



OPEN ACCESS

EDITED BY
Chengyi Pu,
Central University of Finance and
Economics, China

REVIEWED BY
Jiangfeng Dong,
Sichuan University, China
Shen Tong Shen,
Henan University of Urban Construction,
China

*CORRESPONDENCE
Jianhui Dong,
dongjianhui@cdu.edu.cn

SPECIALTY SECTION
This article was submitted to
Environmental Informatics and Remote
Sensing, a section of the journal
Frontiers in Earth Science

RECEIVED 07 November 2022
ACCEPTED 29 December 2022
PUBLISHED 16 January 2023

CITATION
Ma W, Dong J, Wei Z, Peng L, Wu Q,
Wang X, Dong Y and Wu Y (2023),
Landslide susceptibility assessment using
the certainty factor and deep
neural network.
Front. Earth Sci. 10:1091560.
doi: 10.3389/feart.2022.1091560

COPYRIGHT
© 2023 Ma, Dong, Wei, Peng, Wu, Wang,
Dong and Wu. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Landslide susceptibility assessment using the certainty factor and deep neural network

Wenli Ma^{1,2,3}, Jianhui Dong^{4,5*}, Zhanxi Wei^{1,2,3}, Liang Peng⁶,
Qihong Wu^{4,5}, Xiao Wang^{4,5}, Yangdan Dong⁷ and Yuanzao Wu^{1,2,3}

¹Qinghai 906 Engineering Survey and Design Institute Co Ltd, Xining, China, ²Key Laboratory of Environmental Geology of Qinghai Province, Xining, China, ³Qinghai Geological Environment Protection and Disaster Prevention Engineering Technology Research Center, Xining, China, ⁴Sichuan Engineering Research Center for Mechanical Properties and Engineering Technology of Unsaturated Soils, Chengdu University, Chengdu, China, ⁵School of Architecture and Civil Engineering, Chengdu University, Chengdu, China, ⁶Qinghai Hydrogeology and Engineering Geology and Environmental Geology Survey Institute, Xining, China, ⁷State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu, China

Areas with vulnerable ecological environments often breed many geological disasters, especially landslides, which pose a severe threat to the safety of people's lives and property in these areas. To aid in landslide prevention and mitigation, an approach combining the coefficient of determination method (CF) and a deep neural network (DNN) were proposed in this study for landslide susceptibility evaluation. The deep neural network can excavate the deep features of samples and improve the accuracy of the susceptibility model. In addition, the logistic regression model (LRM) and support vector machine (SVM) were selected to create landslide susceptibility maps for comparison, which also involved the coefficient of determination method (CF). Based on landslide remote sensing interpretation and field investigations, a spatial database of mudstone landslides in the Xining area was established. Eight different conditional factors, including the elevation, slope, slope aspect, undulation, curvature, watershed, distance from a fault, and distance from a road, in the study area were selected as the evaluation factors to evaluate the susceptibility. The results revealed that four factors (i.e., the ground elevation, curvature, distance from a fault, and distance from a road) had relatively significant influences on the landslide susceptibility in the study area. Finally, the confusion matrix was used to evaluate the accuracy of the results obtained using the three methods, and the optimal result was selected to evaluate the landslide susceptibility in the study area. It was found that the combined CF-DNN method was more suitable for evaluating the landslide susceptibility in this area. Landslide susceptibility zoning was conducted to divide the study area into four sensitivity levels: low (32.65%), medium (35.12%), high (22.44%), and extremely high (9.79%) susceptibility. The high-risk areas were primarily distributed in the high-elevation areas along the eastern edge of the Huangshui Basin.

KEYWORDS

xining area, landslides, susceptibility assessment, GIS, coefficient of determination method, deep neural network

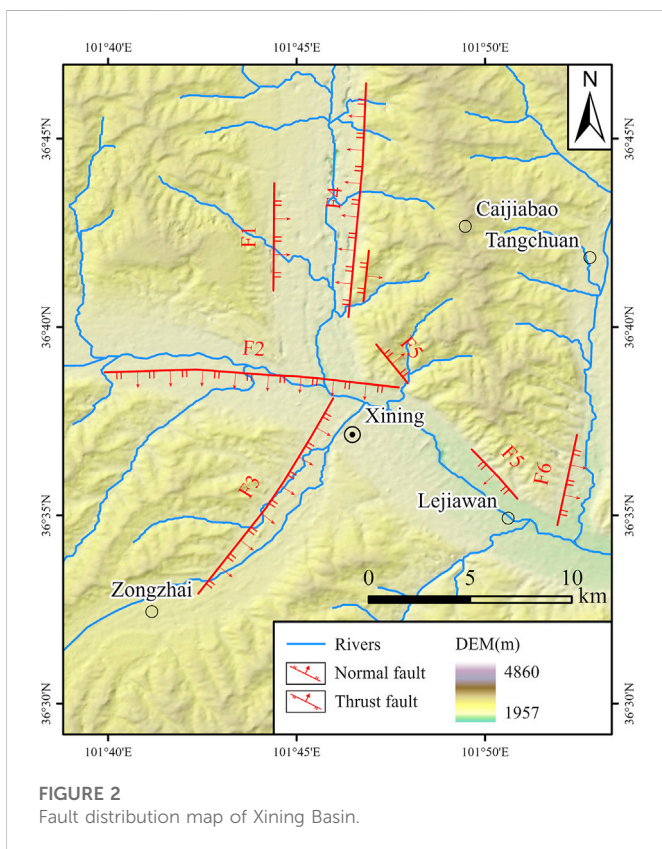
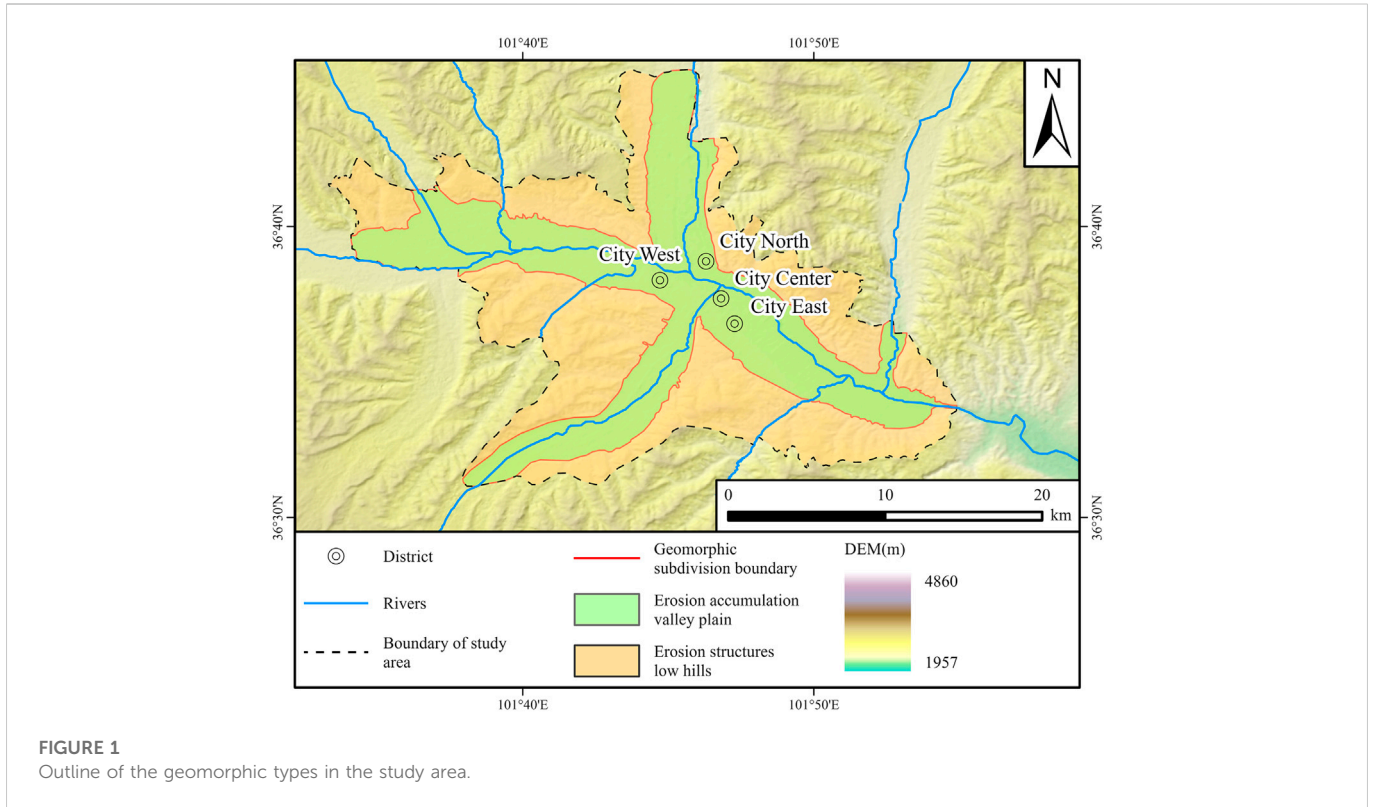
1 Introduction

