sustainability

MDPI

Article

Risk Assessment of Debris Flow Disasters Triggered by an Outburst of Huokou Lake in Antu County Based on an Information Quantity and Random Forest Approach

Qiuling Lang [1,*], Peng Liu [2], Yichen Zhang [1], Jiquan Zhang [3] and Jintao Huang [2]

[1] School of Jilin Emergency Management, Changchun Institute of Technology, Changchun 130012, China; zhangyc@ccit.edu.cn
[2] School of Survey and Surveying Engineering, Changchun Institute of Technology, Changchun 130012, China; liupeng@stu.ccit.edu.cn (P.L.); huangjintao@stu.ccit.edu.cn (J.H.)
[3] School of Environment, Northeast Normal University, Changchun 130117, China; zhangjq022@nenu.edu.cn
* Correspondence: 0215046@ccit.edu.cn; Tel.: +86-151-0441-7089

Abstract: Debris flow disasters frequently occur and pose considerable hazards; thus, it is essential to thoroughly evaluate their risks. This study constructs a database comprising 20 assessment indicators, utilizing comprehensive natural disaster risk assessment theory and incorporating the triggering factors of Huokou Lake in the Changbaishan Mountains. This research employs a hybrid ANP-CRITIC methodology to allocate weights to the assessment indicators efficiently. For hazard assessment, this research utilizes both the Information Quantity and Random Forest models for comparative analysis. The ROC curve was employed to validate the outcomes, ultimately favoring the Random Forest model due to its superior accuracy in assessing debris flow hazards. In this study, the risk of debris flow disasters in Antu County is comparatively assessed under scenarios with and without an outburst event. The findings indicate that areas of high and very high risk are predominantly located within the central regions of economically prosperous and densely populated townships. Additionally, the risk in Erdao Baihe Township escalates significantly when considering the outburst of Huokou Lake. The significance of this study lies in its ability to furnish a robust scientific basis for decision-makers aimed at preventing future debris flow disasters. Furthermore, it serves as a crucial reference for advancing sustainable regional development and facilitates the equilibrium between economic growth and environmental protection within disaster management strategies.

Keywords: information quantity; random forest; debris flow; risk assessment; Mike21; ANP; CRITIC; sustainability

check for updates

Citation: Lang, Q.; Liu, P.; Zhang, Y.; Zhang, J.; Huang, J. Risk Assessment of Debris Flow Disasters Triggered by an Outburst of Huokou Lake in Antu County Based on an Information Quantity and Random Forest Approach. *Sustainability* **2024**, *16*, 9545. https://doi.org/10.3390/su16219545

Academic Editors: Danqing Song and Zhuo Chen

Received: 6 September 2024
Revised: 17 October 2024
Accepted: 29 October 2024
Published: 1 November 2024

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

1. Introduction

Debris flows have historically constituted a prevalent natural disaster in mountainous regions globally [1]. Such events can inflict substantial damage and losses, affecting not only infrastructure, including residential and industrial buildings, but also human lives and livestock through burial and land degradation. Additionally, they impact various sectors such as water conservancy and hydropower, transportation and communication networks, and the broader economy [2].

Research on the evaluation of debris flow disaster risk commenced relatively early on an international scale. Nevertheless, due to the ambiguous definition of disaster risk prevalent at that time, numerous studies failed to gain widespread recognition and dissemination. For instance, a risk evaluation encompassing ten natural hazards, including earthquakes and landslides in California, was carried out in the United States between 1970 and 1973 [3,4]. It was only in 1982, when the United Nations Disaster Relief Office introduced a more widely accepted definition of disaster risk, that notable international geohazard researchers, such as Carrara [5], Varnes [6], and Brabb [7], began to undertake

studies to establish preliminary concepts and propose rudimentary methods. However, the absence of standardized technical methodologies and processes has constrained the practical promotion and application of related risk evaluation studies. Research on debris flow disaster risk in China commenced comparatively late. Despite the emergence of some studies on this subject since the 1980s, a more comprehensive and mature system has yet to be established. It was not until the 1990s that a significant number of Chinese scholars began to systematically investigate debris flow disaster risk, coinciding with the gradual maturation of international research on debris flow risk assessment [8].

In recent years, the rapid advancements in computer and information technology, coupled with the proliferation of remote sensing technologies, software, models, and methodologies, have significantly advanced research in debris flow disaster risk evaluation. In 2016, Korean researchers Kim BJ et al. developed a real-time disaster management Geographic Information System (GIS) utilizing GIS data [9]. American researchers Meehan Christopher L et al. examined debris flows resulting from landslides and surface erosion, simulating the dynamic process of material flow [10]. Their findings indicated that the severity of debris flow hazards increased markedly with the intensity of heavy rainfall, presenting a substantial threat to society. In 2017, Pablo Tierz et al. employed a Bayesian Information Network to assess rainstorm-induced debris flows in Somma-Vesuvius, Italy, focusing on probability analysis and risk assessment [11]. In 2018, Han Xu selected Dongchuan District as the study area to analyze debris flow risk and developed an evaluation system integrating GIS and remote sensing technologies [12]. In 2021, Jiang Tao-tao applied the risk = hazard $\times$ vulnerability model to analyze debris flow risk in Fangshan District, Beijing, China [13]. In 2023, Wu Qingan et al. identified ten indicators as factors for debris flow risk evaluation, considering the formation conditions and characteristics of debris flows in Zhouqu County [14]. By incorporating results from a questionnaire survey, they developed a debris flow risk evaluation model using the gray correlation degree method. They also established a risk degree model based on the product of hazard and vulnerability degrees, deriving risk assessments for each debris flow.

Collectively, the aforementioned studies have established a foundational basis for debris flow disaster risk evaluation; however, several limitations persist in this domain. For instance, the subjective assignment method is inherently influenced by the uncertainty surrounding the decision-maker's knowledge base, cognitive reasoning, and the interrelationships among various qualitative and quantitative indicators, making it challenging to accurately quantify the relative importance of these indicators. Conversely, the objective assignment method determines weights solely based on the objective information of the indicators, relying on robust mathematical theories; however, it fails to account for the subjective intentions of the decision-maker and analysts. Furthermore, the results derived from this method may occasionally diverge from the true significance of the indicators, which can be challenging to rationalize [15]. Through comparative analysis of the assessment accuracy of various machine learning models, scholars generally concur that the random forest (RF) model exhibits superior predictive accuracy and enhanced modeling adaptability, rendering it an accurate and reliable forecasting tool [16]. Additionally, traditional risk assessment primarily focuses on disaster-causing conditions, which are deemed to arise from the interplay of three factors: hazard, exposure, and vulnerability. This approach often overlooks the inherent complexity of the system affected by debris flow disasters, as well as the influence of community preparedness on risk mitigation efforts. To address these limitations, this paper pioneers the application of the comprehensive risk assessment theory of natural hazards as proposed by Zhang et al. (2009) [17]. By integrating the network analytic hierarchy process (ANP) and the criteria importance though intercriteria correlation (CRITIC) methods, it analyzes debris flow disaster risk in Antu County from both subjective and objective perspectives. This theory extends the traditional risk formation principle from three to four elements: hazard, exposure, vulnerability, and emergency response and recovery capacity (ERRC), which is the basic condition and special defense capacity that regional human societies possess to guarantee that disaster-bearing bodies are

free from, or are less threatened by, natural hazards, and is adopted in this study to ensure that the assessment results are more scientific and standardized. For hazard assessment, this study employs both the Information Quantity Model and the Random Forest Model for comparative analysis, thereby enhancing the accuracy of the hazard assessment results. Finally, the visualization of risk using GIS, combined with twenty assessment indicators and up-to-date data, offers a valuable tool for decision-makers to evaluate debris flow disaster risk in Antu County in a comprehensive, scientific, standardized, and user-friendly manner.

2. Overview of the Study Area

Antu County is situated in the northern foothills of the Changbai Mountains, characterized by rolling hills, gullies, and ravines. The Changbai Mountain Range extends from south to north, resulting in a terrain that features high elevations in the south and low elevations in the north, as well as high elevations in the east and low elevations in the west, forming a north–south orientation and a narrow east–west expanse. The region encompasses extensive mountainous areas and numerous rivers, presenting a continental low mountainous landscape. The northern portion of the county consists of low mountains and hills; the central region features mid-mountain and low-mountain areas; and the southern section is characterized by plateaus and tablelands. Influenced by the topography and the interaction of two types of airflow from the continent and the Pacific Ocean, vertical zoning is pronounced in this region. Both temperature and precipitation are significantly affected by altitude; specifically, the temperature decreases by 0.5–0.6 °C for every 100 m in elevation, while annual precipitation increases by 33 mm. Hengdan, located in the central northern part of the Arakou Ridge, serves as a natural geographical demarcation line for the county, distinctly dividing it into two climate zones. The southern region, encompassing the Tianchi area, exhibits a continental climate typical of the middle temperate zone, with an average annual precipitation of 1407.6 mm, primarily concentrated from June to August, and an average annual temperature of −7.3 °C. Conversely, the northern region falls within the northern temperate monsoon climate zone, with an average annual precipitation of 632 mm, also concentrated from July to August, and an average annual temperature of 2.9 °C (Figure 1).

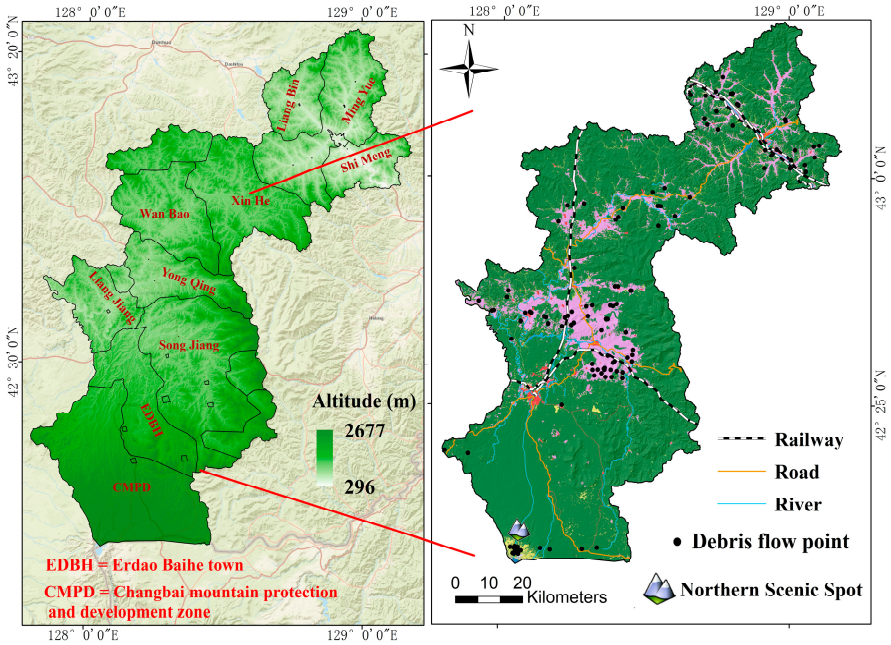


Figure 1. Location and overview of the study area.

3. Data and Methods

3.1. Data Sources and Preparation Techniques

This study is founded on the Geological Hazard Investigation and Regionalization Project for the Changbai Mountain Protected Development Zone in Jilin Province, wherein remote sensing interpretation serves as the preliminary approach, complemented by ground investigations and necessary mapping, physical exploration, and mountain engineering techniques (Figure 2). The data utilized in this study primarily consist of vector and raster datasets encompassing remote sensing, meteorological information, basic geographic data, and attributes pertaining to population and economic factors. The specific sources of all datasets are detailed in Table 1. Data processing was conducted using ArcGIS 10.8.1 software. Slope and topographic relief were directly extracted using ArcGIS software, while the topographic wetness index (TWI) and stream power index (SPI) were derived using the raster calculator; distance from the water system (DFWS) was calculated using Euclidean distance. Vegetation cover was obtained from Landsat 8 remote sensing imagery with a resolution of 30 m, utilizing remote sensing image processing software ENVI 5.6 to generate the NDVI. Appropriate vegetation classification standards were then applied within ArcGIS software to derive the vegetation cover. Building density was extracted from points of interest in the planning cloud for point extraction and subsequently processed using kernel density estimation, while road density was derived through geometric calculations based on the attribute table. The remaining metrics were acquired and imported into ArcGIS software for direct cropping within the study area, with all aforementioned data categorized into intervals using the natural breakpoint method. Figure 3 illustrates the technical procedures employed in this study.



Figure 2. (**a**) North scenic debris flow No. 1 ditch; (**b**) north scenic debris flow No. 2 ditch; (**c**) remote sensing image of debris flow.

Table 1. Data types and sources.

