

Received February 20, 2022, accepted March 17, 2022, date of publication March 28, 2022, date of current version April 11, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3162597

Hazard Assessment of Debris Flow Based on Infinite Irrelevance Method and Probabilistic Neural Network Coupling Model

CUI WANG^{ID}, XUEDONG WANG^{ID}, AND YONGDONG ZHU^{ID}

College of Mining, Liaoning Technical University, Fuxin 123000, China

Corresponding author: Xuedong Wang (jilindaxue318@163.com)

This work was supported in part by the National Natural Science Fund of China under Grant 51604140, in part by the National Key Research and Development Program of China under Grant 2017YFC1503102, in part by the Foundation of Liaoning Province Education Administration under Grant LJ2020FWL006, and in part by the Discipline Innovation Team of Liaoning Technical University under Grant LNTU20TD-07.

ABSTRACT Debris flow is one of the most harmful natural disasters, which seriously damages the ecological environment balance, human life safety and property loss. Therefore, it is necessary to predict and assess the hazard of debris flow. This paper proposes an effective model for predicting the hazard of debris flow, which is a coupled model based on the Infinite Irrelevance Method (IIM) and the Probabilistic Neural Network (PNN). Taking Xiuyan Manchu Autonomous County, Liaoning Province, China as an example, this paper selects 31 key debris flow trenches according to the geological data survey and the characteristics of debris flow disasters, and uses IIM to screen the frequency, watershed area, height difference, debris reserves, rainfall, lithology, slope, population density, NDVI, a total of 9 influencing factors as the sample data set of PNN, and randomly divided into a training set and test set, the ratio is 65% and 35%. Compared with the IIM-GRNN model, the evaluation results show that the accuracy rate of the IIM-PNN model is 91%, which is more excellent than the IIM-GRNN model with an accuracy rate of 82%. The IIM-PNN model can more effectively classify and predict the hazard of debris flow, and can be used in other regions with similar geological environmental characteristics to take measures to manage and prevent the recurrence of the disaster.

INDEX TERMS Hazard assessment, debris flow, infinite irrelevance method, probabilistic neural network.

I. INTRODUCTION

Natural disasters have developed extensively around the world over the past few decades [1], [2] and caused significant losses to humans and the environment [3]. Among them, debris flow disasters have the characteristics of the high frequency of outbreaks and wide areas of occurrence. The combination and superposition of active geological structures, rainstorm conditions, and other factors have caused the wide-spread development of debris flow disasters, which seriously threatened the safety of human life and property, and also hindered the engineering and economic construction of mountainous areas to a certain extent [4], [5]. The diversity of visible factor selection and the complexity of the interaction of geological conditions make it difficult to

assess the hazard of debris flow [6]. Therefore, quantifying the hazard level of potential debris flow before disasters and deploying defenses accordingly is an important way to prevent and control debris flow disasters [7].

At present, the research of debris flow on a global scale has relatively systematic theories and methods, and accurate assessment of the hazard of debris flow is the basis for disaster prevention and management. Regarding the debris flow hazard assessment, the earlier relevant research is the Japanese scholars' judgment on the hazard of debris flow [8]. American geological engineers also put forward a framework for assessing the hazard of debris flow and used soil type, slope, and other factors to predict the debris flow [9]. To improve the accuracy of the determination of debris flow hazard, correlation analysis is used to select characteristic parameters to assess the debris flow hazard [10]. In addition, a GIS-based regional debris flow disaster mapping model [11],

The associate editor coordinating the review of this manuscript and approving it for publication was Szidónia Lefkovits^{ID}.

based on the statistical distribution of severity factors obtained from remote sensing [12], an artificial neural network [13], and the establishment of the frequency-amplitude relationship of debris flow [14] have been applied to the hazard assessment of various debris flow.

At the same time, relevant researches in China are also constantly improving [15]. The grey correlation analysis method is used in the hazard assessment of the single-ditch debris flow [16]. In the identification of the hazard factors of debris flow, grade assignment [17], formula assignment [18], and stepwise discriminant analysis are carried out [19]. Remote sensing interpretation, ground survey, and GIS technology [20], set pairs analysis and modified set pair analysis [21], neural network models such as BP [22], [23], and multiple coupling models [24] have also been well-applied in China's debris flow hazard assessment. General Regression Neural Network (GRNN) has strong nonlinear mapping ability and learning speed, good prediction when sample data is small, and the network can also handle unstable data. For the classification of small data samples, Probabilistic Neural Network(PNN) shows excellent classification prediction performance [25], which is mainly used for rapid training compared to traditional reverse propagation neural network methods [26]. The PNN model also combines Analytic Hierarchy Process [27], remote sensing data analysis [28], [29], and other methods to obtain the relationship between variables and weights, and get higher accuracy. Present research on neural networks has good applications in oil and gas forecasting, traffic intelligence forecasting.

Despite the continuous increase in assessment methods, the current assessment system of debris flow hazard still needs further discussion [30]. Among them, a reasonable choice of influencing factors is essential. For example, BP neural network is an effective method of forecasting complex nonlinear systems [31], but in the assessment of debris flow hazard, when the sample data is reduced, the hazard information cannot be extracted well [32]. In addition, when there is too much overlapping information on the debris flow hazard, it will cause part of the information to be repeatedly enlarged, which will also reduce the effectiveness of the assessment. The current methods used for factor screening are mainly empirical models such as Analytic Hierarchy Process (AHP). For AHP, the data statistics are large, the weights are difficult to determine, and they are susceptible to human subjective influence, so they usually need to be coupled with other models. Of course, some machine learning methods can also be used for the selection of factors, such as Support Vector Machine (SVM), but the classical SVM model is a binary classification model, which needs to be improved if it is to be applicable to the assessment under multiple factors. Therefore, according to the non-linear relationship between primary and secondary factors, the use of multiple correlation coefficients to select secondary factors [33] provides a new direction for the screening of debris flow hazard factors. Infinite Irrelevance Method (IIM) can preserve the physical

meaning of the feature by calculating the correlation degree of the feature [34]. By comparing the complex correlation coefficients, some features of smaller complex correlation coefficients can be selected from all the features, and the selection of factors can be made more accurately. This method is commonly used for the study of model metric performance.

To assess the hazard of debris flow more effectively, In this study, IIM is applied to the screening of influencing factors, so as to determine a complete debris flow hazard assessment factor system, avoid mutual interference between multiple information, and provide a basis for subsequent successful prediction classification. On this basis, the IIM-PNN integrated assessment model is constructed to predict and classify the debris flow hazard in the study area and determine the distribution of hazard areas. Eventually, by comparing the prediction results and accuracy of different models, a new model to improve the accuracy of debris flow hazard assessment is explored, which can provide targeted suggestions for disaster prevention and control in the study area to avoid unnecessary waste.

In this paper, Xiuyan Manchu Autonomous County, Liaoning Province, China is selected as an example. The newly established IIM-PNN integrated models are applied to the hazard assessment of debris flow. The main objectives are: (1) to screen the influencing factors of debris flow hazard in Xiuyan County, clarify the main assessment factors, and classify the hazard prediction in the study area; (2) to compare the prediction results of different models, the prediction performance, determine the optimal model, and achieve a reasonable prediction of debris flow hazard in the study area. The research results provide reference values for the predictive classification of debris flow hazards, prevention and control measures, and further related assessments.

