

Article

Comparison of Machine Learning and Traditional Statistical Methods in Debris Flow Susceptibility Assessment: A Case Study of Changping District, Beijing

Feifan Gu ¹, Jianping Chen ¹, Xiaohui Sun ², Yongchao Li ³, Yiwei Zhang ¹ and Qing Wang ^{1,*}¹ College of Construction Engineering, Jilin University, Changchun 130026, China² Department of Earth Sciences and Engineering, Taiyuan University of Technology, Taiyuan 030024, China³ Key Laboratory of Shale Gas and Geoengineering, Institute of Geology and Geophysics, University of Chinese Academy of Sciences, Beijing 100029, China

* Correspondence: wangqing@jlu.edu.cn

Abstract: As a common geological hazard, debris flow is widely distributed around the world. Meanwhile, due to the influence of many factors such as geology, geomorphology and climate, the occurrence frequency and main inducing factors are different in different places. Therefore, the evaluation of debris flow sensitivity can provide a very important theoretical basis for disaster prevention and control. In this research, 43 debris flow gullies in Changping District, Beijing were cataloged and studied through field surveys and the 3S technology (GIS (Geography Information Systems), GPS (Global Positioning Systems), RS (Remote Sensing)). Eleven factors, including elevation, slope, plane curvature, profile curvature, roundness, geomorphic information entropy, TWI, SPI, TCI, NDVI and rainfall, were selected to establish a comprehensive evaluation index system. The watershed unit is directly related to the development and activities of debris flow, which can fully reflect the geomorphic and geological environment of debris flow. Therefore, the watershed unit was selected as the basic mapping unit to establish four evaluation models, namely ACA-PCA-FR (Analytic Hierarchy Process-Principal Component Analysis-Frequency Ratio), FR (Frequency Ratio), SVM (Support Vector Machines) and LR (Logistic Regression). In other words, this research evaluates debris flow susceptibility by comparing it with two traditional weight methods (ACA-PCA-FR and FR) and two machine learning methods (SVM and LR). The results show that the SVM evaluation model is superior to the other three models, and the value of the area under the receiver-operating characteristic curve (AUC) is 0.889 from the receiver operating characteristic curve (ROC). It verifies that the SVM model has strong adaptability to small sample data. The study was divided into five regions, which were very low, low, moderate, high and very high, accounting for 22.31%, 25.04%, 17.66%, 18.85% and 16.14% of the total study area, respectively, by SVM model. The results obtained in this research agree with the actual survey results, and can provide theoretical help for disaster prevention and reduction projects.

Keywords: debris flow susceptibility assessment; SVM; LR; ACA-PCA-FR; Beijing

Citation: Gu, F.; Chen, J.; Sun, X.; Li, Y.; Zhang, Y.; Wang, Q. Comparison of Machine Learning and Traditional Statistical Methods in Debris Flow Susceptibility Assessment: A Case Study of Changping District, Beijing. *Water* **2023**, *15*, 705. <https://doi.org/10.3390/w15040705>

Academic Editor: Athanasios Loukas

Received: 12 January 2023

Revised: 5 February 2023

Accepted: 7 February 2023

Published: 10 February 2023



Copyright: © 2023 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 flow is a special flood that flows along slopes or gullies with a large amount of sediment, debris and other solid materials mixed with water. Debris flow is also a sudden geological disaster that often occurs in mountainous areas, bringing huge losses to the people. The debris flow in Zhouqu County, Gansu Province, China in 2010 resulted in 1487 deaths, and the direct economic loss alone was up to 400 million Yuan, which was the most disastrous geological disaster in China in the past 20 years [1]. In recent years, the frequency of debris flow in China is high, which seriously threatens the safety of life and property. Zhang et al. [2] counted all 10,927 debris flow disasters in China from 2005 to 2015 and found that the peak occurrence period of debris flow disasters in China was from

May to September each year, which was the same time as the rainy season in most parts of China.

The research on debris flow susceptibility has mainly gone through three stages [3]: (1) In 1976, the United Nations entrusted the International Federation of Engineering Geology to research debris flow susceptibility [4], and since then it has gradually become an important part of the evaluation and prevention of debris flow disaster. KOVACS et al. [5] used the qualitative evaluation method to analyze debris flow susceptibility in the 1980s, which provided ideas for debris flow susceptibility assessment. (2) With the successful application of mathematical statistical methods and the 3S technology [6] in the susceptibility analysis of debris flow, debris flow susceptibility assessment gradually developed towards a comprehensive quantitative evaluation. These methods transition from qualitative to semi-quantitative and quantitative, and have been applied for a long time with a wide range of applications. The commonly used methods include the certainty factor (CF) [7], the analytic hierarchy process (AHP) [8], the information content model (ICM) [9], the principal component analysis (PCA) [10] and the frequency ratio method (FR) [11]. Li et al. [12], based on the analytic hierarchy process and principal component analysis method for elevation, combined eight influencing factors, such as slope, to give weight, and then the results of debris flow susceptibility assessment in Mentougou District of Beijing are obtained by the information content model. Li et al. [13] compared debris flow susceptibility in Pinggu District of Beijing by using the certainty factor (CF), the Information value (IV) model and the logistic regression (LR) method, and the results show that the sensitivity result obtained by the logistic regression (LR) model is the most accurate. The calculation process of these methods is complex and the calculation time is long [14]. Moreover, in the case of many damage points and evaluation factors, the above methods cannot gain advantages in the calculation of complex and big data. (3) With the rapid development of computer technology, more intelligent algorithms were put forward and widely used [15]. Among them, the machine learning algorithms were very widely used, and many scholars had applied them to the study and prediction of debris flow susceptibility. The commonly used machine learning algorithms were: support vector machine (SVM) [16], logistic regression (LR) [17], random forests (RF) [18], decision tree (DT) [19] and artificial neural network (ANN) [20]. Qiu et al. [21] improved and obtained more accurate GA-SVM and CF-GA-SVM based on SVM. By comparing with SVM and LR models, debris flow susceptibility in the Jalong corridor area was researched. The results showed that the SVM model has higher accuracy than the LR model in the case of small sample data, and the improved SVM model has the most accurate prediction result. Ke et al. [22] used LR, SVM, RF and enhanced regression tree (RT) models to research debris flow susceptibility in Sichuan Province, China. The results showed that the enhanced RT model was more accurate than other models in prediction. Compared with mathematical statistical methods, machine learning methods were faster to calculate, capable of self-organizing learning, and more accurate when predicting and evaluating debris flow susceptibility [23]. Ahmad et al. [24] used Logistic Regression (LR), Shannon Entropy (SE), Weights-of-Evidence (WoE), and Frequency Ratio (FR) models to assess the geohazards susceptibility along the Upper Indus Basin; the results showed that the Logistic Regression model performed the best of all the models. Pal et al. [25] selected an ensemble of Bayesian generalized linear model (BGLM), sparse partial least squares (SPLS), boosted tree (BT), and random forest (RF) algorithms to evaluate debris flow and landslide hazards in Iran. Ciccacese et al. [26] put forward Frequency Ratio [FR], Weight of Evidence [WOE] and Logistic Regression [LR] models to map the debris flow hazard in Italy. Vianello et al. [27] used the Rock Engineering System (RES) method for debris flow susceptibility mapping in the Western Italian Alps.

Although there are many debris flow susceptibility assessment models, how to choose the model suitable for the study area is the most important [28]. Based on field investigation, remote sensing interpretation and historical data, the catchment unit was selected as the basic mapping unit in this research. After reading a large amount of scholarly literature on

debris flow susceptibility assessment, rainfall, elevation, slope and another 11 factors were chosen to establish the FR, ACA-PCA-FR, SVM and LR models, and finally the debris flow susceptibility mapping results were obtained. The results show that the SVM model has the best performance among the four models, and the AUC value of the SVM model is the largest, which is 0.889. The result verifies that the SVM model has strong adaptability to small simple data and is very suitable for the evaluation of geological disasters. Finally, according to the results of the SVM model, the study area is divided into five regions, which are very low, low, moderate, high and very high zones, and accounted for 22.31%, 25.04%, 17.66%, 18.85% and 16.14%, respectively, of the total study area. Combined with the actual investigation and relevant data, the susceptibility assessment results obtained in this paper are reliable and can provide better help for disaster prevention and reduction in this region.

2. Study Area

The study area is located in the mountainous area of Changping District, which is located in the northwest of Beijing. The southernmost part of Changping district is only 40 km away from downtown Beijing. Changping District gradually forms a gentle slope zone from northwest to southeast, with mountains and half mountains in the west and north, and plains in the southeast, as presented in Figure 1 [29]. Changping District has a warm-temperate, semi-humid continental monsoon climate. Due to climate influence, the distribution of rainfall time is uneven in the study area. The average annual rainfall is 584.2 mm and is concentrated in summer. The rainfall from June to August is 443.3 mm, accounting for 76% of the rainfall. The temperature and rainfall show similar change laws. The average temperature of Changping District for many years has been 11.5–11.8 °C. The coldest month is January, with an average temperature of −4.1 °C. The hottest month is July, with an average temperature of 25.7 °C. Dolomite, granite, andesite, sandstone and alluvial diluvium are mainly exposed in the study area, as presented in Figure 2. The strength of rock mass is affected by fault and weathering, and broken rock blocks are produced. According to relevant data records, a large amount of rainfall in a short period is the main cause of debris flow outbreak in the study area, and geological and geomorphic conditions are the basic conditions for the debris flow outbreak.

According to historical debris flow gully data, 43 debris flow gullies were investigated in this study, with a maximum area of 5.79 km² and a minimum area of 0.10 km². In the field survey, the vegetation in the study area is relatively lush. There are loose rocks in the gully, mainly from the broken rocks on both sides of the debris flow gully. There are several collapses on the slopes on both sides. There are some villages scattered in the study area, and the population density is relatively low. However, a quarry was found in the survey, which needs attention. The basic situation of debris flow gullies during the field investigation is presented in Figure 3. The basic data of 43 debris flow gullies can be found in Table A1 of Appendix A.

According to incomplete statistics, from 1950 to 1999, there were more than 200 debris flows in 29 times in Beijing, which killed 515 people and destroyed more than 8200 houses, as presented in Figure 1. On average, a disastrous debris flow occurred every 1.8 years. From 21 to 22 July 2012, most regions in China suffered heavy rain, including Beijing and its surrounding areas, which suffered the strongest rainstorm and flood disaster in 61 years. According to survey data, the rainstorm caused 79 deaths, 10,660 houses collapsed, 1.602 million people were affected, and the economic loss was 11.64 billion yuan [30].

