



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

Jianping Chen¹ · Zepeng Wang¹ · Wei Chen² · Changyuan Wan¹ · Yunyan Liu¹ · Junjie Huang¹

Received: 10 August 2022 / Accepted: 17 January 2023 / Published online: 26 January 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The reasonable selection of non-geological disaster samples is of great significance to improve the accuracy of geological disaster assessment, reduce the cost of disaster management, and maintain the sustainable development of ecological environment. Liulin County was selected as the study area. This paper creatively divided non-geological disaster sampling areas by macro-geomorphology, and carried out susceptibility mapping based on random forest (RF) and frequency ratio-random forest (FR-RF) models. The accuracy of each model was evaluated by receiver operating characteristic curve (ROC) combined with the distribution characteristics of geological disasters and the actual urban construction in the study area. The results show that the FR-RF model constructed by selecting non-geological disaster samples in hilly area is most suitable for the susceptibility mapping of this study area. The different results in different sampling areas are mainly due to the great changes in the representativeness of non-geological disaster samples. The distance from the roads is the most important factor affecting the occurrence of disasters in the study area. The statistical results of disaster management cost estimation and gross domestic product (GDP) value show that the disaster management cost of HFR-RF model decreases by 13.45% on average compared with other models, and the ratio of GDP to disaster management cost is relatively high. These research results promote the progress of geological disaster prevention technology, maintain the stability of geological environment, and are of great significance to the stable and sustainable development of local economy.

Keywords Non-geological disaster · Susceptibility · Frequency ratio · Random forest · ROC · Sustainable development

Introduction

In recent years, geological disasters occurred frequently all over the world (Arabameri et al. 2019). China is one of the countries with high frequency of geological disasters. Geological disasters have the characteristics of sudden occurrence and strong destruction, which seriously affect the sustainable development of local economy (Yang and Hu 2019). According to data released by China National Bureau of Statistics in 2021, a total of 100,650 geological

disasters were reported from 2011 to 2021, with 4834 people injured and 2729 killed, leading to direct economic losses of 45.393 billion yuan. At the same time, due to climate change and more extreme rainfall, the rapid urbanization of mountain development and geological disasters caused by human engineering activities are gradually increasing (Froude and Petley 2018; Gariano and Guzzetti 2016; Lin et al. 2017; Zhang et al. 2021). According to the Resources and Environment Science and Data Center of China, geological disasters in Liulin County have threatened 2931 people and 24,2617 million property and destroyed 11 buildings and 200 roads. Scientific and accurate analysis and evaluation of geological disasters in Liulin County and the identification of geological disaster prone areas can reduce the cost of disaster management, maintain the safety of ecological environment, and have very important significance for the prevention and control of local geological disasters (Jiang et al. 2017).

Geological disasters susceptibility is defined as the probability of geological disasters occurring in a region, which is

Responsible Editor: Philippe Garrigues

✉ Zepeng Wang
449139098@qq.com

¹ College of Mining, Liaoning Technical University, FuXin 123000, China

² College of Environment, Liaoning Technical University, FuXin 123000, China

obtained by analyzing the local geographical environment, human engineering activities, and other factors. Researchers around the world have applied a variety of methods, both qualitative and quantitative, to the assessment of geological disasters susceptibility. Qualitative methods are subjective, while quantitative methods involve numerical analysis and calculation (Pawluszek-Filipiak et al. 2020). At present, qualitative methods, such as analytic hierarchy process (AHP) (Pawluszek and Borkowski 2017) and fuzzy logic (FL) (Kanungo et al. 2006), are subject to strong subjective influence, resulting in large error of evaluation results and easy mapping, and their application scope is becoming smaller and smaller. Quantitative methods are mainly statistical analysis and machine learning model, such as information value (IV) (Ba et al. 2017; Du et al. 2019; Sarda and Pandey 2019), frequency ratio (FR) (Khan et al. 2019; Nicu 2017), and weight of evidence (WOE) (Lee and Oh 2019). Support vector machine (SVM) (Huang and Zhao 2018), artificial neural network (ANN) (Aditian et al. 2018), and random forest (RF) (Akinci et al. 2020; Zhao et al. 2020) are the most widely used methods in geological disaster susceptibility mapping (Wang et al. 2022a). However, there is little research on the potential impact of the selection of the spatial location of non-geological disaster samples on the geological disasters susceptibility mapping. Deng selected non-geological disaster samples in the very low and low susceptibility areas of the information value model to conduct susceptibility evaluation (Deng et al. 2021). Li selected non-geological disaster samples in the study area by down-sampling method to carry out the susceptibility evaluation (Li et al. 2018). An uniformly selected non-geological disaster sites 500 m away from known disaster sites to conduct susceptibility evaluation (An and Niu 2016). The common point of these studies is that they set a single sample selection method in advance, without considering whether the non-disaster points selected in this way are representative or whether it is more reasonable to select non-disaster points in other areas. Therefore, this paper puts forward hypotheses: can we get more representative samples by randomly sampling non-geological disaster sample areas divided by macro-geomorphology of the study area, and whether the model has better prediction effect and higher accuracy? Compared with previous studies, the advantage of the hypothesis is that the representativeness of negative samples (non-geological disaster samples) is closely related to the topography and geomorphology of the space, geological environment, and other factors, so the sampling area divided in this way can better reflect the representativeness of negative samples. According to the statistics of disasters in Shanxi Province by Resources and Environmental Sciences and Data Center of Chinese Academy of Sciences, Liulin County, is one of the regions where geological disasters occur more frequently over the years. Geological disasters such as landslide and collapse occur

frequently in the area. Especially in recent years, due to the intensification of human engineering activities such as road excavation and slope cutting, slope stability decreases, and geological disasters occur more frequently (Huang et al. 2016; Li et al. 2022b). The relevant studies have divided the macro-geomorphology in detail, and the accuracy of the model will not be affected by the geomorphology division error. Selecting Liulin County for susceptibility assessment cannot only improve the accuracy of local disaster prevention and mitigation, reduce the cost of disaster prevention and mitigation, and protect the stability of ecological environment, but also provide a theoretical basis for the selection of non-geological disaster samples in the surrounding areas with similar geomorphology. Considering that different results may occur when non-geological disaster samples are selected for model construction in different areas, the FR model is selected to judge such situations. Although the prediction accuracy of FR model is generally weaker than that of machine learning model, it has an advantage that the result of susceptibility mapping is consistent with the actual situation on the whole distribution.

To sum up, Liulin County (Shanxi, China) was selected as the study area. Combined with GIS technology, the sampling area of non-geological disaster samples was divided according to the macro-geomorphology of the study area. In combination with previous experience and field conditions, the evaluation index of geological disaster susceptibility was reasonably selected, and the geological disaster susceptibility mapping was carried out by using RF and FR-RF models. The ROC was used to compare the accuracy of the model constructed from the negative samples selected from different sampling areas, and the most appropriate model was judged by combining the distribution characteristics of geological disasters and field conditions in the study area. By analyzing the causes of different results in different sampling areas, evaluating the key causes of each model and the effects of energy conservation and environmental protection, and determining the key areas for disaster prevention and control in the study area, it can effectively avoid unnecessary economic losses in the process of disaster management and governance, and achieve long-term stable and sustainable development.

Materials and methodology

Study area

Liulin County is located at the western foot of Lvliang Mountain and on the east bank of the Yellow River. It is the west gate of Shanxi. It is surrounded by Lishi District, Lin County, Zhongyang County and Shilou County. It is between 110°39'45" ~ 110°05'33" east longitude and 37°08'53" ~

37°37'28" north latitude (Fig. 1). It has jurisdiction over 8 towns, 7 townships, and 197 administrative villages, with a total population of 346,000 (as of February 15, 2022). The County has a warm temperate continental monsoon climate, with an average precipitation of 456.3 mm, and the extreme values of rainfall over the years are 632 mm and 37.44 mm. Liulin County belongs to the hilly and gully area of the Loess Plateau in the northwest, and its altitude decreases from east to west. The geological tectonic movement uplifted the Lvliang Mountain structure, tilted the rock stratum to the west, cut off the Yellow River, and formed the bedrock of the county seat, which is high in the northeast and low in the southwest. The Quaternary loess layer covered on the surface has been eroded and denuded by rain for a long time, and is gradually cut into complex geomorphic units with undulating beams, ravines, crisscrossing hills, and fragmentation. According to the results of geological disaster investigation and zoning in Liulin County, a total of 151 geological disaster points (including 32 collapses, 71 landslides, 1 debris flow, and 47 unstable slopes) had been identified. According to the classification standard of disaster levels in the technical requirements of geological disaster investigation, there were 6 large disasters, 21 medium disasters, and 124 small disasters in the study area. According to the classification standard of danger, there were 2 large disasters, 4 medium disasters, and 145 small disasters in the study area. The landslides in Chengjiashuang Middle School and Shaqu Village were both of large scale, threatening 1600 people and 7.1 million property. The distribution

of geological disasters was shown in Fig. 1A. Based on the scope of the study area, a raster dataset of Liulin County was created, with a grid cell size of 300 m × 300 m (Huang et al. 2020) and a total of 14,377 grids.

