



# Influence of buffer distance on environmental geological hazard susceptibility assessment

Zepeng Wang<sup>1,2,3,4</sup> · Jianping Chen<sup>2</sup> · Zechen Lian<sup>5</sup> · Fei Li<sup>6</sup> · Lu Pang<sup>2</sup> · Yabo Xin<sup>2</sup>

Received: 23 April 2023 / Accepted: 22 December 2023 / Published online: 9 January 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

Previous researches seldom studied the selection of buffer distance between geological hazards (positive samples) and non-geological hazards (negative samples), and its reasonable selection plays a very important role in improving the accuracy of susceptibility zoning, protecting the environment and reducing the cost of hazard management. Based on GIS technology and random forest (RF) and frequency-ratio random forest (FR-RF) models, this study innovatively explored the influence of randomly selected non-geological hazard samples outside different buffer distances on the susceptibility evaluation results, with buffer distances of 100 m, 500 m, 1000 m and 2000 m in sequence. The results show that through the confusion matrix and ROC curve test, the accuracy of the model increases first and then decreases with the increase of buffer distance. Both RF and FR-RF models have the highest accuracy when the buffer distance is 1000 m, and the accuracy of the RF model is generally higher than that of the FR-RF model under the same buffer distance. Similar attribute values of positive samples and randomly selected negative samples or “extreme” attribute values of negative samples are the main reasons for the differences in evaluation results of different buffer distances. According to the weight analysis of causative factors, the distance from road, the distance from river and the normalized vegetation index (NDVI) are the main factors affecting the occurrence of hazards. The high and very high susceptibility areas in the study area are mainly distributed on both sides of roads and water systems, which are the key areas for hazard prevention and reduction. The HMC of RF-1000m decreased by 3.55% on average compared with other models. The results of this study improve the accuracy of geological hazard susceptibility assessment, maintain the safety of ecological environment, and provide a scientific basis for the selection of buffer distance index in local and surrounding areas in the future.

**Keywords** Geological hazard · Buffer distance · Negative sample · Frequency ratio · Random forest · Environmental safety

Responsible Editor: Philippe Garrigues

✉ Jianping Chen  
13804181164@139.com

- <sup>1</sup> Coal Geological Geophysical Exploration Surveying & Mapping Institute of Shanxi Province, JinZhong 030600, China
- <sup>2</sup> College of Mining, Liaoning Technical University, Fuxin 123000, China
- <sup>3</sup> Key Laboratory of Monitoring and Protection of Natural Resources in Mining Cities, Ministry of Natural Resources, Jinzhong 030600, China
- <sup>4</sup> Shanxi Provincial Key Lab of Resources, Environment and Disaster Monitoring, Jinzhong 030600, China
- <sup>5</sup> Shanxi Huajin Geotechnical Engineering CO., LTD, Taiyuan 030000, China
- <sup>6</sup> Inner Mongolia Shendong Tianlong Group CO., LTD, Ordos 017000, China

## Introduction

With the rapid development of society and economy, geological hazards caused by human engineering construction have increased significantly (Froude and Petley 2018; Zhao et al. 2021). Geological hazards are characterized by complexity, suddenness and inevitability (Yang and Hu 2019). Geological hazards in China are widely distributed and occur frequently, bringing huge economic losses every year. Therefore, it is of great significance to construct a more accurate geological hazard susceptibility model and to partition the study area into geological hazard susceptibility zones in a more reasonable way, so as to reduce the hazard losses and protect the ecological environment (Jiang et al. 2017).

In the work of geological hazard susceptibility assessment, an empirical model (such as analytic hierarchy model) was first used (Pawluszek and Borkowski 2017) and

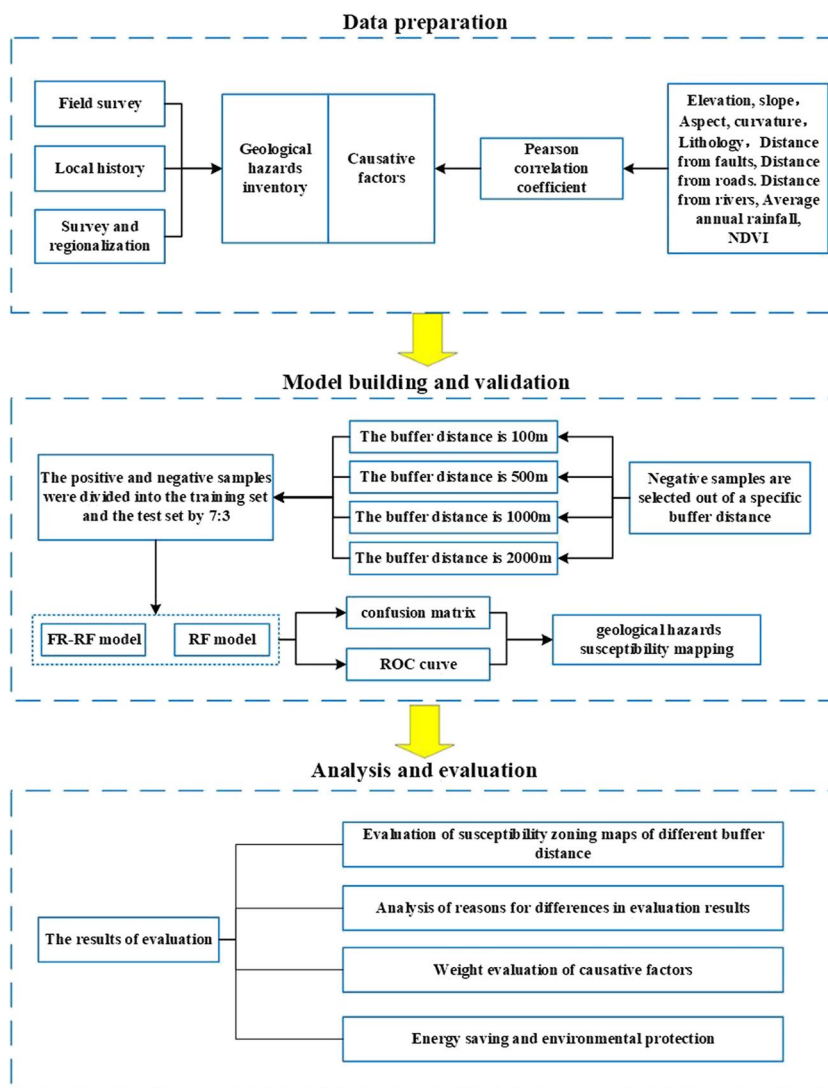
achieved a certain prediction effect. However, this type of model is susceptible to subjective factors, which reduces the evaluation accuracy of the model. With the development of GIS and other technologies, statistical analysis models have been applied to geological hazard susceptibility models (such as information value and frequency ratio model) (Ba et al. 2017; Du et al. 2019; Khan et al. 2019; Nicu 2017; Sarda and Pandey 2019) and have led to a certain improvement in evaluation accuracy. However, this type of model lacks uniform standards and criteria in the hierarchical classification of causative factors, and the prediction results vary considerably (Aditian et al. 2018; Wang et al. 2014). With the rapid development of computers, machine learning models are widely used in the evaluation of geological hazard susceptibility (such as random forest and support vector machine model) (Akinci et al. 2020; Huang and Zhao 2018; Zhao et al. 2020). And the predictive accuracy of the models was greatly improved, generally higher than the first two types of models. At the same time, researchers also found that the most suitable evaluation models are different for different study areas. Therefore, it has become a trend for researchers to use two or more models for comparison in the evaluation of geological hazard susceptibility. In the process of continuous application of machine learning models, the hybrid model composed of a single model has become a new trend in geological hazard susceptibility evaluation (Chen et al. 2019), such as FT (fractal theory)-IV-RF (Zhao et al. 2021), ANN-Bayes (He et al. 2012) and RS-SVM (Chang and Wan 2015). Machine learning models require positive and negative sample data in the process of using them, while previous studies mostly improve the prediction accuracy by improving machine learning models, and few studies related to positive and negative samples. Accurate positive samples can be obtained through remote sensing image interpretation or field surveys (Kalantar et al. 2018). Accurate positive samples can be obtained through remote sensing image interpretation or field investigation (Conoscenti et al. 2016; Mondini and Chang 2014; Posner and Georgakakos 2015). Negative sample acquisition methods vary, including random selection in the study area (Polykretis and Chalkias 2018; Pourghasemi and Rahmati 2018), in the low slope area (Kavzoglu et al. 2014), through special methods (Chen et al. 2021; Liu et al. 2021; Su et al. 2022) and out of the positive sample buffer (Peng et al. 2014; Su et al. 2017). All of these selection methods have yielded better prediction results. But the studies on randomly selecting samples outside the positive sample buffer are not only few, but more importantly, only one particular buffer distance is set in these studies, and multiple buffer distances are not studied. Other studies based on machine learning models also do not divide and study the buffer distance. Park et al. (2019) evaluated the susceptibility of Woomyeon Mountain in Korea with a buffer distance of 100 m. Wang set the buffer distance 500 m and evaluated

