https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions

Regional landslide susceptibility assessment based on Inter.iamb-Tabu algorithm

Chao Yin[1], Xixuan Zhang[1]*, Xuebing Ma[1], Xinliang Liu[1], Shufeng Li[1]
(1 School of Civil Engineering and Geomatics, Shandong University of Technology, Zibo 255049, China)
Correspondence to: Xixuan Zhang (zxxuan0209@163.com)

Abstract: Due to the great differences in geological environment characteristics and landslide disaster mechanism in different regions, the logical structure of each mathematical model is also different. It can only be determined through comparative research. Four improved algorithms based on Bayesian networkwere verified, and the error index was introduced to determine the algorithm with the best modeling effect. The landslide susceptibility probability of 774570 grids in Boshan District was calculated, and the landslide susceptibility distribution map of Boshan District was plotted. Based on the spatial superposition and grid calculator function of GIS, the landslide susceptibility assessment results of each model were compared.

Keywords: landslide susceptibility assessment, hazard factor, Bayesian network, Inter.iamb-Tabu algorithm, frequency ratio method.

1 Introduction

The formation of the landslide is the result of the disaster environment elements exceeding a certain threshold, and its mechanical mechanism is that the shear stress on the penetrating structural surface exceeds the shear strength of the surface (Yin and Zhang 2018; Mao et al. 2022). According to the information released by the China Geological Environment Monitoring Institute, a total of 34,218 geological disasters occurred in China during the “13th Five-Year Plan” period, of which landslides accounted for more than 50%. On average, it causes more than 200 deaths and billions of economic losses every year. Based on the selection and classification of hazard factors, the landslide susceptibility assessment calculates the probability of landslide occurrence and analyzes its spatial differentiation based on the mathematical model, which provides a theoretical basis for landslide prevention policy formulation and land use planning (Chen et al. 2018; Ouyang et al. 2019). The selection and classification of hazard factors and the construction of assessment model are the key steps of landslide susceptibility assessment (Shojaeezadeh et al. 2020).

The methods commonly used for selecting hazard factors include qualitative and quantitative methods. Among them, the qualitative method is based on fully revealing the occurrence law of landslides and the characteristics of disaster-pregnant environment, and combines expert opinions to select hazard factors (Zhang et al. 2020). The quantitative method is to screen the hazard factors based on the disaster-pregnant environment information and the feature extraction algorithm. Its essence is a supervised feature extraction problem (Zheng et al. 2019; Moghaddam et al. 2020; Lin et al. 2021; Zhou et al. 2021). The classification of hazard factors usually adopts empirical judgment method, equidistant division method and statistical analysis method. Among them, empirical judgment method and equidistant division method mainly rely on subjective experience and ignore objective data, and this can easily lead to low accuracy of assessment results (Zhang et al. 2016). Statistical analysis methods include frequency ratio method, information quantity method and entropy index method. Among them, the frequency ratio method is a widely used method to link the hazard factors with the assessment model. It analyzes the impact of different intervals of the hazard factors on the occurrence of landslides by introducing the interval area (ratio) and the landslide area (ratio), thereby reducing the uncertainty caused by its coupling with different models (Tsangaratos and Ilia 2016; Wu et al. 2020).

The commonly used mathematical models for landslide susceptibility assessment are divided into three categories (Mohammady et al. 2012). The first category is data statistical models, such as Logistic regression and binary/multivariate statistical methods (Sun et al. 2020). The second category is machine learning methods, such as

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

artificial neural network (ANN) and support vector machine (SVM). In recent years, deep learning has made great progress, which can avoid the disadvantages of over-fitting and unstable topology in shallow learning methods. Among them, random forest (RF), convolutional neural network (CNN) and Bayesian network are the most widely used. Thirdly, in order to overcome the shortcomings of a certain mathematical model, some scholars have proposed a hybrid algorithm that integrates multiple models. Due to the high generalization ability and robustness of the hybrid algorithm, the landslide susceptibility assessment results are significantly better than the single algorithm (Song et al. 2012). There are great differences in the geological environment characteristics and landslide disaster mechanism in different regions, and the logical structure of each mathematical model is also different. Therefore, the best assessment model that fits the characteristics of the disaster-pregnant environment in a certain area is often unable to be known in advance, and can only be determined through a large number of comparative studies.

The essence of Bayesian network is to use a joint probability distribution to describe the probability dependence between variables (Costache and Bui 2019). The maximum and minimum hill-climbing algorithm (MMHC) is a widely used Bayesian algorithm. MMPC-Tabu, Fast.iamb-Tabu and Inter.iamb-Tabu are hybrid algorithms obtained by improving MMHC through Tabu Search algorithm. Based on the above three algorithms and taking MMHC as a control, this paper carried out a study on landslide susceptibility assessment in Boshan District, China. The main contents include: ① Screen the landslide hazard factors using the single factor Logistic regression method, and classify them by the frequency ratio method. ② Verify The modeling effects of MMHC, MMPC-Tabu, Fast.iamb-Tabu and Inter.iamb-Tabu models, and the best model is determined by its error index. ③ Carry out the landslide susceptibility assessment of Boshan District, and compare the assessment results of each model.

2 Research area and research data

2.1 Research area overview

Boshan District is located in the northern part of the central mountainous area of Shandong Province. It spans 117°43'-118°42' east longitude, 36°16'-36°31' north latitude, and the altitude is between 102m and 1066 m. The geological structure is complex, and there are two main types of geological structure. One is the Archean basement structure, which is dominated by linear tight folds. The axial direction of the fold is consistent with the direction of schistosity, showing 330°~340° distribution. The basement faults are also more developed, but it is difficult to identify due to the regional geological effects of lithology. The second is the cover structure of the Mesozoic and Cenozoic, which is dominated by faults, followed by folds. There is Yaojiayu fault in the north-south direction, with a total length of 60km, a strike of N5°W~N10°E, a tendency of NW, and a dip angle of 55°~75°. The NE-trending fault is the Zihe fault, with a total length of 110km, a width of 400m~1000m, and a surface outcrop of about 60km. It runs through the towns of Boshan, Yuanquan and Chishang, and is 19km long (Gao et al. 2012; Jiang et al. 2014).

The strata in the area are well developed, from southeast to northwest, from old to new. There are four boundaries and seven systems. The Archean Taishan Group (Art) is distributed in Lingxi, Letuan, Boshan Town, Chishang, Lijia and other places, with a thickness of 2700m~15000m. The Cambrian system (∈) is distributed in Shimen, Boshan Town, Lijia, Chishang, Yuanquan, Lingxi and other places, with a thickness of about 600m. The Ordovician (O) is distributed in Shimen, Xiajiazhuang, Shima, Letuan, Badou, Yuanquan, Yuezhuang, Boshan Town and other places, with a thickness of about 800m. The Carboniferous (C) is distributed in Badou, Fushan, Shantou, Yucheng, Baita, Xiajiazhuang, etc., with a thickness of about 140m. The Permian (P) is distributed in Badou, Shantou, Fushan, Xiajiazhuang, Baita, Jiaozhuang, Yucheng, etc., with a thickness of about 430m. The Mesozoic (J) is composed of sandstone and shale, belonging to river, lake and swamp deposits, containing plant fossils, which is about 180m thick. Quaternary is widely distributed in valleys, rivers and low-lying areas. It is dominated by clayey sandy soils, interspersed with gravelly layers, with a thickness of 3m~4m (Sun et al. 2020).