The term “landslide” refers to the phenomenon in which the soil and rock mass constituting the slope disintegrates and slips and/or collapses under the action of gravity. Landslides are regarded as one of the most common types of geological disasters in mountainous areas (Confuorto et al., 2019). They are characterized by wide distribution, high frequency of occurrence, strong concealment, and a large degree of destruction, and thus, they often have catastrophic consequences and pose a huge threat to people’s lives and property, as well as social and economic development (Lin et al., 2008; Wang et al., 2018). Since the middle of the 20th century, with the increasing population and the increasing scope of human engineering activities, more and more landslides have been induced, especially in the Xining area in Qinghai Province. Xining City is one of the areas with the most vulnerable ecological environment in China, located in the transition zone between the western side of the Loess Plateau and the Qinghai-Tibet Plateau. The geological environment in this region is very complex, consisting of undulating terrain, multi-periodic tectonic movements, complex geological structures, weak formation lithology, strong river erosion, leading to a high risk of landslides. With the continuous development of economic construction in Xining, human engineering activities have intensified. Consequently, more and more geological disasters have been induced, especially landslides, seriously threatening people’s lives and property in this area. For example, the Zhangjiawan Brick Factory landslide in 1960 (Bai et al., 2021; Peng et al., 2021), which had a volume of about $2.2 \times 10^7 \text{ m}^3$, the Beishansi landslide in 2001 (Yao et al., 2014), which had a volume of about $1.38 \times 10^6 \text{ m}^3$, and the Hanzhuang community landslide in 2011, which had a volume of about $7.326 \times 10^6 \text{ m}^3$, were larger mudstone landslides that have occurred in this area. According to a geological survey conducted by the staff of the Qinghai Provincial Geological Environment Monitoring Station, mudstone landslides accounted for 78% of the total number of landslides investigated in the study area. The occurrence of landslides in this area is greatly related to the poor permeability of the mudstone, that is, when water collects in the slope is affected by the water resistance of the mudstone, eventually forming a soft slip zone. The deformation mechanisms and formation mechanisms of mudstone landslides in Xining have been investigated in previous studies (Xin et al., 2015). However, the distribution characteristics of mudstone landslides in Xining City have rarely been studied. Assessing the susceptibility of the study area to mudstone landslides is beneficial to providing strong support for the prediction and control of mudstone landslides.

Landslide susceptibility assessment is the basis of landslide risk assessment, which refers to the possibility of landslide occurrence caused by a combination of multiple influencing factors in a specific area (Brabb, 1984). In earlier studies, scholars mainly relied on empirical methods to qualitatively determine landslide susceptibility, such as the analytic hierarchy process (AHP), and fuzzy mathematics (Al-Harbi, 2001; Rozos et al., 2011; Giovanni et al., 2016). For small areas, the method of synthesizing the descriptions provided by experts has a certain reference value for landslide susceptibility rating; however, for large areas, such methods are subjective due to manual determination, which leads to the possibility of large errors in the results. In later studies, scholars began to use statistical methods to analyze and assess landslide susceptibility, such as the information model (IM), certainty

coefficient method (CF), frequency ratio model (FR), and weight of evidence method (WEM) (Pradhan and Lee, 2010; Liu et al., 2014; Pourghasemi et al., 2014; Xu et al., 2016; Dai et al., 2017). Statistical methods have become one of several methods that are commonly used at present. Although the use of such quantitative methods avoids the disadvantages of the subjective judgment of the previous empirical experts, there are still great limitations in the prediction and evaluation of regional landslides using many evaluation indicators and the statistics of a large dataset. Therefore, with the continuous development of machine learning methods, more and more scholars have begun to use machine learning methods to assess landslide susceptibility, such as the logistic regression model (LRM), decision tree model (DTM), random forest model (RFM), support vector machine (SVM), naive bayes model (NBM), artificial neural networks (ANNs), and other methods (Bui et al., 2016; Hong et al., 2016; Wang et al., 2016; Chen et al., 2017; Le et al., 2017; Shirzadi et al., 2017; Chen et al., 2018; Sahin et al., 2018; Dou et al., 2019; Hong et al., 2019; Shariati et al., 2019; Nhu et al., 2020). However, the above models all belong to the category of shallow neural networks. When the landslide susceptibility assessment process involves diverse and interrelated evaluation factors, it is often impossible to explore the relationship between the factors to achieve more accurate results (Wang et al., 2019). Therefore, in recent years, the use of deep learning methods to extract landslide factors and characteristic information about landslides has become one of the important directions in landslide disaster research (Zheng, 2019). The structure of the deep neural network model (DNN) has more layers, leading to stronger learning ability and a stronger ability to express the characteristics of objects. When faced with a large amount of data, it has a higher recognition and evaluation accuracy. For example, training a large number of geological disaster data samples through a deep neural network model (DNN) improves the accuracy of geological disaster risk prediction and early warning (Li et al., 2018). In this study, the unified sample is compared with different models to prove that the deep neural network can still improve the accuracy in small samples.

Therefore, in this study, the DNN model was introduced into the susceptibility modeling of mudstone landslides in Xining City. In addition, based on the quantification of the sensitive factors and their correlations with the occurrence of landslides, scholars have often used two methods or even multiple methods and models coupled to reflect the characteristics of the study area. For example, Fan et al. (2017) used a combination of deterministic factors (CF) and the analytic hierarchy process (AHP) (Fan et al., 2017), Li et al. (2018) used the coefficient of determination (CF) method coupled with support vector machines (SVMs) (Luo et al., 2021), and Luo et al. (2021) coupled the CF and logistic regression model (LRM) with geographic information system (GIS) support (Akgun, 2012). All of the above studies proved that the evaluation results of coupled models are more reasonable than those of a single model. Among them, the deterministic coefficient method, as a bivariate statistical method, can be used to determine the weight of each conditional factor of the landslide to analyze the correlation between each factor and the occurrence of the landslide. Therefore, in this study, a coupled CF-DNN method was developed to analyze the mudstone landslides in Xining City. Given that the accuracy and reliability of logistic regression models and support vector machine models have been verified in previous studies (Akgun, 2012; Li et al., 2018; Luo et al., 2021), in this study, the LRM and



In this study, 1) a landslide database was constructed based on original landslide-related data obtained from landslide remote sensing interpretation and field investigations; 2) based on eight condition factors, a landslide inventory map was created and combined with the deterministic factor results to analyze the main impacts of the landslides factors; 3) a mudstone landslide susceptibility map of Xining was created using the ArcGIS software platform; and 4) the confusion matrix was used to evaluate the accuracy of the CF-DNN model.

2 Study area and data

In this study, Xining City, Qinghai Province, China, was taken as the study area (101°33'45"–101°56'15"E, 28°49'–29°31'N; Figure 1A). The Xining Basin is located in the northeastern part of the Qinghai-Tibet Plateau. Under the superposition of multi-stage tectonic movements, fold and fault structures have formed. The fold structure is mainly dominated by the regional Huangshui anticline (Figure 2). The fault structure is generally characterized by superficial tension. The formation and evolution of the regional landscape pattern are strictly controlled by the dominant NW-NWW trending tectonic system. Under the coupling of multi-factor mechanisms, the study area was finally shaped into the current erosion structure, with low hills and an erosion deposit valley plain (Figure 1B). The low hills of the erosion structure are distributed above the third-level terraces at the edges of the valleys of the Huangshui River and its tributaries. The hilly area covers 218.76 m², accounting for 54.63% of the total area. Due to the strong erosion of the gullies, the terrain is very fragmented. The elevation of the hilly area is 2,450–3,500 m, and the relative height difference is 100–500 m. On

SVM were reasonably used as the reference models of the DNN, which was combined with the deterministic coefficient method (CF).

