk Assessment Index System for Smart Cities. In Proceedings of the 2019 IEEE 2nd International Conference on Electronics Technology (ICET), Chengdu, China, 10–13 May 2019; pp. 609–613. [CrossRef]
23. Wei, Y.; Zhang, J.; Wang, J. Research on Building Fire Risk Fast Assessment Method Based on Fuzzy Comprehensive Evaluation and SVM. *Procedia Eng.* **2018**, *211*, 1141–1150. [CrossRef]
24. Kim, Y.; Cho, Y.; Heo, H.K.; Lim, L. Estimating Casualties from Urban Fires: A Focus on Building and Urban Environment Information. *Sustain. Cities Soc.* **2024**, *115*, 105839. [CrossRef]
25. Oh, H.Y.; Jeong, M.-H. Grid-Based Urban Fire Prediction Using Extreme Gradient Boosting (XGBoost). *Sens. Mater.* **2022**, *34*, 4879. [CrossRef]
26. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 6639–6649. [CrossRef]
27. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [CrossRef]
28. Cao, Y.; Li, F.; Xi, X.; van Bilsen, D.J.C.; Xu, L. Urban Livability: Agent-Based Simulation, Assessment, and Interpretation for the Case of Futian District, Shenzhen. *J. Clean. Prod.* **2021**, *320*, 128662. [CrossRef]
29. Shenzhen Municipal Government Data Open Platform. Available online: https://opendata.sz.gov.cn/ (accessed on 20 March 2026).
30. Geospatial Data Cloud. Available online: https://www.gscloud.cn/ (accessed on 20 March 2026).

31. Yang, J.; Dong, J.; Xiao, X.; Dai, J.; Wu, C.; Xia, J.; Zhao, G.; Zhao, M.; Li, Z.; Zhang, Y.; et al. Divergent Shifts in Peak Photosynthesis Timing of Temperate and Alpine Grasslands in China. *Remote Sens. Environ.* **2019**, *233*, 111395. [CrossRef]
32. Peng, S.; Ding, Y.; Liu, W.; Li, Z. 1 km Monthly Temperature and Precipitation Dataset for China from 1901 to 2017. *Earth Syst. Sci. Data* **2019**, *11*, 1931–1946. [CrossRef]
33. Peng, S. *1-km Monthly Mean Temperature Dataset for China (1901–2024)*; National Tibetan Plateau/Third Pole Environment Data Center: Beijing, China, 2019. [CrossRef]
34. OpenStreetMap. Available online: https://www.openstreetmap.org/ (accessed on 20 March 2026).
35. Shi, Q.; Zhu, J.; Liu, Z.; Guo, H.; Liu, M.; Liu, Z.; Liu, X. *A First High-Quality Vector Data of Buildings in East Asian Countries Based on a Comprehensive Large-Scale Mapping Framework*; Zenodo: Geneva, Switzerland, 2023. [CrossRef]
36. Zhong, X.; Yan, Q.; Li, G. Long Time Series Nighttime Light Dataset of China (2000–2020). *J. Glob. Change Data Discov.* **2022**, *6*, 416–424. [CrossRef]
37. Baidu Maps Open Platform. Map Imagery—Customized Map Solution. Available online: https://lbsyun.baidu.com/products/map (accessed on 20 March 2026).
38. O'Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [CrossRef]
39. Hao, Y.; Li, M.; Wang, J.; Li, X.; Chen, J. A High-Resolution Spatial Distribution-Based Integration Machine Learning Algorithm for Urban Fire Risk Assessment: A Case Study in Chengdu, China. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 404. [CrossRef]
40. Hossain, M.R.; Smirnov, O. Analyzing the Risk Factors of Residential Fires in Urban and Rural Census Tracts of Ohio Using Panel Data Analysis. *Appl. Geogr.* **2023**, *151*, 102863. [CrossRef]
41. Hu, J.; Shu, X.; Xie, S.; Tang, S.; Wu, J.; Deng, B. Socioeconomic Determinants of Urban Fire Risk: A City-Wide Analysis of 283 Chinese Cities from 2013 to 2016. *Fire Saf. J.* **2019**, *110*, 102890. [CrossRef]
42. Rahmawati, D.; Pamungkas, A.; Aulia, B.U.; Larasati, K.D.; Rahadyan, G.A.; Dito, A.H. Participatory Mapping for Urban Fire Risk Reduction in High-Density Urban Settlement. *Procedia Soc. Behav. Sci.* **2016**, *227*, 395–401. [CrossRef]
43. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Routledge: New York, NY, USA, 2018; ISBN 978-1-315-14091-9.
44. Islam, S.M.S.; Islam, K.M.A.; Mullick, M.R.A. Drought Hot Spot Analysis Using Local Indicators of Spatial Autocorrelation: An Experience from Bangladesh. *Environ. Chall.* **2022**, *6*, 100410. [CrossRef]
45. Hu, X.; Ma, C.; Huang, P.; Guo, X. Ecological Vulnerability Assessment Based on AHP-PSR Method and Analysis of Its Single Parameter Sensitivity and Spatial Autocorrelation for Ecological Protection—A Case of Weifang City, China. *Ecol. Indic.* **2021**, *125*, 107464. [CrossRef]
46. He, P.; Zhang, D.; Chen, W.; Jia, B. Spatio-Temporal Distribution Estimation and Analysis of Post-Earthquake Casualties in Urban Areas: A Case Study in Chenghua District, Chengdu. *Int. J. Disaster Risk Reduct.* **2025**, *123*, 105477. [CrossRef]
47. Wu, J.; Chen, X.; Li, R.; Wang, A.; Huang, S.; Li, Q.; Qi, H.; Liu, M.; Cheng, H.; Wang, Z. A Novel Framework for High Resolution Air Quality Index Prediction with Interpretable Artificial Intelligence and Uncertainties Estimation. *J. Environ. Manag.* **2024**, *357*, 120785. [CrossRef] [PubMed]
48. Segoni, S.; Nocentini, N.; Ajin, R.S.; Gatto, A.; Fanti, R. Understanding the Spatial Patterns of Major Geo-Hydrological Disasters in Italy Using the CatBoost Algorithm. *Sci. Total Environ.* **2025**, *999*, 180355. [CrossRef]
49. Yuan, L.; Dan, Y.; Leng, Y.; Zhang, J. Influencing Factors of Urban Safety Perception Based on the Combination of Multi-Source Data and Machine Learning: A Case Study of Nanchang City, China. *Sci. Rep.* **2025**, *16*, 3972. [CrossRef] [PubMed]
50. Zhang, Z.; Guo, Z.; Wu, S.; Tang, P.; Wang, S.; Yang, L. Harnessing Machine Learning and Multi-Source Data Fusion for Urban Fire Risk Assessment: Predictive Analysis of Spatial Heterogeneity. *Nat. Hazards* **2026**, *122*, 262. [CrossRef]
51. Yang, X.; Hao, Y.; Ding, H.; Yu, C.; Liu, J.; Li, L.; Chen, J. Explainable Artificial Intelligence (XAI) Framework Using XGBoost and SHAP for Assessing Urban Fire Risk Based on Spatial Distribution Features. *Int. J. Disaster Risk Reduct.* **2025**, *129*, 105798. [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.

land

MDPI

Article

Generative Regeneration of Historic Urban Fabric: A Framework Based on Deep Learning and Multi-Objective Optimization

Xiaoyu Ying [1,2,3], Shenbo Ni [1], Jiajing Wu [1,2,*], Yujie Zhao [1,3], Haiqiang Liu [4], Rongxin Qiu [5], Te Li [5], Jiamei Bei [1] and Hui Zhao [1]

[1] School of Art and Archaeology, Hangzhou City University, Hangzhou 310015, China; yingxiaoyu@hzcu.edu.cn (X.Y.); 2231003002@stu.hzcu.edu.cn (S.N.); zhaoyj@hzcu.edu.cn (Y.Z.); 2241003019@stu.hzcu.edu.cn (J.B.); 2241003023@stu.hzcu.edu.cn (H.Z.)
[2] Zhejiang Engineering Research Center of Building's Digital Carbon Neutral Technology, Hangzhou 310015, China
[3] Zhejiang Provincial Collaborative Innovation Center for Immovable Cultural Heritage Protection Technology, Hangzhou 310015, China
[4] Department of KANSEI Design Engineering, Faculty of Engineering, Yamaguchi University, Yamaguchi 753-8511, Japan; ryu.kaikyou@yamaguchi-u.ac.jp
[5] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, China; 22312119@zju.edu.cn (R.Q.); llt@zju.edu.cn (T.L.)
* Correspondence: wujj@hzcu.edu.cn; Tel.: +86-13732207710

Abstract

Amid China's rapid urbanization, many historic districts face complex challenges, including fragmented traditional fabrics, disordered spatial morphology, and discontinuous street networks. To tackle these issues, this study proposes a multimodal deep learning framework that combines Generative Adversarial Networks (GANs) and Diffusion Models, establishing an integrated generation-optimization workflow for the renewal of historic districts. The methodology begins by using Pix2PixHD to generate high-precision fabric layouts, followed by fine-tuning a Diffusion Model through Low-Rank Adaptation (LoRA) to achieve diversified morphological expansion. The candidate proposals are quantitatively evaluated using a ten-indicator evaluation matrix that covers both architectural fabric and street network dimensions. Afterwards, these proposals undergo iterative optimization with a multi-objective framework to enhance both urban fabric morphology and network performance. The framework was validated through an empirical study of the Yuehe Historic District in Jiaxing. The results indicate that the generated schemes closely align with the original urban fabric. Compared with the existing expanded area (EA), the weighted comprehensive fitness score of the optimized scheme group improved from 0.66 to 0.89 $\pm$ 0.02 (a 34.8% increase), with the standard deviation decreasing from 0.07 to 0.02, indicating significantly enhanced stability. Deep learning balances morphological authenticity, generative diversity, and performance in historic district preservation and renewal.

Keywords: historic district renewal; historic urban fabric; deep learning; generative design; multi-objective optimization

1. Introduction

China is home to a rich collection of historic districts, but preserving and renewing these areas presents significant challenges [1]. Rapid urbanization places tremendous

pressure on heritage sites, intensifying the conflict between conservation and redevelopment. This issue is especially pronounced in southern China, where dense water networks, narrow alleyways, and compact mixed-use blocks form a unique urban fabric. These features contribute to distinct regional identities, making the protection of these areas an urgent priority.

Conventional large-scale renewal strategies often require extensive land consolidation. However, this approach fragments traditional small-scale morphologies and leads to a decline in spatial identity and vitality. For instance, inappropriate renewal interventions in Nanjing's historic quarters have caused significant morphological degradation [2,3]. Consequently, it is essential to establish an organic renewal mechanism to preserve the continuity of the historic urban fabric while accommodating modern development.

To establish a clear framework for this approach, the core concepts must be explicitly defined within the context of contemporary urban planning. Historic District Renewal refers to a dynamic process that balances heritage conservation with functional modernization to ensure long-term viability. This process revolves around the preservation of the Historic Urban Fabric, which is defined as the physical texture of a city composed of plot patterns, building footprints, street networks, and open spaces that embody its historical evolution [2]. In modern practice, the synthesis of this fabric can be accelerated through Deep Learning, a subset of machine learning that utilizes multi-layered neural networks to recognize complex spatial patterns. When integrated into Generative Design frameworks, these algorithms automatically produce diverse design options based on defined parametric constraints [4–6]. To ensure these designs meet real-world planning standards, the outcomes are processed through Multi-Objective Optimization, a computational method that systematically identifies optimal solutions by balancing multiple conflicting performance goals [7].

Although advanced computational tools are increasingly applied to urban design, current generative frameworks face significant limitations when applied to sensitive heritage environments. These limitations can be grouped into three core issues:

- Issue 1: An epistemological gap between computational metrics and heritage values. In current generative design workflows, "authenticity" is often oversimplified as mere physical similarity or geometric replication of building footprints. However, heritage conservation doctrines—such as the Nara Document on Authenticity [8] and the Historic Urban Landscape (HUL) approach [9]—state that authenticity is a broader, multidimensional concept. It encompasses materiality, utilization, memory, historic context, and intangible human associations. Equating simple geometric resemblance with authentic heritage continuity is a conceptually overstretched claim that risks stripping a district of its deeper historical value.
- Issue 2: Methodological deficiencies in traditional planning tools. Conventional renewal strategies still rely heavily on manual surveying and subjective experiential judgments, as seen in projects like Dujiangyan West Street [10]. These methods lack standardized quantitative tools, which prevents the precise restoration and performance-based evaluation of intricate spatial structures.
- Issue 3: Technical barriers in computational design generation. The application of deep generative models to complex urban forms faces severe algorithmic constraints. Generative Adversarial Networks (GANs) frequently suffer from data scarcity, mode collapse, and unstable training processes [4]. Meanwhile, advanced Denoising Diffusion Probabilistic Models lack explicit topological reasoning, which often results in semantic hallucinations, discontinuous street networks, and a lack of precise metric control [6,11,12].

https://doi.org/10.3390/land15060976

The evolution of neural networks in architectural design generation has progressed from early probabilistic models to advanced deep generative frameworks, most notably Generative Adversarial Networks (GANs). In 2011, researchers at Stanford University pioneered the application of probabilistic methods, specifically Markov Chain Monte Carlo techniques, to architectural layout generation. This work established a foundational framework for data-driven design [13]. The introduction of GANs by Goodfellow et al. in 2014 marked a paradigm shift, offering a new approach to image-driven design synthesis [4,14]. Specialized models like Pix2PixHD have since been used to generate architectural layouts, master plans, and campus designs [5,14–18]. However, using GANs to regenerate historic urban fabric remains difficult due to insufficient training samples and limited morphological variety in the outputs [18–22].

To address these limitations, diffusion models have emerged as a powerful alternative, outperforming GANs in sample fidelity and diversity. Recent architectural studies have combined Latent Diffusion Models (such as Stable Diffusion) with adapter-based techniques like ControlNet and Low-Rank Adaptation (LoRA) to achieve precise stylistic control over facades and floor plans [6,23]. Furthermore, hybrid workflows that integrate diffusion models with optimization algorithms have been developed to predict performance metrics such as daylighting [7]. Yet, because diffusion models operate primarily as visual synthesizers, they struggle to natively incorporate domain-specific engineering knowledge or maintain strict spatial topology [23–25]. This limitation often produces broken street networks and geometric errors in historical contexts. To resolve these conflicts, this study proposes a multimodal deep generative framework that combines the structural stability of GANs with the creative diversity of Diffusion Models. This hybrid architecture utilizes Pix2PixHD to establish high-precision fabric boundaries and fine-tunes a Diffusion Model via LoRA to expand morphological diversity. By embedding a ten-indicator evaluation matrix into a Multi-Objective Optimization workflow, the framework achieves both high-fidelity reconstruction and controllable synthesis. Ultimately, this research establishes a novel methodological paradigm for the digital preservation and adaptive regeneration of cultural heritage, effectively balancing generative creativity with engineering precision.

2. Materials and Methods

To tackle the critical challenge of balancing morphological fidelity and generative diversity in the regeneration of historic urban fabric, this study proposes a novel technical framework that combines deep learning with Multi-Objective Optimization. This framework aims to create precise fabric layouts using GANs while utilizing Diffusion Models to broaden the design space. Subsequently, Multi-Objective Optimization algorithms are applied to systematically enhance the generated solutions, improving the overall spatial quality of both the urban fabric and the street network.

The implementation workflow consists of three core stages (see Figure 1 for the overall framework):

- Initial Layout Synthesis (GAN-based): A constructed multimodal dataset is utilized, employing the Pix2PixHD model to generate high-fidelity initial urban fabric layouts. These layouts adhere strictly to land-use constraints and historical spatial patterns;
- Design Space Expansion (Diffusion-based): To overcome the diversity limitations of GANs, we implement a Stable Diffusion model fine-tuned via LoRA. This stage uses semantic guidance to diversify the initial results, generating a wide variety of candidate solutions (morphological variants);
- Performance-Driven Optimization and Decision-Making: We establish a quantitative evaluation system that includes metrics for fabric morphology and street network efficiency (with detailed definitions in Section 2.1). Using this framework, a custom

https://doi.org/10.3390/land15060976

Python (version 3.10, Python Software Foundation, Wilmington, DE, USA) algorithm integrates with Multi-Objective Optimization strategies for fitness evaluation and iterative refinement. This process filters and selects Pareto-optimal solutions that exhibit superior overall performance, establishing a closed-loop optimization cycle of generation-evaluation-feedback.

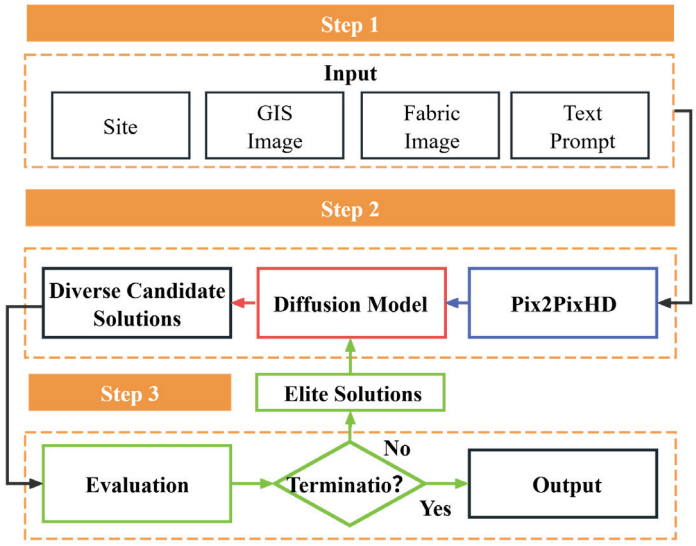


Figure 1. The proposed computational workflow integrating GANs, Diffusion Models, and Multi-Objective Optimization.

This framework systematically integrates the generative capabilities of deep learning with the scientific rigor of Multi-Objective Optimization, providing comprehensive technical support for the renewal of historic districts—from the synthesis of multiple schemes to the selection of optimal solutions. The following sections detail the implementation methodologies for both the generation and optimization modules.

2.1. Deep Learning-Driven Synthesis of Historic District Layouts

2.1.1. Multimodal Dataset Construction and Pre-Processing

To address the complex challenges involved in synthesizing the morphology of historic districts—especially those characterized by the distinctive river–street typologies of the Jiangnan region—this study developed a structured, multimodal dataset for training and validating generative design models. Data acquisition combined Geographic Information System (GIS) data, field survey drawings, archival materials from design institutes, and professional architectural databases to ensure that the samples were highly representative, standardized, and practically feasible.

- A total of 60 representative historic district cases were curated through a stratified sampling strategy based on four rigorous criteria:
- Morphological Integration: building layouts must exhibit a strong spatial interdependence with river corridors;
- Legibility of Fabric: the spatial fabric characteristics must be clearly discernible;
- Heritage Value: the districts must possess significant historical and cultural importance; and Data Completeness: high-quality visual and archival documentation must be available.