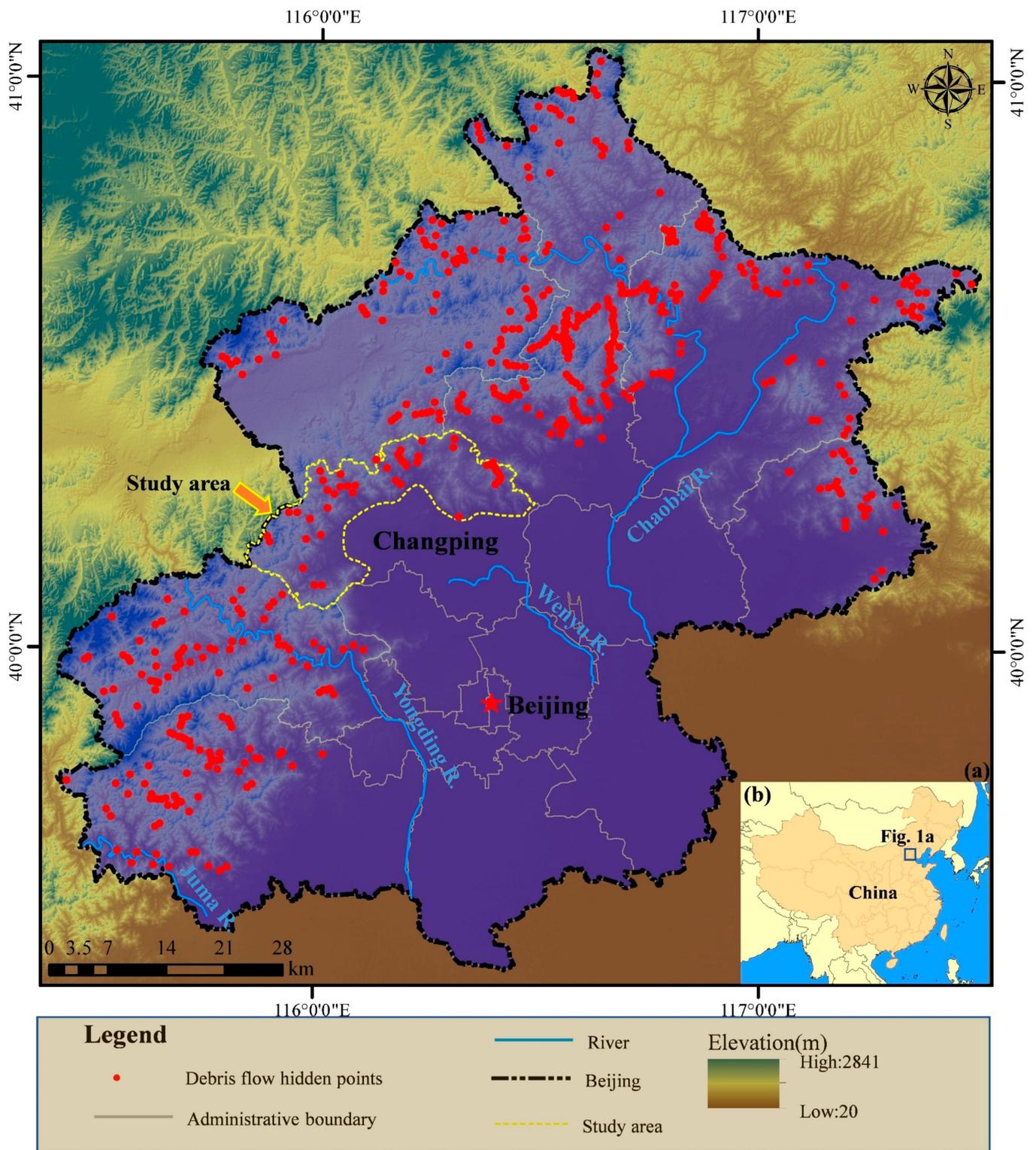


Figure 1. Debris flow distribution in Beijing and geographical location of study area. (a) The location of the research area in the whole of Beijing; (b) The location of the research area in the whole of China.

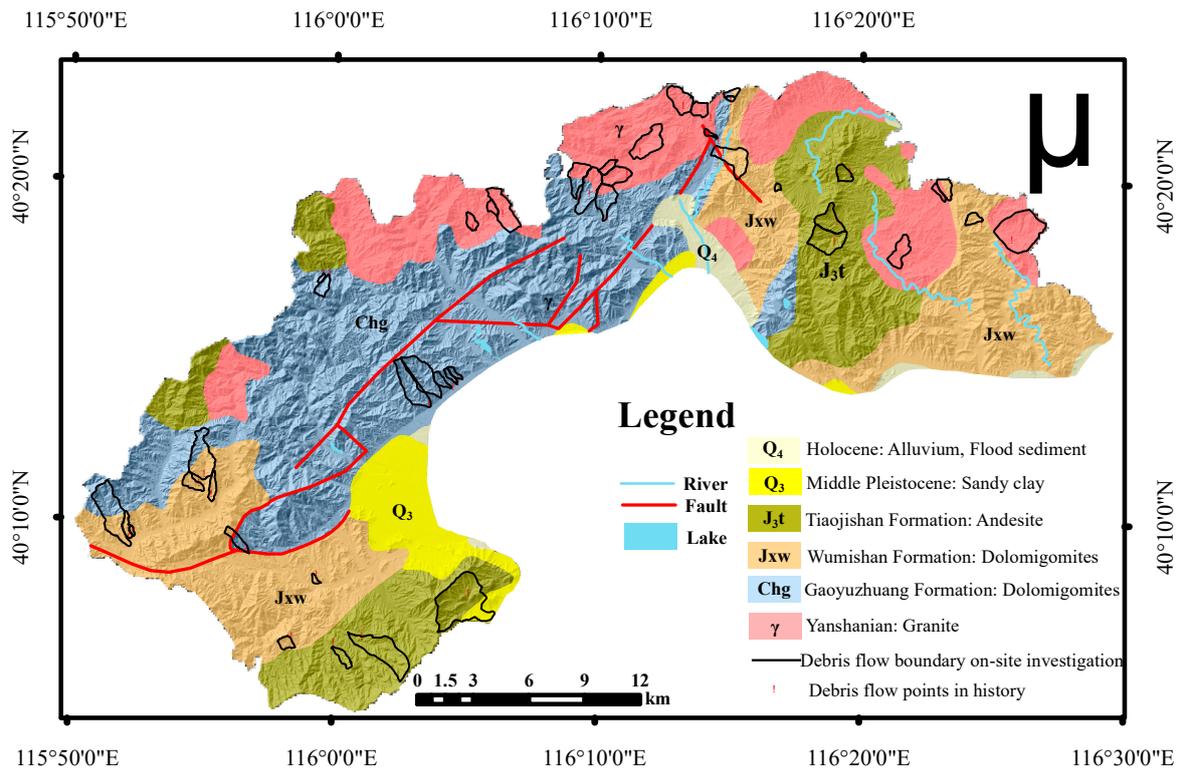


Figure 2. The geological and tectonic map of the study area.



Figure 3. Field investigation in the study area. (a) Waste broken slag; (b) loose material source in the channel; (c) water erosion in the channel; (d) small collapse on both sides of the channel.

3. Materials and Methods

3.1. Modeling Flow Chart

The debris flow susceptibility assessment mainly includes four processes: debris flow extraction and cataloging, model building, susceptibility mapping and result verifi-

ation [31]. Among them, the establishment of the debris flow susceptibility assessment model is very important, which is directly related to the scientific reliability of debris flow assessment results. According to the historical debris flow disaster point data, this paper investigated 43 known debris flow gullies in detail, based on field surveys and the 3S technique, and also used the Borderline-SMOTE algorithm to generate 43 non-debris flow trenches [32], and established four models, FR, ACA-PCA-FR, SVM and LR, respectively. Then, the susceptibility assessment result chart was obtained according to the established model. Finally, the model was compared and verified by receiver-operating characteristic curve (ROC) and area under curve (AUC) values. The modeling flow chart of this research is presented in Figure 4.

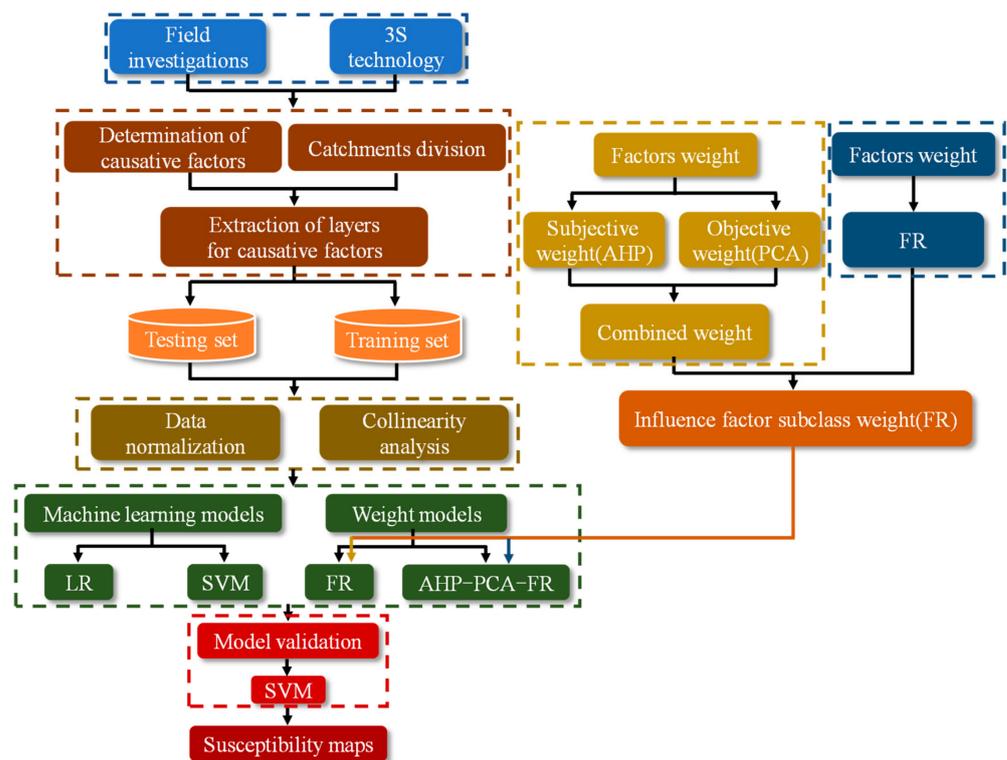


Figure 4. The modeling flow chart of this research.

3.2. Mapping Unit

The previous research on debris flow susceptibility mostly takes the valley watershed as the natural catchment basin [33]. Considering that the formation area and circulation area of debris flow in the study area are not distinguished, and the distribution of material sources is relatively wide, the use of the catchment unit as the evaluation unit can flexibly reflect the regional hydrological characteristics and surface topography by adjusting the water collection, and the evaluation results cover the whole area, with a wider application range [34]. At present, in the study of debris flow, relevant scholars have selected catchment units and achieved good results [35]. Thus, this research selects the catchment unit as the basic mapping unit. The main steps to extract the catchment units in the study area are as follows: (1) Carrying out hydrological analysis on DEM data and dividing the catchment units according to different catchment volumes; (2) Comparing the terrain conditions and combining with the field investigation, finding the most practical extraction threshold; (3) Through manual comparison and identification, adjacent units are merged to obtain the modified division results. In this study, 5000, 10,000, 15,000 and 20,000 were, respectively, set as the thresholds for water-catchment extraction. After comparing the results, it was found that the watershed units divided when the threshold was 10,000 were the most reasonable. Therefore, 10,000 was selected as the extraction threshold, and the study area was, finally, divided into 273 catchment units.