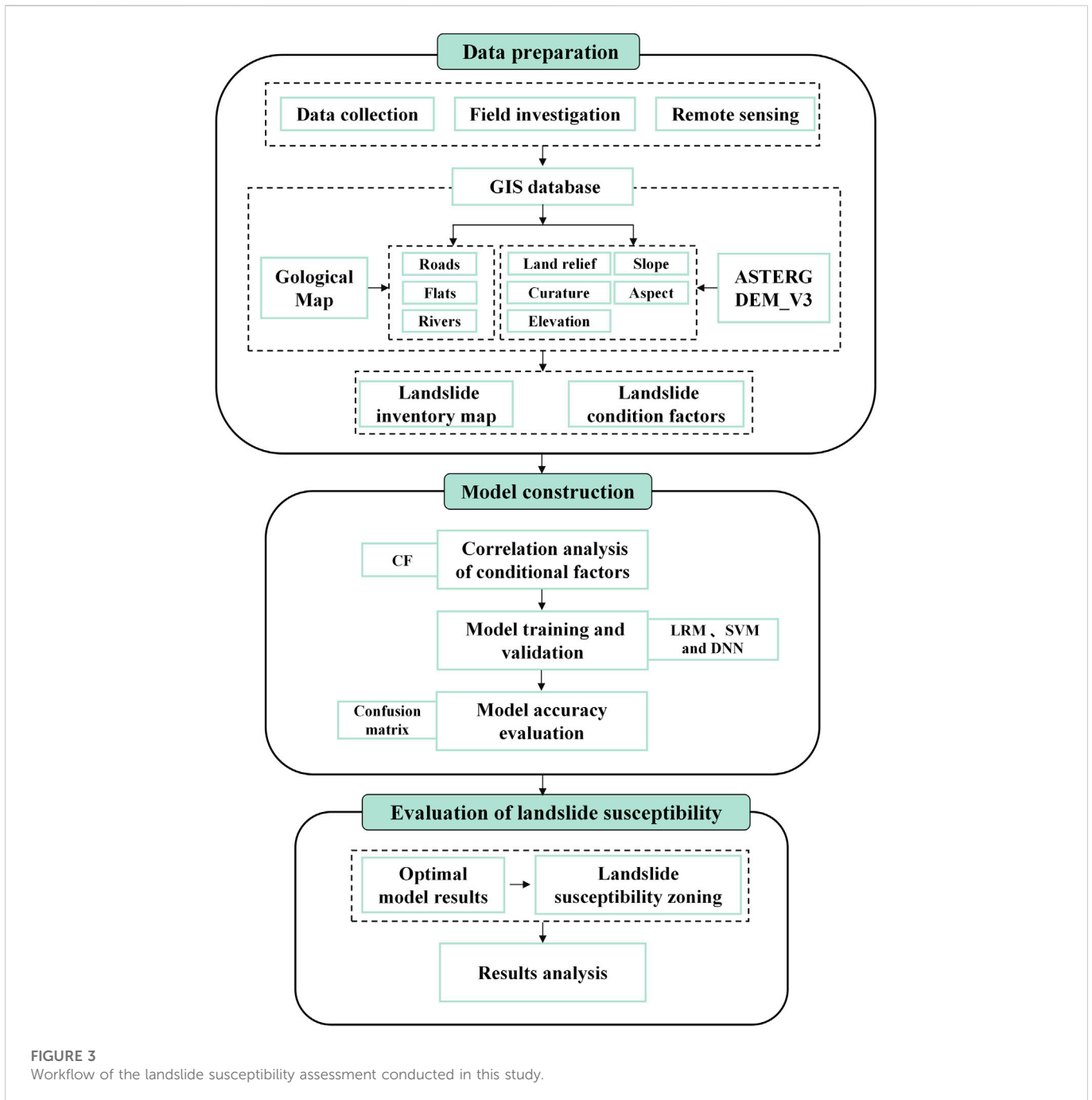


FIGURE 3
Workflow of the landslide susceptibility assessment conducted in this study.

the front edge of low mountains and hills, there are high and steep terrace (grades III to VII) slopes with heights of 100–300 m, and landslides are relatively more developed on these slopes. The erosion deposit valley plain is the urban area of Xining, with an area of 181.65 m², accounting for 45.37% of the study area. The terrain of the river valley area is relatively broad and is distributed in bands along the river. The terrain is high in the west and low in the east, and the terrace surface is slightly inclined toward the riverbed. The elevation is 2,200–2,450 m.

The exposed strata in the study area are relatively complex, including pre-Quaternary strata and Quaternary strata. The pre-Quaternary strata include the Changcheng system, Cretaceous

strata, and Tertiary strata, among which the Tertiary strata are the most exposed and are mainly composed of mudstone. Most of the exposed mudstone has a low strength, high degree of weathering, and poor permeability. When the water collects on the slope surface, the slope material becomes saturated due to its own water-blocking properties, making it easy for a weak sliding zone to form. A mudstone with a high degree of weathering can easily become the source of potential landslides. Thus, in this paper, the mudstone area in Xining was selected as the target research area. In terms of meteorology and hydrology, the study area has a plateau semi-arid continental climate, with a long winter and short summer, a large temperature difference, sparse but concentrated rainfall, and strong

evaporation. The Huangshui Basin is the main river basin in the study area. The Huangshui River is a first-class tributary of the Yellow River, which flows through the urban area from west to east. The river length in the study area is 35 km. The annual average flow at Xining Hydrological Station is 32.8 m³/s, the annual average minimum flow is 4.58 m³/s, and the maximum flow is 698 m³/s.

The unique tectonic, landform, lithology, meteorology, and hydrology of this area have resulted in the occurrence of a variety of disasters, and various adverse geological phenomena induced by external dynamic geological action are very prominent. Currently, according to the field investigation results reported by the staff of the Qinghai Provincial Geological Environmental Monitoring Station, a total of 294 landslide hidden danger points have been investigated, including 184 mudstone landslides, accounting for 62.58% of the total landslide hidden danger points investigated. Based on a GIS platform, five terrain factors (i.e., the elevation, slope, land relief, aspect, and curvature) and three geological factors (distance from a road, distance from a fault, and distance from a river) were selected. Topographic factors can be extracted from DEM data which is derived from the open geospatial data cloud platform. Geological factors are derived from the geological map provided by Qinghai Provincial Geological Environment Monitoring Station.

3 Methods

The basic workflow of the landslide susceptibility assessment conducted in this study is presented in Figure 3.

In this section, we mainly describe the methods involved in the landslide susceptibility assessment process. In this study, three different machine learning models (i.e., a logistic regression model, support vector machine, and deep neural network) were coupled with the certainty coefficient method. The data for the eight conditional factors were processed using a GIS platform. The certainty coefficient (CF) was calculated from the processed data. Then, the certainty factor was used to analyze the correlation between each condition factor and landslide. This information was then plotted on the landslide inventory map, which was convenient for the subsequent modeling of the landslide susceptibility model using *Python*.

3.1 Preparation of the dataset for modeling

The mudstone landslide-related data for Xining City were collected from historical data, field investigation data, and remote sensing interpretation data. A total of 184 mudstone landslides have been identified in the study area. According to the coordinates of the study area, these landslides were marked as points on the 1:10,000 terrain map of the area, and these points are considered to be the location of the geometric center of the slide. Based on the landslide data and 30 m DEM data, a preliminary landslide inventory map containing 184 landslides was generated, these landslides are all mudstone landslides. In addition, using a GIS platform, eight graded maps of the landslide condition factors were compiled in a grid format. The number of pixels of all of the factors for each classification in the reclassified grid data was counted. After conversion to vector data, each classification was connected to the space to count the number of landslides within the classification. The CF value was calculated using the number of statistical pixels and the number of landslides. The

obtained CF values were assigned in the vector layers of each factor. Before the modeling, based on the statistics of landslide scale in the study area, a circular buffer was built on the landslide point with a radius of 300 m, and non-landslide points were randomly generated outside the buffer to generate non-landslide samples equal to the number of landslide samples. The landslide and non-landslide sample points were coded as one and zero. After merging, a total of 368 samples were generated, which were randomly divided into training samples (70%) and validation samples (30%). The final dataset included a landslide inventory map, a hierarchical map of the landslide condition factors with CF values, and landslide and non-landslide samples.

3.2 Model

3.2.1 Logistic regression model

The logistic regression model (LRM) is a commonly used machine learning model for the binary classification of a dependent variable. In this paper, the relationship between the occurrence of landslides (the dependent variable) and multiple hazard factors is described (Menard, 1995; Atkinson and Massari, 1998). The independent variable in this model can be continuous or discrete, and it does not need to satisfy a normal frequency distribution (Bai et al., 2015; Wang et al., 2015). The expression of the logistic regression core function is

$$P = 1 / [1 + e^{(\alpha + \beta_1 x_1 + \dots + \beta_n x_n)}] \quad (1)$$

where P is the probability of landslide occurrence (0–1); α is a constant calculated *via* logistic regression; β is the regression coefficient calculated *via* logistic regression; and i is the number of evaluation factor types. By taking the natural logarithm of both sides of Eq. 1, $\ln(p/(1-p))$ as the dependent variable, and the impact factor x ($i = 1, 2, \dots, n$) as the independent variable, we obtain.

$$\ln(P/(1-P)) = \alpha + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_n x_n = \alpha + \beta x \quad (2)$$

3.2.2 Support vector machine model

The support vector machine is a binary classification prediction model developed based on statistical principles (Vapnik, 1998; Abe, 2010), which is similar to a neural network but differs in that the SVM uses mathematical methods and optimization techniques. SVMs include linear support vector machines and non-linear support vector machines. In the case of linear separability, an optimal classification function is obtained by transforming a constrained extreme value problem into a dual problem. The binary classification prediction problem of landslide susceptibility zoning is often non-linear; thus, it is necessary to map the original data to a high-dimensional feature space and make it linearly separable to identify the optimal classification plane (Brenning, 2005). That is, the optimal classification plane is the plane that maximizes the separation of the data points belonging to two different classes. The core principle of identifying this plane is to introduce the corresponding kernel function after transforming it into a dual problem (Yao et al., 2008), which solves the problem of the increased complexity caused by mapping to a high-dimensional space. At present, the most commonly used kernel functions include the linear kernel function, polynomial function, radial basis function (RBF), and Sigmoid function. However, there is a certain error in the sample after the kernel function is determined. For outliers that do not meet the constraints, the slack

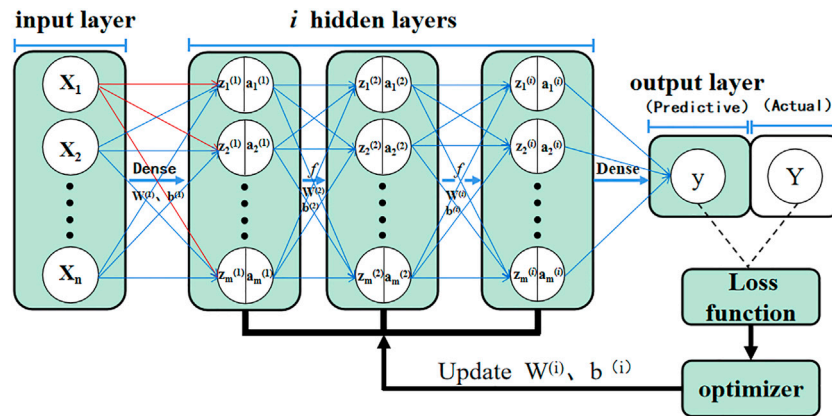


FIGURE 4
Schematic diagram of the DNN.

variable ϵ_i and penalty factor c are introduced for optimization. The ϵ_i value actually represents the outlier distance of the corresponding point. The larger the value is, the farther away the point is, and ϵ_i is equal to zero for non-outlier points. The penalty factor refers to the tolerance of the error, which must be determined in advance. The larger the value is, the greater the loss of the objective function, which easily leads to overfitting. Conversely, the smaller the value is, the easier it is for underfitting to occur.