the susceptibility of geological hazards in Yunyang County (Chongqing, China) (Wang et al. 2020). Wang conducted a multi-model susceptibility evaluation for Helong City (Jilin Province, China) with a buffer distance of 1000 m (Wang et al. 2021b). However, the large variability of sample attributes selected outside different buffer distances may have a large impact on the model accuracy. Based on this problem, this paper innovatively divides the buffer distance according to the scope of the study area and the experience of previous researchers and conducts a detailed analysis and study: to determine whether the buffer distance has an impact on the model accuracy, whether the impact is large, how will the model accuracy change with buffer distance, what is the optimal buffer distance. The main objectives of this study are (1) to map geological hazard susceptibility using RF and FR-RF models with different buffer distances and analyze the distribution of susceptibility zones of different grades in the study area; (2) to analyze whether different buffer distances have an effect on the results of the susceptibility evaluation and how the ROC curves and confusion matrices change with changes in buffer distances to analyze the reasons for such differences; (3) to determine whether there is a significant impact on the model accuracy through model test and economic and environmental losses to determine the optimal buffer distance in the study area; and (4) identify the main causative factors leading to the occurrence of geological hazards in the study area.

The western part of Shanxi Province is the area where geologic hazards occur most frequently. Linfen city is a more serious area in western Shanxi province where geologic hazards occur (Zhao et al. 2016). And Pu County is the area of Linfen City where geologic hazards occur in larger numbers and are more typical (Zhao 2017). Geological hazards pose a great threat to local infrastructure and the safety of people's lives and property. According to the "14th Five-Year Plan for Geological Hazards Prevention and Control in Shanxi Province", a more accurate evaluation of the susceptibility to geological hazards is required than in the past. Therefore, the analysis of buffer distance in this area can not only improve the accuracy of the susceptibility model, but also provide a theoretical basis for the selection of negative samples in the neighbouring areas with similar terrain.

In summary, taking Puxian County as an example, the buffer distance of the study area was divided into 100 m, 500 m, 1000 m and 2000 m, and the RF and FR-RF models were used to evaluate the susceptibility to geologic hazards with different buffer distances. And the results of the study can provide theoretical basis for more accurate hazard prevention and mitigation in the study area. It can also provide a new idea for negative sample selection and a scientific basis for future geological hazard susceptibility evaluation based on machine learning models in buffer distance selection. The research process is shown in Fig. 1.

Fig. 1 Flow chart of this study



## Materials and methods

### Study area

Pu County is located in the southwestern part of Shanxi Province, at the southern end of the Lvliang Mountain Range and on the east bank of the middle reaches of the Yellow River. The county has a warm-temperate continental climate, with an average annual temperature between 8.6 and 12.6 °C and an average annual rainfall of 586 mm. The altitude of the area is shown as high in the northeast, east and south and low in the west. The large geomorphological units can be divided into folded and broken block-stripped high school mountainous area, middle mountainous area, loess plateau and hilly area and intermountain valley area. The Loess Plateau and hilly areas are distributed with a large number of slopes, which have the proximity conditions to produce avalanches and landslide deformation activities. The

terrain of the river valley area is relatively flat, and geological hazards are not developed, but the valley slopes on both sides of the transition to other landforms have steep slopes due to the erosion and cutting influence of the river, and most of them are composed of sub-soft and semi-hard rocks. These valleys and slopes provide favourable conditions for the occurrence of geological hazards due to river erosion and hollowing. Through field survey and data collection, 148 geological hazards were identified (Fig. 2). Based on ArcGIS, the study area was divided into raster units with the size of 30 × 30 m (Liu et al. 2018).

### Causative factors

Based on geological environmental conditions, geological hazard development mechanism and related assessment work in the study area, ten factors including distance from fault (DFF), distance from roads (DFR1), distance from

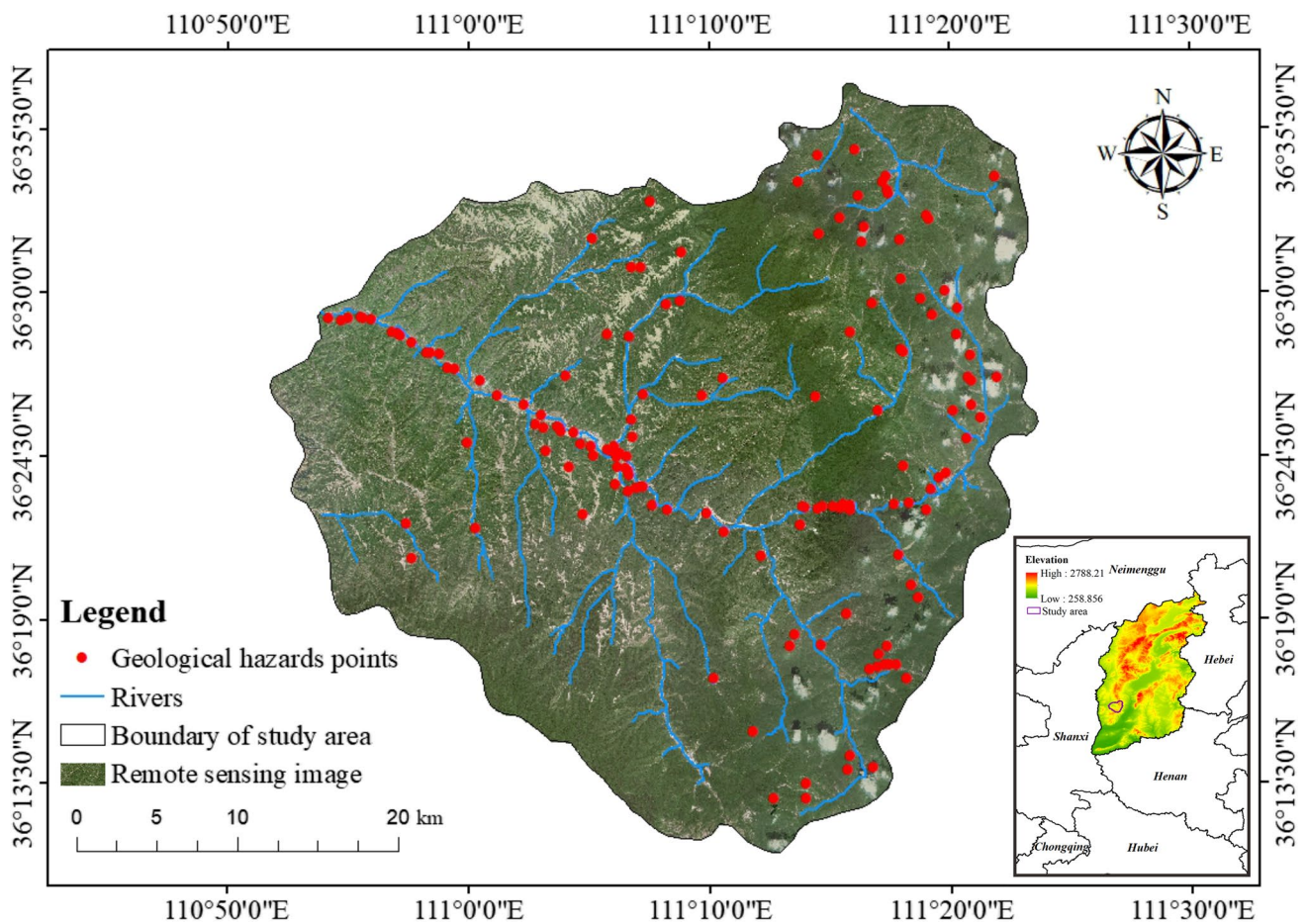


Fig. 2 Geographical location map of Pu County

ivers (DFR2), average annual rainfall (AAR) and normalized vegetation index (NDVI) were selected for correlation analysis and geological hazard susceptibility evaluation.

Elevation is the most commonly used index in geological hazard susceptibility evaluation. Potential energy is more likely to be converted into kinetic energy and geological hazards occur in areas with greater elevation changes. (Xiong et al. 2017). Slope is the embodiment of the steepness of a local area and is one of the key indexes in the evaluation of geological hazard susceptibility (Bordoni et al. 2020; Zhang et al. 2020). The distribution of light duration and intensity and rainfall infiltration was different in different aspect (Erener and Duzgun 2013). Curvature refers to the topographic form of slope, which has an indirect influence on the development range of geological hazards (Pourghasemi et al. 2018).