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
EGU
Discussions

2.2 Research data

The data used in this paper include:

1) GDEMV2 30m resolution digital elevation model, from geospatial data cloud (http://www.gscloud.cn/).

2) 10m resolution land use data of Tsinghua University, from Tsinghua University Geospatial Database (http://data.ess.tsinghua.edu.cn/).

3) Fault data of Shandong Province, from Geological Expertise Service System(http://geol.ckcest.cn/).

4) Boshan District road data, from Openstreetmap network(https://www.openstreetmap.org/#map=5/34.574/113.247).

5) Precipitation data of Boshan District provided by Boshan District Weather Bureau.

6) The prevention and control scheme of geological disasters in Boshan District compiled by Zibo Natural Resources and Planning Bureau.

According to the geological disaster prevention and control plan of Boshan District, the occurrence time and location of 99 historical landslides were determined. The author has carried out on-site reconnaissance of 99 landslides, combined with remote sensing interpretation, and clarified the scale of each landslide. The results show that the total volume of landslides in Boshan District is 2732400m^3, with a total area of 1.027km^2. The largest landslide is located in Boshan Town, with a volume of 74160m^3 and an area of 0.021km^2. Chishang Town has the largest number of landslides (19). The distribution of landslides in Boshan District is shown in Figure 1.

Figure 1 Landslide distribution map of Boshan District

3 Analysis of landslide hazard factors

3.1 Landslide hazard factors

By summarizing the existing literature and taking into account the availability of data, 13 landslide hazard factors (Chen et al. 2018; Khalaj et al. 2020) were selected: slope gradient, elevation, slope direction, land use type, lithology, distance from fault, distance from river, distance from road, cumulative precipitation, plane curvature, profile curvature, normalized difference vegetation index (NDVI), topographic wetness index (TWI) and river dynamic index (SPI).

Among them, the curvature is the degree of bending deformation of a point on the ground surface. When the curvature>0, the slope is a convex slope. When the curvature<0, the slope is a concave slope. Plane curvature is the bending degree of a point on the ground surface on its contour line, that is, the component of curvature in the horizontal direction. The profile curvature is the elevation change rate of the slope from the maximum descent direction, that is, the component of the curvature in the vertical direction (Yilmaz 2009).

NDVI ranged from -1.0 to 1.0. When NDVI<0, the reflection of visible light is high, and the surface is covered with clouds, water, snow, etc. When NDVI=0, the surface is mostly rock or bare soil. When NDVI>0, there is vegetation coverage on the surface, and the vegetation coverage increases with the increase of NDVI (Aditian et al. 2018).

TWI is an index that reflects the influence of regional topography on runoff flow and accumulation, which quantifies the control of topography on basic hydrological processes. Obtaining TWI based on GIS includes five steps: filling the depression, calculating the slope after filling the depression, calculating the flow direction, calculating the flow and calculating the flow per unit area (Bollmann et al. 2019).

SPI is an indicator reflecting the erosion capacity of surface water flow, which can be used to determine the strong flow path formed by water flow convergence and the location where gully erosion may occur. The erosion capacity of surface water flow increases with the increase of SPI (Wang et al. 2019).

3.2 Screening of hazard factors

In Bayesian networks, the number of network nodes has a significant impact on the complexity of structural learning, that is, when the training samples are fixed, too many network nodes will lead to a decrease in the accuracy of the model, and this is not conducive to reflecting the relationship between the main nodes and the outcome (Michalowski and Park 2020). On the other hand, landslide hazard factors are not completely independent. For example, distance from river, TWI and SPI are indicators of surface runoff intensity, and the three contain repetitive information. In order to eliminate the multi-collinearity of landslide hazard factors and improve the modeling efficiency (Nseka et al. 2019), this paper selects landslide hazard factors based on single factor Logistic regression method. The calculation results are shown in Table 1.

Table 1 Calculation results of frequency ratio method

According to Ebid et al. (2014), Kim et al. (2021), Conforti and Ietto (2021) and Table 1, nine hazard factors with *P* values between 0.1 and 0.2 were selected for landslide susceptibility modeling in Boshan District, including slope gradient, elevation, slope aspect, land use type, distance from fault, distance from road, profile curvature, NDVI and SPI. The slope gradient, elevation, slope aspect, profile curvature and SPI information were extracted by using 30m resolution DEM in Boshan District, NDVI information was extracted by using Landsat TM image of Landsat 8 OLI_TIRS satellite, land use type information was extracted by using 10m resolution land use data of Tsinghua University, information of distance from fault was extracted by using fault data of Shandong Province, and information of distance from road was extracted by using road data of Boshan District. Based on the resampling function of ArcGIS10.2, the research unit was processed into 30m×30m grid. The distribution map of the above nine hazard factors was drawn and as shown in Figure 2.

Figure 2 Distribution of hazard factors

3.3 Classification of hazard factors

Based on the historical landslide data of Boshan District, the frequency ratio method is used to classify the landslide hazard factors. The calculation method is shown in Formula 1.

$$\mathrm{FR} = \frac{N_i/N}{M_i/M} \tag{1}$$

In the formula, FR is the frequency ratio, N_i is the landslide area in the *i*-th interval of a hazard factor, N is the total area of the landslide, M_i is the area in the *i*-th interval of a hazard factor, and M is the total area of Boshan District. FR<1 means that this interval has a limiting effect on the landslide. FR>1 means that the interval has a promoting effect on the landslide. FR=1 means that the interval has no obvious effect on the landslide. The FR of each interval of the nine hazard factors is calculated. Limited by the space, only the calculation results of the elevation are listed in this article, as shown in Table 2.

Table 2 Calculation results of elevation frequency ratio

When the elevation is between 0~300m and 450m~600m, FR>1. When the elevation is between 30m~450m and 600m~1200m, FR<1. Thus, the elevation is divided into 4 levels: 0~300m, 300m~450m, 450m~600m and 600m~1200m. The classification results are shown in Table 3.

Table 3 Classification results of hazard factors

4 Assessment of landslide susceptibility in Boshan District

4.1 Model Introduction

4.1.1 Bayesian network

Bayesian network was proposed by Pearl Judea in 1987. Its essence is to use a joint probability distribution to describe the probability dependence between variables. Given a series of random variables X={X_1,..., X_n}, the joint probability P (X_1,..., X_n) can be expressed as a Bayesian network B=(G, θ), where G is the directed acyclic graph of the Bayesian network. The nodes in the graph are a series of random variables, and the directed edges represent the probability dependence between random variables. If there is an edge from X_i to X_j, then Xi is the parent node of X_j and

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

X_j is the child node of X_i (Zhong et al. 2022). θ is a conditional probability distribution table, which quantitatively describes the relationship between random variables and their parent nodes. The construction of Bayesian network is divided into three steps. First, finding the relevant variables and their possible values. Second, obtaining the optimal network structure based on machine learning algorithm, that is, the structure learning process. Third, calculating the conditional probability table of each node, that is, the parameter learning process.