3.3. Determination of Causative Factors

The number of factors that affect the development of debris flow is large and complex, and the factors are interrelated and interact with each other [36]. The causative factors of debris flow in this paper are mainly based on the distribution and development status of debris flow disasters in the study area. After referring to a large number of susceptibility assessment documents, 11 impact factors are selected, including elevation (H), slope, plane curvature (Pl.Cv), profile curvature (Pr.Cv), roundness (Rd), geomorphic information entropy (GIE), TWI, SPI, TCI, NDVI and rainfall (Rf). The elevation data is from the digital elevation data (DEM) of 91 Weitu software; with a resolution of $7.3 \text{ m} \times 7.3 \text{ m}$, the elevation directly reflects the topographic relief changes in the study area, and to a certain extent, reflects the changes in valleys, vegetation and the state of deposits [37]. The slope is one of the important factors affecting the occurrence of debris flow. In general, the probability of debris flow occurring in the area with a steep terrain slope is greater than that in the area with a gentle terrain slope. Secondly, the terrain slope also affects the distribution, migration and accumulation of material sources on the slope, as well as the erosion of runoff on the slope [38]. The slope can be extracted from DEM. Pl.Cv and Pr.Cv represent the concavity and convexity of the terrain surface, which indirectly affect the development range of debris flow [39]. When the Pl.Cv value is positive, it indicates that the terrain surface is convex upwards, while negative values indicate that the terrain is concave downwards. When the value is 0, the surface is horizontal. Pr.Cv is opposite to the plane curvature. Pl.Cv and Pr.Cv can be extracted from DEM. Roundness can be used to reflect the shape characteristics of debris flow basins [40]. It refers to the ratio of the area of a basin to the area of a circle with the same circumference, which can be calculated by the following equation:

$$Rd = \frac{4\pi A}{L^2} \quad (1)$$

where Rd is the roundness of the watershed; A is the area of each catchment unit; and L is the length of each catchment unit. In 1899, Davis et al. [41] put forward the theory of the geomorphic cycle and believed that the evolution process of the surface could be divided into three stages: initial uplift stage, weathering erosion stage and stable stage. These three stages were defined as the young, middle-aged and old age of watershed evolution, respectively. Ai et al. [42] combined analogical information entropy and the Straler integral theory, and proposed the calculation method of geomorphic information entropy GIE :

$$S = \frac{G_{mean} - G_{min}}{G_{max} - G_{min}} \quad (2)$$

$$GIE = S - \ln S - 1 \quad (3)$$

where S is the Straler integral value. G_{mean} is the average elevation of the small watershed; and G_{max} and G_{min} are the highest and lowest elevation values of the small watershed, respectively. The value of GIE is the same as the Straler integral value, which can reflect the development stage of the watershed. The larger the value is, the more stable the basin is; otherwise, the more intense the erosion is. When $GIE < 0.111$, the watershed development stage is juvenile; when $0.111 < GIE < 0.400$, the development stage of the watershed is the prime period; when $GIE > 0.400$, the watershed development stage is the old age stage. The topographic wetness index (TWI) is defined as the theoretical measurement of flow accumulation and soil moisture at a point in the watershed [43]. Topographic moisture indices are commonly used to quantify topographic control over hydrological processes and as indicators of soil conditions and sediment and material accumulation. Therefore, debris flow susceptibility can be estimated as a function of the response of topography to hydrology in a region. Beven and Kirkby et al. [44] (1979) proposed the following formula to calculate TWI :

$$TWI = \ln\left(\frac{A_s}{\tan\alpha}\right) \quad (4)$$

$$SPI = A_s \times \tan S \quad (5)$$

$$TCI = k \times \ln A_s \quad (6)$$

where A_s is the specific catchment area; S is slope Angle; k is curvature; and TWI , SPI and TCI are dimensionless exponents. The stream power index (SPI) reflects the erosion ability of water flow to soil [45]. The higher the SPI value, the more the runoff concentration may lead to soil erosion, and the greater the probability of debris flow. The topographic characteristic index (TCI) mainly represents the migration ability of the river, to reflect the possibility of debris flow [46]. Rainfall is the most important factor for the formation of debris flow in the study area [47]. According to the historical debris flow disaster data, the occurrence of debris flow in the study area is mainly caused by continuous heavy rainfall in a short period. Therefore, this paper obtained the 24 h maximum rainfall data of the study area through Beijing Changping hydrological stations. Then, the rainfall factor map was obtained by the Kriging interpolation method in the Arcgis10.5. The Normalized Difference Vegetation Index ($NDVI$) reflected the size of the vegetation cover density in the study area [48]. The larger the $NDVI$ value, the better the vegetation on the surface, which can effectively preserve the soil and water, so the more unfavorable it is to the occurrence of debris flow. The $NDVI$ data used in this paper were acquired from Landsat8 OLI images with a resolution of 30 m. The spatial distribution of impact factors in this study area is presented in Figure 5.

3.4. Methods

3.4.1. Analytic Hierarchy Process (AHP)

The analytic hierarchy process (AHP) carries out the multi-criteria decision analysis method of relative importance calculation by decomposing all relevant elements in the whole process of weighting [49]. At present, the AHP method has been widely used in geological hazard mapping. When the AHP method is used to evaluate the risk of geological disasters, the corresponding hierarchical model should be established first. Then, the importance of the two influencing factors is compared subjectively and a judgment matrix is established. The judgment matrix can be expressed as:

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \quad (7)$$

where A is the judgment matrix; a_{ij} is the result of the comparison between the importance of factor i and factor j , which has the following characteristics:

$$a_{ij} = \frac{1}{a_{ji}} \quad (8)$$

The relative importance of each factor can be rated on a scale of 1 to 9, from less important to more important. Finally, the following formula is used to check the consistency:

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

$$CR = CI / RI \quad (10)$$

where CR is the random consistency ratio, when $CR < 0.1$, the consistency test passes; CI is the consistency index; n is the order of the judgment matrix; λ_{max} is the largest eigenvalue of the judgment matrix. RI is the average random consistency index, as shown in Table 1.

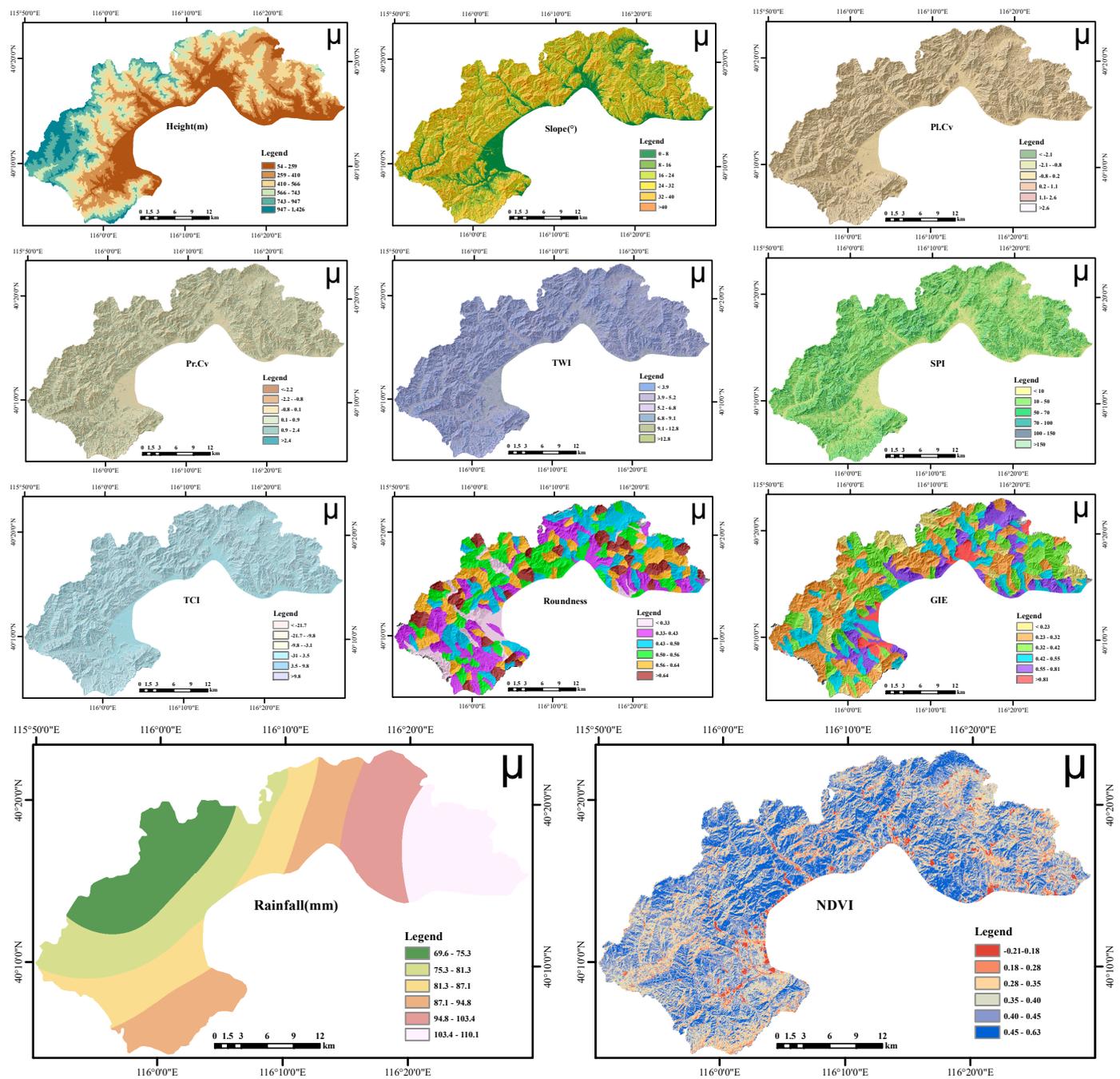


Figure 5. Maps showing causative factors in this study area with GIS.

Table 1. Average random consistency indicator value table.

n	1	2	3	4	5	6	7	8	9	10	11
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49	1.51

3.4.2. Frequency Ratio (FR)

Frequency ratio (FR) is a binary statistical method used to calculate the probability relationship between the dependent variable and the independent variable, and it is a simple and effective zoning model for the debris flow hazard. Secondly, the frequency ratio method can also be used for multi-classification spatial data prediction [50]. This method determines the correlation between the hazard degree of debris flow and different

evaluation indices by calculating the FR value. When the FR value is greater than 1, it indicates that the correlation between the debris flow risk degree and the evaluation index is high; when the FR value is less than 1, it indicates that the correlation between the debris flow risk degree and the evaluation index is small. The FR value can be calculated according to the following formula:

$$FR = \frac{n/N}{m/M} \quad (11)$$

where n is the number of debris flow in the subinterval of the evaluation index; N is the total number of debris flow in the study area; m is the number of grids occupied by each evaluation index subinterval; and M is the total number of grids in the study area.

3.4.3. Principal Component Analysis (PCA)

In the research of practical problems, people always hope to use the factors involved in the problem as much as possible, in order to avoid the problem of unreliable analysis results due to insufficient information. However, the more complicated the consideration problem becomes, the higher the requirement for calculation and modeling analysis. This can also lead to a large amount of redundant information leading to unreliable results, so before establishing a model for data preprocessing, people remove the correlation between sample characteristic dimensions. PCA is a common method of data preprocessing, and it is used to process high-dimensional data in many fields [51]. When processing high-dimensional data, it can be represented by fewer principal components, and the dimensionality of sample data can be reduced under the premise of little data loss, to greatly simplify the cost of data calculation and research.

