sustainability

MDPI

Article

Unraveling Nonlinear and Spatially Heterogeneous Impacts of Urban Pluvial Flooding Factors in a Hill-Basin City Using Geographically Explainable Artificial Intelligence: A Case Study of Changsha

Ziqiang He [1], Yu Chen [1,2,*], Qimeng Ning [1,3], Bo Lu [1,*], Shixiong Xie [4] and Shijie Tang [5]

[1] School of Architecture and Urban Planning, Hunan City University, Yiyang 413000, China; max5516640@gmail.com (Z.H.); ningqimeng@hncu.edu.cn (Q.N.)
[2] Key Laboratory of Key Technologies of Digital Urban-Rural Spatial Planning of Hunan Province, Yiyang 413000, China
[3] Key Laboratory of Urban Planning Information Technology of Hunan Provincial Universities, Yiyang 413000, China
[4] School of Civil and Environmental Engineering, Hunan University of Technology, Zhuzhou 412007, China; 14325@hut.edu.cn
[5] Shanghai Academy of Fine Arts, Shanghai University, Shanghai 200444, China; tangsj@shu.edu.cn
* Correspondence: cy2016@hnu.edu.cn (Y.C.); lubo@hncu.edu.cn (B.L.); Tel.: +86-18874869880 (Y.C.); +86-13357312070 (B.L.)

Abstract

The factors influencing urban pluvial flooding in cities with complex topography, such as hill–basin systems, are highly nonlinear and spatially heterogeneous due to the interplay between rugged terrain and intensive human activities. However, previous research has predominantly focused on plain, mountainous, and coastal cities. As a result, the waterlogging mechanisms in hill–basin areas remain notably understudied. In this study, we developed a geographically explainable artificial intelligence (GeoXAI) framework integrating Geographical Machine Learning Regression (GeoMLR) and Geographical Shapley (GeoShapley) values to analyze nonlinear impacts of flooding factors in Changsha, a typical hill–basin city. The XGBoost model was employed to predict flooding risk (validation AUC = 0.8597, R^2 = 0.8973), while the GeoMLR model verified stable nonlinear driving relationships between factors and flooding susceptibility (test set R^2 = 0.7546)—both supporting the proposal of targeted zonal regulation strategies. Results indicated that impervious surface density (ISD), normalized difference vegetation index (NDVI), and slope are the dominant drivers of flooding, with each exhibiting distinct nonlinear threshold effects (ISD > 0.35, NDVI < 0.70, Slope < 5°) that differ significantly from those identified in plain, mountainous, or coastal regions. Spatial analysis further revealed that topography regulates flooding by controlling convergence pathways and flow velocity, while vegetation mitigates flooding through enhanced interception and infiltration, showing complementary effects across zones. Based on these findings, we proposed tailored zonal management strategies. This study not only advances the mechanistic understanding of urban waterlogging in hill–basin regions but also provides a transferable GeoXAI framework offering a robust methodological foundation for flood resilience planning in topographically complex cities.

Keywords: hill-basin; urban pluvial flooding; GeoXAI; nonlinear threshold; spatial heterogeneity; zonal management

check for updates

Academic Editor: Mike Spiliotis

Received: 18 September 2025
Revised: 27 October 2025
Accepted: 3 November 2025
Published: 5 November 2025

Citation: He, Z.; Chen, Y.; Ning, Q.; Lu, B.; Xie, S.; Tang, S. Unraveling Nonlinear and Spatially Heterogeneous Impacts of Urban Pluvial Flooding Factors in a Hill-Basin City Using Geographically Explainable Artificial Intelligence: A Case Study of Changsha. *Sustainability* **2025**, *17*, 9866. https://doi.org/10.3390/su17219866

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

1. Introduction

Urban pluvial flooding (UPF) represents one of the most severe environmental threats to modern cities, exacerbated by global climate change and rapid urbanization. Increasingly frequent extreme rainfall events have transformed urban flooding into a recurring hazard that compromises infrastructure, endangers public safety, and disrupts economic activities [1,2]. As a major obstacle to sustainable urban development, flooding results in substantial loss of life, property damage, and functional paralysis of urban systems [3,4].

It is widely accepted that flooding arises from complex interactions among extreme precipitation, intensive land development, expansion of impervious surfaces, and insufficient drainage capacity [5,6]. Although factors such as imperviousness and drainage efficiency are recognized as critical drivers [7–9], a comprehensive understanding of their relative contributions and interaction mechanisms across different topographic regions remains inadequate. Urban topography significantly influences flooding spatial patterns by altering surface flow convergence, runoff velocity, and water retention [6,10]. Yet, prior research has largely focused on plain, mountainous, or coastal settings [6,10], with complex hill–basin systems receiving comparatively little attention. Accordingly, elucidating the key drivers and underlying mechanisms in such areas is essential for formulating effective mitigation policies and enhancing urban resilience.

Studies on UPF mechanisms generally align with two perspectives: hydrological and geographical. Hydrological approaches leverage physical or empirical models, such as SWMM [11,12] and SWAT [13,14], to simulate rainfall–runoff processes with high precision. These methods, however, rely heavily on high-resolution input data, including detailed drainage networks, land use changes, and meteorological records, which constrains their application in large-scale or data-scarce regions. In contrast, geographical approaches emphasize the role of surface environmental factors, such as topography [15] and land use [10], in shaping urban pluvial flooding susceptibility (UPFS). This perspective facilitates efficient spatial evaluation of UPFS under conditions of reasonable data availability [8,9], proving highly advantageous for regional flood risk assessment and planning.

Methodologically, research on urban flooding has traditionally relied on spatial econometric models and linear regression. Yet, in practice, flooding mechanisms exhibit pronounced nonlinearity and spatial heterogeneity [9]. The same factor may exert varying magnitudes of influence, directions, and mechanistic roles across different geographical contexts [9], a complexity especially pronounced in cities with diverse topography [8,9]. Conventional linear models are inadequate for capturing such nuanced, nonlinear relationships and spatial heterogeneity [16,17]. Traditional linear models struggle to capture these complexities. In response, spatial econometric models like Geographically Weighted Regression (GWR) [18], and ensemble learning methods such as XGBoost–SHAP [7], have recently been adopted to enhance the interpretability of those linear models. While beneficial, these approaches often fail to fully represent spatial heterogeneity and offer limited insight into interaction mechanisms [9].

Despite these advances, critical research limitations still remain. First, a strong geographical bias exists: most studies concentrate on coastal, mountainous, or plain cities, neglecting complex hill–basin environments such as Changsha. The synergistic effects of unique "stepped" topography and rapid urbanization in these areas require urgent investigation. Second, current models cannot simultaneously decode nonlinear and spatially heterogeneous driver effects, resulting in an incomplete mechanistic understanding and insufficient support for region-specific governance. The emerging paradigm of Geographically Explainable Artificial Intelligence (GeoXAI) aims to overcome the challenges associated with other models [8,9]. By incorporating spatial coordinates into model inputs, combining spatial regression techniques, and utilizing Shapley value-based attribution,

GeoXAI concurrently captures nonlinear behaviors and spatial heterogeneity in flooding factors. This framework overcomes the constraints of traditional models in interpretability and spatial representation, establishing a robust foundation for refined attribution analysis and spatially targeted flood management.

This study selected Changsha, a typical hill–basin city in China, as a case study. We proposed a GeoXAI analytical framework that integrates Geographical Machine Learning Regression (GeoMLR) and Geographical Shapley (GeoShapley) values to systematically examine the influence mechanisms and spatial heterogeneous patterns of natural and built-environment factors on UPFS. This research is guided by three research questions:

(1) How can a unified analytical framework be constructed for nonlinear attribution of pluvial flooding in a hill–basin city?

(2) What are the principal driving factors of flooding in such settings, and what are their characteristic nonlinear responses?

(3) What spatial heterogeneous patterns emerge in flooding risk under multi-factor interactions?

The remainder of this paper is organized as follows: Section 2 describes the materials and methods, Section 3 presents the results, Section 4 discusses the implications and compares findings with existing literature, and Section 5 summarizes the conclusions and suggests future research directions.

2. Materials and Methods

2.1. Study Area

This study selected Changsha (113°08′ E, 28°12′ N), the capital city of Hunan Province in south-central China, as the study area (Figure 1). Covering approximately 11,819 km^2, the city is situated within the core zone of the Yangtze River Mid-Reach urban agglomeration. Changsha represents a typical hill–basin city, characterized by complex geomorphology, a subtropical monsoon climate, and rapid urbanization. The combination of its distinctive topography and accelerated urban expansion makes it an ideal representative for investigating the spatial formation mechanisms of UPF in hill–basin environments.

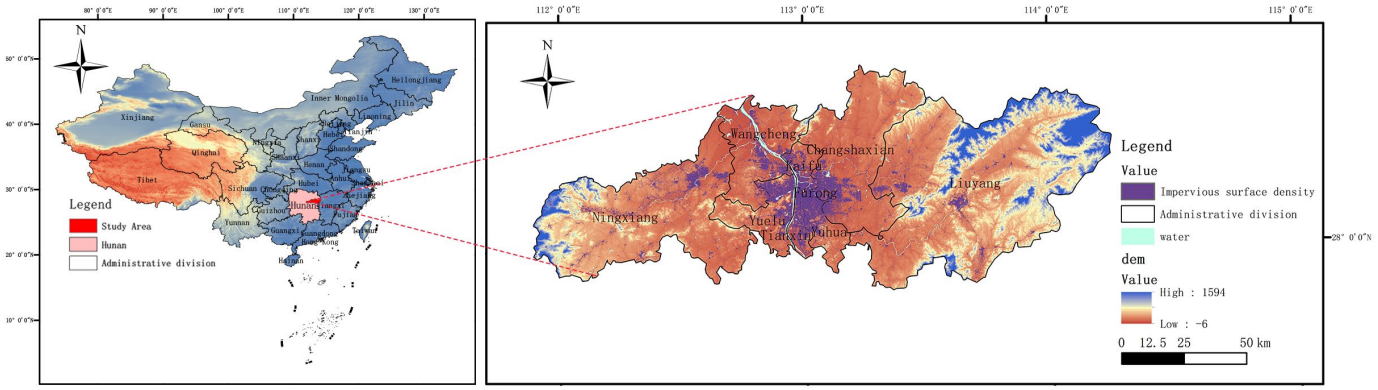

Figure 1. Study area.