II. RELATED WORK

A. STUDY AREA

Xiuyan Manchu Autonomous County belongs to Anshan City, Liaoning Province, China (Fig 1). The geographical coordinate is E122°52' ~123°46', N40°00' ~40°39', with a total area of 4502km². The county is 75.5 km wide from east to west and 91.8 km long from north to south. Because the study area has the basic conditions for the development of debris flow, it is easy to cause serious disasters, so 31 key debris flow in the study area was investigated and studied.

1) GEOMORPHOLOGY AND TOPOGRAPHY

The study area has a higher terrain in the northeast and a gentle slope from Qianshan to Bohai in the west. The low-lying area accounts for 78% of the total area. The landform in the territory is relatively complex, forming a hilly and low mountain landscape. The moderate terrain area has the smallest area, accounting for only 10.8%, but the debris flow distribution density is the largest. As the terrain becomes slower, the debris flow distribution density

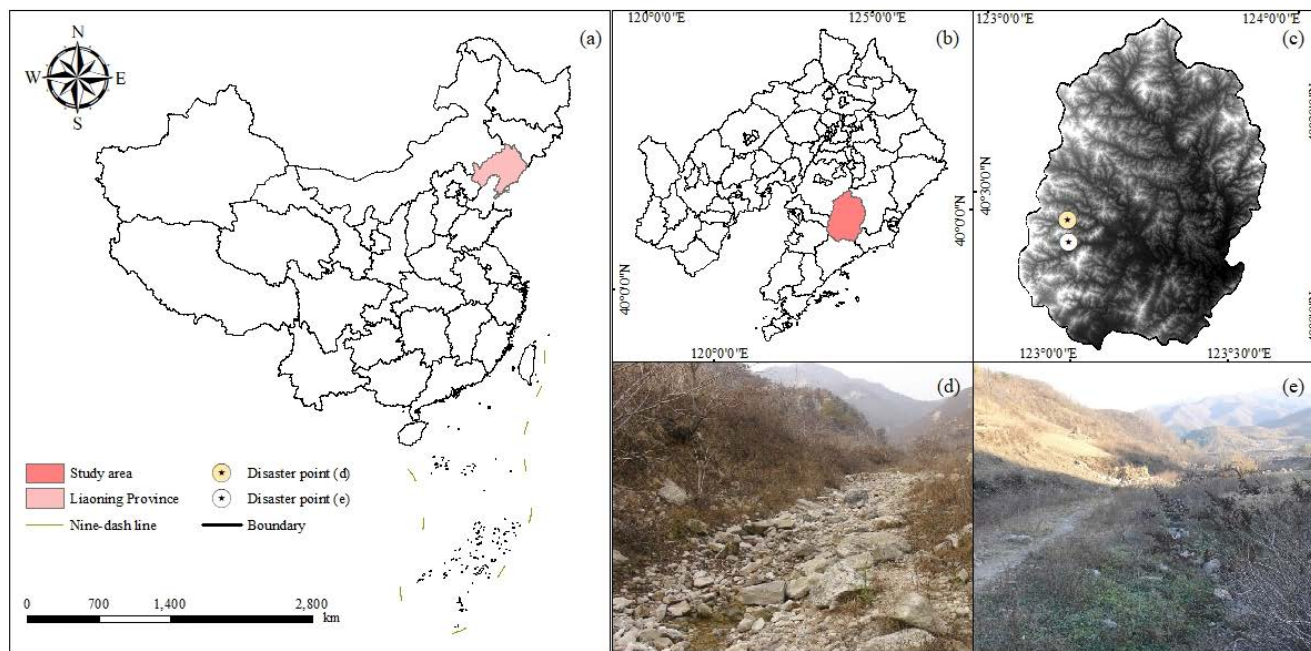


FIGURE 1. Debris flow hazard assessment study area: (a) Map of China; (b) Map of Liaoning Province; (c) Study area; (d), (e) Two debris flow disaster points located in the west of Study area.

gradually decreases (Fig 2a). The larger the terrain elevation difference is, the more likely it is to provide sufficient dynamic conditions for the occurrence of debris flow.

2) METEOROLOGICAL AND HYDROLOGICAL CHARACTERISTICS

The study area is located in a temperate humid monsoon climate zone with an average annual temperature of 7.5°. July has the highest temperature, with an average temperature of 23.0 °C and a maximum temperature of 37.3 °C; the lowest temperature of the year occurs in January, with an average temperature of -10.7 °C. The large annual and daily temperature difference in Xiuyan County promotes the thermal expansion and contraction of rocks, accelerates the weathering of rocks, and provides provenance conditions for the development of debris flow. According to statistics, the average rainfall in this area in the past 30 years has been 763-929 mm, and the distribution of precipitation has shown a decreasing trend from the northeast to the southwest of the study area (Fig 2b).

3) LITHOLOGY

The exposed strata in the study area include the Lower Proterozoic Sinian Liaohe Group (Pth1), Upper Proterozoic Qingbaikou (Qn) strata, Carboniferous-Permian (CP) strata, Upper Jurassic (J1) strata, and the Fourth Series (Q) deposits. According to the types and characteristics of engineering geology of rock and soil, and the engineering mechanical properties of the original rock and the degree of weathering resistance, the exposed rock formations of Xiuyan County are divided into massive hard metamorphic rock intrusive

rock groups (I), fractured soft rock Strongly weathered metamorphic rock intrusive rock group (II), hard carbonate rock group (III), soft and hard interphase clastic rock group (IV) and loose soil (V). Category II is the most widely distributed in the study area (Fig 2c), mainly composed of the pre-seismic Liaohe group metamorphosis rock and Triassic intrusive rock, the surface of this group of strong weathering, loose debris accumulation more, for the development of debris flow to provide material conditions.

4) VEGETATION CONDITIONS

The northern temperate deciduous and broad-leaved forest area of Xiuyan County has a large forest area and superior natural conditions. In recent years, due to the increasing intensity of human engineering activities, the virgin forest in the county has been destroyed. Therefore, vegetation conditions also affect the formation and development of debris flow in the study area. The areas with a lower Normalized Difference Vegetation Index (NDVI) in the area have the highest disaster density (Fig 2d), indicating that places with poorer vegetation conditions are relatively more likely to occur debris flow.

B. INFINITE IRRELEVANCE METHOD (IIM)

In the process of debris flow hazard assessment, it is often affected by different factors. Therefore, the choice of factors becomes very important [35], [36]. IIM is the correlation analysis method used in this paper, which uses the multiple correlation coefficients between factors to select some factors for a comprehensive assessment.