Parameters	Data Types	Source
Debris flow point	Vector data	Geological Hazard Investigation and Regionalization Project for the Changbai Mountain Protected Development Zone in Jilin Province
Zoning	Vector data	https://www.webmap.cn (accessed on 1 March 2024)
DEM	Raster data	http://www.gscloud.cn (accessed on 6 March 2024)
Slope	Raster data	ASTGTM2 DEM30m GIS analysis
Topographic relief	Raster data	ASTGTM2 DEM30m GIS analysis
TWI	Raster data	ASTGTM2 DEM30m GIS analysis
SPI	Raster data	ASTGTM2 DEM30m GIS analysis
Average annual rainfall	Raster data	https://www.resdc.cn/ (accessed on 15 March 2024)
Distance from water system	Vector data	https://www.resdc.cn/ (accessed on 15 March 2024)
Vegetation coverage	Raster data	https://www.databox.store (accessed on 5 March 2024)
Lithology	Vector data	http://dc.ngac.org.cn/Distribute (accessed on 8 March 2024)
Population density	Raster data	World UN population density dataset https://population.un.org/wpp/ (accessed on 11 March 2024)

Table 1. *Cont.*

Parameters	Data Types	Source
Road density	Vector data	https://www.gscloud.cn (accessed on 6 March 2024)
Building density	POI	http://www.guihuayun.com/ (accessed on 20 March 2024)
Land use	Raster data	https://www.esa.int (accessed on 20 March 2024)
Road category	Vector data	https://www.gscloud.cn (accessed on 20 March 2024)
Building stories	POI	http://www.guihuayun.com/ (accessed on 20 March 2024)
POVP	Raster data	Antu County Statistical Yearbook http://www.antu.gov.cn/sj_2368/tjgb/index.html (accessed on 28 March 2024)
GDP	Raster data	https://www.databox.store (accessed on 12 March 2024)
NOBIHW	Raster data	Antu County Statistical Yearbook
NOHCW	Raster data	Antu County Statistical Yearbook
Education status	Raster data	Antu County Statistical Yearbook

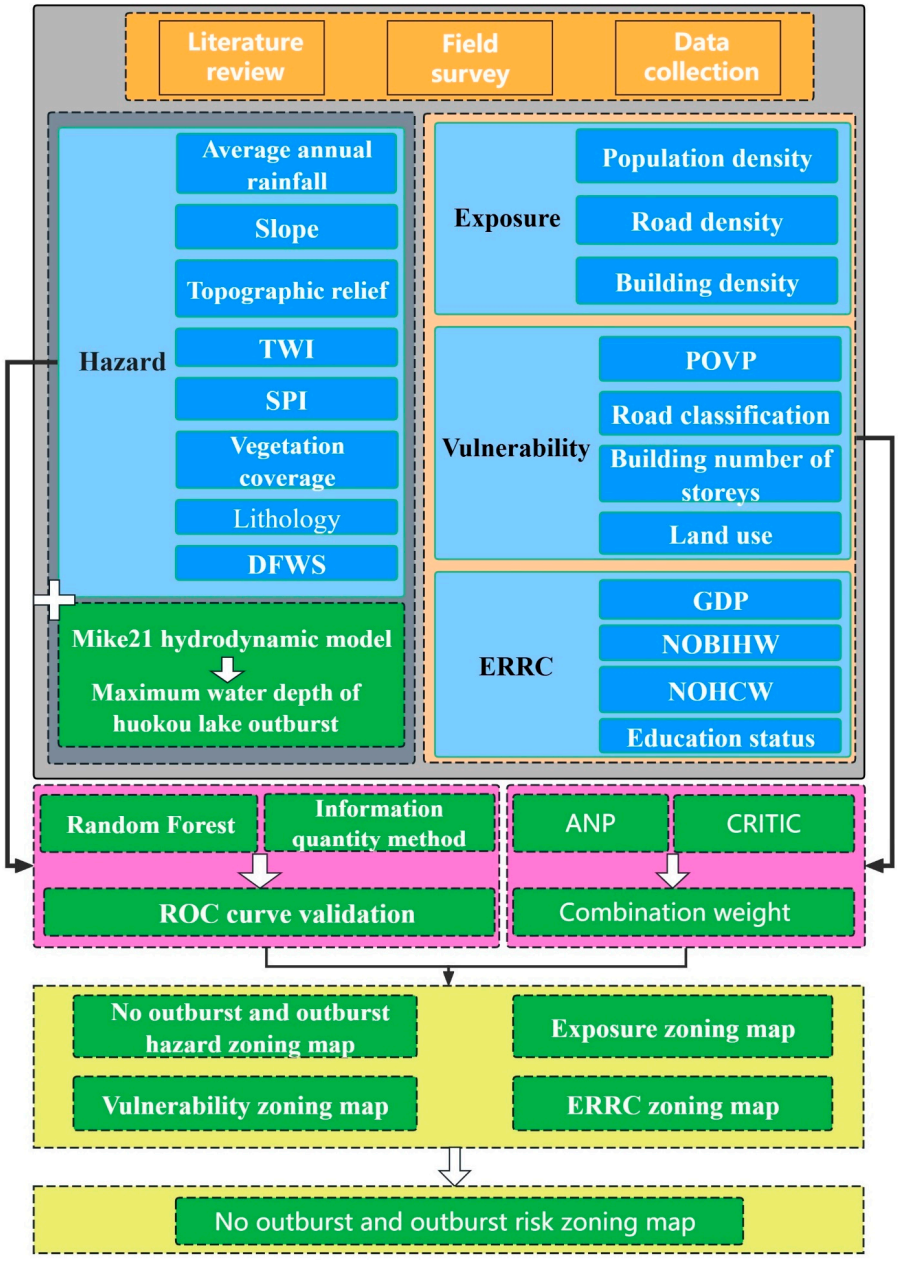


Figure 3. Evaluation flow chart.

3.2. Calculation of Weights

3.2.1. Calculation of Subjective Weight by ANP

The ANP method is an adaptive non-independent hierarchical decision-making method developed on the basis of the analytic hierarchy process (AHP) method. Compared with the AHP, which only emphasizes the unidirectional hierarchical relationship between the decision-making levels, the ANP takes more account of the interactions between the factors or the adjacent levels, and solves the interactions, dependencies, and feedbacks of the factors, so as to make the results more accurately and efficiently described, and it is a kind of subjective empowerment method [18].

(1) Constructing the Network Structure:

In accordance with the principles of the ANP and the evaluation framework, an ANP network evaluation model is developed. This model employs double arrows to represent mutual influences between elements, while single arrows indicate the direction of influence generation (Figure 4).

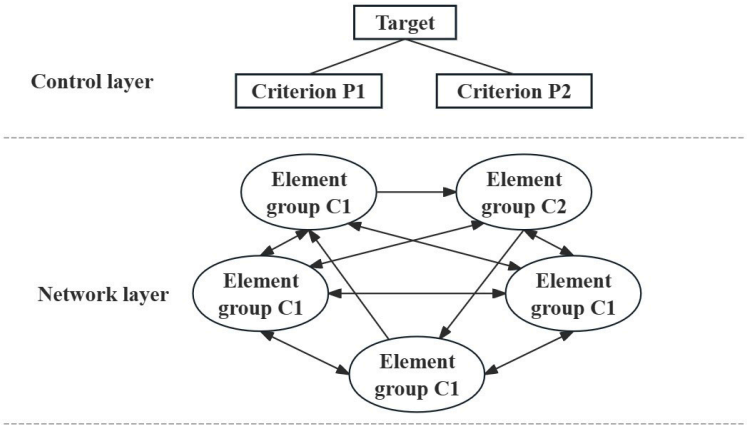


Figure 4. The structure model of ANP.

(2) Pairwise comparisons are conducted based on the interactions among elements within the network model. These comparisons utilize the Saaty 1–9 scale (Table 2) and are derived from expert questionnaires. The expert ratings are then aggregated using the geometric mean method.

Table 2. Saaty 1–9 scale.

Meaning	Relative Importance
A is as important as B	1
A is slightly more important than B	3
A is obviously more important than B	5
A is more important than B	7
A is absolutely more important than B	9
The middle value of the above importance	2, 4, 6, 8
The opposite expression of the above degree of importance	1/3, 1/5, 1/7, 1/9

(3) The unweighted supermatrix is computed, followed by a consistency test. Initially, within the control criterion layer, indirect dominance comparisons among elements are conducted, and weight vectors are determined using the eigenvalue method. Consistency testing is applied to each local weight matrix to ensure accuracy. Subsequently, the internal and external relationships among elements in other groups are sequentially compared to derive the unweighted supermatrix W, which is constructed from the ranking vectors of interacting elements within the network layer.

$$W = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1N} \\ W_{21} & W_{22} & \cdots & W_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ W_{N1} & W_{N2} & \cdots & W_{NN} \end{bmatrix} \quad (1)$$

(4) Each sub-block of the unweighted supermatrix only considers the ordering of the elements within the subsystem (i.e., within the hierarchy) concerning a given criterion and does not estimate the influence of the other subsystems on this criterion, so each column of the unweighted supermatrix is not normalized. For this reason, with each subsystem (or hierarchy) as an element, a two-by-two comparison for a given subsystem yields a weighted matrix A, which weights the elements of W to yield a weighted supermatrix $\overline{W}$.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \quad (2)$$

$$\overline{W} = \begin{bmatrix} a_{11} \times w_{11} & a_{12} \times w_{12} & \cdots & a_{1N} \times w_{1N} \\ a_{21} \times w_{21} & a_{22} \times w_{22} & \cdots & a_{2N} \times w_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ a_{N1} \times w_{N1} & a_{N2} \times w_{N2} & \cdots & a_{NN} \times w_{NN} \end{bmatrix} \quad (3)$$

(5) The limit supermatrix is computed to ascertain the weights of the elements. To account for interdependencies among the elements, the weighted supermatrix is stabilized through iterative self-multiplication until convergence is achieved. This process yields the final weight values for the elements.

$$W^{\infty} = \lim_{k \to \infty} (W)^k \quad (4)$$

3.2.2. Calculation of Objective Weight by CRITIC

The CRITIC method is a weighting method based on the comparative strength and conflict of assessment indicators. Comparative intensity refers to the size of the difference between the values of each assessment program for the same indicator, which is expressed in the form of standard deviation, while the conflict between indicators is expressed by the correlation coefficient. The larger the standard deviation, the greater the fluctuation and the higher the weight; if the correlation coefficient is larger, the smaller the conflict and the lower the weight. This method can fully combine the volatility, correlation, and other characteristics of the data, and the resulting weights are objective and effective, avoiding to a certain extent the influence of the extreme value of the weights of a single piece of data on the whole, and it is an objective assessment method [19]. The fundamental steps are as follows:

(1) Let a total of m evaluation objects and n evaluation indicators be selected, and the original data matrix X is obtained.

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} \quad (5)$$

(2) In order to eliminate the influence of the scale on the evaluation results, the indicators need to be standardized.
Positive indicators are as follows:

$$r_{ij}^{+} = \frac{x_{ij} - x_{\min j}}{x_{\max j} - x_{\min j}} \tag{6}$$

Negative indicators are as follows:

$$r_{ij}^{-} = \frac{x_{\max j} - x_{ij}}{x_{\max j} - x_{\min j}} \tag{7}$$

where x_{ij} is the *jth* indicator for the *ith* object ($0 < i \leq m$, $0 < j \leq n$).

(3) Calculate the variability within each metric.

$$S_j = \sqrt{\frac{\sum_{i=1}^{m} \left(X_{ij} - \overline{X}_j\right)^2}{n-1}} \tag{8}$$

(4) Calculate the conflicting nature of the indicators.

$$R_j = \sum_{i=1}^{m} (1 - r_{ij}) \tag{9}$$

(5) Calculate the amount of information for each indicator.

$$C_j = S_j \sum_{i=1}^{m} (1 - r_{ij}) = S_j \times R_j \tag{10}$$

(6) Calculate the weight of each indicator.

$$w_j = \frac{C_j}{\sum_{j=1}^{n} C_j} \tag{11}$$

3.2.3. Calculation of Combination Weight by Game Theory

After using the ANP method to calculate subjective weights and the CRITIC method to calculate objective weights, this study used game theory to optimize the combination of the results of the two methods. The game theory combination assignment takes the Nash equilibrium as the goal, absorbs the multi-class weight factors, and seeks the optimal combination that takes into account the multi-class weights [20]. The calculation steps are as follows:

(1) Assuming that L methods are used to calculate the weights, any linear combination of these L weights is calculated as follows:

$$W^{*} = \sum_{k=1}^{L} \beta_k W^{(k)T} \tag{12}$$

where β_k is the weighting factor, and the following applies:

$$\sum_{k=1}^{L} \beta_k = 1, \beta_k \rangle 0 \tag{13}$$

(2) The calculation of weighting factors is as follows:

$$\min_i \left\| \sum_{k=1}^{L} \beta_k W^{(k)T} - \omega_i \right\|_2 \tag{14}$$

where ω_i is the weight calculated by the *ith* method. According to the differential properties of the matrix, Formula (14) is expanded as follows:

$$\begin{bmatrix} \omega_1\omega_1^T & \omega_1\omega_2^T & \cdots & \omega_1\omega_L^T \\ \omega_2\omega_1^T & \omega_2\omega_2^T & \cdots & \omega_2\omega_L^T \\ \vdots & \vdots & \ddots & \vdots \\ \omega_L\omega_1^T & \omega_L\omega_2^T & \cdots & \omega_L\omega_L^T \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \omega_1\omega_1^T \\ \omega_2\omega_2^T \\ \vdots \\ \omega_L\omega_L^T \end{bmatrix} \tag{15}$$

(3) Calculate the weight coefficient β_k according to Equation (15) and normalize it into Equation (12) to obtain the optimal weights.

3.3. Information Quantity

The theoretical foundation of the Information Quantity model is grounded in information theory [21], which employs techniques from probability theory and mathematical statistics to examine the debris flow hazard model. This model incorporates various assessment indicators through the lens of information entropy.

By conducting a spatial superposition analysis of various assessment indicators and debris flow hazard points, and utilizing a GIS platform to compute the Information Quantity for each individual indicator, followed by the weighted integration of multiple indicators to derive the comprehensive Information Quantity, the model for debris flow hazard assessment is established.

$$I = \sum_{j=1}^{n} \ln \frac{N_j/N}{S_j/S} \tag{16}$$

where I is the total information weighted by various evaluation indicators, which can be used as the debris flow hazard index; Nj is the number of debris flows contained within a specific grading interval of a single evaluation indicator; N is the total number of debris flows; Sj is the number of rasters within the specific grading interval of a single evaluation indicator; and S is the total number of rasters.

3.4. Random Forest

In the early 1990s, mathematician Leo Breiman [22] introduced the Random Forest algorithm, an ensemble learning method based on Bagging. This algorithm swiftly established itself as a prominent example in the field of ensemble learning and garnered significant attention within the data science community.

The fundamental principle of the Random Forest algorithm involves employing bootstrap sampling (self-sampling method), where K samples are randomly drawn from the original training dataset. For each sample, M features are selected at random for model construction, resulting in K distinct decision trees. The final classification outcome is determined through majority voting across these trees. This ensemble approach effectively leverages the strengths of multiple decision trees, thereby enhancing classification accuracy.