Based on the landform division of Liulin County by Li et al. (2022c), in this paper, the study area is divided into five non-geological disaster data sampling areas, namely the mixed area (the whole study area), mountain area, broken plateau area, hilly area, and river valley area (Fig. 1B).

Methodology

The research process is shown in Fig. 2, which is divided into three stages: (1) geological disaster investigation, determination and collection of geological disaster impact factors; (2) Firstly, the FR model was constructed, and then the RF and FR-RF models were used to map the susceptibility of each sampling area. By comparing with the FR model, the model with relatively reasonable susceptibility mapping results was selected. The accuracy of the model with relatively reasonable partition results was tested by ROC curve. The most suitable sampling area and model were judged according to the distribution characteristics of geological disasters and the actual urban construction. (3) Analyze the reasons for the differences in the model construction results of different sampling areas, evaluate the susceptibility zoning map of the most suitable study area, the key disaster causing factors, and the energy-saving and environmental protection effect of each model.

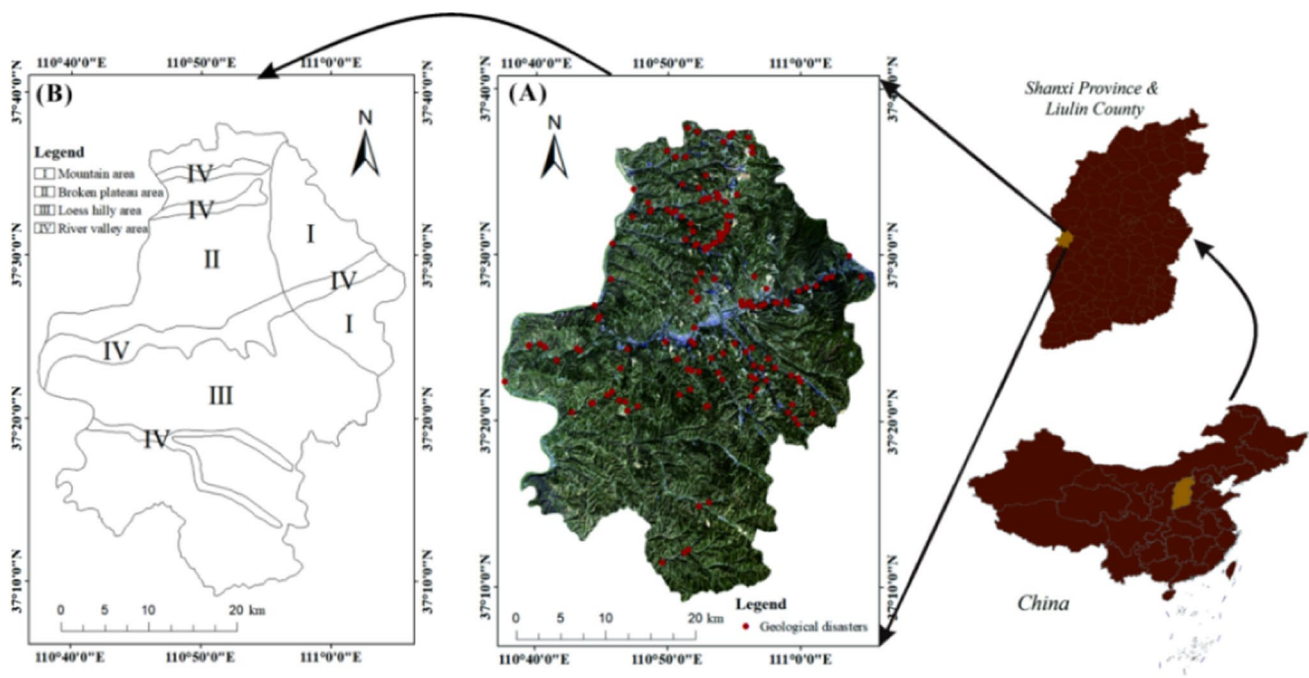
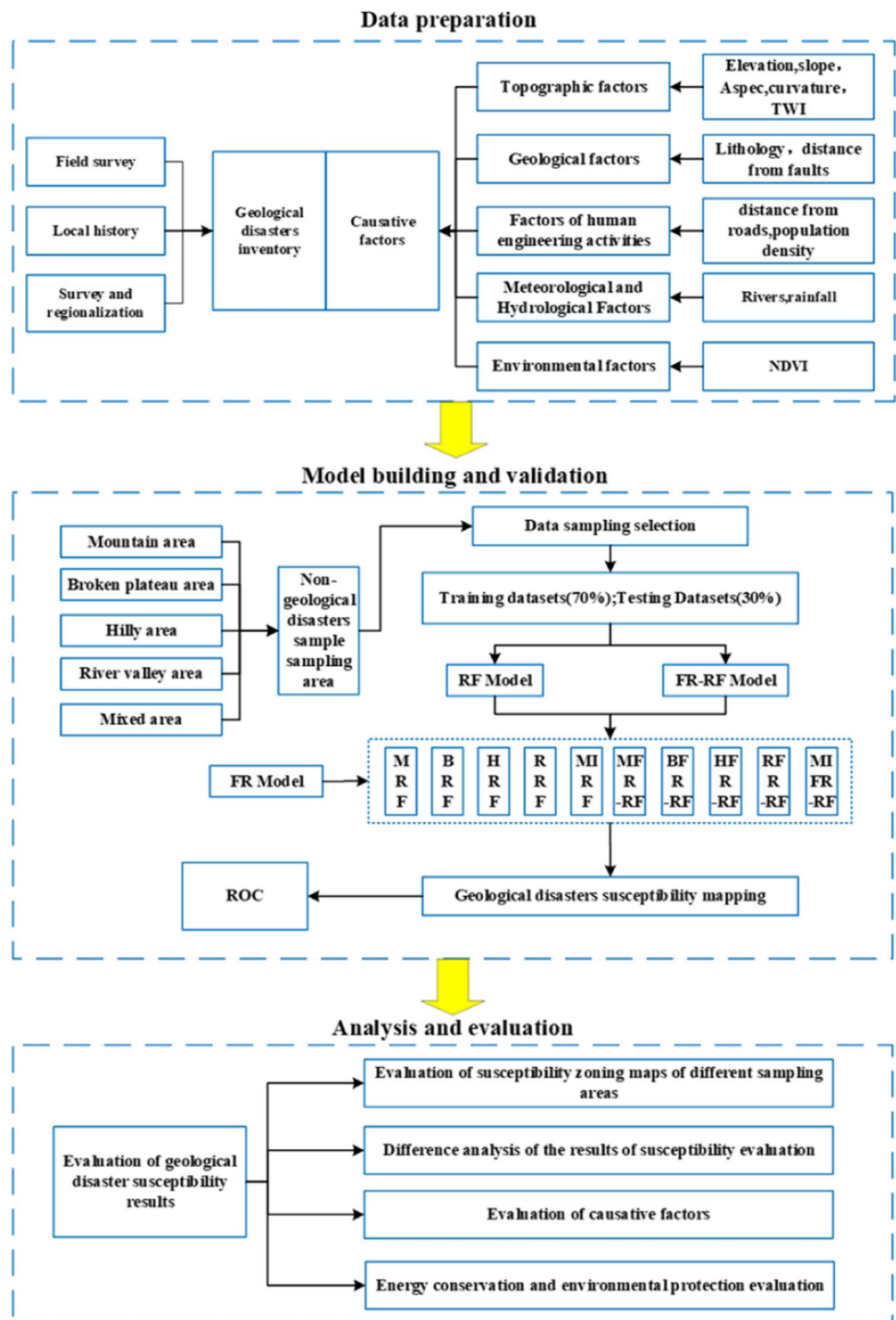


Fig. 1 Location of the study area: **A** This is a map of geological disasters in the study area; **B** This is the macro geomorphology division map of the study area

Fig. 2 Flow chart of this study



Model

Frequency ratio (FR model)

As a statistical analysis model, frequency ratio was widely used in the study of geological disasters susceptibility. Different categories of factors had different contributions to geological disasters development, and FR was used to

quantify this difference. The higher the frequency ratio, the greater the geological disasters susceptibility. The sum of the frequency ratios of each factor subclass was used to determine the susceptibility index value (Aditian et al. 2018). The calculation method in practical application was as follows:

$$FR_{ij} = \frac{N_{ij}}{N_r} / \frac{A_{ij}}{A_r} \tag{1}$$

where FR_{ij} is the frequency ratio; N_{ij} is the number of geological disasters grids at the j th level under the i th influencing factor; N_r is the grid number of geological disasters 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.