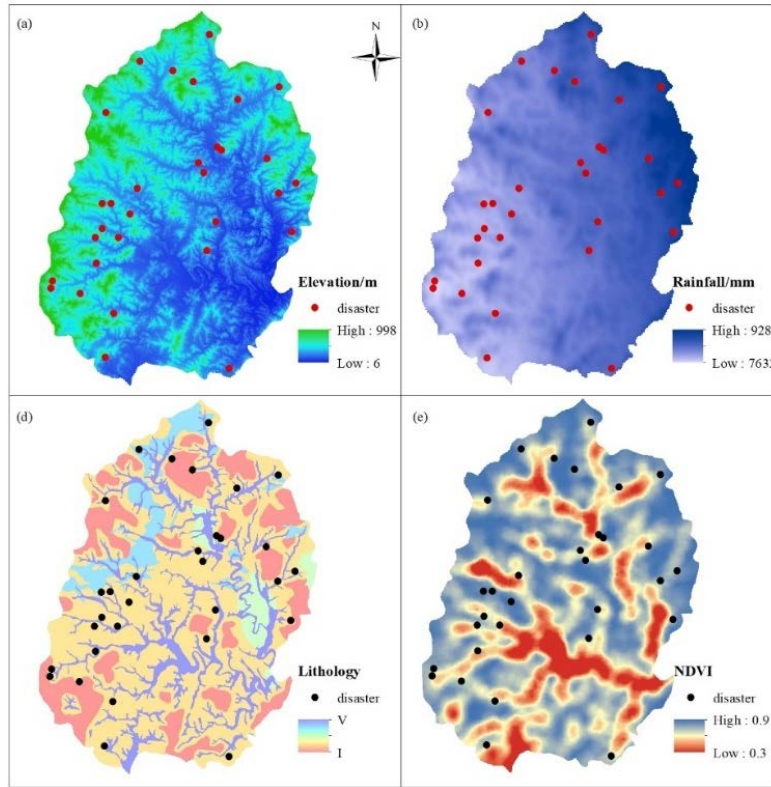


FIGURE 2. Debris flow distribution characteristics: (a) Elevation distribution; (b) Annual average rainfall distribution; (c) Lithological distribution; (d) Vegetation distribution.

According to the determined p factors, the value of each node is calculated separately, and the infinite irrelevant group is obtained through IIM [37]. The solution steps are as follows:

1) DETERMINING ANALYTICAL SAMPLES

Establish the sample matrix. It is assumed that there are n features which can be written as

$$X = (X_1, X_2, \dots, X_n) \tag{1}$$

where $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,p})$, $i = 1, 2, \dots, n$, p is the number of rows in X .

2) SOLVING THE CORRELATION MATRIX OF THE ANALYSIS SAMPLE

Calculate covariance matrix (S) and correlation coefficient matrix (R). Each of the elements in the covariance matrix is the covariance among the random vector X of different components. The covariance is the second order statistical property of the variables. If the correlation between the random vector of different components is low, the result of covariance matrix is almost a diagonal matrix. For some special applications, to shorten the length of the random vector, the principal component analysis method is introduced to guarantee the transformation of the variables in covariance matrix is a completely diagonal matrix. The correlation coefficient matrix is composed of the correlation

coefficients among the columns of matrix. That means, the elements of the i -th row of the correlation matrix are the correlation coefficients between the i -th and j -th column in the original matrix.

$$S_{p \times p} = (S_{ij}) \tag{2}$$

where the variance, the covariance and the mean value can be calculated by formula(3)-(5) respectively.

$$S_{ii} = \frac{1}{n} \sum_{a=1}^n (x_{ai} - \bar{x}_i)^2, \quad i = 1, 2, \dots, p \tag{3}$$

$$S_{ij} = \frac{1}{n} \sum_{a=1}^n (x_{ai} - \bar{x}_i)(x_{aj} - \bar{x}_j), \quad i \neq j; i, j = 1, 2, \dots, p \tag{4}$$

$$\bar{x}_i = \frac{1}{n} \sum_{a=1}^n x_{ai}, \quad i = 1, 2, \dots, p \tag{5}$$

Moreover, the correlation matrix R in formula(6)-(7) can be calculated by

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \tag{6}$$

$$r_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}}, \quad i, j = 1, 2, \dots, p \tag{7}$$

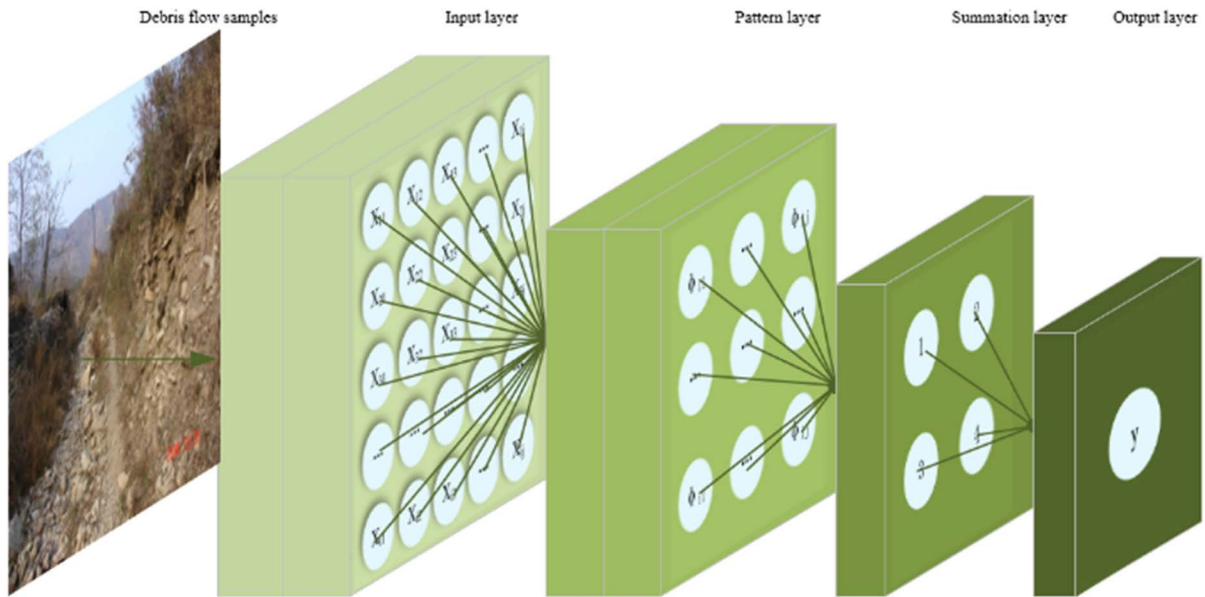


FIGURE 3. Probabilistic neural network structure diagram.

3) SOLVING THE CORRELATION MATRIX OF THE ANALYSIS SAMPLE

The correlation coefficient is a statistical index introduced to reflect the degree of correlation among the variables. The correlation coefficient is calculated by the difference product method. Based on the difference between these two variables and their mean values, the two coefficients are multiplied according to the correlation of two degrees among variables. The complex correlation coefficient is written as

$$\rho_i = \sqrt{1 - (r_{i1}^2) (1 - r_{i2,1}^2) (1 - r_{i3,12}^2) \cdots (1 - r_{ip,12 \dots (p-1)}^2)} \tag{8}$$

From the above three steps, the complex correlation coefficient \$\rho_i\$ is obtained. If the multiple correlation coefficient is larger, it indicates that \$\rho_i\$ is more easily replaced by \$\rho_1, \rho_2, \dots, \rho_n\$, and the effect of the index \$X_i\$ on the assessment target is weaker. After the threshold value \$D\$ is specified, its calculation formula is

$$D = 0.95 \max(\rho_1, \rho_2, \dots, \rho_n) \tag{9}$$

when \$\rho_i > D\$, factor \$X_i\$ can be deleted, and the remaining factors have relatively small multiple correlation coefficients, which can well reflect the original assessment index system [38].

C. PROBABILISTIC NEURAL NETWORK (PNN)

PNN is composed of radial basis neurons and competing neurons and is often used as a tool to classify different populations of work [39]. The PNN model proposed by Specht is a neural architecture implemented using a PDF estimator [40]. This method is usually used for classification problems. It is a non-parametric estimation method based

on Bayesian optimal classification decision theory and probability density function in statistics [41]. PNN is a forward propagation algorithm without feedback (Fig 3), which is characterized by interactive operations between data and more targeted acquisition of cross information about features.