The calculation process of SVM is based on a set of linearly inseparable landslide data x_i ($x_i \in R^d$), where i is the number of samples, d is the dimension, y_i is the corresponding output data, and $y_i \in \{-1, 1\}$, which means the output results represent landslide and non-landslide points. The original data are mapped to a certain feature space through non-linear mapping $\varphi(x)$ and distinguishing landslide categories using the maximum interval $\|\omega\|^2/2$ in the hyperplane equation $\omega \cdot \varphi(x) + b = 0$. That is, $\|\omega\|^2$ needs to be the minimum value, and the following constraints must be satisfied:

$$y_i(\omega \cdot x_i + b) \geq 1 - \epsilon_i (\epsilon_i \geq 0) \tag{3}$$

While solving the hyperplane, the slack variables should be as small as possible, so the problem is transformed into a quadratic programming problem of finding the minimum value of $\|\omega\|^2/2 + c(\sum_{i=1}^n \epsilon_i)$ under the constraints. The kernel function $K(x_i, y_i)$ is introduced to obtain the classification function by solving.

$$f(x) = \text{sgn}\{\sum_{i=1}^n a_i y_i \cdot K(x_i, y_i) + b\} \tag{4}$$

3.2.3 Deep neural network mode

A complete neural network consists of an input layer, a hidden layer, and an output layer. Each layer is built on multiple neurons. Each neuron can be regarded as a simple linear function, and the neurons on adjacent layers are combined in a densely connected manner. The difference is that a deep neural network has multiple hidden layers, and there is no connection within the same layer (between hidden layers). The propagation process of the DNN is divided into forward propagation composed of an input layer-hidden layer-output layer and backward propagation composed of a loss function optimizer. A schematic diagram of the DNN is shown in Figure 4.

Forward propagation refers to the calculation and recognition process of the original data starting from the input layer, through the

hidden layer calculation, finally reaching the output layer, and outputting the result from the output layer. When sample X is input, the DNN automatically assigns the initial value to the weight matrix W , and it continues the calculations in the hidden layer according to Eqs 5–7 until the result y can be output in the output layer.

$$Z^{(i)} = W^{(i)}X + b^{(i)} \tag{5}$$

$$a^{(i)} = f(Z^{(i)}) \tag{6}$$

$$y = W^{(i)}a^{(i-1)} + b^{(i)} \tag{7}$$

where X is the input sample, denoted as (x_1, x_2, \dots, x_n) , and n is characterized number. Z is the output result, denoted as $(Z_1^{(i)}, Z_2^{(i)}, \dots, Z_m^{(i)})$, where i is the number of hidden layers, and m is the number of neurons in the next layer. W is the weight matrix; b is a constant term; and a is the input value of the next layer of the network, that is, substitute it into X in Eq. 5 and continue to calculate forward. f is the activation function, the purpose of which is to convert linear to non-linear after each layer of network calculation is completed to improve the recognition rate, and the Sigmoid function, Tanh function, and rectified linear unit (ReLU) function are commonly used.

The output result y is the predicted value, and there is a significant difference between it and the real value Y . Obtaining the correct W and b is the key factor affecting the output result. The essence of backpropagation is to calculate the difference between the predicted value and the true value according to the loss function to reversely adjust the parameters W and b and select a suitable optimizer to update the parameters. The essence of the neural network is to obtain values of the parameters W and b with higher accuracy through multiple forward and reverse circular training to obtain the final optimized model. For the classification problem of landslide susceptibility zoning, the loss function adopted is generally the cross entropy function, and optimizer methods such as the gradient descent and Adam method can be used.

3.2.4 Certainty coefficient method

The certainty coefficient method, proposed by Shortliff (1975), was used in this study to analyze the correlation between each

condition factor and the occurrence of landslides (Shortliffe and Buchanan, 1975). Although the evaluation process of the certainty coefficient model is relatively simple, the accuracy is high. The premise of the high accuracy is that the geological disasters that have occurred and the disasters that occur in the future occur under the same geological conditions. The calculation formula is as follows:

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_a(1 - PP_s)}, PP_a \geq PP_s \\ \frac{PP_a - PP_s}{PP_s(1 - PP_a)}, PP_a < PP_s \end{cases} \quad (8)$$

where PP_a is the conditional probability of landslides in impact factor classification a , which is expressed as the ratio of the landslide points in classification a to the classification area; and PP_s is the probability of historical landslides in the entire study area, which is expressed as the ratio of historical landslide points to the total area of the study area.

The CF ranges from -1 to 1 . When the CF values are between zero and 1 , landslides are prone to occur in this geological environment. The closer the CF value is to 1 , the more sensitive the unit is to landslides. In contrast, when the CF value is between -1 and zero, landslides are unlikely to occur in this geological environment.

The data mentioned in the formula are all derived from the data in the conditional factor grading diagram, including the number of landslides included in each grading, the number of landslide pixels, the total number of landslides and the total number of grids. cf values calculated according to the formula will be assigned to the attribute table of each conditional factor grading diagram.

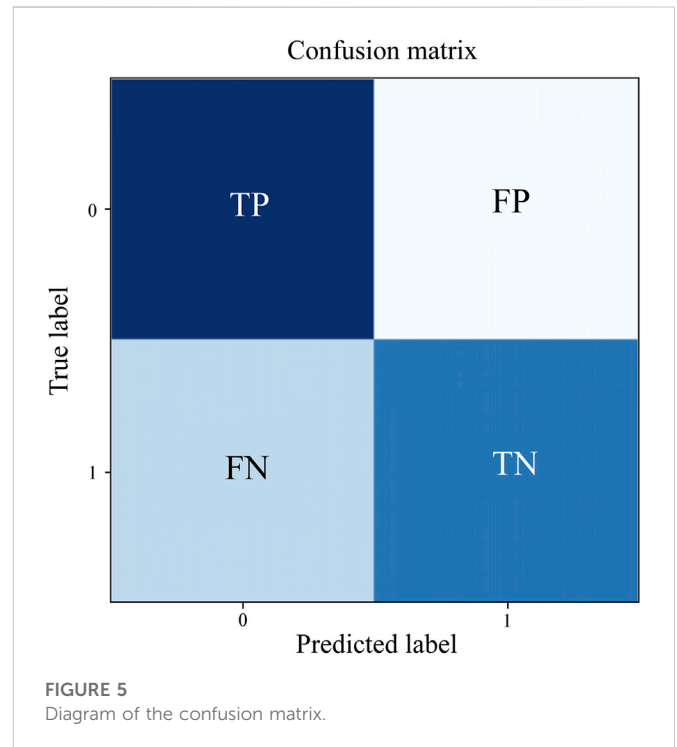
3.3 Factor multicollinearity diagnostics

Multicollinearity refers to the linear correlation between independent variables. If multicollinearity exists, the matrix is irreversible when calculating the partial regression coefficient of the independent variable. The results of variance analysis of the whole model are inconsistent with the test results of regression coefficients of each independent variable. In this study, a variety of different models were used for sensitivity analysis. Here, the variance inflation factor diagnostic method and tolerance value method were combined to determine whether the selected factors could be fully incorporated into the model.

Variance inflation factor (VIF) is a measure of the severity of multicollinearity in a multiple linear regression model. It represents the ratio of the variance of the estimator of the regression coefficient to the variance when the independent variables are assumed to be not linearly correlated. When VIF is large, it indicates that there is multicollinearity between the independent variables. This diagnostic method also has the problem that the threshold value is not easy to determine, so it needs to be considered in combination with the tolerance value method. The tolerance value is actually the inverse of VIF. Its value ranges from 0 to 1 . The closer Tol is to 1 , the weaker the collinearity between independent variables is, indicating that the factor can be completely entered into the model through multiple collinearity diagnosis.