Topographically, the eastern and western regions of Changsha are dominated by hilly and mountainous terrain, with the highest elevation reaching 1607.9 m, while the central area is relatively low and flat, with a minimum elevation of 23.5 m. The Xiang River flows northward through the city, creating a conspicuous "river valley effect". During heavy rainfall events, this topographical configuration promotes rapid convergence of surface runoff within the basin. When drainage capacity is insufficient, this often leads to localized, "step-like spreading" flooding—a spatial pattern notably distinct from the "contiguous waterlogging".

Urbanization in Changsha has seen a marked increase, with the urbanization rate rising from 67.7% in 2010 to 79% in 2023. This rapid urban growth has resulted in a substantial expansion of its Impervious Surface Density (ISD), particularly in eastern newly developed zones such as the Xingsha area. In contrast, the central old urban districts have long been challenged by low drainage system standards (primarily designed for a 1–3 year return period) and aging infrastructure system. This infrastructural "mismatch" between new and old urban zones further exacerbates the spatial disparity of flooding risk, providing valuable comparative opportunities for this research.

2.2. Data Sources and Processing

The data used in this study comprise four primary categories: historical flood records, built-environment characteristics, physical geographical variables, and socioeconomic indicators. A total of 238 urban flooding points were sourced from the Changsha Water Authority. These data require submitting a formal application to local water authorities (with research project approval) for access; this study obtained the data via institutional collaboration. Based on previous research [8,9,19], the following variables were selected: historical flood points (FP), impervious surface density (ISD), population density (POP), normalized difference vegetation index (NDVI), road density (RD), drainage system density (DSD), available water capacity (AWC) of soil, distribution of points of interest (POI), digital elevation model (DEM), slope, curvature, building density (BD), and distance to water (DTW). Among these variables, RD, BD, and DTW are sourced from OpenStreetMap (2022), a globally free and open-source open geospatial database (https://www.openstreetmap.org) where vector data can be downloaded by region for direct spatial analysis; NDVI, DEM, and ISD belong to remote sensing and public datasets, with DEM from Geospatial Data Cloud, NDVI from NASA EarthData, and ISD from [20]—all are international/domestic public datasets downloadable without application and in standardized TIFF format compatible with GIS 10.8 software; AWC and POP are academic public datasets, derived from [21] and [20], respectively.

To align with the 2022 baseline year of this study, the temporal attributes of the aforementioned remote sensing and academic public datasets are further clarified as follows: DEM from Geospatial Data Cloud adopts the 2022 version, NDVI from NASA EarthData uses the 2022 annual composite data, and ISD from [21] corresponds to the 2022 dataset—all three are fully consistent with the 2022 temporal baseline of RD, BD, and DTW (from OpenStreetMap), ensuring no temporal asynchrony between built-environment and physical geographical indicators. AWC from [21] is a static soil property dataset, and its variation over the years around 2022 is negligible, which allows it to stably represent the soil water storage capacity corresponding to the 2022 baseline. POP from [22] is a refined version based on China's 7th National Census in 2020; although its original survey year is 2020, Changsha's urbanization rate only increased by 0.3% between 2020 and 2022, leading to minimal changes in population density. This small variation ensures the dataset can effectively reflect the socioeconomic exposure level associated with urban pluvial flooding under the 2022 baseline.

All datasets underwent systematic preprocessing, including format standardization, coordinate system unification (WGS 1984 UTM Zone 49N), and conversion to raster format. Notably, to align with the study's 2022 baseline year, all time-variant variables (e.g., impervious surface density, road density, normalized difference vegetation index) used in the analysis are 2022 annual average data, ensuring temporal consistency with the contemporary urban conditions targeted by this research. The only exception is population density (POP), which is derived from a refined dataset based on China's 7th National Census (2020). Finally, each layer was resampled to a consistent spatial resolution of 500 m grid cells to

facilitate subsequent integrated analysis. Detailed sources and processing procedures for each variable are summarized in Table 1.

Table 1. Data sources and descriptive statistics.

Variable	Code	Description (Representing)	Unit	Mean	Std	Source
Historical Flood Points	FP	Historical flood locations	number	-	-	requires submitting a data application to local water authorities
Population density	POP	Social exposure	persons/km^2	8.81	31.4529	[22]
Impervious surface density	ISD	Urban expansion intensity	%	9.08	0.2191	[20]
Normalized Difference Vegetation Index	NDVI	Vegetation regulation capacity	Dimensionless	0.822	1261.4034	[23]
Road density	RD	Surface drainage accessibility	m/km^2	1.5	0.0033	OpenStreetMap
Drainage system density	DSD	Drainage capacity	m/km^2	0.2	0.0007	OpenStreetMap
Available Water Capacity	AWC	Surface water storage capacity	mm	131.51	29.5122	[21]
Points of Interest	POI	Land-use functional mix	count/km^2	14.77	73.0838	Amap POI
Digital Elevation Model	DEM	Basic topographic elevation data	m	166.11	159.5089	Geospatial Data Cloud
Slope	SLOPE	Surface slope gradient	°	6.88	6.1223	Calculated by author based on DEM
Curvature	CURV	Surface curvature (identifying catchment areas)	m^{-1}	−0.0006	0.0566	Calculated by author based on DEM
Building Density	BD	Impervious area from building footprint	%	0.72	0.0357	OpenStreetMap
Distance to Water	DTW	Distance to the nearest water body	m	2695.32	2764.8884	OpenStreetMap

2.3. Methods

2.3.1. Overall Research Framework

This study was structured around a three-stage analytical rationale: data fusion, followed by model construction, and concluding with interpretation analysis. As illustrated in Figure 2, the overall research framework consists of three main phases: (1) Urban Pluvial Flooding Susceptibility (UPFS) assessment; (2) Geographical Machine Learning Regression (GeoMLR); and (3) Explainable attribution analysis.

Using the preprocessed multi-source spatial data (refer to Section 2.2), an XGBoost classifier was employed to evaluate UPFS. This model categorizes each grid cell into flood-prone or non-flood-prone categories and estimates its probability of flooding occurrence. This probability served as a continuous UPFS indicator for subsequent attribution analysis.

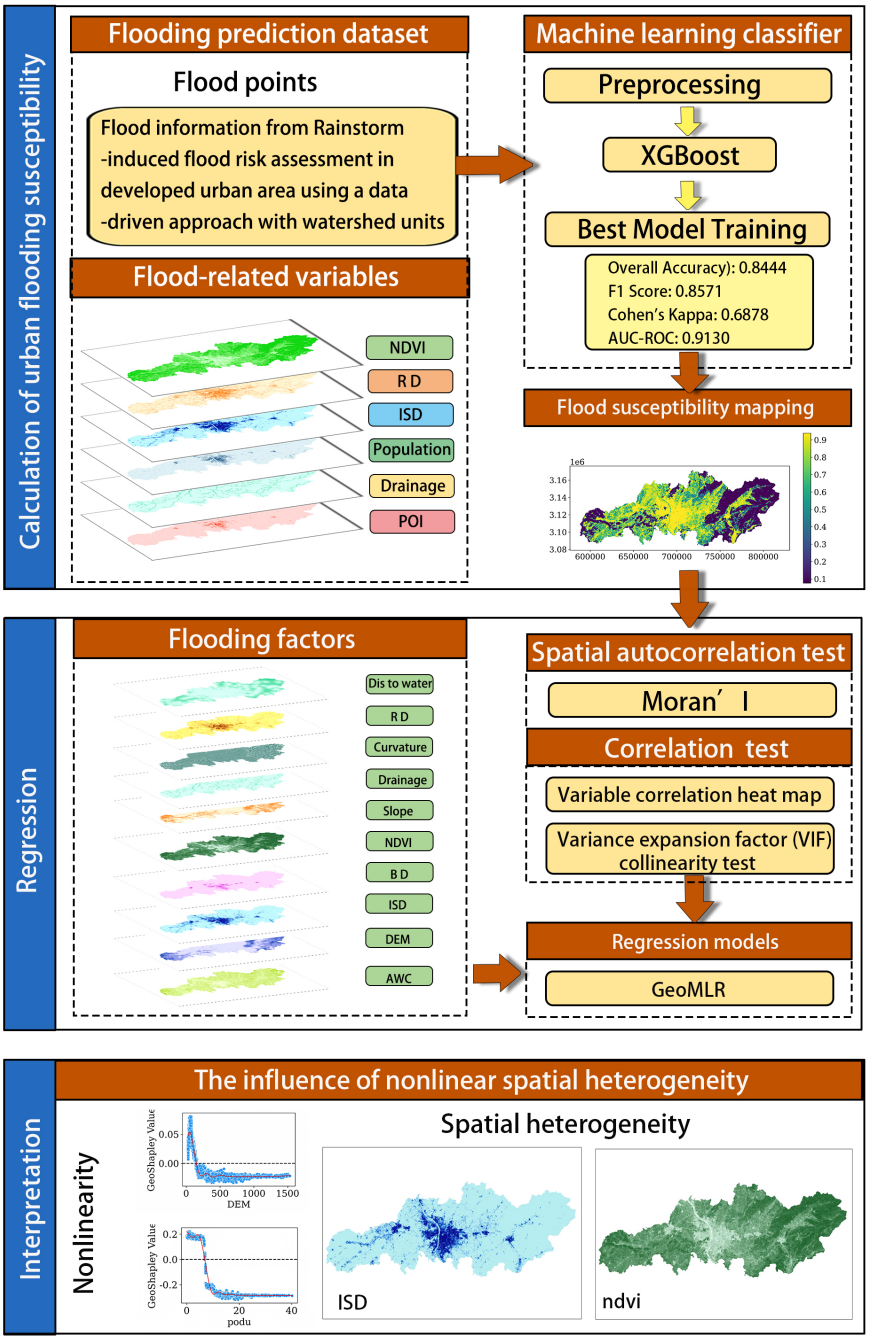

Figure 2. Research framework diagram.

To capture the spatial non-stationarity of driving factors, we introduced Geographical Machine Learning Regression (GeoMLR). This method integrates geographic coordinates (X, Y) as additional features into an XGBoost regression model alongside the original variables, explicitly incorporating spatial dependence within a nonlinear learning framework. This approach enables the modeling of spatially varying relationships between UPFS and various drivers.