The PNN structure consists of an input layer, a hidden layer, a summation layer, and an output layer in sequence. After the input layer passes the data onto the hidden layer, the distance between the input vector and the center is calculated. The vector \$X\$ is used as the input of the hidden layer. The input-output relationship is defined as follows:

$$\Phi_{ij}(X) = \frac{1}{(2\pi)^{\frac{1}{2}}} e^{-\frac{(X-X_{ij})(X-X_{ij})^T}{\sigma^2}} \tag{10}$$

In the formula, \$m\$ is the total number of training samples, and \$X_{ij}\$ is the \$j\$-th attribute of the \$i\$-th sample.

In the summation layer, the number of neurons is the same as the number of input training samples, and the summation layer uses the output of hidden neurons belonging to the same category in the hidden layer as a weighted average:

$$v_i = \frac{\sum_{j=1}^L \Phi_{ij}}{L} \tag{11}$$

where \$V_i\$ represents the output of the \$i\$-th category, \$L\$ represents the number of neurons in the \$i\$-th category, and the largest in the summation layer is taken as the output category in the output layer:

$$y = \operatorname{argmax}(v_i) \tag{12}$$

In the actual calculation, the vector of the input layer is multiplied by the weighting coefficient, and then input into

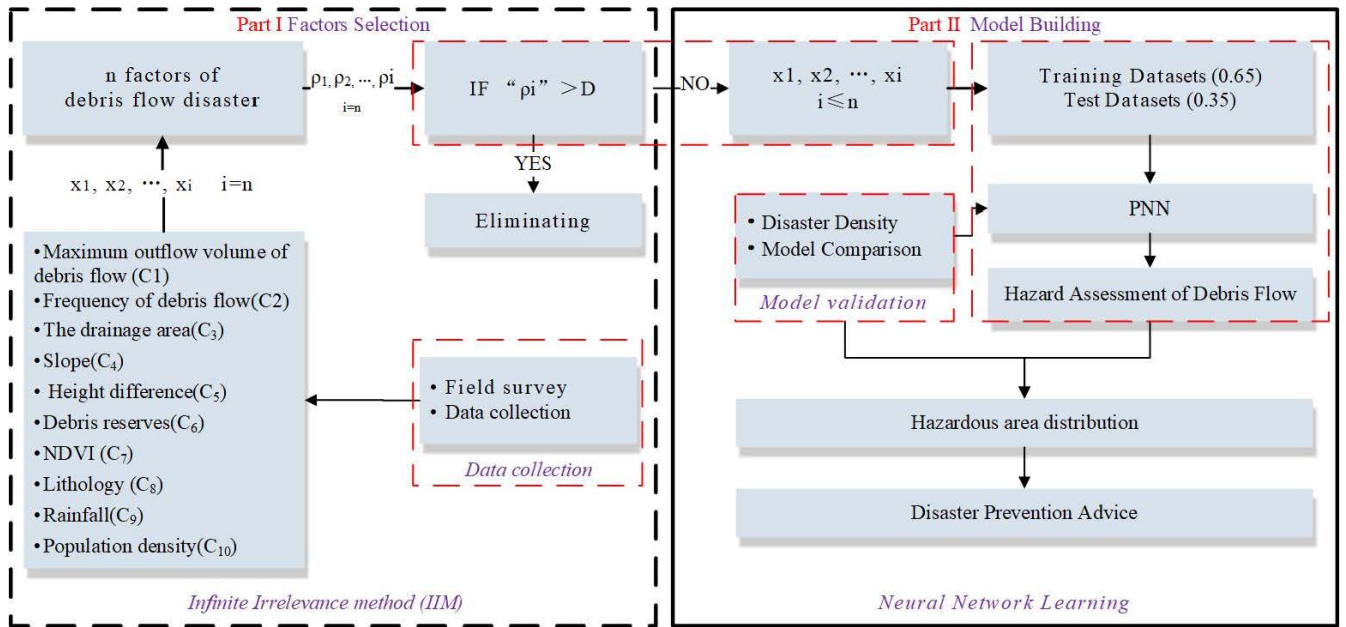


FIGURE 4. Flow chart of debris flow hazard assessment based on IIM-PNN model.

the radial basis function of calculation:

$$Z_i = X\omega_i \tag{13}$$

Assuming that both X and ω have been standardized to a unit length, then the result is subjected to the radial basis operation $exp((Z_i - 1)/\sigma^2)$, which is equivalent to the following formula:

$$exp\left[-\frac{(\omega_i - x)^T(\omega_i - x)}{2\sigma^2}\right] \tag{14}$$

In the formula, σ is the smoothing factor, which plays a vital role in network performance.

D. COMPREHENSIVE ASSESSMENT SYSTEM BASED ON IIM AND PNN

In the hazard assessment of debris flow, the complexity of the factors often affects the accuracy of the assessment, more metrics don't mean better accuracy. Therefore, the correlation between the factors of the assessment system is required to be as small as possible, otherwise, the information provided by the factors overlaps, and the analysis result is easily distorted. This study introduces the IIM to screen the factors. The selection criterion is based on the calculation of the critical value of the multiple correlation coefficients between the factors, which can better represent the mutual influence ability between the factors.

The PNN model is responsible for the supervised learning of training data and obtains probability estimates of various categories. In this model, the cross features of factors are better expressed, thereby improving the classification effect [42].

Based on the ground investigation and sorting into debris flow data in the study area, this study couples IIM and PNN to

establish a new comprehensive assessment system for debris flow. The assessment process is shown in Fig 4. Based on the survey data, the factors that affect the hazard of debris flow in the study area are analyzed and preliminary selected; secondly, IIM is used to calculate the multiple correlation coefficients between the factors, and the more relevant factors are eliminated; the selected factors are input into PNN for classification forecast; finally use the results of the IIM-PNN model to assess the hazard of debris flow, and compare its accuracy with the IIM-GRNN model.

When the data matrix is arbitrarily input in the model network, supervised learning will be generated according to the input signal. Before the model is established, by adjusting the sett of the radial basis function in the model, the learning ability of the model can be relatively improved, thereby further improving the prediction accuracy of the IIM-PNN model. Finally, the results of the model classification and prediction are used to assess the hazard of the study area. The hazard of a debris flow can generally be divided into four levels: slight hazard (I'), low hazard (II'), medium hazard (III'), and high hazard (IV').

III. RESULTS AND DISCUSSION

A. CHOICE OF INFLUENCING FACTORS

Based on the eruption records and the investigation into the debris flow in the study area, which collected relevant eruption data onto 31 key debris flow trenches for assessment. Regarding the outbreak of debris flow, there are many influencing factors [43]. For example, terrain conditions are mainly dynamic conditions for debris flow outbreaks, which provide sufficient material sources and water collection conditions for debris flow outbreaks [44]. Hydrogeological conditions have an important influence on the formation

TABLE 1. Classification of debris flow factors.

Conditions	Factors
Topography	The drainage area(C ₃), Slope(C ₄), Height difference(C ₅), Debris reserves(C ₆) and NDVI (C ₇)
Geology	Lithology (C ₈)
Precipitation	Rainfall(C ₉)
Human activities	Population density(C ₁₀)

of debris flow; the biggest influence on torrential rain-type debris flow is rainfall conditions [45].