Non-debris flow hazard sites, matched in number to historical debris flow hazard sites, were randomly selected as sample data within the study area. Utilizing SPSSPRO 1.1.28 software, the sample data were allocated in a 7:3 ratio and input into the random forest algorithm, with 70% designated for model training and 30% for model testing. Evaluation metrics, including accuracy, precision, and recall, were employed to assess the predictive accuracy of the model on unknown data. Based on the specific context, relevant parameters within the random forest were adjusted to optimize the model's performance. Following testing, the model employed in this study was configured with 100 decision

trees, a maximum of 1000 nodes per decision tree, a maximum depth of 10, and a minimum sub-node size of 1.

3.5. ROC Curves and Sensitivity Analysis

The ROC curve is a graphical representation used to assess the performance of a classification model [23]. The x-axis represents the false positive rate (1-specificity), while the y-axis denotes the true positive rate (sensitivity). The ROC curve effectively illustrates the trade-off between specificity and sensitivity. The area under the ROC curve, denoted as the AUC value, ranges from 0 to 1. When the AUC value falls within the range of 0.5 to 0.6, the discriminatory power is considered inadequate. When the AUC value is situated between 0.6 and 0.7, the discriminatory efficacy is deemed poor. When the AUC value lies between 0.7 and 0.8, the discriminatory effectiveness is categorized as fair. When the AUC value ranges from 0.8 to 0.9, the discriminatory performance is regarded as better. When the AUC value reaches between 0.9 and 1, the discriminatory effect is classified as excellent.

The literature review employed the area under the curve (AUC) to quantify the success rate of hazard prediction. In this approach, the horizontal axis represents the cumulative percentage of hazard zones, while the vertical axis represents the cumulative percentage of debris flow hazard points. The AUC is used as a measure of the predictive success rate.

Sensitivity analysis is a quantitative analytical method that elucidates the degree of influence exerted by input variables on output variables [24]. This methodology is widely applied across various domains, including economics, ecology, engineering, and chemistry. The primary objective of this paper is to ascertain the sensitivity coefficient by systematically removing an influencing factor from the model, thereby evaluating the impact of each factor on the model's assessment results in relation to this sensitivity coefficient. A sensitivity value greater than 0 indicates that the factor positively influences the model's assessment results; furthermore, a larger sensitivity value correlates with a more significant impact on these results. The sensitivity value is calculated using the following formula:

$$RD_i = \frac{AUC - AUC_i}{AUC} \times 100\% \tag{17}$$

where RD_i is the sensitivity coefficient value after removing the i influence factor; AUC is the AUC value obtained from the test based on the premise of all the factors involved in the assessment of the model; AUC_i is the AUC value obtained from the test after removing the i influence factor.

3.6. Model Construction

Taking into account the characteristics of Huokou Lake, this study draws upon concepts presented in the works of Bin Zhou [25], Rayees Ahme [26], and Tandin Wangchuk [27], which form a well-established international framework, to model the potential outburst of Huokou Lake through a series of defined steps.

3.6.1. Modeling of the Tianchi Outburst

Changbaishan Tianchi is a Huokou lake with a water surface elevation of 2189.7 m, a lake area of 9.4 km^2, an average depth of the lake of 204 m, and a maximum depth of 373 m. The topographic conditions surrounding Tianchi dictate that water can only be discharged through the northern side of the lake. Consequently, this discharge can be conceptualized as a dam breach. Utilizing the model presented below allows for effective simplification of this issue, rendering the discharge process from Tianchi more predictable.

Consider Tianchi as an inverted circular frustum, where the radius of the upper base exceeds that of the lower base (Figure 5) [28]. Let R_3 denote the radius of the lake bottom, R the radius of the water surface after a height increase of ΔH, H the original water depth, and ΔV the volume of the water body added due to the height increase ΔH. This volume

ΔV can be determined using the formula for the similarity of triangles and the volume of a circular frustum:

$$\Delta V = \frac{1}{3}\Delta H\pi\left(R^2 + R \cdot R_2 + R_2^2\right) \tag{18}$$

$$\frac{R - R_2}{R_2 - R_3} = \frac{\Delta H + H}{H} \tag{19}$$

In this paper, the volume of the overflowed water body is 0.5 km^3, and the data related to the mouth of Tianchi are brought in to find the lake surface elevation of 49 m.

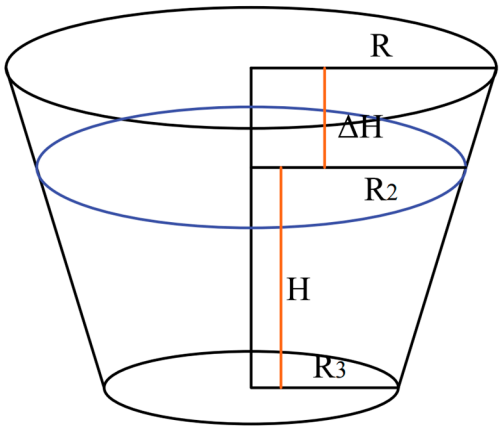


Figure 5. Tianchi round table model.

3.6.2. Mike21 Model

In this study, the MIKE21 hydrodynamic model is employed to simulate the maximum water depth resulting from the outburst of Huokou Lake. Since its inception, MIKE 2014 software has consistently outperformed other hydrodynamic models due to its precision in simulation, reliability in boundary condition input, user-friendly operability, efficient processing capabilities, and streamlined interface [29].

(1) Flow process line design

For the relationship between the size of the parameter α, which is the ratio of the water depth H_0 upstream of the dam site to the water depth H_2 downstream of the dam site before dam failure, and a certain critical value α_c, the dam failure flow can be categorized into three scenarios: continuous wave flow $\alpha < \alpha_c$; critical flow $\alpha = \alpha_c$; and discontinuous wave flow $\alpha > \alpha_c$. (Values of α_c can be found in Table 3)

Table 3. The value of flow parameters under different working conditions.

River Valley Section Form	m	αc	λ	Maximum Flow Calculation Formula
Rectangle	1.0	0.138	0.298	$\frac{8}{27}\sqrt{g}BH_0^{3/2}$
Quartic parabola	1.25	0.205	0.220	$0.220\sqrt{g}BH_0^{3/2}$
Quadratic parabola	1.5	0.272	0.172	$0.172\sqrt{g}BH_0^{3/2}$
Triangle	2.0	0.373	0.116	$0.116\sqrt{g}BH_0^{3/2}$

The uniform formula for maximum flow at the dam site can be found using the instantaneous total failure of Hsieh Renzhi [30]:

$$Q_m = \lambda B_0\sqrt{g}H_0^{3/2} \tag{20}$$

$$\lambda = m^{m-1}\left[\frac{2\sqrt{m} + \frac{u_0}{\sqrt{gH_0}}}{1 + 2m}\right]^{2m+1} \tag{21}$$

where Q_m is the peak flow; B_0 is the width of the water surface; g is the acceleration of gravity; H_0 is the upstream water depth; λ is the flow coefficient. The formula for solving λ is different for different flow regimes; see Table 3. In this paper, $\alpha = 0.5/204 = 0.0024 < \alpha c = 0.272$; therefore, continuous wave flow analysis is chosen.

$$Q_m = 0.172\sqrt{g}BH_0^{3/2} = 44000\ \mathrm{m}^3/s \tag{22}$$

The generalized estimation method is employed to determine the flow process line. This method utilizes the results obtained from a detailed algorithm and model analysis, where the maximum flow rate derived from the flow process line is combined with the reservoir capacity to compute the total duration of the flow process. The flow process line is then approximated as a quadratic parabola or a 2.5-times parabola [31]. The total duration of the dam failure is initially calculated based on the maximum flow rate, Q_m, observed at the dam failure site, and the total volume of water, W, associated with the failure.

$$T = K\frac{W}{Q_m} \tag{23}$$

where K is the coefficient, with 4 times and 2.5 times the parabolic corresponding to K; take the value of 4–5, 3.5. Then, combined with the existing generalized estimation model table to determine the flow process line, this paper selects the 2.5 times parabolic model, and its calculation table is shown in Table 4.

Table 4. Generalized estimated model: 2.5th parabola.

t/T	0	0.01	0.1	0.2	0.3	0.4	0.5	0.65	1.0
Q/Qm	Q_0/Q_m	1.0	0.62	0.45	0.36	0.29	0.23	0.15	Q_0/Q_m

Finally, verify whether the water discharge between the process curve and the $Q = Q_0$ straight line matches the dam-break reservoir capacity or falls within the acceptable error margin. If so, the calculation can be terminated. This study will apply the relevant data to resolve the flow process line, with the details provided in Table 5.

Table 5. Flow process calculation table.

0	397.73	3977.3	7954.6	11,931.9
7.77×10^{-5}	44,000	27,280	19,800	15,840
15,909.2	19,886.5	25,852.45	39,773	
12,760	10,120	6600	7.77×10^{-5}	

(2) Modeling of Flood Evolution

Import the corrected coordinate data in the MIKE21 Grid Generator, generate the reservoir boundary and terrain file, and determine the simulation area. The parameter settings are shown in Table 6.

Table 6. Calculation parameters of flood routing model.

Parameter	Numerical Value of Outburst Flood of Huokou Lake	Meaning
Module	Hydrodynamics	Model module
Bathymetry	Based on the project's existing topographic information, GPS, and other information of the actual measurement information, use the ArcGIS software to fill the topographic file for the water body, turn the point, and generate the XYZ file; the MIKE conversion work will be converted from the topography to the MIKE mesh terrain files.	Model calculation area setting

Table 6. *Cont.*

Parameter	Numerical Value of Outburst Flood of Huokou Lake	Meaning
Simulation time	5 April 2023 0:00:00–5 April 2023 8:04:00	Simulation time
Time step	30 s	Time step
No. of time step	880 step	Calculation steps
Enable flood and drying	Drying depth 0.01 m; flooding depth 0.05 m	Water depth h < 0.01 m; do not participate in the hydrodynamic calculation; when 0.01 m < h < 0.05 m, the flow velocity at the grid point is zero; when h > 0.05 m, regional grid points participate in hydrodynamic calculation.
Resistance	Roughness is a key parameter in hydraulic calculations, reflecting the combined coefficients of the regional surface, channel bottom, and bank slopes that affect water resistance. In the two-dimensional surface flow model, the roughness is expressed as Manning's value (unit $m^{1/3}/s$), which is subdivided into various types of subsurface according to highway, farmland, green land, building land, village houses, woods, water surface, etc., and given different roughness values, respectively. The Manning diagram in the MIKE21 model was made according to the land use type map (Figure 6). Firstly, the Manning value M is added and assigned to different land types in ArcGIS software; the Manning value attribute is selected as a conversion field to convert the land use type map into a roughness raster; the Manning raster is exported as an .asc file, and finally the .asc file is converted into a .dfs2 file by using the Grd2Mike module in the ToolBox of the MIKE software.	The comprehensive coefficient of water resistance affected by regional surface, river bottom, and bank slope

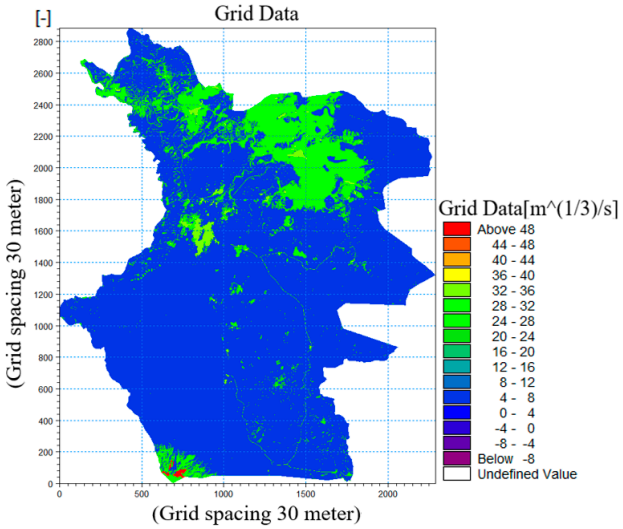


Figure 6. Simulation area roughness Manning diagram.

3.6.3. Risk Evaluation Model

In this study, the natural disaster risk index method and the weighted comprehensive assessment method were employed to develop a debris flow disaster risk assessment model, aimed at quantifying the degree of debris flow risk [32,33]. The specific formulas used in the model are as follows:

$$DFRI = H^{wh} \times E^{we} \times V^{wv} \times (1 - R)^{wr} \quad (24)$$

where *DFRI* is the debris flow hazard risk index; the values of H, E, V, R represent the index of debris flow hazard, exposure, vulnerability, and ERRC factors, respectively; w_h, w_e, w_v, and w_r are the weights assigned to the hazards, exposures, vulnerabilities, and ERRC factors of the debris flow hazards, respectively.

$$E = \sum_{i=1}^{n} W_{ei} X_{ei} \tag{25}$$

$$V = \sum_{i=1}^{n} W_{pi} X_{pi} \tag{26}$$

$$C = \sum_{i=1}^{n} W_{ri} X_{ri} \tag{27}$$

where n is the total number of indicators; i is the *ith* indicator; W_{ei}, W_{pi}, and W_{ri} are the obtained factor weights; and X_{ei}, X_{pi}, and X_{ri} are the quantitative values of the indicators corresponding to exposure, vulnerability, and ERRC, respectively.

4. Result

4.1. Simulation Result of Burst Water Body

The flood map shows that the flood inundation area is 1.71×10^8 m^2 (Figure 7). The inundated area with a maximum water depth < 5 m is 1.27×10^8 m^2, accounting for 74% of the inundated area; the area with a maximum water depth of 5–10 m is 2.83×10^7 m^2, accounting for 17% of the total area; the area with a maximum water depth of 10–20 m is 1.3×10^7 m^2, accounting for 7.2% of the total area, and this water depth is located in the main ditch of Erdao Baihe River; the maximum water depth > 20 m is 3.1×10^6 m^2, accounting for 1.8% of the total area, and this water depth is located in the source of the Tianchi outburst, the Changbai Mountain volcano. After the Tianchi outburst, the floodwaters rushed out along the Erdao Baihe River. They gradually spread outward, passing through the scenic area of the north slope of Changbaishan Mountain, EDBH town.