3.4.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) is based on statistical learning theory [52]. It has the following advantages: (1) low requirements of data volume; (2) strong generalization ability; (3) strong adaptability to high-dimensional samples; and (4) strong learning ability and fast convergence. By transforming each evaluation index from low-dimensional space to high-dimensional space, it realizes the linear segmentation of data. By this method, nonlinear problems in low-dimensional space can be analyzed and evaluated. SVM is widely used in the research of debris flow hazard zoning because of its low requirement for data volume [53–59].

3.4.5. Logistic Regression (LR)

The main idea of the logistic regression (LR) model is to determine the possibility of debris flow in the future after converting each factor into a logical variable [60–65]. Logistic regression uses the maximum likelihood to find the best-fit function. The simplified logistic regression method is shown as follows:

$$P = \frac{1}{(1 + e^{-y})} \quad (12)$$

where P is the sensitivity index of debris flow, whose value is between 0 and 1. The larger the p -value, the greater the possibility of debris flow. y is the weighted sum of each evaluation index, which can be expressed as follows:

$$y = c_0 + c_1x_1 + c_2x_2 \dots + c_jx_j \quad (13)$$

where c_0 is a constant term, c_1, c_2, \dots, c_j is the logistic regression coefficient; x_1, x_2, \dots, x_j is the value of each evaluation index.

4. Result Analysis

4.1. Calculation Results of Weights

The main step of the ACA-PCA-FR model is to obtain the weight of 11 influencing factors through the combination of AHP and PCA. FR is used to obtain the weight of each subclass in each influencing factor. Finally, the susceptibility map is obtained by stacking all factor maps. Based on the historical data and field investigation results of the research area, this paper ranked the 11 influencing factors in order of importance, $R_f > H > \text{slope} > \text{Pl.Cv} > \text{Pr.Cv} > \text{Rd} > \text{GIE} > \text{TWI} > \text{TCI} > \text{SPI} > \text{NDVI}$. The established judgment matrix is shown in Table 2.

Table 2. Judgment matrix of impact factors.

Factors	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	Weight
X ₁	1	2	3	4	4	5	6	5	6	7	3	0.260
X ₂	1/2	1	2	3	3	4	2	2	3	3	3	0.154
X ₃	1/3	1/2	1	2	2	3	2	3	3	4	2	0.118
X ₄	1/4	1/3	1/2	1	2	2	3	2	3	2	3	0.095
X ₅	1/4	1/3	1/2	1/2	1	2	3	3	2	3	2	0.084
X ₆	1/5	1/4	1/3	1/2	1/2	1	2	2	3	2	3	0.066
X ₇	1/6	1/2	1/2	1/3	1/3	1/2	1	2	3	3	2	0.060
X ₈	1/5	1/2	1/3	1/2	1/3	1/2	1/2	1	2	3	4	0.055
X ₉	1/6	1/3	1/3	1/3	1/2	1/3	1/3	1/2	1	4	3	0.044
X ₁₀	1/7	1/3	1/4	1/2	1/3	1/2	1/3	1/3	1/4	1	2	0.031
X ₁₁	1/3	1/3	1/2	1/3	1/2	1/3	1/2	1/4	1/3	1/2	1	0.033

X₁ = R_f, X₂ = H, X₃ = slope, X₄ = Pl.Cv, X₅ = Pr.Cv, X₆ = Rd, X₇ = GIE, X₈ = TWI, X₉ = TCI, X₁₀ = SPI, X₁₁ = NDVI.

Through calculations, the consistency index CI = 0.1208 and the consistency ratio CR = 0.0795 were obtained. Since CR < 0.10, the consistency of the judgment matrix A was acceptable.

The objective weights can be obtained by AHP. The data of 11 influencing factors of 273 water catchment units in the study area were imported into the SPSS software for principal component analysis, and the weights of 11 influencing factors were calculated by the software, as shown in Table 3. Finally, the combined weight value is obtained by averaging the weights obtained by the two methods. It can be seen from the results that the combined weight is more reasonable than the weight value obtained by AHP and PCA, as shown in Figures 6 and 7.

Table 3. Results of combined weights of each impact factor.

Factors	Weight		
	AHP	PCA	Average
R _f	0.260	0.083	0.172
H	0.154	0.101	0.127
Slope	0.118	0.091	0.104
Pl.Cv	0.095	0.082	0.088
Pr.Cv	0.084	0.079	0.082
Rd	0.066	0.092	0.079
GIE	0.060	0.086	0.073
TWI	0.055	0.105	0.080
TCI	0.044	0.099	0.072
SPI	0.031	0.088	0.059
NDVI	0.033	0.093	0.063

4.2. Distribution of Debris Flow in Different Classes of Factors

In the process of FR and ACA-PCA-FR modeling, it is necessary to count the ratio of the debris flow grid number in each category and the ratio of each category to the total grid number in each influencing factor. In this paper, each influencing factor is divided into

six categories according to the natural discontinuous method, and the frequency ratio of each subcategory of each influencing factor is obtained through the distribution of 43 debris flow trenches investigated in the field, as shown in Figure 8 and Table 4. The FR model is used to make a statistical calculation of the frequency ratio method for 11 evaluation factors by using the probability statistics method. Then, it makes a superposition of the evaluation factors to get the debris flow sensitivity susceptibility map.

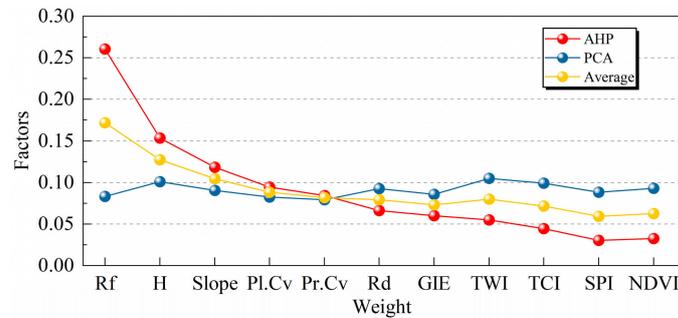


Figure 6. Results of combined weights of each impact factor.

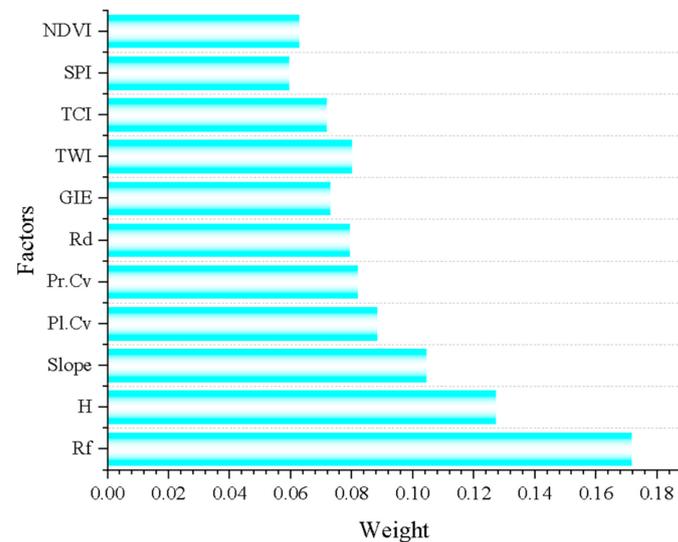


Figure 7. Weights of each impact factor.

4.3. Correlation Analysis

Before the susceptibility assessment of debris flow, many studies carried out independent tests for the evaluation indices selected. However, many susceptibility assessment models are sensitive to the multicollinearity of evaluation indices, such as the LR model. The linear correlation of each evaluation index will increase the prediction error of debris flow susceptibility assessment. Some studies use independent tests to test the mutual independence of each evaluation index [66]. For example, by using the variance inflation factor (VIF) and the conditional independence test [67], evaluation indices with high linear correlation can be proposed. However, compared with these methods, the PCA method can also eliminate the multicollinearity problem among evaluation indices, and has less information loss on the original evaluation indices, that is, the principal components obtained by this method are obtained through all the pre-selected price indices instead of reducing the dimension by directly eliminating some evaluation indicators. Therefore, the PCA method was used to reduce the dimension of the pre-selected evaluation indices, and the re-selected evaluation indices were transformed to make them independent of each other. Then, the sensitivity zoning of debris flow is carried out by using the evaluation index after the dimensionality reduction to eliminate the influence of linear correlation

among factors on the prediction results. The correlation matrix of impact factors is shown in Table 5.

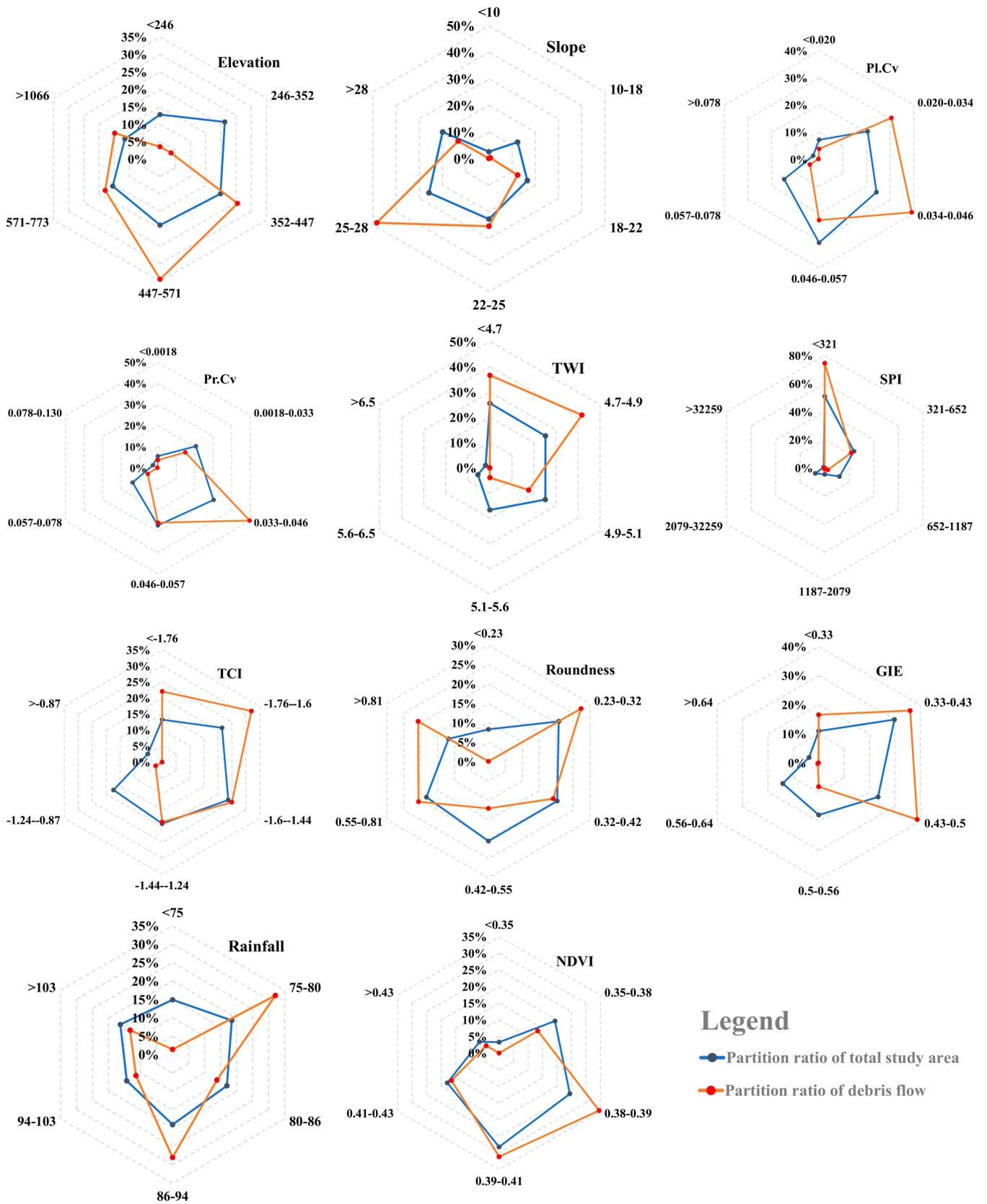


Figure 8. Distribution of debris flow in different classes of factors.