The hazard assessment of the debris flow in China is represented by “Single Gully debris flow Hazard Calculation” [46]. The factors’ system for the hazard of debris flow is based on “Maximum outflow volume of debris flow (C1) and Frequency of debris flow (C2)” [47]. C1 and C2 represent the size and frequency of debris flow. Under the same frequency of debris flow, the larger the scale of debris flow, the greater the hazard; under the same scale of debris flow, the higher the frequency of debris flow, the greater the hazard. Therefore, theoretically speaking, the hazard degree of debris flow can be expressed as the definite integral of the scale of the debris flow-frequency curve. The frequency and scale of debris flow are determined by the topographic and geomorphological conditions, geological conditions, precipitation conditions, and human activity conditions in the basin. Therefore, combined with the results of “Geological Hazard Investigation and Zoning in Xiuyan Manchu Autonomous County, Liaoning Province, China”, the remaining 8 evaluation indicators were also selected (Table 1).

As shown in the indicators in Table 1, C3 not only affects the collection of rainfall in the watershed, but also directly determines the scale of sand production. the larger the C3, the more rainwater collection, the wider the range of sources provided, and the greater the mass of loose solids involved in debris flow movement under the excitation of floods. The mass of loose solids, in turn, affects the maximum flushing out volume of a debris flow and is closely related to the debris flow activity. In particular, the size of runoff and flood peaks within the debris flow outbreak zone is directly related to the initiation of loose material in the zone and the ability to participate in debris flow activity; For debris flow, C4 is one of the important topographic conditions for debris flow formation, with the increase of C4 watershed erosion mode from surface erosion and gully erosion to landslide and collapse, the formation capacity of loose material in the watershed gradually increases, and the intensity of watershed erosion gradually increases; C5 reflects the amount and stability of the potential energy of debris fluid converted into kinetic energy to a certain extent, the area with relatively high C5 is about favorable for the conversion of potential energy into kinetic energy;

The presence of sufficient C6 in the debris flow channel is one of the necessary conditions for whether a single trench is a debris flow trench, and is important for the discernment of debris flow trench. When there is an appropriate amount of C6 in the channel as the source of debris flow and the rainfall reaches a certain value, debris flow will break out; Vegetation can enhance land stability, promote groundwater infiltration, and play a role in inhibiting soil erosion. C7 refers to the Normalized Difference of Vegetation Index (NDVI), which is a factor of vegetation cover with values distributed in $[-1, 1]$, with negative values representing features with high reflection of visible light such as clouds, water, and snow, and positive values reflect the high vegetation cover; C8 is based on the rock and soil structure. Type indicates that the rock mechanical parameters under a different rock and soil structure types affect the shear strength of the soil and control its stability; Abundant water is a major cause of debris flow hazards, and C9 represents the average annual rainfall in the study area, which provides sufficient hydrodynamic conditions for debris flow formation and eruption; C10 reflects the intensity of unreasonable human engineering and economic activities, affects the natural environment, and also causes slope instability, and this leads to the occurrence of debris flow.

C3, C5, C6, C8, and C10 are derived from the measurement and calculation results of measuring tools such as GPS locators and laser rangefinders in the study area; C4 is derived from the Digital Elevation Model (DEM) with a spatial resolution of 30 m on the Geospatial Data Cloud website Data analysis and calculation [48], [49]; C7 is derived from satellite remote sensing data of the Geographical National Conditions Monitoring Cloud Platform; C9 is derived from the statistics of annual average rainfall raster data in the past 30 years. To improve the accuracy of interpretation results, the lithological indicators in the debris flow ditch in the study area were quantitatively allocated, and the average value of the remaining indicators was taken as the sample data onto the study area. The quantitative results are shown in Table 2. Among them, lithology index I-V was quantified as >0.9 (V), $0.9-0.6$ (IV), $0.6-0.3$ (III), $0.3-0.1$ (II), and <0.1 (I).

B. FACTOR OPTIMIZATION BASED ON IIM

The accuracy of the model is largely affected by the original influencing factors, and also strongly depends on

TABLE 2. The quantitative results of early factors.

Num	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Hazard
1	1456.6	1	7.6	34	552.2	265.6	0.636	0.428	94	26.3	I'
2	2265.9	5	12.1	28.9	531.1	352.1	0.731	0.385	96.3	57.9	I'
3	2096.3	3	3.1	33.1	534	267.2	0.635	0.615	84.7	14.5	I'
4	164.8	3	3.9	39.5	584.6	183.3	0.948	0.362	91.8	71.8	I'
5	1992.5	2	17.4	34.9	404	334.4	0.521	0.33	120.4	48.9	I'
6	2726.1	2	10.6	35	564.7	368	0.861	0.488	105.9	56.6	II'
7	5460.9	3	10.8	37.5	565.6	408.1	0.814	0.651	99.7	29.6	II'
8	1031.9	1	4.8	30.1	363.5	294.6	0.61	0.583	89.4	79.2	II'
9	892.6	3	13.6	23.1	572.4	339.5	0.978	0.614	103	51.5	II'
10	1441	7	9.7	33.6	498.7	299.3	0.768	0.579	108.1	61.9	II'
11	1952.9	3	5.6	28.4	383.1	564.5	0	0.735	87.6	115.9	III'
12	722.2	2	18.9	32.9	695.3	269.6	0.505	0.339	139.6	42.3	III'
13	2127.9	3	17.3	26.4	465.5	431.9	0.752	0.596	138.5	75.1	III'
14	1783.2	3	9.2	28.9	378.7	516.9	0.798	0.656	117.4	65.2	III'
15	1480.5	5	20.8	33	544.3	411.4	0.739	0.438	114.9	108.7	III'
16	2338.4	4	12.5	33.8	539.5	453.9	0.806	0.326	121.5	52	III'
17	2503.7	4	8.1	33.5	412.7	411	0.822	0.526	110.1	123.5	III'
18	996.3	1	6	21.3	272.6	427.1	0.758	0.7	110.7	76.7	III'
19	1982.5	8	8.4	27.2	568.7	453.9	0.013	0.709	84.9	125	IV'
20	770.8	3	25.4	34.2	716.6	373.6	1	0.465	137.3	52	IV'
21	3143.1	4	14.7	30.4	714	524.6	0.534	0.475	118.3	61.2	IV'
22	1046.7	3	28.3	33.1	687.3	345.3	0.916	0.576	134.5	32.5	IV'
23	777.8	10	19.8	37.1	701.5	409.1	0.754	0.708	113.2	75.8	IV'
24	1141.4	8	11.9	25.2	765.4	377.7	0.907	0.631	130	59.7	IV'
25	554.5	1	35.2	24.2	581.7	493	0.847	0.527	159	73.9	IV'
26	995.4	4	24.1	40	713.2	253.7	0.324	0.489	137	29	IV'
27	439.2	7	19.7	30.2	547.2	462	0.464	0.459	114.9	106.6	IV'
28	1409.6	4	20.9	33.9	696.7	421.4	0.752	0.362	140	76.6	IV'
29	929.5	3	21.9	33	744.1	454	0.187	0.714	115.2	63.3	IV'
30	380.1	6	20.8	22.9	790.5	478.2	0.059	0.614	117.3	62.5	IV'
31	243.3	4	16.7	26.5	571	448	0.756	0.424	125.1	89.8	IV'

the correlation between factors. According to the IIM, the main factors were screened using formulas (1)-(8), and the multiple correlation coefficients ρ of the initial 10 factors were calculated and obtained, which were 1, 0.626, 0.920, 0.671, 0.756, 0.611, 0.926, 0.799, 0.758, 0.822. Substituting the maximum value 1 in the complex correlation coefficient into equation (4), the critical value $D = 0.95$ can be calculated, where $\rho_{C1} > 0.95$, so this index is eliminated.