Finally, the Shapley value (SHAP) method was applied to interpret the model, quantifying both individual and interactive contributions of each driving factor. The GeoShapley method was further utilized to aggregate the contributions of the geographic coordinates into a unified spatial effect term, facilitating the analysis of regional heterogeneity among factors. Through this integrated framework, the study achieves not only accurate UPFS prediction, but also a systematic interpretation of the nonlinear and spatial heterogeneous

impacts underlying UPFS mechanism, offering insights into flood causation in hill-basin complex topographic urban environments.

2.3.2. Predictive Analysis of Urban Flooding Susceptibility

This study, based on hydrological mechanisms and data-driven modeling requirements [9], selected the following 6 core indicators to predict UPFS: Population Density (POP), representing social exposure [9]; Points of Interest (POI) Density, reflecting urban development level and human activity intensity [9,15]; Impervious Surface Density (ISD), characterizing surface hardening degree and runoff enhancement effect [10]; Normalized Difference Vegetation Index (NDVI), embodying vegetation's functions of interception, flow retardation, and infiltration promotion [10,15]; Road Density (RD) and Drainage System Density (DSD), representing surface convergence capacity and municipal drainage capacity, respectively [8,10,24].

In this study, UPFS is calculated using variables related to flooding. It is important to note that while these variables may not directly cause flooding, they are highly correlated with its occurrence, such as characteristics of human activities [8–10]. Utilizing these variables allows for a high-precision estimation of UPFS. Therefore, the following features have been selected as flood-related variables. These variables possess clear hydrological significance and have demonstrated strong predictive power in machine learning models.

This study adopted the XGBoost classification algorithm to build the prediction model. The data were first cleaned and then split into training and validation sets in a 7:3 ratio. Hyperparameter optimization was performed via random search, and the probability of flooding within a grid cell was output as the UPFS. Model performance was comprehensively evaluated using Overall Accuracy (OA), the Kappa coefficient, the F1-score, and the Area Under the Curve (AUC) values for both the training and validation sets [25–29].

To evaluate the spatial pattern of the UPFS prediction results, we quantified its spatial characteristics. The Global Moran's I index was employed to test the spatial autocorrelation of the predicted UPFS. This index verifies whether UPFS exhibits significant clustering patterns in space [30–32].

To assess multi-collinearity among the predictor variables, correlation analysis and Variance Inflation Factor (VIF) tests were conducted prior to modeling [33,34]. A heatmap of Pearson correlation coefficients between variables was plotted to visually identify and remove highly correlated redundant variables. A VIF test was further used to quantify the degree of multicollinearity, and variables were screened according to the commonly accepted criterion (VIF < 5) to ensure the robustness and interpretative reliability of the subsequent regression model.

The XGBoost model was trained by iteratively optimizing the following objective function (Equation (1)), which includes the loss function and regularization term:

$$\ell^{(t)} = \sum_{i=1}^{n} \left[y_i \ln(1 + e^{-\bar{y}_i^{(t)}}) + (1 - y_i) \ln(1 + e^{\bar{y}_i^{(t)}}) \right] + \sum_{k=1}^{t} \Omega(f_k) \quad (1)$$

where $y_i \in \{0, 1\}$ is the true label of the ith sample (1 = flood point, 0 = non-flood point); $\hat{y}_i^t = \sum_{k=1}^{t} f_k(x_i)$ is the unnormalized prediction value (i.e., accumulated leaf node scores) at the tth iteration; $\Omega(f_k) = \gamma T_k + \frac{1}{2}\lambda \|w_k\|^2$ is the regularization term for the kth tree; T_k is the number of leaf nodes (controlling model complexity), w_k is the leaf node weight vector (penalizing extreme values), and $\Omega(f_k)$ constrains model complexity to prevent overfitting to the urban flooding system.

Subsequently, gradients are calculated in each iteration to efficiently split nodes:

$$g_i = \frac{\partial \ell}{\partial \hat{y}^{(t-1)}} = \hat{p}_i - y_i,\ h_i = \frac{\partial^2 \ell}{\partial \hat{y}^{(t-1)} \partial^2} = \hat{p}_i(1 - \hat{p}_i) \quad (2)$$

where $\hat{p}_i = \frac{1}{1 + e^{-q_i^t(t-1)}}$ is the predicted probability (i.e., UPFS value) for the *i*th sample; g_i is the first-order gradient (residual) of the logistic loss; h_i is the second-order gradient (Hessian), representing the confidence of the probability estimate.

Finally, the leaf weight sum is transformed into a probability via the Sigmoid function:

$$\mathrm{UPFS}_i = \frac{1}{1 + exp\left(-\sum_{k=1}^{K} f_k(\boldsymbol{x}_i)\right)} \quad (3)$$

where $x_i = [\mathrm{ISD}_i, \mathrm{NDVI}_i, \ldots]$ is the feature vector of the *i*th grid cell (see Table 1); $\mathrm{UPFS}_i \in [0, 1]$ represents the probability of flooding susceptibility for the cell, with higher values indicating greater risk.

2.3.3. Attribution Analysis of Urban Flooding Based on Explainable Machine Learning

In the regression analysis, UPF factors refer to those variables that have the potential influences to cause flooding and exhibit a certain causal relationship with it. To explain the occurrence of urban flooding from a geographical perspective, this study selected appropriate flooding drivers based on the physical attributes and land cover characteristics of the land surface [8,9].

The physical attributes of the surface are divided into two major categories: topographic features and soil texture. Topographic features directly influence the spatial distribution and movement of accumulated water, while soil texture determines the amount of surface runoff generated by controlling rainwater absorption and infiltration processes; thus, both were included in the core research scope [9]. Among topographic features, the Digital Elevation Model (DEM) was used as a fundamental topographic variable because it could accurately reflects ground elevation differences and is key to identifying low-lying catchment areas and potential water accumulation spaces [15,35]; Curvature quantifies the convex or concave form of the surface (positive curvature for convex, negative for concave, zero for flat). It influences the formation of water accumulation areas by altering the convergence or dispersion of water flow, making it an important supplementary variable [19,36]; Slope is a core element regulating the rhythm of water movement, directly determining the speed and extent of flooding formation, significantly impacting flood convergence efficiency and risk level, and is thus critically incorporated [19]. Regarding soil texture, Available Water Capacity (AWC) was selected as a representative variable because it directly reflects the soil's infiltration and water storage capacity, closely related to the rainwater infiltration process, and is a key indicator for measuring the soil's regulatory effect on surface runoff [9].

Land cover characteristics influence UPF by altering surface hydrological processes. Based on their direct effects on rainfall interception, infiltration, and discharge, this study selected the following core indicators: The Normalized Difference Vegetation Index (NDVI) was included because it characterizes vegetation coverage. Vegetation plays a key role in water retention, reducing soil erosion, and improving local microclimates, thereby reducing surface runoff by intercepting rainfall and enhancing infiltration [27,37]. Impervious Surface Density (ISD) was included because it directly affects rainwater infiltration efficiency. Hardened surfaces significantly impede infiltration and increase surface runoff, making it an important influencing factor [38]. Building Density (BD) influences runoff concen-

tration by altering the degree of surface hardening and spatial structure. Dense buildings can exacerbate rainwater accumulation, thus it is a key variable [7]. Distance to Water (DTW) is critically considered because it is directly related to drainage efficiency where areas near water are susceptible to backwater or inundation effects, leading to increased flooding risk [19,37]. Drainage density was included as a core indicator as it reflects the completeness of a region's drainage system, directly affecting rainwater removal efficiency. High drainage density can effectively reduce flooding risk and enhance a region's ability to withstand rainstorms [7,19].

It is noteworthy that continuing to use these indicators in the UPFS regression attribution analysis stage remains justified: UPFS itself is learned or calculated based on these input variables. Consistent variable semantics ensure a closed loop in prediction and explanation logic. Using explainable machine learning methods like SHAP and GeoMLR allows for quantifying the contribution degree of each input feature to the UPFS prediction results, thus achieving closure from prediction to causal explanation. For example, the SHAP method can calculate the contribution value of each feature to the model output [27], while GeoMLR + GeoShapley further incorporates spatial attributes to reveal the spatial heterogeneity and interaction effects of variable influences [9,27].

To investigate the driving mechanisms of each factor on UPFS, this study employed explainable machine learning methods for attribution analysis of the XGBoost model [8,27,39,40]. Firstly, the Shapley value (SHAP) method was used to quantify feature contributions. Shapley values decompose the contribution of each feature to the model prediction result into the sum of the feature's independent contribution and its interactive contribution with other features. A positive Shapley value for a feature indicates that a larger value for that feature increases UPFS; a negative value indicates a suppressive effect. By plotting feature importance ranking plots and partial dependence plots (PDPs), the nonlinear response trends of each factor to UPFS and second-order interaction effects can be intuitively understood.

Furthermore, considering the spatial heterogeneity of flooding influencing factors, this study introduces the Geographic Explainable AI (GeoXAI) framework and implements spatial attribution through the GeoMLR and GeoShapley methods [8,27,33,39,40] GeoMLR involves adding geographic coordinates as features to the XGBoost regression model to capture spatial dependency. The GeoShapley method merges the Shapley contributions of the X and Y coordinates into a unified spatial effect variable and quantifies the interactive contributions between this spatial variable and other factors. Through this GeoXAI analysis pipeline, the differential effects of driving factors in different regions can be revealed. This in-depth analysis, achieved through spatial embedding and SHAP interaction [8,9], aids in understanding the nonlinear driving mechanisms and regional characteristics of urban flooding, providing a quantitative basis for zonal management and risk control.

The following core equations illustrate the theoretical foundation of our model construction:

First is the decomposition of the Shapley value. For each prediction unit i, the Shapley value $\phi_j^{(i)}$ for the j-th feature is defined as:

$$\phi_j^{(i)} = \sum_{S \subset F \setminus \{j\}} \frac{|S|!(M-|S|-1)!}{M!} \left[f_{S\cup\{j\}}\left(x_{S\cup\{j\}}^{(i)}\right) - f_S\left(x_S^{(i)}\right) \right] \quad (4)$$

where F is the set of all features, $M = |F|$; S is a subset of features not containing feature j; f_s represents the model output when only the subset S of features is included.

The overall prediction of the model can be decomposed as:

$$\hat{y}^{(i)} = \phi_0 + \sum_{j=1}^{M} \phi_j^{(i)} \tag{5}$$

where ϕ_0 is the base prediction value (global mean).