Table 4. The frequency ratio of different classes of influence factors.

Factor	Class	Study Area		Debris Flows Area		FR
		Count	Ratio (%)	Count	Ratio (%)	
Elevation (m)	<246	1,763,582	12.77	32,373	3.54	0.278
	246–352	2,943,221	21.30	32,857	3.60	0.169
	352–447	2,749,383	19.90	232,606	25.47	1.280
	447–571	2,615,987	18.94	314,927	34.48	1.821
	571–773	2,147,022	15.54	164,597	18.02	1.160
	>1066	1,596,180	11.55	136,007	14.89	1.289
Slope (°)	<10	355,183	2.57	328	0.04	0.014
	10–18	1,706,124	12.35	5503	0.60	0.049
	18–22	2,289,905	16.58	113,069	12.38	0.747
	22–25	3,142,477	22.75	232,494	25.45	1.119
	25–28	3,564,604	25.80	441,176	48.30	1.872
	>28	2,757,082	19.96	120,797	13.23	0.663
Pl.Cv	<0.020	989,053	7.16	34,787	3.81	0.532
	0.020–0.034	2,834,110	20.51	278,744	30.52	1.488
	0.034–0.046	3,350,592	24.25	357,437	39.13	1.614
	0.046–0.057	4,254,670	30.80	204,989	22.44	0.729
	0.057–0.078	2,036,610	14.74	35,319	3.87	0.262
	>0.078	350,340	2.54	2091	0.23	0.090
Pr.Cv	<0.0018	776,090	5.62	34,546	3.78	0.673
	0.0018–0.033	2,850,849	20.64	135,740	14.86	0.720
	0.033–0.046	4,174,347	30.22	454,510	49.76	1.647
	0.046–0.057	3,751,047	27.15	235,962	25.83	0.951
	0.057–0.078	1,892,812	13.70	50,186	5.49	0.401
	>0.078	370,230	2.68	2423	0.27	0.099
TWI	<4.7	3,530,258	25.55	334,721	36.65	1.434
	4.7–4.9	3,486,820	25.24	381,479	41.77	1.655
	4.9–5.1	3,477,971	25.17	161,005	17.63	0.700
	5.1–5.6	2,304,522	16.68	35,393	3.88	0.232
	5.6–6.5	748,518	5.42	441	0.05	0.009
	>6.5	267,286	1.93	326	0.04	0.018
SPI	<321	7,044,594	50.99	681,341	74.6	1.463
	321–652	3,294,696	23.85	197,750	21.65	0.908
	652–1187	1,661,894	12.03	23,440	2.57	0.213
	1187–2079	626,998	4.54	8392	0.92	0.202
	2079–32,259	1,063,258	7.70	353	0.04	0.005
	>32,259	123,935	0.90	2091	0.23	0.255
TCI	<−1.76	1,830,826	13.25	201,634	22.08	1.666
	−1.76–1.6	2,956,309	21.40	291,421	31.91	1.491
	−1.6–1.44	3,267,552	23.65	228,035	24.97	1.056
	−1.44–1.24	2,652,379	19.20	170,634	18.68	0.973
	−1.24–0.87	2,401,896	17.39	21,315	2.33	0.134
	>−0.87	706,413	5.11	328	0.04	0.007
Roundness	<0.23	1,149,635	8.32	576	0.06	0.008
	0.23–0.32	2,869,935	20.77	249,892	27.36	1.317
	0.32–0.42	2,807,727	20.32	174,583	19.11	0.941
	0.42–0.55	2,830,499	20.49	110,037	12.05	0.588
	0.55–0.81	2,530,300	18.32	188,650	20.65	1.128
	>0.81	1,627,279	11.78	189,629	20.76	1.763
GIE	<0.33	1,518,871	10.99	151,473	16.58	1.508
	0.33–0.43	4,122,934	29.84	328,981	36.02	1.207
	0.43–0.5	3,229,952	23.38	354,525	38.82	1.660
	0.5–0.56	2,467,237	17.86	74,420	8.15	0.456
	0.56–0.64	1,953,781	14.14	3199	0.35	0.025
	>0.64	522,600	3.78	769	0.08	0.022

Table 4. Cont.

Factor	Class	Study Area		Debris Flows Area		FR
		Count	Ratio (%)	Count	Ratio (%)	
Rainfall (mm)	<74.5	2,053,685	14.87	12,687	1.39	0.093
	74.5–80.4	2,559,075	18.52	293,259	32.11	1.733
	80.4–86.1	2,344,335	16.97	126,493	13.85	0.816
	86.1–94.2	2,629,605	19.03	255,501	27.97	1.470
	94.2–103.4	1,974,843	14.29	104,323	11.42	0.799
	>103.4	2,253,822	16.31	121,104	13.26	0.813
NDVI	<0.35	454,500	3.29	0	0	0.000
	0.35–0.38	2,669,162	19.32	122,156	13.37	0.692
	0.38–0.39	3,380,910	24.47	315,819	34.58	1.413
	0.39–0.41	3,899,555	28.23	284,587	31.16	1.104
	0.41–0.43	2,474,058	17.91	150,116	16.44	0.918
	>0.43	937,190	6.78	40,689	4.45	0.657

Table 5. Parameter correlation matrix.

Factors	GIE	H	NDVI	Pl.Cv	Pr.Cv	Slope	Rf	SPI	TCI	TWI	Rd
GIE	1.000	−0.550	−0.317	0.199	0.188	−0.727	0.191	0.166	0.658	0.696	−0.257
H	−0.550	1.000	0.155	−0.170	−0.139	0.603	−0.535	−0.115	−0.488	−0.524	0.210
NDVI	−0.317	0.155	1.000	−0.029	−0.041	0.378	−0.277	−0.116	−0.383	−0.432	0.116
Pl.Cv	0.199	−0.170	−0.029	1.000	0.980	−0.112	0.133	0.277	−0.019	−0.040	−0.032
Pr.Cv	0.188	−0.139	−0.041	0.980	1.000	−0.089	0.105	0.311	−0.046	−0.056	−0.024
Slope	−0.727	0.603	0.378	−0.112	−0.089	1.000	−0.299	−0.005	−0.821	−0.904	0.388
Rf	0.191	−0.535	−0.277	0.133	0.105	−0.299	1.000	−0.143	0.183	0.130	−0.041
SPI	0.166	−0.115	−0.116	0.277	0.311	−0.005	−0.143	1.000	0.078	0.132	−0.021
TCI	0.658	−0.488	−0.383	−0.019	−0.046	−0.821	0.183	0.078	1.000	0.923	−0.300
TWI	0.696	−0.524	−0.432	−0.040	−0.056	−0.904	0.130	0.132	0.923	1.000	−0.362
Rd	−0.257	0.210	0.116	−0.032	−0.024	0.388	−0.041	−0.021	−0.300	−0.362	1.000

5. Discussion

In this paper, the natural break point method was used to classify the susceptibility assessment results of debris flow obtained by the four models into five susceptibility areas, which were very low, low, moderate, high and very high susceptibility areas. Finally, the susceptibility assessment map was obtained, as shown in Figure 9. The percentage of the five susceptibility areas in the total area can be obtained by extraction, as shown in Figure 10. As can be seen from the figure, the susceptibility results of FR and ACA-PCA-FR are very similar. On the whole, the areas of high and very high susceptibility areas in the two methods are large, accounting for about 50% of the total area. The 43 debris flow gullies in the field survey are all located in the highly sensitive area, which is reasonable.

SVM and LR are two machine learning languages widely used in the evaluation of geological hazards. In this study, the results obtained by SVM and LR models are similar. On the whole, the areas of high and very high susceptibility areas in these two methods are significantly smaller than the traditional weight assignment methods, accounting for 34.99% and 29.95% of the total area, respectively. However, the 43 debris flow gullies in the field survey are located in the high susceptibility area, so it is also reasonable.

However, in the process of field investigation, the vegetation in the study area is very dense, and the sources in the gullies are relatively few. There are relatively more sources in the gullies with a large area. Therefore, from the perspective of field investigation, the results obtained by SVM and LR models are more reasonable than the sensitivity results obtained by FR and ACA-PCA-FR models. In order to further verify the accuracy of the four sensitivity models, ROC curves were drawn, and AUC values were calculated for the susceptibility results obtained from the four models.

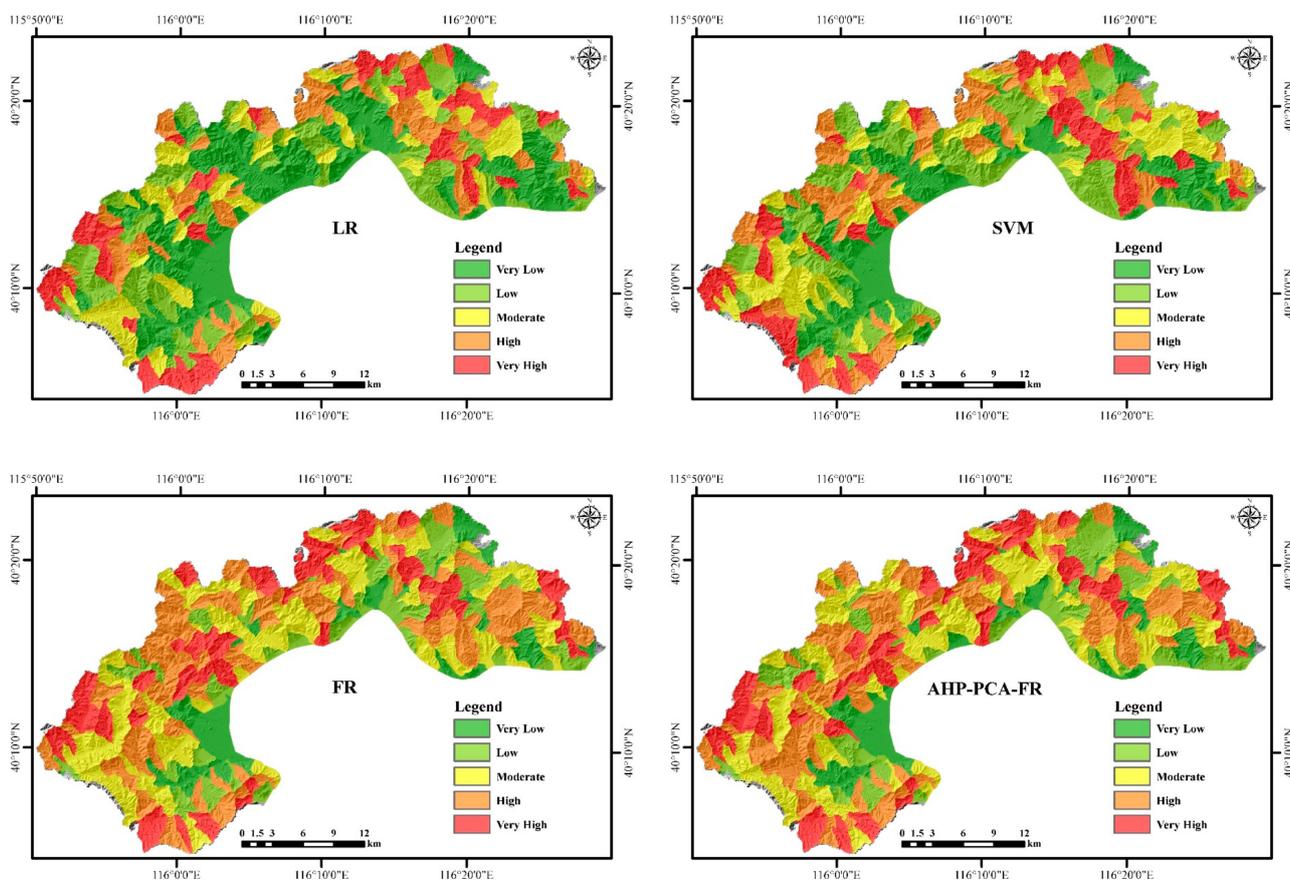


Figure 9. Debris flow susceptibility maps produced by four models.