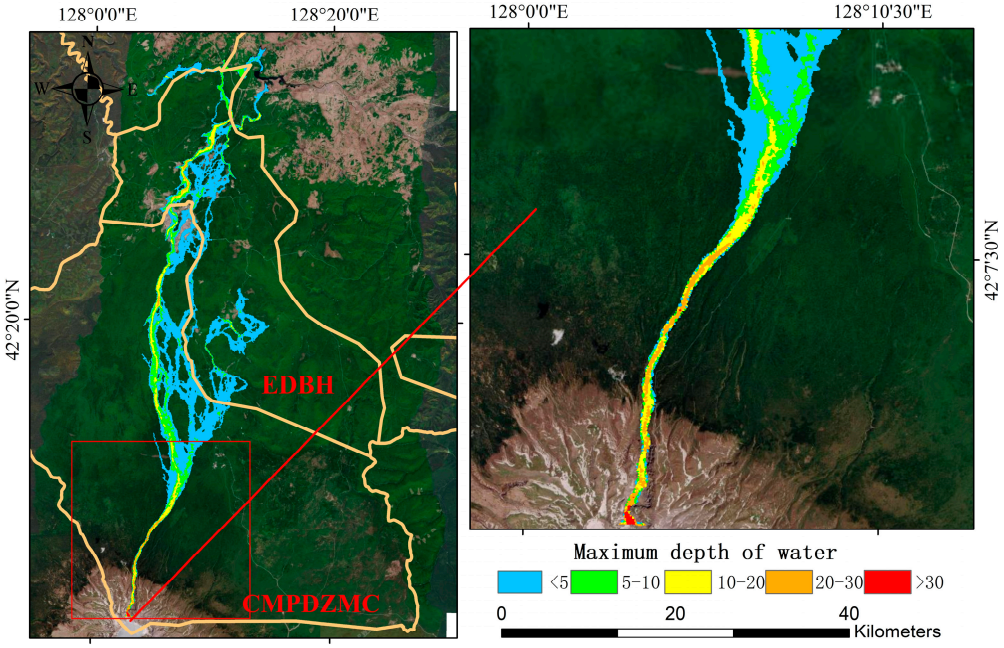


Figure 7. Maximum water depth distribution of flood routing.

4.2. Parameter Introduction and Processing Results

In this study, we employed a multi-dimensional indicator assessment to develop the indicator system for evaluating debris flow disaster risk in Antu County. Drawing upon the theory of comprehensive risk assessment for natural disasters, we categorized the assessment into four elements: hazard, exposure, vulnerability, and ERRC. Regarding the hazard component, we not only reviewed extensive relevant literature but also integrated the geological characteristics of the study area and the mechanisms underlying debris flow disaster occurrence. For the exposure, vulnerability, and ERRC, we considered indicators related to demographics, economy, and education from a societal perspective, grounded in the population, economic, and social context of the study area. In summary, we identified a total of 20 assessment indicators, including maximum water depth, average annual rainfall, slope, topographic relief, TWI, SPI, DFWS, vegetation cover, lithology, population density, road density, building density, percentage of vulnerable population (POVP), road category, building stories, land use, GDP, number of beds in hospitals and welfare centers (NOBIHW), number of healthcare workers (NOHCW), and education status, as the indicators utilized in this study.

4.2.1. Hazard Indicator

Average Annual Rainfall: Water is a critical factor in debris flow formation, serving as both a stimulus and a fundamental component. It plays a key role in the debris flow process by contributing to the conditions necessary for its occurrence. The formation of debris flow requires substantial surface runoff, which acts as a crucial trigger for debris flow events. Surface runoff on slopes and gully beds continually induces erosion and hillside collapse. This process mobilizes a significant amount of solid materials, such as sand and stones, which are transported by the water flow and driven by gravity and inertia under conditions of rapid runoff [34].

Slope: Slope denotes the average gradient of the terrain on either side of a debris flow channel, and it is a critical indicator for assessing geological hazards. Generally, a gentle slope does not favor the occurrence of debris flows. In contrast, a steeper slope increases the potential energy of downward-moving material, leading to slope instability and a higher probability of debris flow [35]. However, a steeper slope does not always correspond to a higher probability of debris flow. Within a certain range of gradients, if the slope is too shallow, there may be insufficient potential energy to mobilize the material. Conversely, if the slope is too steep, it may hinder the accumulation of debris.

Vegetation Coverage: Vegetation plays a crucial role in water retention, reducing surface runoff, stabilizing the soil, and enhancing overall soil stability [36,37]. In areas with sparse vegetation cover, rock and soil weathering is more pronounced, and erosion by rivers and rainfall is more intense. Consequently, soil stability is compromised, increasing the likelihood of debris flow disasters.

Lithology: The lithology of stratigraphic layers plays a critical role in the formation of debris flows. Variations in the physical and mechanical properties of rocks significantly influence this process [38]. The weathering and erosion of resistant rock formations and certain loose accumulation layers can supply ample material sources, thereby facilitating the development of debris flows. In this study, lithologies are classified based on their relative softness or hardness. Specifically, granite and basalt are categorized as very hard, andesite and diorite as hard, coarser facies and metamorphic rocks as moderately hard, gneisses and sandstones as soft, and cherts and shales as very soft.

DFWS: River scouring can lead to the transport of weathered topsoil and loose accumulations into the river channel. Prolonged transport can result in sedimentation, which increases the likelihood of debris flow formation during periods of heavy rainfall [39].

Topographic Relief: Topographic relief is a crucial parameter for quantitatively describing landform characteristics. Originating from the concept of "depth of terrain cut" introduced by the Institute of Geography of the Academy of Sciences of the former USSR, it is now a key indicator for classifying landform types [40]. A higher degree of topographic

relief correlates with the increased potential energy of the landform, which, to some extent, enhances the likelihood of geological hazards.

TWI: The TWI measures the degree of soil wetness or dryness [41]. Higher soil moisture content increases the likelihood of saturation and runoff, while lower moisture content indicates drier soil. The formula is as follows:

$$TWI = \ln\left(\frac{A_s}{\tan\beta}\right) \tag{28}$$

where A_s is the unit catchment area and β is the slope.

SPI: The erosive power of surface water flows is typically quantified using the SPI. This index can be used to identify the pathways of intense flows formed by water accumulation on natural slopes and pinpoint locations susceptible to gully-head erosion, which may lead to debris flows [42]. The formula is calculated as follows:

$$SPI = A_s \tan\beta \tag{29}$$

where A_s is the unit catchment area and β is the slope.

The spatial distribution of hazard indicators is shown in Figure 8.

4.2.2. Exposure Indicator

Population Density: Population density is a critical exposure indicator for debris flow disasters [43]. Higher population density and more frequent human activities increase the potential loss of life and the exposure of individuals to debris flow hazards.

Building Density: Buildings, both public and private, are fundamental to daily life and play a crucial role in the material economy [44]. Higher building density indicates a greater concentration of structures within a given area.

Road Density: Road transportation serves as a crucial link in social production, circulation, and consumption, becoming an indispensable element of social infrastructure [45]. Higher road density reflects a greater concentration of roads within a given area.

The spatial distribution of exposure indicators is shown in Figure 9.

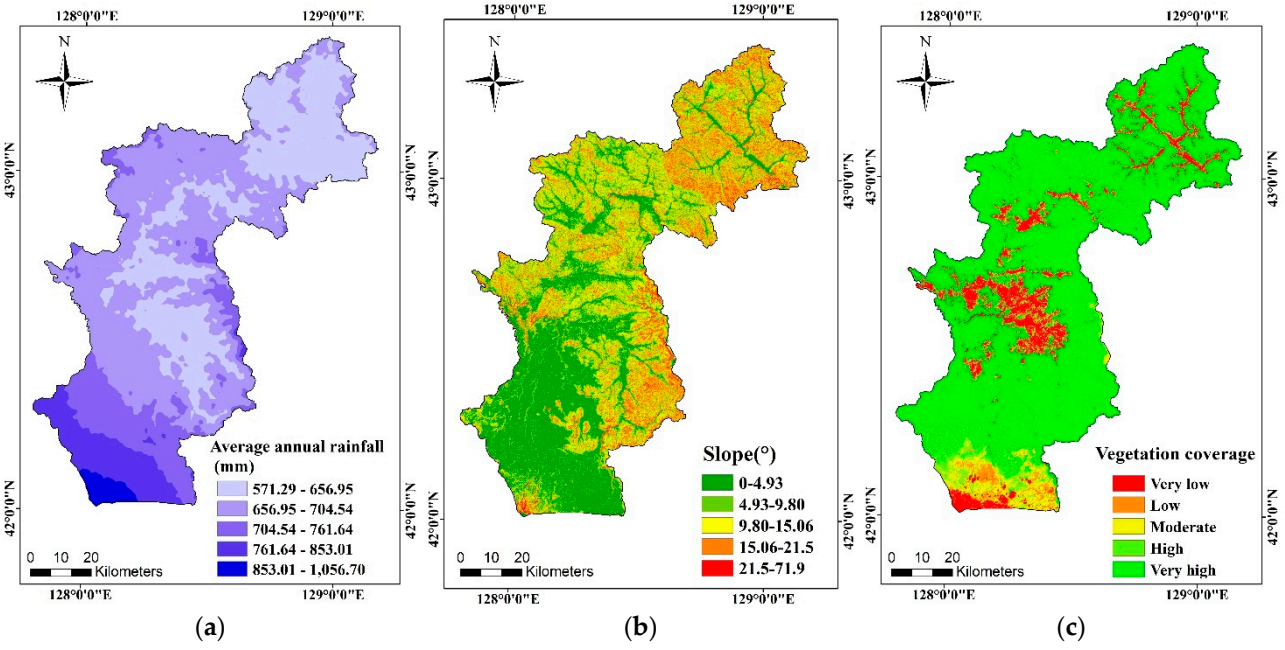


Figure 8. *Cont.*

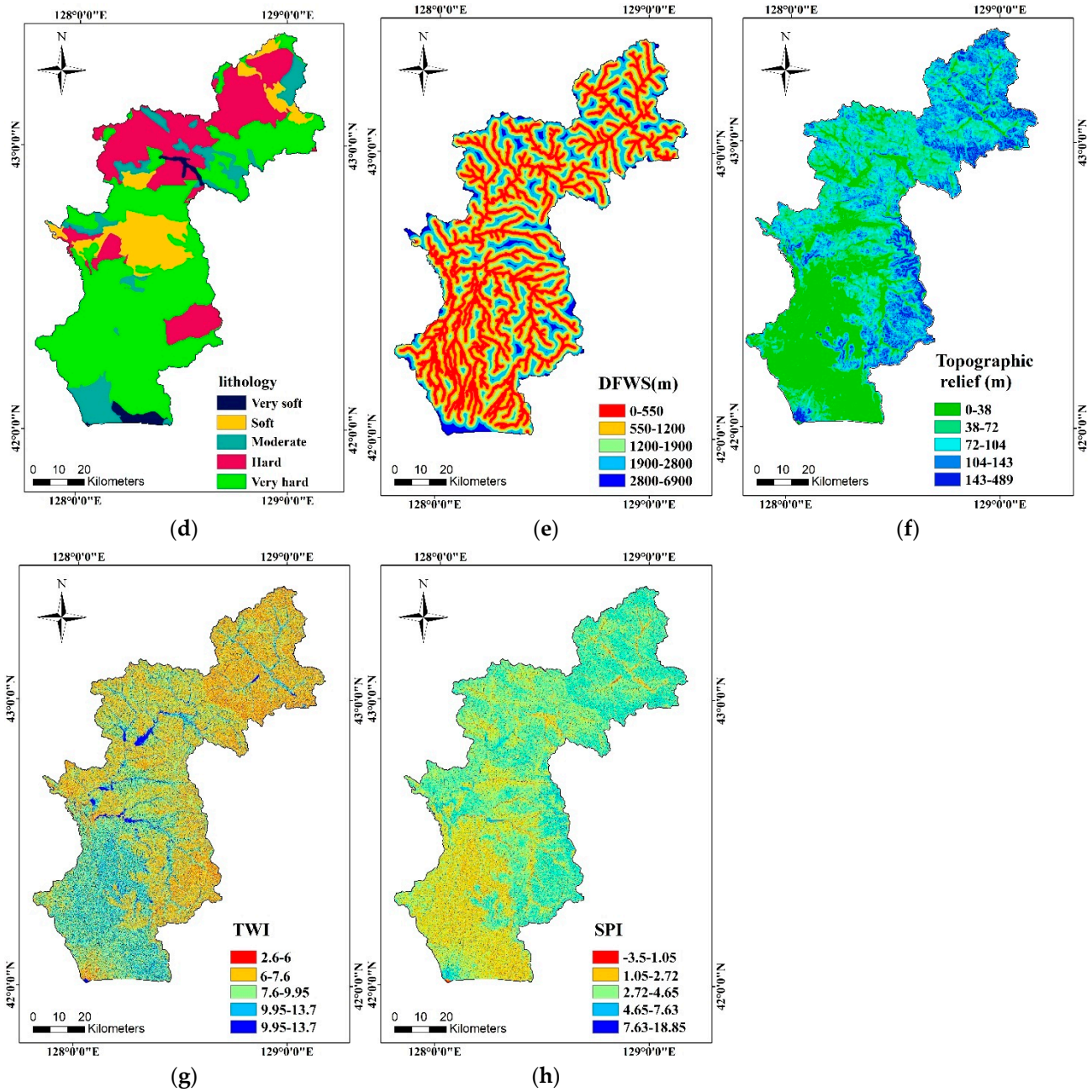


Figure 8. (**a**) Average annual rainfall; (**b**) slope; (**c**) vegetation coverage; (**d**) lithology; (**e**) DFWS; (**f**) topographic relief; (**g**) TWI; (**h**) SPI.