Random forest (RF model)

Originally developed by Breiman (2001), random forest was an ensemble model that combines multiple decision trees (DTs) for classification and prediction (Rabby et al. 2020). Taalab pointed out that there was no rule for determining the amount of numbers in the RF, but the accuracy of the model did not increase with the number of trees (Taalab et al. 2018). The RF algorithm was explained in detail by Catani et al. (2013). The random forest algorithm included the following steps: (1) obtain the original training data and resample it multiple times; (2) at each resampling, select a set of random features; (3) given the resampling and a set of random features, estimate the decision tree; (4) aggregate the set of estimated decision trees in order to get a single decision tree. Different from the learning model of other machines, the randomly sampled samples and features in the random forest model reduced the sensitivity of data noise and outliers in the classification process, and effectively avoided over fitting.

One of the characteristics of random forest was that the importance of the susceptibility evaluation factor can be given. The reduction of the Gini coefficient was used to calculate the importance of each factor to the susceptibility classification result. The formula is as follows:

$$P_r = \frac{\sum_{i=1}^k \sum_{j=1}^t D_{Grij}}{\sum_{r=1}^m \sum_{i=1}^k \sum_{j=1}^t D_{Grij}} \quad (2)$$

where m , k , and t are the total number of evaluation factors, the number of classification trees, and the number of nodes in a single tree respectively; D_{Grij} is the reduction value of the Gini coefficient in the i th evaluation factor at the j th node of the r th tree; P_r is the importance of the r th evaluation factor among all evaluation factors.

Coupling model

Pourghasemi, Zhou, and Chowdhuri obtained high prediction results by mapping and evaluating geological disaster susceptibility through coupled models (Chowdhuri et al. 2020; Pourghasemi and Kerle 2016; Zhou et al. 2021b). As a statistical analysis model, FR can be connected with environmental factors, and reassign environmental factors. The assigned

factor can be used as the input variable of the machine learning model to build a new model (Li et al. 2021).

Model validation method

Previous researchers mostly used receiver operating characteristic (ROC) and area under the curve (AUC) to test and evaluate the effect of the model comprehensively (Ayalew et al. 2005; Chen et al. 2020; Zhang et al. 2019). The ROC curve reflected the relationship between “TPR” and “FPR”.

$$TPR = \frac{TP}{TP + FN} \quad (3)$$

$$FPR = \frac{FP}{FP + TN} \quad (4)$$

where TP is the true positive rate; FN is the false negative rate; TN is the true negative rate; FP is the false positive rate. The value of AUC ranged from 0.5 to 1. The higher the value, the higher the accuracy of the model.

Geologic disaster causative factors

An important stage for geological disasters susceptibility analysis was to determine the causative factor that affected the occurrence of disasters in the research area. Vakhshoori and others pointed out that there were no standard rules for the selection of disaster causing factors, and it can be selected according to the study area scale, the topographic and geological environmental conditions, geological disaster occurrence mechanism, and data availability in the research area (Vakhshoori et al. 2019). According to the survey results of the study area, the comprehensive action of landform, stratigraphic lithology and structure, meteorology and hydrology, vegetation, and human engineering activities led to the occurrence of geological disasters such as collapse and landslide. Considering these standards and combining previous experience (Duan et al. 2022b; Wang 2019), a total of 12 causative factors were used through GIS platform, namely, elevation, slope, aspect, curvature, lithology, distance from faults, distance from roads, population density, distance from rivers, rainfall, normalized difference vegetation index (NDVI), and topographic wetness index (TWI).

The elevation (DEM) data (which can generate slope, aspect, curvature, water system, and topographic wetness index data), road vector data, and remote sensing data (which can generate normalized vegetation cover index) of Liulin County were obtained through Geospatial Data Cloud. The tomographic data of the study area were obtained through satellite images; the average rainfall and population density data over the years were obtained

through the Resource and Environmental Science and Data Center of the Chinese Academy of Sciences; the lithology data were obtained from the 1:100,000 geological map of Lvliang City (Wang 2019).

Topographic factors

Elevation was the most direct representation of a region's topography and was frequently used to analyze geological disasters susceptibility (Hu et al. 2020; Shou and Lin 2020; Wu et al. 2020). The elevation of the study area was ranged between 597 and 1467 m, and the terrain was high in the east and low in the west. In ArcGIS, it was divided into 7 grades according to the natural discontinuity method (Fig. 3A), and the geological disasters were most distributed between 840 and 922 m, accounting for about 42.38% of the total.

Slope was one of the key slope stability factors and one of the most commonly used parameters in geological disasters susceptibility analysis (Bordoni et al. 2020; Zhang et al. 2020). The slope of the study area was ranged between 0 and 78°, and it was divided into 7 grades according to the natural discontinuity method (Fig. 3B). The geological disasters were concentrated between 14 and 33°, accounting for about 60.26% of the total.

The slope aspect affected the duration and intensity of sunshine, the distribution characteristics of rainfall on the slope (Erener and Duzgun 2013). Starting from 22.5°, every 45° was a direction (337.5 ~ 360° and 0 ~ 22.5° were the north direction); - 1 was set as the plane direction, 9 directions in total. Geological disasters were not distributed in the plane, but evenly distributed in the other 8 directions (Fig. 3C).

Curvature is one of the factors associated with the occurrence of geological disasters, representing the shape and terrain of the slope (Pourghasemi et al. 2018). The curvature of the research area was divided into level 4 (Fig. 3D), and geological disasters were mostly distributed between - 2 and 0, accounting for about 66.89% of the total.

Topographic wetness index (TWI) is an indicator of the influence of regional topography on runoff direction and water accumulation, which indicates the change of soil moisture (Saleem et al. 2019). The TWI of the study area is divided into seven levels (Fig. 3L), and the most geological disasters are distributed between 8.1 and 10, accounting for about 48.34% of the total.

Geological factors

Lithology determined the physical and chemical properties of rock mass (Sun et al. 2018), and was an important factor in the formation and evolution of geological disaster. According to the mechanical and other properties of geotechnical, the study area was divided into 4 categories (Fig. 3E), among which geological disasters were most distributed in loess, accounting for about 57.62% of the total, and no distribution in magmatic rocks.

Geological tectonic movements deteriorated the mechanical properties of rocks (Hong et al. 2015). There were only three faults in the eastern part of the study area, and the distances were divided into 7 grades according to the natural discontinuity method (Fig. 3F). The farther from the fault, the more geological disasters were distributed. The distribution of geological disasters beyond 5 km accounted for 94.04% of the total.

Factors of human engineering activities

Factors of human engineering activities reflected the scope and intensity of human influence on the natural environment. Frequent interference, especially large-scale human construction activities may cause slopes to lose stability (Kanwal et al. 2017). Most geological disasters occur in high population density or within 1 km from the roads. With the increase of distance, the number of geological disasters decreased rapidly (Wang et al. 2021). The population density in the study area was divided into 4 grades (Fig. 3H), among which geological disasters were most distributed between population densities of 200 and 300, accounting for about 90.73% of the total. The distance from roads was divided into 7 grades according to the natural discontinuity method (Fig. 3G), among which the geological disasters were most distributed between 437 m from the road, accounting for 78.81% of the total.

Meteorological and hydrological factors

The flushing of the rock and soil on the surface of the slope by rivers was one of the important reasons for the occurrence of geological disasters (Meinhardt et al. 2015). The distance analysis was carried out in ArcGIS and divided into 7 grades (Fig. 3I), among which the geological disasters distribution frequency within 391 m from rivers was the highest, accounting for about 59.6% of the total.

Rainfall was one of the most important factors that induce geological disasters. The surface water formed by rainfall would not only wash the slope, but also penetrate and soften the rock and soil (Wang et al. 2020), reducing the stability of the slope. Usually, the greater the rainfall, the higher the frequency of disasters. Based on the annual average data of rainfall in the study area, the interpolation analysis was carried out based on ArcGIS and divided into 7 grades according to the natural discontinuity method (Fig. 3J), among which the geological disasters were most distributed within the annual rainfall of 470 to 478 mm, accounting for 42.38% of the total.