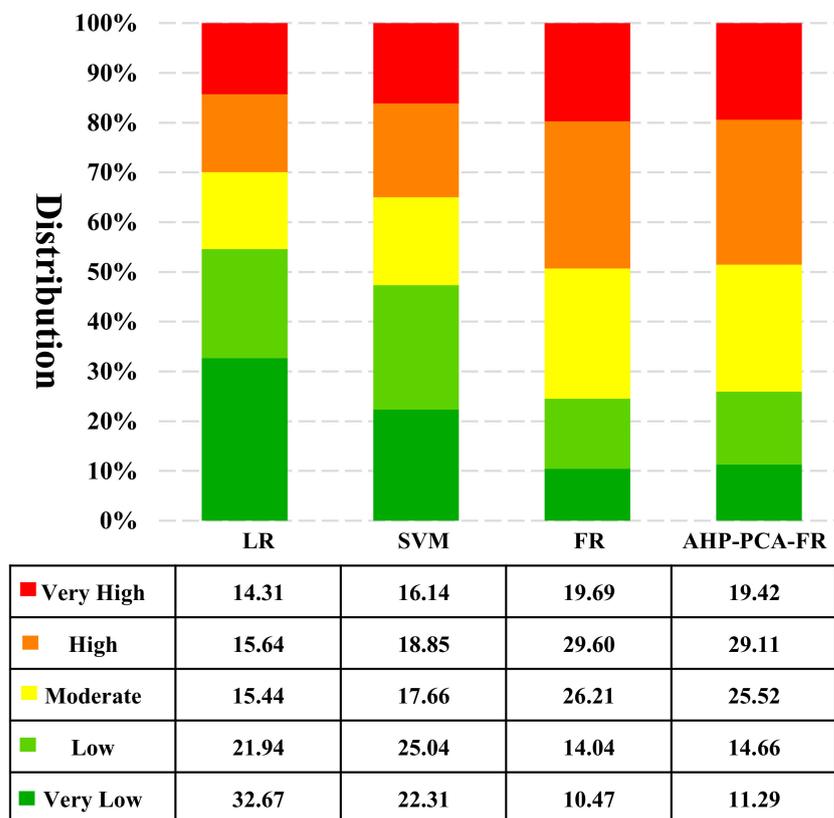


Figure 10. Distribution of the different debris flow susceptibility classes from the four models.

The ROC curve, also known as the receiver characteristic curve, is widely used in medical laboratories and disease prediction [68]. The individual of each disaster point and its evaluation factor subgroup is equivalent to the subject in the medical laboratory. The occurrence and non-occurrence of debris flow are taken as two categories (positive category and negative category) for analysis [69]. The ROC curve and the area under the curve (AUC) corresponding to each mode were obtained. AUC is the standard to judge the advantages and disadvantages of the four models used. When AUC = 0.5, it means that the model results have no reference value, while when it is less than 0.5, it means that the model does not conform to the real situation. When AUC is greater than 0.5, the closer its value is to 1, the more accurate the model effect is.

The ROC curve and AUC value drawn by the four models are shown in Figure 11. It can be seen from Figure 11 that the prediction accuracy of the SVM model is better than the other three models, with the largest AUC value of 0.889, followed by LR (AUC = 0.842), ACA-PCA-FR (AUC = 0.829), FR (AUC = 0.797). The results show that SVM has low requirements on the amount of data and can still maintain very high accuracy in the case of a small sample data, which is very suitable for the evaluation of geological disasters. Meanwhile, the AUC values of the SVM and LR models are larger than those of LR and ACA-PCA-FR, indicating that the performance of the machine learning algorithm is significantly better than that of the traditional weight method. In addition, the main difference between FR and ACA-PCA-FR is that when determining the weight of each influence factor, the FR model adopts the frequency ratio method, while ACA-PCA-FR determines the weight through the combination of the subjective weight determined by AHP and objective weight determined by PCA. According to the AUC value, it can be concluded that the method of subjective and objective combination weighting is better than the method of objective weighting.

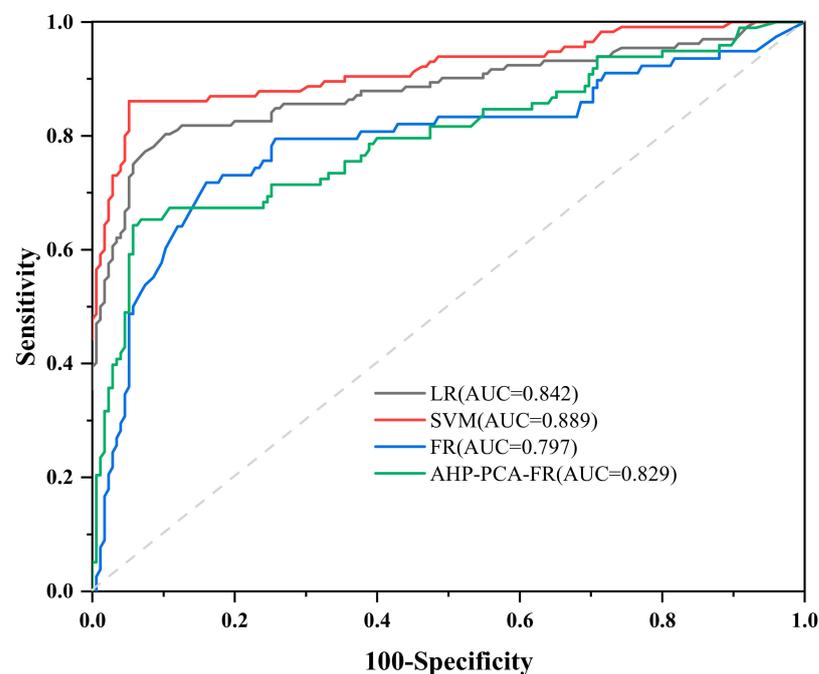


Figure 11. The ROC curves of four debris flow susceptibility models.

According to the above results, the susceptibility results obtained from the SVM model were taken as the final evaluation results, among which the very low, low, moderate, high and very high susceptibility zones accounted for 22.31%, 25.04%, 17.66%, 18.85% and 16.14%, respectively, of the total area of the study area.

The main advantages of logical regression are simple implementation, wide application, small amount of calculation, fast speed, low storage resources, and the ability to solve multicollinearity problems. The main disadvantage is that it is easy to be under-fitted, and

the accuracy is generally not high. It can not handle a large number of multi-class features or variables well. For nonlinear features, it needs to be converted.

The main advantages of the support vector machine are high classification accuracy, accurate classification in small sample size, good generalization ability, and easy to solve nonlinear problems, classification and regression problems of high-dimensional features. The main disadvantage is that there is no universal and effective standard for the selection of kernel functions of support vector machines. At the same time, because there is no reliable sample selection mode, the sample selection method also has a great impact on the evaluation results.

Compared with LR model, SVM model has better accuracy and reliability, and also has better classification ability: SVM tends to divide more watershed units into unstable units, and the over-prediction characteristics of SVM should be fully considered in practical work; however, LR is relatively conservative, and tends to predict more basin units as stable units, and it is easier to divide dangerous areas into stable areas. Therefore, the characteristics of the two models should be fully considered in practical work. Both models have certain limitations, and the single model has defects in varying degrees. In the study area, for the advantages, disadvantages and application conditions of different models, using multiple models to solve problems will also be a problem that needs further study in the future.

Comparing this paper with the research of Ahmad et al. [24], the theme of the two articles is to compare with machine learning and mathematical statistics methods to assess debris flow susceptibility. Among them, there are FR model and LR model in both articles, and seven of the influence factors used are the same. Both articles have concluded that the machine learning model is better than the mathematical statistics model, which also verifies that the conclusion of this paper is indeed correct. The difference is that this paper also makes a deeper comparative study between different machine learning models and different mathematical statistical models, in order to better apply them to the evaluation of geological hazards.

6. Conclusions

Based on field investigations and the 3S technology, this research selects 11 influence factors, such as elevation and slope, and selects the catchment unit that is suitable as the basic mapping unit. By setting different flow thresholds for comparative research, the study area is finally divided into 273 catchment units. Then, four models, R, ACA-PCA-FR, LR and SVM, are established, respectively. In the LR and SVM models, 43 debris flow gullies surveyed in the field are taken as known debris flows. The same number of 43 non-debris flow gullies were obtained based on the Borderline-SMOTE algorithm. These two kinds of data were substituted into the model as test data. Finally, the sensitivity evaluation results of the study area were obtained, and the results were verified by the ROC curve and AUC value, and the following conclusions were drawn:

1. Among the four models, the SVM model has the best performance and the highest prediction accuracy, with AUC = 0.889, followed by LR (AUC = 0.842), ACA-PCA-FR (AUC = 0.829) and FR (AUC = 0.797). The results show that SVM can still maintain very high prediction accuracy in the case of small sample data, learning can be strong and have a fast convergence, and has strong adaptability to high-dimensional samples, which is very suitable for the evaluation and analysis of geological disasters.
2. Among the four models, the results of the FR and ACA-PCA-FR models are relatively similar. These two methods are traditional weight evaluation methods. According to the field survey results and AUC values, the accuracy of these two methods is relatively low. The results of LR and SVM, as two widely-used machine learning algorithms, are similar, more consistent with the field survey results, and the AUC value is relatively high, so in this study, the machine learning algorithm is more accurate and reasonable.

At the same time, this paper also has some shortcomings. Because it is difficult to obtain historical data and the number of known debris flow gullies is small, this directly

affects the reliability of machine learning algorithms. However, there is no doubt that more reasonable and efficient algorithms can be put forward to assess debris flow susceptibility in the future.

Author Contributions: Formal analysis, F.G.; funding acquisition, J.C.; methodology, F.G.; project administration, J.C.; software, Y.L.; supervision, Q.W.; validation, F.G.; visualization, Y.Z.; writing—original draft, F.G.; writing—review and editing, J.C., X.S. and Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the key project of the National Natural Science Foundation of China (Grant no. U1702241).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank anonymous reviewers for their comments and suggestions, which helped to improve this paper.