4.2.3. Vulnerability Indicator

Road Category: The category of a road is a significant determinant of its vulnerability to damage from debris flow disasters [46]. Higher-grade roads are generally more structurally robust, costlier to construct, and exhibit lower vulnerability. In contrast, rural roads, which are typically of lower grade, are more susceptible to damage during such events.

POVP: Individuals under the age of 14 and over the age of 60 generally possess lower capacity to respond to and evade disasters compared to adults between these ages. Consequently, this demographic is at a higher risk of harm during debris flow disasters [47].

Land Use: The value of land resources significantly impacts the extent of loss during a debris flow disaster [48]. By assigning values to different land use types, it is evident that higher compensation for land occupation corresponds to higher land value, which in turn results in a higher vulnerability value.

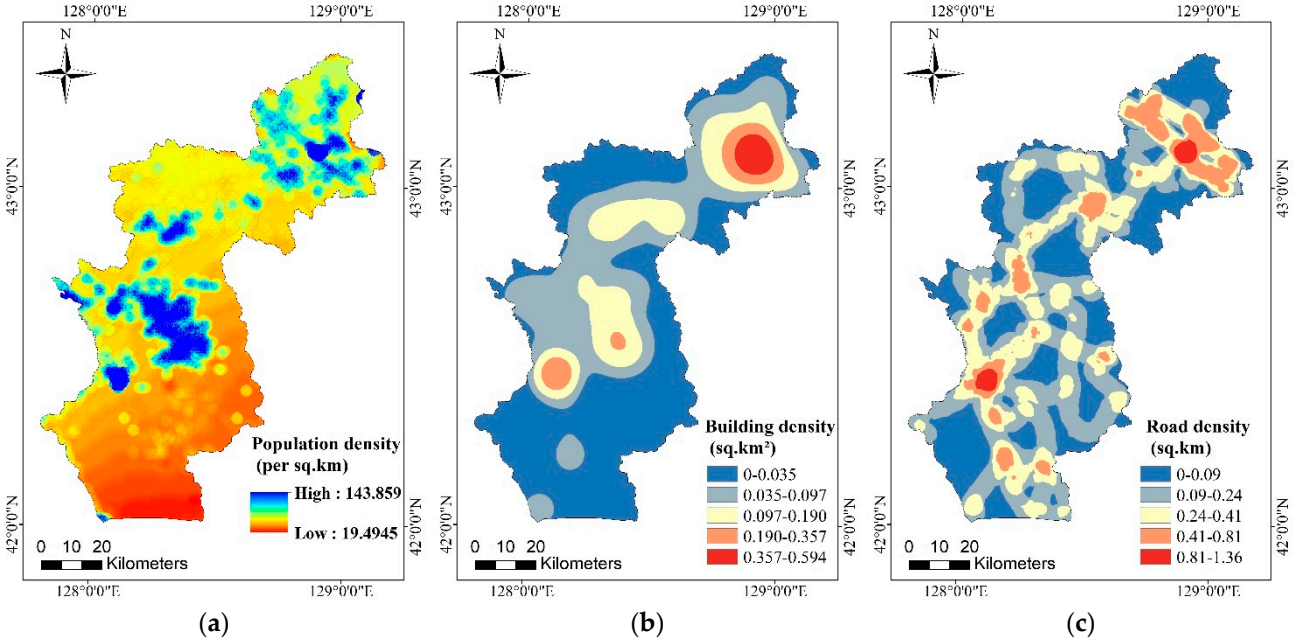


Figure 9. (**a**) Population density; (**b**) building density; (**c**) road density.

Building Stories: The number of stories in a building can reflect the level of economic development to some extent [49]. In terms of vulnerability assessment, the number of stories directly correlates with the potential loss a building may incur during a disaster. Generally, buildings with fewer stories are more susceptible to being severely damaged or buried in the event of a disaster.

The spatial distribution of vulnerability indicators is illustrated in Figure 10.

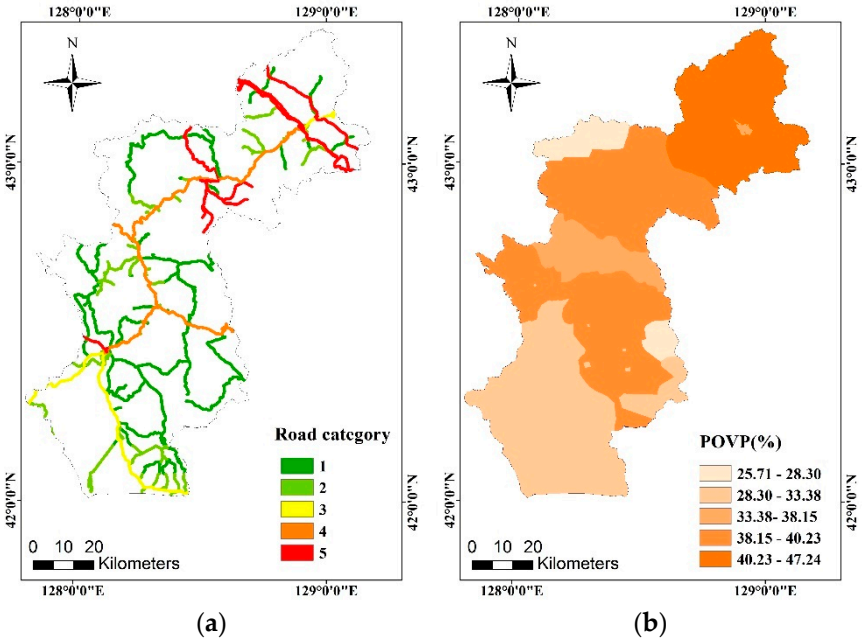


Figure 10. *Cont.*

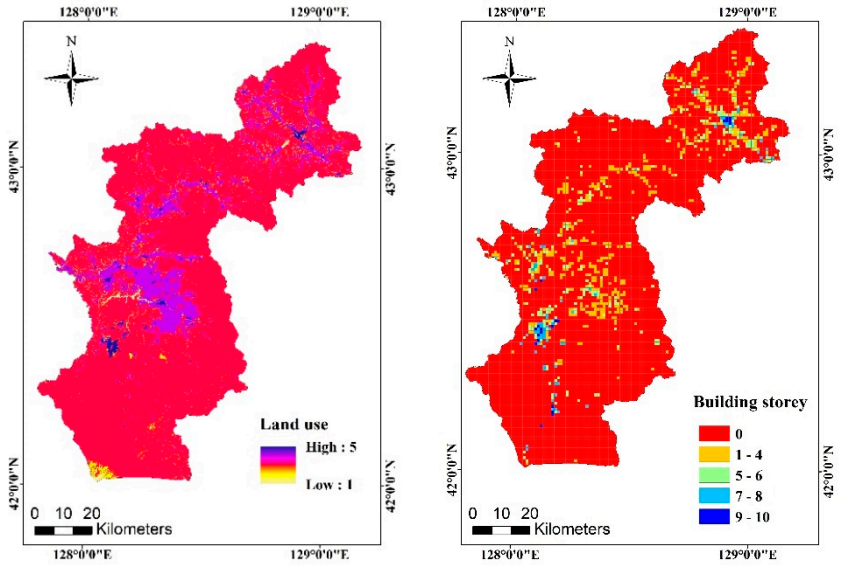


Figure 10. (**a**) Road category; (**b**) POVP; (**c**) land use; (**d**) building stories.

4.2.4. ERRC Indicator

GDP: GDP refers to the gross domestic product, which can also refer to the total value created by the people in a certain time and space. It directly reflects the degree of development of a region [50]. The higher the GDP, the more developed the region is and the stronger its ability to recover from the debris flow disaster.

Education Status: Education enhances disaster awareness and knowledge, thereby improving individuals' capacity to effectively manage and respond to disasters [51]. Individuals with higher levels of education generally possess superior disaster-coping abilities.

NOHCW/NOBIHW: The capacity for relief in a region is crucial for effective disaster response and recovery. Regions with a higher number of healthcare workers, hospitals, and orphanage beds are better equipped to manage and recover from disasters [52].

The spatial distribution of ERRC indicators is shown in Figure 11.

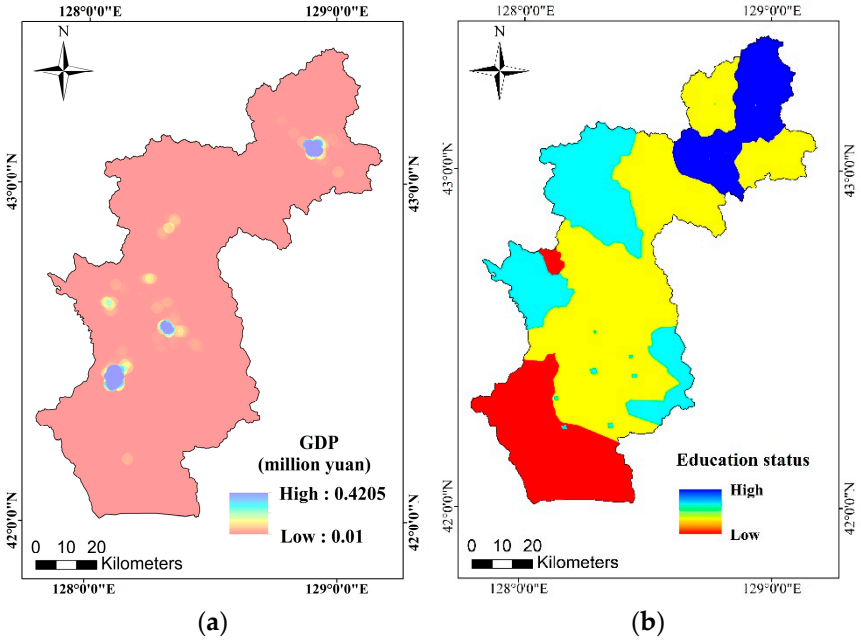


Figure 11. *Cont.*

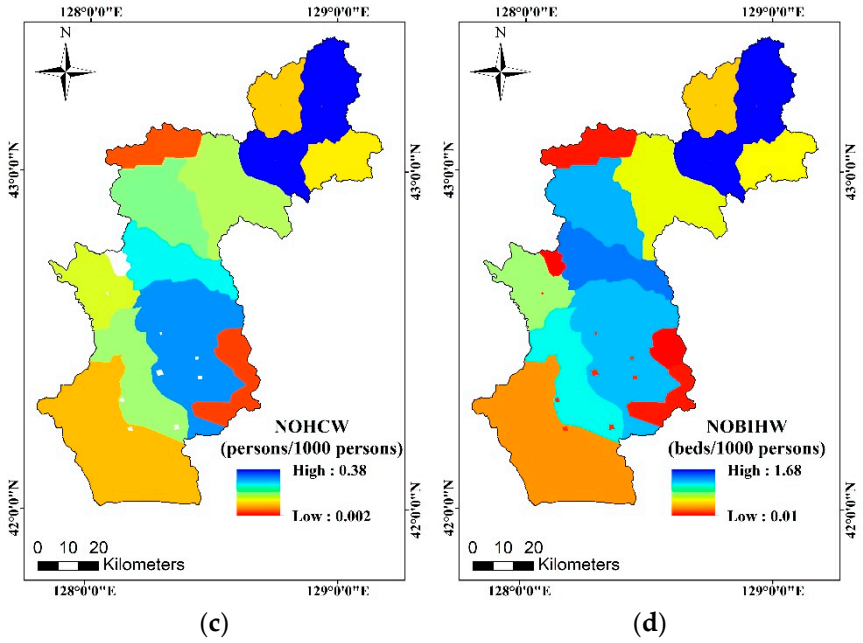


Figure 11. (**a**) GDP; (**b**) education status; (**c**) NOHCW; (**d**) NOBIHW.

4.3. Model Validation and Sensitivity Analysis Results

For hazard assessment, the accuracy of the two models used in the study was judged based on the AUC value of the area under the ROC curve. The AUC values range from 0 to 1, whereby a value approaching 1 signifies superior discrimination capabilities. The AUC value for the Information Quantity model presented in this study (Figure 12a) is 0.87, while the AUC value for the Random Forest model is 0.90, suggesting that the discriminative performance of both models is acceptable, with the Random Forest model demonstrating enhanced precision. Furthermore, a comparative analysis of the AUC values of the Random Forest model with those from other similar studies (Table 7) reveals that the Random Forest model exhibits the highest precision. Consequently, this study adopts the Random Forest model for assessing the hazard associated with debris flow.