4.1.2 MMHC algorithm

The MMHC algorithm is a widely used Bayesian algorithm (Song et al. 2022). The first stage uses the maximum and minimum heuristic search strategy, and uses the MaxMinHeuristic function to obtain the candidate parents and children (CPC) of each variable. The MaxMinHeuristic function calculates the minimum correlation value between the target node *T* and all other nodes, and then selects the nodes corresponding to the maximum value of these correlation values to enter the CPC. Under the condition that all subsets of CPC are given, if the remaining nodes are independent of the target node *T*, the first stage ends. The second stage is to use Ind(X; $T|Z$) function removes the nodes that should not enter the CPC in the first stage. If the node *X* and the target node *T* have an independent relationship when *Z* is known, *X* is removed from the CPC. Ind(X; $T|Z$) function is used to determine the conditional independence between nodes. If *X* and *T* are conditionally independent of each other when *Z* is given, the return value of Ind(X; $T|Z$) is true.

4.1.3 Improvement of MMHC algorithm

MMPC-Tabu, Fast.iamb-Tabu and Inter.iamb-Tabu are hybrid algorithms obtained by improving MMHC by Tabu Search algorithm. The Iamb (Incremental association Markov blanket) algorithm consists of two phases: forward (Growing) and backward (Shirinking) (Yang et al. 2019). In a Bayesian network, the Markov blanket node MB(*T*) is the parent node of the target node *T*, the child node and other parent nodes of the child node. All the estimated values of MB(*T*) constitute CMB(*T*), but it is easy to produce false positive relationship in the forward process, that is, some elements of CMB(*T*) are not the estimated values of MB(*T*).

4.2 Comparison of model performance

Two standard networks (Chest clinic Network and TANK Network) were selected from the Bayesian network resource library to verify the modeling effects of MMHC, MMPC-Tabu, Fast.iamb-Tabu, and Inter.iamb-Tabu, and the algorithm with the best performance was selected to carry out the landslide susceptibility assessment in Boshan District. Among them, Chest clinic Network, also known as Asia Network, contains 8 nodes and 8 directed edges. TANK Network contains 14 nodes and 20 directed edges (Zhou et al. 2022). The directed acyclic graphs of Chest clinic Network and TANK Network are shown in Figure 3 and Figure 4.

Figure 3 Directed acyclic graphs of Chest clinic Network

Figure 4 Directed acyclic graph of TANK Network

Chest clinic Network and TANK Network were used to generate data sets with a sample size of 1200, 1500, 1800 and 2100, respectively. The data sets were used as training samples. Bayesian networks were generated based on MMHC, MMPC-Tabu, Fast.iamb-Tabu and Inter.iamb-Tabu and compared with the directed acyclic graphs of Chest clinic Network and TANK Network. The number of missing edges, error edges and reverse edges of the newly generated network is recorded and the error index is calculated. The error index calculation method is shown in Formula 2 (Mukhammadzoda et al. 2021; Li et al. 2022), and the comparison result is shown in Table 4.

$$E_n = n_a + n_b + 0.5n_c \quad (2)$$

In the formula: E_n is the error index, n_a is the number of missing edges, n_b is the number of error edges, n_c is the number of reverse edges.

Table 4 Comparison results of different sample sizes and algorithms

Seen from Table 4, when the number of samples is 1800 and 2100 respectively, the error indexes of MMHC,

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

MMPC-Tabu, Fast.iamb-Tabu and Inter.iamb-Tabu for Chest clinic Network and TANK Network are equal, indicating that the number of samples reaches 1800 to meet the training requirements of the model, and continuing to increase the number of samples has no significant effect on improving the accuracy of modeling. When the number of samples is 1800, the error index of Inter.iamb-Tabu algorithm for Chest clinic Network and TANK Network is 0, while the error index of other algorithms for Chest clinic Network and TANK Network is not 0, indicating that when the number of samples is 1800, the Inter.iamb-Tabu algorithm has been trained and its performance is significantly better than other algorithms. This paper selects Inter.iamb-Tabu to carry out landslide sensitivity evaluation in Boshan District (Pan 2019).

4.3 Landslide susceptibility assessment based on Inter.iamb-Tabu in Boshan area

4.3.1 Model establishment

The 30m×30m resolution grid unit is used as the basic unit of landslide susceptibility assessment (Ambrosi et al. 2018; Asdar et al. 2021). According to the grid calculator statistics of ArcGIS10.2, there are 1849 grids in 99 landslides in Boshan District. In addition, 1849 grids are randomly selected in the non-landslide area to establish the landslide grid data set and the non-landslide grid data set. The training sample set is constructed based on 900 landslide grids and 900 non-landslide grids. The landslide susceptibility model of Boshan District is constructed by Inter.iamb-Tabu, and the conditional probability of each node is calculated according to the maximum likelihood assessment method. The computer hardware environment for modeling is CPU i7-6700 processor, 8G memory, GTX1050 Ti-8G graphics card, and the software environment is the Bayesian network learning package in R3.5.0 software. The constructed Bayesian network and the conditional probability of each hazard factor are shown in Figure 5.

Figure 5 Landslide susceptibility model of Boshan District

It can be seen in Figure 5 that the landslide susceptibility model of Boshan District includes 10 nodes (9 hazard factor nodes and 1 ending node) and 14 directed edges. Slope gradient, elevation, land use type and distance from fault are the parent nodes of landslides, which have a direct inducing effect on landslides. The profile curvature and NDVI are the mutual feedback nodes of the landslide, that is, the profile curvature and NDVI have a direct inducing effect on the landslide, and the occurrence of the landslide may change the profile curvature and NDVI of the disaster-pregnant environment. Slope aspect, distance from road and SPI have a mutual feedback relationship with other hazard factors, and play an indirect role in landslide occurrence through this mutual feedback relationship (Li et al. 2021; Yin et al. 2020).

4.3.2 Model verification

Based on the remaining 949 landslide grids and 949 non-landslide grids, the validation sample set was constructed. Method of the area under the receiver operating characteristic (ROC) curve (AUC) was used to verify the landslide susceptibility model in Boshan District. The ROC curve is a curve drawn with the true positive rate (susceptibility) as the ordinate and the false positive rate (1-specificity) as the abscissa. Among them, the true positive rate is the probability that the model judgment and the actual situation are both landslides. The false positive rate is the probability of the model is judged as a landslide but actually non-landslide (Zêzere et al. 2017). AUC is used to represent the area value under the ROC curve. When AUC is between 0.5 and 0.7, the accuracy of the assessment results is lower. When AUC is between 0.7 and 0.9, the assessment results have higher accuracy. When AUC is between 0.9 and 1.0, the accuracy of the results is very high (Berhane et al. 2020). Figure 6 shows the verification results of the landslide susceptibility model in Boshan District. Since the AUC value reaches 0.907, the model accuracy is high.