Environmental factors

Vegetation improved the slope stability by absorbing water and strengthening rock and soil. Generally speaking, the higher the vegetation coverage, the stronger

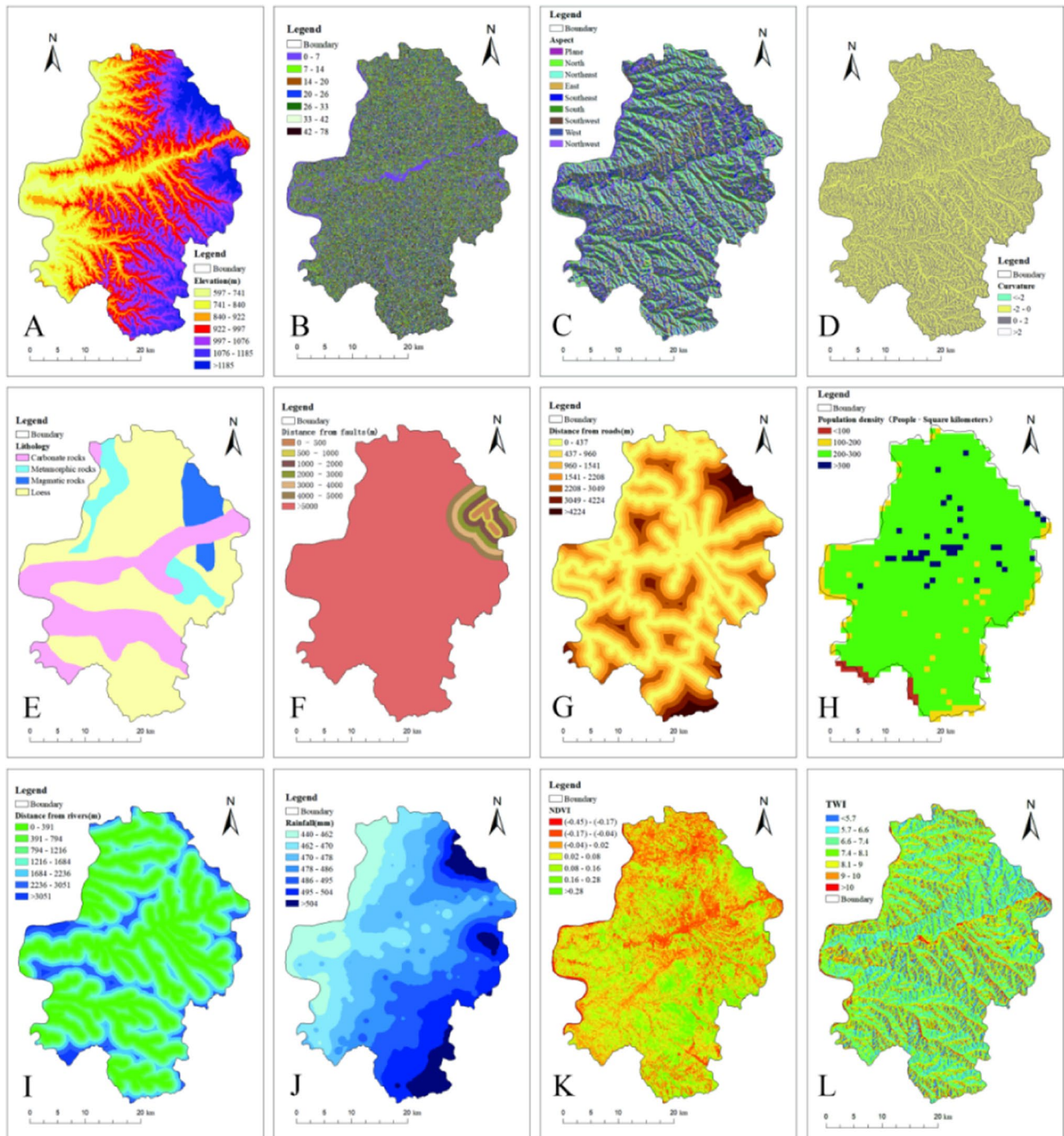


Fig. 3 Factors classification: **A** Elevation; **B** Slope; **C** Aspect; **D** Curvature; **E** Lithology; **F** Distance from faults; **G** Distance from roads; **H** Population density; **I** Distance from rivers; **J** Rainfall; **K** NDVI; **L** TWI

the slope stability (Chen et al. 2017). The NDVI was a digital representation of the surface vegetation density, which ranged from -1 to 1 , and the closer to 1 the value was, the lush the vegetation coverage in the area. The vegetation coverage index was obtained after analyzing

and processing the images of remote sensing. According to the natural discontinuity method, it was divided into 7 grades (Fig. 3K). The geological disasters were most distributed between -0.17 and 0.02 , accounting for about 60.93% of the total.

Results

Model building

According to Wang’s (2019) research experience in this area, the sample selection ratio of geological disasters and non-geological disasters is 1:1. Combined with non-geological disasters samples (151 samples from each sampling area) randomly selected in five sampling areas (mountain area, broken plateau area, hilly area, river valley area, and mixed area) to construct ten models of mountain RF (MRF), broken plateau area RF (BRF), hill RF (HRF), river valley RF (RRF), mixed RF (MIRF), mountain FR-RF (MFR-RF), broken plateau area FR-RF (BFR-RF), hill FR-RF (HFR-RF), river valley FR-RF (RFR-RF), and mixed FR-RF (MIF-RF). Susceptibility zoning map is divided into five levels according to the natural discontinuity method, namely very low, low, moderate, high, and very high. Among them, the distance between non-geological disasters samples should be more than 500 m, and 500 m away from the known geological disasters samples.

Geological disasters susceptibility mapping by FR model

FR model is one of the statistical analysis models. According to Eq. (1), calculate the frequency ratio of each classification in each causative factor (Table 1). All the calculated

frequency ratios are loaded into the raster data of the study area, so that each raster will store the frequency ratios of the corresponding classification of the 12 causative factors. The total frequency ratio of the grid can be obtained by summing the frequency ratios of all the disaster causing factors in the grid. According to the interpolation analysis of the total frequency ratio of each grid, the susceptibility map of the study area was obtained and divided into five levels (Fig. 4, Fig. 5).

Geological disasters susceptibility mapping by RF model

RF model is a machine learning algorithm that requires input of initial variables. Attribute values of each evaluation index of disaster and non-disaster samples were extracted based on ArcGIS platform. The sample was divided into training sets and test sets according to 7:3 (Hussain et al. 2021; Sahin et al. 2020; Zhou et al. 2021a). A total of 302 data samples (151 disaster data and 151 selected non-disaster samples) from five sampling areas were successively brought into the RF model for training and testing. The test accuracy was all higher than 0.9, indicating that the model had high accuracy. By bringing the original value of each evaluation index of 14,377 units in the study area into the trained model, the probability of geological disaster occurrence in each unit can be obtained. Based on the probability of geological disaster occurrence in each unit, the zoning diagram of the vulnerability of each model can be obtained by interpolation analysis according to the natural discontinuity method (Fig. 4).

Table 1 Causative factors classification range and FR values

	Classification range	FR values
Elevation (m)	597 ~ 741; 741 ~ 840; 840 ~ 922; 922 ~ 997; 997 ~ 1076; 1076 ~ 1185;> 1185	1.5177; 0.9293; 2.0969; 1.0041; 0.4739; 0.1146; 0
Slope (°)	0 ~ 6; 6 ~ 10; 10 ~ 13; 13 ~ 17; 17 ~ 21; 21 ~ 26;> 26	1.4023; 1.3351; 0.8429; 0.9190; 0.7801; 0.8115; 0.3808
Aspect (°)	Plane; North; Northeast; East; Southeast; South; Southwest; West; Northwest	0; 1.2603; 0.9210; 1.2007; 1.3458; 0.8592; 0.7007; 0.7563; 1.2695
Curvature	< -2; -2 ~ 0; 0 ~ 2;> 2	0; 1.4960; 0.5999; 0
Lithology	Carbonate rocks; Metamorphic rocks; Magmatic rocks; Loess	0.9510; 1.8600; 0; 1.0140
Distance from faults (m)	0 ~ 500; 500 ~ 1000; 1000 ~ 2000; 2000 ~ 3000; 3000 ~ 4000; 4000 ~ 5000;> 5000	0; 1.4994; 0.8352; 1.1852; 0; 0.8035; 1.0250
Distance from roads	0 ~ 437; 437 ~ 960; 960 ~ 1541; 1541 ~ 2208; 2208 ~ 3049; 3049 ~ 4224;> 4224	2.6823; 0.2617; 0.3569; 0.3189; 0.3886; 0; 0
Population density (/km ²)	< 100; 100 ~ 200; 200 ~ 300;> 300	0.6521; 0.5005; 0.9984; 2.1158
Distance from rivers (m)	0 ~ 391; 391 ~ 794; 794 ~ 1216; 1216 ~ 1684; 1684 ~ 2236; 2236 ~ 3051;> 3051	2.4118; 0.4460; 0.6174; 0.5260; 0.6493; 0.5391; 0
Rainfall (mm)	440 ~ 462; 462 ~ 470; 470 ~ 478; 478 ~ 486; 486 ~ 495; 495 ~ 504;> 504	1.2067; 0.9045; 1.8261; 1.2475; 0.4966; 0.2548; 0
NDVI	- 0.45 ~ - 0.17;- 0.17 ~ - 0.04;- 0.04 ~ 0.02; 0.02 ~ 0.08; 0.08 ~ 0.16; 0.16 ~ 0.28;> 0.28	0.9918; 2.2916; 1.1364; 0.6615; 0.7839; 0.2355; 0.4481
TWI	3.7 ~ 5.7; 5.7 ~ 6.6; 6.6 ~ 7.4; 7.4 ~ 8.1; 8.1 ~ 9; 9 ~ 10; 10 ~ 15.2	0.8741; 0.8839; 0.6176; 0.7542; 1.4154; 1.6543; 0.6960