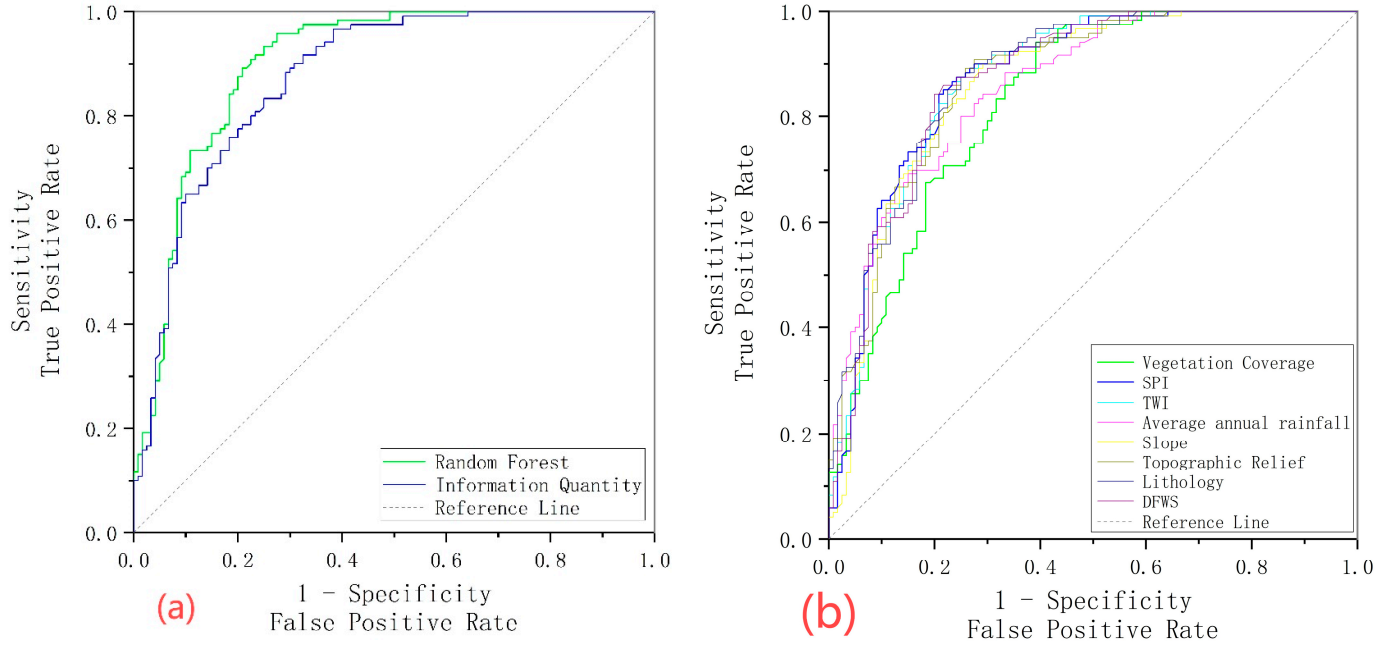


Figure 12. (**a**) ROC curve; (**b**) sensitivity analysis of indicators.

Table 7. Comparison of AUC values from other studies.

Research	Luo et al., 2022 [38]	Sun et al., 2024 [53]	Li et al., 2023 [54]	Yuan et al., 2022 [55]	Yu et al., 2021 [56]	Ours
AUC value	0.737	0.796	0.824	0.894	0.889	0.900

The sensitivity analysis of the indicators pertaining to the Random Forest model was conducted using Formula (17), with the results illustrated in Figure 12b and Table 8. It is evident that all factor RDs exceed 0, indicating that each of these indicators positively influences the model's assessment outcomes. Notably, the removal of the vegetation cover indicator exerts the most significant impact on the assessment results. The deletion of the lithology indicator demonstrates the least impact on the model's assessment outcomes.

Table 8. Sensitivity coefficient values.

Indicator	Average Annual Rainfall	SPI	TWI	Slope	Topographic Relief	Vegetation Coverage	Lithology	DFWS
RD	3.3	2.7	2.7	4.3	4.2	7.6	2.4	3.1

4.4. Hazard, Exposure, Vulnerability, and ERRC Results

The information values of the hazard indicators and the calculation results of the weights of each indicator are shown in Tables 9 and 10. The results of the four-factor zoning map are shown in Figure 13 by using the raster calculator in GIS.

Table 9. Information quantities of the indicators.

Indicator Layer	Category	Information Quantity
Slope	0–4.93	−0.5951
	4.93–9.80	0.2155
	9.80–15.06	0.2132
	15.06–21.5	0.5824
	21.5–71.9	1.5738
Topographic relief	0–38	−0.916500757
	38–72	0.537534352
	72–104	0.120183604
	104–143	−0.240809566
	143–489	0.461616993
Average annual rainfall	571–656	0.68347837
	656–704	−0.634114802
	704–761	−1.270701842
	761–853	−0.745041941
	853–1056	1.159032524
TWI	2.6–6	−0.038017119
	6–7.6	−0.047757929
	7.6–9.95	−0.07962952
	9.95–13.7	0.494845358
	13.7–25	−0.055833464
SPI	−3.5–1.05	−1.077750746
	1.05–2.72	−0.072137699
	2.72–4.65	0.235479555
	4.65–7.63	0.336167532
	7.63–18.85	0.793671632

Table 9. *Cont.*

Indicator Layer	Category	Information Quantity
DFWS	0–550	0.368379884
	550–1200	0.020267322
	1200–1900	−0.566280743
	1900–2800	−0.883742455
	2800–6900	0.693889346
Vegetation coverage	0–0.16	1.787859015
	0.16–0.45	1.218030363
	0.45–0.70	−0.358028398
	0.70–0.90	−0.263473695
	0.9–1	−1.256118074
Lithology	Very soft	0.383702631
	soft	0.823669846
	Moderate	0.256887989
	Hard	−0.256430435
	Very hard	−0.272373367

Table 10. Weight of each evaluation indicator.

Target Layer	Target Layer Weight	Indicator Layer	Indicator Layer Weight
Hazard	0.426	Maximum water depth	0.102
		Average annual rainfall	0.232
		TWI	0.086
		SPI	0.079
		Distance to water system	0.061
		Slope	0.103
		Topographic relief	0.103
		Vegetation coverage	0.086
		Lithology	0.084
Exposure	0.201	Population density	0.448
		Road density	0.189
		Building density	0.363
Vulnerability	0.221	Land use	0.290
		Road category	0.244
		Building stories	0.174
		POVP	0.290
ERRC	0.153	GDP	0.390
		NOBIHW	0.223
		NOHCW	0.211
		Education status	0.177

The results of the hazard assessment conducted using the Random Forest model are presented in Figure 14 and Table 11. The analysis in conjunction with the hazard zoning map reveals that higher hazard areas in Antu County are predominantly located in the southern and northern regions, whereas the central area exhibits relatively lower hazard levels. This pattern is attributed to the central region's relatively gentle topography and lower average annual rainfall. Notably, incorporating the factor of the Tianchi outburst leads to an increased proportion of high and very high hazard areas in Erdao Baihe Township, situated in the central downward area. The statistical results for these hazard zoning subdivisions are detailed in Table 12.

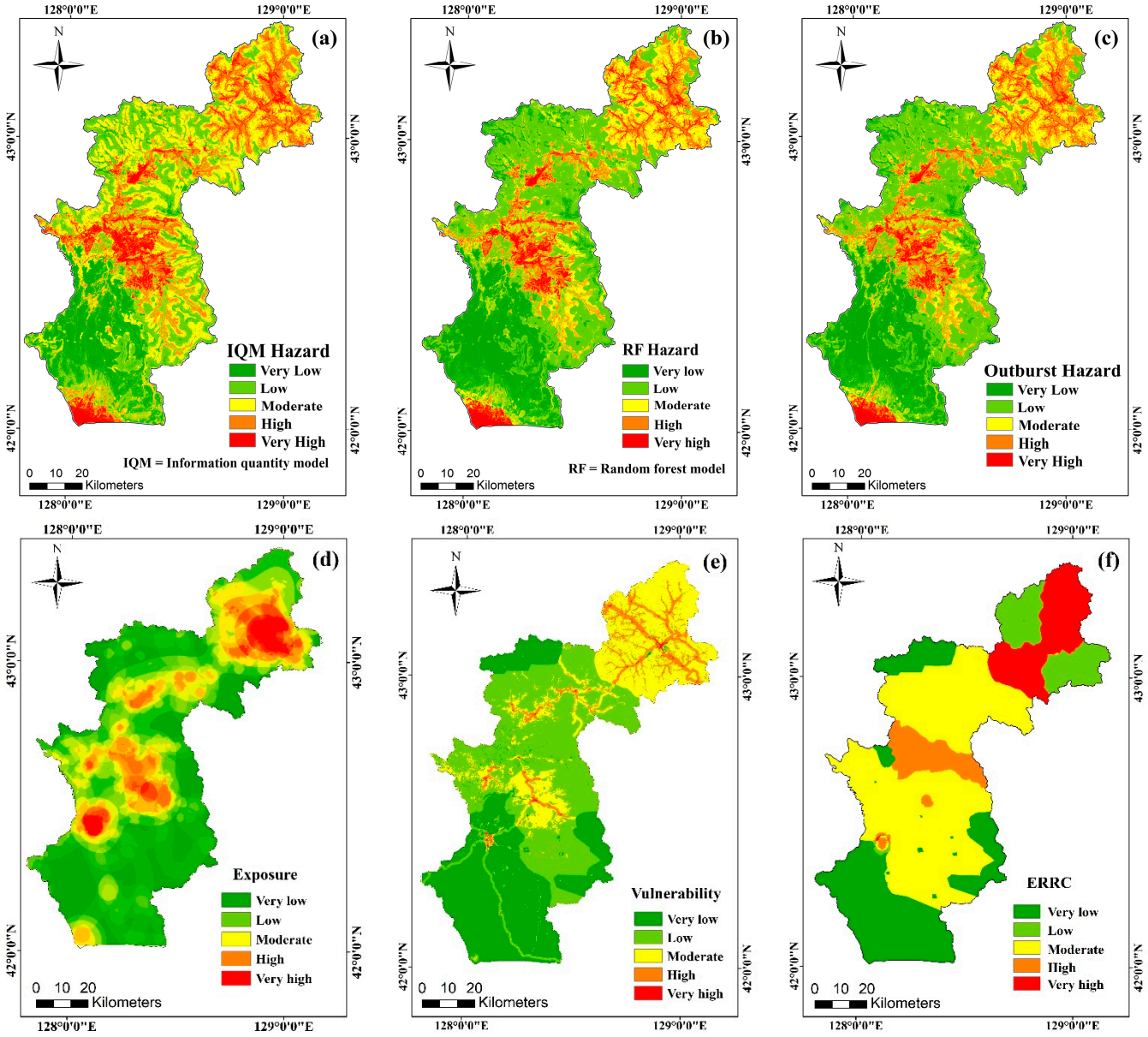


Figure 13. (**a**) Information quantity hazard; (**b**) random forest hazard; (**c**) outburst hazard; (**d**) exposure; (**e**) vulnerability; (**f**) ERRC.

Table 11. Random forest model evaluation results.

		Accuracy	Recall Ratio	Precision Ratio	F1
No outburst hazard	Training sets	0.994	0.994	0.994	0.994
	Testing sets	0.903	0.903	0.903	0.903
Outburst hazard	Training sets	0.982	0.982	0.982	0.982
	Testing sets	0.903	0.903	0.906	0.903

The exposure zoning map indicates that the regions with the highest exposure levels are predominantly situated in the northern part of Antu County, primarily due to Mingyue Township being the central hub of the county. This centrality attracts individuals from neighboring townships seeking development opportunities, resulting in increased population density and, consequently, higher demand for buildings. Thus, both population density and building density are relatively elevated in this area. Additionally, Erdao Baihe

Township exhibits high exposure levels because it is closest to the northern scenic area of Changbaishan Mountain, where road density is high and there is a significant movement of people.

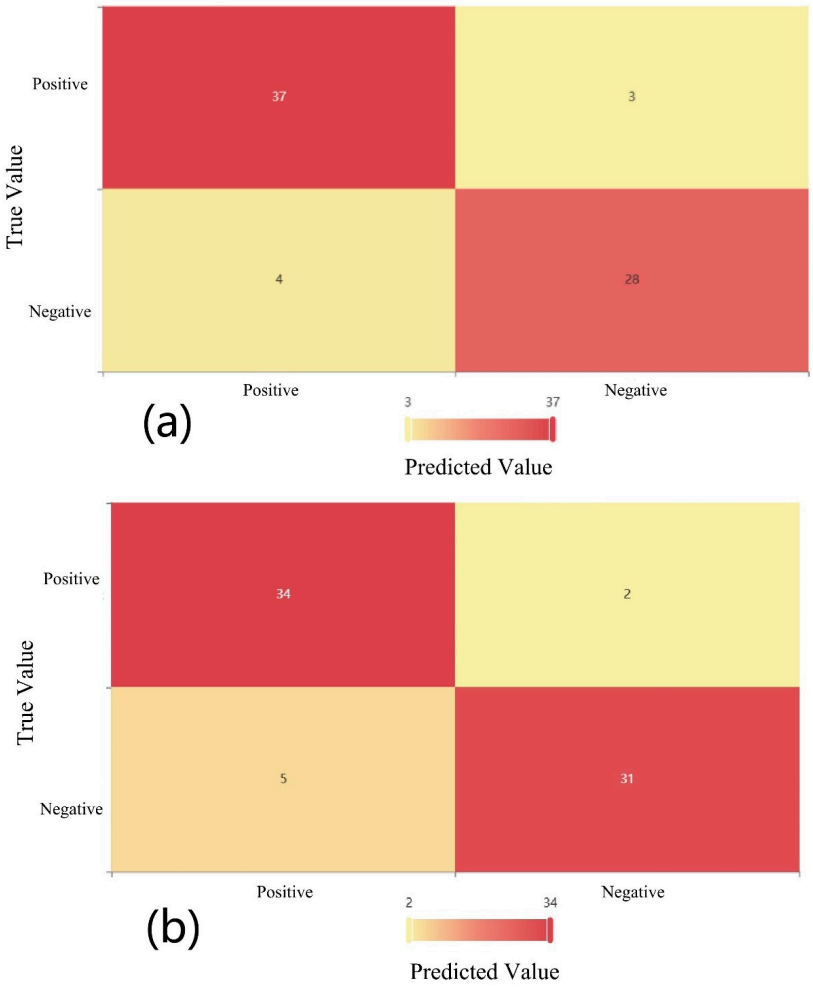


Figure 14. Random forest confusion matrix heat map: (**a**) no outburst hazard; (**b**) outburst hazard.

Table 12. Comparative analysis of hazard risk.

