Early warning of debris flow using optimized self-organizing feature mapping network
Xuedong Wang, Cui Wang and Chaobiao Zhang

ABSTRACT

Early warning of debris flow is one of the core contents of disaster prevention and mitigation work for debris flow disasters. There are few early warning methods based on the combination of rainfall threshold and geological environment conditions. In this paper, we presented an early warning method for debris flow based on the infinite irrelevance method (IIM) and self-organizing feature mapping (SOFM), and applied it to Liaoning Province, China. The proposed model consisted of three stages. Firstly, eight geological environmental conditions and two rainfall-inducing conditions were selected by analyzing the factors affecting the development of debris flow in the study area, and the rainfall threshold for debris flow outbreak was 150 mm. Secondly, the correlation between various factors was analyzed by IIM, which prevented the blindness of parameter selection and improved the prediction accuracy of the model. Finally, SOFM was employed to predict the test data. Experimental results showed that the IIM-SOFM model had a strong early warning ability. When 25 samples of low-frequency debris flow area were selected, the accuracy rate of the IIM-SOFM model with optimized network structure parameters was 100%, which it was obviously superior to the rainfall threshold method, BP neural network and competitive neural network. Consequently, it is feasible to use the IIM-SOFM model for early warning of debris flow, outperforming traditional machine learning methods.

Key words | debris flow, early warning, infinite irrelevance method, self-organizing feature mapping

HIGHLIGHTS

• A novel hybrid IIM-SOFM early warning model is proposed.
• Eight geological environmental conditions and two rainfall-inducing conditions are selected by analyzing the factors affecting the development of debris flow.
• The correlation between various factors is analyzed by IIM, and it improves the prediction accuracy of the model.
• The IIM-SOFM model is superior to the rainfall threshold method, BP neural network, and competitive neural network.
• The IIM-SOFM model has high accuracy and stability in a low-frequency debris flow area.

INTRODUCTION

China is a country with serious debris flow disasters. Since the 1970s, the annual death toll caused by debris flow has exceeded 3,700 (Niu et al. 2014). Effective debris flow early warning has become a key task of disaster prevention and mitigation. Its research has always been a hot topic in academia, but there are still some shortcomings in the study of debris flow (Ma et al. 2017).

Rainfall conditions are the most direct conditions for inducing debris flow. Previously there has been a lot of research on rainfall conditions during the occurrence of
debris flow, and the rainfall threshold has become the most widely used early warning method (Rosi et al. 2015; Pan et al. 2018). To analyze the primary causes of debris flow occurrence, it was necessary to understand the relationship between rainfall threshold and debris flow initiation. Local rainfall threshold was proposed as a result of field investigations of specific debris flows (Tang et al. 2011). Empirical approaches on historical and statistical bases were also used. The rainfall intensity-duration (I-D) was the most common type of threshold which was widely applied in different climatic and geological settings (Guzzetti et al. 2008; Saito et al. 2010). Guo et al. (2016) looked at 518 debris flow events that occurred in the Wenchuan earthquake-stricken area in order to analyze their spatial characteristics. Such results are useful for debris flow early warning based on empirical rainfall threshold.

The method of determining the rainfall threshold based on historical data and statistical theory brings a lot of uncertainty, and some new research methods have emerged. Chen et al. (2019) defined an elastic wave velocity threshold for rainfall-induced landslide prediction and early warning. Rosatti et al. (2019) proposed a new rainfall threshold methodology based on volumetric relations deriving from a simplified description of the dynamic of debris flow. Study of the effect of spatial variability of soil deposits and the related stratigraphic settings also was carried out to assess rainfall conditions leading to debris flow (Tufano et al. 2016). In addition, Huang et al. (2019) presented critical pore pressure threshold in combination with rainfall factors for gully-type debris flow early warning. Physical methods (e.g. numerical simulation and model test simulation) were used to find relationships among rainfall, detritus material properties, and pore pressure and their contributions to the outbreak of debris flow (Li et al. 2018). Zhao et al. (2019) proposed a probabilistic threshold to integrate antecedent soil moisture conditions with rainfall threshold. Also, 3D WebGIS and GIS based on a combination of wireless sensor network platforms matured in debris flow early warning (Tiranti et al. 2014; Huang et al. 2015). The rainfall threshold method combined with other methods obtained a more scientific and effective early warning method. Therefore, it is necessary to include the early warning model of geological environment development conditions for debris flow development.

As an economic, effective, and advanced prevention method, debris flow monitoring and early warning has been widely of concern to scholars and government departments, domestic and abroad. The main monitoring contents include deformation, displacement rate, source accumulation velocity, soil water content, groundwater level, pore water pressure, pre-rainfall, rainfall, rainfall intensity, rainfall duration, etc. (Baum & Godt 2010). Based on a correlation study between monitoring data and debris flow outbursts, it was determined that the criteria model for monitoring and early warning of debris flow was the core of the technology (Naidu et al. 2018). The sensor system was the key component in monitoring and early warning applications, and its precision and reliability significantly influenced the effectiveness of an early warning system (Arattano & Marchi 2008). However, conventional sensors for debris flow monitoring had several drawbacks (Zhang et al. 2019); the service life of the sensors was less than the recurrence interval of a debris flow, which increased the cost of monitoring (Chen et al. 2016). Even wireless transmission technologies based on fiber Bragg gratings, micro-drones and small aperture arrays were widely used, and lightning and other effects could greatly affect the stability of signal transmission (Zhang & Chen 2017; Nagatani et al. 2018; Marchetti et al. 2019). Obviously, monitoring and early warning methods are uneconomical and impractical for many debris flows in a low-frequency debris flow area.