Balanced geographic representation was achieved by drawing cases from ten distinct cities in the Jiangnan region, including Hangzhou, Jiaxing, Suzhou, and Shaoxing, with each city contributing between four and eight cases to prevent single-city dominance. The

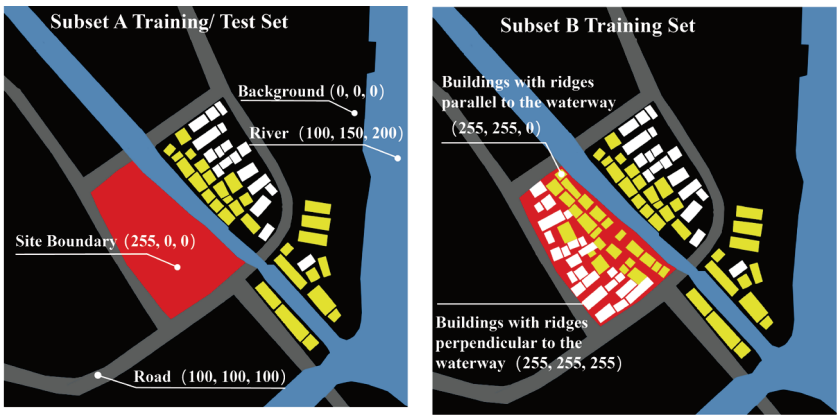
https://doi.org/10.3390/land15060976

dataset focuses on historic districts shaped or renewed from the early years of the People's Republic of China to the present. However, the surviving architectural fabric within these districts spans a much longer period. Many building remains date back to the Ming and Qing dynasties and have persisted through various phases of urban change into the present day. The dataset encompasses both well-preserved cores and moderately transformed buffer zones to reflect real-world conservation and urban change conditions.

The final dataset consists of 60 cases. This number reflects a practical reality. Well-preserved historic districts of the specific river–street typology are scarce. The collected cases represent nearly all available samples that meet the rigorous selection criteria. A larger dataset of the same typology is not easily obtainable. However, because the typology is highly coherent, the morphological variation across cases is limited. The 60 samples are therefore sufficient to capture the essential spatial patterns of this fabric. To further address the sample size, data augmentation and transfer learning strategies were applied, as described below.

A 5-fold cross-validation strategy was adopted to evaluate the generalizability of the deep learning model and to reduce the risk of overfitting caused by the limited sample size. The complete dataset contained 60 historic district cases. From these, 10 cases were randomly selected as a fixed test set. The remaining 50 cases were used for the 5-fold cross-validation. They were randomly partitioned into five disjoint subsets of equal size. In each fold, one subset served as the validation set, and the other four subsets were used for training. The complete Pix2PixHD-LoRA pipeline was trained and evaluated in every fold. The results are summarized in Section 3.1 and confirm the stability of the proposed framework.

In compliance with the input specifications of the deep learning architecture, all image samples underwent standardized pre-processing (detailed in Appendix A). Images were resampled to a uniform resolution of 1024 × 1024 pixels at a cartographic scale of 1:2000, which standardizes the covered ground area to approximately 442 m × 442 m. Following standard semantic segmentation protocols, distinct spatial elements were encoded using specific Red-Green-Blue (RGB) values (Figure 2): background (0, 0, 0), land-use boundaries (255, 0, 0), waterways (100, 150, 200), street networks (100, 100, 100), buildings with ridge orientations parallel to the watercourse (255, 255, 0), and buildings with ridge orientations perpendicular to the watercourse (255, 255, 255). For each case, a paired dataset was generated for the Pix2PixHD algorithm, consisting of the Source Domain (Domain A, site constraints with existing buildings removed) and the Target Domain (Domain B, ground-truth historic urban fabric).



(**a**) (**b**)

Figure 2. Schematic diagram of the pre-processed semantic samples: (**a**) Subset A Training | Test Set; (**b**) Subset B Training Set.

https://doi.org/10.3390/land15060976

To enhance model generalization and reduce overfitting due to data scarcity, two data augmentation strategies were implemented: Geometric Symmetry Augmentation and Noise Injection. Horizontal flipping [26] reinforces the model's ability to learn directional invariance in morphology, particularly relevant for the mirrored layouts commonly found in riverside architecture, while vertical flipping [27] helps better extract local fabric features by introducing top-down symmetry. Additionally, zero-mean Gaussian noise [28] was introduced to simulate potential image degradation or sensor artifacts encountered in real-world applications, compelling the model to learn more robust feature representations.

The curated dataset was split into a training set (50 cases) and a testing set (10 cases). The testing set was strictly isolated from the training process and deliberately designed to contain unseen, challenging morphologies—such as irregular water edges, highly fragmented parcels, and dead-end alley configurations—to rigorously evaluate the model's generalization capability within the regional typology. While the raw sample size appears modest, several design choices ensure its adequacy for the generative task: first, each case contains dense spatial information with hundreds of building footprints and street segments, providing a rich supervision signal; second, the geometric symmetry augmentation expands the training set fourfold to 200 equivalent samples; third, the LoRA fine-tuning strategy leverages the strong visual priors of a large-scale pre-trained Stable Diffusion model, substantially reducing the demand for massive domain-specific data. Moreover, confining the dataset to a single coherent typology—the Jiangnan river–street fabric—limits stylistic variance and allows the model to efficiently learn deep morphological patterns, aligned with sample sizes adopted in comparable architectural generation workflows [15,17,18]. Consequently, this structured database establishes a solid foundation for efficient neural network training and reliable evaluation.

2.1.2. High-Resolution Synthesis of Historic Urban Fabric via Pix2PixHD

To achieve high-fidelity synthesis of historic district layouts, this study employs the Pix2PixHD model [5] as the primary generative architecture. As illustrated in Figure 3, this model utilizes an advanced Conditional Generative Adversarial Network framework optimized for high-resolution synthesis. Its primary innovation lies in the integration of a coarse-to-fine generator coupled with a multi-scale discriminator, effectively addressing challenges in detail preservation and training stability.

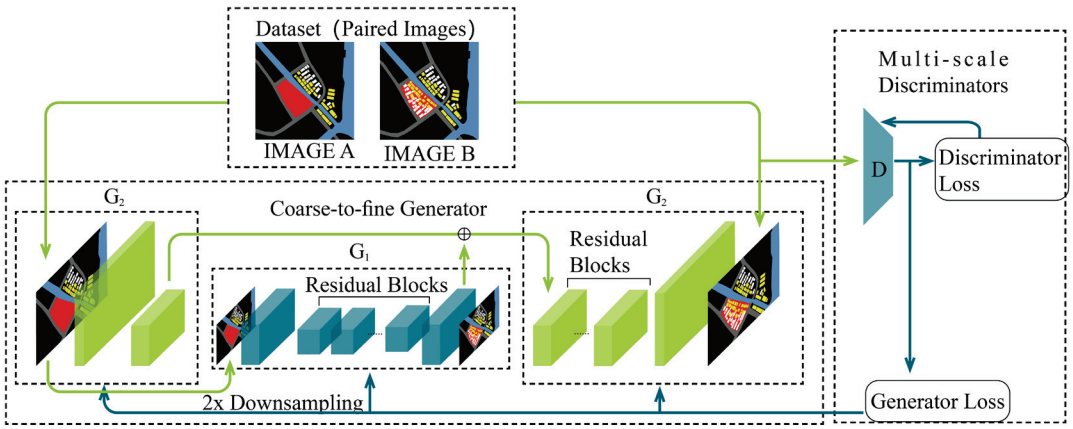


Figure 3. The network architecture of the Pix2PixHD framework.

The generator (G) adopts a two-stage cascaded architecture $G = \{G_1, G_2\}$. The Global Generator G_1 first constructs macro-structural features at a resolution of 512 × 512. Subsequently, the Local Enhancer Network (G_2) performs 4× upsampling and local refinement on the output of G_1. By integrating global contextual information, G_2 synthesizes the final

https://doi.org/10.3390/land15060976

high-resolution result (e.g., 1024 × 1024), ensuring both high-frequency detail fidelity and structural consistency. Complementarily, the discriminator ($D = \{D_1, D_2, D_3\}$) evaluates the input at three distinct scales to assess image authenticity across different perceptual fields.

The training process is optimized via a composite objective function:

$$\mathcal{L}_{total}\ (G, D_k) = \min_G \left(\left(\max_{D_1,D_2,D_3} \sum_k \mathcal{L}_{GAN}(G, D_k) \right) + \lambda_{FM} \sum_k \mathcal{L}_{FM}(G, D_k) + \lambda_{VGG} \mathcal{L}_{VGG}(y, G) \right) \quad (1)$$

where $\mathcal{L}_{GAN}$ is the adversarial loss, which drives the synthesis to approximate the real data distribution; $\mathcal{L}_{FM}$ denotes the feature matching loss, which stabilizes training by minimizing feature discrepancies within the discriminator's intermediate layers; $\mathcal{L}_{VGG}$ signifies the perceptual loss, which utilizes a pre-trained Visual Geometry Group network to constrain the semantic content in the feature space; $\mathcal{L}_{FM}$ and $\mathcal{L}_{VGG}$ are weighting coefficients empirically set to 10 to balance each loss term's contribution; and y represents the ground-truth image used for training.

Ultimately, the Pix2PixHD model synthesizes high-fidelity pre-synthesized layouts for historic districts. These layouts serve as the foundational structural guidance to drive the subsequent phase of Diffusion Model-based design space expansion.

2.1.3. Diversified Layout Synthesis via Diffusion Models and LoRA

To achieve a diversified synthesis of historic district layouts (Figure 4), this study employs a Diffusion Model integrated with LoRA for efficient domain adaptation. By embedding lightweight, trainable rank-decomposition matrices into the pre-trained backbone, this approach enables rapid adaptation for generating historic urban fabrics with limited computational resources [21].

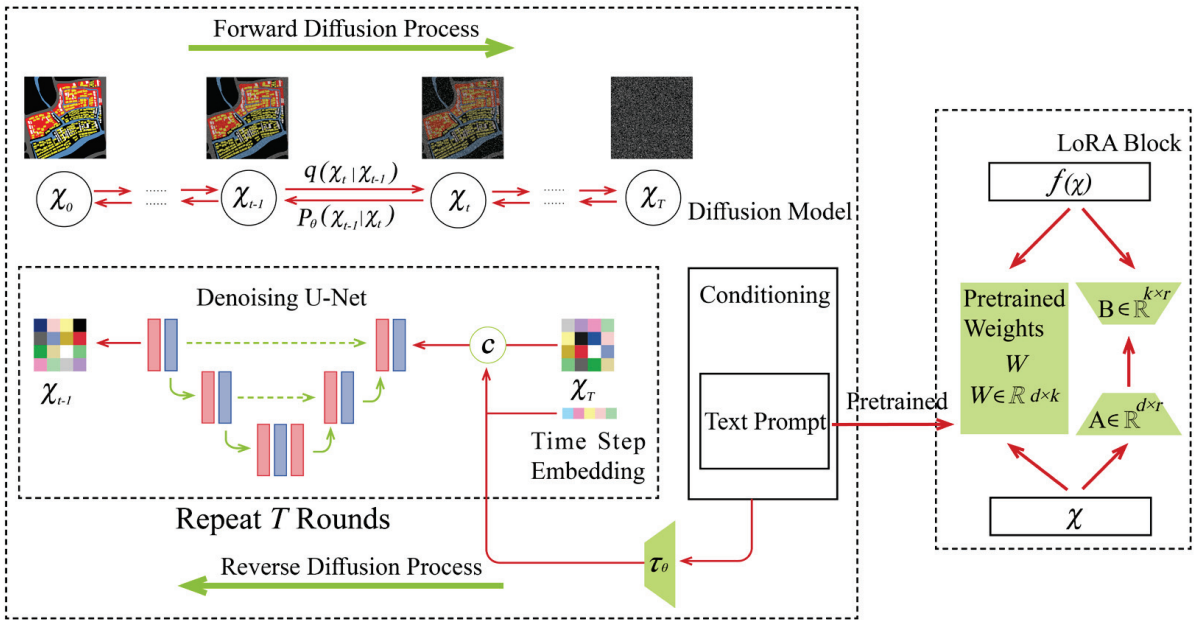


Figure 4. Architecture of the Diffusion Model integrated with LoRA.

The generative mechanism is based on reversing a forward noise injection process. The forward process gradually corrupts the raw data x_0 with Gaussian noise $x_T \sim \mathcal{N}(0, I)$ using a fixed variance schedule $\beta_1, \ldots, \beta_t$, adhering to the Markov chain property:

$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{2}$$

Conversely, the reverse process trains a neural network ϵ_θ, typically a U-Net, to iteratively reconstruct a coherent image from this noise:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{3}$$

To adapt the pre-trained model (Stable Diffusion v1.5) efficiently, this study utilizes LoRA fine-tuning. This technique approximates the weight updates ΔW using low-rank decomposition:

$$W = W_0 + \Delta W = W_0 + BA \tag{4}$$

The training objective is to minimize the mean squared error between the predicted and the actual noise:

$$\mathcal{L}_{LDM} = \mathbb{E}_{x,\varepsilon\sim\mathcal{N}(0,1),t}\left[\|\varepsilon - \varepsilon_\theta(x_t, t, c)\|_2^2\right] \tag{5}$$

where x_t denotes the latent representation of the fabric layout at timestep t; ε and ε_θ represent the actual added noise and the noise predicted by the U-Net, respectively; c signifies the conditional input, which, in this study, is the pre-synthesized layout from the Pix2PixHD stage; $W_0 \in \mathbb{R}^{d\times k}$ is the pre-trained weight matrix, whereas $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$ are the low-rank matrices with rank $r \ll \min(d, k)$; and $\mathcal{L}_{LDM}$ is the loss function used to optimize the LoRA parameters while preserving generative priors.

This integrated framework leverages the robustness of diffusion models for synthesizing complex topological structures while balancing computational efficiency and morphological fidelity.

2.1.4. Training and Implementation Details

The Pix2PixHD model was trained using the official PyTorch (version 2.0.1, Meta Platforms, Inc., Menlo Park, CA, USA) implementation. Training was conducted for 200 epochs with a batch size of 1, due to the high resolution of the input images. The Adam optimizer was adopted with a learning rate of 0.0002, β_1 = 0.5, and β_2 = 0.999. The loss weighting coefficients $\mathcal{L}_{FM}$ and $\mathcal{L}_{VGG}$ (corresponding to λ_1 and λ_2 in Equation (1)) were both set to 10. The generator followed a coarse-to-fine architecture, while the discriminator operated at three scales. A single NVIDIA GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA) with 24 GB memory was used.

For the diffusion-based expansion, the pre-trained Stable Diffusion v1.5 (Stability AI, London, UK) checkpoint was fine-tuned using the Diffusers library (version 0.25.0, Hugging Face, Inc., New York, NY, USA). LoRA was applied to the attention layers of the U-Net, with rank r = 16 and alpha = 16. The model was trained for 20,000 steps, with a batch size of 4 and a learning rate of 1×10^{-4}. All input images were resized to 1024 × 1024 pixels. The conditioning text prompt was: "historic water town fabric, traditional Jiangnan architecture, dense alleyways, riverside buildings". The classifier-free guidance scale was set to 7.5. During inference, 50 denoising steps were performed using the DDIM scheduler. All LoRA experiments ran on a single NVIDIA GeForce RTX 4090 GPU with 24 GB memory.

https://doi.org/10.3390/land15060976

A fixed random seed of 42 was used across all experiments to ensure reproducibility. Generated images were post-processed by thresholding the semantic colour channels to extract binary masks for buildings, streets, and water. These masks were then vectorised into GIS-ready polygon and polyline layers using standard computer vision and geospatial libraries.

The code developed for this study (Python 3.10, PyTorch 2.0.1, Diffusers 0.25.0) will be publicly released on GitHub upon acceptance. The training dataset cannot be shared due to privacy and archival restrictions. A detailed description of data sources and pre-processing is provided in Appendix A.

2.2. Evaluation Metrics and Multi-Objective Optimization Framework

Deep learning models can effectively create multiple design scenarios for historic districts, offering a variety of options for organic renewal. However, the inherent stochasticity of the generation process leads to significant variations in output quality. On one hand, the model can generate high-fidelity designs that align exceptionally well with site-specific constraints. On the other hand, it often produces a substantial number of suboptimal solutions that may suffer from poor spatial organization, a lack of coherence in the urban fabric, and issues with street network integrity. Relying exclusively on manual expert judgment to screen and assess large-scale generative datasets is not only computationally inefficient but also lacks objective, standardized evaluation criteria.

To address these challenges, this study establishes a comprehensive evaluation system that quantitatively decodes the morphological rules and spatial logic of historic districts. Grounded in empirical research on the Yuehe Historic District in Jiaxing, Zhejiang, this system facilitates automated fitness evaluation and controlled, iterative optimization of the generated outcomes. Ultimately, it constitutes a deep learning framework characterized by both precision and adaptability, providing robust technical support for the preservation and adaptive regeneration of historic district.

2.2.1. Evaluation Indicators

This study utilizes established frameworks in urban morphology and heritage conservation to identify two fundamental physical dimensions that shape the historical context and functionality of historic districts: architectural fabric and street networks [25,29]. While each of these dimensions has been extensively examined in isolation in the existing literature, their integration into a cohesive, synergistic evaluation framework tailored for deep learning-driven optimization has not been adequately addressed.

To address this gap, this study establishes a comprehensive assessment system focused on these two primary dimensions: architectural fabric and street network configuration. This system is explicitly grounded in the unique spatial characteristics and stringent conservation requirements of traditional Jiangnan water towns (traditional waterfront towns south of the Yangtze River). Furthermore, the framework is broken down into 10 specific, quantifiable performance indicators (as detailed in Table 1). This metric system aims to translate abstract, qualitative conservation needs into computable objectives compatible with Multi-Objective Optimization algorithms. Consequently, it directs the generative process towards an optimal balance between preserving the continuity of historical context and ensuring spatial rationality.

- Architectural Fabric Dimension

https://doi.org/10.3390/land15060976

Table 1. The quantitative evaluation indicator system for historic district regeneration.

Dimension	Evaluation Indicator		Scoring Criteria & Ideal Thresholds	Weight (%)
Architectural Fabric	A_{FCR}	Fabric Coverage Rate	Opt. *: ≥50%. Pen.: −0.1 per 5% decrease.	11.8
	A_{NSU}	No. of Small Fabric Units	Opt.: 0 units. Pen.: −0.1 per 5 additional units.	9.5
	A_{NLU}	No. of Large Fabric Units	Opt.: 0 units. Pen.: −0.1 per 5 additional units.	7.4
	A_{AFS}	Average Fabric Size	Opt.: 120 m^2. Pen.: −0.1 per 10 m^2 deviation.	12.1
	A_{SDF}	Standard Deviation of Fabric Size	Opt.: 120 m^2. Pen.: −0.1 per 10 m^2 deviation.	7.9
Street Network	A_{SND}	Street Network Density	Opt.: ≥1000 m/ha. Pen.: −0.05 per 100 m/ha decrease.	11.1
	A_{MCP}	Minimum Cost Path	Opt.: Euclidean distance. Pen.: −0.1 per 10 m excess.	8.3
	A_{STA}	Spatio-Temporal Accessibility	Range: 0–1.	12.6
	A_{ASW}	Average Street Width	Opt.: 6 m. Limit: Linear scale to 6–10 m.	10.6
	A_{SDW}	Standard Deviation of Street Width	Opt.: 2 m. Limit: Linear scale to 2–6 m.	8.7

* Opt. (Optimal Value) represents the benchmark for a full score; Pen. (Penalty) denotes the point deduction per unit of deviation. This weighting scheme is grounded in Jiangnan morphological theory and confirmed by expert consensus.