The basic property of lithology determines the initial state of slope and the ability to resist weathering and erosion, which is also the most important factor for geological

hazards (Lin et al. 2021; Sun et al. 2018). Geological structure movement deteriorates the mechanical properties of rock and soil mass, directly or indirectly leading to the occurrence of geological hazards (Hong et al. 2015). Human engineering construction has great influence on the slope; frequent disturbance may lead to slope instability (Kanwal et al. 2017). Geological hazards are widely distributed in the distance near or very close to the roads, and the number of geological hazards decreases rapidly as the distance becomes longer (Wang et al. 2021b).

Rainfall not only erodes slopes but also penetrates and softens rocks and soil, providing sufficient hydrodynamic conditions for the occurrence of geological hazards (Li et al. 2020). The erosion of slope surface rock and soil by water system is one of the important causes of geological hazards (Meinhardt et al. 2015). Vegetation can prevent soil erosion and help strengthen the stability of the slope. Generally, the more lush the vegetation, the less prone to slope instability (Chen et al. 2017).



## Methods

### Pearson correlation coefficient

Pearson's correlation coefficient can be calculated to evaluate the overall linear relationship (Biswas and Si 2011), usually expressed as  $r$ . It can also represent the positive and negative correlation between variables (Hu et al. 2021). In general, the higher the correlation between variables, the closer the absolute value of  $r$  is to 1. The degree of correlation between variables are divided into small, low significant correlation and the strong correlation, the  $r$  value range of the corresponding to  $|r| < 0.3$ ,  $0.3 \leq |r| < 0.5$ ,  $0.5 \leq |r| < 0.8$ ,  $|r| \geq 0.8$  (Guo 2020).

$$r = \frac{\sum_{i=1}^k (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^k (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^k (Y_i - \bar{Y})^2}} \quad (1)$$

where  $k$  is the number of samples,  $X_i, Y_i$  are the observed values of sample variables, and  $\bar{X}, \bar{Y}$  represent the mean values of samples.

### FR model

FR model is a method of mathematical calculation and analysis, and its size represents the contribution degree of factors at various levels to the occurrence of hazards after classification (Ma et al. 2022). The sum of the frequency ratios of all subcategories of causative factors is the final probability index of hazard occurrence (Aditian et al. 2018). The calculation method has been explained in previous articles (Chen et al. 2023).

$$FR_{ij} = \frac{N_{ij}}{N_r} \div \frac{A_{ij}}{A_r} \quad (2)$$

where  $FR_{ij}$  is the frequency ratio;  $N_{ij}$  is the number of geological hazards grids at the  $j$ th level under the  $i$ th influencing factor;  $N_r$  is the grid number of geological hazards in the study area;  $A_{ij}$  is the number of regional grids of the  $j$ th level under the  $i$ th influencing factor;  $A_r$  is the total number of grids in the study area.

### RF model

The random forest, originally developed by Breiman (Breiman 2001; Nun et al. 2014), is a multi-variable classification machine learning model (Catani et al. 2013). Due to variance and other problems, a single DT shows a weak prediction effect (Taalab et al. 2018). Therefore, the RF model combines multiple DT for classification and prediction, which greatly improves the accuracy of prediction. The

calculation method has been explained in previous articles (Chen et al. 2023).

### Hybrid model

The FR model can re-quantify the attributes of each causal factor and input them as initial variables in the form of frequency ratios into the RF model to construct a new model (Li et al. 2021).

### Model validation method

**Confusion matrix** Due to the extreme imbalance between geological hazard and non-geological hazard samples, it is poorly applicable to judge the model prediction precision accuracy by statistical methods only (Yu et al. 2021). Previous researchers have used confusion matrices for geological hazard susceptibility evaluation (Frattini et al. 2010).

Combined with Table 1, the accuracy (ACC) calculation formula is shown as follows:

$$ACC = \frac{F_{11} + F_{22}}{F_{11} + F_{12} + F_{21} + F_{22}} \quad (3)$$

**ROC curve** The ROC curve of each model can be plotted and the AUC value (generally between 0.5 and 1) can be calculated by Python language. Many studies evaluate the accuracy of the model through them (Chen et al. 2020). An AUC value greater than 0.7 indicates good prediction effect, and an AUC value greater than 0.8 indicates excellent prediction effect (Wang et al. 2021a). The calculation method has been explained in previous articles (Chen et al. 2023).

## Result and discussion

### Screening of causative factors

The table of factor correlation coefficients (Table 2) was obtained by Pearson correlation coefficient analysis. From the table, it can be seen that AAR is strongly correlated with elevation. It can also be seen from Fig. 3 that the trend of rainfall change is almost consistent with the trend of

**Table 1** Confusion matrix

ACC		True value	
		Positive sample	Negative sample
Predicted value	Positive sample	$F_{11}$	$F_{12}$
	Negative sample	$F_{21}$	$F_{22}$

**Table 2** Factor phase relation number table

	Elevation	Slope	Aspect	Curvature	Lithology	DFF	DFR1	DFR2	AAR	NDVI
Elevation	1									
Slope	-0.052	1								
Aspect	-0.277	0.131	1							
Curvature	0.050	0.109	-0.043	1						
Lithology	-0.002	-0.214	-0.005	-0.036	1					
DFF	-0.350	-0.070	0.070	0.010	0.307	1				
DFR1	0.309	0.033	-0.028	0.017	0.227	-0.091	1			
DFR2	0.315	0.088	0.013	0.081	0.214	-0.094	0.436	1		
AAR	0.826	-0.008	-0.217	0.061	-0.297	-0.601	0.207	0.271	1	
NDVI	0.488	0.209	-0.112	0.095	-0.061	-0.224	0.246	0.351	0.415	1

elevation change. Except for the strong correlation between elevation and AAR, the correlation between elevation and the rest of the factors is less than 0.5, which is in the low and weak correlation. On the other hand, the correlation between AAR and DFF is higher than 0.5, with significant correlation. Therefore, the AAR factor is excluded in this paper.

### Susceptibility map of geological hazard based on the RF and FR-RF model

Although the RF model has been well evaluated in previous studies, it is unreasonable to use only one method to evaluate the susceptibility of geological hazards in the study area (Wang et al. 2022). In this paper, the RF and FR-RF models are used to model the susceptibility of geological hazards. The ratio of positive and negative samples is 1:1 (Jiang et al. 2017; Tsangaratos and Iliia 2016; Zhang et al. 2023). In order to avoid the negative sample selection process, the distance is too close, and the attribute values are too similar which reduces the model accuracy. Therefore, the negative samples are selected based on ArcGIS so that the distance between them is greater than 500 m. RF is a machine learning algorithm that can classify multiple variables, so it needs to input initial variables. Based on ArcGIS platform, the attribute values of each causative factor of positive and negative samples are extracted as input variables of RF model. The frequency ratio of each causative factor is used as the input variable of the FR-RF model. According to previous research experience, the samples are divided into training set and test set by 7:3 (Hussain et al. 2021; Sahin et al. 2020; Zhou et al. 2021). Similarly, attribute values and FR values of causative factors are assigned to the 30 × 30 m grids divided in the study area and input into each trained model as initial variables to obtain the susceptibility probability of each grid in the research area. The natural discontinuity method can be used to divide the susceptibility probability of the study area into 5 levels and draw the susceptibility zoning maps (Fig. 4). Table 3

### Model validation

#### Confusion matrix

The confusion matrix is used to observe the performance of the model on each category. It calculates the accuracy of the model corresponding to each category, making the categories more distinguishable. Based on Python,  $F_{11}$ ,  $F_{12}$ ,  $F_{21}$  and  $F_{22}$  are calculated for each model and ACC values are calculated for each model according to Table 1. Through Table 4, it is learnt that with the increase of buffer distance, the change trend of RF and FR-RF models remains consistent, both showing the trend of increasing first and then decreasing, and the ACC values all reach the maximum when the buffer distance is 1000 m. The ACC values of RF models constructed with different buffer distances are always higher than those of FR-RF models. The results of the confusion matrix indicate that the model constructed by RF with negative samples randomly selected outside the buffer distance of 1000 m is more accurate.