In recent years, many researchers have developed new models of intelligent algorithms due to the frequent occurrence of debris flows. Back propagation (BP) neural networks were widely used in early warning models (Li & Wang 2007), but their convergence tended to be slow and easily fell into a local minimum (Ji & Zhou 2011). A new early-warning system based on the radial basis function (RBF) network, using rainfall indices and soil–water index, was established in Japan in 2005 and has operated since 2007 (Osanai et al. 2010). But RBF networks need to know the number of implicit layer nodes and cluster centers in advance, which affects the effect of the application. The self-organizing feature mapping (SOFM) artificial neural network was first proposed by Kohonen in the 1980s (Kohonen 1981). The most critical feature of SOFM is its ability to provide a topology that preserves the mapping of high-dimensional inputs to low-dimensional meshes. SOFM is particularly well suited for data clustering and predictive analysis because
it facilitates the use of unique human insights (Boniecki et al. 2019). Network structures and parameters also have a large impact on their accuracy (Liu et al. 2019). Therefore, the optimized SOFM model can be well applied for debris flow warning.

Based on previous studies, an IIM-SOFM early warning model is proposed for low-frequency debris flow areas. To accomplish this, the infinite irrelevance method can reduce the dimensionality between factors, and the optimized IIM-SOFM model can effectively achieve early warning.

STUDY AREA

Overview of the study area

Xiuyan Manchu Autonomous County is affiliated to Anshan City, Liaoning Province, China (Figure 1).

The geographical coordinates are E122°52’–123°46’, N40°00’–40°39’, with a total area of 4,502 km², and 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 (Figure 2), serious disasters are easily caused. In a debris flow survey, 358 debris flows were found.

Meteorological and hydrological characteristics

The study area is located in the temperate humid monsoon climate zone, with an average annual temperature of 7.5 °C, and the highest temperature in July, when the average temperature is 23.0 °C and the highest temperature is 37.3 °C; and with the lowest temperature in January, when the average temperature is −10.7 °C and the lowest temperature is −30.9 °C. The average annual precipitation is 896 mm, the annual maximum rainfall is 1,451.3 mm (1964), and the minimum rainfall is 172.2 mm (1962). The precipitation during the year is highly concentrated in June–September. Precipitation accounts for 73.5% to 80.2% of the annual precipitation. Heavy rains often occur in July–August. Precipitation with a daily rainfall of ≥50 mm occurs 2.3–3.9 times per year in the study area. The rainfall in the study area is concentrated and heavy rains have become the predisposing factor for geological disasters in debris flows.

General characteristics of debris flow development

Topography, geology, and hydrogeology are the three main influencing factors for debris flow outbreaks (Liu 2002). Under the combined action of rainfall and geological
environment conditions, 358 potential debris flows in the study area have the following developmental characteristics. The outbreak of debris flow is mainly caused by continuous rainfall and torrential rain, and the accumulation of loose debris must be determined after an outbreak. At the time, the outbreak of debris flow showed a certain cumulative growth. Therefore, the frequency of debris flow outbreaks decreased, belonging to the type of low-frequency debris flow, but often reflected the characteristics of mass-generation. Once the outbreak occurred, the scale was large and the damage was great.

**METHODOLOGY**

**Infinite irrelevance method**

Attributes that affect debris flow early warning have a high dimensionality and include numerous irrelevant attributes. Indicator selection becomes crucial, and the core issue is to reduce the dimension, for the study of indicators with high relevance (Yager & Filev 1999; Xu & Da 2005).

The analytical method used in this paper is an infinite irrelevance method. The infinite irrelevance method starts from the analysis of the multi-correlation between the indicators, and uses the multi-correlation to remove those from the p indicators $x_1, x_2, \ldots, x_p$ which can be replaced by the remaining indicators, therefore, some of the indicators are selected to comprehensively reflect the original p indicators, making the evaluation results more realistic and effective.

According to the determined p indicators, the values of each node are calculated separately, and the infinite irrelevance group is obtained by the infinite irrelevance method. The solution steps are as follows (Zhang 1999).

1. Determining analytical samples

   The analysis sample refers to n groups of evaluation object values of p indicators. The constructed sample X is expressed as follows: $X = (x_1, x_2, \ldots, x_p)$.

2. Solving the correlation matrix of the analysis sample

   The covariance matrix and correlation matrix of X are recorded as $V(x)$ and $R(x)$, respectively:

   $V(x) = \begin{bmatrix} \frac{1}{\sqrt{\sigma_{11}}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\sigma_{22}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{\sigma_{pp}}} \end{bmatrix}$

   $R(x) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$

   (1)

   In the formula: $\sigma_{ij} = \frac{1}{n} \sum_{i=1}^{n} (x_{ai} - \bar{x}_i)(x_{aj} - \bar{x}_j)$, when $i = j$, $\sigma_{ij}$ represents the variance of $x_i$; when $i \neq j$, $\sigma_{ij}$ is the covariance of $x_i$ and $x_j$.