This dimension encompasses five key indicators designed to assess the morphological integrity of the built form:

1. Fabric Coverage Rate [30]

To evaluate the spatial compactness and land-use intensity of the historic district, this indicator is defined as the ratio (percentage) of the total building footprint area to the total site area. Its theoretical basis stems from the classical measurement of the Building Coverage Ratio in urban morphology, serving as a critical metric for assessing the preservation integrity of traditional building clusters and the potential intrusiveness of modern developments. In the scoring criteria, a score of 100% is awarded for values $\geq$ 50%. For every 5% decrease, 0.1 points are deducted. This threshold is empirically derived from systematic surveys of historic districts in the Jiangnan region and aligns with the characteristic building density of 45–55% found in traditional settlements.

$$A_{FCR} = \frac{\sum S_{bi}}{S_{\text{total}}} \times 100\% \quad (6)$$

where A_{FCR} is the Fabric Coverage Rate; S_{bi} represents the footprint area of the i-th building unit; and S_{total} represents the total area of the site (red line boundary).

Number of Small Fabric Units [30]

To measure the level of spatial fragmentation within the historic district, this indicator counts the total number of independent building units with a footprint of less than 30 m^2. It helps to identify the distribution of fine-grained architectural elements. In historic districts, this metric reveals the remnants of traditional alleyways, ancillary structures, and courtyard layouts, enabling us to assess whether the original capillary-like spatial fabric remains intact. In the scoring system, full points are awarded for 0 units, which indicates a consolidated fabric. For each additional 5 units, 0.1 points are deducted. This penalty

https://doi.org/10.3390/land15060976

mechanism is designed to prevent excessive spatial fragmentation from weakening the coherence of the historic fabric.

$$A_{NSU} = \sum_{i=1}^{n} f(S_{bi}),\ f(S_{bi}) = \begin{cases} 1, & S_{bi} < 30\ \text{m}^2 \\ 0, & S_{bi} \geq 30\ \text{m}^2 \end{cases} \tag{7}$$

where A_{NSU} is the total count of independent building units smaller than 30 m^2, and S_{bi} is the footprint area of an individual building unit i.

2. Number of Large Fabric Units [30]

To identify large-scale volumes inconsistent with an area's historic character, this indicator counts the total number of independent building units exceeding 240 m^2. In historic district studies, this measure primarily serves to detect the introduction of incompatible, large-scale modern constructions. It provides a crucial foundation for assessing scalar conflicts with the historical environment. The scoring system begins at 0 points, with 0.1 points deducted for every additional 5 units. Its purpose is to strictly limit building volumes that do not align with the historical context.

$$A_{NLU} = \sum_{i=1}^{n} f(S_{bi}),\ f(S_{bi}) = \begin{cases} 1, & S_{bi} > 240\ \text{m}^2 \\ 0, & S_{bi} \leq 240\ \text{m}^2 \end{cases} \tag{8}$$

where A_{NLU} is the total count of independent building units exceeding 240 m^2, and S_{bi} is the footprint area of an individual building unit i.

3. Average Fabric Size [30,31]

To evaluate the overall scalar coordination of the built form within the historic district, this metric calculates the arithmetic mean of all building unit footprints (m^2). This average reflects the area's foundational spatial granularity. It allows comparison of scale characteristics across different historical periods and helps determine whether any modifications or additions have distorted the original morphological dimensions. In the scoring system, the maximum value is set at 120 m^2, which earns full marks. For each deviation of 10 m^2 from this optimal value, there is a deduction of 0.1 points. This benchmark is based on the typical scale of traditional vernacular dwellings in the Jiangnan region.

$$A_{AFS} = \frac{1}{n} \sum_{i=1}^{n} S_{bi} \tag{9}$$

where A_{AFS} is the arithmetic mean of all building unit footprints (Mean Grain Size); n is the total number of building units within the district; and S_{bi} is the area of building unit i.

4. Standard Deviation of Fabric Size [30,31]

To measure the homogeneity of the historic urban fabric and its spatial rhythm, this metric measures the dispersion of building unit areas (m^2). The methodology is based on statistical methods for spatial form analysis, enabling the identification of regional morphological variations. In the scoring system, the target value for full marks is set at 120 m^2. For every deviation of 10 m^2 from this target, there is a deduction of 0.1 points. This approach aims to maintain the district's spatial rhythm, ensuring that architectural groups exhibit variation within a unified scale.

$$A_{SDF} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (S_{bi} - A_m)^2} \tag{10}$$

https://doi.org/10.3390/land15060976

where A_{SDF} is the standard deviation of building footprint areas, representing morphological homogeneity, and A_m is the average fabric size calculated in the previous indicator A_{AFS}.

- Street Network Dimension

This dimension encompasses five key indicators designed to assess the connectivity and accessibility of the spatial structure:

1. Street Network Density [29,30]

To quantify the coverage and connectivity potential of the historic district's transportation network, this indicator calculates the total street length per unit area (m/ha). This method is based on traditional urban morphology research, which suggests that a high network density corresponds to the dense street network characteristic of historical settlements. In the scoring system, a maximum score is assigned for densities of 1000 m/ha or more. For every 100 m/ha decrease, 0.05 points are subtracted, reflecting the high-density spatial characteristics typical of Jiangnan water towns.

$$A_{SND} = \frac{\sum L_{sj}}{S_{ha}} \tag{11}$$

where A_{SND} is the street network density (m/ha); L_{sj} denotes the length of the j-th street segment; and S_{ha} denotes the total area of the study region in hectares.

2. Minimum Cost Path [30,32]

To objectively evaluate the accessibility efficiency of the historic district, this metric uses graph-theoretic principles to analyze the shortest path length (m) between the two farthest nodes in the network. It effectively identifies connectivity barriers caused by dead-end streets and tortuous alleyways. In the scoring system, full points are awarded for distances equal to or less than the straight-line (Euclidean) distance between the primary entrances. For every additional 10 m beyond this distance, 0.1 points are deducted. This approach balances accessibility efficiency with the preservation of traditional spatial complexity.

$$A_{MCP} = \min \sum_{j \in P_{uv}} l_j \tag{12}$$

where A_{MCP} is the shortest path length between the two most distant nodes u and v in the network, and l_j represents the length of the individual segments j that constitute path P_{uv}.

3. Spatio-Temporal Accessibility [30,32,33]

To evaluate the walkability and human-scale design of the historic district, this metric uses time geography as its theoretical foundation. It measures the percentage of street areas that are accessible within a defined walking timeframe (e.g., 2 h). This method effectively evaluates how conveniently residents and visitors can reach service facilities and cultural attractions. In the scoring system, the percentage value is normalized to a score ranging from 0 to 1, highlighting the pedestrian-oriented spatial character of historic districts.

$$A_{STA} = \frac{N_{\text{reachable}}}{N_{\text{total}}} \times 100\% \tag{13}$$

where A_{STA} is the percentage of the street area or nodes accessible within the defined walking timeframe (e.g., 2 h); $N_{\text{reachable}}$ is the number of nodes reachable within the time threshold; and N_{total} is the total set of nodes within the network.

https://doi.org/10.3390/land15060976

4. Average Street Width [30,31]

To identify instances of inappropriate widening of historic streets and alleys for modern traffic demands, this indicator calculates the arithmetic mean (m) of all street widths. This calculation reflects the overall spatial scale of the streetscape and is vital for preserving the original spatial proportions, specifically the Distance/Height ratio, in historic districts. The scoring standard sets 6 m as the optimal width for full points. Widths of 10 m receive 0 points, with a linear decrease in points for widths between 6 m and 10 m. This range aligns perfectly with the typical scale of traditional streets and alleys.

$$A_{ASW} = \frac{1}{m}\sum_{k=1}^{m} w_k \tag{14}$$

where A_{ASW} is the arithmetic mean of all street widths; w_k represents the width of the k-th street segment; and m is the total number of street segments.

5. Standard Deviation of Street Width [30,31]

To capture the spatial heterogeneity of the streetscape within the historic district, this metric assesses the statistical dispersion of street widths (m). A lower standard deviation generally indicates well-planned, well-preserved traditional grids. The scoring system assigns 2 m for full points and 6 m for zero points, with a linear scale in between. This approach aims to balance street uniformity with organic diversity.

$$A_{SDW} = \sqrt{\frac{1}{m}\sum_{k=1}^{m}(w_k - A_w)^2} \tag{15}$$

where A_{SDW} is the standard deviation of street widths, reflecting spatial heterogeneity; and A_w is the average street width calculated in the previous indicator A_{ASW}.

A pool of candidate indicators was first identified from the literature on urban morphology and heritage conservation. Each indicator was assessed for importance, sensitivity, and measurability. A structured expert consultation questionnaire was then developed.

A panel of 20 experts was invited. The selection followed three criteria:

- A minimum of three years of professional experience in historic district conservation, urban design, or related fields;
- An intermediate or senior professional title;
- Direct involvement in at least one historic district project or published research in a relevant area.

The panel included architects, urban morphologists, municipal planning officials, and heritage managers, ensuring representation from both academic and practice-oriented perspectives.

Two rounds of consultation were conducted. In the first round, each expert received a questionnaire containing all candidate indicators with their definitions, measurement formulas, and scoring criteria. Experts rated each indicator on a 1–5 scale (1 = not important, 5 = extremely important) and provided written justifications for their ratings. The completed questionnaires were collected, and the mean scores, coefficients of variation, and all written comments were compiled.

https://doi.org/10.3390/land15060976

In the second round, the aggregated results were returned to the panel. Each expert received a summary that included the group mean score for each indicator and the anonymized written justifications from the first round. Experts were asked to review the feedback and revise their initial ratings where appropriate. In this round, experts also completed pairwise comparisons between indicators to support the subsequent weight derivation.

Both rounds achieved a 95% response rate (19 valid questionnaires each), indicating high expert positivity. Expert concentration was assessed by the mean score and the proportion of full marks for each indicator, which showed clear differentiation among the indicators.

The Kendall's W coefficient of concordance was 0.84 ($p < 0.001$), showing strong consensus among the experts. Internal consistency reliability was high, with Cronbach's alpha exceeding 0.90 for both the architectural fabric and street network dimensions. The item-level content validity index (I-CVI) ranged from 0.85 to 1.00, and the overall scale-level CVI was 0.94, confirming the relevance of all ten indicators. Indicators with a mean importance score below 4.0 or a coefficient of variation above 0.25 were considered for removal or revision. Ratings of 4 or 5 were considered relevant for the calculation of I-CVI.

Several written justifications provided insight into the experts' reasoning. One architect noted that "the Standard Deviation of Fabric Size is particularly useful for distinguishing between organically evolved historic fabric and uniform modern redevelopment." A heritage manager commented that "Spatio-Temporal Accessibility reflects the pedestrian-oriented character of historic districts and should carry substantial weight." A municipal planning official remarked that "the penalty mechanism for Large Fabric Units is well-designed, as oversized modern buildings are among the most disruptive elements in historic contexts." These comments supported the final indicator selection and confirmed that the evaluation framework captured the key concerns of historic district renewal.

Final weights were derived using the Analytic Hierarchy Process (AHP), based on the pairwise comparisons collected in the second round. Two comparison matrices were constructed: one for the five indicators within the architectural fabric dimension and one for the five indicators within the street network dimension. A separate comparison between the two dimensions determined their relative importance. Experts used Saaty's 1–9 scale for all pairwise comparisons. The individual judgments were aggregated using the geometric mean method. The consistency ratio was checked for each matrix and found to be below 0.10, indicating acceptable consistency. The resulting weight set is shown in Table 1.

2.2.2. Multi-Objective Optimization Framework

To thoroughly evaluate the performance of the generated historic urban fabric, this study establishes a systematic Multi-Objective Optimization framework. The main goal of the framework is to achieve synergistic optimization by balancing architectural morphology and street network efficiency through iterative evolutionary screening.

- Fitness Function Formulation

The design selection problem is formulated as a multi-objective optimization. Because the expert panel produced a clear, consensual set of trade-off preferences, the problem is scalarized into a single fitness function using the weighted-sum method:

https://doi.org/10.3390/land15060976

$$F(\mathrm{X}) = \sum_{i=1}^{n} w_i \cdot \overline{f}_i(\mathrm{X}) \tag{16}$$

where X represents a candidate design solution generated by the Pix2PixHD-LoRA hybrid model; n denotes the total number of morphological indicators, which is set at 10; and $\overline{f}_i(\mathrm{X}) \in [0, 1]$ refers to the normalized score of the i-th indicator, ensuring that each metric's maximum contribution is limited to 1; w_i is the weight from Table 1 (sum of $w_i = 1$).

- Evolutionary Operations and Convergence

The optimization process begins with an initial population of 100 candidate solutions. Each iterative generation carries out the following core operations:

1. Elitism and Selection: The algorithm ranks the population in detail. An elite retention strategy selects and advances the top-performing solutions directly to the next generation, preserving optimal genetic traits.
2. Stochastic Variation: Using the elite phenotypes, new candidate solutions are synthesized via genetic operators such as crossover and mutation. This introduces morphological innovation and maintains diversity within the population.
3. Convergence Criteria: To ensure computational efficiency, the algorithm monitors the relative fitness gain over a sliding window of three consecutive generations. The process concludes when the improvement rate falls below a predetermined threshold, $\tau = 0.01$ (1%), indicating that the system has reached stochastic stabilization.

2.3. Case Study

This study utilizes the principle of theoretical sampling, focusing on the Yuehe Historic District in Jiaxing, Zhejiang, China, as its empirical case. The selection of this case is based on its dual representativeness. First, it serves as a typical example of a traditional Historic District in the Jiangnan water towns, featuring an intact spatial layout, well-preserved architecture, and a unique fabric of streets formed by water and markets shaped by rivers. This makes it an ideal spatial paradigm for research. Second, the district currently faces significant fabric decay, including fabric fragmentation, scale imbalance, and landscape disconnection. This scenario is a critical illustrative case that highlights the universal contradictions involved in preserving and developing historic districts amid rapid urbanization. By conducting an in-depth examination of this case, we aim to effectively validate the proposed methodology and uncover the intrinsic patterns that govern the evolution and preservation of the historic district's fabric. This will provide a solid empirical foundation for developing related theoretical frameworks.

The specific spatial scope of the empirical study is outlined in Figure 5. Within Jiaxing City's regulatory planning framework, this area is designated as a Construction Control Area rather than a highly restrictive Core Protection Area. This classification strategically positions the site as a critical transitional area for our research. Situated at the intersection of heritage conservation and urban development, the area confronts the common and urgent practical challenges in harmonizing architectural character and integrating the urban fabric. Consequently, selecting this district for empirical analysis robustly validates the applicability and operational efficacy of the proposed methodology in addressing complex real-world contradictions.

https://doi.org/10.3390/land15060976

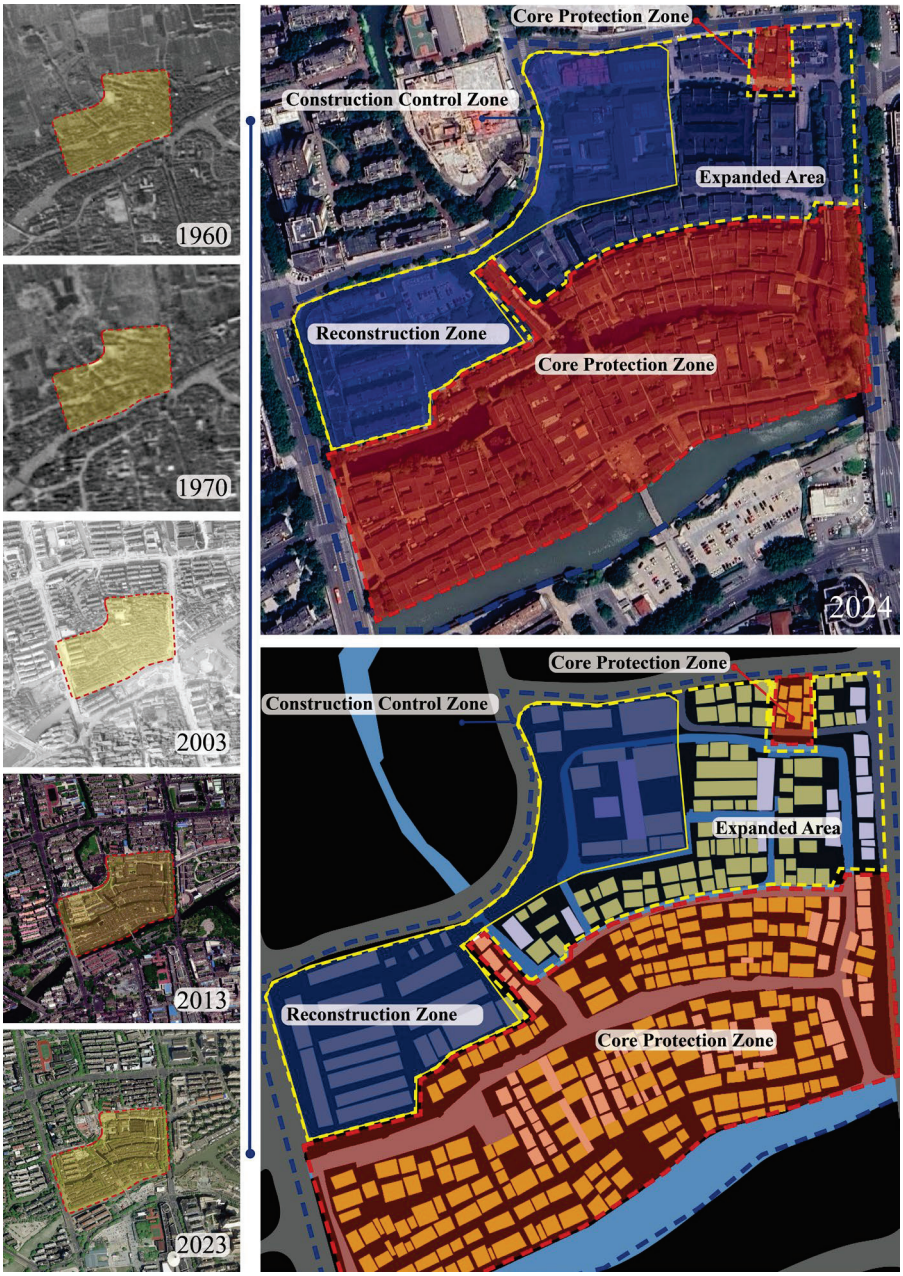


Figure 5. The designated renovation area within the Yuehe Historic District, Jiaxing.

3. Results

3.1. Comparative Evaluation of the Original Area, Expanded Area, and Optimal Generative Scheme

This study adopted an experimental approach to progressive fabric generation, focusing on the Construction Control Area of the Yuehe Historic District in Jiaxing. The workflow began with the Pix2PixHD network, which produced a high-precision initial morphological layout. A LoRA-tuned diffusion model was then applied to expand the solution space, generating 100 morphologically diverse candidates. Finally, a Multi-Objective Optimization algorithm iteratively enhanced each scheme. The optimization aimed to improve spatial performance while preserving the continuity of the historical urban context.

Before the comparative evaluation, a 5-fold cross-validation was performed on the 50 training cases to assess the stability of the generation and evaluation pipeline. In each fold, the Pix2PixHD-LoRA model was trained on four subsets and used to generate 100 candidate schemes for the held-out validation subset. All schemes were evaluated

using the weighted comprehensive score. Across the five folds, the comprehensive scores ranged from 0.83 to 0.85, with a mean of 0.84 and a standard deviation below 0.02. These small variations indicate that model performance does not depend strongly on a specific data split, which confirms the stability of the evaluation framework.