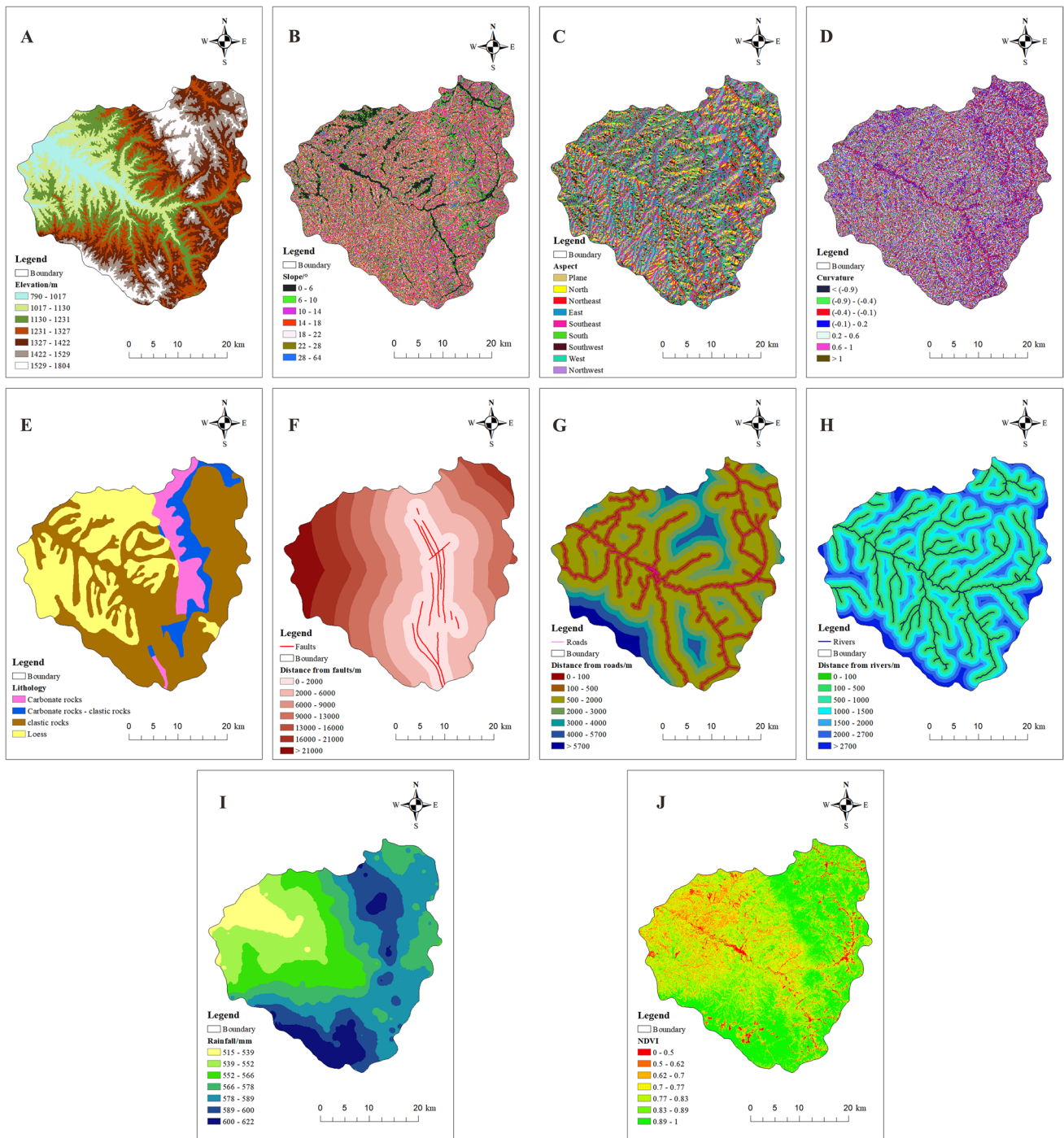
#### ROC curves

According to Fig. 5, the prediction effect of the models constructed outside different buffer distances is better. The variation trend of AUC values in RF and FR-RF models was consistent with that of ACC, both of which showed a trend of first rising and then declining, and AUC values reached the maximum when the buffer distance was 1000 m. The results of ROC curve show that the accuracy of the model constructed by RF is higher when the buffer distance is 1000 m.

### Model evaluation

#### Evaluation of susceptibility zoning maps of different buffer distance

According to Fig. 4, the spatial distribution of the model susceptibility zoning maps is basically consistent with

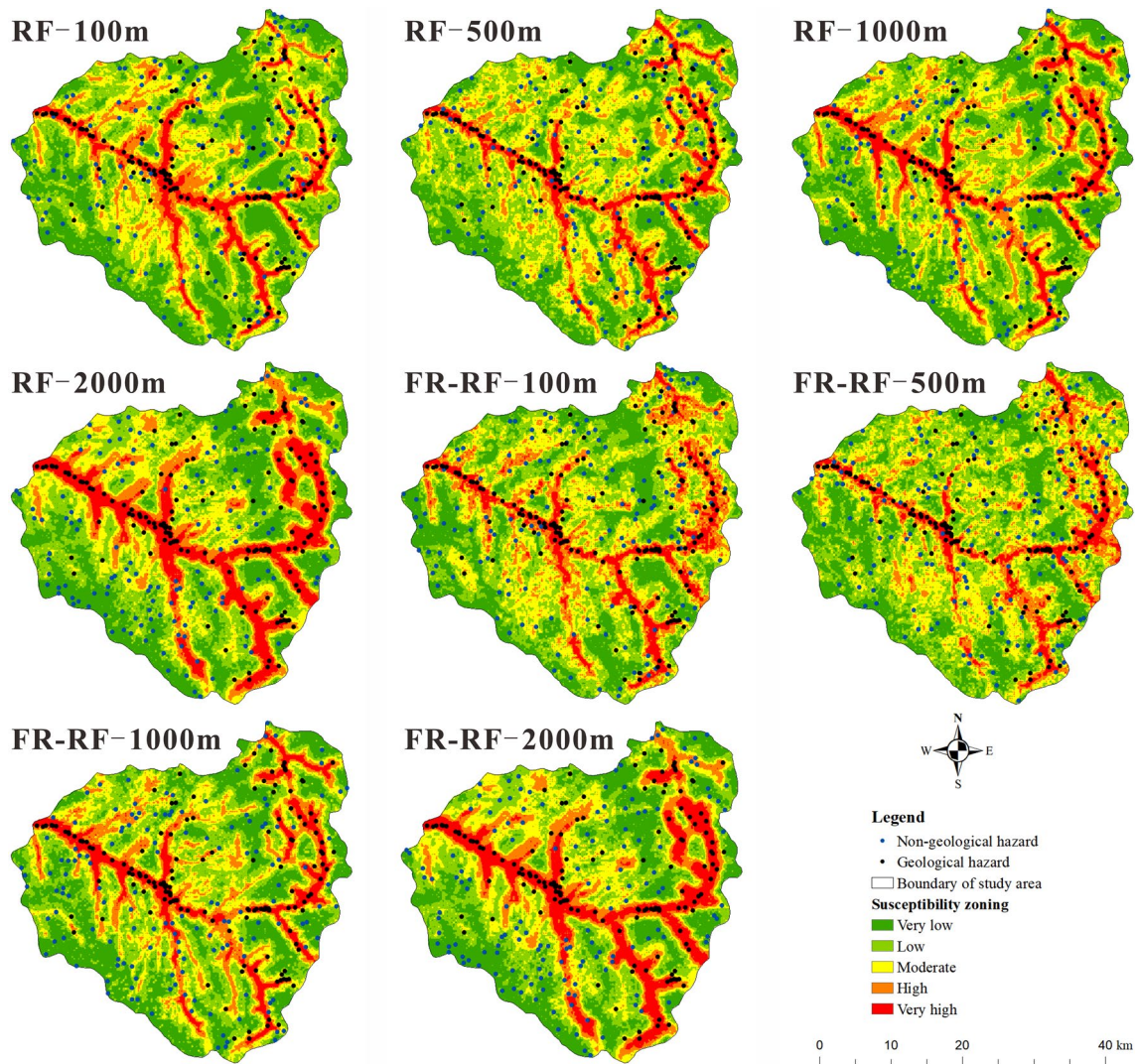


**Fig. 3** Causative factors attribute value classification: **A** elevation; **B** slope; **C** aspect; **D** curvature; **E** Lithology; **F** DFF; **G** DFR1; **H** DFR2; **I** AAR; **J** NDVI

the actual investigation. Very low susceptibility zones are widely distributed in the southern border area of the study area, near Cao Village in the southeast and in the very high altitude area in the northeast. The low susceptibility zone is widely distributed in the area near the centre of Pu County within the very low susceptibility zone in the south and

northeast. In the south, on both sides of the Nanchuan River and in the north-west, moderate susceptibility zones are mainly distributed. The high susceptibility zone is mainly distributed on both sides of rivers and roads in the study area. Very high susceptibility zones are mainly distributed in Xinshui River, Nanchuan River, Miangou River and on