3. Calculating the multi-correlation coefficient between each indicator and other $p - 1$ indicators

   Substituting $R$ for $R'$ replaces the row i and column j of the matrix R with the last row and the last column of the matrix. Considering the degree of linear correlation between a variable $x_i$ and the remaining $p - 1$ variables, called the multi-correlation coefficient, usually denoted as $r_{xi}$, now r is divided into blocks:

   $r(x) = \begin{pmatrix} R_{x,i} & r_i \\ r_i^T & 1 \end{pmatrix}$

   (2)

   $R_{x,i}$ is the correlation matrix with the variable $x_i$ removed. $R_{x,i}$ is a matrix of elements of $R'$ from 1 to $p - 1$ row and from 1 to $p - 1$ column, $r_i$ is a matrix of $R'$'s $p$ column and 1 to $p - 1$ row, and $r_i^T$ is a matrix of $R'$'s $p$ rows and 1 to $p - 1$ column elements. Currently, the multi-correlation coefficient formula between $x_i$ and the remaining $p - 1$ variables is:

   $r_{x_i}^2 = r_i^T R_{x-i} r_i$

   (3)

   If the multi-correlation coefficient is larger, it indicates that $x_i$ is more easily replaced by $x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p$, and the effect of the index $x_i$ on the evaluation target is weaker. After the threshold value D is specified, when $D > D$, the indicator $x_i$ can be deleted, and the remaining indicators have a relatively small multi-correlation coefficient, which can well reflect the original evaluation index system.
Self-organizing feature mapping network

Self-organizing feature mapping (SOFM) has been widely studied and applied in vector quantization and pattern recognition (Su & Wah 2003). Usually, if you enter a mode arbitrarily in the network, it will generate a new area based on the input signal for self-organizing learning. At this time, the input mode will be affected, resulting in different responses. There are basically three types of self-organizing networks, each with a competitive layer.

SOFM is achieved by the neural architecture shown in Figure 3, and the visual representation produced by it can be used to form assumptions about the topology (Pal et al. 2005).

The SOFM network learning process consists of the following steps. For a given $n$-dimensional input modes $X_k = (x_1, x_2, \ldots, x_m)$, $k = 1, 2, \ldots, n$; suppose the competition layer contains $R$ neurons, then SOFM has been studied. The process is roughly as follows.

1. Weight initialization
   
   Give the initial weight vector $IW_{ij} (i = 1, 2, \ldots, S^1, j = 1, 2, \ldots, R)$ ($IW_{ij}$ represents the connection weight between the $i$-th neuron of the competition layer and the $j$-th neuron of the output layer), giving a small random number; $n^1$ is the winning neuron, $a^1$ is the output of the competitive layer neurons. At the same time, the initial neighboring threshold $N_C(t)$ of the output node is determined, the maximum number of iterations is $T$, and the initial value of the number of iterations is $N = 1$.

2. Randomly select a sample $x$ as input in the sample set
   
   For the $k$-th input vector $X_k = (x_1, x_2, \ldots, x_m)$, do the following calculation:

   $\circ$ Find the output neuron $C$ that best matches $X_k$. Generally, distance matching is used to find $C$ to minimize the distance between $W_C$ and $X_k$: $C: X_k = \text{MIN} \{X_k - W_j, j = 1, 2, \ldots, R\}$.

   $\circ$ Determine the neighborhood interaction function, and calculate the excitability $F_C(j)$ of each output neuron in the neighborhood of the ‘side suppression’ feedback region with the radius of $N_C(t)$ as the center.

   $\circ$ Adjusting weight

   \[ IW_j(t + 1) = W_j(t) + \eta(t)F_C(j)(X_k) - W_j(t), \quad j \in N_C(t) \]
   \[ IW_j(t) = W_j(t), \quad j \notin N_C(t) \] (4)

   $\eta(t)$ is called the learning rate, and $0 < \eta(t) < 1$, $\eta(t)$ and $N_C(t)$ decrease with the increase of the number of learnings.

3. Learning rate and neighborhood update
   
   After the winning neurons and their neighbors have updated the weights, the learning rate and the neighborhood need to be updated before entering the next iteration, namely:

   \[ \eta = \eta(1 - N/T); \quad N_C = [N_C(1 - N/T)] \] (5)

   [ ]Indicates rounding up

4. End of iteration judgment
   
   The second step of the calculation is repeated until the predetermined number of learnings is reached, or the amount of change in weight during each learning is less than a given threshold.

Establishing a comprehensive early warning system based on IIM and SOFM

A nonlinear SOFM based on improved IIM is studied in this research. An early warning model of debris flow was established based on field surveys, interpretation of...
meteorological data and remote sensing images of the study area. Figure 4 shows a flow chart of the IIM-SOFM model. The warning steps for the IIM-SOFM model are as follows.