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

Appendix A

The basic statistics of 43 debris flow gullies have been given in Table A1.

Table A1. The basic statistics of 43 debris flow gullies.

No.	H (m)	Slope (°)	Pl.Cv	Pr.Cv	TWI	SPI	TCI	Rd	GIE	Rf	NDVI
1	243.65	20.13	0.04	0.18	5.04	51.76	−1.55	0.42	0.66	85.36	0.41
2	505.38	25.92	0.05	0.04	4.63	219.43	−1.78	0.37	0.42	89.65	0.38
3	1038.43	25.57	0.04	0.06	4.67	206.57	−1.84	0.43	0.21	75.51	0.39
4	602.15	26.64	0.03	0.06	4.70	381.37	−1.90	0.46	0.39	90.60	0.38
5	931.01	26.22	0.04	0.05	4.71	135.54	−1.74	0.43	0.21	77.37	0.40
6	749.67	30.09	0.05	0.00	4.56	219.89	−1.58	0.42	0.47	80.98	0.44
7	1008.40	25.66	0.01	0.01	4.84	273.88	−1.46	0.43	0.21	77.23	0.41
8	967.57	27.50	0.03	0.06	4.77	286.41	−1.65	0.41	0.31	80.01	0.41
9	869.67	26.78	0.03	0.06	4.72	156.73	−1.74	0.45	0.52	80.43	0.41
10	784.33	26.03	0.00	0.04	4.48	80.38	−2.18	0.45	0.53	80.58	0.41
11	437.44	24.66	−0.02	0.09	4.81	119.62	−2.06	0.56	0.47	87.57	0.46
12	269.94	28.29	−0.01	0.08	4.52	148.31	−2.30	0.68	0.46	79.67	0.48
13	263.18	27.44	−0.06	0.04	4.54	153.34	−2.46	0.43	0.45	79.76	0.48
14	418.28	27.99	0.06	−0.01	4.54	309.73	−1.56	0.38	0.36	78.21	0.44
15	738.46	26.25	−0.02	0.09	4.70	160.45	−2.17	0.63	0.19	75.40	0.46
16	820.86	27.17	0.04	0.02	4.64	187.79	−1.68	0.63	0.19	76.35	0.41
17	546.49	27.14	0.00	0.00	4.54	123.96	−1.85	0.39	0.36	74.29	0.46
18	392.30	27.96	0.01	0.07	4.72	295.73	−1.97	0.46	0.35	78.19	0.43
19	307.44	28.72	−0.02	0.07	4.49	171.67	−2.53	0.68	0.46	79.39	0.46
20	630.33	27.45	0.03	0.06	4.70	213.02	−1.87	0.50	0.25	88.02	0.42
21	388.86	26.87	0.02	0.09	4.68	174.03	−1.96	0.43	0.39	83.89	0.43
22	565.62	29.21	0.02	0.01	4.64	568.68	−2.01	0.55	0.25	70.20	0.35
23	473.61	27.67	0.04	0.06	4.74	265.07	−1.73	0.56	0.25	85.87	0.38
24	532.11	31.21	0.02	−0.01	4.64	196.51	−1.52	0.41	0.16	82.79	0.42
25	389.12	23.22	0.05	0.06	4.73	163.41	−1.84	0.59	0.25	83.82	0.42
26	432.29	23.28	0.05	0.05	4.88	249.88	−1.75	0.57	0.33	82.15	0.40
27	598.21	26.25	−0.04	0.01	4.69	97.54	−1.89	0.45	0.29	92.20	0.36
28	532.19	27.01	0.04	0.08	4.72	246.06	−1.84	0.36	0.33	80.85	0.41
29	397.58	29.74	−0.13	0.14	4.47	137.77	−3.18	0.45	0.36	91.34	0.40
30	564.05	29.76	−0.11	0.05	4.42	108.60	−2.72	0.45	0.29	92.24	0.34
31	411.35	25.69	0.03	0.07	4.77	320.24	−1.93	0.55	0.32	94.10	0.38
32	525.73	25.93	0.07	0.08	4.70	140.00	−1.75	0.45	0.29	89.90	0.41

Table A1. Cont.

No.	H (m)	Slope (°)	Pl.Cv	Pr.Cv	TWI	SPI	TCI	Rd	GIE	Rf	NDVI
33	410.43	15.58	0.07	0.04	5.21	81.75	−1.08	0.54	0.34	106.72	0.41
34	404.58	13.64	0.08	0.12	5.39	138.10	−1.26	0.48	0.34	106.59	0.41
35	461.08	25.49	0.02	0.11	4.70	123.40	−2.14	0.52	0.30	102.13	0.40
36	508.76	26.66	0.03	0.06	4.66	261.97	−1.91	0.60	0.19	102.29	0.39
37	453.55	23.81	0.02	0.04	4.97	223.28	−1.41	0.72	0.38	109.66	0.37
38	655.98	28.49	0.00	−0.02	4.67	222.83	−1.62	0.55	0.25	70.36	0.36
39	379.86	23.35	0.06	0.13	4.82	148.32	−1.74	0.55	0.34	108.49	0.40
40	387.34	20.13	0.06	0.05	4.99	117.33	−1.16	0.58	0.38	105.98	0.41
41	470.47	24.17	0.01	0.07	4.80	162.16	−1.93	0.69	0.36	102.31	0.40
42	411.80	24.66	0.02	0.03	5.03	319.65	−1.32	0.62	0.30	90.49	0.39
43	413.66	20.88	0.01	0.11	4.91	68.87	−1.87	0.70	0.32	98.69	0.43

References

- Chong, Y.; Chen, G.; Meng, X.; Yang, Y.; Shi, W.; Bian, S.; Zhang, Y.; Yue, D. Quantitative analysis of artificial dam failure effects on debris flows—A case study of the Zhouqu '8.8' debris flow in northwestern China. *Sci. Total Environ.* **2021**, *792*, 148439. [[CrossRef](#)] [[PubMed](#)]
- Zhang, S.; Zhang, L.M.; Chen, H.X. Relationships among three repeated large-scale debris flows at Pubugou Ravine in the Wenchuan earthquake zone. *Can. Geotech. J.* **2014**, *51*, 951–965. [[CrossRef](#)]
- Ma, C.; Deng, J.Y.; Wang, R. Analysis of the triggering conditions and erosion of a runoff-triggered debris flow in Miyun County, Beijing, China. *Landslides* **2018**, *15*, 2475–2485. [[CrossRef](#)]
- Dormann, C.F.; Elith, J.; Bacher, S.; Buchmann, C.; Carl, G.; Carré, G.; Marquéz, J.R.G.; Gruber, B.; Lafourcade, B.; Leitão, P.J.; et al. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **2013**, *36*, 27–46. [[CrossRef](#)]
- Robert, H. Soil slumps and debris flows: Prediction and Protection. *Bull. Assoc. Eng. Geol.* **1981**, *18*, 17–28.
- Feizizadeh, B.; Roodposhti, M.S.; Blaschke, T.; Aryal, J. Comparing GIS-based support vector machine kernel functions for landslide susceptibility mapping. *Arab. J. Geosci.* **2017**, *10*, 1–13. [[CrossRef](#)]
- Devkota, K.C.; Regmi, A.D.; Pourghasemi, H.R.; Yoshida, K.; Pradhan, B.; Ryu, I.C.; Dhital, M.R.; Althuwaynee, O.F. Landslide susceptibility mapping using certainty factor, index of entropy and logistic regression models in GIS and their comparison at Mugling-Narayanghat road section in Nepal Himalaya. *Nat. Hazards* **2013**, *65*, 135–165. [[CrossRef](#)]
- Chen, X.; Chen, H.; You, Y.; Liu, J. Susceptibility assessment of debris flows using the analytic hierarchy process method—a case study in Subao river valley, China. *J. Rock Mech. Geotech. Eng.* **2015**, *7*, 404–410. [[CrossRef](#)]
- Xu, W.; Yu, W.; Jing, S.; Zhang, G.; Huang, J. Debris flow susceptibility assessment by GIS and information value model in a large-scale region, Sichuan Province (China). *Nat. Hazards* **2013**, *65*, 1379–1392. [[CrossRef](#)]
- Shi, M.Y.; Chen, J.P.; Song, Y.; Zhang, W.; Song, S.Y.; Zhang, X.D. Assessing debris flow susceptibility in Heshigten Banner, Inner Mongolia, China, using principal component analysis and an improved fuzzy C-means algorithm. *Bull. Eng. Geol. Environ.* **2016**, *75*, 909–922. [[CrossRef](#)]
- Lee, S.; Sambath, T. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environ. Geol.* **2006**, *50*, 847–855. [[CrossRef](#)]
- Li, Z.H.; Chen, J.P.; Tan, C.; Zhou, X.; Li, Y.C.; Han, M.X. Debris flow susceptibility assessment based on topo-hydrological factors at different unit scales: A case study of Mentougou district, Beijing. *Environ. Earth Sci.* **2021**, *80*, 365. [[CrossRef](#)]
- Li, Y.C.; Chen, J.P.; Tan, C.; Li, Y.; Gu, F.F.; Zhang, Y.W.; Mehmood, Q. Application of the borderline-SMOTE method in susceptibility assessments of debris flows in Pinggu District, Beijing, China. *Nat. Hazards* **2021**, *105*, 2499–2522. [[CrossRef](#)]
- Kang, S.H.; Lee, S.R. Debris flow susceptibility assessment based on an empirical approach in the central region of South Korea. *Geomorphology* **2018**, *308*, 1–12. [[CrossRef](#)]
- Liang, Z.; Wang, C.M.; Zhang, Z.M. A comparison of statistical and machine learning methods for debris flow susceptibility mapping. *Stoch. Environ. Res. Risk A* **2020**, *34*, 1887–1907. [[CrossRef](#)]
- Lin, G.F.; Chang, M.J.; Huang, Y.C.; Ho, J.Y. Assessment of susceptibility to rainfall induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. *Eng. Geol.* **2017**, *224*, 62–74. [[CrossRef](#)]
- Cao, C.; Zhang, W.; Chen, J.P.; Shan, B.; Song, S.Y.; Zhan, J.W. Quantitative estimation of debris flow source materials by integrating multi-source data: A case study. *Eng. Geol.* **2021**, *291*, 106222. [[CrossRef](#)]
- Sun, X.H.; Chen, J.P.; Bao, Y.D.; Han, X.D.; Zhan, J.W.; Peng, W. Landslide Susceptibility Mapping Using Logistic Regression Analysis along the Jinsha River and Its Tributaries Close to Derong and Deqin County, Southwestern China. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 438. [[CrossRef](#)]
- McSherry, D. Strategic induction of decision trees. *Knowl. Based Syst.* **1988**, *12*, 269–275. [[CrossRef](#)]