**Fig. 4** Geological hazard susceptibility zoning maps

**Table 3** Classification of causative factors and FR values

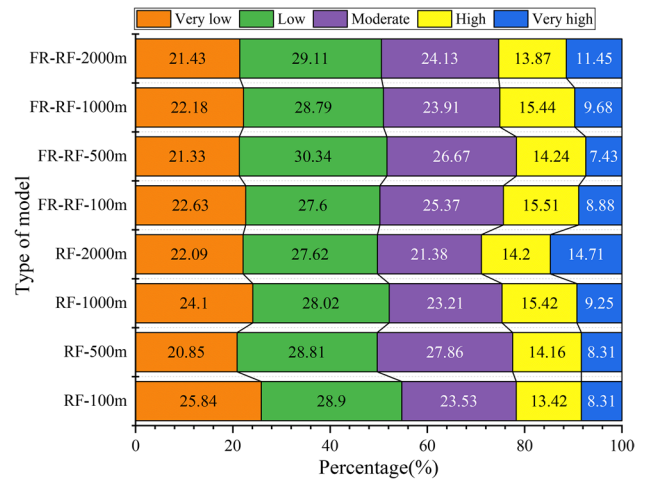
Classes	FR value
Elevation (m) 790~1017; 1017~1130; 1130~1231; 1231~1327; 1327~1422; 1422~1529; 1529~1804	3.7477; 0.7075; 0.9414; 1.1760; 0.7268; 0.4671; 0.1834
Slope (°) 0~6; 6~10; 10~14; 14~18; 18~22; 22~28; 28~64	1.4932; 0.8572; 1.0509; 0.9368; 0.7555; 0.8510; 0.7389
Aspect Plane; North; Northeast; East; Southeast; South; Southwest; West; Northwest	0.1192; 0.1126; 0.1077; 0.0771; 0.0690; 0.1092; 0.1608; 0.0842; 0.0829
Curvature <(-0.9); (-0.9)~(-0.4); (-0.4)~(-0.1);(-0.1)~0.2; 0.2~0.6; 0.6~1; >1	0.6015; 0.7685; 1.1502; 1.0727; 0.9640; 0.8763; 0
Lithology Carbonate rocks; carbonate rocks-clastic rocks; clastic rocks; Loess	1.2647; 0.6318; 1.4290; 0.2999
DFP (m) 0~2000; 2000~6000; 6000~9000; 9000~13000; 13,000~16,000; 16,000~21,000; >21,000	0.6923; 0.9636; 1.4658; 1.4915; 0.3044; 1.0921; 0.3786
DFR1 (m) 0~100; 100~500; 500~2000; 2000~3000; 3000~4000; 4000~5700; >5700	4.2560; 2.3498; 0.5613; 0.3161; 0.1639; 0.4042; 0
DFR2 (m) 0~100; 100~500; 500~1000; 1000~1500; 1500~2000; 2000~2700; >2700	3.0703; 1.9886; 0.8910; 0.4371; 0.1926; 0.3732; 0
NDVI 0~0.5; 0.5~0.62; 0.62~0.7; 0.7~0.77; 0.77~0.83; 0.83~0.89; 0.89~1	7.5335; 2.1665; 1.6758; 0.8388; 0.7454; 1.0147; 0.3983



**Table 4** Different buffer distance ACC values

	Buffer distance				
	100 m	500 m	1000 m	2000 m	300 m
RF	0.753	0.798	0.843	0.820	0.77
FR-RF	0.719	0.742	0.831	0.798	0.77

both sides of S329 road. According to Fig. 6, the proportion of very high susceptibility zones increases with the increase of buffer distance and reaches the highest at a buffer distance of 2000 m. The proportion of high susceptibility zone fluctuates and the RF and FR-RF models reach the highest at a buffer distance of 1000 m and 100 m, respectively. The proportion of moderate susceptibility zone shows an overall trend of increasing and then decreasing and reaches the highest at a buffer distance of 500 m. The RF model showed a decreasing trend in the proportion of low susceptibility zones, reaching the maximum at a buffer distance of 100 m, and the FR-RF model showed a fluctuating change, reaching the maximum at a buffer distance of 500 m. The very low susceptibility zone is highest at a buffer distance of 100 m and lowest at 500 m. At a buffer distance of 100 m, the proportion of very low and low susceptibility zones of the RF model is higher than that of the FR-RF model, and the moderate, high and very high susceptibility zones are higher than that of the FR-RF model. At a buffer distance of 500 m, the proportion of very low, low and high susceptibility zones of the RF model is lower than that of the FR-RF model, and the moderate and very high susceptibility zones are higher than that of the FR-RF model. At a buffer distance of 1000 m, the proportion of very low susceptibility zones of the RF model is higher than that of the FR-RF model, and the rest of the susceptibility zones are lower than that of the FR-RF model. At a buffer distance of 2000 m, the very low, high and high susceptibility zones of the RF model are higher

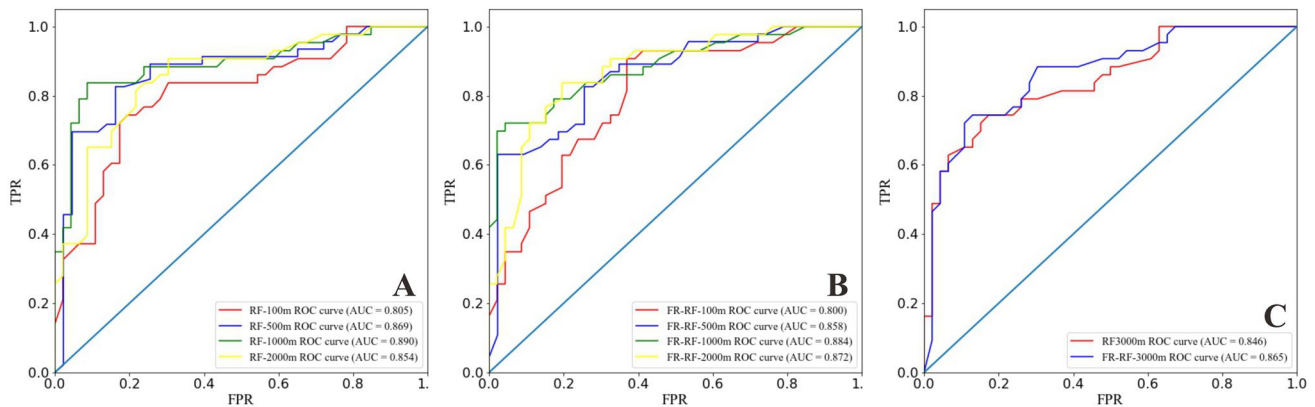


**Fig. 6** The proportion of susceptibility zoning of different models

than the FR-RF model, and the low and moderate susceptibility zones are higher than the FR-RF model.

**Analysis of reasons for differences in evaluation results**

According to Figs. 4 and 5 and Table 1, there are large differences in the results of geological hazard susceptibility when negative samples are selected outside different buffer distances. Since the RF model is the result of comprehensive decision-making by multiple decision trees, when the spatial location of the sample changes, its initial attribute values change along with it, and the results produced by the features selected by each decision tree will change. When all the decision results are aggregated together, there are differences in the judgement of the susceptibility level within each grid. In terms of buffer distance variation, the smaller the buffer distance, the closer the randomly created negative samples are to the positive samples. This results in the



**Fig. 5** ROC curve and AUC value: A ROC curves and AUC values under RF modeling, B ROC curves and AUC values under FR-RF modeling, C ROC curves and AUC values at a buffer distance of 3000 m

negative sample being more likely to have similar or consistent values for certain attributes as the positive sample. At this point, the negative samples are not representative. When the RF model is trained, it will confuse the features of the positive and negative samples, especially the features with higher weights, and therefore, the accuracy will decrease. As the buffer distance becomes larger, there are fewer cases where the attributes of the positive and negative samples are similar, in which case the negative samples selected are more representative. When buffer distances are particularly large, the randomly selected negative sample attribute values become more “extreme”. For example, the DFR1 is generally long. Therefore, when constructing the model, areas that do not meet these characteristics are considered to be highly vulnerable to geological hazards. Therefore, if negative samples are selected outside the buffer distance of 2000 m for model construction, the number of high- and very high-susceptibility zones will increase, and the ACC and AUC values will decrease, which is not in line with the actual situation. The difference between RF and FR-RF models is mainly because the FR-RF model inputs are calculated FR values, while the RF model inputs the real attribute values of each causal factor.

In order to further verify whether the samples will be more “extreme” and less “representative” as the buffer distance increases, we use 3000 m as the new buffer distance to construct and test the RF and FR-RF models. Fig. 5C and Table 4 show that when the buffer distance is 3000 m, the models are less effective, the ACC value of RF and FR-RF models is 0.77, and the AUC value is 0.846 and 0.865, respectively, which are smaller than those of 2000 m. Therefore, it can be inferred that the buffer distance continues to decrease, the “representativeness” of RF and FR-RF models will be more “extreme”, and we use 3000 m as the new buffer distance to construct RF and FR-RF models and test them. Therefore, it can be inferred that as the buffer distance continues to decrease, the model accuracy will also continue to decrease.