Figure 6 Verification results of landslide susceptibility model in Boshan District

4.4 Landslide susceptibility zoning in Boshan District

Using the grid calculator of ArcGIS10.2, the landslide susceptibility model of Boshan District is calculated

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

according to Figure 5. 32.78 hours later, the probability of landslide susceptibility of 774570 grids is obtained, and the probability distribution map of landslide susceptibility in Boshan District is drawn, as shown in Figure 7.

Figure 7 Probability distribution map of landslide susceptibility in Boshan area

As seen in figure 7, the highest probability of landslide susceptibility in Boshan District is 0.809 and the lowest is 0.287. The areas with high probability are the western part of the town, Chishang Town, the southern part of Boshan Town and the mountainous area of Shima Town. The areas with low probability are the plains of the northern part of Yuanquan Town, Baita Town and Boshan City, Badou Town and Gushan Town. The landslide susceptibility zoning of Boshan District is carried out by using the natural discontinuity point method, and it is divided into five sensitivity intervals: very high sensitivity area, high sensitivity area, medium sensitivity area, low sensitivity area and very low sensitivity area. The discontinuity points are 0.730,0.615,0.505 and 0.395 respectively (Du et al. 2020). The landslide susceptibility zoning of Boshan District is shown in Figure 8.

Figure 8 Landslide susceptibility zoning map of Boshan District

It can be seen from Figure 8 that the extremely high sensitive area, high sensitive area, medium sensitive area, low sensitive area and extremely low sensitive area account for 7.3% (49.8006km^2), 16.5% (112.563km^2), 26.1% (178.0542km^2), 33.2% (226.4904km^2) and 16.9% (115.2918km^2) of the total area of Boshan District, respectively. There are 67, 22, 7, 2 and 1 landslides located in the extremely high sensitive area, high sensitive area, medium sensitive area, low sensitive area and extremely low sensitive area, respectively. They accounted for 67.21%, 22.95%, 4.92%, 3.28% and 1.64% of the total number of landslides in Boshan District respectively, and they accounted for 85.32% (1391569m^2), 10.58% (172560m^2), 2.17% (35393m^2), 1.09% (17778m^2) and 0.84% (13700m^2) of the total area of landslides.

4.5 Comparison of assessment results of different models

Taking the Inter.iamb-Tabu model as the benchmark model, based on the spatial superposition and raster calculator function of ArcGIS10.2, the landslide susceptibility assessment results of the benchmark model are compared with other models, and the underestimation area, equal area and overestimation area of the MMHC, MMPC-Tabu and Fast.iamb-Tabu models are obtained respectively, as shown in Figure 9.

a) Comparison between MMHC model and benchmark model

b) Comparison between MMPC-Tabu model and baseline model

c) Comparison between Fast.iamb-Tabu model and benchmark model

Figure 9 Comparison of landslide susceptibility assessment results with different models

It can be seen from Figure 9 that the area of the landslide susceptibility assessment results of the MMHC, MMPC-Tabu, and Fast.iamb-Tabu models equal to the benchmark model. It is much larger than the area of the under-estimated area or the over-estimated area (Zhang et al. 2023). The MMHC model has the most erroneous assessment areas, most of which are over-estimated, showing a scattered distribution. The error estimation area of Fast.iamb-Tabu model is the least, which is basically within the error estimation area of other models. The underestimation area of Fast.iamb-Tabu model is superimposed with the distance from road, and the overestimation area is superimposed with the distance from fault, as shown in Figure 10 and Figure 11.

Figure 10 The superposition of under-estimated area and distance from road

Figure 11 The superposition of over-estimated area and distance from fault

From Figure 10 and Figure 11, it can be seen that the underestimation area is distributed in the area near the road, the distribution of the overestimation area is scattered, and the coincidence rate with the area near the fault is higher. The results show that compared with the Inter.iamb-Tabu model, other models weaken the effect of distance from road and overestimate the effect of distance from fault. In summary, the MMHC model, MMPC-Tabu model and Fast.iamb-Tabu model are easy to discard the feature information of some factors in the sample to achieve the optimal overall accuracy of the model.

5 Conclusion and discussion

1) The landslide hazard factors in Boshan area include slope gradient, elevation, slope direction, land use type, distance from fault, distance from road, profile curvature, NDVI and SPI. The landslide susceptibility model based on Inter.iamb-Tabu in Boshan area is the best. The model contains 10 nodes (9 hazard factor nodes and 1 ending node) and 14 directed edges. Boshan area is divided into extremely high sensitive area, high sensitive area, medium sensitive area, low sensitive area and extremely low sensitive area, accounting for 7.3% (49.8006km^2), 16.5% (112.563km^2), 26.1% (178.0542km^2), 33.2% (226.4904km^2) and 16.9% (115.2918km^2) of the total area of Boshan respectively. There are 67, 22, 7, 2 and 1 landslides located in extremely high sensitive area, high sensitive area, medium sensitive area, low sensitive area and extremely low sensitive area respectively.

2) MMHC is a widely used Bayesian algorithm. Inter.iamb-Tabu is a hybrid algorithm obtained by improving MMHC by Tabu Search algorithm. In this paper, based on the frequency ratio method and the Inter.iamb-Tabu algorithm, the landslide susceptibility assessment in Boshan area is carried out, and the accuracy of results is high. When the MMHC model, MMPC-Tabu model and Fast.iamb-Tabu model are trained, it is easy to abandon the feature information of some factors in the sample to achieve the optimal overall accuracy of the model. This paper does not carry out landslide susceptibility assessment based on deep learning algorithms such as decision tree, RF, CNN and deep belief network and their hybrid algorithms. The accuracy of related algorithms needs to be verified in future research.

Funding:

This research was supported by the National Natural Science Foundation of China (Grant No. 51808327) and Natural Science Foundation of Shandong Province (Grant No. ZR2019PEE016).

Conflict of interest:

The authors have not disclosed any competing interests.

Ethical approval:

I certify that this manuscript is original and has not been published and will not be submitted elsewhere for publication. And the study is not split up into several parts to increase the quantity of submissions and submitted to various journals or to one journal over time. No data have been fabricated or manipulated (including images) to support our conclusions. No data, text or theories by others are presented as if they were our own.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

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.

Ambrosi C, Strozzi T, Scapozza C, Wegmüller U (2018) Landslide hazard assessment in the Himalayas (Nepal and Bhutan) based on Earth-Observation data. Engineering Geology 237:217-228.

An N, Teng Y, Yang JY, Li L (2018) Bayesian network structure learning method based on causal effect. Application Research of Computers 35(12):3609-3613.

Asdar, Arsyad U, Soma AS, Mappangaja B, Wahyuni, Amaliah R, Nurdin PF (2021) Analysis of the landslides vulnerability level using frequency ratio method in Tangka Watershed. Earth and Environmental Science 870:012013.

Berhane G, Kebede M, Alfarah N, Hagos E, Grum B, Giday A, Abera T (2020) Landslide susceptibility zonation mapping using GIS-based frequency ratio model with multi-class spatial data-sets in the Adwa-Adigrat mountain chains, northern Ethiopia. Journal of African Earth Sciences 164:103795.