In addition, the correlation matrix of step 2 in IIM can also be used to correlate the influencing factors and the hazard degree. The bubble diagram (Fig 5) not only reflects the correlation between the influencing factors, but also helps to initially determine the dominant factors of debris flow outbreak in the study area. When the bubbles are larger and darker, the correlation is greater, with red indicating positive correlation and purple indicating negative correlation. As shown in the figure, the correlation coefficients of C3 and C9 are the largest (0.82), and the rainfall conditions have a great influence on the formation of the whole watershed scale, reflecting the characteristics of interaction between factors; the influencing factors with high correlation with the hazard degree are C3 (0.61), C9 (0.59)

and C6 (0.58) at one time, and the remaining influencing factors are relatively low. These three influencing factors of preliminary judgment precisely reflect the channel, water source conditions and physical source conditions of debris flow formation, which are consistent with the actual situation, indicating that the selection of influencing factors is relatively reasonable.

C. ASSESSMENT MODEL BASED ON IIM-PNN

The factors' system optimized by IIM is more effective and provides a more reliable calculation basis of the PNN model. In this study, the data set was randomly divided into a training set (0.65) and test set (0.35), and the IIM-PNN model with the optimal input parameters was used to classify and predict the hazards of debris flow, and IIM-GRNN model of the method was used as a control, and the accuracy of two neural network tests of the model was compared (Fig 6).

As shown in Fig 6, there are 11 samples used as the test set, and the hazard degree are I'(Slight), II'(Low), III'(Medium) and IV'(High). The test results are Medium and High, which is due to the fact that the training and test sets are randomly divided, which tends to lead to a relative concentration of samples at a certain level. The accuracy

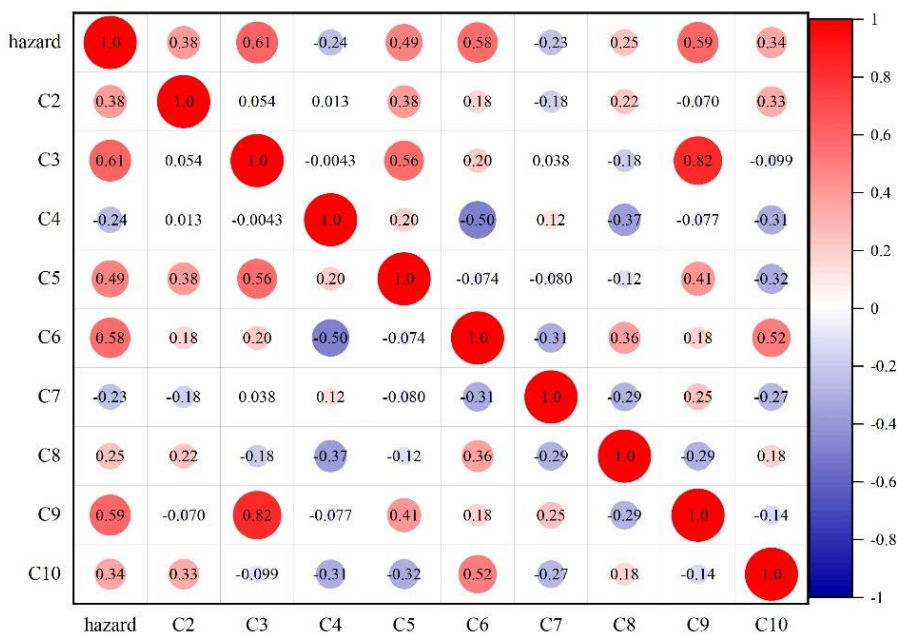


FIGURE 5. The influencing factors are related to the hazard of debris flow.

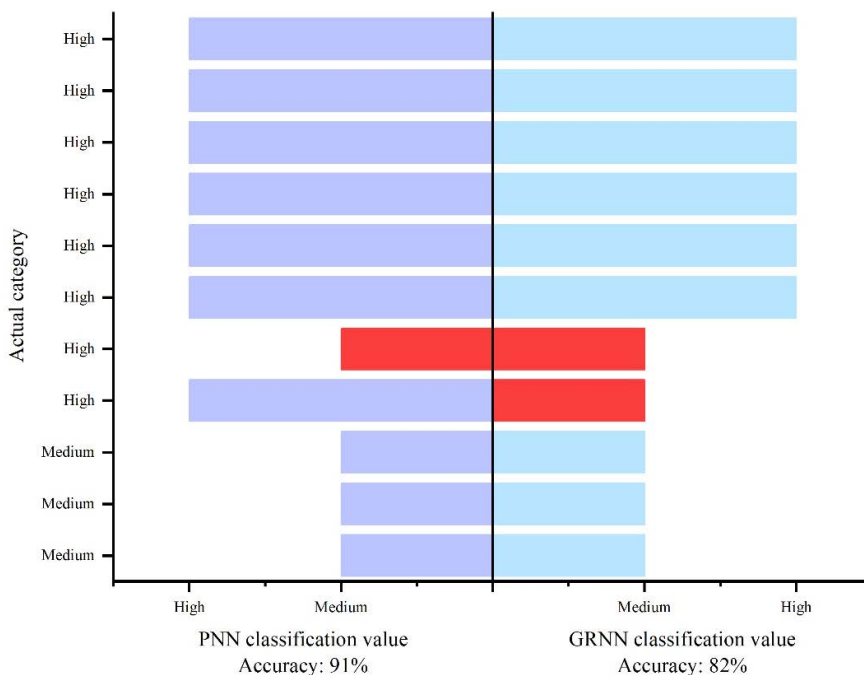


FIGURE 6. Model accuracy comparison chart.

rate under PNN classification (91%) is higher than that under GRNN classification (82%), which indicates that the whole IIM-PNN comprehensive assessment model is more advantageous.

D. COMPREHENSIVE HAZARD ASSESSMENT OF DEBRIS FLOW

Among the 31 debris flow samples of this study, 5 were slight-hazard, 5 were low-hazard, 8 were medium-hazard,

and 13 were high-hazard. The debris flow density of each hazard level is shown in Figure 7. The debris flow disasters in the study area are more harmful, and the high hazard debris flow density is 0.42, which is the key target for the prevention and control of debris flow disasters in this study area. The resulting density plots also demonstrate that the density is greater for medium and high hazards, and so are the predicted results, indicating that the predictions of the two models mentioned above are reasonable.

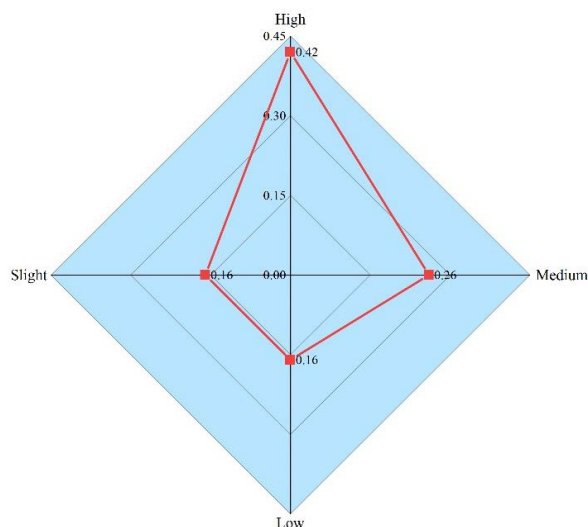


FIGURE 7. Debris flow density map of each grade.