Step 1: Combine survey data and meteorological data, analyze and evaluate the factors affecting the outbreak of debris flow in the study area, and initially select the main influencing factors.

Step 2: Combine statistical knowledge to study the relationship between rainfall and debris flow outbreak, and obtain the rainfall threshold of debris flow outbreak in the study area.

Step 3: Use IIM to analyze the correlation of primary selection factors, eliminate the more relevant factors, and obtain the best early warning indicators.

Step 4: The above optimized index data are subjected to SOFM training, and the predicted results of the training model are tested in combination with the known debris flow development result data, and combined with the rainfall threshold to form the final debris flow early warning binary model.

Step 5: The debris flow early warning is carried out using the optimized IIM-SOFM model.

DETERMINATION OF RAINFALL THRESHOLD AND SELECTING THE INFLUENCING FACTORS

The existing debris flow warning methods mainly include three categories, namely monitoring and early warning, rainfall threshold warning and comprehensive early warning methods. The monitoring and early warning method is to pre-set monitoring equipment (dynamic monitoring of deformation, earth pressure, pore water pressure, etc.) in the potential debris flow ditch for early warning (Mohanty et al. 2019). The rainfall threshold is the main early warning method for regional debris flow. By studying the correlation between rainfall and debris flow outburst, the critical rainfall threshold of debris flow outbreak is determined for early warning. The comprehensive early warning method is to comprehensively consider the factors of debris flow outbursts, and early warning is also the development direction of the debris flow warning method (Segoni et al. 2018).

Applicability of monitoring method for early warning

Debris flow monitoring and early warning methods are the most effective method for single-channel debris flow

---

**Figure 4** | The flowchart of the early warning model: (a) monitor; (b) outbreak record; (c) field investigation; (d) remote sensing; (e) radio broadcast; (f) telephone; (g) manual transfer; (h) TV forecast.
warning. According to the survey, there were 358 potential debris flow ditches in the study area. The statistics of the outbreaks of debris flows collected in the past 30 years are shown in Figure 5.

Combining Figure 5 and the People’s Republic of China industry standard ‘Destruction Code for Debris Flow Disaster Prevention Engineering’ DZ/T0220-2006, only one of the 18 known debris flow gullies belongs to the intermediate-frequency debris flow ditch, and the rest belong to the typical low-frequency debris flow. The largest deposit volume is 164,800 m$^3$, which is a medium-sized and above debris flow, and the risk of debris flow outbreak is high. Therefore, it is obviously uneconomical and unrealistic to adopt a monitoring and early warning method in the study area with many debris flows, low frequency, and high risk. It is feasible to seek a safe, effective, and economical method.

Determinant of rainfall threshold

The debris flow rainfall threshold early warning is the most popular regional debris flow warning method because it can be synchronized with meteorological early warning (Mirus et al. 2018). The general early warning level of debris flow can be divided into five levels. Generally, the third level and above indicate the possibility of debris flow outbreak, and the warning information is released. The meaning of each level is shown in Table 1 below.

The rainfall indicators generally considered for the debris flow rainfall threshold include 24-hour rainfall, hourly rain intensity, and pre-rainfall (the first five days of rainfall and the first three days of rainfall, etc.). The range of debris flow outbreak from upstream to downstream can be divided into formation area, circulation area and accumulation area. The characteristics of debris in each area can reflect the hydrodynamic conditions of debris flow formation. Each of the three groups of debris source in the area and the accumulation area was subjected to a screening test and averaged, and the obtained results are shown in Figure 6. The coefficient of nonuniformity in the formation area, flow area and accumulation area is less than 5, and the coefficient of curvature is greater than 3, the grading of the debris is poor, the fine particles are accumulated, the debris is well permeable and the reserves are large, and the previous period is attenuated. The five-day compression and the first three days of reduction, etc. have less impact on the debris flow outbreaks in the region and can be ignored.
Therefore, the debris flow of the Xiuyan is a rare river-valley-type debris flow. The excitation condition is heavy rain. Based on the data recorded for debris flow in the past 30 years, the correlation between rainfall conditions and debris flow outbreaks has been studied (Cong et al. 2006). The Logistic regression model was combined with the pre-effective rainfall. It is considered that the critical rainfall threshold for the debris flow in Xiuyan County is 150 mm. When the rainfall reaches this magnitude, there should be valleys with a poor geological environment within the rainfall range. Be vigilant to reduce the damage caused by mudslide disasters to people’s lives and property.

Although some research results have achieved certain early warning effects, there are still certain problems. We had statistics on the relationship between mudslide outbreaks in the study area and rainfall (24-hour rainfall and hourly rain) over the past 30 years, as shown in Figure 7.

It can be seen from Figure 7 that when the rainfall threshold is 150 mm, the 11th debris flow ditch with rainfall of 104.8 mm and hourly rain intensity of 40.6 mm also has a mudslide disaster. Eight groups of 18 known outbreaks of mudslides had early warning errors, and the warning error rate was as high as 44.44%. This is because the debris flow outburst is affected by many factors. It is
obviously one-sided to use the rainfall threshold as the early warning indicator. Therefore, it is necessary and feasible to find a multi-factor debris flow warning method combined with rainfall conditions and geological environment conditions.