### Weight evaluation of causative factors

According to the RF-1000m model, to analyze the weight shares causative factors (Fig. 7). DFR1, DFR2, NDVI and elevation are the most important factors leading to the occurrence of geological hazards in the study area. According to the geological hazard survey, the occurrence of 124 geological hazards was related to human activities, which accounted for 83.78% of the total number, and most of them were distributed on both sides of the roads and rivers, especially the S329 road and Xinshui River. Due to historical reasons and the limitation of topographic and geomorphological conditions, local villagers living in the loess area are accustomed to building houses and digging kilns to live at the foot of

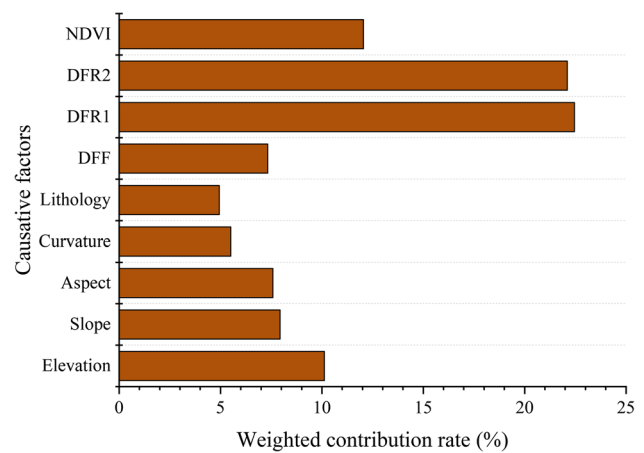


Fig. 7 Weight proportion of causative factors

slopes and opening up mountains to build roads. Such engineering activities often form high and steep slopes, breaking the natural equilibrium of the slope; with the passage of time and rainfall scouring, the slope soil body becomes broken, losing the original stability, the formation of landslides, collapses and unstable slopes. This type of engineering activity often occurs on both sides of the road. River development in the study area is widely distributed. On both sides of Xinshui River, Nanchuan River and Miangou River, bedrock is mostly exposed, the downward erosion of the river is blocked, the erosion on both sides is strong, which has obvious influence on the stability of the valley slopes, and it has triggered more landslides in the history. In some gullies, mainly loess gullies, the downward erosion of flowing water and both sides of the erosion exist. The valley slopes on both sides are steep and are still under erosion by flowing water, making them prone to landslides and avalanches. Vegetation plays the role of slope protection and preventing soil erosion, which has a certain influence on the evolution and stability of slopes. From Fig. 3H and J, it is seen that the vegetation cover on both sides of the river is generally low, which, together with the effects of river erosion and rainfall scouring, results in this being a high incidence area for geological hazards. The north-eastern and southern parts of the area have higher vegetation cover, and the distribution of geological hazards is minimal. Geological hazards are mainly distributed in areas of lower elevation, and almost none in areas of very high elevation. This is mainly due to the fact that there is less engineering construction in the high-altitude areas, and the erosion of the water system is not serious. Meanwhile, although the rainfall is high, it is mostly absorbed by the vegetation or flows to the low altitude areas, so geological hazards are not likely to occur in this area. On the other hand, the lithology of the low-altitude areas is more fragile, mostly loess and clastic rocks, the water system is widely distributed, human engineering is frequent, and the

**Table 5** Economic and loss rate indexes

	RF-100m	RF-500m	RF-1000m	RF-2000m	FR-RF-100m	FR-RF-500m	FR-RF-1000m	FR-RF-2000m
DGC	0.8699	0.9328	0.9178	1.0000	0.9336	0.9086	0.9408	0.9408
GDP loss value	0.8698	0.9263	0.9139	1.0000	0.9256	0.9047	0.9336	0.9460
Loss ratios	1.0000	0.9930	0.9958	1.0000	0.9915	0.9957	0.9923	0.9863

vegetation cover is low, so the combination of many factors leads to geological hazards in this area.

The mechanism of geological hazards is mainly related to the stability of the slope body. When the shear strength of the slope body is less than the shear force, landslides, collapse, debris flows and other slope-based geological hazards will occur. The shear strength of the slope body is closely related to the mechanical properties of the geotechnical body, slope, aspect, moisture content and other factors. Therefore, when these factors change in the direction of prompting the stability of the slope body to decrease, it will lead to the occurrence of geological hazards, especially when induced by rainfall and other factors. Through the above geological hazard susceptibility zoning and weighting analysis of causing factors, the geological hazard-prone areas in the study area are mainly on both sides of the river and the road. The erosion of the river on the slopes on both sides of the river destroys the physical and chemical properties of the geotechnical body, increases its water content, and reduces the shear strength of the slope, thus making it more prone to geological hazards. Secondly, with the development of human society, it is forced to carry out infrastructural construction in these gullies and ravines, such as cutting the slope to build houses and opening up mountains to build roads. These activities force the originally stable slopes to become unstable by physical means, and then geological hazards occur under the influence of river erosion and other factors. River and human activities are the most important factors, and NDVI, slope and other factors also play an important role in the occurrence of geological hazards.

### Economic and environmental loss assessment

Research shows that land use and GDP indicators are very important for maintaining environmental safety and sustainable economic development (Wang et al. 2021b; Zhao et al. 2006). The more accurate the geological hazard susceptibility model is, the more reasonable the zoning will be, and the lower the economic loss and treatment cost will be. HMC and loss rates (expressed as GDP/HMC in this article) can be used as objective indicators to assess the extent of damage. (Yum et al. 2020). The calculation method has been explained in previous articles (Chen et al. 2023).

Economic and loss rate indicators of models constructed with different buffer distances (Table 5) show that

FR-RF-1000m HMC and GDP loss ratio is lower than other models, and the economic benefits are better (only worse than RF-100m and FR-RF-500m). FR-RF-1000m is the model with the highest prediction degree in the study area. Compared with other models, the governance cost of FR-RF-1000m decreases by 3.55% on average, which indicates that the RF model can effectively reduce the cost of hazard management by using sampling outside 1000m buffer distance.

### Conclusion

Accurate and scientific geological hazard susceptibility analysis and obtaining a scientific and reliable buffer distance index are the key steps to improve the accuracy of susceptibility zoning. This paper takes Puxian County as the research object and constructs the susceptibility model by dividing different buffer distances. Through the study, whether it is the RF or FR-RF model, the change of buffer distance will have an impact on the accuracy of the model. Therefore, buffer distance is a necessary consideration when using machine learning methods to construct highly accurate geological hazard susceptibility models. Each area should have a buffer distance that is most suitable for it. Through this study, we found that the buffer distance is too large or too small, which will lead to the sample not being “representative”, and it should be in an “intermediate value”. Through the weighting analysis of causative factors, DFR1, DFR2 and NDVI are the main factors leading to the occurrence of hazards in the study area. By comparing the economic and environmental losses, the average cost of hazard management was 3.55% higher in the other models than in the RF-1000m model. This study is of great significance to promote the sustainable development of economy and ecological environment in the geological hazard susceptibility areas and also provides scientific basis for the selection of buffer distance in the future evaluation of geological hazard susceptibility in Puxian County and the surrounding areas.

**Author contribution** Conceptualization, ZW; methodology, ZW; software, ZW; validation, ZW; formal analysis, LP; investigation, ZL and YX; resources, ZL; data curation, LP; writing—original draft preparation, ZW; writing—review and editing, JC; visualization, FL; supervision, ZW; project administration, JC; funding acquisition,



JC. All authors have read and agreed to the published version of the manuscript.

**Funding** This work was supported by National Natural Science Foundation of China (grant numbers: 51604140).

**Data availability** The data that support the findings of this study are available from the corresponding author, Jianping Chen, upon reasonable request.