Combining with Figure 2 above, it can be seen that debris flow is mainly distributed among the northeast and west of the study area. Among them, the hazards are mainly distributed in places with large elevation changes, increased average annual rainfall, complex lithological changes and low vegetation cover. High-hazard debris flow are often distributed in places with large watershed areas, large topographic elevation differences, broken rock structures, and high rainfall, and this feature is more obvious in the northeastern part of the study area; medium-hazard debris flow are more active and often occur in places with more debris storage, less vegetation cover, and large elevation differences, and the western part of the study area conforms to this feature; low-hazard debris flow often occur in places with small watershed areas and well-recovered vegetation, and the study area The southern part of the study area, where only isolated debris flow hazards occur, fits this characteristic; the remaining areas have relatively low likelihood of debris flow, and if they do occur, they are relatively minor.

To sum up, this study proposes that the IIM-PNN model is feasible for the hazard assessment of debris flow. The network training is easy, the convergence speed is fast, and it has strong fault tolerance. It does not need to be repeated for a long time when increasing or decreasing the category mode. Learn. The IIM-PNN model can provide reasonable assistance and prevent unnecessary debris flow disaster losses, such as strengthening civil engineering and biological engineering, strengthening dynamic monitoring during the rainy season, and taking avoidance measures for retail residents threatened by debris flow. However, due to the limitations of data acquisition and selection of influencing factors in different regions, it is difficult to form a relatively unified assessment model. Therefore, how to extract the information of each factor more accurately is the focus of future research work and a difficult problem that needs to be broken in the field of debris flow hazard assessment.

IV. CONCLUSION

In this paper, a debris flow hazard assessment method based on the IIM-PNN model is proposed. The model training and actual testing results show that this method has the following advantages:

1) The IIM-PNN model comprehensively considers factors such as topography, geology, precipitation, and human activities, and effectively screens a variety of indicators to analyze the importance of different factors of the hazard of debris flow, which helps to improve the accuracy of training results.

2) The IIM-PNN model does not need to set the factor weights in advance, and the classification results can be obtained directly. The accuracy rate is 91%, and the model is stable. The debris flow hazard in the study area is divided into four categories: high, medium, low, and micro. The results of the actual field investigation have a good effect on the hazard assessment of debris flow in the study area.

3) Compared with the traditional neural network (GRNN), the accuracy of the IIM-PNN model is increased by 10%, which improves the effectiveness of the evaluation and enables targeted prevention and control measures.

REFERENCES

- [1] X. Bai, J. Jian, S. He, and W. Liu, "Dynamic process of the massive Xinmo landslide, Sichuan (China), from joint seismic signal and morphodynamic analysis," *Bull. Eng. Geol. Environ.*, vol. 78, no. 5, pp. 3269–3279, Jul. 2019.
- [2] A. Arabameri, M. Yamani, B. Pradhan, A. Melesse, K. Shirani, and D. Tien Bui, "Novel ensembles of COPRAS multi-criteria decision-making with logistic regression, boosted regression tree, and random forest for spatial prediction of gully erosion susceptibility," *Sci. Total Environ.*, vol. 688, pp. 903–916, Oct. 2019.
- [3] M. J. Froude and D. N. Petley, "Global fatal landslide occurrence from 2004 to 2016," *Natural Hazards Earth Syst. Sci.*, vol. 18, no. 8, pp. 2161–2181, Aug. 2018.
- [4] W. Tang, H.-T. Ding, N.-S. Chen, S.-C. Ma, L.-H. Liu, K.-L. Wu, and S.-F. Tian, "Artificial neural network-based prediction of glacial debris flows in the ParlungZangbo basin, southeastern Tibetan plateau, China," *J. Mountain Sci.*, vol. 18, no. 1, pp. 51–67, Jan. 2021.
- [5] H. Luo, L. Zhang, H. Wang, and J. He, "Multi-hazard vulnerability of buildings to debris flows," *Eng. Geol.*, vol. 279, Dec. 2020, Art. no. 105859.
- [6] Z. Kang, C. Li, A. Ma, and J. Luo, *Research on Debris Flow in China*. Beijing, China: Science Press, 2004.
- [7] Q. Yin and N. Su, "Debris-flow risk assessment based on DEA redundancy analysis," *Chin. J. Geol. Hazard Control*, vol. 31, no. 3, pp. 30–34, Jun. 2020.
- [8] A. Katsuji, T. Kunio, and N. Akato, "Determination of the risk of earth-rock flow," *New Sand Control*, vol. 30, pp. 7–16, 1977.
- [9] R. Hollingsworth and G. S. Kovacs, "Soil slumps and debris flows: Prediction and protection," *Environ. Eng. Geosci.*, vol. 38, no. 1, pp. 17–28, Feb. 1981.
- [10] I. S. Zektser, A. P. Belousova, and V. Y. Dudov, "Regional assessment and mapping of groundwater vulnerability to contamination," *Environ. Geol.*, vol. 25, no. 4, pp. 225–231, Jun. 1995.
- [11] R. J. Hofmeister and D. J. Miller, "GIS-based modeling of debris-flow initiation, transport and deposition zones for regional hazard assessments in western Oregon, USA," *Debris-Flow Hazards Mitigation: Mech., Predict., Assessment*, vol. 1, no. 2, pp. 1141–1149, 2003.
- [12] D. M. Staley, "Data used to characterize the historical distribution of wildfire severity in the western United States in support of pre-fire assessment of debris-flow hazards," United States Geol. Surv., Commonwealth Virginia, Tech. Rep., 2018.