Based on this stable model, the optimal candidate scheme (GA_{opt}), which achieved a comprehensive score of 9.16, was selected from the optimization results (Figure 6). A tripartite quantitative comparison was then conducted among GA_{opt}, the Original Area, and the existing Expanded Area (Table 2).

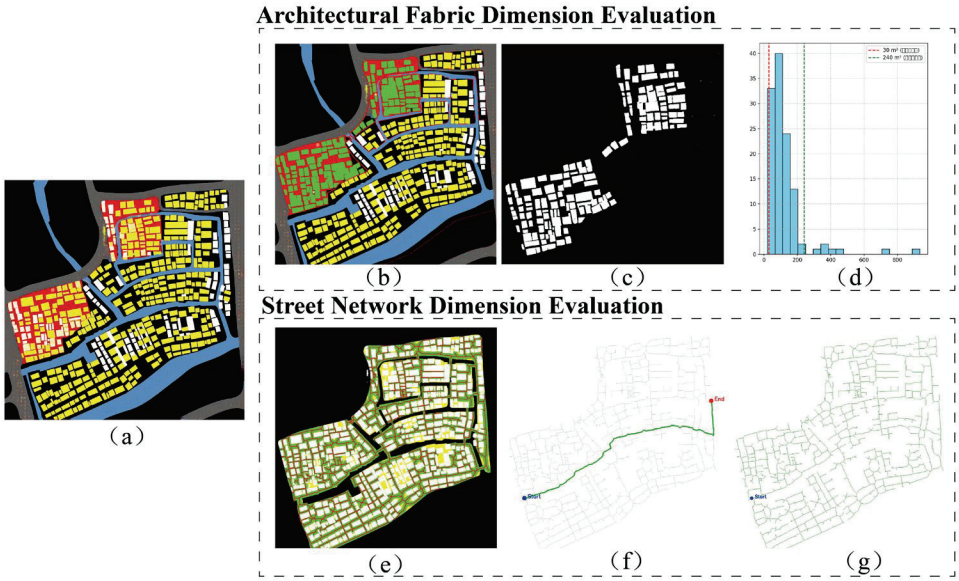


Figure 6. Quantitative evaluation of the architectural fabric and street network for the optimal generation scheme: (**a**) Original Site Conditions (red area: renovative area); (**b**) Fabric Coverage Rate; (**c**) Average Fabric Size and Standard Deviation; (**d**) Number of Small (Large) Fabric Units; (**e**) Average Street Width and Size Standard Deviation; (**f**) Minimum Cost Path; (**g**) Street Network Density and Spatio-Temporal Accessibility.

In the context of architectural fabric logic, the quantitative results precisely identify the degradation of historical morphological integrity caused by the existing Expanded Area while also showcasing the significant advantages of deep learning-driven reconstruction. Morphological measurements indicate that, while the Expanded Area attempts to replicate traditional scales, with a score of 0.98 for the Number of Small Fabric Units, its score for Standard Deviation of Fabric Size drops to 0. This indicates that the redevelopment in the Expanded Area adopted a highly repetitive, uniform modular building system, leading to a loss of the organic character associated with natural historical growth. In contrast, the optimal scheme, GA_{opt}, achieved a score of 0.99 in Standard Deviation of Fabric Size. This not only corrects the rigidity present in the Expanded Area but also exceeds the original average level of the Original Area (0.74). This shows that the coupled Pix2PixHD-LoRA model effectively extracted and reinforced scale fluctuation patterns within the spatial genotype, facilitating a transition from mechanical repetition to organic complexity. Furthermore, the Average Fabric Size score for GA_{opt} (0.99) significantly surpasses that of the Expanded Area (0.43), indicating that the deep learning model integrates necessary spatial hierarchies while preserving small-scale urban fabric.

https://doi.org/10.3390/land15060976

Table 2. Comparison of Quantitative Evaluation Scores among the Original Area, Expanded Area, and GA_{opt}.

Dimension	Evaluation Indicator	Original Area		Expanded Area		GA_{opt}	
		Raw *	Wtd. *	Raw	Wtd.	Raw	Wtd.
Architectural Fabric	Fabric Coverage Rate	0.93	0.11	0.91	0.11	1.00	0.12
	No. of Small Fabric Units	0.92	0.09	0.98	0.09	0.92	0.09
	No. of Large Fabric Units	0.88	0.07	0.58	0.04	0.86	0.06
	Average Fabric Size	0.89	0.11	0.43	0.05	0.99	0.12
	Standard Deviation of Fabric Size	0.74	0.06	0	0.00	0.99	0.08
Street Network	Street Network Density	0.91	0.10	0.87	0.10	0.95	0.11
	Minimum Cost Path	0.72	0.06	0.62	0.05	0.70	0.06
	Spatio-Temporal Accessibility	1	0.13	0.95	0.12	1.00	0.13
	Average Street Width	0.98	0.10	0.64	0.07	1.00	0.11
	Standard Deviation of Street Width	0.72	0.06	0.38	0.03	0.74	0.06
Total Score	\	8.69	0.89	6.36	0.66	9.16	0.94

* Raw = normalized indicator score (0–1). Wtd. = Raw × Weigh (rounded to two decimal places). The weighted total is the sum of Wtd. scores. Weights follow the updated Table 1. GA_{opt} improves over the Expanded Area by 42.4% in weighted total.

The evaluation of street network performance indicators further validates the effectiveness of the generated schemes in enhancing spatial accessibility and fostering a human-centric environment. Spatial analysis shows that the Expanded Area scores poorly on Average Street Width (0.64) and Standard Deviation of Street Width (0.38). This reflects that the existing developments, which prioritize modern motorized traffic, have adopted simplified and homogenized street-widening methods. As a result, this approach undermines the quality of the narrow street and dense network found in historical alleys, which serve as vital spaces for social interaction. In contrast, the optimal candidate scheme, GA_{opt}, shows a clear return to traditional pedestrian-centric logic in the street network. It achieves perfect scores for Spatio-temporal Accessibility and Average Street Width, both reaching the theoretical peak of 1.00, while Street Network Density (0.95) surpasses that of the Original Area (0.91). The data demonstrate that the optimization algorithm effectively filters out the rigid vehicular logic observed in the Expanded Area. Consequently, the new street network is characterized by high permeability and a natural rhythmic variation in street widths, reflected in a Street Width Standard Deviation of 0.74.

The comprehensive performance evaluation reveals that the Expanded Area has a total score of only 0.66, indicating a significant misalignment between fabric diversity and street network performance. In contrast, the GA_{opt} scheme achieved a total score of 0.94, marking a 42.4% improvement over the Expanded Area and exceeding the average quality of the Original Area (0.89). This result underscores a key academic finding: deep learning should not be viewed merely as a passive imitation of historical forms but rather as a performance-enhanced translation. By capturing the spatial genotype of the Original Area through deep learning and applying extreme-value screening via Multi-Objective Optimization, it is possible to develop optimal schemes that both respect historical context and outperform existing or even original conditions in spatial performance, including fabric and street network hierarchy. This approach provides a quantifiable, empirical means to address the contradictions between the urban context and functional degradation in the Construction Control Area of historic urban areas.

https://doi.org/10.3390/land15060976

3.2. Comparative Performance Analysis of Generative Paradigms

This section compares three generative architectures: the standalone Pix2PixHD (deterministic), the standalone LoRA (stochastic), and the proposed Pix2PixHD–LoRA hybrid. Figure 7 presents a stacked bar chart comparing the evaluation metrics across the Original Area, Expanded Area, and Generated Area.

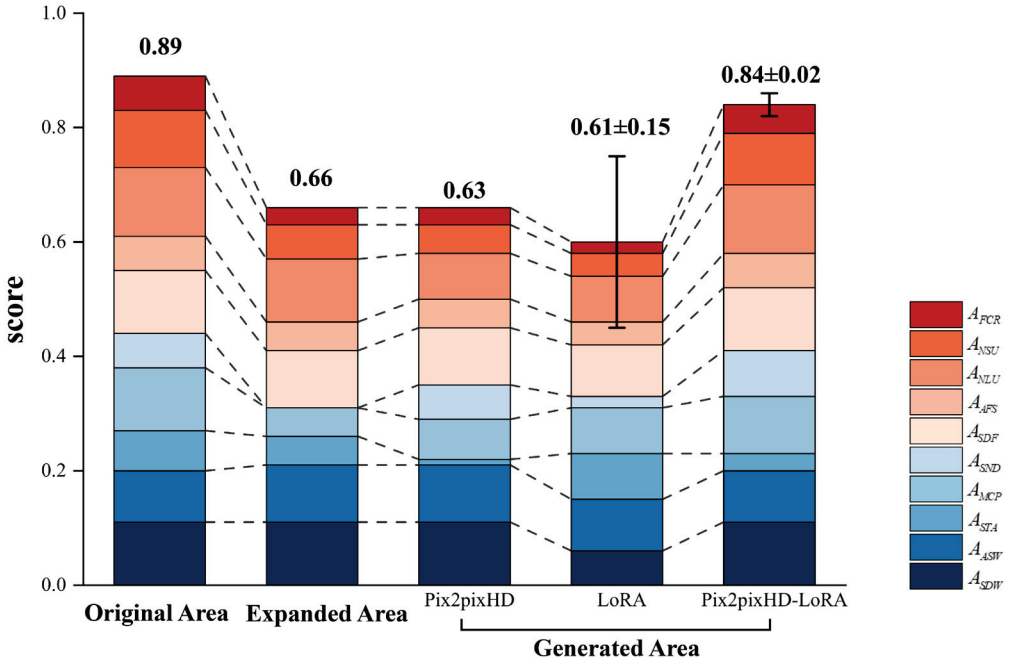


Figure 7. Stacked bar chart comparing evaluation metrics across the Original Area, Expanded Area, and Generated Area.

The robustness of the stochastic methods was assessed through multiple independent runs. Table 3 summarizes the weighted comprehensive scores. For the hybrid model, the mean score across 100 generated samples is 0.84. The intra-run standard deviation is ±0.07, and the scores range from 0.45 to 0.90. The cross-run standard deviation is ±0.02, indicating stable model performance across different training runs.

Table 3. Stability of the weighted comprehensive score for each generative method based on independent runs.

Statistic	Pix2PixHD (n = 1)	LoRA (n = 40)	Pix2PixHD–LoRA (n = 100)
Mean score	0.63	0.61	0.84
Intra-run standard deviation	–	±0.15	±0.07
Intra-run range (min–max) *	–	0.39–0.79	0.45–0.90
Cross-run standard deviation	–	–	±0.02

* Pix2PixHD is deterministic and yields a single output. LoRA and Pix2PixHD–LoRA statistics are computed over 40 and 100 independent samples, respectively. Intra-run standard deviation and range reflect the diversity of generated candidates within a single generation run. Cross-run standard deviation (±0.02 for the hybrid model) reflects the stability of the mean score across multiple complete training runs, as estimated by 5-fold cross-validation (Section 3.1).

Detailed per-indicator scores, benchmarked against the Original Area and the Expanded Area, are reported in Table 4. This analysis examines the trade-offs between morphological fidelity, functional plausibility, and generative diversity (see Appendix A).

https://doi.org/10.3390/land15060976

Table 4. Comparison of Evaluation Indicator Scores for the Pix2PixHD, LoRA, and Pix2PixHD-LoRA.

Dimension	Evaluation Indicator	Pix2PixHD		LoRA		Pix2PixHD-LoRA	
		Raw *	Wtd. *	Raw	Wtd.	Raw	Wtd.
Architectural Fabric	Fabric Coverage Rate	0.88	0.10	0.46 ± 0.15	0.05 ± 0.02	0.91 ± 0.02	0.11 ± 0.00
	No. of Small Fabric Units	0.96	0.09	0.88 ± 0.12	0.08 ± 0.01	0.90 ± 0.06	0.09 ± 0.01
	No. of Large Fabric Units	0.10	0.01	0.98 ± 0.02	0.07 ± 0.00	0.33 ± 0.06	0.02 ± 0.00
	Average Fabric Size	0.55	0.07	0.64 ± 0.17	0.08 ± 0.02	0.85 ± 0.11	0.10 ± 0.01
	Standard Deviation of Fabric Size	0.71	0.06	0.26 ± 0.15	0.02 ± 0.01	0.95 ± 0.05	0.08 ± 0.00
Street Network	Street Network Density	0.83	0.09	0.79 ± 0.13	0.09 ± 0.01	0.94 ± 0.01	0.10 ± 0.00
	Minimum Cost Path	0.66	0.05	0.47 ± 0.32	0.04 ± 0.03	0.73 ± 0.01	0.06 ± 0.00
	Spatio-Temporal Accessibility	0.66	0.08	0.63 ± 0.34	0.08 ± 0.04	0.96 ± 0.07	0.12 ± 0.01
	Average Street Width	0.48	0.05	0.35 ± 0.36	0.04 ± 0.04	0.88 ± 0.05	0.09 ± 0.01
	Standard Deviation of Street Width	0.33	0.03	0.21 ± 0.24	0.02 ± 0.02	0.60 ± 0.08	0.05 ± 0.01
Total Score	\	6.17	0.63	5.38 ± 1.64	0.61 ± 0.15	8.06 ± 0.19	0.84 ± 0.02

* Raw = normalized indicator score (0–1). Wtd. = Raw × Weigh (rounded to two decimal places). The weighted total is the sum of Wtd. scores. Values are mean ± standard deviation over 100 candidate solutions per generation. LoRA and Pix2PixHD-LoRA values represent mean ± standard deviation over 40 and 100 generated samples, respectively. The Pix2PixHD-LoRA total score standard deviation (±0.02) reflects cross-run model stability (see Table 5 for intra-population diversity).

Table 5. Evaluation Comparison of Pix2PixHD-LoRA Multi-Objective Optimization Results.

Dimension	Evaluation Indicator	Pix2PixHD-LoRA-1		Pix2PixHD-LoRA-2		Pix2PixHD-LoRA-3	
		Raw *	Wtd. *	Raw	Wtd.	Raw	Wtd.
Architectural Fabric	Fabric Coverage Rate	0.94 ± 0.05	0.11 ± 0.01	0.99 ± 0.02	0.12 ± 0.00	0.99 ± 0.02	0.12 ± 0.00
	No. of Small Fabric Units	0.87 ± 0.09	0.08 ± 0.01	0.9 ± 0.06	0.09 ± 0.01	0.88 ± 0.08	0.08 ± 0.01
	No. of Large Fabric Units	0.79 ± 0.06	0.06 ± 0.00	0.8 ± 0.06	0.06 ± 0.00	0.82 ± 0.06	0.06 ± 0.00
	Average Fabric Size	0.81 ± 0.16	0.10 ± 0.02	0.87 ± 0.12	0.11 ± 0.01	0.86 ± 0.11	0.10 ± 0.01
	Standard Deviation of Fabric Size	0.68 ± 0.26	0.05 ± 0.02	0.69 ± 0.23	0.05 ± 0.02	0.65 ± 0.23	0.05 ± 0.02
Street Network	Street Network Density	0.91 ± 0.05	0.10 ± 0.01	0.94 ± 0.01	0.10 ± 0.00	0.95 ± 0.01	0.11 ± 0.00
	Minimum Cost Path	0.67 ± 0.13	0.06 ± 0.01	0.7 ± 0	0.06 ± 0.00	0.7 ± 0	0.06 ± 0.00
	Spatio-Temporal Accessibility	0.9 ± 0.14	0.11 ± 0.02	0.96 ± 0.03	0.12 ± 0.00	0.98 ± 0.03	0.12 ± 0.00
	Average Street Width	0.91 ± 0.15	0.10 ± 0.02	1 ± 0	0.11 ± 0.00	1 ± 0	0.11 ± 0.00
	Standard Deviation of Street Width	0.65 ± 0.14	0.06 ± 0.01	0.74 ± 0.03	0.06 ± 0.00	0.75 ± 0.02	0.07 ± 0.00
Total Score *	\	8.08 ± 0.95	0.84 ± 0.07	8.58 ± 0.32	0.89 ± 0.03	8.58 ± 0.33	0.89 ± 0.02

* Raw = normalized indicator score (0–1). Wtd. = Raw × Weigh (rounded to two decimal places). The weighted total is the sum of Wtd. scores. Values are mean ± standard deviation over 100 candidate solutions per generation. Data represent mean ± standard deviation across the population of 100 candidate solutions at each generation. Generation 1 corresponds to the initial Pix2PixHD-LoRA outputs, Generations 2 and 3 are the first and second rounds of multi-objective optimization. Data represents the mean ± standard deviation of multiple runs of the hybrid model ($n = 100$).

The standalone Pix2PixHD model operates as a deterministic generator, producing a single output, highlighting its fundamental limitations. While it partially reduces the morphological homogeneity of the Expanded Area—restoring the fabric size standard deviation from 0.00 to 0.71 and maintaining high fabric coverage (0.88)—it struggles to replicate the complex scalar hierarchy of the historic tissue. Specifically, the generation of large fabric units is abnormally low (0.10), and the average fabric size (0.55) remains sub-optimal, indicating an inherent deficiency in capturing low-frequency structural patterns.

https://doi.org/10.3390/land15060976

Critically, its performance in street network synthesis falls short of the Original Area in both spatio-temporal accessibility (0.66) and street-width control (0.48). This suggests that the model has difficulty coordinating complex functional dependencies. Furthermore, the monotony of its outputs—providing no alternative design scenarios—severely limits its practical use in design decision support.

On the other hand, the standalone LoRA model exhibits robust generative diversity, producing a wide variety of morphologically distinct phenotypes. However, quantitative metrics reveal critical instability (Total Score: 0.61 ± 0.15). The model not only significantly underperforms the Original Area but also, in key metrics such as fabric coverage (0.46 ± 0.15), falls short of the Expanded Area. The substantial standard deviations observed across street network indicators (consistently exceeding ± 0.24) imply that outputs fluctuate dramatically between complete structural failure and accidental viability. While LoRA effectively serves as a creative engine for stylistic exploration, its inherent geometric inaccuracies—resulting in blurred building boundaries and topological discontinuities in the street network—render it unreliable for autonomous planning tasks.

The proposed Pix2PixHD-LoRA hybrid framework successfully addresses the existing dichotomy in image generation, achieving a unified equilibrium of fidelity, diversity, and stability.

- Morphological Fidelity: Pix2PixHD-LoRA significantly outperforms Pix2PixHD across all key metrics, closely matching the benchmarks of the Original Area. Notably, the average fabric size score of 0.85 represents a 54.5% improvement over Pix2PixHD. Furthermore, the standard deviation of fabric size (0.95 ± 0.05) exceeds that of the Original Area (0.74), showcasing a superior ability to generate heterogeneous yet harmonious fabric patterns.
- Functional Integrity: In terms of street network functionality, Pix2PixHD-LoRA marks a significant advancement. Both spatio-temporal accessibility (0.96) and network density (0.94) approach the optimal values found in the Original Area. The average street width score of 0.88 reflects an impressive 83.3% improvement over Pix2PixHD, indicating precise control over spatial scale.
- Stability and Practicality: Most importantly, Pix2PixHD-LoRA achieves a high overall score of 0.84 with an exceptionally low standard deviation of ± 0.02 across 100 generated iterations.