Method used to measure the influencing factors

As mentioned earlier, there are three main influencing factors for debris flow outbreaks. The maximum (possible) flush out and the frequency of debris flows were the two most important indicators of debris flow hazard. The topographical conditions were mainly the dynamic conditions of the debris flow outbreak, providing sufficient material sources and catchment conditions for the debris flow outbreak. Hydrogeological conditions had an important impact on the formation of debris flows, and the greatest impact on storm-type debris flows was rainfall conditions (Liu et al. 2009). Therefore, based on the comprehensive consideration of the dangerous conditions, provenance conditions, dynamic conditions and induced conditions of the debris flow outbreak in the study area, combined with the results of the ‘Geological disaster investigation and zoning in Xiuyan Manchu Autonomous County, Anshan City, Liaoning Province, China’, the preliminary selection of the four types of impact conditions, there are a total of ten early warning indicators, as shown in Table 2.

Firstly, based on the outbreak records and investigations of debris flows in the study area for nearly 30 years, the largest (possible) displacement and frequency data of 25 debris flow outbreaks were collected. On this basis, the rainfall monitoring data of ten hydrological stations and rainfall stations in Xiuyan County were collected for nearly 30 years. Comparing this with the recorded debris flow, it is easy to extract data on the corresponding 24 h maximum rainfall and hourly rainfall intensity.

Secondly, according to the Chinese National Standard (Slope and Debris Flow Disaster Investigation Procedures (1:50000; DD2008-02)), the remaining six influencing factors, namely the drainage area, the ditch slope, the vegetation coverage, and the loose material are explained by System Probatoire d’ Observation de la Terre 5 (SPOT5) image and the digital elevation model (DEM) can estimate the reserves per unit area, the average slope and relative height difference of the grooved bed.

Zhang et al. (2011) provided details for implementing this process. To improve the accuracy of the interpretation results, based on a standard large-scale topographic map (1:1000), 358 mudslide trenches were investigated in the field. Influencing factors throughout the study area were measured using three measurement tools: a GPS locator, tape measure and laser rangefinder. In addition, drilling was performed at prearranged points. Therefore, according to image interpretation and field investigation, the analysis and quantitative assignment of the debris flow warning indicators in the study area are realized.

RESULTS AND DISCUSSION

Quantitative results for indicators

In order to verify the validity of the IIM-SOFM model prediction, the debris flows of 25 known outbreak data in the past 30 years in the study area were selected; 20 debris flows were randomly selected for training tests and five for testing. The quantitative results for the early warning indicators are shown in Table 3.
Indicator optimization based on IIM

The early warning accuracy of the model is affected by the original influencing factors, and the prediction accuracy is also strongly dependent on the correlation of the predictive indicators. The following paragraphs focus on the selection and optimization of primary selection indicators.

Based on the IIM method, 20 groups were randomly selected as training data, and the last five groups were used as test samples. The optimized indicator was the input value of the neural network, and the predicted level of the debris flow degree was the output value. The primary factors were screened using Equations (1)–(5) and MATLAB software. Considering the size and frequency of debris flows as the main intrinsic factor for the evaluation of debris flow risk, the empirical criteria of $D = 0.9$ and ‘screening to scale and frequency’ were adopted (Wang et al. 2012). The multi-correlation coefficients $D$ of ten indicators were calculated and obtained, which were 0.69, 0.4279, 0.7392, 0.643, 0.4302, 0.4456, 0.7042, 0.7247 and 0.519. The maximum multi-correlation coefficient $D$ is 0.7392 of the basin area index, indicating that the selected ten early warning indicators have little correlation with each other and can be used for early warning.