3.4 Accuracy evaluation method

There are various methods for evaluating the accuracy of landslide susceptibility results. In this study, the confusion matrix was chosen as



the method for evaluating the model's accuracy. The confusion matrix is an effective tool in machine learning for evaluating the accuracy of each iteration of the model, that is, it can be used to evaluate the performance of a classification model based on a set of test data with known true values. For a binary classification problem such as landslide *versus* non-landslide, when judging samples, zero is used to represent landslide, and one is used to represent non-landslide. In the early stage, we know which data are landslide and non-landslide data in the real situation through sample collection, and thus, we know the predicted value of the data through the results of the sample data output by the classification model. Therefore, four first-level indicators TP, TN, FP, and FN are formed according to the combination of the actual values of one and zero and the predicted values of one and zero (Figure 5).

For a predictive scoring model, the higher the TP and TN values are, and the lower the FP and FN values are, the higher the model accuracy is. When using a large amount of data, it is difficult to measure the reliability based only on the first-level basic indicators. Therefore, four indicators (ACC, PPV, TPR, and TNR) are added to the results of the basic statistical data of the confusion matrix. These indicators are calculated as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$PPV = \frac{TP}{TP + FP} \quad (10)$$

$$TPR = \frac{TP}{TP + FN} \quad (11)$$

$$TNR = \frac{TN}{TN + FP} \quad (12)$$

where ACC is the accuracy rate, which refers to the proportion of all of the results that are judged to be correct by the classification model compared to the total observations; PPV is the positive predictive

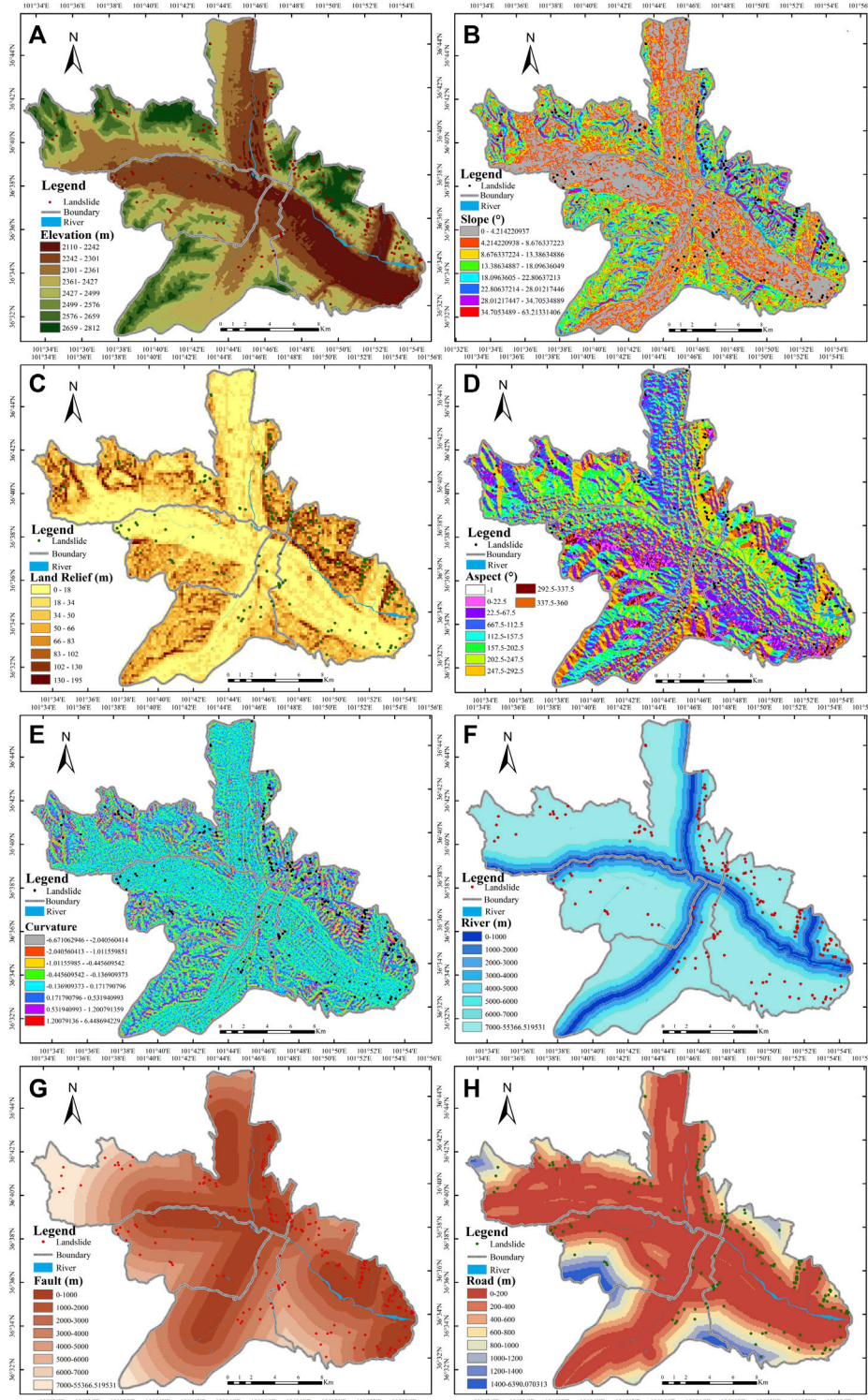


FIGURE 6 Grading maps of the evaluation factors for landslide susceptibility assessment in Xining City: (A) Elevation; (B) Slope; (C) Land relief; (D) Aspect; (E) Curvature; (F) Distance from a river; (G) Distance from a fault; and (H) Distance from a road.

value, which refers to the proportion of all of the results with a predicted value of landslide that the model predicts correctly; *TPR* is the true positive rate, which refers to the proportion of all of the result with a true value of landslide that the model predicts correctly; and

TNR is the true negative rate, which refers to the proportion of all of the result with a true value of non-landslide that the model predicts correctly. The larger the values of these four secondary indicators are, the better the results of the model are.

TABLE 1 Grading of factors and CF values.

Evaluation factor	Elevation (m)									
Grading	2,110–2,242	2,242–2,301	2,301–2,361	2,361–2,427	2,427–2,499	2,499–2,576	2,576–2,659	2,659–2,812		
Landslide proportion	10.32	19.02	10.32	15.21	23.91	7.07	8.15	5.98		
Grading area ratio	11.54	20.29	16.29	15.67	12.59	10.03	8.98	4.60		
CF	-0.1053	-0.0627	-0.3662	-0.0292	0.4737	-0.2955	-0.0921	0.2299		
Evaluation factor	Slope (°)									
Grading	Flat	North	Northeast	East	Southeast	South	Southwest	West	Northwest	North
Landslide proportion	0	2.17	14.67	11.96	8.70	13.59	12.50	20.65	8.70	5.98
Grading area ratio	0.17	6.77	14.80	15.03	13.00	12.67	11.73	11.19	9.77	4.87
CF	-1	-0.6752	0.0021	-0.1959	-0.3239	0.0780	0.0718	0.4641	-0.1001	0.1951
Evaluation factor	Land relief (m)									
Grading	0–18	18–34	34–50	50–66	66–83	83–102	102–130	130–195		
Landslide proportion	18.48	15.22	11.96	11.41	16.3	11.96	9.78	4.89		
Grading area ratio	30.89	19.11	14.25	13.77	10.33	6.77	3.79	1.1		
CF	-0.675	-0.5672	-0.5442	-0.5496	-0.1657	-0.0412	0.2871	0.5866		
Evaluation factor	Aspect (°)									
Grading	0–4.21	4.21–8.68	8.68–13.39	13.39–18.10	18.10–22.81	22.81–28.01	28.01–34.71	34.71–63.2		
Landslide proportion	21.20	13.59	10.87	9.78	16.30	13.04	10.87	3.26		
Grading area ratio	27.36	23.59	15.63	12.96	9.63	6.27	3.42	1.14		
CF	-0.5743	-0.6836	-0.618	-0.5856	-0.0756	0.1253	0.4283	0.3645		
Evaluation factor	Curvature									
Grading	-6.67–-2.04	2.04–-1.01	-1.01–-0.45	-0.45–-0.14	-0.14–0.17	0.17–0.53	0.53–1.20	1.20–6.45		
Landslide proportion	1.09	1.63	8.7	23.37	32.07	23.37	7.61	2.17		
Grading area ratio	0.13	1.25	7.28	21.55	39.46	21.36	8.09	0.89		
CF	0.7872	-0.4078	-0.5392	-0.6972	-0.5584	-0.6823	-0.4888	0.2477		
Evaluation factor	Distance from a river (m)									
Grading	0–200	200–400	400–600	600–800	800–1,000	1,000–1,200	1,200–1,400	1,400–6,390		
Landslide proportion	1.09	1.09	4.89	10.87	10.33	13.59	7.61	50.54		
Grading area ratio	6.29	7.62	6.22	6.52	6.13	5.2	4.9	57.12		
CF	-0.8279	-0.8580	0.2144	0.4022	0.4086	0.6203	0.3585	-0.1157		
Evaluation factor	Distance from a fault (m)									
Grading	0–1,000	1,000–2,000	2,000–3,000	3,000–4,000	4,000–5,000	5,000–6,000	6,000–7,000	7,000–1		
Landslide proportion	45.65	17.39	14.13	9.24	7.07	3.26	1.09	2.17		
Grading area ratio	25.02	20.9	17.53	14.62	8.9	5.33	3.17	4.54		
CF	0.4544	0.1686	0.1948	0.3694	0.2067	0.3889	-1.923	0.5227		
Evaluation factor	Distance from a road (m)									
Grading	0–500	500–1,000	1,000–1,500	1,500–2,000	2,000–2,500	2,500–3,000	3,000–3,500	3,500–		
Landslide proportion	34.24	40.76	13.04	6.52	2.72	2.17	0.54	0		
Grading area ratio	55.52	16.05	9.37	7.39	5.11	3.28	1.88	1.4		
CF	0.3846	0.6095	0.2833	-0.1178	-0.4696	-0.3388	-0.7125	-1		

TABLE 2 The factor multicollinearity diagnostics.