This outcome holds substantial methodological significance: it shows that the hybrid framework does not sacrifice quantity for quality. Instead, the high-precision structural prior provided by Pix2PixHD effectively constrains the LoRA model's latent space, guiding sampling towards high-quality regions of the solution space. Consequently, designers receive not a chaotic assortment of random images but a carefully curated set of valid design alternatives that uphold high morphological consistency and functional rigor.

Comparative analysis indicates that both Pix2PixHD and LoRA exhibit significant limitations when used alone. Pix2PixHD struggles with mode collapse, leading to a lack of diversity in its outputs. On the other hand, LoRA faces issues with geometric hallucination and instability. The proposed hybrid model, with its cascaded architecture, achieves a synergistic effect beyond a simple combination of the two systems. It provides a superior balance across three critical areas: morphological accuracy, functional rationality, and generative diversity. This innovative framework creates a new digital generation paradigm that harmonizes algorithmic creativity with engineering precision, effectively addressing the dual challenges of preserving traditional urban fabric and adapting to modern functional requirements in historic district regeneration.

https://doi.org/10.3390/land15060976

3.3. Analysis of Multi-Objective Optimization Results

To rigorously validate the comprehensive efficacy of the multi-objective iterative optimization framework in enhancing generative schemes for historic districts, this study conducted multiple rounds of evolutionary searches using an equal-weighted evaluation system. The goal was to synergistically optimize both the architectural fabric and street network performance, utilizing an initial population of candidate solutions generated by the Pix2PixHD-LoRA hybrid model.

After three generations of iteration, the optimization process met the convergence criteria, with the final intergenerational improvement rate of 0.7%, which is below the threshold of $\tau = 0.01$. This demonstrates systematic improvements across all dimensional metrics (Detailed in Table 5 and Appendix B).

In the architectural fabric dimension, the fabric coverage score increased from 0.94 ± 0.05 to 0.99 ± 0.02, while the average fabric size score rose from 0.81 ± 0.16 to 0.86 ± 0.11. Notably, the mean metric score for the Standard Deviation of Fabric Size remained within the stable range of 0.65–0.69, indicating a more rational fabric composition and enhanced morphological order. The score for the number of large fabric units also improved from 0.79 ± 0.06 to 0.82 ± 0.06, suggesting further enhancement in functional integrity.

In terms of the street network, the density score improved from 0.91 ± 0.05 to 0.95 ± 0.01, and spatio-temporal accessibility increased from 0.90 ± 0.14 to 0.98 ± 0.03. The average street width score reached its maximum value of 1.00, up from 0.91 ± 0.15, while the standard deviation of the street width score also improved to 0.75 ± 0.02. These results demonstrate an overall optimization of the street network's structure with respect to connectivity efficiency and scalar coordination.

Through this optimization process, the comprehensive fitness score of the solution population increased from 0.84 ± 0.07 to 0.89 ± 0.02. This significant increase, along with a notable reduction in standard deviation, indicates that the optimization framework not only enhances overall spatial performance but also improves the stability and reliability of the generative outputs.

Conclusion: The integrated generative-optimization framework developed in this study effectively guides the initial solution set toward a Pareto-optimal frontier, balancing spatial performance and historical authenticity through a multi-objective iterative mechanism. This approach provides a technically feasible and controllable digital paradigm for the organic renewal of historic districts. The results demonstrate that the proposed methodology significantly enhances the comprehensive performance of generated solutions across multiple dimensions—including morphological rationality, street network efficiency, and output stability—while strictly preserving morphological fidelity.

4. Discussion

Comparison with previous studies. Earlier work often relied on a single generative model. For example, some studies used Pix2Pix or Pix2PixHD to produce urban layouts. These models achieved reasonable structural accuracy, but their outputs showed limited diversity [15,17,18]. Other researchers applied standalone diffusion models for floor plans or façades. These models created varied styles but often lost precise geometric control [6,23]. Our hybrid approach combines the two. It keeps the structural stability of GANs and adds diverse variations in diffusion models. The performance-driven optimization further selects the best-fitting schemes. The results show that our generated schemes outperform the expanded area in several metrics, such as standard deviation of fabric size and street network density. The key new finding is that a high-fidelity structural prior from Pix2PixHD can guide the diffusion model's latent space. This guidance produces stable, high-quality

https://doi.org/10.3390/land15060976

candidates, not random images. This moves beyond passive imitation and achieves a measurable improvement in urban fabric quality.

The generative design methodology proposed in this study has shown promising capabilities for synthesizing historic urban fabric and street network morphology. However, several limitations remain, highlighting important areas for future research.

Limitations of model comparison. Our experiments compared only three architectures: Pix2PixHD, LoRA, and the proposed hybrid model. Many other advanced generative models exist, such as ControlNet-guided diffusion models, transformer-based layout generators, and graph neural networks for spatial synthesis. These models were not tested. Therefore, the current study cannot show how our framework performs relative to the latest alternatives. This limitation should be addressed in future work by including a broader benchmark.

Technological Workflow and Computational Efficiency: Currently, the framework operates across heterogeneous software ecosystems, integrating Pix2PixHD training, LoRA fine-tuning, and optimization algorithms across different platforms. This fragmentation results in operational complexity and hinders seamless automation. Future work should focus on developing a unified computational framework (Integrated Platform) that streamlines module coordination through interoperable protocols. From a computational perspective, training high-fidelity GANs (e.g., Pix2PixHD) is intensive. Although LoRA reduces the adaptation overhead, overall latency still poses a challenge for real-time interaction. Future research will investigate techniques for model lightweighting—such as knowledge distillation or network pruning—to create more efficient generative architectures suitable for rapid iterative design.

Data Generalizability and Multidimensional Evaluation: While the current dataset systematically captures the river–street typologies of the Jiangnan region, its geographical scope is limited. This constrains the model's generalization to other regional architectural styles (e.g., Northern courtyard dwellings or mountainous settlements). Future efforts should expand the dataset to include cross-regional and multicultural samples to enhance the model's domain adaptation capabilities. Regarding the evaluation system, current metrics focus primarily on morphological aspects. Key performance indicators, such as disaster resilience (specifically emergency evacuation efficiency in narrow alleys), economic feasibility, and microclimatic comfort (e.g., wind and thermal performance), have not yet been fully integrated. Existing research provides a foundation for quantifying such environmental and behavioral dimensions, from pedestrian wind environments and layout patterns [34–36] to public willingness for sustainable transitions [37] to indoor environmental quality, occupant satisfaction [38,39], HVAC-related energy consumption [38,40], and health outcomes such as sick building syndrome in various building types [39,41]. Developing a holistic evaluation framework that encompasses social, economic, and environmental dimensions is essential. Furthermore, the optimization phase currently uses a conservative elite retention rate (1%). Future iterations should consider increasing this parameter (e.g., to 20%) to better balance global exploration (diversity) and local exploitation (stability), thereby enhancing the robustness of optimization convergence.

By addressing these limitations through systematic improvements, generative design methodologies can advance toward greater integration, universality, efficiency, and comprehensiveness in the field of urban heritage conservation.

https://doi.org/10.3390/land15060976

5. Conclusions

This study proposes an innovative computational framework that synergizes deep learning with Multi-Objective Optimization to address the complex digital design challenges posed by the preservation and regeneration of historic urban fabric. By systematically integrating the high-fidelity generative capabilities of Pix2PixHD, the diverse expressive potential of Diffusion Models, and the goal-oriented decision-making of optimization algorithms, this framework effectively bridges the longstanding methodological disconnect between morphological accuracy, generative diversity, and functional rationality.

The key findings and contributions are summarized as follows:

- Establishment of a Closed-Loop Generative-Optimization Framework

This study addresses the critical dilemma in current digital design methodologies: the trade-offs between morphological fidelity and functional performance. It establishes a comprehensive, reproducible technical pathway that systematically integrates multimodal dataset construction, dual-stage model synergism, quantitative evaluation, and iterative optimization. This framework not only ensures high fidelity between the synthesized outcomes and the authentic historical fabric but also provides robust methodological support for human–machine collaborative decision-making.

- Superiority of the Hybrid Generative Architecture

The proposed Pix2PixHD-LoRA hybrid model shows significant advantages over standalone paradigms. Compared to the deterministic Pix2PixHD model, the hybrid approach maintains high historical fidelity and geometric precision while substantially enhancing generative diversity, effectively addressing the mode collapse issue commonly found in GANs. Compared with standalone Diffusion/LoRA approaches, the hybrid method effectively constrains topological inconsistencies and geometric distortions by incorporating structural priors from Pix2PixHD, achieving an organic integration of algorithmic creativity and engineering controllability.

- Efficacy of Performance-Driven Optimization

The Multi-Objective Optimization process substantially improved the overall spatial performance of the generated solutions. By coupling a quantitative evaluation system based on ten morphological indicators with an evolutionary strategy, the framework achieved synergistic improvements across conflicting objectives, such as fabric compactness and street network efficiency. The top-ranked individual design (GA_{opt}) achieved a weighted comprehensive score of 0.94, a 42.4% improvement over the Expanded Area (0.66). The optimized population reached a mean weighted score of 0.89 ± 0.02, representing a 34.8% gain over the Expanded Area and a 6.0% increase over the initial generated baseline (0.84 ± 0.07). Furthermore, the optimization process significantly enhanced output stability, reducing the standard deviation from 0.07 to 0.02, indicating high reliability.

Limitations. First, the model comparison was limited to three architectures. Future benchmarks should include more advanced generative models. Second, the dataset represents a single regional typology, the Jiangnan river–street fabric. Broader cultural and geographic coverage is needed. Third, the evaluation metrics focus on physical form. Social, economic, and environmental dimensions, such as disaster resilience and microclimatic comfort, are not yet integrated. Fourth, the computational workflow is not yet unified, which limits interactive speed.

https://doi.org/10.3390/land15060976

Future research. Future work should test a wider set of generative models, expand the dataset to multiple regions, and develop an integrated evaluation framework that includes disaster, comfort, and cost indicators. Model lightweighting techniques and a unified software platform will be explored to enable real-time, interactive design.

Implications. This framework can be adapted to other historic environments that face similar fabric-based challenges. The method provides a quantifiable, repeatable way to balance preservation and renewal. Planners can use the generated and optimized alternatives as design references. This makes the decision process more transparent and evidence-based. The approach can also be extended to other urban design tasks where morphological authenticity and performance must be balanced.

In summary, the generative design framework developed in this study achieves an effective equilibrium among morphological fidelity, design diversity, and performance optimization. It provides both theoretical insights and a practical digital paradigm for the scientific preservation and adaptive regeneration of historic districts amid rapid urbanization.

Author Contributions: Conceptualization, X.Y. and J.W.; methodology, X.Y. and Y.Z.; software, S.N. and R.Q.; validation, Y.Z. and J.B.; formal analysis, Y.Z. and T.L.; investigation, S.N. and R.Q.; resources, H.L., R.Q. and H.Z.; data curation, S.N. and T.L.; writing—original draft preparation, X.Y. and S.N.; writing—review and editing, J.W., H.L., and H.Z.; visualization, S.N. and J.B.; supervision, J.W.; project administration, J.W.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant numbers 52508143 and 52478034), the Joint Fund of Natural Science Foundation of Zhejiang Province (grant number LLSSZ26F020002, LHZY24A010003 and LHZQN25D010005).

Data Availability Statement: The data presented in this study are not publicly available due to privacy and ethical restrictions.

Acknowledgments: I acknowledge the use of Gemini in assisting with the preparation and/editing of the language and grammar in this paper.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

GANs	Generative Adversarial Networks
LoRA	Low-Rank Adaptation
EA	Expanded Area
GIS	Geographic Information System
GA_{opt}	Optimal Generative Scheme

Appendix A

Appendix A.1

This appendix presents the processed training images (Domain B) utilized for model training.

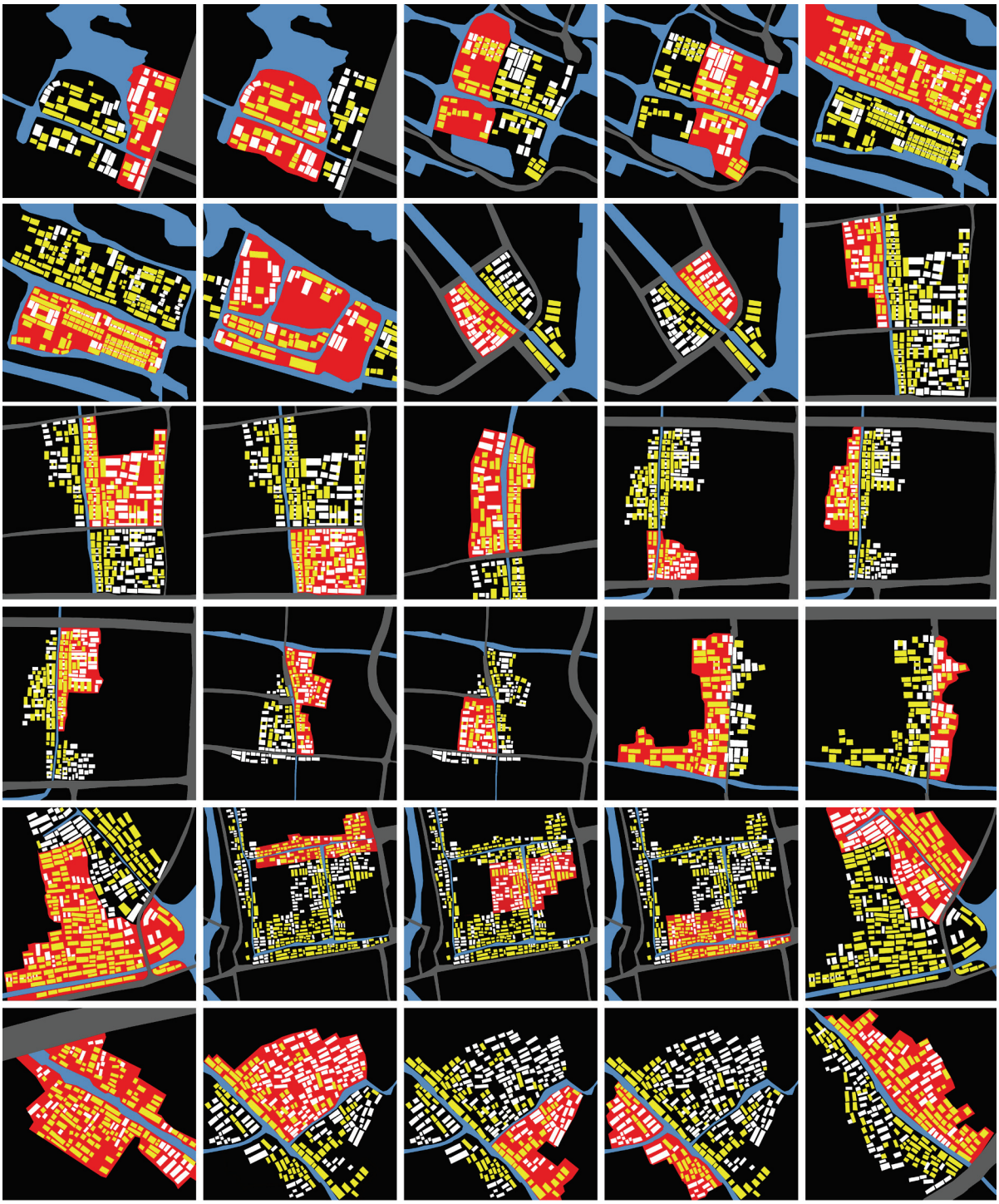


Figure A1. A subset of Dataset Preparation. Overview of the generation process and color-coding scheme for the paired dataset. The multi-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for background, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

Appendix A.2

This appendix provides a visual comparison of generation outputs across different model architectures.

https://doi.org/10.3390/land15060976



Figure A2. A subset of generation Result of Pix2PixHD. Overview of the generation process and color-coding scheme for the paired dataset. The mul-ti-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for back-ground, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

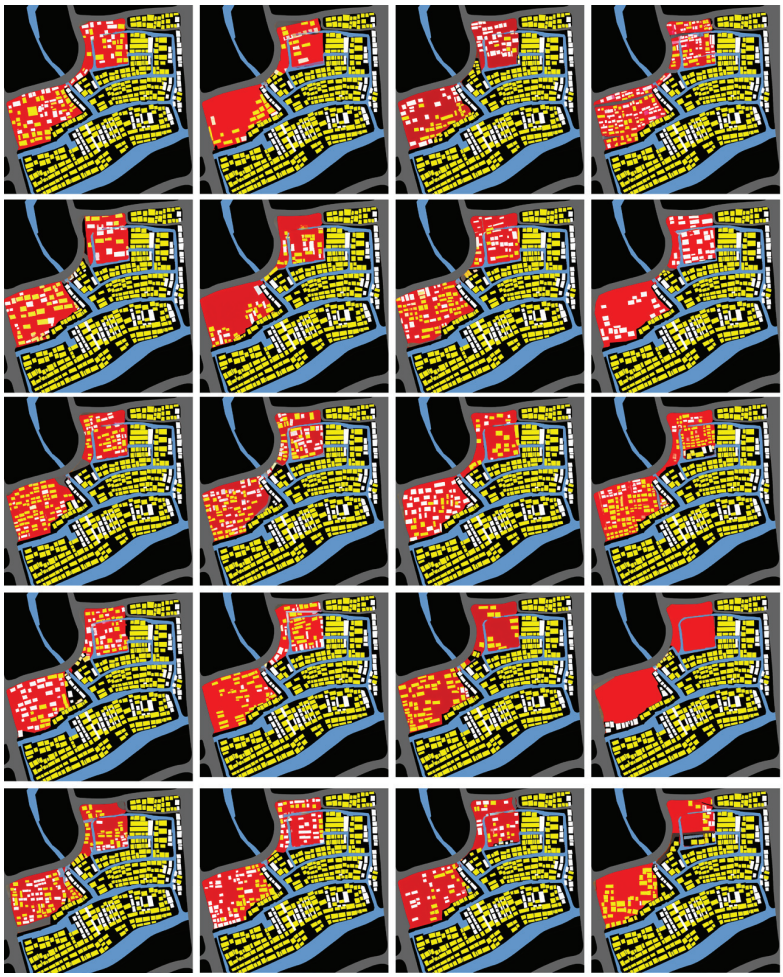


Figure A3. A subset of generation Results of LoRA. Overview of the generation process and color-coding scheme for the paired dataset. The multi-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for background, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