Table 3  The quantitative results of early warning indicators

<table>
<thead>
<tr>
<th>Ditch</th>
<th>Maximum displacement $10^3$ m$^3$</th>
<th>Frequency/ 100 years</th>
<th>The drainage area km$^2$</th>
<th>Slope of ditch bank $%$</th>
<th>Vegetation coverage</th>
<th>Loose material reserves per unit area $10^3$ m$^3$</th>
<th>Average slope of trench bed $10^3$ m$^3$</th>
<th>Relative height difference $%$</th>
<th>Hour rain intensity mm</th>
<th>24 h maximum rainfall mm</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>315.9</td>
<td>1</td>
<td>7.6</td>
<td>128.4</td>
<td>552.2</td>
<td>34</td>
<td>265.6</td>
<td>79</td>
<td>21.5</td>
<td>94</td>
<td>I</td>
</tr>
<tr>
<td>2</td>
<td>164.8</td>
<td>3</td>
<td>3.9</td>
<td>167</td>
<td>584.6</td>
<td>39.5</td>
<td>183.3</td>
<td>87.7</td>
<td>28.7</td>
<td>91.8</td>
<td>I</td>
</tr>
<tr>
<td>3</td>
<td>243.5</td>
<td>3</td>
<td>3.1</td>
<td>166.9</td>
<td>534</td>
<td>33.1</td>
<td>267.2</td>
<td>76.2</td>
<td>25.2</td>
<td>84.7</td>
<td>I</td>
</tr>
<tr>
<td>4</td>
<td>1,031.9</td>
<td>8</td>
<td>8.4</td>
<td>158</td>
<td>568.7</td>
<td>27.2</td>
<td>453.9</td>
<td>67.5</td>
<td>32.2</td>
<td>84.9</td>
<td>II</td>
</tr>
<tr>
<td>5</td>
<td>2,503.7</td>
<td>4</td>
<td>14.7</td>
<td>111.6</td>
<td>714</td>
<td>30.4</td>
<td>524.6</td>
<td>80.2</td>
<td>47.9</td>
<td>118.3</td>
<td>II</td>
</tr>
<tr>
<td>6</td>
<td>1,992.5</td>
<td>2</td>
<td>12.6</td>
<td>134.4</td>
<td>726</td>
<td>23</td>
<td>444.2</td>
<td>62.2</td>
<td>38.7</td>
<td>103.7</td>
<td>II</td>
</tr>
<tr>
<td>7</td>
<td>1,480.5</td>
<td>5</td>
<td>10.7</td>
<td>153.3</td>
<td>720.7</td>
<td>25</td>
<td>421.6</td>
<td>75.2</td>
<td>35.6</td>
<td>95.7</td>
<td>II</td>
</tr>
<tr>
<td>8</td>
<td>1,456.6</td>
<td>10</td>
<td>19.8</td>
<td>118.9</td>
<td>701.5</td>
<td>37.1</td>
<td>409.1</td>
<td>69.1</td>
<td>28.2</td>
<td>113.2</td>
<td>III</td>
</tr>
<tr>
<td>9</td>
<td>722.2</td>
<td>5</td>
<td>20.8</td>
<td>86.4</td>
<td>544.3</td>
<td>33</td>
<td>411.4</td>
<td>61.3</td>
<td>28.5</td>
<td>114.9</td>
<td>III</td>
</tr>
<tr>
<td>10</td>
<td>1,441</td>
<td>3</td>
<td>16.5</td>
<td>53.4</td>
<td>363.3</td>
<td>29.1</td>
<td>296.8</td>
<td>77</td>
<td>47.2</td>
<td>128.5</td>
<td>III</td>
</tr>
<tr>
<td>11</td>
<td>1,441</td>
<td>3</td>
<td>16.5</td>
<td>53.4</td>
<td>363.3</td>
<td>18.6</td>
<td>296.8</td>
<td>40.6</td>
<td>43.5</td>
<td>104.8</td>
<td>III</td>
</tr>
<tr>
<td>12</td>
<td>770.8</td>
<td>2</td>
<td>18.9</td>
<td>110.4</td>
<td>695.3</td>
<td>32.9</td>
<td>269.6</td>
<td>90.8</td>
<td>37.9</td>
<td>139.6</td>
<td>IV</td>
</tr>
<tr>
<td>13</td>
<td>777.8</td>
<td>1</td>
<td>6</td>
<td>59.3</td>
<td>272.6</td>
<td>21.3</td>
<td>427.1</td>
<td>76.1</td>
<td>44.6</td>
<td>110.7</td>
<td>IV</td>
</tr>
<tr>
<td>14</td>
<td>996.3</td>
<td>4</td>
<td>20.9</td>
<td>88.2</td>
<td>696.7</td>
<td>33.9</td>
<td>421.4</td>
<td>76.3</td>
<td>58.6</td>
<td>140</td>
<td>IV</td>
</tr>
<tr>
<td>15</td>
<td>777.8</td>
<td>1</td>
<td>6</td>
<td>59.3</td>
<td>272.6</td>
<td>23.6</td>
<td>427.1</td>
<td>46.3</td>
<td>59.8</td>
<td>132.1</td>
<td>IV</td>
</tr>
<tr>
<td>16</td>
<td>1,992.5</td>
<td>2</td>
<td>12.6</td>
<td>134.4</td>
<td>358.2</td>
<td>37.4</td>
<td>444.2</td>
<td>53.8</td>
<td>54.2</td>
<td>168.5</td>
<td>IV</td>
</tr>
<tr>
<td>17</td>
<td>2,321.3</td>
<td>3</td>
<td>13.5</td>
<td>88.9</td>
<td>263.8</td>
<td>26.7</td>
<td>348.3</td>
<td>56.5</td>
<td>73.8</td>
<td>184.6</td>
<td>IV</td>
</tr>
<tr>
<td>18</td>
<td>1,141.4</td>
<td>3</td>
<td>10.8</td>
<td>108.8</td>
<td>634</td>
<td>42.6</td>
<td>408.1</td>
<td>40.8</td>
<td>55.3</td>
<td>186.8</td>
<td>V</td>
</tr>
<tr>
<td>19</td>
<td>239.6</td>
<td>2</td>
<td>10.6</td>
<td>131.3</td>
<td>564.7</td>
<td>24.8</td>
<td>368</td>
<td>43.5</td>
<td>50.4</td>
<td>241.2</td>
<td>V</td>
</tr>
<tr>
<td>20</td>
<td>1,783.2</td>
<td>4</td>
<td>12.5</td>
<td>73.9</td>
<td>539.5</td>
<td>38.1</td>
<td>453.9</td>
<td>53.2</td>
<td>53.5</td>
<td>352.6</td>
<td>V</td>
</tr>
<tr>
<td>21</td>
<td>1,480.5</td>
<td>5</td>
<td>10.7</td>
<td>153.3</td>
<td>720.7</td>
<td>48.2</td>
<td>421.6</td>
<td>68.3</td>
<td>63.3</td>
<td>317.6</td>
<td>V</td>
</tr>
<tr>
<td>22</td>
<td>2,127.9</td>
<td>4</td>
<td>24.1</td>
<td>72.8</td>
<td>713.2</td>
<td>22.8</td>
<td>253.7</td>
<td>38.1</td>
<td>63.4</td>
<td>318</td>
<td>V</td>
</tr>
<tr>
<td>23</td>
<td>3,143.1</td>
<td>7</td>
<td>19.7</td>
<td>82.9</td>
<td>547.2</td>
<td>32.7</td>
<td>462</td>
<td>54.7</td>
<td>64</td>
<td>358</td>
<td>V</td>
</tr>
<tr>
<td>24</td>
<td>164.8</td>
<td>3</td>
<td>3.9</td>
<td>167</td>
<td>584.6</td>
<td>31.6</td>
<td>183.3</td>
<td>50.3</td>
<td>65.6</td>
<td>264.8</td>
<td>V</td>
</tr>
<tr>
<td>25</td>
<td>2,726.1</td>
<td>3</td>
<td>25.4</td>
<td>87.4</td>
<td>498.7</td>
<td>23.7</td>
<td>373.6</td>
<td>42.8</td>
<td>70.3</td>
<td>287.4</td>
<td>V</td>
</tr>
</tbody>
</table>
Prediction result of IIM-SOFM model

