

Research Article

Risk Assessment of Debris Flow in Huyugou River Basin Based on Machine Learning and Mass Flow

Jiazheng Li  and Yiqing Lv 

School of Mining Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Correspondence should be addressed to Yiqing Lv; lvqiqing@tyut.edu.cn

Received 12 July 2022; Revised 22 July 2022; Accepted 28 July 2022; Published 24 August 2022

Academic Editor: Muhammad Zakarya

Copyright © 2022 Jiazheng Li and Yiqing Lv. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The Huyugou river basin is a typical debris flow river basin in the Shanxi Province, which has great harm after the outbreak and seriously affects the safety of people's lives and property. Therefore, it is urgent to carry out debris flow risk assessment. In this paper, a machine learning algorithm is implemented to assess the disaster susceptibility of each branch gully in a river basin of the Huyugou. Furthermore, its high-susceptibility branch gully and main gully were selected as the starting points of debris flow simulation for numerical simulation. The machine learning algorithm is implemented in a cloud-edge platform to minimize the model training and prediction times. Under the simulated rainfall conditions of major debris flow disasters, e.g., the one that occurred in 1996, the accuracy rate reached 84%. The results show that the debris flow susceptibility of each branch gully in the study area is mainly affected by the peak flow rate of the river basin, the length of the main gully, and the relative height difference of the river basin. The total risk area of debris flow is $1.91 \times 10^5 \text{ m}^2$, and the high-risk area accounts for 52.18% of the total area. It is mainly located in the upper part of the main gully accumulation area and the confluence of each channel and the main gully. The middle-risk area accounts for 36.14% of the total area, and the low-risk area accounts for less. We also observed significant reduction, from 34.68% to 36.98%, in the training and prediction times of the machine learning models when implemented over the proposed edge-cloud framework. The reappearance of debris flow in the study area is relatively accurate, which provides a certain scientific basis for the risk assessment of debris flow in the future.

1. Introduction

The Shanxi Province is located in the west side of Taihang Mountain, in the middle of Loess Plateau, and the eastern edge of Ordos Basin. The structure is complex, the altitude gap is relatively large, and therefore many tragedies are likely to happen [1]. Although the incidence of debris flow disasters is far less than other disasters, it poses a serious threat to people's health and property safety, and the degree of risk is self-evident [2]. Similarly, the rise of state-of-the-art computational technologies such as big data, Internet of things, and cloud computing can enhance the safety through introducing some sort of monitoring system. With the growth of big data and storage technologies that are becoming more and more mature, the mining data acquisition becomes simple and convenient. However, the data analysis of mining disaster datasets is no longer limited to the simple

statistical analysis. In fact, machine learning approaches such as logistic regression, decision tree, and XGBoost algorithms become more essential. The main problem with the machine learning methods is the time required to train the model and then predict the disaster that should be minimized. The new concept of edge computing can solve the issues related to data analysis.

The Huyu River Basin in the Taiyuan City is the main river connected with the Xishan coalfield. In 1996, a major debris flow disaster occurred. The flood caused by rainstorm and the debris flow mixed together with solid deposits hit down, rushed through the coal mine and coal power group [3]. After the debris flow left the mountain area, it continued to move eastward along the street and finally entered the Fenhe River. The affected area was 15 km long from east to west, and the affected area was about 8 square kilometers. The disaster caused many people to be killed, and more than

100 people were trapped underground. The trapped time was very long. The direct economic loss caused by the damaged houses and roads is as high as CNY 240 million [4]. With the harm of debris flow to modern society, the problem of Huyugou has seriously affected the urbanization construction of the Taiyuan City, so it is of great significance to study the river basin quickly to monitor and prevent debris flow disasters [5, 6].

In this paper, based on the traditional geographic information system (GIS) evaluation system, combined with numerical simulation method, field investigation, and multiparty data analysis, the random forest (RF) method in machine learning was used to evaluate the susceptibility of debris flow in the study area [7]. Moreover, we also used other methods to assess the correctness and precision of the proposed system. The suggested RF method has been implemented in different modules that run on different layers of the edge computing model. On this basis, the numerical simulation of major debris flow, which occurred in 1996, was reproduced and compared with field data and the risk zoning was carried out to deliver practical basis for the monitoring, avoidance, and control of local debris flow tragedies. The major contributions of this study are as follows.