Subsequently, for the GeoMLR extension and GeoShapley spatial attribution. In GeoMLR, geographic coordinates (X_i,Y_i) are introduced as additional features into the model:

$$\hat{y}^{(i)} = \beta_0 + \sum_{j=1}^{M} \beta_j x_j^{(i)} + \beta_X X_i + \beta_Y Y_i + \varepsilon_i \tag{6}$$

The GeoShapley method further merges the Shapley values of the X and Y coordinates to form a spatial effect variable $\phi_{\text{space}}^{(i)}$:

$$\phi_{\text{space}}^{(i)} = \phi_X^{(i)} + \phi_Y^{(i)} \tag{7}$$

Finally, the interactive contribution between $\phi_{\text{space}}^{(i)}$ and the Shapley values of other features is analyzed:

$$\text{Interaction}(p, q) = \phi_{p,q}^{(i)} - \phi_p^{(i)} - \phi_q^{(i)} \tag{8}$$

3. Results

3.1. Model Performance Results

3.1.1. UPFS Assessment Results

After hyperparameter optimization via random search (n_iter = 100), the optimal parameter set for the XGBoost classifier was determined as follows: subsample = 0.6, reg_lambda = 50, n_estimators = 150, min_child_weight = 3, max_depth = 6, learning_rate = 0.15, gamma = 0.1, and colsample_bytree = 0.8. The model exhibited strong performance during testing: overall accuracy (OA) reached 0.9259, indicating that over 92% of flood-prone and non-flood-prone units were correctly predicted(Table 2): the Kappa coefficient was 0.8519, reflecting substantial agreement between predictions and observations; and the F1-score reached 0.9231, demonstrating a effective balance between precision and recall while minimizing both false positives and false negatives. Furthermore, the area under the curve (AUC) values were 0.9568 for the training set and 0.9882 for the validation set, exceeding performance levels reported in other comparable studies. These results demonstrate that the classifier possesses strong discriminative power for assessing flooding susceptibility.

Table 2. Performance of the prediction model.

Metric	Training Set	Validation Set	Training-Validation Gap
Accuracy	0.8667	0.7556	0.1111
Precision	0.8545	0.7037	0.1508
Recall	0.8868	0.8636	0.0232
F1-Score	0.8704	0.7755	0.0949
AUC	0.9318	0.8597	0.0721
Log Loss	0.5223	0.5694	0.0471

To address the overfitting risk implied by initial high metrics, we reconstructed the model with enhanced regularization and a stratified 7:3 training-validation split (ensuring consistent label distribution between sets, critical for imbalanced flood point data). The model's performance on both training and validation sets is realistic and stable, with no severe overfitting (training-validation metric gaps are controlled within reasonable ranges):

Key observations confirming no severe overfitting:

The AUC gap between training and validation sets is only 0.0721, far below the threshold (>0.2) that indicates significant overfitting; Recall (ability to identify actual flood points) remains high and stable (training: 0.8868, validation: 0.8636), with a minimal gap of 0.0232—proving the model captures the intrinsic relationship between flood drivers and UPFS, rather than fitting noise; Log loss values for both sets are close (gap: 0.0471), reflecting the model's consistent probability estimation ability across datasets [15].

Based on the high-performance classification results, the spatial distribution of UPFS across Changsha was illustrated (Figure 3). The map indicated distinct spatial patterns of flood risk. Areas of high susceptibility are predominantly concentrated within the hill–basin areas along banks of the Xiang River in central Changsha. This region is characterized by relatively low-lying topography interspersed with small hills, combined with a high proportion of impervious surfaces resulting from intensive urban development, which significantly reduces rainwater infiltration capacity. The backwater effect caused by rising water levels in the Xiang River further increases flooding vulnerability in this area. Additionally, several areas with high topographic relief in Ningxiang City (western region) and Liuyang City (eastern region) also exhibit high flooding susceptibility. Certain parts of Ningxiang, situated at the confluence of hilly and plain terrains, experience reduced rainwater convergence efficiency due to complex topographic variations. In Liuyang, local river distribution and high rainfall intensity contribute significantly to the observed flooding risks.

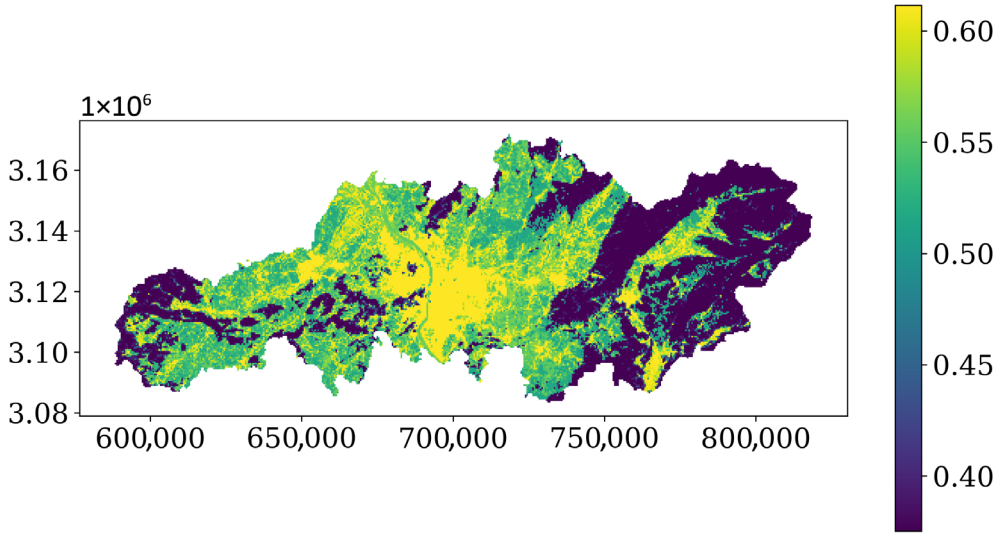

Figure 3. UPFS distribution map of Changsha. Notes: (1) X (Easting)/Y (Northing): meters, WGS 1984 UTM Zone 49N; (2) Color: UPFS (0–1), higher = greater susceptibility.

3.1.2. Regression Analysis Model Evaluation

The results of the Moran's I analysis showed that the global Moran's I for UPFS in Changsha was 0.8257 ($p < 0.001$), indicating a significant positive spatial autocorrelation. This implies that the UPFS in Changsha exhibited a pronounced clustering characteristic in space, which is consistent with findings on spatial risk patterns in similar cities reported in existing studies [30–32].

Prior to regression modeling, a correlation test was conducted on all candidate variables to avoid the interference of multi-collinearity on the interpretation of results. Preliminary correlation heatmap revealed that the correlation coefficient between BD and Building Area exceeded 0.75, indicating high information redundancy between these two variables. Consequently, the Building Area variable was removed, retaining only BD as the representative indicator. After this screening step, the correlation coefficients among the

remaining variables were all below 0.75 (Figure 4), effectively retaining core information while reducing the potential impact of redundant features on model stability.

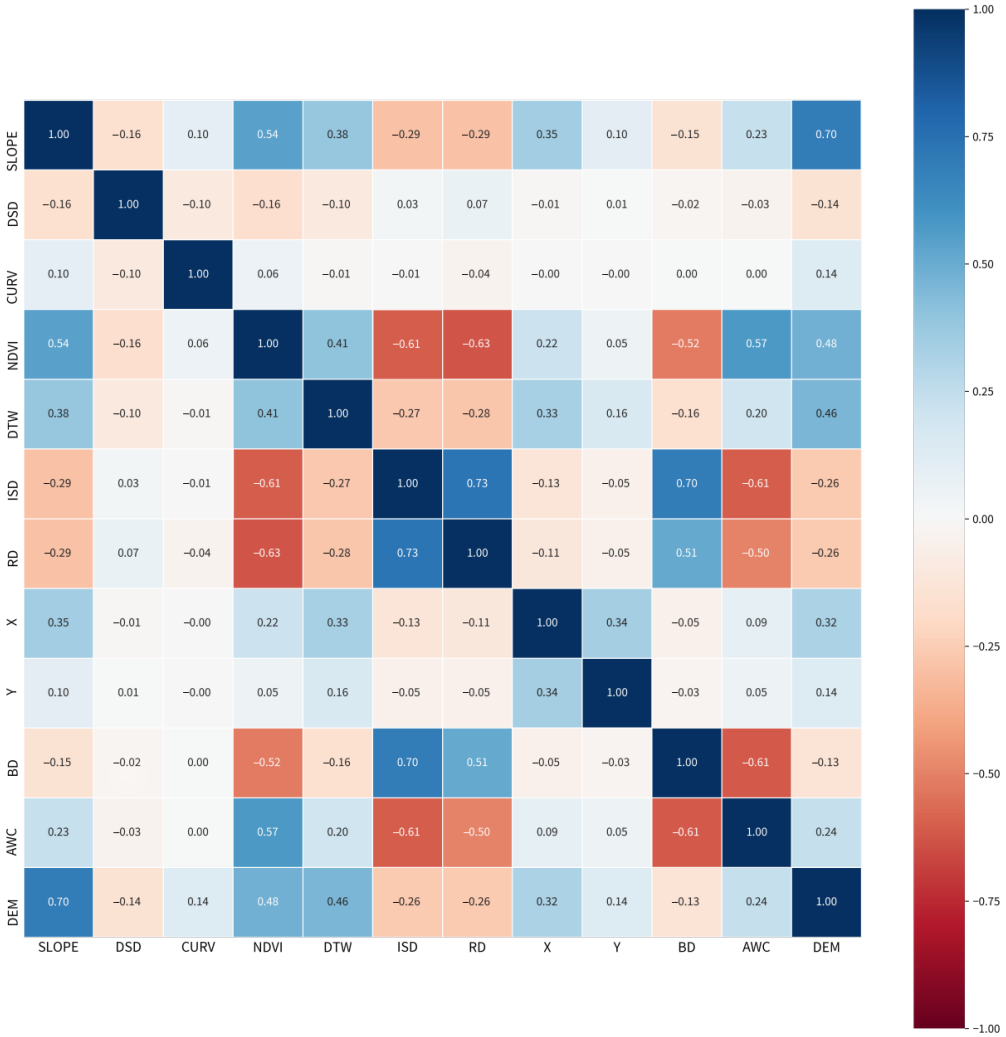

Figure 4. Heatmap of multicollinearity among independent variables.