Bollmann S, Kristensen MH, Larsen MS, Olsen MV Pedersen MJ, Østergaard LR, O'Brien K, Langkammer C, Fazlollahi A, Barth M (2019) SHARQnet-Sophisticated harmonic artifact reduction in quantitative susceptibility mapping using a deep convolutional neural network. Zeitschrift für Medizinische Physik 29(2):139-149.

Chen W, Peng JB, Hong HY, Shahabi H, Pradhan B, Liu JZ, Zhu AX, Pei XJ, Duan Z (2018) Landslide susceptibility modeling using GIS-based machine learning techniques for Chongren County, Jiangxi Province, China. Science of the Total Environment 626:1121-1135.

Chen W, Xie XS, Peng JB, Shahabi H, Hong H, Bui DT, Duan Z, Li S, Zhu AX (2018) GIS-based landslide susceptibility evaluation using a novel hybrid integration approach of bivariate statistical based random forest method. CATENA 164:135-149.

Conforti M, Ietto F (2021) Modeling Shallow Landslide Susceptibility and Assessment of the Relative Importance of Predisposing Factors, through a GIS-Based Statistical Analysis. Geosciences 11(8):333.

Costache R, Bui DT (2019) Spatial prediction of flood potential using new ensembles of bivariate statistics and artificial intelligence: A case study at the Putna river catchment of Romania. Science of the Total Environment 691:1098-1118.

Du J, Glade T, Woldai T, Chai B, Zeng B (2020) Landslide susceptibility assessment based on an incomplete landslide inventory in the Jilong Valley, Tibet, Chinese Himalayas. Engineering Geology 270:105572.

Ebid AHI, Ali ZT, Ghobary MAF (2014) Blood pressure control in hypertensive patients: impact of an Egyptian pharmaceutical care model. Journal of Applied Pharmaceutical Science 4(9):93-101.

Gao XM, Qin ZL, Wang LJ, Chen LX, Ma SQ, Yang KD (2012) The Climatic Characteristics of Geological Calamity in the Mountainous Area of the Middle Part of Shandong Province. Science and Technology Review 30(4):55-60.

He DL, Cheng Y, Zhao RL (2022) Study on learning Bayesian network structure based on MMHC algorithm. Journal of Beijing Technology and Business University (Natural Science Edition) 26(3):43-48. (in Chinese)

Jiang C, Qu BJ, Dong Q, Lin MS (2014) Construction of Geological Disasters Warning System Based on GIS in Zibo City. Shandong Land and Resources 30(3):89-91. (in Chinese)

Khalaj S, Toroody FB, Abaei MM, Bahootoroody A, Abbassi R (2020) A methodology for uncertainty analysis of landslides triggered by an earthquake. Computers and Geotechnics 117:103262.

Kim H, Lee JH, Park HJ, Heo JH (2021) Assessment of temporal probability for rainfall-induced landslides based on nonstationary extreme value analysis. Engineering Geology 294:106372.

Li M, Zhang R, Liu KF (2021) A New Marine Disaster Assessment Model Combining Bayesian Network with Information Diffusion. Journal of Marine Science and Engineering 9(6):640.

Lin QG, Lima P, Steger S, Glade T, Jiang T, Zhang J, Liu T, Wang Y (2021) National-scale data-driven rainfall induced landslide susceptibility mapping for China by accounting for incomplete landslide data. Geoscience Frontiers 12:101248.

Li TT, Zhou YZ, Zhao Y, et al. Zhang C, Zhang X (2022) A hierarchical object oriented Bayesian network-based fault diagnosis method for building energy systems. Applied Energy 306(B):118088.

Mao YM, Mwakapesa DS, Li YC, Xu KB, Nanehkaran YA, Zhang MS (2022) Assessment of landslide susceptibility using DBSCAN-AHD and LD-EV methods. Journal of Mountain Science 19(1):184-197.

Michalowski RL, Park D (2020) Stability assessment of slopes in rock governed by the Hoek-Brown strength criterion. International Journal of Rock Mechanics and Mining Sciences 127:104217.

Moghaddam SHA, Mokhtarzade M, Beirami BA (2020) A feature extraction method based on spectral segmentation and integration of hyperspectral images. International Journal of Applied Earth Observation and Geoinformation 89:102097.

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Mohammady M, Pourghasemi HR, Pradhan B (2012) Landslide susceptibility mapping at Golestan Province, Iran: A comparison between frequency ratio, Dempster-Shafer, and weights-of-evidence models. Journal of Asian Earth Sciences 61:221-236

Mukhammadzoda S, Shohnavaz F, Ilhomjon O, Zhang GC (2021) Application of Frequency Ratio Method for Landslide Susceptibility Mapping in the Surkhob Valley, Tajikistan. Journal of Geoscience and Environment Protection 9:168-189.

Nseka D, Kakembo V, Bamutaze Y, Mugagga F (2019) Analysis of topographic parameters underpinning landslide occurrence in Kigezi highlands of southwestern Uganda. Natural Hazards 99(2):973-989.

Ouyang CJ, Wang ZW, An HC, Liu XR, Wang DP (2019) An example of a hazard and risk assessment for debris flows-A case study of Niwan Gully, Wudu, China. Engineering Geology 263:105351.

Shojaeezadeh SA, Nikoo MR, Mirchi A, Mallakpour I, AghaKouchak A, Sadegh M (2020) Probabilistic hazard assessment of contaminated sediment in rivers. Science of the Total Environment 703:134875.

Song YQ, Gong JH, Gao S, Sheng G, Wang D, Cui T, Yi L, Wei B (2012) Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. Computers & Geosciences 42:189-199.

Sun DL, Wen HJ, Wang DZ, Xu JH (2020) A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. Geomorphology 362:107201.

Sun Q, Shi QM (2020) Study on the risk zoning of urban earthquake disaster based on GIS: taking Zibo city as an example. Earthquake Research in Sichuan 175:19-24. (in Chinese)

Tsangaratos P, Ilia I (2016) Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size. CATENA 145:164-179.

Wang Y, Fang ZC, Hong HY (2019) Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the Total Environment 666:975-993.

Wu ZN, Shen YX, Wang HL, Wu M (2020) Urban flood disaster risk evaluation based on ontology and Bayesian Network. Journal of Hydrology 583:124596.

Yang XL, Wang YJ, Ou Y, Tong Y (2019) Three-Fast-Inter Incremental Association Markov Blanket learning algorithm. Pattern Recognition Letters 122:73-78.

Yin C, Li HR, Che F, Li Y, Liu D (2020) Susceptibility mapping and zoning of highway landslide disasters in China. PLoS ONE 15(9): e0235780.

Yin C, Zhang JL (2018) Hazard regionalization of debris-flow disasters along highways in China. Natural Hazards 91(2):1-19.

Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). Computers & Geosciences 35(6):1125-1138.

Zhang H, Yin C, Wang SP, Guo B (2023) Landslide susceptibility mapping based on landslide classification and improved convolutional neural networks. Natural Hazards 116:1931-1971.