## Declarations

**Ethics approval and consent to participate** Not applicable

**Consent for publication** Not applicable

**Competing interests** The authors declare no competing interest.

## References

- Adition A, Kubota T, Shinohara Y (2018) Comparison of GIS-based landslide susceptibility models using frequency ratio, logistic regression, and artificial neural network in a tertiary region of Ambon, Indonesia. *Geomorphology* 318:101–111. <https://doi.org/10.1016/j.geomorph.2018.06.006>
- Akinci H, Kilicoglu C, Dogan S (2020) Random forest-based landslide susceptibility mapping in coastal regions of Artvin, Turkey. *Isprs Int J Geo-Inf* 9:553. <https://doi.org/10.3390/ijgi9090553>
- Ba QQ, Chen YM, Deng SS, Wu QJ, Yang JX, Zhang JY (2017) An improved information value model based on gray clustering for landslide susceptibility mapping. *Isprs Int J Geo-Inf* 6:18. <https://doi.org/10.3390/ijgi6010018>
- Biswas A, Si BC (2011) Scales and locations of time stability of soil water storage in a hummocky landscape. *J Hydrol* 408:100–112. <https://doi.org/10.1016/j.jhydrol.2011.07.027>
- Bordoni M, Galanti Y, Bartelletti C, Persichillo MG, Barsanti M, Giannecchini R, Avanzi GD, Cevasco A, Brandolini P, Galve JP, Meisina C (2020) The influence of the inventory on the determination of the rainfall-induced shallow landslides susceptibility using generalized additive models. *Catena* 193:104630. <https://doi.org/10.1016/j.catena.2020.104630>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Catani F, Lagomarsino D, Segoni S, Tofani V (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Nat Hazard Earth Sys* 13:2815–2831. <https://doi.org/10.5194/nhess-13-2815-2013>
- Chang SH, Wan SA (2015) Discrete rough set analysis of two different soil-behavior-induced landslides in National Shei-Pa Park, Taiwan. *Geosci Front* 6:807–816. <https://doi.org/10.1016/j.gsf.2013.12.010>
- Chen JP, Wang ZP, Chen W, Wan CY, Liu YY, Huang JJ (2023) The influence of the selection of non-geological disasters sample spatial range on the evaluation of environmental geological disasters susceptibility: a case study of Liulin County. *Environ Sci Pollut R*:36697985. <https://doi.org/10.1007/s11356-023-25454-2>
- Chen S, Miao ZL, Wu LX, Zhang AS, Li QR, He YG (2021) A one-class-classifier-based negative data generation method for rapid earthquake-induced landslide susceptibility mapping. *Front Earth Sc-Switz* 9. <https://doi.org/10.3389/feart.2021.609896>
- Chen W, Chen YZ, Tsangaratos P, Ilia I, Wang XJ (2020) Combining evolutionary algorithms and machine learning models in landslide susceptibility assessments. *Remote Sens-Basel* 12:3854. <https://doi.org/10.3390/rs12233854>
- Chen W, Panahi M, Tsangaratos P, Shahabi H, Ilia I, Panahi S, Li SJ, Jaafari A, Bin Ahmad B (2019) Applying population-based evolutionary algorithms and a neuro-fuzzy system for modeling landslide susceptibility. *Catena* 172:212–231. <https://doi.org/10.1016/j.catena.2018.08.025>
- Chen W, Xie XS, Wang JL, Pradhan B, Hong HY, Bui DT, Duan Z, Ma JQ (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* 151:147–160. <https://doi.org/10.1016/j.catena.2016.11.032>
- Conoscenti C, Rotigliano E, Cama M, Caraballo-Arias NA, Lombardo L, Agnesi V (2016) Exploring the effect of absence selection on landslide susceptibility models: a case study in Sicily, Italy. *Geomorphology* 261:222–235. <https://doi.org/10.1016/j.geomorph.2016.03.006>
- Du GL, Zhang YS, Yang ZH, Guo CB, Yao X, Sun DY (2019) Landslide susceptibility mapping in the region of eastern Himalayan syntaxis, Tibetan Plateau, China: a comparison between analytical hierarchy process information value and logistic regression-information value methods. *B Eng Geol Environ* 78:4201–4215. <https://doi.org/10.1007/s10064-018-1393-4>
- Erener A, Duzgun HBS (2013) A regional scale quantitative risk assessment for landslides: case of Kumluca watershed in Bartin, Turkey. *Landslides* 10:55–73. <https://doi.org/10.1007/s10346-012-0317-9>
- Frattoni P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. *Eng Geol* 111:62–72. <https://doi.org/10.1016/j.enggeo.2009.12.004>
- Froude MJ, Petley DN (2018) Global fatal landslide occurrence from 2004 to 2016. *Nat Hazard Earth Sys* 18:2161–2181. <https://doi.org/10.5194/nhess-18-2161-2018>
- Guo YJ (2020) Assessment of landslide susceptibility based on ensemble learning algorithm in Xi'an city. Dissertation, Xi'an University of Science and Technology. <https://doi.org/10.27397/d.cnki.gxaku.2020.000053>
- He SW, Pan P, Dai L, Wang HJ, Liu JP (2012) Application of kernel-based Fisher discriminant analysis to map landslide susceptibility in the Qinggan River delta, Three Gorges, China. *Geomorphology* 171:30–41. <https://doi.org/10.1016/j.geomorph.2012.04.024>
- Hong HY, Pradhan B, Xu C, Tien Bui D (2015) Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena* 133:266–281. <https://doi.org/10.1016/j.catena.2015.05.019>
- Hu XD, Huang C, Mei HB, Zhang H (2021) Landslide susceptibility mapping using an ensemble model of Bagging scheme and random subspace-based naive Bayes tree in Zigui County of the Three Gorges Reservoir Area, China. *B Eng Geol Environ* 80:5315–5329. <https://doi.org/10.1007/s10064-021-02275-6>
- Huang Y, Zhao L (2018) Review on landslide susceptibility mapping using support vector machines. *Catena* 165:520–529. <https://doi.org/10.1016/j.catena.2018.03.003>
- Hussain MA, Chen ZL, Wang R, Shoaib M (2021) PS-InSAR-based validated landslide susceptibility mapping along Karakorum Highway, Pakistan. *Remote Sens-Basel* 13:4129. <https://doi.org/10.3390/rs13204129>
- Jiang WG, Rao PZ, Cao R, Tang ZH, Chen K (2017) Comparative evaluation of geological disaster susceptibility using multi-regression methods and spatial accuracy validation. *J Geogr Sci* 27:439–462. <https://doi.org/10.1007/s11442-017-1386-4>
- Kalantar B, Pradhan B, Naghibi SA, Motevalli A, Mansor S (2018) Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural

- networks (ANN). *Geomat Nat Haz Risk* 9:49–69. <https://doi.org/10.1080/19475705.2017.1407368>
- Kanwal S, Atif S, Shafiq M (2017) GIS based landslide susceptibility mapping of northern areas of Pakistan, a case study of Shigar and Shyok Basins. *Geomat Nat Haz Risk* 8:348–366. <https://doi.org/10.1080/19475705.2016.1220023>
- Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11:425–439. <https://doi.org/10.1007/s10346-013-0391-7>
- Khan H, Shafique M, Khan MA, Bacha MA, Shah SU, Calligaris C (2019) Landslide susceptibility assessment using frequency ratio, a case study of Northern Pakistan. *Egypt J Remote Sens* 22:11–24. <https://doi.org/10.1016/j.ejrs.2018.03.004>
- Li CX, Ma Y, He YX (2020) Sensitivity analysis of debris flow to environmental factors: a case of Longxi River basin in Duijiangyan, Sichuan Province. *Chinese J Geol Hazard Contr* 31:32–39. <https://doi.org/10.16031/j.cnki.issn.1003-8035.2020.05.05>
- Li WB, Fan XM, Huang FM, Wu XL, Yin KL, Chang ZL (2021) Uncertainties of landslide susceptibility modeling under different environmental factor connections and prediction models. *Earth Sci* 46:3777–3795. <https://doi.org/10.3799/dqkx.2021.042>
- Lin JH, Chen WH, Qi XH, Hou HR (2021) Risk assessment and its influencing factors analysis of geological hazards in typical mountain environment. *J Clean Prod* 309. <https://doi.org/10.1016/j.jclepro.2021.127077>
- Liu J, Li SL, Chen T (2018) Landslide susceptibility assesment based on optimized random forest model. *Geomat Inf Sci Wuhan Univ* 43:1085–1091. <https://doi.org/10.13203/j.whugis20160515>
- Liu MM, Liu JP, Xu SH, Zhou T, Ma Y, Zhang FH, Konecny M (2021) Landslide susceptibility mapping with the fusion of multi-feature SVM model based FCM sampling strategy: a case study from Shaanxi Province. *Int J Image Data Fus* 12:349–366. <https://doi.org/10.1080/19479832.2021.1961316>
- Ma X, Wang NQ, Li XK, Yan D., Li JL (2022) Assessment of landslide susceptibility based on RF-FR model: taking Lueyang County as an example. *Northwest Geo* 55: 335-344. <https://doi.org/10.19751/j.cnki.61-1149/p.2022.03.028>
- Meinhardt M, Fink M, Tunschel H (2015) Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: comparison of a new method to calculate weighting factors by means of bivariate statistics. *Geomorphology* 234:80–97. <https://doi.org/10.1016/j.geomorph.2014.12.042>
- Mondini AC, Chang KT (2014) Combining spectral and geoenvironmental information for probabilistic event landslide mapping. *Geomorphology* 213:183–189. <https://doi.org/10.1016/j.geomorph.2014.01.007>
- Nicu IC (2017) Frequency ratio and GIS-based evaluation of landslide susceptibility applied to cultural heritage assessment. *J Cult Herit* 28:172–176. <https://doi.org/10.1016/j.cuther.2017.06.002>
- Nun I, Pichara K, Protopapas P, Kim DW (2014) Supervised detection of anomalous light curves in massive astronomical catalogs. *ApJ* 793. <https://doi.org/10.1088/0004-637X/793/1/23>
- Park S, Hamm SY, Kim J (2019) Performance evaluation of the GIS-based data-mining techniques decision tree, random forest, and rotation forest for landslide susceptibility modeling. *Sustainability-Basel* 11. <https://doi.org/10.3390/su11205659>
- Pawluszek K, Borkowski A (2017) Impact of DEM-derived factors and analytical hierarchy process on landslide susceptibility mapping in the region of Roznow Lake, Poland. *Nat Hazards* 86:919–952. <https://doi.org/10.1007/s11069-016-2725-y>
- Peng L, Niu RQ, Huang B, Wu XL, Zhao YN, Ye RQ (2014) Landslide susceptibility mapping based on rough set theory and support vector machines: a case of the Three Gorges area, China. *Geomorphology* 204:287–301. <https://doi.org/10.1016/j.geomorph.2013.08.013>
- Polykretis C, Chalkias C (2018) Comparison and evaluation of landslide susceptibility maps obtained from weight of evidence, logistic regression, and artificial neural network models. *Nat Hazards* 93:249–274. <https://doi.org/10.1007/s11069-018-3299-7>
- Posner AJ, Georgakakos KP (2015) Normalized landslide index method for susceptibility map development in El Salvador. *Nat Hazards* 79:1825–1845. <https://doi.org/10.1007/s11069-015-1930-4>
- Pourghasemi HR, Gayen A, Park S, Lee CW, Lee S (2018) Assessment of landslide-prone areas and their zonation using logistic regression, LogitBoost, and Naive Bayes machine-learning algorithms. *Sustainability-Basel* 10:3697. <https://doi.org/10.3390/su10103697>
- Pourghasemi HR, Rahmati O (2018) Prediction of the landslide susceptibility: which algorithm, which precision? *Catena* 162:177–192. <https://doi.org/10.1016/j.catena.2017.11.022>
- Sahin EK, Colkesen I, Kavzoglu T (2010) A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *GeoIn* 35:341–363. <https://doi.org/10.1080/10106049.2018.1516248>
- Sarda VK, Pandey DD (2019) Landslide susceptibility mapping using information value method. *Jordan J Civ Eng* 13:335–350
- Su CX, Wang BJ, Lv YH, Zhang MP, Peng DL, Bate B, Zhang S (2022) Improved landslide susceptibility mapping using unsupervised and supervised collaborative machine learning models. *Georisk*. <https://doi.org/10.1080/17499518.2022.2088802>
- Su QM, Zhang J, Zhao SM, Wang L, Liu J, Guo JL (2017) Comparative assessment of three nonlinear approaches for landslide susceptibility mapping in a coal mine area. *ISPRS Int J Geo-Inf* 6. <https://doi.org/10.3390/ijgi6070228>
- Sun XH, Chen JP, Bao YD, Han XD, Zhan JW, Peng W (2018) 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* 7:438. <https://doi.org/10.3390/ijgi7110438>
- Taalab K, Cheng T, Zhang Y (2018) Mapping landslide susceptibility and types using random forest. *Big Earth Data* 2:159–178. <https://doi.org/10.1080/20964471.2018.1472392>
- Tsangaratos P, Ilia I (2016) Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. *Landslides* 13:305–320. <https://doi.org/10.1007/s10346-015-0565-6>
- Wang C, Wang XD, Zhang HY, Meng FQ, Li XL (2022) Assessment of environmental geological disaster susceptibility under a multimodel comparison to aid in the sustainable development of the regional economy. *Environ Sci Pollut R*. <https://doi.org/10.1007/s11356-022-22649-x>
- Wang JJ, Kl Y, Xiao LL (2014) Landslide susceptibility assessment based on GIS and weighted information value: a case study of Wanzhou district, Three Gorges Reservoir. *Chinese J Rock Mech Eng* 33:797–808. <https://doi.org/10.13722/j.cnki.jrme.2014.04.012>
- Wang JZ, Gao YC, Tie YB, Xu W, Bai YJ, Zhang YF (2021a) Evaluation of the susceptibility to landslide in mountainous towns based on slope unit: taking Kangding as an example. *Sediment Geol Tethyan Geol*:1–17. <https://doi.org/10.19826/j.cnki.1009-3850.2021.03001>
- Wang XD, Zhang CB, Wang C, Liu GW, Wang HX (2021b) GIS-based for prediction and prevention of environmental geological disaster susceptibility: from a perspective of sustainable development. *Ecotox Environ Safe* 226:112881. <https://doi.org/10.1016/j.ecoenv.2021.112881>
- Wang Y, Sun DL, Wen HJ, Zhang H, Zhang FT (2020) Comparison of random forest model and frequency ratio model for landslide susceptibility mapping (LSM) in Yunyang County (Chongqing, China). *Int J Env Res Pub He* 17. <https://doi.org/10.3390/ijerph17124206>
- Xiong JN, Wei FQ, Liu ZQ (2017) Hazard assessment of debris flow in Sichuan Province. *J Geo-inform Sci* 19:1604–1612 <https://kns.cnki.net/kcms/detail/11.5809.P.20171228.1447.020.html>

- Yang Y, Hu N (2019) The spatial and temporal evolution of coordinated ecological and socioeconomic development in the provinces along the Silk Road Economic Belt in China. *Sustain Cities Soc* 47:101466. <https://doi.org/10.1016/j.scs.2019.101466>
- Yu JN, Liu K, Zhang BY, Huang Y, Fan CY, Song CQ, Tang GA (2021) Vertical accuracy assessment and applicability analysis of TanDEM-X 90 m DEM in China. *J Geo-inform Sci* 23:646–657
- Yum SG, Ahn S, Bae J, Kim JM (2020) Assessing the risk of natural disaster-induced losses to tunnel-construction projects using empirical financial-loss data from South Korea. *Sustainability-Basel* 12:8026. <https://doi.org/10.3390/su12198026>
- Zhang Q, Yu H, Li ZN, Zhang GH, Ma DT (2020) Assessing potential likelihood and impacts of landslides on transportation network vulnerability. *Transport Res D-Tr E* 82:102304. <https://doi.org/10.1016/j.trd.2020.102304>
- Zhang WA, He YW, Wang LQ, Liu SL, Meng XY (2023) Landslide susceptibility mapping using random forest and extreme gradient boosting: a case study of Fengjie. *GeolJ, Chongqing*. <https://doi.org/10.1002/gj.4683>
- Zhao BB, Ge YF, Chen HZ (2021) Landslide susceptibility assessment for a transmission line in Gansu Province, China by using a hybrid approach of fractal theory, information value, and random forest models. *Environ Earth Sci* 80. <https://doi.org/10.1007/s12665-021-09737-w>
- Zhao LR, Wu XL, Niu RQ, Wang Y, Zhang KX (2020) Using the rotation and random forest models of ensemble learning to predict landslide susceptibility. *Geomat Nat Haz Risk* 11:1542–1564. <https://doi.org/10.1080/19475705.2020.1803421>
- Zhao Q (2017) Types and causes of geological calamity of mountain region of Linfen in Shanxi Province. *J Taiyuan Normal Univ (NaturalScience Edition)* 16:93–96
- Zhao YZ, Zou XY, Cheng H, Jia HK, Wu YQ, Wang GY, Zhang CL, Gao SY (2006) Assessing the ecological security of the Tibetan plateau: methodology and a case study for Lhaze County. *J Environ Manage* 80:120–131. <https://doi.org/10.1016/j.jenvman.2005.08.019>
- Zhao ZB, Han JQ, Chen S (2016) Analysis of geological disasters and its harmfulness in Linfen. *J Shanxi Norm Univ(Nat Sci Edit)* 30:117–123. <https://doi.org/10.16207/j.cnki.1009-4490.2016.02.024>
- Zhou XT, Wu WC, Lin ZY, Zhang GL, Chen RX, Song Y, Wang ZL, Lang T, Qin YZ, Ou PH, Wenchao HF, Zhang Y, Xie LF, Huang XL, Fu X, Li J, Jiang JH, Zhang M, Liu YX et al (2021) Zonation of landslide susceptibility in Ruijin, Jiangxi, China. *Int J Env Res Pub He* 18:5906. <https://doi.org/10.3390/ijerph18115906>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.