Subsequently, a VIF test was performed on the retained variables. The results showed that the VIF values for all variables were below 5 (Table 3), which is under the commonly used threshold of 10 in academia and also satisfies the stricter criterion (VIF < 5) often employed in geographical and environmental science research [9]. This confirms that no severe multi-collinearity situation among the variables, allowing them to proceed directly to the regression analysis stage.

On this basis, the GeoMLR model was employed to investigate the relationship between urban flooding drivers and UPFS. The optimal hyperparameters were determined via the random search method (n_iter = 100) as follows: n_estimators = 300, learning_rate = 0.1, max_depth = 6, min_child_weight = 2, subsample = 0.8, and colsample_bytree = 0.9. The model's robustness was enhanced through 5-fold cross-validation. To comprehensively evaluate the model's fitting precision, prediction error, and generalization ability (with a focus on detecting potential overfitting), key performance metrics were compared between the training set and test set, and the results are summarized in Table 4.

Table 3. Variance Inflation Factor multicollinearity test.

No.	Variable	VIF
1	ISD	4.324877
2	NDVI	3.62101
3	BD	2.220939
4	RD	2.236096
5	AWC	1.906527
6	DEM	3.262558
7	Slope	3.412544
8	Dis_to_water	1.438868
9	X	1.326951
10	Y	1.140251
11	Drainage	1.069217
12	Curvature	1.038764

Table 4. Performance comparison of the regression models on the training set and test set.

Evaluation Metric	Training Set	Test Set	Metric Difference (Training-Test)
Mean Squared Error (MSE)	0.0007	0.0011	−0.0004
Root Mean Squared Error (RMSE)	0.0260	0.0335	−0.0075
Mean Absolute Error (MAE)	0.0179	0.0232	−0.0053
Coefficient of Determination (R^2)	0.8522	0.7546	0.0976

The evaluation results demonstrate that the GeoMLR model achieved a high coefficient of determination (R^2 = 0.7546) on the test set, with a Mean Absolute Error (MAE) of 0.0232 and a Mean Squared Error (MSE) of 0.0011, indicating high fitting precision and low prediction error. More importantly, there is an extremely small difference and consistent trend between the metrics of the training set and test set, confirming the absence of obvious overfitting. Among the error-based metrics (MSE, RMSE, MAE), the test set values are slightly higher than those of the training set, but the differences are only 0.0004, 0.0075, and 0.0053, respectively—all within a very narrow range that reflects the normal performance of the model on unseen data. For the R^2 value, the training set (0.8522) and test set (0.7546) differ by only 0.0976 without a significant gap, which indicates that the model has fully learned effective information from the training data while avoiding overfitting to its specific characteristics, thus possessing robust generalization ability. Compared to existing literature utilizing machine learning methods for studying the spatial distribution of flooding [7,15,41], the performance of our model is among the superior levels, validating the effectiveness and applicability of GeoMLR in characterizing nonlinear spatial driving mechanisms.

3.2. Influence of Flooding Factors on Urban Pluvial Flooding Susceptibility

3.2.1. Global Influence of Flooding Pluvial Factors

The GeoShapley Summary Plot (Figure 5) visually represented the global contribution of each factors and their interaction terms to the prediction outcome, i.e., UPFS. This facilitates the analysis of mechanism through which different features operate within geographical space. Features were ranked vertically from top to bottom based on their contribution magnitude, reflecting the strength of both independent and synergistic effects. The horizontal axis corresponded to the GeoShapley value, where a positive value indicated that an increase in the feature enhances UPFS, and a negative value suggested a suppressing effect. Point color reflects the original value of the feature, with blue representing low values

and red high values, aiding in interpreting the relationship between feature magnitude and influence direction.

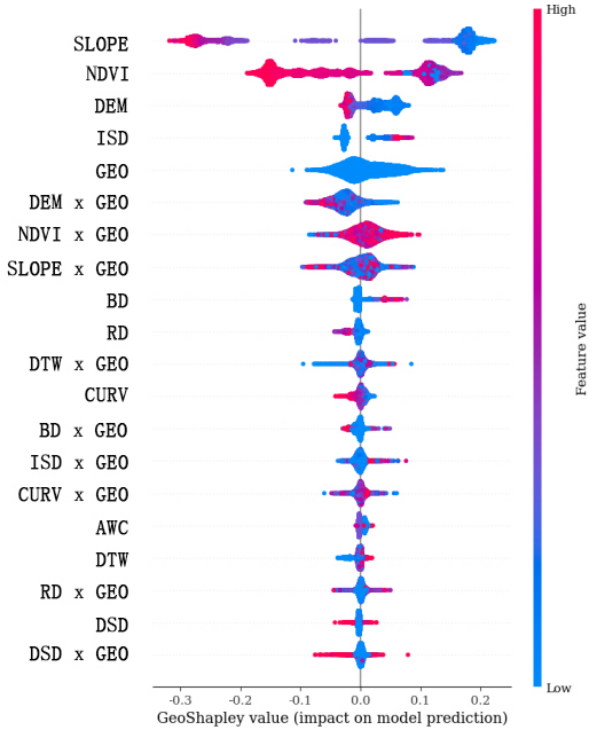

Figure 5. Global variable SHAP summary plot and ranking.

Among the most influential factors, high values of Slope were largely distributed in the negative value range, indicating that steeper slopes facilitate rainwater drainage, reduce water accumulation, and consequently lower flooding risk. This finding aligns with existing understanding of slope's role in controlling surface runoff paths and detention time [10]. Furthermore, this study revealed that even in areas with gentle slopes, the flood-suppressing effect could be attenuated by high impervious surface coverage, a nuance seldom captured in traditional statistical models.

High values of the NDVI were concentrated in the negative range, implying that greater vegetation coverage reduces flooding risk by intercepting precipitation, promoting infiltration, and diminishing surface runoff. This result correlates with existing research on the stormwater regulation capacity of urban ecological infrastructure [27]. In addition, this study identifies a spatial divergence in NDVI's mitigating effect: flood suppression is more pronounced in suburban woodlands and contiguous green spaces than in fragmented urban vegetation.

Low values of the DEM clustered in the positive range, confirming that low-lying terrain (e.g., valleys and depressions) naturally promotes water convergence and increases flood propensity [41]. Beyond conventional DEM analysis, incorporating geographic coordinates via the GeoShapley method allows this study to reveal district-level variations in flood-promoting intensity. For example, low-lying areas in Yuelu District, characterized by a dense river network, demonstrate a stronger flood-enhancing effect than eastern industrial suburbs at comparable elevations.

High values of ISD show a clear positive correlation with UPFS, indicating that extensive impervious areas inhibit rainwater infiltration, increase surface runoff, and elevate flooding risk, consistent with established hydrological studies [7,19]. Notably, the interaction between ISD and RD exerted a stronger flood-promoting effect in high-density

urban settings than in low-density suburban areas, underscoring the modulating role of spatial context in flood generation processes.

Among less influential factors, BD, though globally less dominant, still exhibited a slight risk-increasing effect at high values, mainly by reducing stormwater storage space and intensifying runoff concentration, albeit weaker than primary factors such as ISD. This study also found that in high-slope areas, BD's flood-promoting role was partially offset by efficient drainage [19]. RD sporadically appeared in the positive range, suggesting that as an extension of impervious surfaces, it contributed to runoff convergence, though often masked by the effect of ISD. Curvature showed limited correlation with GeoShapley values, implying only minor influence on flow paths in localized micro-terrains such as narrow valleys. Low values of Available Water Capacity (AWC) were occasionally associated with positive contributions, indicating that poor soil water retention may modestly increase surface runoff, though this effect is often mitigated by other factors such as vegetation and slope [9].

3.2.2. Nonlinear Influence of Flooding Factors

The set of GeoShapley Partial Dependence Plots (PDPs) (Figure 6) presents a detailed analysis of the nonlinear influence of each feature on UPFS. The feature values are plotted on the x-axis, and the corresponding GeoShapley values, which quantify the marginal contribution of each feature to the prediction, on the y-axis. A positive GeoShapley value indicates an increase in flooding risk, while a negative value suggests a decrease. The red average trend curve, combined with blue scatter points representing sample-level distribution, demonstrates that most factors exhibit pronounced nonlinear behavior.

Specifically, Analysis indicated that slope was associated with increased flooding risk in areas with gentle slopes (<5°) due to slow drainage. When slope exceeded approximately 5°, however, flooding risk decreases markedly as steeper terrain accelerates runoff. This flood-suppression threshold of 5° is consistent with findings from hilly regions [19] but higher than that reported for plain cities such as Wuhan (>2°) [42], highlighting how medium slopes in transitional zones (e.g., between western hills and eastern plains) facilitate drainage, whereas extensive basin areas accumulate risk.

NDVI was linked to higher flooding risk (positive GeoShapley values) in low-value areas (NDVI < 0.7, typically barren or sparsely vegetated land) owing to poor water retention. When ND exceeds 0.7, suggesting dense vegetation, flood risk declined substantially, owing to the enhanced interception and infiltration capacity. This suppression threshold of 0.7 aligns with findings in [9], likely resulting from the high biomass and improved soil properties under Changsha's subtropical monsoon climate [43], illustrating how interactions among vegetation and climate factors shape threshold behavior.

For the DEM, GeoShapley values were predominantly positive at elevations below 150 m, particularly along the Xiang River and eastern alluvial plain. This confirms that low-lying areas act as natural convergence zones with high flood propensity. Above 150 m, flooding risk decreased due to improved drainage aided by topographic uplift. The inflection point at 150 m aligns closely with local geomorphology, influenced by the river valley effect and water level patterns of the Xiang River [44]. These values differ from those reported in other settings [8,9], highlighting the importance of local hydrologic and topographic conditions.

ISD was negatively correlated with flooding risk (near-zero or negative GeoShapley values) at low values (<0.35) where infiltration was favorable. However, Beyond 0.35, characteristic of urban built-up areas, surface runoff increases sharply, elevating flood risk. In older urban districts such as Furong, even ISD values slightly below 0.35 may approach this threshold due to aging drainage systems designed for low return periods (1–3 years).

This suggests that the condition of infrastructure influence the effective imperviousness threshold [8,9].