20. Sun, X.H.; Chen, J.P.; Li, Y.R.; Rene, N.N. Landslide Susceptibility Mapping along a Rapidly Uplifting River Valley of the Upper Jinsha River, Southeastern Tibetan Plateau, China. *Remote Sens.* **2022**, *14*, 1730. [[CrossRef](#)]
21. Qiu, C.C.; Su, L.J.; Zou, Q.; Geng, X.Y. A hybrid machine-learning model to map glacier-related debris flow susceptibility along Gyirong Zangbo watershed under the changing climate. *Sci. Total Environ.* **2022**, *818*, 151752. [[CrossRef](#)]
22. Ke, X.; Basanta, R. Comparison of Different Machine Learning Methods for Debris Flow Susceptibility Mapping: A Case Study in the Sichuan Province, China. *Remote Sens.* **2020**, *12*, 295.
23. Elkadiri, R.; Sultan, M.; Youssef, A.M.; Elbayoumi, T.; Chase, R.; Bulkhi, A.B.; Al-Katheeri, M.M. A remote sensing-based approach for debris-flow susceptibility assessment using artificial neural networks and logistic regression modeling. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4818–4835. [[CrossRef](#)]
24. Ahmad, H.; Chen, N.S.; Rahman, M.; Islam, M.M.; Pourghasemi, H.R.; Hussain, S.F.; Habumugisha, J.M.; Liu, E.L.; Zheng, H.; Ni, H.Y.; et al. Geohazards Susceptibility Assessment along the Upper Indus Basin Using Four Machine Learning and Statistical Models. *ISPRS Int. Geo-Inf.* **2021**, *10*, 315. [[CrossRef](#)]
25. Pal, S.C.; Chakraborty, R.; Saha, A.; Bozchaloei, S.K.; Pham, Q.B.; Linh, N.T.T.; Anh, D.T.; Janizadeh, S.; Ahmadi, K. Evaluation of debris flow and landslide hazards using ensemble framework of Bayesian- and tree-based models. *Bull. Eng. Geol. Environ.* **2022**, *81*, 5. [[CrossRef](#)]
26. Ciccarese, G.; Mulas, M.; Corsini, A. Combining spatial modelling and regionalization of rainfall thresholds for debris flows hazard mapping in the Emilia-Romagna Apennines (Italy). *Landslides* **2021**, *18*, 3513–3529.
27. Vianello, D.; Vagnon, F.; Bonetto, S.; Mosca, P. Debris flow susceptibility mapping using the Rock Engineering System (RES) method: A case study. *Landslides* **2022**, 1–22. [[CrossRef](#)]
28. Cao, J.; Zhang, Z.; Du, J.; Zhang, L.; Song, Y.; Sun, G. Multi-geohazards susceptibility mapping based on machine learning—A case study in Jiuzhaigou, China. *Nat. Hazards* **2020**, *102*, 851–871. [[CrossRef](#)]
29. Ma, C.; Wang, Y.J.; Du, C.; Wang, Y.Q.; Li, Y.P. Variation in initiation condition of debris flow in the mountain regions surrounding Beijing. *Geomorphology* **2016**, *273*, 323–334. [[CrossRef](#)]
30. Li, Y.; Chen, J.; Zhang, Y.; Song, S.; Han, X.; Ammar, M. Debris flow susceptibility assessment and runout prediction: A case study in Shiyang Gully, Beijing, China. *Int J. Environ. Res.* **2020**, *14*, 365–383. [[CrossRef](#)]
31. Cheng, W.M.; Wang, N.; Zhao, M.; Zhao, S.M. Relative tectonics and debris flow hazards in the Beijing mountain area from DEM-derived geomorphic indices and drainage analysis. *Geomorphology* **2016**, *257*, 134–142. [[CrossRef](#)]
32. Han, H.; Wang, W.Y.; Mao, B.H. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. *Lect. Notes Comput. Sci.* **2005**, *3644*, 878–887.
33. Qin, S.; Lv, J.; Cao, C.; Ma, Z.; Hu, X.; Liu, F.; Qiao, S.; Dou, Q. Mapping debris flow susceptibility based on watershed unit and grid cell unit: A comparison study. *Geomat. Nat. Hazards Risk* **2019**, *10*, 1648–1666. [[CrossRef](#)]
34. Ayalew, L.; Yamagishi, H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **2005**, *65*, 15–31. [[CrossRef](#)]
35. Tiranti, D.; Cavalli, M.; Crema, S.; Zerbato, M.; Graziadei, M.; Barbero, S.; Cremonini, R.; Silvestro, C.; Bodrato, G.; Tresso, F. Semi-quantitative method for the assessment of debris supply from slopes to river in ungauged catchments. *Sci. Total Environ.* **2016**, *554*, 337–348. [[CrossRef](#)]
36. Zhang, Y.; Ge, T.; Tian, W.; Liou, Y.-A. Debris flow susceptibility mapping using machine-learning techniques in Shigatse Area, China. *Remote Sens.* **2019**, *11*, 2801. [[CrossRef](#)]
37. Yilmaz, I. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: Conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ. Earth Sci.* **2010**, *61*, 821–836. [[CrossRef](#)]
38. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification*; Springer: Boston, MA, USA, 2016; pp. 207–235.
39. Xu, C.; Dai, F.; Xu, X.; Lee, Y.H. GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China. *Geomorphology* **2012**, *145–146*, 70–80. [[CrossRef](#)]
40. Li, T.; Qiu, S.; Mao, S.X.; Bao, R.; Deng, H.B. Evaluating water resource accessibility in Southwest China. *Water* **2019**, *11*, 1708. [[CrossRef](#)]
41. Wang, Q.; Guo, Y.; Li, W.; He, J.; Wu, Z. Predictive modeling of landslide hazards in Wen County, northwestern China based on information value, weights-of-evidence, and certainty factor. *Geomat. Nat. Hazards Risk* **2019**, *10*, 820–835. [[CrossRef](#)]
42. Ballabio, C.; Sterlacchini, S. Support vector machines for landslide susceptibility mapping: The Staffora river basin case study, Italy. *Math. Geosci.* **2012**, *44*, 47–70. [[CrossRef](#)]
43. Moore, I.D.; Grayson, R.; Ladson, A. Digital terrain modeling: A review of hydrological, geomorphological, and biological applications. *Hydrol Process* **1991**, *5*, 3–30. [[CrossRef](#)]
44. Beven, K.J.; Kirkby, M.J. A physically based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [[CrossRef](#)]
45. Wilson, J.P.; Gallant, J.C. Digital Terrain Analysis. In *Terrain Analysis: Principles and Applications*; Wilson, J.P., Gallant, J.C., Eds.; Wiley: New York, NY, USA, 2000; pp. 1–27.
46. Park, S.J.; McSweeney, K.; Lowery, B. Identification of the spatial distribution of soils using a process-based terrain characterization. *Geoderma* **2001**, *103*, 249–272. [[CrossRef](#)]

47. Tang, C.; Zhu, J.; Li, W.L. Rainfall-triggered debris flows following the Wenchuan earthquake. *Bull. Eng. Geol. Environ.* **2009**, *68*, 187–194. [[CrossRef](#)]
48. Tehrany, M.S.; Pradhan, B.; Mansor, S.; Ahmad, N. Flood susceptibility assessment using GIS based support vector machine model with different kernel types. *Catena* **2015**, *125*, 91–101. [[CrossRef](#)]
49. Saaty, T.L. Modeling unstructured decision problems—The theory of analytical hierarchies. *Math. Comput. Simul.* **1978**, *20*, 147–158. [[CrossRef](#)]
50. Chen, W.; Wang, J.; Xie, X.; Hong, H.; Van Trung, N.; Bui, D.T.; Wang, G.; Li, X. Spatial prediction of landslide susceptibility using integrated frequency ratio with entropy and support vector machines by different kernel functions. *Environ. Earth Sci.* **2016**, *75*, 1344. [[CrossRef](#)]
51. Sabokbar, H.F.; Roodposhti, M.S.; Tazik, E. Landslide susceptibility mapping using geographically-weighted principal component analysis. *Geomorphology* **2014**, *226*, 15–24. [[CrossRef](#)]
52. Zhou, C.; Yin, K.; Cao, Y.; Ahmed, B. Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. *Eng. Geol.* **2016**, *204*, 108–120. [[CrossRef](#)]
53. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [[CrossRef](#)]
54. Zhang, X.; Wu, Y.; Zhai, E.; Ye, P. Coupling analysis of the heat-water dynamics and frozen depth in a seasonally frozen zone. *J. Hydrol.* **2021**, *593*, 125603. [[CrossRef](#)]
55. Wu, X.; Ren, F.; Niu, R. Landslide susceptibility assessment using object mapping units, decision tree, and support vector machine models in the Three Gorges of China. *Environ. Earth Sci.* **2013**, *71*, 4725–4738. [[CrossRef](#)]
56. Bui, D.T.; Tuan, T.A.; Hoang, N.D.; Thanh, N.Q.; Nguyen, D.B.; Van Liem, N.; Pradhan, B. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. *Landslides* **2016**, *14*, 447–458.
57. Bui, D.T.; Pradhan, B.; Lofman, O.; Revhaug, I. Landslide susceptibility assessment in Vietnam using support vector machines, decision tree, and Naïve Bayes models. *Math. Probl. Eng.* **2012**, *2012*, 1–26.
58. Zhang, X.; Zhai, E.; Wu, Y.; Sun, D.; Lu, Y. Theoretical and Numerical Analyses on Hydro–Thermal–Salt–Mechanical Interaction of Unsaturated Salinized Soil Subjected to Typical Unidirectional Freezing Process. *Int. J. Geomech.* **2021**, *21*, 04021104. [[CrossRef](#)]
59. Pradhan, B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
60. Umar, Z.; Pradhan, B.; Ahmad, A.; Jebur, M.N.; Tehrany, M.S. Earthquake induced landslide susceptibility mapping using an integrated ensemble frequency ratio and logistic regression models in West Sumatera Province, Indonesia. *Catena* **2014**, *118*, 124–135. [[CrossRef](#)]
61. Bui, D.T.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2015**, *13*, 361–378.
62. Pradhan, B.; Lee, S. Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* **2010**, *25*, 747–759. [[CrossRef](#)]
63. Pham, B.T.; Bui, D.T.; Prakash, I.; Nguyen, L.H.; Dholakia, M.B. A comparative study of sequential minimal optimization-based support vector machines, vote feature intervals, and logistic regression in landslide susceptibility assessment using GIS. *Environ. Earth Sci.* **2017**, *76*, 371. [[CrossRef](#)]
64. Zhang, X.; Ye, P.; Wu, Y. Enhanced technology for sewage sludge advanced dewatering from an engineering practice perspective: A review. *J. Environ. Manag.* **2022**, *321*, 115938. [[CrossRef](#)] [[PubMed](#)]
65. Meng, Q.; Miao, F.; Zhen, J.; Wang, X.; Wang, A.; Peng, Y.; Fan, Q. GIS-based landslide susceptibility mapping with logistic regression, analytical hierarchy process, and combined fuzzy and support vector machine methods: A case study from Wolong Giant Panda Natural Reserve, China. *Bull. Eng. Geol. Environ.* **2015**, *75*, 923–944. [[CrossRef](#)]
66. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženilek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [[CrossRef](#)]
67. Mather, P.M. The use of backpropagating artificial neural networks in land cover classification. *Int. J. Remote Sens.* **2003**, *24*, 4907–4938.
68. Li, L.; Lan, H.; Guo, C.; Zhang, Y.; Li, Q.; Wu, Y. A modified frequency ratio method for landslide susceptibility assessment. *Landslides* **2017**, *14*, 727–741. [[CrossRef](#)]
69. Kavzoglu, T.; Sahin, E.K.; Colkesen, I. Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* **2013**, *11*, 425–439. [[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.