https://doi.org/10.3390/land15060976

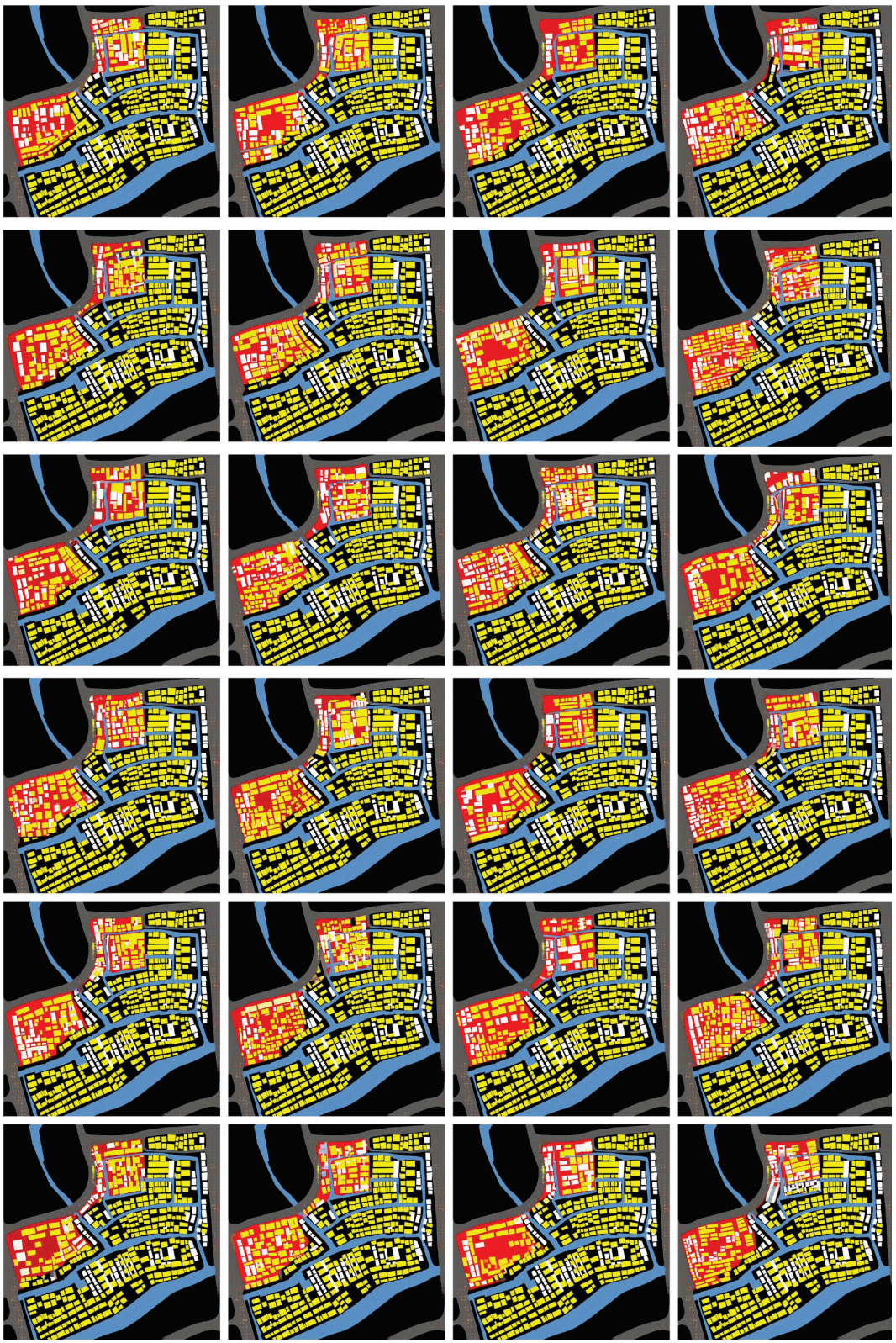


Figure A4. A subset of generation Results of Pix2PixHD-LoRA-1. Overview of the generation process and color-coding scheme for the paired dataset. The multi-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for background, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

https://doi.org/10.3390/land15060976

Table A1. Generation Results of Deep Learning Model.

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
A1	0.88	0.96	0.1	0.55	0.71	0.83	0.66	0.66	0.48	0.33	6.17
B1	0.58	0.90	0.96	0.94	0.54	0.93	0.72	1.00	0.73	0.44	7.73
B2	0.57	1.00	0.94	0.92	0.42	0.91	0.70	1.00	0.70	0.49	7.65
B3	0.63	0.92	0.98	0.72	0.38	0.94	0.72	1.00	0.82	0.54	7.64
B4	0.55	0.96	0.96	0.79	0.28	0.94	0.71	1.00	0.78	0.54	7.49
B5	0.61	0.90	0.98	0.73	0.27	0.92	0.71	1.00	0.76	0.48	7.36
B6	0.65	0.82	0.94	0.55	0.32	0.92	0.72	0.98	0.82	0.51	7.23
B7	0.55	0.86	1.00	0.52	0.20	0.93	0.71	1.00	0.78	0.52	7.08
B8	0.64	0.72	1.00	0.44	0.10	0.93	0.72	1.00	0.87	0.55	6.95
B9	0.56	0.84	1.00	0.43	0.09	0.93	0.71	1.00	0.80	0.49	6.86
B10	0.57	0.50	0.98	0.36	0.33	0.95	0.72	1.00	0.86	0.53	6.80
B11	0.62	1.00	0.98	0.63	0.23	0.85	0.72	0.88	0.58	0.30	6.79
B12	0.57	0.52	1.00	0.32	0.07	0.97	0.71	1.00	0.90	0.64	6.71
B13	0.59	0.52	1.00	0.31	0.06	0.97	0.71	1.00	0.91	0.63	6.69
B14	0.58	0.88	1.00	0.53	0.16	0.87	0.72	0.95	0.57	0.40	6.66
B15	0.54	0.90	0.98	0.65	0.26	0.85	0.72	0.94	0.48	0.26	6.59
B16	0.44	0.92	0.96	0.69	0.39	0.85	0.72	0.95	0.40	0.26	6.58
B17	0.55	0.78	1.00	0.45	0.11	0.89	0.71	0.98	0.64	0.38	6.50
B18	0.52	0.86	0.98	0.66	0.53	0.84	0.71	0.84	0.41	0.08	6.42
B19	0.55	0.74	0.98	0.42	0.17	0.86	0.72	0.93	0.59	0.39	6.35
B20	0.36	0.94	1.00	0.51	0.12	0.84	0.70	0.76	0.30	0.10	5.63
B21	0.33	0.96	0.98	0.94	0.34	0.63	0.61	0.28	0.00	0.00	5.06
B22	0.26	0.96	0.90	0.71	0.71	0.64	0.46	0.28	0.00	0.00	4.92
B23	0.38	0.86	0.98	0.73	0.28	0.58	0.62	0.28	0.00	0.00	4.71
B24	0.33	0.92	1.00	0.50	0.11	0.65	0.67	0.33	0.00	0.00	4.51
B25	0.36	0.92	0.94	0.83	0.55	0.62	0.07	0.22	0.00	0.00	4.51
B26	0.12	0.94	1.00	0.59	0.23	0.57	0.74	0.25	0.00	0.00	4.45
B27	0.06	1.00	1.00	0.75	0.09	0.57	0.62	0.29	0.00	0.00	4.39
B28	0.28	0.92	1.00	0.66	0.17	0.62	0.19	0.25	0.00	0.00	4.10
B29	0.06	0.98	1.00	0.49	0.09	0.58	0.52	0.27	0.00	0.00	3.99
B30	0.45	1.00	1.00	0.96	0.31	0.78	0.00	0.44	0.01	0.00	3.72
B31	0.47	0.96	0.96	0.86	0.32	0.71	0.00	0.29	0.00	0.00	3.57
B32	0.41	0.98	0.96	0.79	0.42	0.75	0.00	0.41	0.00	0.00	3.56
B33	0.40	0.98	1.00	0.84	0.24	0.72	0.00	0.28	0.00	0.00	3.46
B34	0.49	0.92	1.00	0.81	0.23	0.77	0.00	0.44	0.10	0.00	3.44
B35	0.46	0.90	0.98	0.71	0.29	0.75	0.00	0.28	0.00	0.00	3.33
B36	0.49	0.90	0.98	0.62	0.31	0.76	0.00	0.32	0.09	0.00	3.30
B37	0.52	0.88	1.00	0.60	0.26	0.78	0.00	0.27	0.16	0.00	3.25
B38	0.41	0.92	1.00	0.62	0.24	0.60	0.00	0.23	0.00	0.00	3.19
B39	0.40	0.90	0.98	0.60	0.21	0.78	0.00	0.46	0.04	0.00	3.08
B40	0.36	0.86	1.00	0.48	0.16	0.75	0.00	0.29	0.00	0.00	2.85
C1	0.98	0.84	0.82	0.97	0.97	0.94	0.70	0.99	1.00	0.73	8.94
C2	1.00	0.96	0.82	0.84	0.97	0.94	0.70	0.98	1.00	0.74	8.94
C3	1.00	0.80	0.80	0.97	0.98	0.95	0.70	0.93	1.00	0.78	8.90
C4	1.00	0.82	0.76	0.95	0.97	0.94	0.70	0.97	1.00	0.70	8.82
C5	0.99	0.98	0.76	0.86	0.95	0.94	0.70	0.97	1.00	0.67	8.82
C6	0.92	0.80	0.86	0.88	0.99	0.94	0.70	1.00	1.00	0.73	8.82
C7	0.89	0.92	0.78	0.99	0.93	0.94	0.70	0.98	0.99	0.70	8.81
C8	0.94	0.82	0.74	0.96	0.96	0.94	0.70	1.00	1.00	0.73	8.78
C9	0.93	0.86	0.82	0.89	0.92	0.95	0.70	0.98	1.00	0.73	8.78

https://doi.org/10.3390/land15060976

Table A1. *Cont.*

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C10	0.96	0.74	0.84	0.91	0.96	0.95	0.70	0.88	1.00	0.77	8.71
C11	0.92	0.84	0.74	0.91	0.92	0.95	0.70	1.00	1.00	0.73	8.70
C12	0.94	0.84	0.86	0.95	0.96	0.93	0.70	0.90	0.98	0.66	8.70
C13	1.00	0.96	0.78	0.93	0.68	0.94	0.70	0.97	1.00	0.73	8.68
C14	0.91	0.86	0.90	0.88	0.74	0.95	0.70	1.00	1.00	0.73	8.67
C15	0.87	0.88	0.84	1.00	0.81	0.94	0.70	0.97	0.97	0.69	8.67
C16	1.00	0.90	0.82	0.96	0.62	0.93	0.70	0.99	1.00	0.74	8.65
C17	0.95	0.84	0.86	0.90	0.81	0.93	0.70	0.94	1.00	0.72	8.65
C18	1.00	0.96	0.66	0.71	0.94	0.94	0.70	1.00	1.00	0.74	8.65
C19	0.97	0.88	0.82	0.98	0.65	0.94	0.70	0.99	1.00	0.71	8.65
C20	0.98	0.94	0.72	0.89	0.74	0.93	0.70	0.99	1.00	0.76	8.63
C21	0.98	0.94	0.82	0.85	0.69	0.94	0.70	1.00	1.00	0.70	8.63
C22	1.00	0.82	0.84	0.79	0.84	0.95	0.70	0.89	1.00	0.79	8.63
C23	0.98	0.84	0.88	0.96	0.64	0.94	0.70	0.93	1.00	0.74	8.61
C24	0.93	0.98	0.84	0.97	0.51	0.95	0.70	1.00	1.00	0.71	8.60
C25	0.98	0.96	0.80	0.92	0.61	0.94	0.70	0.96	1.00	0.72	8.60
C26	1.00	0.80	0.84	0.96	0.58	0.95	0.70	0.98	1.00	0.78	8.58
C27	0.99	0.92	0.78	0.92	0.62	0.93	0.70	0.99	1.00	0.71	8.57
C28	0.94	0.84	0.88	0.98	0.52	0.95	0.70	1.00	1.00	0.76	8.57
C29	0.98	0.86	0.82	0.88	0.92	0.91	0.70	0.90	0.96	0.63	8.55
C30	0.86	0.94	0.82	0.86	0.86	0.94	0.70	1.00	0.94	0.63	8.55
C31	0.96	0.92	0.84	0.89	0.60	0.95	0.70	0.96	1.00	0.74	8.55
C32	0.84	0.86	0.84	0.99	0.75	0.95	0.70	0.96	0.97	0.69	8.55
C33	0.83	0.92	0.84	0.89	0.87	0.94	0.69	0.97	0.95	0.64	8.54
C34	0.93	0.88	0.80	0.88	0.67	0.94	0.70	1.00	1.00	0.73	8.53
C35	0.96	0.68	0.84	0.81	0.98	0.95	0.70	0.84	1.00	0.76	8.53
C36	0.93	0.92	0.84	0.89	0.63	0.94	0.70	0.96	1.00	0.71	8.52
C37	0.98	0.88	0.68	0.69	0.98	0.93	0.71	0.94	1.00	0.72	8.49
C38	0.90	0.98	0.82	0.71	0.84	0.94	0.70	0.97	0.94	0.69	8.49
C39	0.96	0.88	0.82	0.84	0.72	0.93	0.70	1.00	0.99	0.66	8.49
C40	0.93	0.76	0.84	0.91	0.84	0.92	0.70	0.97	0.97	0.66	8.48
C41	0.89	0.98	0.84	0.89	0.77	0.92	0.70	0.98	0.90	0.61	8.48
C42	0.96	0.74	0.76	0.93	0.89	0.93	0.70	0.94	0.98	0.65	8.48
C43	0.91	0.84	0.78	0.97	0.74	0.93	0.68	0.95	0.98	0.68	8.47
C44	0.92	0.80	0.80	0.90	0.73	0.95	0.70	0.96	1.00	0.70	8.47
C45	0.98	0.64	0.82	0.94	0.76	0.94	0.70	0.94	1.00	0.74	8.46
C46	0.92	0.76	0.84	0.71	0.89	0.95	0.70	0.96	1.00	0.73	8.46
C47	1.00	0.86	0.66	0.86	0.82	0.92	0.70	0.93	1.00	0.69	8.44
C48	0.94	0.94	0.88	1.00	0.38	0.93	0.70	0.95	0.99	0.73	8.44
C49	0.98	0.90	0.64	0.82	0.74	0.94	0.70	0.94	1.00	0.73	8.39
C50	0.86	0.86	0.86	0.97	0.59	0.94	0.70	0.98	0.96	0.65	8.38
C51	1.00	0.86	0.86	0.77	0.55	0.94	0.70	0.96	1.00	0.73	8.38
C52	1.00	0.90	0.76	0.69	0.76	0.92	0.70	0.88	1.00	0.75	8.36
C53	1.00	1.00	0.76	0.50	0.68	0.94	0.70	1.00	1.00	0.75	8.33
C54	1.00	0.74	0.82	0.75	0.66	0.94	0.70	0.93	1.00	0.76	8.31
C55	0.93	0.90	0.78	0.76	0.86	0.91	0.70	0.94	0.92	0.60	8.30
C56	1.00	0.66	0.88	0.80	0.67	0.93	0.70	0.95	1.00	0.71	8.30
C57	0.97	0.70	0.92	0.78	0.51	0.95	0.70	0.98	1.00	0.74	8.26
C58	0.95	0.74	0.84	0.90	0.93	0.86	0.71	0.78	0.88	0.66	8.25
C59	0.93	0.76	0.78	0.79	0.65	0.94	0.70	0.97	1.00	0.74	8.25
C60	0.89	0.92	0.84	0.99	0.73	0.87	0.70	0.94	0.78	0.58	8.24
C61	0.93	0.86	0.76	0.86	0.93	0.87	0.70	0.87	0.83	0.61	8.22

https://doi.org/10.3390/land15060976

Table A1. *Cont.*

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C62	0.90	0.96	0.82	0.83	0.63	0.91	0.70	0.98	0.91	0.58	8.21
C63	0.94	0.70	0.84	0.68	0.65	0.96	0.70	0.98	1.00	0.75	8.20
C64	0.93	0.90	0.78	0.97	0.85	0.85	0.70	0.88	0.77	0.57	8.20
C65	0.96	0.64	0.72	0.81	0.78	0.93	0.70	0.97	0.99	0.69	8.18
C66	0.95	0.76	0.74	0.93	0.96	0.87	0.70	0.80	0.84	0.62	8.17
C67	0.89	0.86	0.72	0.83	0.62	0.93	0.70	1.00	0.93	0.62	8.11
C68	0.96	0.90	0.66	0.67	0.71	0.93	0.70	0.92	0.97	0.68	8.10
C69	0.92	0.96	0.76	0.67	0.55	0.93	0.70	0.97	0.97	0.64	8.07
C70	0.93	0.86	0.92	0.99	0.00	0.95	0.71	0.97	1.00	0.73	8.05
C71	0.90	0.72	0.94	0.70	0.46	0.95	0.70	0.91	1.00	0.71	7.99
C72	1.00	0.90	0.74	0.63	0.37	0.93	0.70	0.95	1.00	0.74	7.96
C73	0.95	0.88	0.78	0.89	0.91	0.83	0.70	0.80	0.72	0.47	7.92
C74	0.94	0.98	0.74	0.65	0.32	0.93	0.70	0.98	0.98	0.68	7.90
C75	0.88	0.88	0.78	0.90	0.99	0.83	0.70	0.70	0.72	0.49	7.89
C76	1.00	0.86	0.72	0.74	0.72	0.85	0.70	0.83	0.84	0.60	7.86
C77	0.89	0.84	0.72	0.66	0.65	0.91	0.70	0.98	0.91	0.58	7.85
C78	0.87	0.90	0.88	0.83	0.53	0.86	0.70	0.90	0.76	0.57	7.81
C79	0.92	0.96	0.82	0.85	0.51	0.85	0.70	0.88	0.76	0.56	7.80
C80	1.00	0.84	0.78	0.72	0.24	0.93	0.70	0.89	1.00	0.71	7.80
C81	0.97	0.84	0.72	0.76	0.59	0.85	0.70	0.92	0.82	0.58	7.76
C82	0.88	0.94	0.72	0.80	0.88	0.84	0.70	0.74	0.72	0.49	7.70
C83	0.96	0.66	0.78	0.81	0.11	0.94	0.70	0.98	1.00	0.74	7.67
C84	0.94	0.96	0.72	0.56	0.16	0.92	0.70	0.98	0.97	0.71	7.62
C85	1.00	0.98	0.78	0.49	0.00	0.93	0.70	0.95	1.00	0.68	7.50
C86	1.00	0.68	0.76	0.81	0.04	0.93	0.70	0.83	1.00	0.74	7.50
C87	0.86	0.82	0.78	0.91	0.91	0.83	0.70	0.62	0.68	0.35	7.47
C88	1.00	0.92	0.68	0.60	0.71	0.82	0.70	0.70	0.74	0.51	7.38
C89	0.89	0.88	0.78	0.81	0.71	0.82	0.70	0.70	0.65	0.38	7.31
C90	0.90	0.88	0.80	0.95	0.92	0.80	0.41	0.59	0.58	0.41	7.23
C91	0.99	0.92	0.82	0.26	0.00	0.92	0.70	0.92	0.97	0.70	7.19
C92	0.95	0.82	0.76	0.92	0.80	0.79	0.41	0.69	0.55	0.30	7.01
C93	0.92	0.90	0.74	0.59	0.45	0.82	0.70	0.81	0.67	0.41	7.01
C94	0.90	0.96	0.74	0.63	0.84	0.80	0.41	0.69	0.56	0.44	6.97
C95	0.89	0.98	0.78	0.66	0.73	0.79	0.41	0.77	0.55	0.33	6.90
C96	0.91	0.86	0.74	0.50	0.00	0.83	0.70	0.74	0.71	0.52	6.50
C97	0.94	0.92	0.68	0.23	0.00	0.84	0.70	0.86	0.75	0.56	6.49
C98	0.78	0.92	0.86	0.88	0.67	0.78	0.00	0.35	0.41	0.21	4.12
C99	0.82	1.00	0.74	0.58	0.88	0.76	0.00	0.24	0.35	0.14	4.02
C100	0.85	0.98	0.80	0.17	0.00	0.76	0.00	0.28	0.37	0.15	2.80

Note: The results are categorized into three groups: Group A (pix2pixHD generation), Group B (LoRA generation), and Group C (pix2pixHD-LoRA model generation). A_{FCR}: Fabric Coverage Rate; A_{NSU}: No. of Small Fabric Units; A_{NLU}: No. of Large Fabric Units; A_{AFS}: Average Fabric Size; A_{SDF}: Standard Deviation of Fabric Size; A_{SND}: Street Network Density; A_{MCP}: Minimum Cost Path; A_{STA}: Spatio-Temporal Accessibility; A_{ASW}: Average Street Width; A_{SDW}: Standard Deviation of Street Width.