Zhang J, Yin KL, Wang JJ, Liu L, Huang FM (2016) Evaluation of landslide susceptibility for Wanzhou district of Three Gorges Reservoir. Chinese Journal of Rock Mechanics and Engineering 35(2):284-296. (in Chinese)

Zhang TY, Mao ZA, Wang T (2020) GIS-based evaluation of landslide susceptibility using a novel hybrid computational intelligence model on different mapping units. Journal of Mountain Science 17(12):2929-2941.

Zheng H, Cheng G, Li Y, Chang L (2019) A new fault diagnosis method for planetary gear based on image feature extraction and bag-of-words model. Measurement 145:1-13.

Zêzere JL, Pereira S, Melo R, Melo R, Oliveira SC, Garcia RAC (2017) Mapping landslide susceptibility using data-driven methods. Science

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

of the Total Environment 589:250-267.

Zhong KH, Chen YW, Qin XL (2022) Sub-BN-Merge based Bayesian network structure learning algorithm. Computer Science 49(S2):64-70.

Zhou XZ, Wen HJ, Zhang YL, Wu J, Zhang W (2021) Landslide susceptibility mapping using hybrid random forest with GeoDetector and RFE for factor optimization. Geoscience Frontiers 12:101211.

Zhou YS, Li X, FaiYuen K (2022) Holistic risk assessment of container shipping service based on Bayesian Network Modelling. Reliability Engineering & System Safety 220:108305.

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Discussions

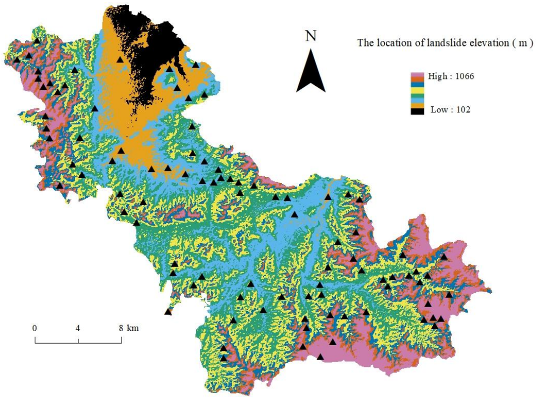


Figure 1 Landslide distribution map of Boshan District

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Discussions

Table 1 Calculation results of frequency ratio method

Hazard factors	Classification	Interbasin area/m^2	Landslide area/m^2	Area ratio	χ^2	P
Slope gradient/°	[0, 10)	94078800	39122	0.000415843	4.081	0.146
	[10, 25)	151664400	513003	0.003382488		
	[25, 90]	89056800	78875	0.000885671		
Elevation/m	[0, 300)	20422800	49218	0.002409954	0.536	0.182
	[300, 450)	113832000	150809	0.001324838		
	[450, 600)	182466000	413936	0.002268565		
	[600, 1200]	18079200	17037	0.000942354		
Slope aspect/°	[0, 90)	85039200	195610	0.002300233	0.486	0.125
	[90, 210)	108810000	164060	0.001507766		
	[210, 270)	75999600	198134	0.002607040		
	[270, 360]	64951200	73196	0.001126938		
Land use type	Construction and road sites	66960000	203813	0.003043802	1.457	0.156
	Cultivated land and grassland	24440400	45432	0.001858889		
	Undeveloped land and other land	243399600	381755	0.001568429		
Lithology	Limestone layer mudstone layer	79012800	34705	0.000439233	10.547	0.278
	Hazle interbedding	34484400	68148	0.001976198		
	Sandy soil and silty soil	221302800	528147	0.002386536		
Distance from fault/m	[0m, 2000m)	188492400	468833	0.002487278	11.481	0.128
	[2000m, 4000m)	62272800	123045	0.001975903		
	[4000m, 8000m]	84034800	39122	0.000465545		
Distance from river/m	[0, 1200)	151999200	330644	0.002175301	1.745	0.396
	[1200, 2400)	124880400	234732	0.001879654		
	[2400, 3200]	57920400	65624	0.001133003		
Distance from road/m	[0, 600)	126889200	281426	0.002217888	0.974	0.173
	[600, 1200)	75999600	141975	0.001868102		
	[1200, +∞)	131911200	207599	0.001573778		
Accumulation precipitation/mm	[600, 700)	135928800	235363	0.001731517	0.347	0.047
	[700, 800]	198871200	395637	0.001989413		
Plane curvature	(-∞, -0.5)	135259200	178573	0.001320228	1.267	0.264
	[-0.5, 0.5)	106131600	198765	0.001872816		
	[0.5, +∞)	93409200	253662	0.002715600		
Profile curvature	(-∞, -0.5)	86378400	264389	0.003060823	2.784	0.145
	[-0.5, 0.5)	112158000	209492	0.001867829		
	[0.5, +∞)	136263600	157119	0.001153052		
NDVI	[-1, -0.1)	158695200	329382	0.002075564	1.465	0.171
	[-0.1, 0.05)	129567600	240411	0.001855487		
	[0.05, 1]	46537200	61207	0.001315227		
TWI	[0, 4)	80686800	111056	0.001376384	1.063	0.036
	[4, 8)	107470800	203813	0.001896450		
	[8, 12]	146642400	316131	0.002155795		
SPI	[0, 30)	120193200	176680	0.001469967	1.928	0.162
	[30, 70)	100105200	188669	0.001884707		
	[70, +∞)	114501600	265651	0.002320064		