The test data were input into the IIM-SOFM model and a prediction of the debris flow warning was obtained. In addition, the results were compared with literature results based on the rainfall threshold for debris flows throughout the region. The warning results are shown in Table 4.

Table 4 shows that the prediction accuracy of the IIM-SOFM model is 100%. When the rainfall threshold model tests five sets of data, one group of prediction results is wrong, and the accuracy rate reaches 80%. In this paper, only 25 data sets with small sample size are validated. The IIM-SOFM early warning model based on rainfall and geological environment conditions still has high precision, which indicates that the proposed method is superior to the traditional rainfall threshold model in processing. In a particular debris flow development area, it can be very difficult to obtain large amounts of basic data. The IIM-SOFM model shows good application prospects in this case and represents a new rapid warning method for early warning of debris flow hazards in under-researched areas.

Comparison and discussion

In this section, the early warning performance between the IIM-SOFM model and traditional machine learning methods will be compared. Rainfall threshold, back propagation (BP) neural networks and competitive neural networks are often used as early warnings for debris flows (Wei et al. 2017). Therefore, we tested these models by using the same data sets and features and compared their results with the findings of the IIM-SOFM model approach presented here. The prediction results and errors are shown in Figure 8.

Figure 8 shows the early warning results for four different classification models, including the IIM-SOFM model approach presented in this paper, while taking advantage of the same data sets and features. IIM-SOFM performs best in test accuracy. The IIM-SOFM model has a test accuracy of 100%, while the other methods have an accuracy of less than 80%. These results show that the IIM-SOFM model method has high efficiency and reliability for debris flow early warning. Based on the field survey data and meteorological data, the intrinsic relationship between the debris flow outburst and the geological environment conditions was studied, and the main factors of the debris flow warning were obtained, which provided a scientific basic data guarantee for the early warning (Shen et al. 2018). The infinite irrelevance method could judge the correlation of the primary selection indicators, scientifically reduced the effective dimension of the indicators, and provided reliable indicators for the SOFM early warning model (Yin et al. 2017).

SOFM has strong nonlinear prediction and clustering capabilities. The reason is that the SOFM system usually consists of an input layer and a 2D contention layer. The input layer receives the input signal and sends it to the competition layer, and performs self-organizing competition on the competition layer composed of the two-dimensional neuron array. The winner of the competition will have the right to respond to the input mode while others are restricted. SOFM can generate feature maps formed by similar sets of input patterns through unsupervised learning, which is usually done by similarity analysis between neurons and input patterns on the competition layer. The weight

<table>
<thead>
<tr>
<th>Number</th>
<th>Extent of outbreak</th>
<th>Level</th>
<th>Rainfall threshold model Level</th>
<th>IIF-SOFM model Level</th>
<th>Accuracy rate/%</th>
<th>The former/%</th>
<th>The latter/%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No outbreak</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>No outbreak</td>
<td>II</td>
<td>II</td>
<td>II</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Small-scale outbreak</td>
<td>III</td>
<td>II</td>
<td>III</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Medium-scale outbreak</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Large-scale outbreak</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
vector is updated by the self-organizing learning process, which results in neuron rearrangement based on patterns in the input space. Therefore, the SOFM model can effectively achieve clustering and prediction of debris flow prediction levels.

Nonlinear regression methods and rainfall threshold methods may be more accurate and efficient when there is sufficient training data. In contrast, those methods may not be a good choice in areas of low-frequency debris flow that lack sufficient training data (Wang & Peng 2018). In this paper, 25 data sets with small sample size are verified, and the IIM-SOFM early warning model still has high accuracy, which indicates that the proposed method is suitable for early warning of low-frequency debris flow area. The IIM-SOFM early warning model shows good application prospects in this case and represents an accurate new method for early warning debris flow disasters.