Appendix B

This appendix documents the complete quantitative evaluation scores and the identified Multi-objective Optimization results.

https://doi.org/10.3390/land15060976



Figure A5. A subset of generation Results of Pix2PixHD-LoRA-2. Overview of the generation process and color-coding scheme for the paired dataset. The multi-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for back-ground, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

https://doi.org/10.3390/land15060976

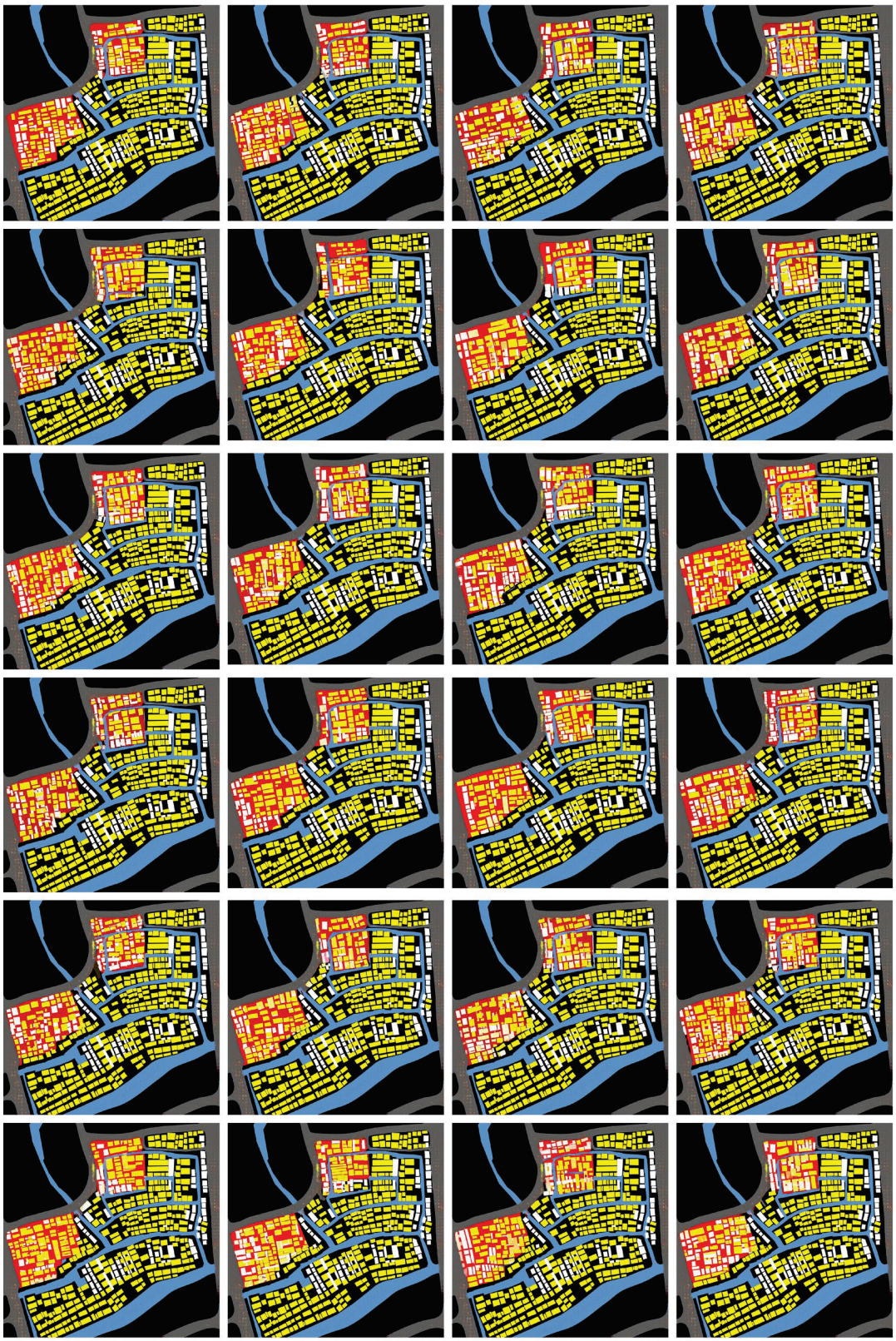


Figure A6. A subset of generation Results of Pix2PixHD-LoRA-3. Overview of the generation process and color-coding scheme for the paired dataset. The multi-channel site constraints are color-coded based on specific RGB values: black (0, 0, 0) for background, red (255, 0, 0) for land-use boundaries, blue (100, 150, 200) for waterways, grey (100, 100, 100) for street networks, yellow (255, 255, 0) for buildings with ridge orientations parallel to the watercourse, and white (255, 255, 255) for buildings with ridge orientations perpendicular to the watercourse.

https://doi.org/10.3390/land15060976

Table A2. Generation Results of Multi-objective Optimization.

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C101	1.00	0.86	0.84	0.98	0.93	0.95	0.70	1.00	1.00	0.76	9.01
C102	1.00	0.96	0.86	0.93	0.87	0.94	0.70	0.99	1.00	0.75	9.00
C103	1.00	0.90	0.84	0.99	0.91	0.95	0.70	0.98	1.00	0.74	8.98
C104	1.00	0.90	0.82	0.96	0.98	0.95	0.69	0.92	1.00	0.75	8.97
C105	1.00	0.90	0.88	0.92	0.94	0.94	0.70	0.98	1.00	0.71	8.97
C106	1.00	0.94	0.78	0.94	0.93	0.95	0.70	0.98	1.00	0.75	8.96
C107	1.00	0.80	0.82	0.99	0.96	0.94	0.69	0.98	1.00	0.76	8.95
C108	1.00	0.98	0.76	0.91	0.91	0.95	0.70	0.98	1.00	0.74	8.94
C109	1.00	0.92	0.82	0.99	0.83	0.94	0.70	0.95	1.00	0.75	8.89
C110	1.00	0.92	0.76	0.95	0.90	0.95	0.69	0.98	1.00	0.74	8.89
C111	0.96	0.90	0.86	0.97	0.84	0.97	0.70	0.97	1.00	0.72	8.88
C112	1.00	0.92	0.82	0.99	0.81	0.94	0.70	0.95	1.00	0.74	8.86
C113	1.00	0.96	0.72	0.87	0.92	0.95	0.70	0.98	1.00	0.77	8.86
C114	1.00	0.90	0.80	0.97	0.82	0.94	0.70	0.99	1.00	0.74	8.85
C115	1.00	0.90	0.86	0.90	0.87	0.93	0.70	0.95	1.00	0.74	8.85
C116	1.00	0.88	0.82	0.90	0.91	0.94	0.70	0.96	1.00	0.75	8.85
C117	0.97	0.94	0.76	0.90	0.91	0.95	0.70	1.00	1.00	0.72	8.85
C118	1.00	0.96	0.84	0.81	0.96	0.93	0.70	0.94	1.00	0.70	8.84
C119	1.00	0.94	0.76	0.82	0.95	0.94	0.70	0.97	1.00	0.76	8.84
C120	1.00	0.94	0.76	0.95	0.93	0.93	0.70	0.91	1.00	0.71	8.84
C121	1.00	0.90	0.78	0.95	0.95	0.94	0.70	0.91	1.00	0.72	8.84
C122	0.97	0.92	0.82	1.00	0.83	0.92	0.69	0.97	0.99	0.70	8.83
C123	1.00	0.88	0.76	0.93	0.88	0.94	0.70	1.00	1.00	0.74	8.83
C124	1.00	0.96	0.74	0.98	0.72	0.96	0.70	1.00	1.00	0.75	8.81
C125	1.00	0.98	0.80	0.90	0.76	0.94	0.70	0.97	1.00	0.75	8.80
C126	1.00	0.88	0.68	0.91	1.00	0.94	0.70	0.96	1.00	0.74	8.80
C127	1.00	0.86	0.80	0.98	0.80	0.95	0.69	0.95	1.00	0.77	8.79
C128	1.00	0.90	0.86	0.90	0.79	0.94	0.70	0.94	1.00	0.75	8.78
C129	1.00	0.96	0.86	0.91	0.87	0.93	0.70	0.95	0.99	0.61	8.78
C130	1.00	1.00	0.78	0.74	0.92	0.93	0.70	0.99	1.00	0.70	8.77
C131	1.00	1.00	0.78	0.93	0.74	0.94	0.70	0.93	1.00	0.73	8.76
C132	1.00	0.94	0.78	0.76	0.96	0.93	0.70	0.94	1.00	0.76	8.76
C133	1.00	0.94	0.84	0.90	0.69	0.94	0.70	0.98	1.00	0.74	8.73
C134	1.00	0.84	0.80	0.97	0.85	0.94	0.70	0.92	1.00	0.71	8.73
C135	1.00	0.92	0.76	0.76	0.94	0.94	0.69	0.97	1.00	0.75	8.73
C136	1.00	0.92	0.88	0.98	0.60	0.95	0.70	0.93	1.00	0.75	8.70
C137	1.00	0.90	0.82	0.93	0.73	0.95	0.70	0.93	1.00	0.75	8.70
C138	0.98	0.88	0.84	0.99	0.64	0.95	0.69	1.00	1.00	0.73	8.69
C139	1.00	0.78	0.76	0.94	0.87	0.94	0.70	0.95	1.00	0.75	8.68
C140	1.00	0.90	0.74	0.90	0.79	0.94	0.69	1.00	1.00	0.71	8.68
C141	1.00	0.90	0.74	0.94	0.74	0.94	0.69	0.97	1.00	0.76	8.67
C142	1.00	0.90	0.84	0.96	0.56	0.95	0.70	1.00	1.00	0.75	8.67
C143	1.00	0.92	0.80	0.95	0.66	0.95	0.69	0.96	1.00	0.74	8.67
C144	1.00	0.96	0.66	0.75	0.93	0.93	0.70	0.96	1.00	0.77	8.67
C145	0.97	0.82	0.78	0.97	0.80	0.94	0.70	0.99	1.00	0.70	8.67
C146	1.00	0.96	0.78	1.00	0.56	0.95	0.70	0.96	1.00	0.78	8.67
C147	1.00	0.94	0.76	0.82	0.80	0.93	0.70	0.96	1.00	0.74	8.65
C148	1.00	0.92	0.76	0.91	0.73	0.93	0.70	0.94	1.00	0.76	8.64
C149	1.00	0.88	0.90	0.83	0.68	0.95	0.70	0.96	1.00	0.74	8.64
C150	1.00	0.84	0.78	0.95	0.80	0.94	0.70	0.92	1.00	0.71	8.63

https://doi.org/10.3390/land15060976

Table A2. *Cont.*

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C151	0.98	0.94	0.80	0.90	0.63	0.95	0.70	1.00	1.00	0.74	8.63
C152	1.00	0.88	0.80	0.94	0.62	0.94	0.70	0.99	1.00	0.76	8.63
C153	0.96	0.90	0.74	0.97	0.68	0.95	0.70	1.00	1.00	0.73	8.62
C154	1.00	0.88	0.76	0.99	0.65	0.94	0.70	0.95	1.00	0.74	8.62
C155	1.00	0.98	0.70	0.88	0.74	0.94	0.70	0.93	1.00	0.75	8.62
C156	1.00	0.90	0.90	0.94	0.50	0.96	0.70	0.96	1.00	0.75	8.61
C157	1.00	0.92	0.80	0.90	0.77	0.92	0.70	0.92	1.00	0.69	8.61
C158	1.00	0.86	0.88	0.87	0.61	0.94	0.70	0.98	1.00	0.76	8.60
C159	1.00	0.86	0.90	0.88	0.64	0.95	0.70	0.96	1.00	0.73	8.60
C160	1.00	0.90	0.74	0.71	0.89	0.94	0.70	0.97	1.00	0.75	8.60
C161	1.00	0.98	0.84	0.92	0.55	0.94	0.70	0.95	1.00	0.72	8.60
C162	1.00	0.92	0.74	0.91	0.66	0.94	0.70	0.98	1.00	0.75	8.59
C163	0.96	0.96	0.84	0.97	0.59	0.94	0.70	0.97	1.00	0.65	8.58
C164	1.00	0.90	0.82	0.86	0.60	0.95	0.70	1.00	1.00	0.75	8.57
C165	0.96	1.00	0.92	0.95	0.43	0.94	0.70	0.97	1.00	0.69	8.56
C166	1.00	0.88	0.68	0.82	0.83	0.93	0.70	0.98	1.00	0.74	8.56
C167	0.95	0.84	0.84	0.93	0.61	0.95	0.70	1.00	1.00	0.74	8.55
C168	0.94	0.84	0.90	0.79	0.73	0.94	0.69	1.00	1.00	0.71	8.55
C169	1.00	0.82	0.88	0.88	0.62	0.94	0.70	0.93	1.00	0.75	8.53
C170	1.00	0.96	0.66	0.71	0.86	0.93	0.70	0.97	1.00	0.72	8.51
C171	1.00	0.96	0.70	0.64	0.88	0.95	0.70	0.96	1.00	0.72	8.51
C172	1.00	0.98	0.84	0.89	0.45	0.96	0.70	0.89	1.00	0.77	8.49
C173	1.00	0.88	0.84	0.82	0.62	0.95	0.70	0.92	1.00	0.75	8.48
C174	0.95	0.84	0.84	0.84	0.72	0.95	0.70	0.96	1.00	0.69	8.47
C175	1.00	0.90	0.70	0.84	0.72	0.94	0.70	0.95	1.00	0.72	8.46
C176	1.00	0.68	0.74	0.89	0.75	0.96	0.69	0.99	1.00	0.75	8.46
C177	1.00	0.84	0.72	0.88	0.61	0.94	0.70	0.98	1.00	0.77	8.44
C178	0.97	0.84	0.82	0.85	0.53	0.96	0.69	1.00	1.00	0.75	8.42
C179	1.00	0.82	0.82	0.87	0.52	0.94	0.69	0.98	1.00	0.76	8.41
C180	0.93	0.88	0.88	0.92	0.46	0.95	0.70	0.98	1.00	0.70	8.41
C181	1.00	0.96	0.64	0.64	0.77	0.94	0.70	0.99	1.00	0.75	8.40
C182	1.00	0.94	0.74	0.71	0.75	0.93	0.70	0.90	1.00	0.74	8.40
C183	1.00	0.82	0.82	0.83	0.57	0.94	0.70	0.92	1.00	0.75	8.36
C184	1.00	0.80	0.84	0.91	0.52	0.93	0.70	0.86	1.00	0.77	8.33
C185	0.93	0.86	0.86	0.97	0.40	0.94	0.70	1.00	0.99	0.68	8.31
C186	0.95	0.84	0.88	0.81	0.49	0.95	0.70	1.00	1.00	0.69	8.30
C187	1.00	0.98	0.76	0.51	0.61	0.94	0.70	1.00	1.00	0.76	8.26
C188	0.96	0.80	0.88	0.78	0.41	0.97	0.70	1.00	1.00	0.74	8.24
C189	0.97	0.80	0.86	0.79	0.48	0.96	0.69	0.93	1.00	0.75	8.23
C190	1.00	0.66	0.86	0.73	0.65	0.95	0.70	0.90	1.00	0.76	8.22
C191	1.00	0.86	0.84	0.74	0.46	0.95	0.70	0.89	1.00	0.75	8.20
C192	1.00	0.74	0.82	0.76	0.51	0.95	0.70	0.98	1.00	0.74	8.20
C193	0.98	0.82	0.88	0.71	0.41	0.96	0.70	0.98	1.00	0.75	8.19
C194	1.00	0.82	0.80	0.83	0.32	0.94	0.70	0.99	1.00	0.76	8.16
C195	1.00	0.86	0.80	0.87	0.00	0.93	0.69	0.96	1.00	0.75	7.87
C196	1.00	0.98	0.72	0.70	0.08	0.94	0.70	0.94	1.00	0.75	7.80
C197	0.99	0.94	0.80	0.69	0.00	0.94	0.69	0.96	1.00	0.73	7.73
C198	1.00	0.80	0.74	0.65	0.08	0.96	0.70	0.98	1.00	0.77	7.69
C199	1.00	0.94	0.78	0.46	0.00	0.93	0.70	0.95	1.00	0.74	7.49
C200	1.00	0.96	0.74	0.31	0.00	0.93	0.70	0.94	1.00	0.72	7.30
C201	1.00	0.92	0.86	0.99	0.99	0.95	0.70	1.00	1.00	0.74	8.94
C202	1.00	0.98	0.88	0.93	0.95	0.95	0.70	1.00	1.00	0.75	8.94