- (1) A machine learning algorithm is implemented to assess the disaster susceptibility of each branch gully in a river basin of Huyugou.
- (2) The high-susceptibility branch gully and main gully were selected as the starting points of debris flow simulation for numerical simulation.
- (3) The machine learning algorithm is implemented in a cloud-edge platform to improve the training and prediction times.

The rest of the paper is structured as follows. In Section 2, we give an overview of the study area that was used in this research. In Section 3, evaluation of the debris flow susceptibility is presented which is based on the random forest method. Risk assessment of the debris flow in Huyugou based on mass flow is deliberated in Section 4. Evaluation of the proposed methods, obtained findings, and discussion are given in Section 5. As a final point, Section 6 completes this study and offers directions for future research.

2. Overview of the Study Area

The debris flow in the study area is located in the southwest mountainous area of the west mountain in Wanbailin District, Taiyuan City, Shanxi Province. In fact, it is a small river basin within the scope of the Huyugou river basin, with an area of about 12.1695 km², and is dominated by low-middle mountains. The terrain is gradually reduced from west to east. The elevation of the main gully in the study area is about 1585.6 m, and the elevation at the gully mouth is about 1070 m, and the relative elevation difference is approximately 515.6 m. The study area is located in the interior of the continent, far from the ocean, and the monsoon climate is obvious. The maximum annual rainfall can reach

800 mm, and the spatial and temporal distribution of rainfall is uneven, mostly concentrated in summer. The rainfall can reach 80% of the total annual rainfall. The average temperature in the region is low, about 2°C~6°C, and the lowest temperature can reach -7°C, and the highest temperature can reach as high as 22.7°C. The study area is located in the Duerping-South Korea fault zone, with a length of about 26 km. The fault zone is formed by the Duerping fault and the Yayadi fault, toward northeast. The fault generally strikes northeast, which is a normal fault. This should be noted that the study area is distributed by Carboniferous, Permian clastic rocks, a small amount of Ordovician carbonate rocks, and Holocene gravel.

There are abundant sources in the basin, and the main solid source is the product of weathering of clastic rock layer. The solid material source comes from the loose accumulation caused by a large amount of slope instability that is subsequently caused by coal mining, road construction, bridge construction, and other activities in the region. Furthermore, it also includes domestic waste, cinder, and stone slag, which provide a large amount of material source for debris flow. In summary, the debris flow in Huyugou mining area has been an important area of debris flow disaster prevention and mitigation in the Taiyuan City. It is a major task to study and analyze, simulate, and promote the implementation of debris flow protection measures. Figure 1 shows a view of the river basin diagram of the study area.

3. Evaluation of Debris Flow Susceptibility Based on Random Forest Method

In this section, we first illustrate the proposed edge intelligence framework that is used to implement the machine learning algorithms. The main purpose of the edge computing is to bring computation closer to where the data is produced. In this way, the data can be preprocessed and can be fitted well for training purposes. The entire framework is shown in Figure 2 below. In the proposed framework there are three layers, namely, the IoT layer, the edge layer, and the cloud layer. The IoT sensors may include cameras and other data collection devices. Once the data is gathered, it could be preprocessed over the edge devices because the IoT devices have very low processing capabilities. The preprocessing may include data aggregation methods that can remove duplicate and unnecessary data. This duplication may occur when data from overlapping regions are collected. This should be noted that, due to (i) no availability of duplicate entries and (ii) small size of the dataset, we do not use any aggregation technique in this work. Largely, the well-known Euclidean distance equation is used to identify whether two particular collected data points (through sensors) belong to either the same region or two different regions, which is used for data aggregation purpose [8, 9]. The processed data is then moved to the cloud for long-term storage. It should be noted that, in the proposed framework, machine learning algorithms can be used in three different manners: (i) perform the prediction at the edge; (ii) perform the prediction at the cloud; and (iii) train the model on cloud and perform prediction on the edge [10]. However, in case of (i) different algorithms