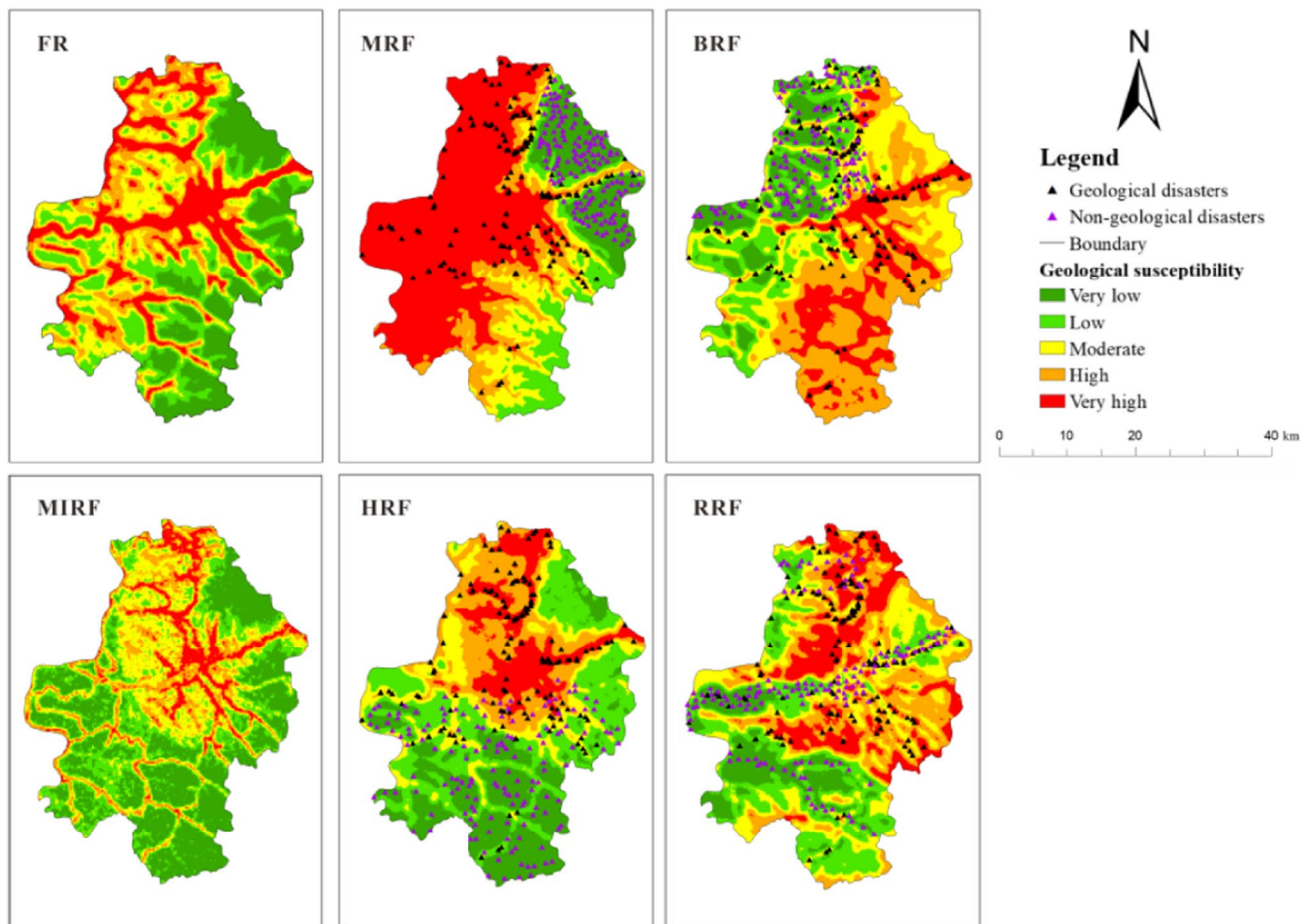


Fig. 4 The geological disasters susceptibility zoning maps by RF model

Geological disasters susceptibility mapping by FR-RF model

The FR model can be connected with the causative factor, and the causative factor can be reassigned. The assigned factor can be used as the input variable of the machine learning model to build a new model. The frequency ratio of each factor was taken as the input variable into the RF model, which was divided into training set and test set by 7:3. Similarly, the 14,377 assessment units in the study area were also reassigned with FR, and the assigned 14,377 assessment units were brought into the previously trained model to obtain the probability of occurrence of geological disasters in each unit. The vulnerability map can be obtained by interpolation analysis of generated probability in ArcGIS. The research can be divided into five levels by natural discontinuity method, namely, very low, low, moderate, high, and very high (Fig. 5).

Model validation

ROC curve is simple and intuitive, which is a typical method to evaluate the prediction ability of models. When AUC

value of the area under ROC curve is greater than 0.8, the prediction effect of the model is good; when AUC value is greater than 0.9, the prediction effect is excellent. By comparing the models in Figs. 4 and 5 with the FR model, it can be seen that the prediction results of BRF, RRF, BFR-RF, and HFR-RF have large errors with the results of the FR model, so the ROC curve test of these four models is not carried out. Figure 6 shows that AUC values of all models are higher than 0.8, and the prediction effect is good.

Discussion

Evaluation of susceptibility maps of different area

It can be seen from Fig. 4 and Fig. 5 that when non-geological disaster samples are sampled in the valley area, most of the valley areas are low and very low susceptibility areas, and there are a large number of areas above the low susceptibility areas in the northeast and southeast mountain areas. When non-geological disaster samples are sampled in the broken