In addition, the existing SOFM model results show that the network structure and training steps of SOFM have a significant impact on the prediction results (Kim et al. 2019; Mishra et al. 2019). The effect of the SOFM model network structure parameters on the early warning results is shown in Figure 9.

Number of dead neurons: The early warning accuracy results of the 5×3 SOFM algorithm and the 5×4 traditional SOFM algorithm under different training steps are shown in Figure 9. When the training steps are 200 and 500, the warning accuracy of both is 100%. Therefore, the training steps of the early warning model are 200 or 500, but the 5×4 traditional SOFM algorithm still has the possibility of early warning errors because of the non-uniqueness of the corresponding results of dead neurons and neurons (Yang et al. 2019).

Figure 9 and Table 5 are a comparison of the average number of dead neurons in the training process between the 5×3 SOFM algorithm and the 5×4 traditional SOFM algorithm. There are no dead neurons in the 5×3 SOFM algorithm and the winning neurons in each warning level are not repeated. The accuracy of the warning is 100%. There are five dead neurons in the 5×4 traditional SOFM algorithm, which account for 25%, and the tenth winning neurons correspond to the III and IV warning levels, which increases the possibility of incorrect model warning results (Pulakka & Kujanpaa 1998). Therefore, the optimal SOFM network structure parameter in this model is 5×3, and the training steps are 200 or 500.

Therefore, the method proposed in this study is feasible for early warning of debris flows. The IIF-SOFM model can provide reasonable assistance and prevent unnecessary debris flow disaster losses. However, from a statistical perspective, the red, orange, and yellow early warning levels are not absolute, and the early warning level of an area can change as debris flows develop. Thus, it is very necessary to continuously update the early warning indicator data through field investigation and monitoring to obtain real-time and accurate debris flow early warning levels.
In this study, a novel hybrid IIM-SOFM early warning model is proposed. Records of debris flow outbreak in Xiuyan County in the past 30 years are selected as an example, and a total of ten impact factors, including the hazardous conditions, material source conditions, dynamic conditions, and induced conditions, are selected for model prediction.

**Figure 9** The influence of network structure parameters on early warning results: (a) influence of network structure parameters of SOFM model on early warning; (b) 5*3 network structure winning neurons; (c) 5*4 network structure winning neurons.

**Table 5** Different network structure winning neurons of IIM-SOFM model

<table>
<thead>
<tr>
<th>Early warning level</th>
<th>Winning neuron number</th>
<th>Dead neurons</th>
<th>Winning neuron number</th>
<th>Dead neurons</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>15</td>
<td>0</td>
<td>20</td>
<td>5 (2,9,12,13,15)</td>
</tr>
<tr>
<td>II</td>
<td>3,14</td>
<td>0</td>
<td>17,18,19</td>
<td>5 (2,9,12,13,15)</td>
</tr>
<tr>
<td>III</td>
<td>8,9,13</td>
<td>0</td>
<td>10,11,16</td>
<td>5 (2,9,12,13,15)</td>
</tr>
<tr>
<td>IV</td>
<td>4,5,7</td>
<td>0</td>
<td>4,5,10,14</td>
<td>5 (2,9,12,13,15)</td>
</tr>
<tr>
<td>V</td>
<td>1,2,6,10,11,12</td>
<td>0</td>
<td>1,3,6,7,8</td>
<td>5 (2,9,12,13,15)</td>
</tr>
</tbody>
</table>

**CONCLUSION**

In this study, a novel hybrid IIM-SOFM early warning model is proposed. Records of debris flow outbreak in Xiuyan County in the past 30 years are selected as an example, and a total of ten impact factors, including the hazardous conditions, material source conditions, dynamic conditions, and induced conditions, are selected for model prediction.
The IIM method effectively reduced the dimensionality of the input factors and improved the stability and accuracy of the SOFM model. In the case of a small amount of data for a low-frequency debris flow area, the IIM-SOFM model exhibits good accuracy and stability and provides an effective method for early warning of debris flows.

The IIM-SOFM method has a simple structure, is simple to train, and the application results showed that the early warning accuracy rate is 100%. The IIM-SOFM model with optimized network structure parameters is obviously superior to the rainfall threshold method, BP neural network, and competitive neural network. This method has high accuracy and stability in a low-frequency debris flow area.

Therefore, it is feasible to use the IIM-SOFM model for early warning of debris flow. However, early warning for debris flows is affected by many factors, and it is difficult to form unified and accurate early warning models. This paper represents exploratory research, and further research must be done to deepen and improve this work to achieve better application effects.

DATA AVAILABILITY STATEMENT
All relevant data are included in the paper or its Supplementary Information.

REFERENCES
Kim, H. I., Keum, H. J. & Han, K. Y. 2019 Real-time urban inundation prediction combining hydraulic and probabilistic methods. Water 11 (2), 293.


First received 25 March 2020; accepted in revised form 16 June 2020. Available online 1 July 2020