Type	Level	Area (km^2)	Proportion (%)
No outburst hazard	Very low	566.83	7.8
	Low	1021.60	14.07
	Moderate	2088.06	28.76
	High	2412.42	33.23
	Very high	1170.34	16.12
Outburst hazard	Very low	547.51	7.5
	Low	1011.08	13.93
	Moderate	2051.53	28.26
	High	2474.46	34.09
	Very high	1174.68	16.18

The vulnerability zoning map reveals that the highest vulnerability levels are present across all townships. These high-vulnerability areas are characterized by a prevalence of low-rise buildings, dense rural road networks, and a substantial vulnerable population. The elevated overall vulnerability in the northern region can be attributed to Mingyue

Township, situated in the north and serving as the central hub of Antu County, which has significant land use value throughout the northern area.

The analysis in conjunction with the ERRC zoning map reveals that regions with higher disaster prevention and mitigation capacities are consistently associated with higher GDP levels. Specifically, Mingyue Township demonstrates a high capacity for disaster prevention and mitigation, attributed to its superior medical facilities, concentration of educational resources, advanced educational status, and a greater number of healthcare professionals.

4.5. Debris Flow Risk Result

As illustrated in Figure 15, the risk of debris flow disasters in Antu County was assessed in accordance with Equation (24) and subsequently categorized into five levels utilizing the natural breakpoint method. In conjunction with Figure 16 and Table 13, it is evident that, excluding the consideration of the Huokou Lake outburst, the proportions of high-risk and very high-risk zones are 7.2% and 2.5%, respectively. Despite the relatively low proportions of these zones, they warrant attention due to their concentration in economically prosperous and densely populated township centers. The primary factors contributing to the high- and very high-risk areas are the concentration of agricultural, construction, and other developmental activities in these regions, which is further exacerbated by prolonged rainfall and the incompatibility of human engineering with the geological environment. Road construction and urban development frequently necessitate extensive land excavation and vegetation removal, thereby undermining the natural capacity for soil and water conservation, consequently heightening the risks of soil erosion and water loss. Additionally, urban development is typically accompanied by population concentration, which may exacerbate the impact of disasters during debris flow events. Elevated population densities not only amplify the pressure on emergency response efforts but also complicate rescue and recovery operations.

Table 13. Comparative analysis of risk.

Type	Level	Area (km^2)	Proportion (%)
No outburst risk	Very low	3112.26	42.90
	Low	2098.73	28.93
	Moderate	1341.64	18.49
	High	524.53	7.2
	Very high	178.04	2.5
Outburst risk	Very low	3101.92	42.75
	Low	2098.06	28.92
	Moderate	1276.95	17.60
	High	542.55	7.48
	Very high	235.72	3.25

When considering the Huokou Lake outburst, the proportions of high-risk and very high-risk areas are 7.48% and 3.25%, and the proportions have increased by 0.28% and 0.75%, respectively. Through comparative analysis, it was found that the increase in the proportion of high-risk and very high-risk areas was due to the increase in the risk level of Erdao Baihe Township and the northern scenic area of Changbaishan Mountain. Erdao Baihe Township is an important gateway to Changbaishan Mountain, attracting a large number of tourists and promoting local economic development, while Changbaishan North Scenic Area has important ecological and cultural values, which leads to the higher exposure and vulnerability of these two areas themselves. When considering the Huokou Lake outburst, the increase in hazard risk in these two regions leads to an increase in high- and very high-risk areas. The marked difference between the two scenarios highlights the critical importance of including extreme event scenarios in risk assessments, and the inclusion of outburst scenarios in risk assessments not only highlights the complexity of

disaster risk, but also emphasizes the need for a comprehensive and adaptive disaster management framework.

After counting, the number of debris flows in the very low-risk area is one, which is 0.83%; the number of debris flows in the low-risk area is six, which is 5%; the number of debris flows in the medium-risk area is fifteen, which is 12.5%; the number of debris flows in the high-risk area is thirty-six, which is 30%; and the number of debris flows in the very high-risk area is sixty-two, which is 51.67%. This shows that the results of this study are reliable.

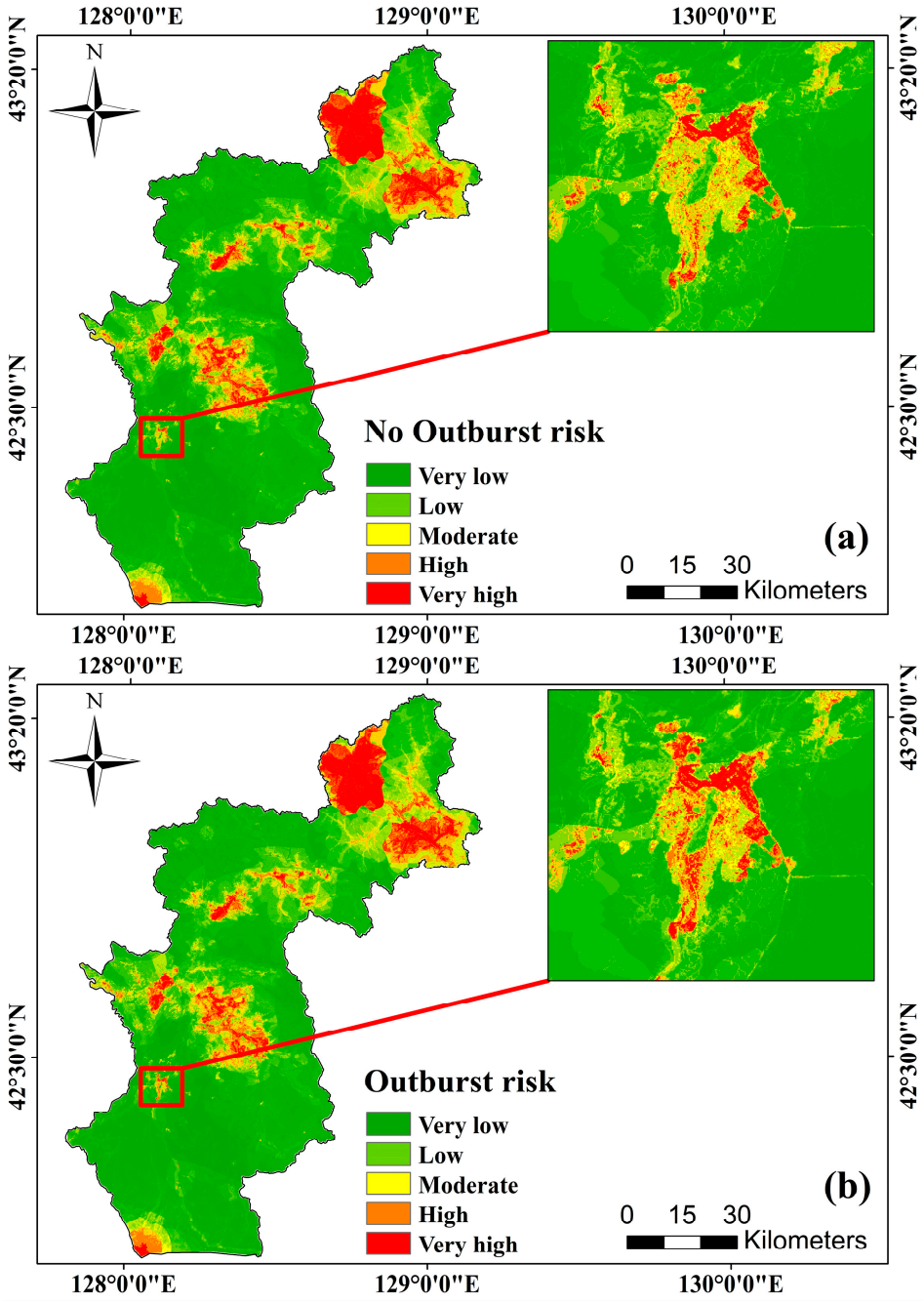


Figure 15. Results of debris flow disaster risk assessment: (**a**) no outburst risk; (**b**) outburst risk.

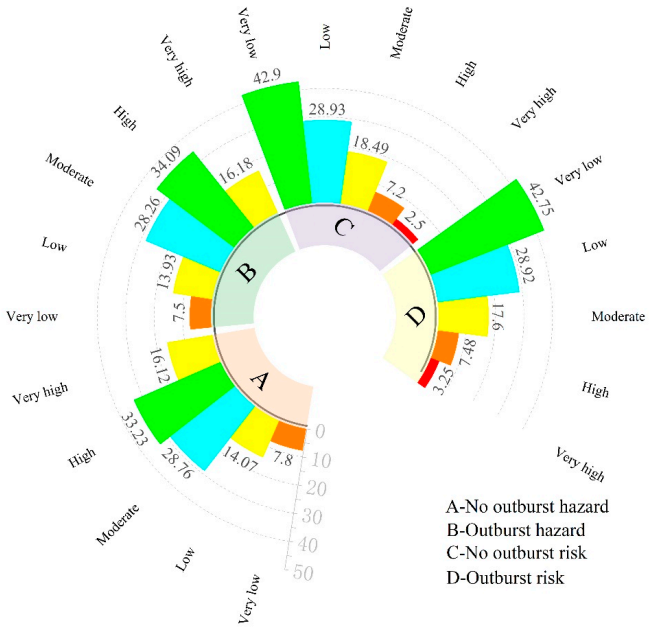


Figure 16. Proportion of hazard and risk areas.

5. Discussion and Conclusions

5.1. Discussion

The CRITIC method is a quantitative approach that determines the weights of indicators based solely on objective data. While it is grounded in a robust mathematical theoretical framework, it is susceptible to fluctuations in objective data and has a limited scope of application. The ANP method involves the assignment of subjective weights by the decision analyst, reflecting the degree of emphasis placed on each indicator. While this approach captures the subjective intentions of the decision-maker, it also introduces a higher degree of subjectivity and arbitrariness. When the decision-maker's assessment of the importance of the indicators is either excessively high or unduly low, it fails to accurately represent the objective reality. This study integrates the ANP-CRITIC approach, grounded in game theory. In comparison to utilizing a single method, this combined approach ensures both the objectivity and scientific rigor of risk zoning while allowing for subjective analysis based on the specific context of the study area.

In recent years, the continuous enrichment and enhancement of geohazard big data, coupled with advancements in computational power, have made the analysis of geohazards using machine learning algorithms a prominent research focus. In the domain of debris flow research, machine learning is predominantly utilized in assessing susceptibility and hazard risk, primarily employing models such as the random forest, support vector machine, and logistic regression. Liu Yongyao et al. [57] employed the random forest algorithm to assess the susceptibility of debris flow in the Wenchuan earthquake region, achieving an average AUC value of 0.84, which indicates the high stability and accuracy of the model. Li Kun et al. [58] utilized both the random forest algorithm and the support vector machine algorithm to evaluate the susceptibility of debris flow in Dongchuan, with results indicating that the random forest algorithm is more suitable for assessing the susceptibility of mountain debris flows. Consequently, this study employed the random forest algorithm to enhance the accuracy of the assessment results concerning debris flow hazards.

Anto County has been designated by UNESCO and WWF as one of the 40 nature reserves within the global "Man and the Ecosphere Conservation Network". The significance of assessing debris flow disaster risk in Antu County lies in its ability to proactively identify potential debris flow hazards and implement appropriate disaster prevention and mitigation measures. This approach aims to safeguard both human lives and property, while

also preserving the region's vital non-renewable resources. Furthermore, the methodology presented in this study may serve as a reference for similar assessments in other regions.

The assessment of debris flow disaster risks encompasses not only the rigorous scientific evaluation of natural hazards but also establishes a direct link to sustainable socio-economic development. Through the quantification and monitoring of debris flow risks, this study provides a robust scientific foundation that enables policymakers to devise more effective legislation aimed at safeguarding the ecological environment and ensuring resident safety. This process not only facilitates disaster risk reduction but also fosters the judicious use of resources and sustainable regional development, thereby supporting the attainment of more comprehensive sustainability objectives.

This study aims to holistically evaluate the assessment indicators pertinent to debris flow disasters through the lens of the four elemental categories. Nonetheless, the data utilized in this study are derived from existing datasets, which may lead to the oversight of certain critical factors and indicators due to limitations in data availability and completeness. For instance, the collection of data regarding the proportion of vulnerable populations and indicators of emergency response and recovery capacity proved to be particularly challenging. Consequently, we opted to initiate data mapping at the township level, which subsequently impacted the accuracy of our data representations. Future research should prioritize the collection of higher-quality data, including statistics from community units, to enhance the accuracy and reliability of the findings.

5.2. Conclusions

In this study, we employed a multi-criteria assessment framework to evaluate the risk of debris flow disasters, focusing on factors such as hazard, exposure, vulnerability, and ERRC. Additionally, we synthesized multi-dimensional indicators encompassing geophysical, socio-economic, and demographic aspects, including population and households. In the process of selecting indicators, we meticulously considered the specific context of the study area while also accounting for the triggering factor related to the outburst of Huokou Lake in the Changbai Mountain region.

In the assessment of debris flow hazards, this study employs the Information Quantity model and the Random Forest model for comparative analysis. The ROC curve was utilized to validate the hazard assessment results. The accuracy of the Information Quantity model was found to be 0.87, while the accuracy of the Random Forest model was determined to be 0.90. The accuracy of the Random Forest model was superior, leading to the conclusion that this model was ultimately employed for the assessment of debris flow hazards.

In the realm of disaster risk assessment, this study employs natural disaster risk theory in conjunction with game theory to optimize both the subjective and objective weights of the ANP and the CRITIC method. Ultimately, a debris flow disaster risk map was generated through the overlay analysis and visualization capabilities of the GIS platform. The results indicate that high-risk and very high-risk areas are predominantly concentrated in the central regions of economically prosperous and relatively densely populated townships, accounting for 9.7% of the total area. The risk associated with Erdao Baihe Township and the Changbaishan North Scenic Area is projected to escalate when accounting for the potential outburst of Fukou Lake, resulting in an increase of high-risk and very high-risk areas to 10.73%.