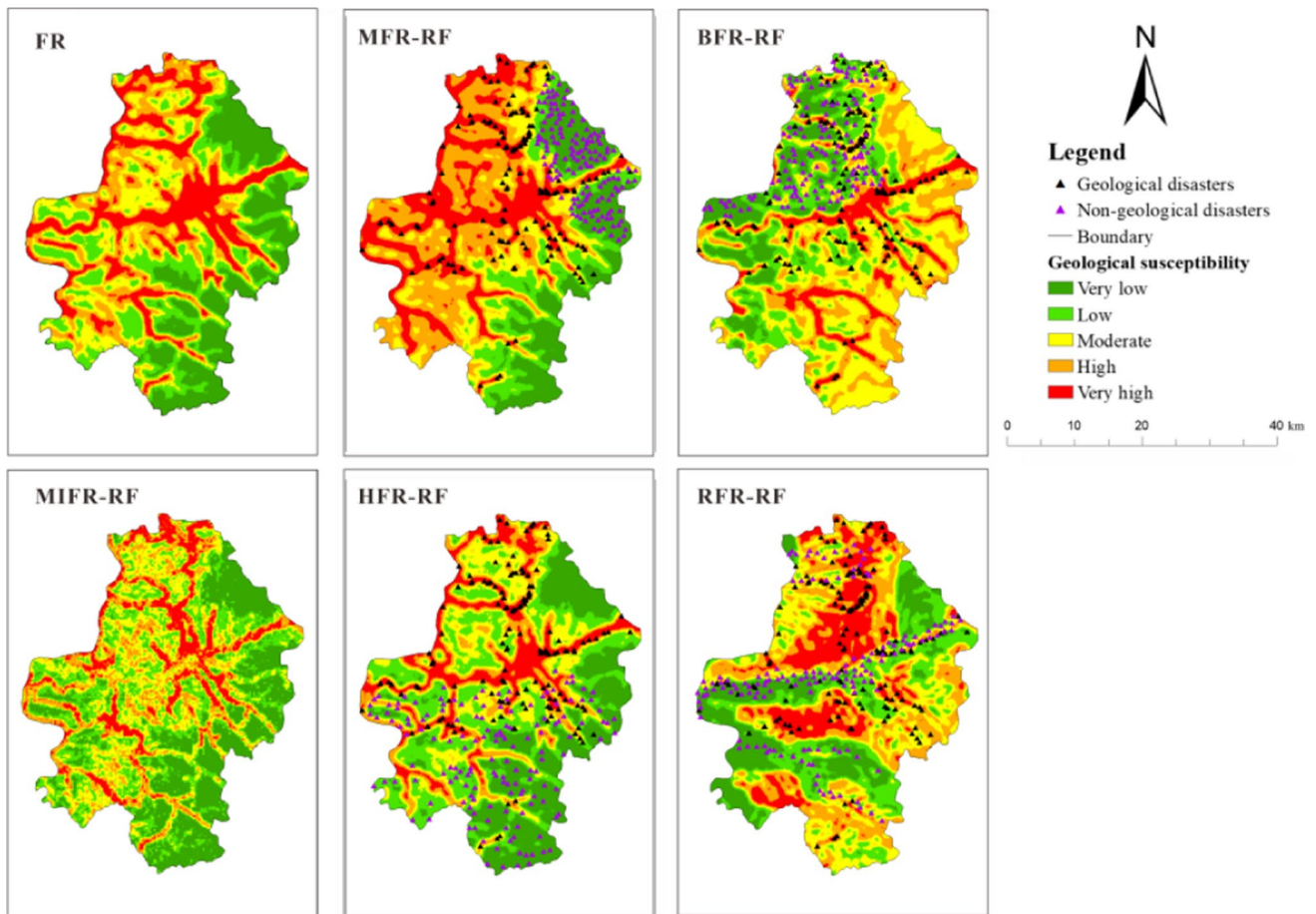


Fig. 5 The geological disasters susceptibility zoning maps by FR-RF model

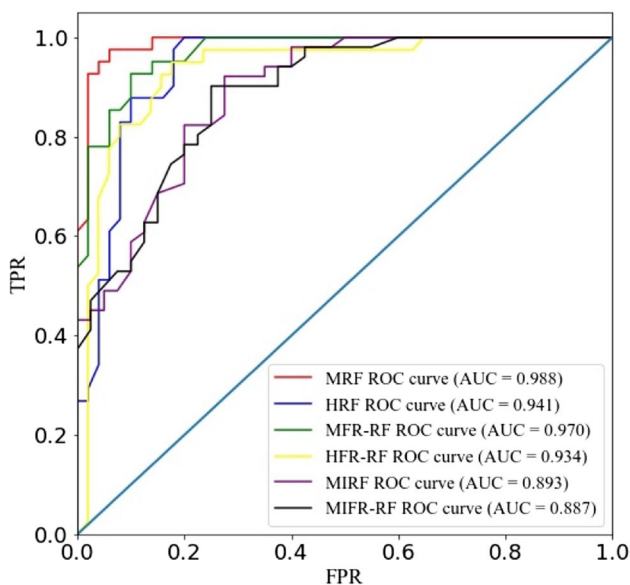


Fig. 6 ROC curves

plateau area, most of the northeast and southeast areas are above the middle susceptibility areas. Compared with the FR model, the spatial distribution error of the model constructed in these two sampling areas is great. Therefore, it is not suitable for the study area to select non-geological disaster samples in the broken plateau and river valley as sampling areas. When non-geological disaster samples are sampled in hilly area, mountainous areas and mixed area, and river valley areas were mostly high and very high susceptibility areas, while mountainous areas are mostly low and very low susceptibility areas. Compared with FR model, the overall spatial distribution is basically consistent. As can be seen from the MRF susceptibility zoning map, the study area has a large area and a large area of continuous distribution, which does not conform to the actual situation of the study area. The large area and continuous distribution in HRF susceptibility zoning map has decreased a lot, but the whole north, especially the northwest, is almost all above the low susceptibility area. From the perspective of urban construction, if most of the research area is in the very high susceptibility area of geological disasters, it will not be suitable for urban construction and human production and life. But there are many towns and



Fig. 7 3D map of northwest

even city centers in the west and middle. As can be seen from the susceptibility zoning map of FR model in the northwest region, there are scattered very low and low susceptibility areas, so the prediction accuracy of MRF and HRF models is low. According to the susceptibility zoning maps of MFR-RF, HFR-RF, and MIRF and MIFR-RF, the situation of large area continuous distribution has basically disappeared, which is closer to the actual situation than the prediction results of MRF and HRF models. However, MFR-RF is still moderate and high susceptibility areas in most northwest areas. Many villages and towns are distributed in the northwest, such as Mengmen Town, Chengjiazhuang Town, Houfengjiagou Village, Yaotou Village, and Majiata Village (Fig. 7). Similarly, the FR model shows that the northwest is scattered with very low and low susceptibility areas, which is also in contradiction with the reality. The FR model showed high and very high susceptibility area along the central road and on both sides of the Sanchuan River, while the MIRF and MIFR-RF models showed low and moderate susceptibility areas, which were inconsistent with each other. The HFR-RF model is the most consistent model with the distribution of FR model susceptibility zoning map. Similarly, the HFR-RF model also reflects the advantages of the machine learning model, that is, it is more accurate and more accurate than the statistical analysis model. For example, there are more low susceptibility areas in the northwest region, while the high and very high susceptibility areas in the central and northern regions are much reduced. Fifty-one percent of the HFR-RF model belongs to low and very low susceptibility areas, and there is no large and continuous distribution of high and very high susceptibility areas. At the same time, AUC value is higher than MIRF and MIFR-RF. The susceptibility zoning map is reasonable and meets the requirements of urban construction. Therefore, the FR-RF model with non-geological disaster samples sampled in hilly areas is most suitable for the

susceptibility mapping of this study area. Li et al. (2022a) selected 902 negative samples unconditionally and randomly in the study area to evaluate the susceptibility of geological disasters combined with IV and IV-LG models, and the accuracy of the two models was almost the same, without any improvement. Duan et al. (2022a) randomly generated an equal amount of non-disaster points 100 m away from the disaster point, and the susceptibility was evaluated by RF and SVM, and the AUC value of ROC curve test was not more than 0.72. An and Niu (2016) combined with SVM to evaluate the susceptibility beyond 500 m of the disaster point, and the model training accuracy was 81.41%. Kavzoglu et al. selected random negative samples in river channels and low slope areas to evaluate landslide susceptibility by MCDA and SVR methods, and the results showed that the predicted results of these two methods were higher than those of LG model (Kavzoglu et al. 2014). Wang Shibao et al. selected negative samples in areas with low information value model and very low susceptibility, and used CNN model to evaluate landslide susceptibility. Compared with ANN model, AUC value only increased by 0.03 (Wang et al. 2022b). There are also some methods to select negative samples by special methods (Chen et al. 2021; Su et al. 2022). Although these studies have obtained good results, there is no discussion on whether the accuracy of negative sample model selection can be improved by other methods. Through this study, it can be seen that the selection accuracy of negative samples in different regions varies greatly, and the selection of negative samples in a reasonable region will greatly improve the accuracy of the model. Therefore, the reasonable selection of non-geological disaster samples is of great significance to improve the accuracy of the susceptibility evaluation results.

According to the HFR-RF susceptibility zoning map in Fig. 5, very low susceptibility areas are concentrated in the south, southeast, and northeast regions. These areas have

higher elevations and steeper slopes, and rain flows naturally to lower elevations. The vegetation is luxuriant, the ability of soil fixation is strong, and the rainwater that penetrates into the soil is also absorbed by the vegetation. At the same time, it is far away from densely populated areas, and human engineering is very little, which makes it a very low risk area for geological disasters. The low susceptibility area is within the very low susceptibility area, mainly distributed in the southwest and west, scattered in the northwest. In these areas, the elevation, rainfall and NDVI are lower than those in the very low susceptibility area, and human engineering activities are still less. The moderate susceptibility areas are mainly distributed in the central and northern areas, and scattered in the southwest. These areas have been close to densely populated areas, with more roads and the elevation is in the transition zone of the study area. Rainfall has decreased a lot, vegetation coverage has become low, and some areas have been in very low vegetation coverage area, and water system has become more. The high susceptibility areas are mainly distributed in the central and northern regions. The surface layer of these areas is mostly covered by loess, with very low vegetation coverage, weak soil fixation ability, minimal rainfall, and serious soil erosion. At the same time, it is located in the low-elevation area of the study area, close to the roads, strong human activities and developed water system, making it a highly susceptibility area. High rock along the roads and river on both sides of the concentration distribution in study area, especially in central and northern is most concentrated. These areas are human concentration areas, roads is more, human engineering activities destruction of slope stability. Although the rainfall is small or very small, the vegetation coverage is very low, and the altitude is low. Most of them are hilly gully and valley terrace. The mountain rainfall flows to the region, and the rainwater seeps into the rock and soil, making the slope that was originally damaged by the project more unstable. Moreover, there are many water systems, which have a strong erosion capacity on the slope. Many factors together to make these areas become extremely susceptibility area and key area of Liulin County geological disaster control and prevention.

Difference analysis of the results of susceptibility evaluation

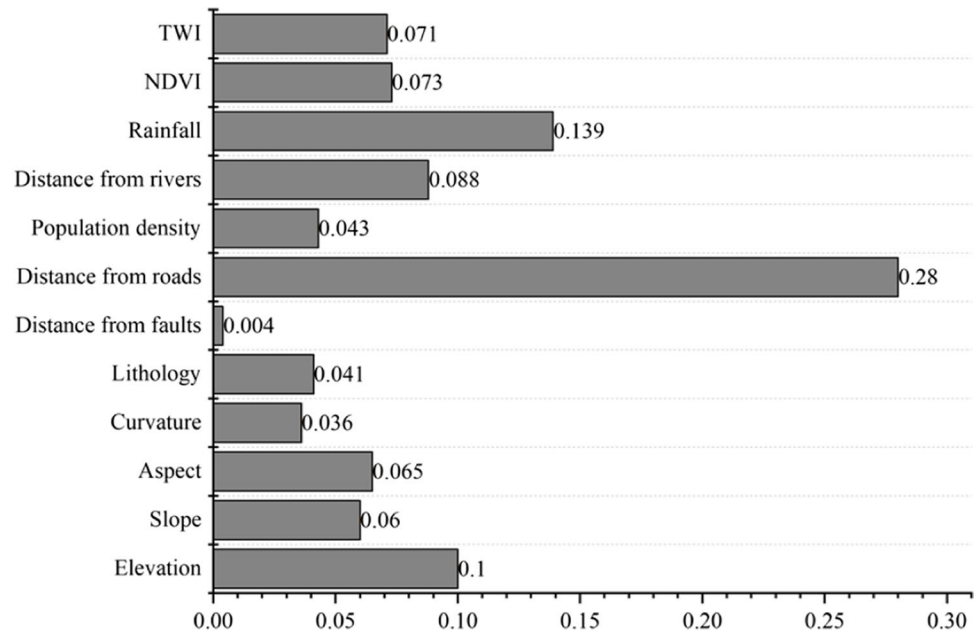
Negative samples were randomly selected from different sampling area to build the geological disasters susceptibility model have great difference. From the perspective of RF model itself, RF model is the result of classification and comprehensive evaluation of many decision trees. When the training set of sample quantity change, each the result of a decision tree to select feature will change. When many trees are clustered, the model judgment will be different, that is, the probability of geological disasters in each grid will