Variable	TOL	VIF
Elevation	0.919,236	1.087859
Slope	0.955,998	1.046027
Land relief	0.482,606	2.072085
Aspect	0.903,503	1.106,803
Curvature	0.435,019	2.298,749
Distance from a river	0.943,796	1.059551
Distance from a fault	0.895,788	1.116,335
Distance from a road	0.952,347	1.050038

4 Results and discussion

4.1 Impact factor analysis

Based on the original landslide-related data obtained from landslide remote sensing interpretation and field investigations, eight condition factors, including elevation, slope, land relief, aspect, curvature, distance from a river, distance from a fault, and distance from a road, were determined. Using a GIS platform, the data for these eight conditional factors were used to create a graded map of landslide susceptibility evaluation factors (Figure 6), which was combined with the deterministic factor results (Table 1) to analyze the influence of each factor on the occurrence of landslides. All of these eight factors have passed the multicollinearity diagnosis and can be completely entered into different models. See Table 2 for details.

4.1.1 Elevation

The elevation reflects the land changes to a certain extent and is an important factor affecting the stress value of the slope. Based on the DEM data, in this study, the natural discontinuity method was used to divide the elevation. The natural break method considers variation in groups of LSI values by calculating the minimum sum of variance of different groups for the optimal classification of LSI values and can be conveniently applied in ArcGIS software. In this method, statistical principle is considered to avoid subjectivity in obtaining susceptibility zonation. Combined with the chart, we can see that the terrain of the Xining urban area is relatively flat and broad and distributed in a band along the valley. The terrain is high in the west and low in the east. The terrace surface slopes slightly towards the riverbed. The altitude is higher farther away from the urban area, and this is also the main distribution area of landslides.

4.1.2 Slope

The slope represents the steepness of the surface. The size of the slope not only affects the stress distribution inside the slope but also controls the surface runoff, groundwater recharge, and rainfall infiltration. In this study, based on the DEM data, the slope analysis function in ArcGIS based on the gridded surface was used to analyze the slope in the study area. Combined with the chart, we can see that the landslides are mainly distributed in the slope range of 28.01°–63.2°. The greater the slope is, the greater the probability of landslide occurrence when the slope is affected by adverse external environmental factors.

TABLE 3 Secondary index values of each model in the confusion matrix.

	LRM	SVM	DNN	Best result
ACC	0.7368	0.7456	0.7719	DNN
PPV	0.8245	0.7719	0.8701	DNN
TPR	0.7014	0.7333	0.7541	DNN
TNR	0.7872	0.7592	0.7924	DNN

4.1.3 Land relief

The land relief factor represents the difference between the highest point and the lowest point in a certain area. The greater the land relief is, the greater the possibility of landslide occurrence. Based on the DEM data, in this study, the block statistics and raster calculator functions in ArcGIS were used to analyze the land relief in the study area. Combined with the chart, we can see that the area most prone to landslides is mainly distributed in the land relief range of 102–195 m, and the area with this land relief range accounts for 4.89% of the total area of the study area, including 14.67% of the landslide area. The CF value becomes closer to one as the land relief increases, which is consistent with the principle that landslides are prone to occur in areas with a large land relief.

4.1.4 Aspect

Different slopes receive different sunshine hours and solar radiation intensity. This leads to differences in the soil moisture content and weathering degree on the slope, which ultimately affects the stability of the slope. Based on the DEM data, in this study, the aspect analysis function of the grid surface in ArcGIS was used to analyze the slope aspect in the area. Combined with the chart, we can see that the slopes most prone to landslides mainly had due west aspects (247.5°–292.5°), accounting for 11.19% of the study area, including 20.65% of the landslide area. As can be seen from the CF values, in the study area, the slopes with aspects ranging from the south to west (sunny slope) were more susceptible to landslides. This is related to the higher degree of weathering of the rock and soil mass due to prolonged exposure to sunlight.

4.1.5 Curvature

The curvature represents the degree of deformation of a point on the surface of the slope. A positive value indicates a convex slope, a negative value indicates a concave slope, and a value of zero or close to zero indicates a relatively flat slope. Based on the DEM data, in this study, the curvature analysis function of the ArcGIS grid surface was used to analyze the curvature of the area. Combined with the chart, we can see that the areas most prone to landslides mainly have curvatures of -6.67 to -2.04 and 1.2 – 6.45 m^{-1} , and the area with this curvature accounts for 1.02% of the study area, including 3.26% of the landslide area. The closer the curvature is to zero, the closer the CF value is to -1 , indicating that the greater the curvature of the slope is, the greater the probability of landslide occurrence.

4.1.6 Distance from a river

Different degrees of scouring and erosion on both sides of the river lead to instability of the slope foot, which eventually induces landslides. Based on the geological map data, in this study, the Euclidean distance function in ArcGIS was used to analyze the

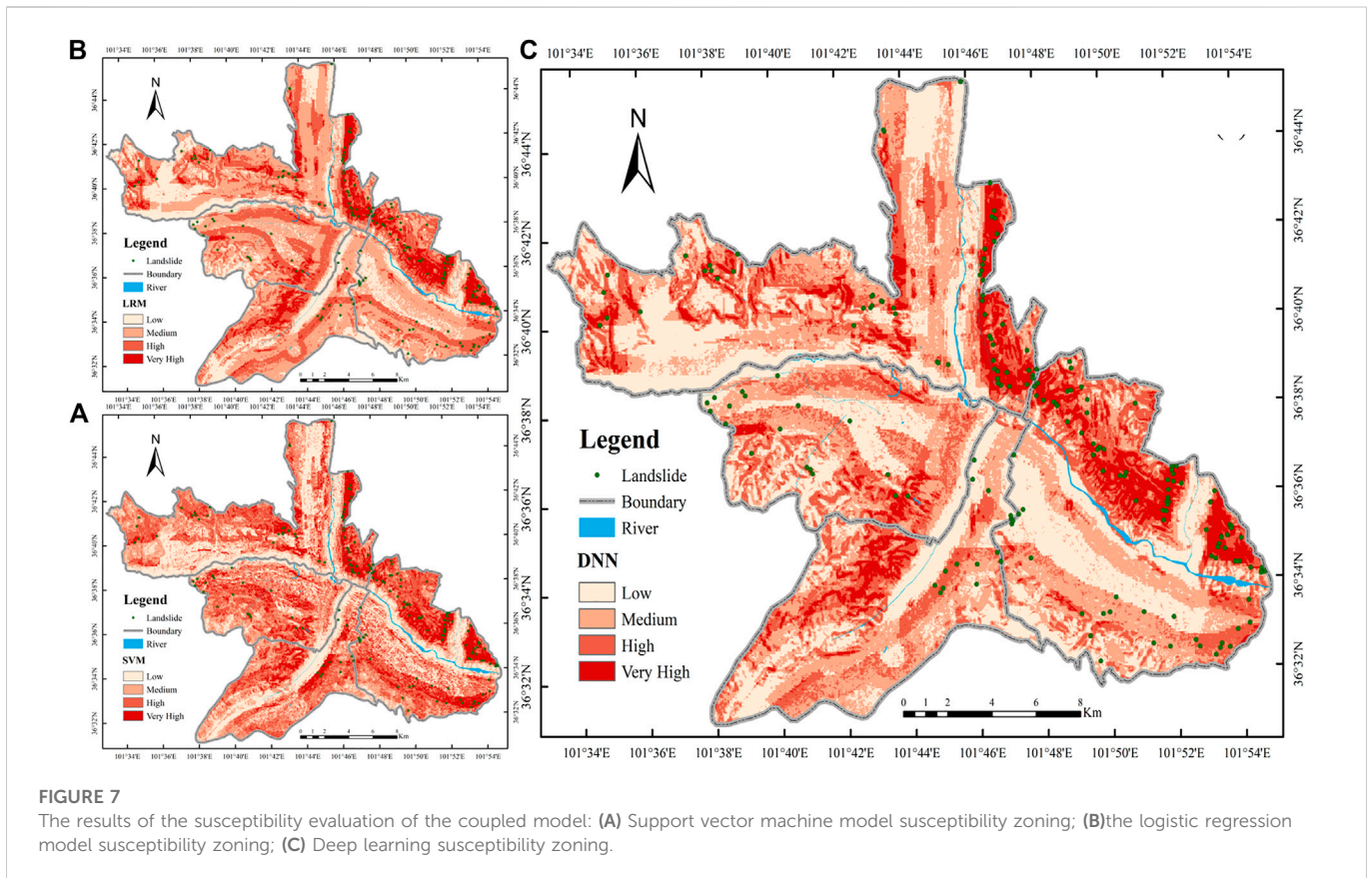


FIGURE 7 The results of the susceptibility evaluation of the coupled model: (A) Support vector machine model susceptibility zoning; (B) the logistic regression model susceptibility zoning; (C) Deep learning susceptibility zoning.

water system in the study area, and the distance from a river was divided into eight grades with 200 m intervals (Table 3). Combined with the chart, we can see that the areas most prone to landslides are mainly distributed in the range of 1,000–1,200 m. The area within this range accounts for 5.2% of the total area of the study area, including 13.59% of the landslide area. Based on the CF values, the 400–1,400 m area is prone to the occurrence of landslides.