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions

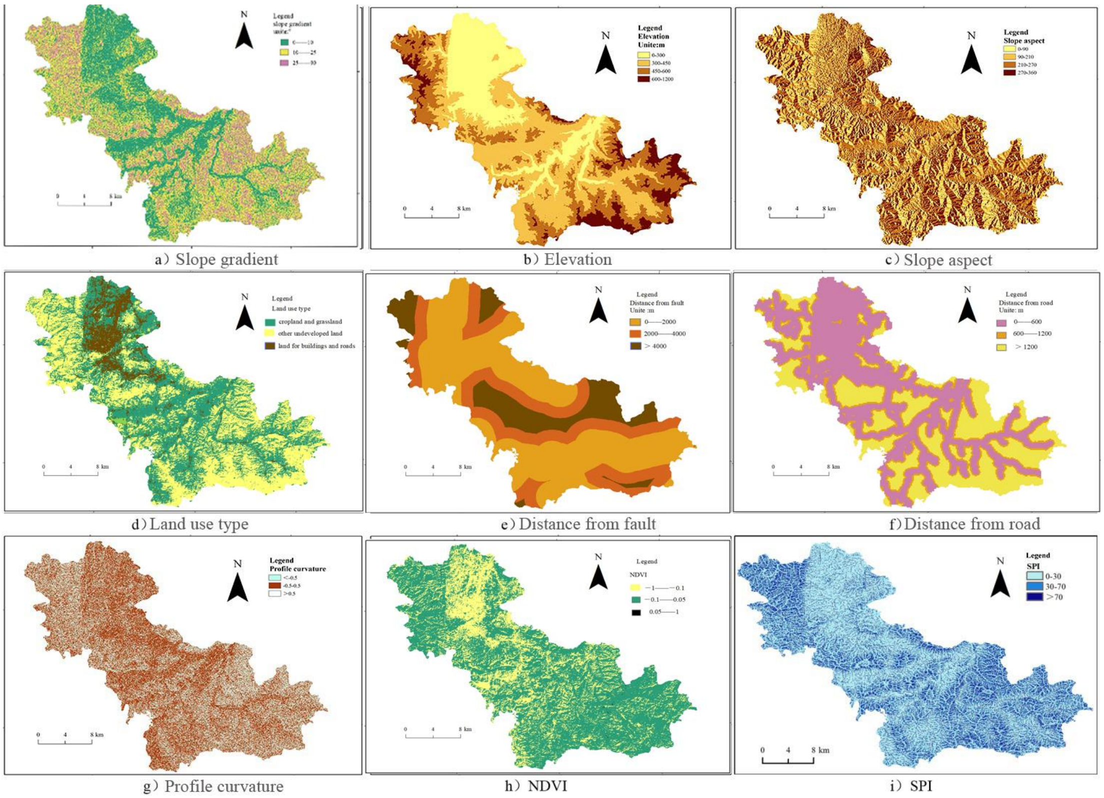


Figure 2 Distribution of hazard factors

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Table 2 Calculation results of elevation frequency ratio

Hazard factor	Interbasin area (ratio)	Landslide area (ratio)	FR
Elevation/m	[0, 150): 4.6873km^2 (0.014);	[0, 150): 10727m^2 (0.017);	[0, 150): 1.2143;
	[150, 300): 15.7356km^2 (0.047);	[150, 300): 38491m^2 (0.061);	[150, 300): 1.2979;
	[300, 400): 54.2376km^2 (0.162);	[300, 400): 71303m^2 (0.113);	[300, 400): 0.6975;
	[400, 450): 59.5944km^2 (0.178);	[400, 450): 79506m^2 (0.126);	[400, 450): 0.7079;
	[450, 480): 74.9952km^2 (0.224);	[450, 480): 165953m^2 (0.263);	[450, 480): 1.1741;
	[480, 520): 64.6164km^2 (0.193);	[480, 520): 148285m^2 (0.235);	[480, 520): 1.2176;
	[520, 560): 25.4448km^2 (0.076);	[520, 560): 60576m^2 (0.096);	[520, 560): 1.2632;
	[560, 600): 17.4096km^2 (0.052);	[560, 600): 39122m^2 (0.062);	[560, 600): 1.1923;
	[600, 900): 11.0484km^2 (0.033);	[600, 900): 15775m^2 (0.025);	[600, 900): 0.7576;
	[900, 1200]: 7.0308km^2 (0.021).	[900, 1200]: 1262m^2 (0.002).	[900, 1200]: 0.0952.

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Table 3 Classification results of hazard factors

Hazard factor	Classification
Slope gradient/°	I: [0, 10); II: [10, 25); III: [25, 90].
Elevation/m	I: [0, 300); II: [300, 450); III: [450, 600); IV: [600, 1200].
Slope aspect/°	I: [0, 90); II: [90, 210); III: [210, 270); IV: [270, 360].
Land use type	I: land for road and architecture, II: Cultivated land and grassland, III: Undeveloped land and other land.
Distance from fault/m	I: [0, 2000); II: [2000, 4000); III: [4000, 8000].
Distance from road/m	I: [0, 600); II: [600, 1200); III: [1200, +∞).
profile curvature	I: (-∞, -0.5); II: [-0.5, 0.5); III: [0.5, +∞).
NDVI	I: [-1, -0.1); II: [-0.1, 0.05); III: [0.05, 1].
SPI	I: [0, 30); II: [30, 70); III: [70, +∞).

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Discussions

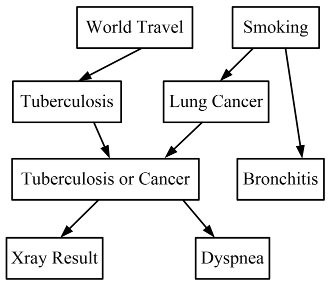


Figure 3 Directed acyclic graphs of Chest clinic Network

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

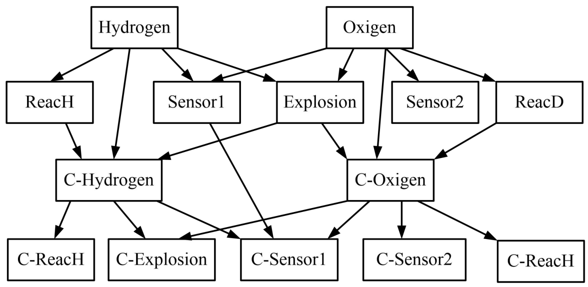


Figure 4 Directed acyclic graph of TANK Network

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions

Table 4 Comparison results of different sample sizes and algorithms

Sample size	Algorithm	Chest clinic Network				TANK Network			
		n_a	n_b	n_c	E_n	n_a	n_b	n_c	E_n
1200	MMHC	2	2	1	4.5	5	4	4	11
	MMPC-Tabu	2	1	1	4	4	4	3	9.5
	Fast.iamb-Tabu	2	1	1	3.5	3	4	2	8
	Inter.iamb-Tabu	1	1	0	2	3	3	2	7
1500	MMHC	2	1	1	3.5	3	3	4	8
	MMPC-Tabu	1	1	1	2.5	3	3	3	7.5
	Fast.iamb-Tabu	1	0	1	1.5	2	2	3	5.5
	Inter.iamb-Tabu	1	0	0	1	1	2	2	4
1800	MMHC	1	1	1	2.5	2	2	2	5
	MMPC-Tabu	0	1	1	1.5	2	1	1	3.5
	Fast.iamb-Tabu	0	1	0	1	1	1	0	2
	Inter.iamb-Tabu	0	0	0	0	0	0	0	0
2100	MMHC	1	1	1	2.5	2	2	2	5
	MMPC-Tabu	0	1	1	1.5	2	1	1	3.5
	Fast.iamb-Tabu	0	1	0	1	1	1	0	2
	Inter.iamb-Tabu	0	0	0	0	0	0	0	0