change accordingly. The most important reason is whether the randomly selected negative samples are representative. The negative samples selected by BRF, BFR-RF, RRF, and RFR-RF models are located in the broken plateau and river valleys, which are the high-frequency areas of the distribution of geological disaster points. Therefore, the causative factor attribute values stored by negative samples are very similar to those of positive samples. The characteristics of positive and negative samples may be confused in the process of RF model construction. The opposite is actually the case (Fig. 4BRF). The MRF model will have a continuous large area distribution, because the negative sample selected in the mountainous area has a relatively “extreme” attribute value of the causative factor. For example, rainfall, elevation, slope, and NDVI are generally the highest in the study area, while lithology is generally the most stable, and it is generally far away from roads and water systems. So when modeling, areas that do not fit these characteristics are considered not very low or low susceptibility areas. However, the negative samples selected in the hilly area are representative, and their attribute characteristics include the characteristics of the whole study area, and there is no case of “extreme” attribute value of disaster causing factors. Secondly, it is not in the high frequency area of the distribution of geological disaster points, so there will be no situation contrary to the actual situation, which is also consistent with the research results of Wang et al. (2022a). The prediction effect of FR-RF model is better than that of RF model, because the coupled model can eliminate the differences in dimensions, properties and other aspects of RF model, and ensure the objectivity of causative factors in model analysis.

Evaluation of causative factors

According to the weight contribution of the disaster causing factors of the FR-RF model sampled by non-geological disaster samples in hilly areas (Fig. 8), the distance from the road, the distance from the river, elevation, and rainfall are the most important factors affecting the occurrence of geological disasters in the study area. Among them, the greatest contribution is the distance from the road. According to Fig. 3, Fig. 5, and Table 1, the areas with very high susceptibility are almost all distributed on both sides of the road. The farther away from roads, the fewer disaster points, and the trend declined rapidly. From this, it can be judged that the construction of roads in the study area was the most important factor affecting the occurrence of geological disasters. Elevation and rainfall were also the key factors for it. Since the occurrence of geological disasters was the result of the joint action for multiple factors, the study area was not an area that the higher the elevation, the greater the rainfall, the more susceptible to disasters. For example, the impact of rainfall on geological disasters

Fig. 8 Contributions of causative factors



should be considered together with NDVI. The northeast and southeast regions were the areas with the highest elevations and the most rainfall in the study area. However, these areas were covered with lush vegetation, which had a strong soil-fixing ability and absorbed more rainfall. At the same time, due to the high elevation and relatively large slope aspect, the rainfall also flowed to the low-elevation areas. In addition, there was less human engineering and the population density was relatively low, so it became a very low susceptibility zoning in the study area. However, the central and north-central regions were in the transition zones of elevation and rainfall in research area. Due to the sparse coverage of vegetation, the ability to solidify the soil was not strong. And there were many mountain streams and valley bottoms, so the rainwater flowed to these areas. At the same time, the population density was large and so was the amount of human engineering. Although the slope was not very steep, it was the most susceptible factor to geological disaster in the research area. River was also one of the most important factors in the research area. For example, the Sanchuan River and its tributaries crossed the central regions of the research area. Then combining the common effects of other factors, it would make the area susceptible for geological disaster.

Energy saving and environmental protection

Land-use and GDP are important evaluation indicators of ecological environmental security and sustainable economic development (Zhao et al. 2006). The cost of disaster management and potential economic loss are closely related to the accuracy of geological disaster susceptibility zoning

map, and the accuracy has become the focus of geological disaster prevention and control (Pourghasemi et al. 2019). In the process of GDP rasterization, the GDP distribution weights of land use types, night light brightness, and residential density are mainly considered, and the calculation formula is as follows:

$$GDP_{ij} = GDP \times \left(\frac{Q_{ij}}{Q} \right) \quad (5)$$

where DP_{ij} is the value of grid cells after spatialization; GDP is the GDP statistic value of the county district unit where the grid cell is located; Q_{ij} is the total weight of land use type, night light brightness and residential density of the grid cell. Q is the total weight of land use type, night light brightness and residential density of the county administrative unit where the grid cell is located.

The cost of disaster management is mainly estimated through labor costs (LC), material costs (MC), and construction machinery usage costs (CMUC) (Eq. 6), and the damage rate (in this paper, it is expressed as GDP/disaster management cost) is used as an objective quantitative indicator of damage degree (Yum et al. 2020). The susceptibility zoning map, land-use, and GDP distribution map were superimposed to calculate the total GDP value of each model and the potential threat of soil resource information, and to comprehensively analyze the potential impact of regional geological disasters on economic development.

$$DGC = LC + MC + CMUC \quad (6)$$

where DGC is the disaster governance costs, LC is the labor costs, $LC = \sum$ working hours \times unit price, MC is the

Table 2 Ratio of economic data

	MFR-RF	HFR-RF	MRF	HRF	MIRF	MIFR-RF
DGC	0.819	0.603	1.000	0.621	0.533	0.604
GDP loss value	0.822	0.607	1.000	0.630	0.537	0.593
Damage ratios	1.003	1.007	1.000	1.015	1.007	0.982

material costs, $MC = \sum$ material consumption \times unit price, $CMUC$ is the construction machinery usage costs, $CMUC = \sum$ machinery consumption \times unit price.

Under each model, the ratio of disaster management costs, total GDP, and loss rate in different sampling areas (Table 2) showed that the ratio of disaster management costs and GDP loss of HFR-RF model is the lowest, and its economic benefit is the best. In line with the above, HFR-RF model is the most suitable model in the study area, and its management costs decreases by 13.45% on average compared with other models. This indicates that using FR-RF model to sample non-geological disaster samples in hilly areas has better effect on energy conservation and environmental protection, and can effectively reduce the cost of disaster management.

Conclusion

Accurate and scientific analysis of geological disaster susceptibility is the key step to ensure the safety of ecological environment and sustainable development of regional economy. This study innovatively evaluated the susceptibility of negative samples by dividing them into macro-geomorphic areas, and the results showed that the method was feasible. However, the susceptibility results of negative samples selected in different sampling areas are quite different, mainly because of the different representativeness of negative samples selected. Therefore, appropriate topographic and geomorphic areas must be selected for sampling by this method. In this paper, the prediction effect of negative samples selected in hilly area is excellent, which is higher than the prediction accuracy of other sampling areas and previous researchers. By comparing energy conservation and environmental protection, the disaster management cost of HFR-RF model is reduced by 13.45% on average compared with other models, and its GDP/disaster management cost ratio is significantly increased. This study can provide a scientific basis for the selection of negative samples for geological hazard susceptibility assessment in the Loess Plateau of North China.

Author contribution Conceptualization, Jianping Chen, Zepeng Wang; methodology, Jianping.Chen; software, Zepeng Wang; validation, Wei Chen; formal analysis, Changyuan Wan and Yunyan Liu; investigation, Junjie Huang; resources, Jianping Chen; data curation, Zepeng Wang.