4.1.7 Distance from a fault

The regional fault structure plays a controlling role in the development of joints and fissures in the geological bodies. Based on the geological map data, in this study, the Euclidean distance function in ArcGIS was used to perform fault analysis in the study area, and the distance from a fault was divided into eight grades with 1,000 m intervals (Table 1). Combined with the chart, we can see that the areas most prone to landslides are mainly distributed in the range of up to 1,000 m. The area within this range accounts for 25.02% of the total area of the study area, including 45.65% of the landslides. It can be seen that the fault factor has a great influence on landslide susceptibility. As the distance from a fault increased, the proportion of landslides and the CF value gradually decrease, indicating that the more fragmented the rock mass is, the more prone to landslides it is.

4.1.8 Distance from a road

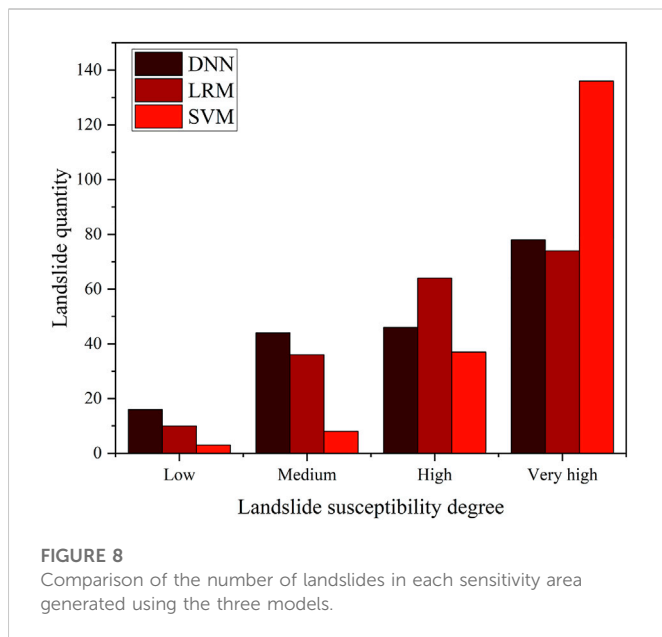
The large-scale cutting and excavation of roads in urban areas has changed the stress state of the cut slopes and aggravated the occurrence of landslides. Based on the geological map data, in this

study, the Euclidean distance function in ArcGIS was used to analyze the roads in the study area, and the distance from a road was divided into eight grades with 500 m intervals (Table 1). Combined with the chart, we can see that the areas most prone to landslides are mainly distributed in the range of 500–1,000 m. The area within this range accounts for 16.05% of the total area of the study area, including 40.76% of the landslide area. As the distance from a road increases, the landslide proportion and CF value gradually decrease, indicating that human activities such as road construction have a certain impact on the development of landslides.

4.2 Model accuracy evaluation and verification

After grading and quantifying each evaluation factor and using the certainty coefficient model to calculate the CF value of each grading index, the logistic regression model (LR), the support vector machine model (SVM), and the deep neural network model (DNN) were used for the coupled calculations. The evaluation results are presented in Figure 7. In the verification, 70% of the 368 landslide and non-landslide samples were used as training samples for the model calculation, and 30% were used as test samples. The confusion matrix was used to compare the results of the above three models, and the reliability was verified. The results are presented in Table 2.

It can be seen from the comprehensive results presented in Table 2 that the ACC, PPV, TPR, and TNR of the deep neural network are



greater than those of the other two models, so it is the optimal model among the three models. This study proves that the model accuracy of deep neural network in small samples is still improved. Of course, for small samples, the improved accuracy is smaller than that of large samples.

4.3 Sensitivity analysis

The principles of the three models were outlined in Section 3, and their algorithms can all be implemented using *Python*. First, the model was constructed using the training dataset, and then, the prediction ability of the trained model was tested using the validation dataset. In addition, the trained model was applied to calculate the landslide sensitivity index (LSI) for each raster layer in the study area. Theoretically, the LSI ranges from zero to 1, reflecting the probability of landslide occurrence. In this case, the LSI values calculated using the three models did not reach the breakpoint value. The LSI values calculated using the SVM model ranged from zero to 0.993, the LSI values calculated using the LR model ranged from zero to 0.595, and the LSI values calculated using the DNN model ranged from 0.146 to 0.853. To avoid subjectivity in obtaining the sensitivity partitions, the most widely used Jenks natural discontinuity method was used in this study to classify the landslide sensitivity categories as extremely high, high, medium, and low (Figure 7).

The results presented in Section 4.2 show that the DNN produced the optimal results for this study area, so the sensitivity analysis was carried out based on the evaluation results of this model. According to the landslide sensitivity zoning map obtained using the CF-DNN (Figure 7C), 32.1% of the study area had an extremely high and high landslide susceptibility, with 124 landslides in these areas. The extremely high and high susceptibility areas were mainly distributed in the higher elevation areas along the eastern edge of the Huangshui watershed, which is in good agreement with the landslide

distribution observed in the field. In addition, the areas with moderate susceptibility accounted for about 35.1% of the study area and included 25% of the landslide area. The low susceptibility area accounted for 32.6% of the study area and 8.7% of the landslide area. The low and medium sensitivity zones accounted for a large proportion of the area and were mainly distributed within the urban area of Xining City. Figure 8 presents a comparison of the model results.

5 Conclusion

In this study, based on remote sensing interpretation and field investigations, eight evaluation factors were selected: elevation, slope, land relief, aspect, curvature, distance from a river, distance from a fault, and distance from a road. According to the landslide susceptibility assessment factor classification map of these eight condition factors and the certainty factor results, the influence of each factor on the occurrence of landslides in the study area was analyzed. The results showed that four factors—elevation, curvature, distance from a fault, and distance from a road—had relatively significant influences on the landslide susceptibility in the study area.

The certainty coefficient method was used to couple the logistic regression, support vector machine, and deep neural network models, and the confusion matrix was used to evaluate the accuracy of these three models. The results indicated that the combined CF-DNN method was the most suitable for evaluating the landslide susceptibility in the study area.

According to the evaluation results obtained using the coupled CF and deep neural network model, the study area was divided into four landslide susceptibility zones: low (32.65%), medium (35.12%), high (22.44%), and extremely high (9.79%). The low and medium susceptibility areas accounted for the largest proportions and were distributed in the urban area of Xining; while the extremely high susceptibility areas were distributed in the high-elevation areas along the eastern edge of the Huangshui Basin.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

Writing—original draft preparation, WM; writing—review and editing, ZW; writing—review and editing, JD; resources, LP; supervision, QW; methodology, XW; visualization, YD; investigation, YW; All authors have read and agreed to the published version of the manuscript.

Funding

Supported by the Open Fund of Sichuan Provincial Engineering Research Center of City Solid Waste Energy and Building Materials

Conversion and Utilization Technology (GF2022YB005). Study on deformation evolution and stability criterion of Mining slope with Gentle stratified structure under complex conditions (41877273); Study on Unloading Law of Overlying Sandstone in Steeply inclined Coal Goaf under High and Steep Slope (2021YJ0053).

Conflict of interest

WM, ZW, and YW were employed by Qinghai 906 Engineering Survey and Design Institute Co Ltd.