- [13] A. A. Khan, A. Jamil, D. Hussain, M. Taj, G. Jabeen, and M. K. Malik, "Machine-learning algorithms for mapping debris-covered glaciers: The Hunza basin case study," *IEEE Access*, vol. 8, pp. 12725–12734, 2020.
- [14] M. Jakob, E. Mark, S. McDougall, P. Friele, C.-A. Lau, and S. Bale, "Regional debris-flow and debris-flood frequency–magnitude relationships," *Earth Surf. Processes Landforms*, vol. 45, no. 12, pp. 2954–2964, 2020.
- [15] B. Tan, "Quantitative comprehensive evaluation of the severity of debris flow gully," *Bull. Soil Water Conservation*, no. 1, pp. 51–57 and 44, Feb. 1986.
- [16] X. Liu and C. Tang, *Risk Assessment of Debris Flows*. Beijing, China: Science Press, 1995.
- [17] X. Liu, "Debris flow hazard assessment in China: A review and perspective," *J. Natural Disasters*, vol. 11, pp. 1–8, Feb. 2002.
- [18] X. Liu, "Quantitative assessment on site-specific debris flow hazard and application," *J. Disaster Prevention Mitigation Eng.*, vol. 30, no. 3, pp. 241–245, Jun. 2010.
- [19] F. Meng, G. Li, and M. Li, "Application of stepwise discriminant analysis to screening evaluation factors of debris flow," *Rock Soil Mech.*, vol. 31, pp. 2925–2929, 2010.
- [20] D. Xue, Z. He, S. Tao, and D. Zhang, "The '5.12' earthquake after regional hazard assessment of geological disasters—Case study in Qingchuan and Pingwu County," *Advances Earth Sci.*, vol. 26, no. 3, pp. 311–318, 2011.
- [21] F.-G. Yang, Y. Liang, V. P. Singh, W.-S. Wang, X.-Q. Zhou, X.-N. Liu, S.-Y. Cao, E. Huang, and Y.-H. Wu, "Debris flow hazard assessment using set pair analysis models: Take Beichuan County as an example," *J. Mountain Sci.*, vol. 11, no. 4, pp. 1015–1022, Jul. 2014.
- [22] T.-C. Chang, "Risk degree of debris flow applying neural networks," *Natural Hazards*, vol. 42, no. 1, pp. 209–224, May 2007.
- [23] X. Li, J. Kong, C. Li, J. Kong, and C. Li, "Application of BP neural network to the hazard assessment of debris flow based on MATLAB," *Geotech. Invest. Surveying*, vol. 38, no. 1, pp. 47–50, 2010.
- [24] W. Lin, W. Zhou, J. Li, L. Yang, and Z. Tang, "Debris flow hazard assessment based on combination of flow-R and FLO-2D models," *Yangtze River*, to be published.
- [25] D. P. Berrar, C. S. Downes, and W. Dubitzky, "Multiclass cancer classification using gene expression profiling and probabilistic neural networks," in *Proc. Pacific Symp. Biocomput. Pacific Symp. Biocomput.*, 2003, pp. 5–16.
- [26] F. Ancona, A. M. Colla, S. Rovetta, and R. Zunino, "Implementing probabilistic neural networks," *Neural Comput. Appl.*, vol. 5, no. 3, pp. 152–159, Sep. 1997.
- [27] R. Jena, B. Pradhan, G. Beydoun, A. Alamri, and A. Shanableh, "Spatial earthquake vulnerability assessment by using multi-criteria decision making and probabilistic neural network techniques in Odisha, India," *Geocarto Int.*, to be published.
- [28] S. A. Drivelos, G. P. Danezis, M. Halagarda, S. Popek, and C. A. Georgiou, "Geographical origin and botanical type honey authentication through elemental metabolomics via chemometrics," *Food Chem.*, vol. 338, Feb. 2021, Art. no. 127936.
- [29] A. Upadhyay, P. Tantry, and A. Varade, "Classification of seagrass habitat using probabilistic neural network," in *Proc. 10th Int. Conf. Innov. Bio-Inspired Comput. Appl.*, 2021, pp. 250–257.
- [30] F. Chen, L. Yao, and H. Zhao, "Discussion on the risk assessment of debris flow," *Sci. Technol. Eng.*, vol. 18, no. 32, pp. 114–123, Nov. 2018.
- [31] M. Wang, "Debris flow risk division based on neural network," *Hydrogeol. Eng. Geol.*, no. 2, pp. 18–19, Feb. 2000.
- [32] X. Li and D. Mo, "Application of GIS and ann technology to evaluate the hazard degree of debris-taking Yunnan Province as an example," *Res. Soil Water Conservation*, vol. 12, no. 4, pp. 7–9, 2005.
- [33] P. Chen, J. Qiao, and Z. Peng, "Screening of debris flow risk factors and risk evaluation based on rank correlation," *Rock Soil Mech.*, vol. 34, pp. 1409–1415, Jan. 2013.
- [34] X. Wang, G. Li, and B. You, "Debris flow hazard assessment based on Monte Carlo simulation," in *Proc. Int. Conf. Multimedia Technol.*, Jul. 2011, pp. 3912–3915.
- [35] R. R. Yager and D. P. Filev, "Induced ordered weighted averaging operators," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 29, no. 2, pp. 141–150, Apr. 1999.
- [36] Z. Xu and Q. L. Da, "An overview of operators for aggregating information," *Int. J. Intell. Syst.*, vol. 18, no. 9, pp. 953–969, 2003.
- [37] Y. Zhang, *Introduction to Multivariate Statistical Analysis*. Beijing, China: Science Press, 1999.
- [38] Y. Kong, Z. Liu, and L. Liu, "Evaluation method of hydropower project investment decision-making," *Water Resource Power*, vol. 31, no. 1, pp. 141–143, 2013.
- [39] L. Shi and X. Wang, "Application of probabilistic neural network in engine fault diagnosis," *Autom. Instrum.*, vol. 32, pp. 33–35, Jun. 2011.
- [40] D. F. Specht, "Probabilistic neural networks," *Neural Netw.*, vol. 3, no. 1, pp. 109–118, 1990.
- [41] Z. Bian, *Pattern Recognition*. Beijing, China: Tsinghua Univ. Press, 2000.
- [42] C. Mi, C. Zhao, Q. Deng, and X. Deng, "Prediction of chilling damage risk in maize growth period based on probabilistic neural network approach," *Int. J. Agricult. Biol. Eng.*, vol. 14, no. 2, pp. 120–125, 2021.
- [43] J. Liu and G. Ou, "New opinion on debris flows hazard assessment," *J. Geol. Hazards Environ. Preservation*, vol. 15, no. 1, pp. 5–8, Mar. 2004.
- [44] X. Yin, B. Zhang, Z. Zhou, X. Han, Z. Wang, and G. Hu, "A new health estimation model for CNC machine tool based on infinite irrelevance and belief rule base," *Microelectron. Rel.*, vol. 84, pp. 187–196, May 2018.
- [45] C.-N. Liu, J.-J. Dong, Y.-F. Peng, and H.-F. Huang, "Effects of strong ground motion on the susceptibility of gully type debris flows," *Eng. Geol.*, vol. 104, nos. 3–4, pp. 241–253, Mar. 2009.
- [46] X. Liu, "Quantitative assessment on site-specific debris flow hazard and application," *J. Disaster Prevention Mitigation Eng.*, vol. 30, no. 3, pp. 241–245, 2010.
- [47] P. Chen, H. Yu, and Y. Liu, "Evaluation of debris flow risk based on independent information data fluctuation weighting method," *Rock Soil Mech.*, vol. 34, pp. 449–454, Jun. 2013.
- [48] Z. Wen, "Application of FCM based on principal components analysis in debris flow classification," *J. Jilin Univ. Earth Sci. Ed.*, vol. 40, no. 2, pp. 368–372, 2010.
- [49] W. Zhang, H.-Z. Li, J.-P. Chen, C. Zhang, L.-M. Xu, and W.-F. Sang, "Comprehensive hazard assessment and protection of debris flows along Jinsha river close to the Wudongde dam site in China," *Natural Hazards*, vol. 58, no. 1, pp. 459–477, Jul. 2011.



CUI WANG is currently pursuing the master's degree with the College of Mining, Liaoning Technical University. His research interests include disaster geology and mine geological environment protection and restoration.



XUEDONG WANG received the Ph.D. degree from the Construction Engineering College, Jilin University, in 2013. He is currently an Associate Professor with the College of Mining, Liaoning Technical University. He is also a Postdoctoral Fellow in the research station on mechanics for postdoctoral fellows at Liaoning Technical University. His research interests include disaster geology and mine geological environment protection and restoration.



YONGDONG ZHU is currently pursuing the master's degree with the College of Mining, Liaoning Technical University. His research interests include disaster geology and mine geological environment protection and restoration.

...