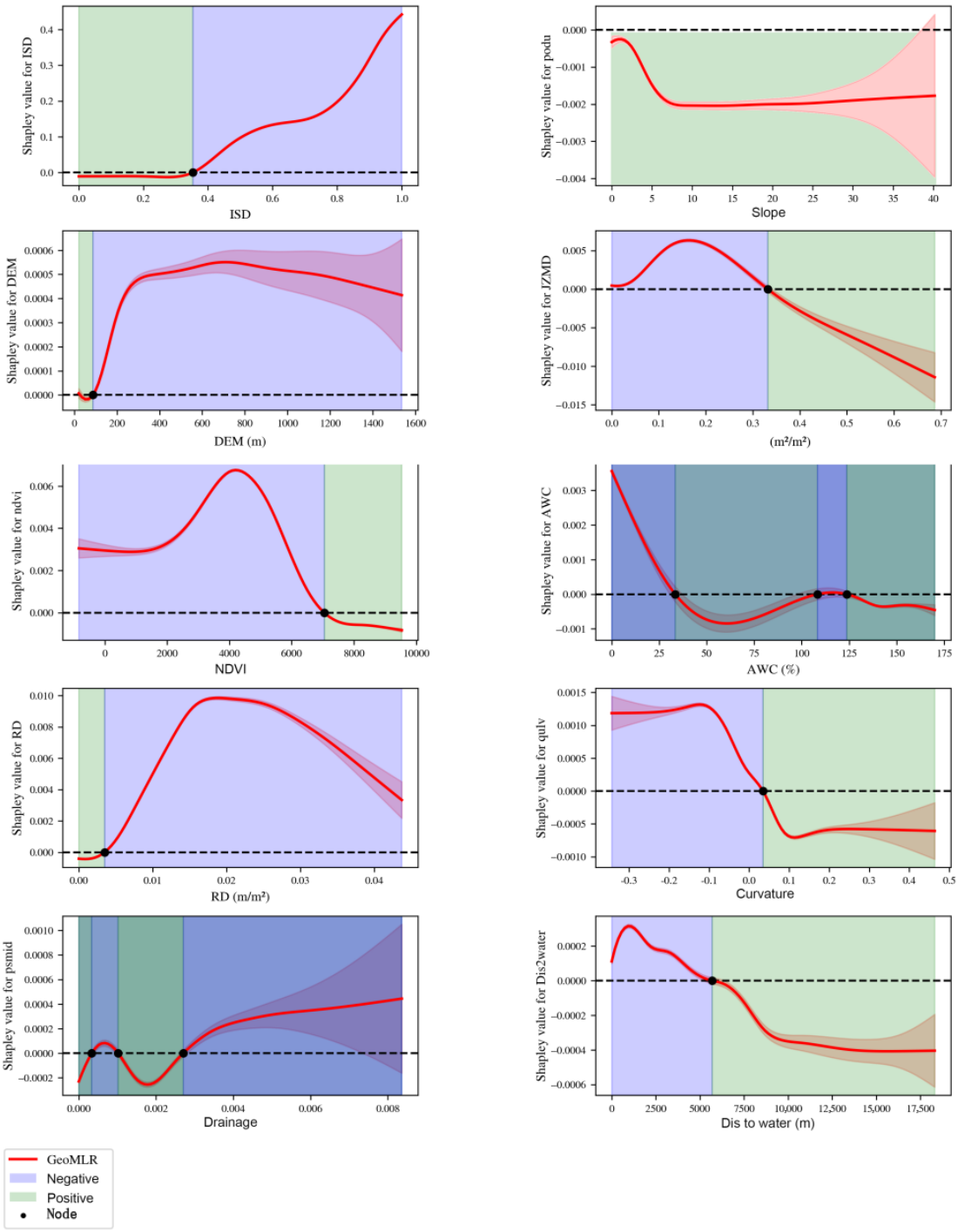

Figure 6. Partial dependence plots (PDPs) illustrating the nonlinear effects of driving factors on UPFS.

3.2.3. Spatial Heterogeneity of Flooding Factors

To quantify the spatially heterogeneous contributions of various driving factors to UPFS, this study applied an independent normalization method to the GeoShapley values, using TwoSlopeNorm symmetric normalization based on the global range of each factor's Shapley values. A red-blue color scheme was used to represent flood-promoting and flood-suppressing effects, respectively. This approach effectively eliminated scale

differences and clearly revealed the spatial patterns and directional characteristics of factor contributions (Figure 7).

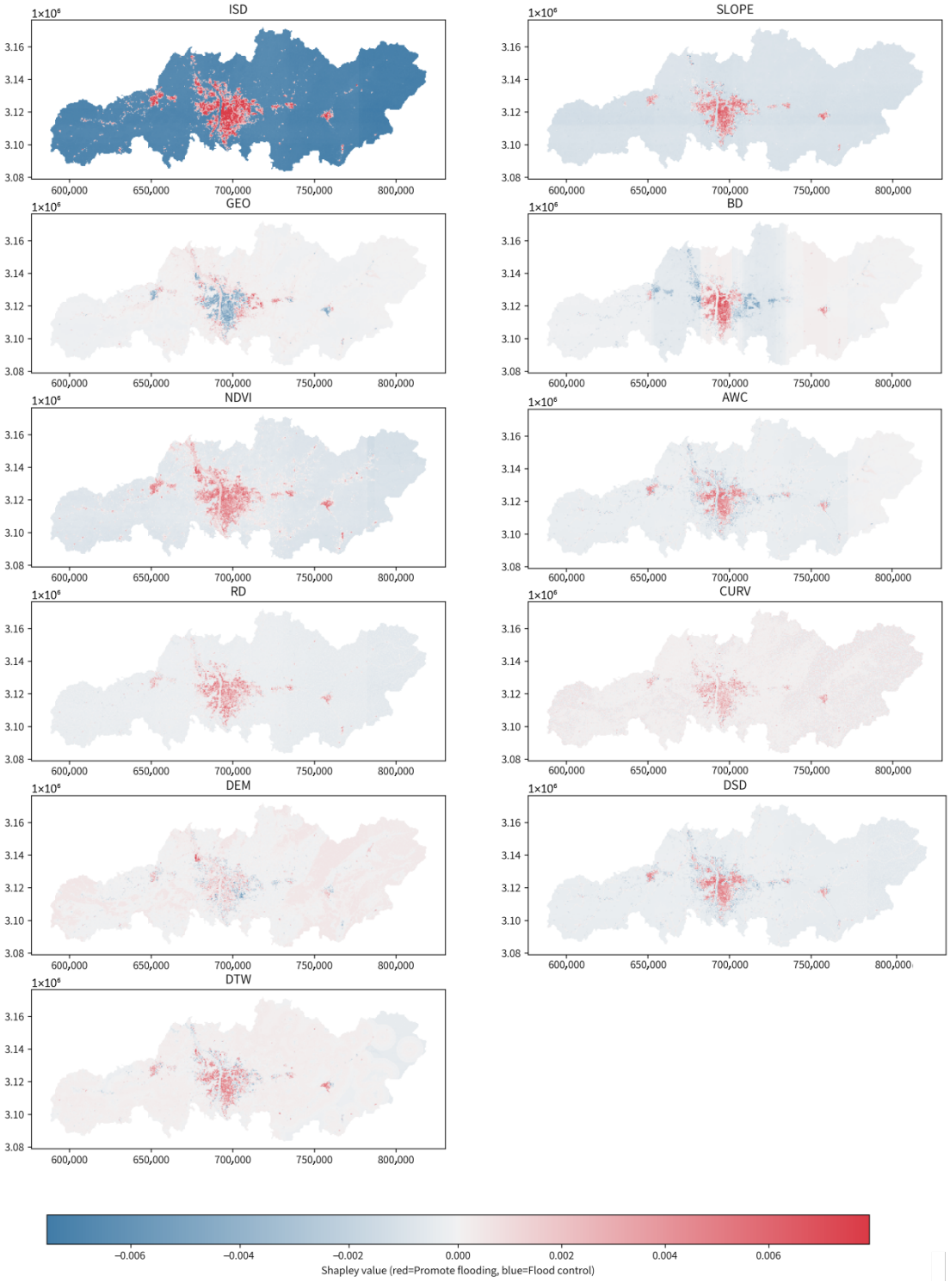

Figure 7. Spatial heterogeneity of flooding factor contributions to UPFS. Notes: X/Y (Easting/Northing): meters, WGS 1984 UTM Zone 49N; subfigures = key factors; color = GeoShapley values (red = promote flooding, blue = suppress flooding).

The analysis showed that areas with a strong flood-promoting effect of ISD (ISD > 0.35) are concentrated along the Xiang River in Furong, Tianxin, and northern Kaifu District, which align closely with high-density built-up areas. Compared with studies from Guangzhou [7] and the Pearl River Delta [41], the flood-promoting effect of ISD

in low-lying areas along the Xiang River in Changsha is enhanced due to the combined influence of topographic backwater effects and impervious surfaces [9]. Although the overall influence of Road Density (RD) was limited, localized clusters of higher flood promotion were observed near major roads in eastern Yuhua District, displaying a spatial pattern similar to that of ISD. This underscores a synergistic flood-amplifying mechanism between road networks and impervious surface expansion [10], which contrasts with the moderating role of building morphology reported by Zhou et al. [19].

Natural regulating factors displayed clear spatial complementarity and topographic modulation. The flood-suppressing effect of the NDVI (NDVI > 0.7) was prominent in the woodlands of Yuelu Mountain and contiguous farmland in northwestern Wangcheng District, consistent with findings in tropical watersheds [37] and Poyang Lake region [27]. This study further revealed that fragmented urban green spaces (e.g., parks in Yuhua District) exhibit lower flood mitigation efficiency compared with contiguous suburban vegetation [9], highlighting the importance of vegetation spatial configuration. The flood-suppressing effect of Slope (Slope > 5°) was distinct in western hilly areas, but minimal in eastern plains. This "strong in the west and weak in the east" divergence contrasts with the more uniform flood mitigation by gentle slopes observed in the Hefei plain [10], reflecting the regulatory effect of hill-plain transition topography on runoff convergence [15]. Contributions from Elevation (DEM) showed "bipolar differentiation": negative in western high-altitude areas (DEM > 150 m) and positive in low-lying zones along the Xiang River (DEM < 50 m). This differs notably from the gradual elevation effects documented in plain cities [18], demonstrating the structural control of terrace topography on flooding drivers in Changsha.