References

- Abe, S. (2010). *Support vector machines for pattern classification*. London: Springer.
- Akgun, A. (2012). A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at izmir, Turkey. *Landslides* 9 (1), 93–106. doi:10.1007/s10346-011-0283-7
- Al-Harbi, K. M. A. S. (2001). Application of the AHP in project management. *Int. Proj. Manage* 19, 19–27. doi:10.1016/S0263-7863(99)00038-1
- Atkinson, P. M., and Massari, R. (1998). Generalized linear modeling of susceptibility to landsliding in the central Apennines, Italy. *Comput. Geosciences* 24 (4), 373–385. doi:10.1016/S0098-3004(97)00117-9
- Bai, C. N., Peng, L., and Shen, Y. (2021). Characteristics and mechanism of large landslide in Zhangjiawan. *Xining. Sci. Technol. Eng.* 21 (03), 927–934.
- Bai, S. B., Lu, P., and Wang, J. (2015). Landslide susceptibility assessment of the Youfang catchment using logistic regression. *J. Mt. Sci-engl* 12 (4), 816–827. doi:10.1007/s11629-014-3171-5
- Brabb, E. E. (1984). “Innovative approaches to landslide hazard and risk mapping,” in Proc Fourth International Symposium on landslide, Toronto, Canada, 23–31 August 1985 (Toronto: Canadian Geotechnical Society), 307–324.
- Brenning, A. (2005). Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Nat. Hazards. E. Sys.Sci* 5, 853–862. doi:10.5194/nhess-5-853-2005
- Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., and Revhaug, I. (2016). Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* 13 (2), 361–378. doi:10.1007/s10346-015-0557-6
- Chen, W., Xie, X., Peng, J., Shahabi, H., Hong, H., Bui, D. T., et al. (2018). GIS-based landslide susceptibility evaluation using a novel HybridIntegration approach of bivariate statistical based random forest method. *Catena* 164, 135–149. doi:10.1016/j.catena.2018.01.012
- Chen, W., Xie, X., Peng, J., Wang, J., Duan, Z., and Hong, H. (2017). GIS-Based landslide susceptibility modelling: A comparative assessment of kernel logistic regression, naïve-bayes tree, and alternating decision tree models. *Nat. Hazrisk* 8 (2), 950–973. doi:10.1080/19475705.2017.1289250
- Confuorto, P., Di Martire, D., Infante, D., Novellino, A., Papa, R., Calcaterra, D., et al. (2019). Monitoring of remedial works performance on landslideaffected areas through ground- and satellite-based techniques. *Catena* 178, 77–89. doi:10.1016/j.catena.2019.03.005
- Dai, L. X., Qiang, X., Xuanmei, F., Ming, C., Qin, Y., Fan, Y., et al. (2017). Preliminary study on spatial distribution and susceptibility evaluation of earthquake-induced geological disasters in Jiuzhaigou, Sichuan province on August 8, 2017. *Chin.J. Eng. Geol.* 25 (4), 1151–1164. doi:10.13544/j.cnki.jeg.2017.04.030
- Dou, J., Yunus, A. P., Tien Bui, D., Sahana, M., Chen, C. W., Zhu, Z., et al. (2019). Evaluating GIS-based multiple statistical models and data mining for earthquake and rainfall-induced landslide susceptibility using the LiDAR DEM. *Remote Sens.* 11 (6), 638. doi:10.3390/rs11060638
- Fan, W., Wei, X., Cao, Y., and Zheng, B. (2017). Landslide susceptibility assessment using the certainty factor and analytic hierarchy process. *J. Mt. Sci-engl* 14 (5), 906–925. doi:10.1007/s11629-016-4068-2
- Giovanni, L., Rocco, P., and Francis, C. (2016). Landslide susceptibility mapping using a fuzzy approach. *Procedia Eng.* 161, 380–387. doi:10.1016/j.proeng.2016.08.578
- Hong, H., Pradhan, B., Jebur, M. N., Bui, D. T., Xu, C., and Akgun, A. (2016). Spatial prediction of landslide hazard at the luxi area (China) using support vector machines. *Environ. Earth. Sci.* 75, 40. doi:10.1007/s12665-015-4866-9
- Hong, H., Shahabi, H., Shirzadi, A., Chen, W., Chapi, K., Ahmad, B. B., et al. (2019). Landslide susceptibility assessment at the wuning area, China: A comparison between multi-criteria decision making, bivariate statistical and machine learning methods. *Nat. Hazards.* 96 (1), 173–212. doi:10.1007/s11069-018-3536-0
- Le, L., Lin, Q., and Wang, Y. (2017). Landslide susceptibility mapping on a global scale using the method of logistic regression. *Nat. Hazards Earth Syst. Sci.* 17 (8), 1411–1424. doi:10.5194/nhess-17-1411-2017
- Li, Y. Y., Mei, H. B., and Ren, X. J. (2018). Geological disaster susceptibility evaluation based on certainty factor and support vector machine. *J. Geo. Information. Sci.* 20 (12), 1699–1709.
- Lin, W. T., Chou, W. C., and Lin, C. Y. (2008). Earthquake-induced landslide hazard and vegetation recovery assessment using remotely sensed data and a neural network based classifier: A case study in central taiwan. *Nat. Hazards.* 47 (3), 331–347. doi:10.1007/s11069-008-9222-x
- Liu, L. N., Xu, C., and Chen, J. (2014). Sensitivity analysis of landslide factors in 2013 Lushan earthquake based on CF method supported by GIS. *Chin.J. Eng. Geol.* 22 (6), 1176–1186.
- Luo, L. G., Pei, X. J., and Huang, R. Q. (2021). Evaluation of landslide susceptibility in Jiuzhaigou Scenic Area based on CF and Logistic regression model with GIS support. *J. Eng. Geol.* 29 (02), 526–535. doi:10.13544/j.cnki.jeg.2019-202
- Menard, S. (1995). *Applied logistic regression analysis*. Berlin: Sage University Paper Series on Quantitative.
- Nhu, V. H., Shirzadi, A., Shahabi, H., Singh, S. K., Al Ansari, N., Clague, J. J., et al. (2020). Shallow landslide susceptibility mapping: A comparison between logistic model tree, logistic regression, naïve bayes tree, artificial neural network, and support vector machine algorithms. *Int. J. Environ. Res. Public Health.* 17 (8), 2749. doi:10.3390/ijerph17082749
- Peng, L., Du, W. X., and Tian, H. (2021). *Sci. Technol. Eng.* 21 (18), 7806–7813.
- Pourghasemi, H. R., Moradi, H. R., Fatemi Aghda, S. M., Gokceoglu, C., and Pradhan, B. (2014). GIS-based landslide susceptibility mapping with probabilistic likelihood ratio and spatial multi-criteria evaluation models (North of Tehran, Iran). *Arab. J. Geosci.* 7 (5), 1857–1878. doi:10.1007/s12517-012-0825-x
- Pradhan, B., and Lee, S. (2010). Landslide susceptibility assessment and factor effect analysis: Backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environ. Model. Softw.* 25 (6), 747–759. doi:10.1016/j.envsoft.2009.10.016
- Rozos, D., Bathrellos, G. D., and Skilodimou, H. D. (2011). Comparison of the implementation of rock engineering system and analytic hierarchy process methods, upon landslide susceptibility mapping, using GIS: A case study from the eastern achaia county of peloponnesus, Greece. *Environ. Earth. Sci.* 63 (1), 49–63. doi:10.1007/s12665-010-0687-z
- Sahin, E. K., Colkesen, I., and Kavzoglu, T. (2018). A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto Int.* 1–23, 341–363. doi:10.1080/10106049.2018.1516248
- Shariati, M., Mafipour, M. S., Mehrabi, P., Bahadori, A., Zandi, Y., Salih, M. N. A., et al. (2019). Application of a hybrid artificial neural network-particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete. *Appl. Sci.* 9 (24), 5534. doi:10.3390/app9245534
- Shirzadi, A., Bui, D. T., Pham, B. T., Solaimani, K., Chapi, K., Kavian, A., et al. (2017). ShallowLandslide susceptibility assessment using a NovelHybrid intelligence approach. *Environ.E. Sci.* 76, 60. doi:10.1007/s12665-016-6374-y

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Shortliffe, E. H., and Buchanan, G. G. (1975). A model of inexact reasoning in medicine. *Math. Biosci.* 23 (3-4), 351–379. doi:10.1016/0025-5564(75)90047-4
- Vapnik, V. N. (1998). *Statistical learning theory*. Hoboken, New Jersey, United States: Wiley-Interscience.
- Wang, L. J., Guo, M., Sawada, K., Lin, J., and Zhang, J. (2016). A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosci. J.* 20 (1), 117–136. doi:10.1007/s12303-015-0026-1
- Wang, T., Wu, S., Shi, J., Xin, P., and Wu, L. (2018). Assessment of the effects of historical strong earthquakes on large-scale landslide groupings in the Wei river midstream. *Eng. Geol.* 235, 11–19. doi:10.1016/j.enggeo.2018.01.020
- Wang, Y., Fang, Z. C., and Hong, H. Y. (2019). Comparison of convolutional neural networks for landslide susceptibility mapping in yanshan county, China. *Sci. Total Environ.* 666, 975–993. doi:10.1016/j.scitotenv.2019.02.263
- Wang, Y. T., Seijmonsbergen, A. C., Bouten, W., and Chen, Q. (2015). Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data. *J. Mt. Sci-engl* 12 (2), 268–288. doi:10.1007/s11629-014-3134-x
- Xin, P., Wang, T., and Wu, S. R. (2015). Study on the formation mechanism of mudstone multistage rotaryLandslide in hanjiashan, datong county, xining, Qinghai province. *J. E. Sci.* 36 (06), 771–780.
- Xu, Y. Z., Lu, Y. N., and Li, D. Y. (2016). Evaluation of landslide susceptibility of granite distribution area in guangxi based on GIS and information quantity model. *Chin.J. Eng. Geol.* 24 (4), 693–703.
- Yao, S. H., Li, Z. M., and Zhang, J. Q. (2014). Study on the relationship between the beishan landslide in xining and the fault on the north bank of huangshui river. *Sci. Technol. Eng.* 14 (04), 161–163+169.
- Yao, X., Tham, L. G., and Dai, F. C. (2008). Landslide susceptibility mapping based on support vector machine: A case study on natural slopes of Hong Kong, China. *Geomorphology* 101, 572–582. doi:10.1016/j.geomorph.2008.02.011
- Zheng, Y. (2019). *Geological hazard prediction and evaluation based on multi-source data and deep learning*. China: Hefei University of Technology.