In light of the study's findings, we advocate for policymakers to implement initiatives that foster sustainable development practices while restricting activities that contribute to erosion and ecosystem degradation. Additionally, it is essential to develop community-specific disaster preparedness plans tailored to the unique needs of densely populated, high-risk areas. This encompasses allocating resources to emergency services to guarantee that vulnerable populations receive timely information and support. Moreover, it is imperative to invest in upgrading infrastructure to endure extreme events, including the establishment of specialized monitoring sites in the Changbai Mountains and the development of the Changbai Mountain Forecast and Warning Model.

Author Contributions: Conceptualization, P.L.; data curation, P.L. and Q.L.; formal analysis, P.L. and J.Z.; methodology, P.L., J.H. and Q.L.; writing—original draft, P.L.; funding acquisition, Y.Z. and J.Z.; writing—review and editing, P.L. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: Jilin Department of Education Science and Technology Research Project and Jilin Science and Technology Development Plan key Research and Development Project (JJKH20210688KJ, 20220203187SF).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

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. IPCC. Summary for Policymakers. In *Working Group I Contribution to the IPCC Fifth Assessment Report Climate Change 2013: The Physical Science Basis*; Cambridge University Press: Cambridge, UK, 2013.
2. Aruth. Risk Assessment of Debris Flow Disaster Chain Induced by Multiple Disaster-Causing Factors on the Northern Slope of Changbai Mountain Based on Multi-Model Coupling. Ph.D. Thesis, Northeast Normal University, Changchun, China, 2021. [CrossRef]
3. Guzzetti, F.; Ardizzone, F.; Cardinali, M.; Rossi, M.; Valigi, D. Landslide volumes and landslide mobilization rates in Umbria, central Italy. *Earth Planet. Sci. Lett.* **2009**, *279*, 222–229. [CrossRef]
4. Arattano, M. On the Use of Seismic Detectors as Monitoring and Warning Systems for Debris Flows. *Nat. Hazards* **1999**, *20*, 197–213. [CrossRef]
5. Carrara, A. Multivariate models for landslide hazard evaluation. *J. Int. Assoc. Math. Geol.* **1983**, *15*, 403–426. [CrossRef]
6. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; UNESCO: Paris, France, 1984; pp. 1–63.
7. Brabb, E.E. Innovative approaches to landslide hazard mapping. In Proceedings of the Fourth International Symposium on Landslides; Canadian Geotechnical Society: Toronto, ON, Canada, 1984; Volume 1, pp. 307–324.
8. He, S.; Jiang, Y.; Ji, X. Review on the economic loss of debris flow disaster. *Disasterology* **2019**, *34*, 153–158.
9. Kim, B.J.; Yune, C.Y.; Kim, K.S.; Lee, S.D. Risk assessment of debris flow on expressway in Korea using GIS. In *Landslides and Engineered Slopes, Proceedings of the 12th International Symposium on Landslides, Napoli, Italy, 12–19 June 2016*; CRC Press: Boca Raton, FL, USA, 2016; pp. 1183–1187.
10. Meehan, C.L.; Kumar, S.; Pando, M.A.; Coe, J.T. (Eds.) Geo-Congress 2019: Geotechnical Materials, Modeling, and Testing. In *Geo-Congress 2019, Proceedings of the Eighth International Conference on Case Histories in Geotechnical Engineering, Philadelphia, PA, USA, 24–27 March 2019*; American Society of Civil Engineers: Reston, VA, USA, 2019.
11. Tierz, P.; Woodhouse, M.J.; Phillips, J.C.; Sandri, L.; Selva, J.; Marzocchi, W.; Odbert, H.M. A Framework for Probabilistic MultiHazard Assessment of Rain-Triggered Lahars Using Bayesian Belief Networks. *Front. Earth Sci.* **2017**, *5*, 73. [CrossRef]
12. Han, X. Risk Assessment of Debris Flow Disaster in Dongchuan District of Yunnan Province. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2018.
13. Jiang, T. Risk Assessment and Early Warning Model of Debris Flow in Shijiaying Township, Beijing. Master's Thesis, University of Geosciences (Beijing), Beijing, China, 2022. [CrossRef]
14. Wu, Q.; Zhang, J.; Liu, X.; Ma, X.; Huang, J.; Wang, L. Risk assessment of debris flow disaster in Zhouqu small watershed based on grey correlation degree. *Sci. Technol. Eng.* **2023**, *23*, 10713–10719.
15. Jiang, Y.; Cui, J.; Zhang, L.; Yu, Z. Evaluation method and application of rock mass quality based on weight fusion-improved extension cloud model. *Min. Res. Dev.* **2023**, *43*, 84–90. [CrossRef]
16. Dou, J.; Yunus, A.P.; Bui, D.T.; Merghadi, A.; Sahana, M.; Zhu, Z.; Chen, C.-W.; Khosravi, K.; Yang, Y.; Pham, B.T. Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci. Total Environ.* **2019**, *662*, 332–346. [CrossRef]
17. Zhang, J.-Q.; Liang, J.-D.; Liu, X.-P.; Tong, Z.-J. GIS-Based Risk Assessment of Ecological Disasters in Jilin Province, Northeast China. *Hum. Ecol. Risk Assess.* **2009**, *15*, 727–745. [CrossRef]
18. Zhang, Y.; Wang, Y.; Yang, N. Research on risk assessment method of transponder system based on ANP and fuzzy evidence theory. *J. Saf. Environ.* **2018**, *18*, 434–440. [CrossRef]
19. Liu, W.; Zhang, J. Zoning evaluation of landslide susceptibility in Zhuxi County. *J. Nat. Disasters* **2024**, *33*, 175–185. [CrossRef]
20. Yu, L.; Guo, R.; Zhang, R. Landslide hazard assessment based on game theory and cloud model. *J. Nat. Disasters* **2024**, *33*, 186–196. [CrossRef]
21. Nie, Z.; Lang, Q.; Zhang, Y.; Zhang, J.; Ke, K. Risk assessment of collapse disasters along national highways based on information model and geodetector. *Disasterology* **2023**, *38*, 193–199.

22. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
23. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [CrossRef]
24. Cai, Y.; Xing, Y.; Hu, D. Review of Sensitivity Analysis. *J. Beijing Norm. Univ. (Nat. Sci. Ed.)* **2008**, *1*, 9–16.
25. Zhou, B.; Zou, Q.; Jiang, H.; Yang, T.; Zhou, W.T.; Chen, S.Y.; Yao, H.K. Process-driven susceptibility assessment of glacial lake outburst debris flow in the Himalayas under climate change. *Adv. Clim. Chang. Res.* **2024**, *15*, 500–514. [CrossRef]
26. Ahmed, R.; Rawat, M.; Wani, G.F.; Ahmad, S.T.; Ahmed, P.; Jain, S.K.; Meraj, G.; Mir, R.A.; Rather, A.F.; Farooq, M. Glacial Lake Outburst Flood Hazard and Risk Assessment of Gangabal Lake in the Upper Jhelum Basin of Kashmir Himalaya Using Geospatial Technology and Hydrodynamic Modeling. *Remote Sens.* **2022**, *14*, 5957. [CrossRef]
27. Wangchuk, T.; Tsubaki, R. A glacial lake outburst flood risk assessment for the Phochhu river basin, Bhutan. *Nat. Hazards Earth Syst. Sci.* **2024**, *24*, 2523–2540. [CrossRef]
28. Li, M. Numerical Simulation of Changbai Mountain Tianchi Lake Collapse Flood Based on Digital Elevation Model. Master's Thesis, Ocean University of China, Qingdao, China, 2015.
29. Hu, J. Application of MIKE21 Hydrodynamic Model in Flood Control and Dispatching Scheme in Western Mountainous Area of Beijing. Master's Thesis, Tsinghua University, Beijing, China, 2016.
30. Xie, R. *Dam-Break Hydraulics*; Shandong Science and Technology Press: Jinan, China, 1993; pp. 51.7–51.9.
31. Li, W. *Hydraulic Calculation Handbook*; China Water Conservancy and Hydropower Press: Beijing, China, 2006; pp. 434–462.
32. Chen, J. Research on Dynamic Risk Assessment of Debris Flow Disaster Induced by Extreme Rainfall in Mountainous Areas. Master's Thesis, Northeast Normal University, Changchun, China, 2019.
33. Duan, C.; Zhang, J.; Chen, Y.; Lang, Q.; Zhang, Y.; Wu, C.; Zhang, Z. Comprehensive Risk Assessment of Urban Waterlogging Disaster Based on MCDA-GIS Integration: The Case Study of Changchun, China. *Remote Sens.* **2022**, *14*, 3101. [CrossRef]
34. Qian, D.; Ma, H.; Zhu, Z. Spatial distribution and analysis of debris flow in Jishishan County, Gansu Province. *Bull. Surv. Mapp.* **2022**, *7*, 107–111+117. [CrossRef]
35. Ning, N.; Ma, J.; Zhang, P.; Qi, S.; Tian, L. Risk Assessment of Debris Flow Disaster in Bailong River Basin of Southern Gansu Province Based on GIS and Information Method. *Resour. Sci.* **2013**, *35*, 892–899.
36. Gao, Y.; Li, M.; Wang, D.; Bai, Y. Preliminary study on the inhibitory effect of vegetation on different types of debris flow. *Soil Water Conserv. Res.* **2013**, *20*, 291–295.
37. Shao, Z. Analysis of the influence of vegetation cover on debris flow development. *Value Eng.* **2017**, *36*, 151–152.
38. Luo, L.; Song, G.; Tang, L.; Xiang, L.; Li, X.; Liang, M. Distribution and risk assessment of debris flow in Wudu section of Bailong River. *People's Yangtze River* **2022**, *53*, 135–142. [CrossRef]
39. Shi, Z.; Zhu, H.; Wang, J.; Xin, C. Discussion on the susceptibility of soil landslide in Xiangxi from the perspective of coupled model. *Study Soil Water Conserv.* **2021**, *28*, 377–383. [CrossRef]
40. Alam, R.; Quayyum, Z.; Moulds, S. Dhaka City Water logging Hazards: Area Identification and Vulnerability Assessment through GIS-Remote Sensing Techniques. *Res. Sq.* **2021**, *1*, 27. [CrossRef]
41. Li, Z.; Chen, N.; Hou, J.; Wu, M.; Zhang, Y.L.; Du, P. Evaluation of debris flow susceptibility in loess area of Yili River Valley based on machine learning. *Chin. J. Geol. Disasters Prev.* **2024**, *35*, 129–140. [CrossRef]
42. Wang, X. Risk Assessment of Multiple Disasters in Tianshui Basin. Master's Thesis, Lanzhou University, Lanzhou, China, 2021. [CrossRef]
43. Li, J. Risk Assessment of Geological Disasters Based on PCA and Improved AHP-CRITIC Method. Master's Thesis, Northwest University, Xi'an, China, 2022. [CrossRef]
44. Wan, P. Risk Assessment of Geological Disasters in Liuhe County, Jilin Province. Master's Thesis, Jilin Jianzhu University, Changchun, China, 2023. [CrossRef]
45. Hu, Y.; Lv, L.; Chen, J.; Zhao, Z.; Gao, Y.; Ren, J.; Hu, S. Road connectivity and its influence by geological disasters in Yadong County, Tibet. *J. Yunnan Univ. (Nat. Sci. Ed.)* **2023**, *45*, 399–408.
46. Zhao, Z.; Chen, J.; Xu, K. A spatial case-based reasoning method for regional landslide risk assessment. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102381. [CrossRef]
47. Bruce, L. Risk Analysis of Debris Flow in Wudu Section of Bailong River Basin. Master's Thesis, Chongqing Jiaotong University, Chongqing, China, 2022. [CrossRef]
48. Yang, X. Geospatial Vulnerability Assessment of Coastal Building Hazard-Affected Bodies. Master's Thesis, Nanning Normal University, Nanjing, China, 2022. [CrossRef]
49. Wang, Z. Study on the Characteristics and Risk Assessment of Landslide Disaster in Chongren Town. Master's Thesis, China University of Mining and Technology, Beijing, China, 2023. [CrossRef]
50. Huang, X. Application of MaxEnt Model and AHP-Entropy Method in Geological Disaster Risk Assessment. Master's Thesis, Northwest University, Xi'an, China, 2022. [CrossRef]
51. Singkran, N.; Kandasamy, J. Developing a strategic flood risk management framework for Bangkok, Thailand. *Nat. Hazards* **2016**, *84*, 933–957. [CrossRef]
52. Zhang, Y.; Lang, Q.; Chen, Y. Provincial geological disaster risk zoning method based on natural disaster risk assessmet framework:A case sthdy in Jilin Province. *Chin. J. Geol. Hazard Control* **2020**, *31*, 104–110.
53. Sun, L.; Ge, Y.; Chen, X.; Zeng, L.; Liang, X.; Feng, X. Hazard of debris flow in the Jinsha River basin assessment. *Geol. Bull.* **2024**, 1–16. Available online: https://link.cnki.net/urlid/11.4648.p.20240222.2025.005 (accessed on 5 September 2024).

54. Li, J. GIS-Based Assessment of Debris Flow Hazard in Lushui City, Yunnan Province. Master's Thesis, Guizhou University for Nationalities, Guiyang, China, 2023. [CrossRef]
55. Yuan, G. Assessment of Debris Flow Hazard in Dechang County. Master's Thesis, Southwest University of Science and Technology, Mianyang, China, 2022. [CrossRef]
56. Yu, C. Research on Risk Zoning of Typical Geologic Hazards and Comprehensive Assessment of Geologic Environment Carrying Capacity in He Long City. Ph.D. Thesis, Jilin University, Jilin, China, 2021. [CrossRef]
57. Liu, Y.; Di, B.; Zhan, Y.; Stamatopoulos, C.A. Debris flows susceptibility assessment in Wenchuan earthquake areas based on random forest algorithm model. *Mt. Res.* **2018**, *36*, 765–773.
58. Li, K.; Zhao, J.; Lin, Y. Debris-flow susceptibility assessment in Dongchuan using stacking ensemble learning including multiple heterogeneous learners with RFE for factor optimization. *Nat. Hazards* **2023**, *118*, 2477–2511. [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.