https://doi.org/10.3390/land15060976

Table A2. *Cont.*

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C203	1.00	0.96	0.84	0.97	0.88	0.97	0.70	0.97	1.00	0.79	8.90
C204	1.00	0.98	0.86	0.96	0.79	0.96	0.70	1.00	1.00	0.76	8.82
C205	0.98	0.90	0.82	0.96	0.94	0.95	0.70	0.98	1.00	0.77	8.82
C206	1.00	0.96	0.82	0.99	0.85	0.95	0.70	0.98	1.00	0.74	8.82
C207	1.00	0.86	0.78	0.95	1.00	0.94	0.71	0.98	1.00	0.76	8.81
C208	1.00	0.90	0.82	0.98	0.80	0.95	0.70	1.00	1.00	0.78	8.78
C209	1.00	0.96	0.86	0.87	0.94	0.93	0.70	0.95	1.00	0.75	8.78
C210	0.99	0.78	0.82	0.94	0.99	0.95	0.70	0.99	1.00	0.78	8.71
C211	1.00	0.90	0.84	0.97	0.83	0.95	0.70	0.98	1.00	0.76	8.70
C212	1.00	0.86	0.76	0.97	0.93	0.95	0.70	0.99	1.00	0.77	8.70
C213	1.00	0.96	0.78	0.94	0.84	0.94	0.70	0.99	1.00	0.77	8.68
C214	1.00	0.98	0.80	0.89	0.92	0.95	0.71	0.93	1.00	0.76	8.67
C215	1.00	0.96	0.76	0.90	0.93	0.95	0.70	0.96	1.00	0.76	8.67
C216	1.00	0.96	0.78	0.81	0.98	0.94	0.70	1.00	1.00	0.75	8.65
C217	0.98	0.86	0.78	0.93	0.99	0.95	0.70	1.00	1.00	0.70	8.65
C218	1.00	0.90	0.78	0.81	0.96	0.96	0.70	0.99	1.00	0.79	8.65
C219	1.00	1.00	0.82	1.00	0.69	0.94	0.71	1.00	1.00	0.74	8.65
C220	1.00	0.88	0.74	0.94	0.96	0.94	0.70	0.97	1.00	0.75	8.63
C221	1.00	0.98	0.78	0.91	0.82	0.94	0.70	1.00	1.00	0.75	8.63
C222	1.00	0.92	0.74	0.98	0.86	0.95	0.70	0.96	1.00	0.75	8.63
C223	1.00	0.94	0.82	0.94	0.74	0.95	0.70	0.98	1.00	0.78	8.61
C224	1.00	0.94	0.88	1.00	0.67	0.94	0.71	0.99	1.00	0.73	8.60
C225	1.00	0.90	0.70	0.96	0.87	0.95	0.71	0.99	1.00	0.76	8.60
C226	1.00	0.86	0.80	0.98	0.84	0.94	0.70	1.00	1.00	0.72	8.58
C227	0.98	0.86	0.86	0.90	0.90	0.94	0.70	1.00	1.00	0.68	8.57
C228	1.00	0.80	0.74	0.96	0.93	0.94	0.69	0.98	1.00	0.75	8.57
C229	1.00	0.94	0.80	0.81	0.85	0.94	0.70	0.98	1.00	0.75	8.55
C230	1.00	0.92	0.68	0.94	0.83	0.94	0.70	1.00	1.00	0.76	8.55
C231	1.00	0.82	0.84	0.96	0.71	0.96	0.70	0.99	1.00	0.79	8.55
C232	1.00	0.78	0.82	0.96	0.79	0.95	0.70	0.99	1.00	0.76	8.55
C233	1.00	0.86	0.86	0.96	0.75	0.94	0.70	0.90	1.00	0.77	8.54
C234	1.00	0.92	0.84	0.95	0.67	0.95	0.71	0.98	1.00	0.75	8.53
C235	1.00	0.86	0.88	0.90	0.71	0.94	0.70	0.98	1.00	0.76	8.53
C236	1.00	0.90	0.76	1.00	0.68	0.95	0.70	1.00	1.00	0.75	8.52
C237	0.95	0.90	0.88	0.88	0.71	0.96	0.71	1.00	1.00	0.76	8.49
C238	0.99	0.88	0.88	0.83	0.78	0.96	0.71	0.95	1.00	0.77	8.49
C239	0.95	0.96	0.90	1.00	0.57	0.95	0.70	1.00	1.00	0.70	8.49
C240	1.00	0.94	0.76	0.98	0.66	0.95	0.71	1.00	1.00	0.74	8.48
C241	1.00	0.72	0.82	0.86	0.92	0.95	0.71	0.98	1.00	0.77	8.48
C242	1.00	0.88	0.78	0.97	0.64	0.96	0.70	1.00	1.00	0.78	8.48
C243	1.00	0.80	0.72	1.00	0.84	0.93	0.70	0.97	1.00	0.74	8.47
C244	1.00	0.78	0.82	0.85	0.85	0.95	0.71	1.00	1.00	0.75	8.47
C245	1.00	0.86	0.78	0.96	0.66	0.95	0.71	1.00	1.00	0.78	8.46
C246	0.92	0.88	0.88	0.98	0.62	0.96	0.71	1.00	1.00	0.75	8.46
C247	0.99	0.90	0.86	0.99	0.59	0.96	0.70	0.96	1.00	0.74	8.44
C248	1.00	0.92	0.78	0.92	0.65	0.94	0.71	1.00	1.00	0.77	8.44
C249	1.00	0.84	0.78	0.95	0.72	0.95	0.70	0.99	1.00	0.75	8.39
C250	1.00	0.82	0.84	0.89	0.68	0.96	0.71	1.00	1.00	0.75	8.38
C251	1.00	0.94	0.84	0.96	0.53	0.94	0.70	0.98	1.00	0.75	8.38
C252	1.00	0.86	0.82	0.90	0.63	0.95	0.71	1.00	1.00	0.76	8.36
C253	1.00	0.92	0.90	0.94	0.45	0.95	0.70	0.96	1.00	0.78	8.33
C254	1.00	0.74	0.84	0.85	0.74	0.96	0.69	1.00	1.00	0.76	8.31

https://doi.org/10.3390/land15060976

Table A2. *Cont.*

Case	A_{FCR}	A_{NSU}	A_{NLU}	A_{AFS}	A_{SDF}	A_{SND}	A_{MCP}	A_{STA}	A_{ASW}	A_{SDW}	Score
C255	0.99	0.90	0.84	0.93	0.55	0.95	0.70	0.98	1.00	0.75	8.30
C256	0.99	0.74	0.86	0.75	0.85	0.96	0.70	0.97	1.00	0.77	8.30
C257	1.00	0.84	0.84	0.88	0.66	0.94	0.70	0.96	1.00	0.75	8.26
C258	1.00	0.74	0.74	0.87	0.81	0.95	0.70	0.99	1.00	0.76	8.25
C259	1.00	0.98	0.86	0.87	0.41	0.96	0.70	1.00	1.00	0.76	8.25
C260	0.98	0.92	0.90	0.96	0.41	0.95	0.70	1.00	1.00	0.69	8.24
C261	1.00	0.88	0.82	0.78	0.65	0.95	0.70	0.97	1.00	0.76	8.22
C262	1.00	0.84	0.88	0.80	0.57	0.96	0.71	0.96	1.00	0.78	8.21
C263	1.00	0.96	0.78	0.87	0.45	0.96	0.70	1.00	1.00	0.77	8.20
C264	1.00	0.90	0.80	0.85	0.54	0.95	0.71	1.00	1.00	0.77	8.20
C265	1.00	1.00	0.72	0.61	0.75	0.94	0.70	0.99	1.00	0.77	8.18
C266	1.00	0.82	0.88	0.78	0.56	0.95	0.70	0.99	1.00	0.78	8.17
C267	0.92	0.88	0.80	0.89	0.60	0.96	0.70	1.00	1.00	0.70	8.11
C268	0.98	0.94	0.88	0.90	0.51	0.94	0.70	0.85	1.00	0.74	8.10
C269	1.00	0.64	0.80	0.89	0.72	0.95	0.70	1.00	1.00	0.76	8.07
C270	0.98	0.78	0.88	0.74	0.65	0.96	0.71	1.00	1.00	0.75	8.05
C271	1.00	0.74	0.78	0.88	0.66	0.95	0.70	0.97	1.00	0.74	7.99
C272	1.00	0.98	0.86	0.72	0.49	0.95	0.70	0.96	1.00	0.76	7.96
C273	0.99	0.78	0.88	0.85	0.50	0.97	0.70	1.00	1.00	0.76	7.92
C274	0.94	0.82	0.90	0.74	0.60	0.95	0.71	1.00	1.00	0.77	7.90
C275	0.98	0.92	0.84	0.83	0.47	0.96	0.70	0.96	1.00	0.75	7.89
C276	1.00	0.78	0.76	0.89	0.63	0.95	0.70	0.89	1.00	0.76	7.86
C277	0.94	0.80	0.84	0.77	0.62	0.95	0.70	1.00	1.00	0.74	7.85
C278	1.00	0.82	0.76	0.92	0.51	0.93	0.70	0.97	1.00	0.71	7.81
C279	0.98	0.90	0.86	0.78	0.36	0.96	0.70	1.00	1.00	0.76	7.80
C280	0.93	0.88	0.92	0.84	0.39	0.95	0.71	0.96	1.00	0.73	7.80
C281	0.93	0.94	0.90	0.65	0.46	0.96	0.70	1.00	1.00	0.75	7.76
C282	0.92	0.86	0.90	0.78	0.57	0.95	0.70	0.95	1.00	0.66	7.70
C283	0.96	0.76	0.88	0.74	0.48	0.97	0.70	1.00	1.00	0.77	7.67
C284	1.00	0.90	0.76	0.76	0.43	0.94	0.70	0.99	1.00	0.77	7.62
C285	1.00	0.76	0.78	0.83	0.44	0.95	0.71	0.97	1.00	0.78	7.50
C286	1.00	0.88	0.92	0.77	0.31	0.96	0.70	0.95	1.00	0.73	7.50
C287	0.99	0.72	0.90	0.67	0.47	0.98	0.71	1.00	1.00	0.76	7.47
C288	1.00	1.00	0.72	0.51	0.61	0.93	0.70	0.99	1.00	0.73	7.38
C289	0.99	0.78	0.90	0.73	0.41	0.95	0.70	0.96	1.00	0.74	7.31
C290	1.00	0.92	0.80	0.74	0.27	0.94	0.70	0.98	1.00	0.78	7.23
C291	0.98	0.68	0.86	0.73	0.52	0.95	0.70	0.94	1.00	0.74	7.19
C292	1.00	0.92	0.78	0.71	0.34	0.93	0.70	0.98	1.00	0.74	7.01
C293	0.95	0.72	0.94	0.74	0.44	0.94	0.70	0.92	1.00	0.71	7.01
C294	1.00	0.92	0.76	0.71	0.27	0.94	0.69	0.96	1.00	0.77	6.97
C295	1.00	0.84	0.68	0.75	0.36	0.93	0.70	0.92	1.00	0.75	6.90
C296	1.00	0.90	0.74	0.72	0.17	0.94	0.70	0.95	1.00	0.75	6.50
C297	1.00	0.96	0.78	0.75	0.00	0.93	0.71	0.96	1.00	0.76	6.49
C298	1.00	0.94	0.68	0.59	0.17	0.93	0.70	0.95	1.00	0.76	4.12
C299	1.00	0.98	0.76	0.64	0.04	0.92	0.70	0.92	1.00	0.73	4.02
C300	1.00	0.84	0.72	0.76	0.00	0.93	0.71	0.94	1.00	0.76	2.80

Note: C1–100 denote the first-generation data of multi-objective optimization; C101–200 represent the second-generation data; and C201–300 represent the third-generation data of multi-objective optimization. A_{FCR}: Fabric Coverage Rate; A_{NSU}: No. of Small Fabric Units; A_{NLU}: No. of Large Fabric Units; A_{AFS}: Average Fabric Size; A_{SDF}: Standard Deviation of Fabric Size; A_{SND}: Street Network Density; A_{MCP}: Minimum Cost Path; A_{STA}: Spatio-Temporal Accessibility; A_{ASW}: Average Street Width; A_{SDW}: Standard Deviation of Street Width.

https://doi.org/10.3390/land15060976

References

1. Yuan, X. Protection and Development of Historical and Cultural Blocks. *Sustain. Dev.* **2024**, *14*, 185–192. [CrossRef]
2. Liu, P.; Neppl, M. A Research on the Conservation of Plot Pattern of Chinese Historic Cities: Connotations, Transformations and Strategies. *Urban Plan. Forum* **2020**, *5*, 92–99. [CrossRef]
3. Zhang, C.; Guo, W.; Xie, J.; Wang, W. Reconstruction of City Historical Block Spatial Texture Based on Parametric Technology: A Case Study on Nanjing Lotus Pond. *Art Des.* **2019**, *3*, 80–83. (In Chinese) [CrossRef]
4. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680. Available online: https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf (accessed on 1 April 2026).
5. Wang, T.-C.; Liu, M.-Y.; Zhu, J.-Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. [CrossRef]
6. Shim, J.; Moon, J.; Kim, H.; Hwang, E. FloorDiffusion: Diffusion Model-Based Conditional Floorplan Image Generation Method Using Parameter-Efficient Fine-Tuning and Image Inpainting. *J. Build. Eng.* **2024**, *95*, 110320. [CrossRef]
7. Hu, X.; Zheng, H.; Lai, D. Prediction and Optimization of Daylight Performance of AI-Generated Residential Floor Plans. *Build. Environ.* **2025**, *279*, 113054. [CrossRef]
8. ICOMOS. The Nara Document on Authenticity. Available online: https://www.icomos.org/charters/nara-e.pdf (accessed on 29 May 2026).
9. UNESCO. Recommendation on the Historic Urban Landscape. Available online: https://whc.unesco.org/en/hul/ (accessed on 29 May 2026).
10. Kou, H.; Zhou, J.; Chen, J.; Zhang, S. Conservation for Sustainable Development: The Sustainability Evaluation of the Xijie Historic District, Dujiangyan City, China. *Sustainability* **2018**, *10*, 4645. [CrossRef]
11. Liu, X.; Wu, L.; Ye, M.; Liu, Q. Let Us Build Bridges: Understanding and Extending Diffusion Generative Models. *arXiv* **2022**. [CrossRef]
12. Balloni, E.; Paolanti, M.; Uggeri, J.; Zingaretti, P.; Pierdicca, R. Enhancing Cultural Heritage with Generative AI: A Comparative Framework for the Evaluation of 3D Model Accuracy and Visual Fidelity. *Digit. Herit.* **2025**, 1–10. [CrossRef]
13. Merrell, P.; Schkufza, E.; Koltun, V. Computer-Generated Residential Building Layouts. *ACM Trans. Graph.* **2010**, *29*, 181. [CrossRef]
14. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
15. Yang, Y.; Wu, Y.; Liu, Y.; Zhang, Y.; Xu, Z. Automatic Layout of Building Functions Based on Generative Adversarial Network. In Proceedings of the China Civil Engineering Society, Beijing, China, 20–22 September 2020. (In Chinese) [CrossRef]
16. Wang, S.; Zeng, W.; Chen, X.; Ye, Y.; Qiao, Y.; Fu, C.-W. Actfloor-GAN: Activity-Guided Adversarial Networks for Human-Centric Floorplan Design. *IEEE Trans. Vis. Comput. Graph.* **2021**, *29*, 1610–1624. [CrossRef]
17. Deng, Q.; Lin, W.; Liu, Y.; Liang, L. Exploration of Generative Design of Campus General Layout Based on Generative Adversarial Network: Taking Primary School Campuses as Example. *World Archit.* **2021**, *9*, 115–119. (In Chinese) [CrossRef]
18. Liu, Y.; Stouffs, R.; Yang, Y. Urban Design Process with Conditional Generative Adversarial Networks. *Archit. J.* **2018**, *9*, 108–113. (In Chinese). Available online: https://scholarbank.nus.edu.sg/handle/10635/195769 (accessed on 29 May 2026).
19. Tian, R. Suggestive Site Planning with Conditional GAN and Urban GIS Data. In *Proceedings of the 2020 DigitalFUTURES*; Springer: Singapore, 2020; pp. 103–113. [CrossRef]
20. Liu, Y.; Hu, K.; Deng, Q. Artificial Intelligence-Assisted Case-Based Design: A Case Study on Urban Texture Darning Surrounding the Ancient City of Nantou in Shenzhen. *South Archit.* **2023**, *10*, 20–27. (In Chinese) [CrossRef]
21. Carbone, D.; Hua, M.; Coste, S.; Vanden-Eijnden, E. Efficient Training of Energy-Based Models Using Jarzynski Equality. *J. Stat. Mech.* **2024**, *2024*, 104019. [CrossRef]
22. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv* **2020**. [CrossRef]
23. Zhang, J.; Huang, Y.; Li, Z.; Li, Y.; Yu, Z.; Li, M. Development of a Method for Commercial Style Transfer of Historical Architectural Facades Based on Stable Diffusion Models. *J. Imaging* **2024**, *10*, 165. [CrossRef]
24. Sharma, H.; Kumar, R.; Gupta, M.; Chilluri, V.S.B. Boosting GAN Performance Through Dataset Augmentation with Denoising Diffusion Models. In Proceedings of the 2025 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 7–8 March 2025. [CrossRef]
25. Lin, B.; Jabi, W.; Corcoran, P.; Lannon, S. The Application of Deep Generative Models in Urban Form Generation Based on Topology: A Review. *Archit. Sci. Rev.* **2023**, *67*, 189–204. [CrossRef]
26. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
27. Perez, L.; Wang, J. The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. *arXiv* **2017**. [CrossRef]

28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
29. Usui, H. Optimisation of Building and Road Network Densities in Terms of Variation in Plot Sizes and Shapes. *Environ. Plan. B* **2020**, *48*, 1263–1278. [CrossRef]
30. Zhao, J.; Hao, X.; Yang, Y. Research on Urban Sustainability Indicators Based on Urban Grain: A Case Study in Jinan, China. *Sustainability* **2023**, *15*, 13320. [CrossRef]
31. Usui, H.; Asami, Y. Size Distribution of Building Lots and Density of Buildings and Street Networks: Theoretical Derivation Based on Gibrat's Law and Empirical Study of Downtown Districts in Tokyo. *Int. Reg. Sci. Rev.* **2019**, *43*, 229–253. [CrossRef]
32. Rebecchi, A.; Buffoli, M.; Dettori, M.; Appolloni, L.; Azara, A.; Castiglia, P.; D'Alessandro, D.; Capolongo, S. Walkable Environments and Healthy Urban Moves: Urban Context Features Assessment Framework Experienced in Milan. *Sustainability* **2019**, *11*, 2778. [CrossRef]
33. Iamtrakul, P.; Chayphong, S.; Gao, W. Assessing Spatial Disparities and Urban Facility Accessibility in Promoting Health and Well-Being. *Transp. Res. Interdiscip. Perspect.* **2024**, *25*, 101126. [CrossRef]
34. Ying, X.; Gao, J.; Liu, Z.; Qin, X.; Chen, J.; Shen, L.; Han, X. Investigation of Pedestrian-Level Wind Environment with Skyline Quantitative Factors. *Buildings* **2022**, *12*, 792. [CrossRef]
35. Ying, X.; Wang, Y.; Li, W.; Liu, Z.; Ding, G. Group Layout Pattern and Outdoor Wind Environment of Enclosed Office Buildings in Hangzhou. *Energies* **2020**, *13*, 406. [CrossRef]
36. Ying, X.-Y.; Ding, G.; Hu, X.-J.; Zhang, Y.-Q. Developing Planning Indicators for Outdoor Wind Environments of High-Rise Residential Buildings. *J. Zhejiang Univ.-Sci. A* **2016**, *17*, 378–388. [CrossRef]
37. Tan, Y.; Ying, X.; Gao, W.; Wang, S.; Liu, Z. Applying an Extended Theory of Planned Behavior to Predict Willingness to Pay for Green and Low-Carbon Energy Transition. *J. Clean. Prod.* **2023**, *387*, 135893. [CrossRef]
38. Wu, J.; Chen, S.; Ying, X.; Shu, J. Influencing Factors on Air Conditioning Energy Consumption of Naturally Ventilated Research Buildings Based on Actual HVAC Behaviours. *Buildings* **2023**, *13*, 2710. [CrossRef]
39. Wu, J.; Qiu, R.; Ying, X.; Chen, S.; Zhao, X. Optimising Behavioural Control Based on Actual HVAC Use in Naturally Ventilated Buildings. *Energies* **2025**, *18*, 6130. [CrossRef]
40. Weng, J.; Zhang, Y.; Chen, Z.; Ying, X.; Zhu, W.; Sun, Y. Field Measurements and Analysis of Indoor Environment, Occupant Satisfaction, and Sick Building Syndrome in University Buildings in Hot Summer and Cold Winter Regions in China. *Int. J. Environ. Res. Public Health* **2023**, *20*, 554. [CrossRef] [PubMed]
41. Weng, J.; Huang, F.; Lin, J.; Wang, Q.; Ying, X.; Sun, Y.; Tan, Y. Sick Building Syndrome: Prevalence and Risk Factors Among Medical Staff in Chinese Hospitals. *Buildings* **2025**, *15*, 1397. [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.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Land Editorial Office
E-mail: land@mdpi.com
www.mdpi.com/journal/land



Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest Editor. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editor and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.





Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-9018-7