In summary, areas with intensive human activity (dominated by ISD/RD) primarily exhibited flood-promoting effects, while zones where natural factors (NDVI/Slope/DEM) tended to suppress flooding, forming a distinct spatial interplay. Compared to coastal cities [38], natural suppression zones in Changsha show greater spatial continuity (e.g., the contiguous western hill area), supporting the delineation of ecological flood control barriers. This study highlights unique flooding driver patterns in hill–plain cities: the efficacy of natural factors is influenced by topographic continuity, anthropogenic risks are amplified in river valley lowlands, and outdated infrastructure in older urban areas can alter safety thresholds.

4. Discussion

This study applied a GeoXAI framework to systematically analyze the driving mechanisms of urban flooding susceptibility in Changsha, a representative city within a hill–basin transition zone. The results demonstrated that flooding is influenced by the joined effects of natural factors (e.g., topography and vegetation) and human activities (e.g., impervious surfaces), consistent with existing literature [8,10,15,27]. More importantly, this research revealed distinct nonlinear thresholds and spatial heterogeneity in flooding drivers, providing new evidence to deepen the mechanistic understanding of flood causation and inform targeted mitigation strategies.

The findings indicate that flooding in such hill-basin settings results from the synergistic effects of natural topography, vegetation coverage, and surface imperviousness—specifically, the overlap of flood-promoting threshold ranges of key drivers. The identified nonlinear thresholds for key variables, which directly drive flooding when in the stated ranges, are: (1) Impervious Surface Density (ISD) > 0.35 (inhibits rainwater infiltration, accelerates runoff to promote flooding); (2) Normalized Difference Vegetation Index (NDVI) < 0.70 (weakens vegetation's rainwater interception and infiltration capacity, losing flood mitigation to promote flooding); (3) Slope < 5° (retards runoff velocity, causes water accumulation to promote

flooding). These thresholds not only reflect Changsha's region-specific hydro-geomorphic response but also underscore flood resilience variations across urban types. These insights enhance understanding of the transferability of nonlinear thresholds within urban disaster prevention frameworks.

Regarding influential factors, Impervious surface density was identified as a key driver increasing flood risk, whereas vegetation coverage and slope showed suppressive effects, a pattern consistent with numerous international and domestic studies [15,33]. For example, Qin et al. highlighted that urbanization-driven impervious expansion amplifies runoff generation and concentration [15], while Li et al. emphasized the role of NDVI and slope in reducing peak flow and runoff [33]. Unlike previous studies focused on plain, mountainous, or coastal cities, this work reveals a geomorphic modulation effect specific to hill–plain settings. Specifically, high impervious surface density exhibited stronger flood-promoting effects in low-lying convergence zones, while the mitigating effect of NDVI is more pronounced in hilly and upland areas.

In terms of threshold effects, this study identified critical values (ISD < 0.35, NDVI > 0.70, Slope > 5°) beyond which the marginal effects of these factors change substantially. These align generally with the "risk inflection point intervals" reported in limited existing work on nonlinear flood influences (e.g., ISD: 0.20–0.35; NDVI: 0.65–0.75) [8,27,39] A key contribution is the demonstration of the spatial dependency of these threshold effects. For instance, the flood-promoting threshold of impervious surface density is more sensitive in low-lying areas along the Xiang River, while the suppression threshold of NDVI is clearer near foothills. This challenges the conventional use of universal thresholds based on global averages [9,33] and supports a shift toward customized, zonal governance strategies for flood management.

From a methodological perspective, although traditional hydrological models offer detailed process representation, they face data and parameterization constraints in large-scale, multi-factor contexts [11–14]. Spatial statistical models such like GWR model capture local variations but often fail to account for complex nonlinearities and interactions [18]. The GeoXAI framework (XGBoost + GeoMLR + GeoShapley) adopted in this study overcomes these limitations, maintaining high predictive performance while identifying nonlinear thresholds, factor interactions, and spatial differentiations. This aligns with recent efforts to integrate explainability into machine learning for disaster susceptibility research [7,19,39,40]. Compared to deep learning methods, the present approach does not autonomously extract spatiotemporal features but offers superior interpretability and policy relevance, effectively bridging theoretical modeling and practical governance.

Additionally, factors such as road density, building density, curvature, and soil available water capacity showed limited global influence. Nonetheless, they may still amplify runoff or modulate flooding at localized scales, such as in high-density urban blocks, road catchments, or micro-topographic settings [8,10,15].

5. Conclusions

This study developed a framework based on Geographically Explainable Artificial Intelligence (GeoXAI) to analyze nonlinear spatial driving mechanisms of urban flooding in cities with complex hill–basin topography. The framework was used to investigate compound driving factors, threshold behaviors, and spatial heterogeneous patterns in flood formation. Compared to previous studies, this work offers insightful methodologies and perspectives for quantitatively examining flooding mechanisms in complex topographic settings. It also demonstrates the notable potential of GeoXAI in elucidating nonlinear driving relationships within urban development area. The results identify key flooding drivers and their nonlinear effects, and propose a refined, targeted management strategy

for urban flooding that accounts for multi-factor interactions between topography and the environment. By dividing the region into high-risk zones, ecological barrier areas, and basin–hill conjunction zones based on risk-exceeding thresholds, this approach supports a shift from uniform management to precision regulation. The framework is not only applicable to Changsha but also offers a quantifiable and transferable model for flood prevention in other cities with analogous topographic conditions. It provides a useful reference for designing targeted policies across different climate zones and stages of urbanization.

This study also acknowledges several limitations. Among these, two data-related limitations merit attention. On the one hand, the multi-source data used are primarily static; notably, short-duration meteorological elements (e.g., extreme rainfall) were not incorporated, which limits the model's applicability for dynamic early warning and real-time risk assessment. On the other hand, temporal asynchrony in individual datasets introduces potential uncertainty: while 2022 was uniformly adopted as the baseline year (with 2022 annual average data for most time-variant indicators), the Population Density (POP) dataset relies on a refined 2020 National Census version. This is due to China's ten-year census cycle (next in 2030), making 2022 census data unavailable. Even though Changsha's 2020–2022 urbanization rate growth and population density changes are minimal, slight discrepancies between 2020-based POP data and 2022's actual socioeconomic exposure may marginally affect flood risk assessment accuracy related to population vulnerability.

Beyond data constraints, the study faces limitations in method and generalization. First, the GeoShapley method has limited capability to resolve higher-order interaction effects, such as the synergistic mechanisms among impervious surface density, drainage density, and rainfall intensity. At present, it can only detect two-factor interactions, which restricts comprehensive understanding of multi-factor compound effects. Second, Another methodological limitation relates to the validation scheme of the GeoMLR model. While we incorporated geographic coordinates (X/Y) as features to capture spatial non-stationarity—critical for analyzing the spatial heterogeneity of flooding factors in hill-basin cities—we initially adopted random 5-fold cross-validation rather than spatially blocked validation (e.g., spatial k-fold, leave-location-out). This approach may have introduced potential bias: tree-based models are prone to learning location-specific patterns of flood points (e.g., clustering of flood events along the Xiang River) rather than the intrinsic hydrological processes (e.g., how ISD and slope synergistically drive runoff).

Although the model's training-test metric differences (Table 4: R^2 gap = 0.0976, MSE gap = 0.0004) suggest no severe overfitting, spatially blocked validation would more strictly isolate training and validation samples by geographic units, avoiding spatial autocorrelation-induced metric inflation. This limitation slightly affects the model's ability to fully decouple 'location effects' from 'process effects' when interpreting spatial heterogeneity, and we have tempered our conclusions on 'spatial dependency of threshold effects' (Section 4) to emphasize that these patterns should be interpreted alongside local topographic and hydrological contexts rather than relying solely on model outputs.

Future research could prioritize developing interpretable models capable of handling higher-order interactions, coupling hydrodynamic processes with indicators of implicit hydrological connectivity, and integrating real-time meteorological monitoring data. These improvements would enhance the model's predictive accuracy and mechanistic interpretability under complex scenarios. Expanding the empirical scope to include cities with varied geomorphologies and climate zones is also essential for validating the robustness and applicability of the proposed framework, thereby supporting the advancement of urban flood risk modeling from mechanism-based analysis toward general theory establishment.