and Yunyan Liu.; writing—original draft preparation, Jianping Chen and Zepeng Wang; writing—review and editing, Jianping Chen and Wei Chen; visualization, Zepeng Wang; supervision, Junjie Huang; project administration, Yunyan Liu.; funding acquisition, Jianping Chen; All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by the 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, Zepeng Wang, upon reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Aditian 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>
- An KQ, Niu RQ (2016) Landslide susceptibility assessment using support vector machine based on weighted-information model. *JYRSR* 33:47–51+58. <https://doi.org/10.11988/ckyyb.20150311>
- Arabameri A, Yamani M, Pradhan B, Melesse A, Shirani K, Bui DT (2019) Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility. *Sci Total Environ* 688:903–916. <https://doi.org/10.1016/j.scitotenv.2019.06.205>
- Ayalew L, Yamagishi H, Marui H, Kanno T (2005) Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Eng Geol* 81:432–445. <https://doi.org/10.1016/j.enggeo.2005.08.004>
- 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>
- 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>
- 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>
- 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 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* 9:609896. <https://doi.org/10.3389/feart.2021.609896>
- Chowdhuri I, Pal SC, Arabameri A, Ngo PTT, Chakraborty R, Malik S, Das B, Roy P (2020) Ensemble approach to develop landslide susceptibility map in landslide dominated Sikkim Himalayan region, India. *Environ Earth Sci* 79:476. <https://doi.org/10.1007/s12665-020-09227-5>
- Deng ND, Shi H, Wen Q, Li YQ, Cao XF (2021) Collapse susceptibility evaluation of random forest model supported by information value model. *Sci Techn Eng* 21:2210–2217. <https://doi.org/10.3969/j.issn.1671-1815.2021.06.012>
- 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>
- Duan YY, Tang J, Liu YG, Gao XJ, Duan YX (2022a) Spatial sensitivity evaluation of loess landslide in Liulin County, Shanxi based on random forest. *S Geog Sin* 42:343–351. <https://doi.org/10.13249/j.cnki.sgs.2022.02.016>
- Duan YY, Tang J, Liu YG, Gao XJ, Duan YX (2022b) Spatial sensitivity evaluation of loess landslide in Liulin County, Shanxi based on random forest. *S Geog Sin* 42:343–351. <https://doi.org/10.13249/j.cnki.sgs.2022.02.016>
- 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>
- 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>
- Gariano SL, Guzzetti F (2016) Landslides in a changing climate. *Earth-Sci Rev* 162:227–252. <https://doi.org/10.1016/j.earscirev.2016.08.011>
- 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 Q, Zhou Y, Wang SX, Wang FT (2020) Machine learning and fractal theory models for landslide susceptibility mapping: case study from the Jinsha River Basin. *Geomorphology* 351:106975. <https://doi.org/10.1016/j.geomorph.2019.106975>
- 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>
- Huang QB, Kang XS, Wang QY, Li W (2016) Types and characteristics of loess landslides and collapses in Loliang area of Shanxi Province. *J Eng Geol* 24:64–72. <https://doi.org/10.13544/j.cnki.jeg.2016.01.008>
- Huang P, Peng L, Pan HY (2020) Linking the Random Forests Model and GIS to Assess Geo-Hazards Risk: A Case Study in Shifang County, China. *Ieee* 28033–28042. <https://doi.org/10.1109/ACCESS.2020.2972005>
- 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>
- Kanungo DP, Arora MK, Sarkar S, Gupta RP (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Eng Geol* 85:347–366. <https://doi.org/10.1016/j.enggeo.2006.03.004>
- 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>
- Lee S, Oh HJ (2019) Landslide susceptibility prediction using evidential belief function, weight of evidence and artificial neural network models. *Korean J Remote Sens* 35:299–316. <https://doi.org/10.7780/kjrs.2019.35.2.9>
- Li YY, Mei HB, Ren XJ, Hu XD, Li MD (2018) Geological disaster susceptibility evaluation based on certainty factor and support vector machine. *J Geo-inform Sci* 20:1699–1709. <https://doi.org/10.12082/dqxkx.2018.180349>
- 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>
- Li X, Xue GC, Liu CZ, Xia N, Yang YP, Yang F, Wang XL, Chang ZY (2022a) Evaluation of geohazard susceptibility based on information value model and information value-logistic regression model: a case study of the central mountainous area of Hainan Island. *J Geomech* 28:294–305. <https://doi.org/10.12090/j.issn.1006-6616.2021111>
- Li YJ, Tang YM, Deng YH, Song YX, Mu HD, Shan C, Cui SY (2022b) Hazard assessment of shallow loess landslides induced by rainfall: a case study of Liulin County of Hanxi Province. *CJGHC* 33:105–114. <https://doi.org/10.16031/j.cnki.issn.1003-8035.2022.02-13>
- Li YJ, Tang YM, Deng YH, Song YX, Mu HD, Shan C, Cui SY (2022c) Hazard assessment of shallow loess landslides induced by rainfall: a case study of Liulin County of Shanxi Province. *CJGHC* 33:105–114. <https://doi.org/10.16031/j.cnki.issn.1003-8035.2022.02-13>

- Lin QG, Wang Y, Liu TX, Zhu YQ, Sui Q (2017) The vulnerability of people to landslides: a case study on the relationship between the casualties and volume of landslides in China. *Int J Env Res Pub He* 14:212. <https://doi.org/10.3390/ijerph14020212>
- 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>
- 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>
- Pauluszek 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>
- Pauluszek-Filipiak K, Orenczak N, Pasternak M (2020) Investigating the effect of cross-modeling in landslide susceptibility mapping. *Appl Sci-Basel* 10:6335. <https://doi.org/10.3390/app10186335>
- Pourghasemi HR, Kerle N (2016) Random forests and evidential belief function-based landslide susceptibility assessment in Western Mazandaran Province, Iran. *Environ Earth Sci* 75:185. <https://doi.org/10.1007/s12665-015-4950-1>
- Pourghasemi HR, Gayen A, Park S, Lee CW, Lee S (2018) Assessment of landslide-prone areas and their zonation using logistic regression, LogitBoost, and NaiveBayes machine-learning algorithms. *Sustainability-Basel* 10:3697. <https://doi.org/10.3390/su10103697>
- Pourghasemi HR, Gayen A, Panahi M, Rezaie F, Blaschke T (2019) Multi-hazard probability assessment and mapping in Iran. *Sci Total Environ* 692:556–571. <https://doi.org/10.1016/j.scitotenv.2019.07.203>
- Rabby YW, Ishtiaque A, Rahman MS (2020) Evaluating the effects of digital elevation models in landslide susceptibility mapping in Rangamati District, Bangladesh. *Remote Sens-Basel* 12:2718. <https://doi.org/10.3390/rs12172718>
- Sahin EK, Colkesen I, Kavzoglu T (2020) 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>
- Saleem N, Huq ME, Twumasi NYD, Javed A, Sajjad A (2019) Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: a review. *Isprs Int J Geo-Inf* 8:545. <https://doi.org/10.3390/ijgi8120545>
- Sarda VK, Pandey DD (2019) Landslide susceptibility mapping using information value method. *Jordan J Civ Eng* 13:335–350
- Shou KJ, Lin JF (2020) Evaluation of the extreme rainfall predictions and their impact on landslide susceptibility in a sub-catchment scale. *Eng Geol* 265:105434. <https://doi.org/10.1016/j.enggeo.2019.105434>
- 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* 1–19. <https://doi.org/10.1080/17499518.2022.2088802>
- 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>
- Vakhshoori V, Pourghasemi HR, Zare M, Blaschke T (2019) Landslide susceptibility mapping using GIS-based data mining algorithms. *Water-Sui* 11:2292. <https://doi.org/10.3390/w11112292>
- 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* 17:4206. <https://doi.org/10.3390/ijerph17124206>
- Wang XD, Zhang CB, Wang C, Liu GW, Wang HX (2021) 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 CH, Lin QG, Wang LB, Jiang T, Su BD, Wang YJ, Mondal SK, Huang JL, Wang Y (2022a) The influences of the spatial extent selection for non-landslide samples on statistical-based landslide susceptibility modelling: a case study of Anhui Province in China. *Nat Hazards* 112:1967–1988. <https://doi.org/10.1007/s11069-022-05252-8>
- Wang SB, Zhuang JQ, Zheng J, Mu KQ, Wang Y, Fu YT (2022b) Landslide susceptibility evaluation based on deep learning along Kangding-Litang section of Sichuan-Tibet Railway. *J Eng Geol* 30:908–919. <https://doi.org/10.13544/j.cnki.jeg.2021-0115>
- Wang CM (2019) Study on characteristics of Lv Liang Loess landslide and rainfall early warning Model in Shanxi Province. Dissertation, Chengdu University of Technology, Chengdu. <https://doi.org/10.26986/d.cnki.gcdlc.2019.000804>
- Wu YL, Ke YT, Chen Z, Liang SY, Zhao HL, Hong HY (2020) Application of alternating decision tree with AdaBoost and bagging ensembles for landslide susceptibility mapping. *Catena* 187:104396. <https://doi.org/10.1016/j.catena.2019.104396>
- 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>
- 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 YH, Ge TT, Tian W, Liou YA (2019) Debris flow susceptibility mapping using machine-learning techniques in Shigatse Area, China. *Remote Sens-Basel* 11:2801. <https://doi.org/10.3390/rs11232801>
- 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 Y, Wang Y, Chen Y, Xu YJ, Zhang GM, Lin QG, Luo RH (2021) Projection of changes in flash flood occurrence under climate change at tourist attractions. *J Hydrol* 595:126039. <https://doi.org/10.1016/j.jhydrol.2021.126039>
- 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 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>
- 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, Peng SL, Shao CJ, Bai YH, Zhang XF, Liu XT, Liu WH (2021a) Zonation of landslide susceptibility in Ruijin, Jiangxi, China. *Int J Env Res Pub He* 18:5906. <https://doi.org/10.3390/ijerph18115906>
- Zhou XZ, Wen HJ, Zhang YL, Xu JH, Zhang WG (2021b) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. *Geosci Front* 12:101211. <https://doi.org/10.1016/j.gsf.2021.101211>

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.