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Discussions

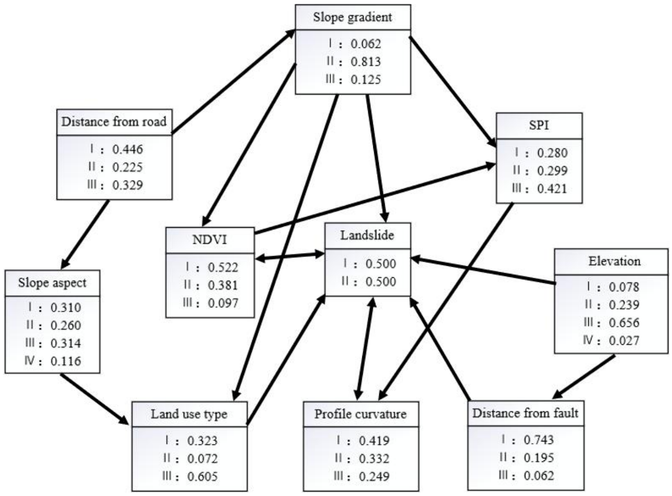


Figure 5 Landslide susceptibility model of Boshan District

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions

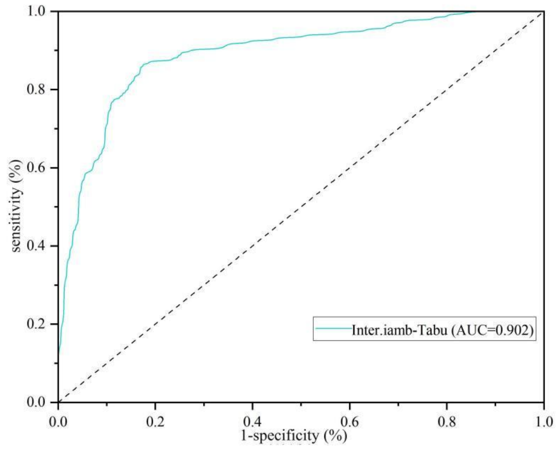


Figure 6 Verification results of landslide susceptibility model in Boshan District

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions

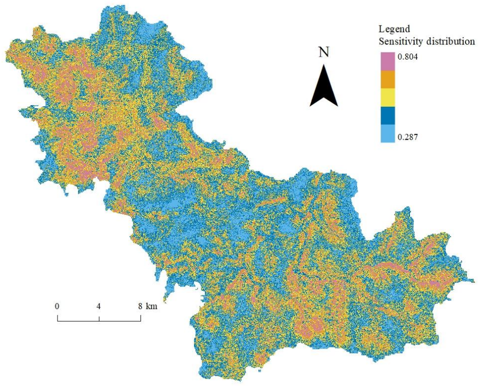


Figure 7 Probability distribution map of landslide susceptibility in Boshan area

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Discussions

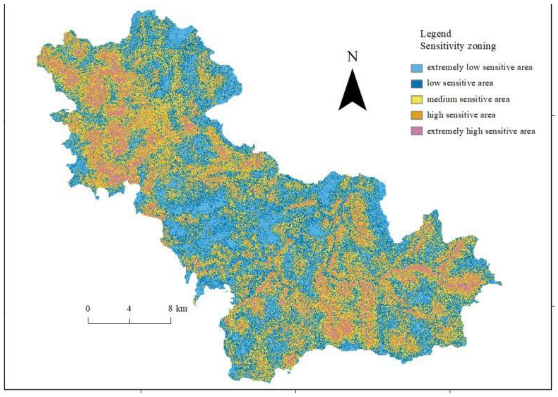
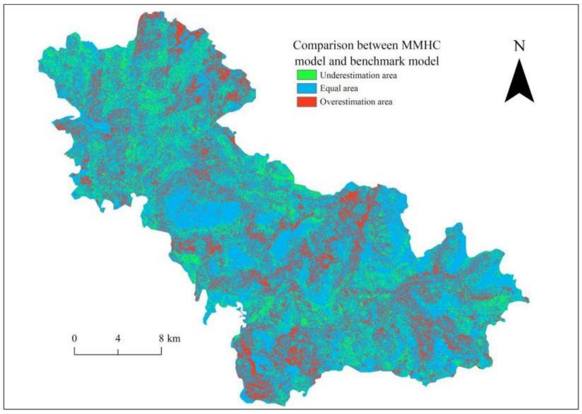


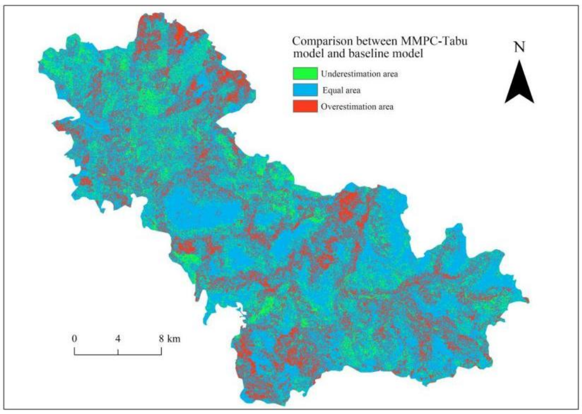
Figure 8 Landslide susceptibility zoning map of Boshan District

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

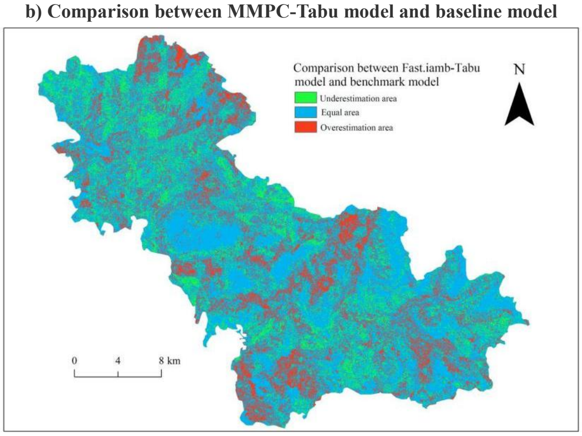
Natural Hazards and Earth System Sciences
Open Access
Discussions



a) Comparison between MMHC model and benchmark model



b) Comparison between MMPC-Tabu model and baseline model

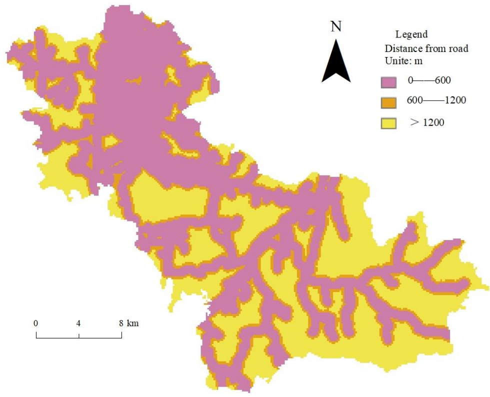


c) Comparison between Fast.iamb-Tabu model and benchmark model

Figure 9 Comparison of landslide susceptibility assessment results with different models

https://doi.org/10.5194/nhess-2023-153
Preprint. Discussion started: 5 October 2023

Natural Hazards and Earth System Sciences
Open Access
Discussions



1
2 **Figure 10 The superposition of under-estimated area and distance from road**
3

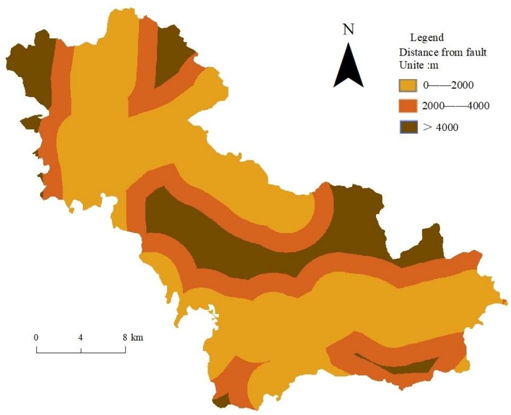


Figure 11 The superposition of over-estimated area and distance from fault