Author Contributions: Conceptualization, Z.H. and Y.C.; Methodology, Z.H., B.L. and S.X.; Software, Q.N.; Validation, Y.C. and B.L.; Formal analysis, S.X.; Investigation, S.X. and S.T.; Resources, Q.N. and S.T.; Data curation, Y.C.; Writing—original draft, Z.H.; Writing—review & editing, Z.H., B.L. and S.T.; Visualization, Z.H.; Supervision, Y.C. and B.L.; Funding acquisition, Q.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Hunan Provincial Postgraduate Scientific Research Innovation Project (No. CX20240985), the Key Project of Hunan Provincial Department of Education, China (No. 24A0567), and the Teaching Reform Research Project of Ordinary Colleges and Universities in Hunan Province, China (No. HNJG-2022-0996).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data sources have been provided in Table 1. The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Chen, D.; Pan, C.; Qiao, S.; Zhi, R.; Tang, S.; Yang, J.; Feng, G.; Dong, W. Evolution and prediction of the extreme rainstorm event in July 2021 in Henan province, China. *Atmos. Sci. Lett.* **2023**, *24*, e1156. [CrossRef]
2. Nandargi, S.; Gaur, A.; Mulye, S.S. Hydrological analysis of extreme rainfall events and severe rainstorms over Uttarakhand, India. *Hydrol. Sci. J.* **2016**, *61*, 2145–2163. [CrossRef]
3. Li, Y.; Huang, T.; Li, H.; Li, Y. Multi-Scenario Urban Waterlogging Risk Assessment Study Considering Hazard and Vulnerability. *Water* **2025**, *17*, 783. [CrossRef]
4. Xiao, H.; Zhang, J.; Wang, F.; Kong, L. A new framework for early-warning classification of urban waterlogging risk based on waterlogging risk angle. *Sustain. Cities Soc.* **2025**, *127*, 106460. [CrossRef]
5. Nie, Y.; Chen, J.; Xiong, X.; Wang, C.; Liu, P.; Zhang, Y. Formation Mechanism and Response Strategies for Urban Waterlogging: A Comprehensive Review. *Appl. Sci.* **2025**, *15*, 3037. [CrossRef]
6. Mo, G.; Xing, Z.; Zhu, Y.; Huang, L.; Chen, W.; Li, J. Social Media Data-Based Rapid Hazard Assessment of Urban Waterlogging Event: A Case Study of Guilin 6.19 Waterlogging. *Water* **2025**, *17*, 354. [CrossRef]
7. Wang, M.; Li, Y.; Yuan, H.; Zhou, S.; Wang, Y.; Ikram, R.M.A.; Li, J. An XGBoost-SHAP approach to quantifying morphological impact on urban flooding susceptibility. *Ecol. Indic.* **2023**, *156*, 111137. [CrossRef]
8. Liu, L. An ensemble framework for explainable geospatial machine learning models. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *132*, 104036. [CrossRef]
9. Ke, E.; Zhao, J.; Zhao, Y. Investigating the influence of nonlinear spatial heterogeneity in urban flooding factors using geographic explainable artificial intelligence. *J. Hydrol.* **2025**, *648*, 132398. [CrossRef]
10. Xing, Z.; Lyu, G.; Yao, Y.; Liu, Z.; Zhang, X. Fine-grained analysis and mapping of urban flood susceptibility with interpretable machine learning: A case study of Hefei, China. *J. Hydrol. Reg. Stud.* **2025**, *60*, 102501. [CrossRef]
11. Sañudo, E.; Cea, L.; Puertas, J. Modelling Pluvial Flooding in Urban Areas Coupling the Models Iber and SWMM. *Water* **2020**, *12*, 2647. [CrossRef]
12. Assaf, M.N.; Manenti, S.; Creaco, E.; Giudicianni, C.; Tamellini, L.; Todeschini, S. New optimization strategies for SWMM modeling of stormwater quality applications in urban area. *J. Environ. Manag.* **2024**, *361*, 121244. [CrossRef]
13. Rajib, A.; Liu, Z.; Merwade, V.; Tavakoly, A.A.; Follum, M.L. Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP. *J. Hydrol.* **2020**, *581*, 124406. [CrossRef]
14. Chunn, D.; Faramarzi, M.; Smerdon, B.; Alessi, D.S. Application of an Integrated SWAT–MODFLOW Model to Evaluate Potential Impacts of Climate Change and Water Withdrawals on Groundwater–Surface Water Interactions in West-Central Alberta. *Water* **2019**, *11*, 110. [CrossRef]
15. Qin, X.; Wang, S.; Meng, M.; Long, H.; Zhang, H.; Shi, H. Enhancing urban resilience through machine learning-supported flood risk assessment: Integrating flood susceptibility with building function vulnerability. *NPJ Urban Sustain.* **2025**, *5*, 19. [CrossRef]
16. Zhang, D.; Tang, P.; Tang, C.; Lai, X. Interpretable machine learning unveils nonlinear drivers of global energy risk spillovers: A TVP-VAR approach. *Econ. Model.* **2025**, *151*, 107178. [CrossRef]
17. He, X.; Lang, Q.; Zhang, J.; Zhang, Y.; Jin, Q.; Xu, J. Interpretable Machine Learning for Explaining and Predicting Collapse Hazards in the Changbai Mountain Region. *Sensors* **2025**, *25*, 1512. [CrossRef]

18. Littidej, P.; Buasri, N. Built-Up Growth Impacts on Digital Elevation Model and Flood Risk Susceptibility Prediction in Muaeng District, Nakhon Ratchasima (Thailand). *Water* **2019**, *11*, 1496. [CrossRef]
19. Zhou, S.; Liu, Z.; Wang, M.; Gan, W.; Zhao, Z.; Wu, Z. Impacts of building configurations on urban stormwater management at a block scale using XGBoost. *Sustain. Cities Soc.* **2022**, *87*, 104235. [CrossRef]
20. Gong, P.; Li, X.; Wang, J.; Bai, Y.; Cheng, B.; Hu, T.; Liu, X.; Xu, B.; Yang, J.; Zhang, W.; et al. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens. Environ.* **2020**, *236*, 111510. [CrossRef]
21. Sinitambirivoutin, M.; Milne, E.; Schiettecatte, L.-S.; Tzamtzis, I.; Dionisio, D.; Henry, M.; Brierley, I.; Salvatore, M.; Bernoux, M. An updated IPCC major soil types map derived from the harmonized world soil database v2.0. *Catena* **2024**, *244*, 108258. [CrossRef]
22. Chen, Y.; Xu, C.; Ge, Y.; Zhang, X.; Zhou, Y. A 100 m gridded population dataset of China's seventh census using ensemble learning and big geospatial data. *Earth Syst. Sci. Data* **2024**, *16*, 3705–3718. [CrossRef]
23. 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]
24. Tian, J.; Chen, Y.; Yang, L.; Li, D.; Liu, L.; Li, J.; Tang, X. Enhancing Urban Flood Susceptibility Assessment by Capturing the Features of the Urban Environment. *Remote Sens.* **2025**, *17*, 1347. [CrossRef]
25. Nikraftar, Z.; Parizi, E.; Saber, M.; Boueshagh, M.; Tavakoli, M.; Esmaeili Mahmoudabadi, A.; Ekradi, M.H.; Mbuvha, R.; Hosseini, S.M. An Interpretable Machine Learning Framework for Unraveling the Dynamics of Surface Soil Moisture Drivers. *Remote Sens.* **2025**, *17*, 2505. [CrossRef]
26. Shampa; Nasir, N.N.; Winey, M.M.; Dey, S.; Zahid, S.M.T.; Tasnim, Z.; Islam, A.K.M.S.; Hussain, M.A.; Hossain, M.P.; Muktadir, H.M. Integration of Remote Sensing and Machine Learning Approaches for Operational Flood Monitoring Along the Coastlines of Bangladesh Under Extreme Weather Events. *Water* **2025**, *17*, 2189. [CrossRef]
27. Li, M.; Zhu, Z.; Deng, J.; Zhang, J.; Li, Y. Geospatial Explainable AI Uncovers Eco-Environmental Effects and Its Driving Mechanisms—Evidence from the Poyang Lake Region, China. *Land* **2025**, *14*, 1361. [CrossRef]
28. Wang, J.; Sanderson, J.; Iqbal, S.; Woo, W.L. Accelerated and Interpretable Flood Susceptibility Mapping Through Explainable Deep Learning with Hydrological Prior Knowledge. *Remote Sens.* **2025**, *17*, 1540. [CrossRef]
29. Liao, Y.; Miao, S.; Fan, W.; Liu, X. A Novel Hybrid Fuzzy Comprehensive Evaluation and Machine Learning Framework for Solar PV Suitability Mapping in China. *Remote Sens.* **2025**, *17*, 2070. [CrossRef]
30. Xiong, Y.; Li, C.; Zou, M.; Xu, Q. Investigating into the Coupling and Coordination Relationship between Urban Resilience and Urbanization: A Case Study of Hunan Province, China. *Sustainability* **2022**, *14*, 5889. [CrossRef]
31. Wei, F.; Zhao, L. The Effect of Flood Risk on Residential Land Prices. *Land* **2022**, *11*, 1612. [CrossRef]
32. Geng, J.; Yu, K.; Xie, Z.; Zhao, G.; Ai, J.; Yang, L.; Yang, H.; Liu, J. Analysis of Spatiotemporal Variation and Drivers of Ecological Quality in Fuzhou Based on RSEI. *Remote Sens.* **2022**, *14*, 4900. [CrossRef]
33. Li, Z. GeoShapley: A Game Theory Approach to Measuring Spatial Effects in Machine Learning Models. *Ann. Am. Assoc. Geogr.* **2024**, *114*, 1365–1385. [CrossRef]
34. Wu, R.; Yu, G.; Cao, Y. The impact of industrial structural transformation in the Yangtze River economic belt on the trade-offs and synergies between urbanization and carbon balance. *Ecol. Indic.* **2025**, *171*, 113165. [CrossRef]
35. Lyu, R.; Pang, J.; Tian, X.; Zhao, W.; Zhang, J. How to optimize the 2D/3D urban thermal environment: Insights derived from UAV LiDAR/multispectral data and multi-source remote sensing data. *Sustain. Cities Soc.* **2023**, *88*, 104287. [CrossRef]
36. Liu, F.; Liu, X.; Xu, T.; Yang, G.; Zhao, Y. Driving Factors and Risk Assessment of Rainstorm Waterlogging in Urban Agglomeration Areas: A Case Study of the Guangdong-Hong Kong-Macao Greater Bay Area, China. *Water* **2021**, *13*, 770. [CrossRef]
37. Sugianto, S.; Deli, A.; Miswar, E.; Rusdi, M.; Irham, M. The Effect of Land Use and Land Cover Changes on Flood Occurrence in Teunom Watershed, Aceh Jaya. *Land* **2022**, *11*, 1271. [CrossRef]
38. Islam, T.; Zeleke, E.B.; Afroz, M.; Melesse, A.M. A Systematic Review of Urban Flood Susceptibility Mapping: Remote Sensing, Machine Learning, and Other Modeling Approaches. *Remote Sens.* **2025**, *17*, 524. [CrossRef]
39. Chen, Y.; Ye, Y.; Liu, X.; Yin, C.; Jones, C.A. Examining the nonlinear and spatial heterogeneity of housing prices in urban Beijing: An application of GeoShapley. *Habitat Int.* **2025**, *162*, 103439. [CrossRef]
40. Xiao, L.; Wu, M.; Weng, Q.; Li, Y. Unraveling Street Configuration Impacts on Urban Vibrancy: A GeoXAI Approach. *Land* **2025**, *14*, 1422. [CrossRef]
41. Chen, J.; Huang, G.; Chen, W. Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning models. *J. Environ. Manag.* **2021**, *293*, 112810. [CrossRef] [PubMed]
42. Du, W.; Gong, Y.; Chen, N. PSO-WELLSVM: An integrated method and its application in urban waterlogging susceptibility assessment in the central Wuhan, China. *Comput. Geosci.* **2022**, *161*, 105079. [CrossRef]

43. Chen, Z.; Nong, X.; Zang, C.; Ou, W.; Qiu, L. Evolution of evapotranspiration in the context of land cover/climate change in the Han River catchment of China. *Hydrol. Process.* **2024**, *38*, e15265. [CrossRef]
44. Xiong, Y.; Zhang, F. Effect of human settlements on urban thermal environment and factor analysis based on multi-source data: A case study of Changsha city. *J. Geogr. Sci.* **2021**, *31*, 819–838. [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.