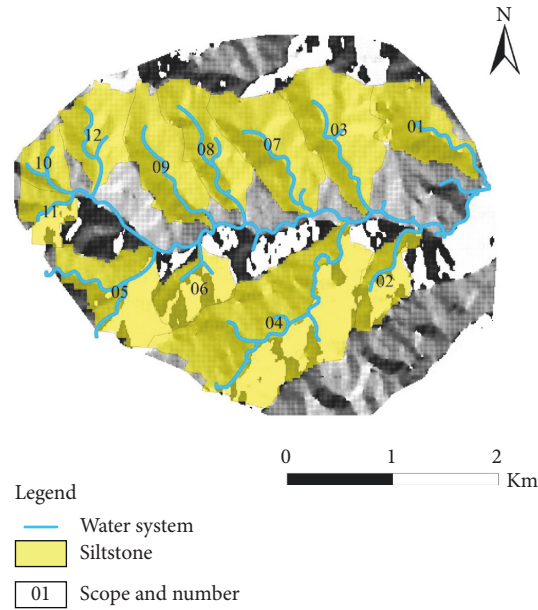


FIGURE 1: River basin diagram of the study area.

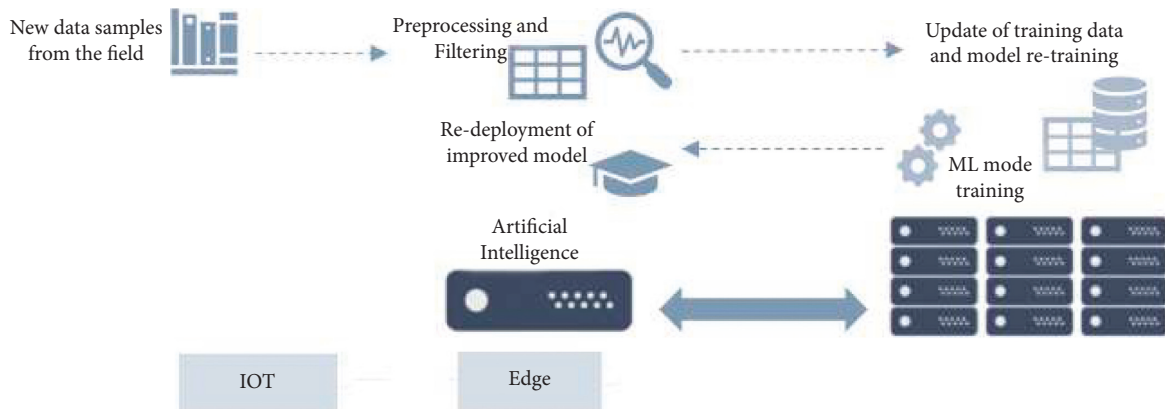


FIGURE 2: Edge computing for machine learning.

have different computational times and it might not be possible for the edge (limited resources) to compute quickly. In the case of (ii), networks are the bottlenecks and it will take quite long time, dependent on the data size, to do predictions. In the case of (iii), the model is trained at regular intervals to make sure that prediction outcomes are more accurate.

Figure 3 shows the flow of data between the edge and cloud in terms of machine learning. The lower part illustrates the scenario when edge computing is not taken into account. This type of setup might be helpful in offline learning, but for real-time online learning this might not be a good option. The upper part describes two situations: (i) when machine learning methods are used over the stored data while preprocessing happens at the edge and (ii) when machine learning is used over the reproduced data over the cloud and the preprocessing along with data aggregation method is used at the edge. The machine learning algorithm is then run in two different modules. The first module is the training that

runs on the cloud. In case that enough data is not available, then more data can be produced through synthesized workloads [11]. Also, the IoT sensors continuously collect data and send it to the edge for preprocessing. Subsequently, the processed data is moved to the cloud for training purposes. The second module runs in the edge and predicts the unseen situations based on the data stored and trained model. It should be noted that, to reduce the training time, the amount of data can also be reduced through data aggregation techniques such as Euclidian distance. In this work, we do not suggest any data reduction mechanism.

3.1. Principle of Random Forest Method. The RF (Random Forests) is one of the most popular algorithms used to solve multiclassification and prediction problems [12–14]. It is an integrated method of binary decision trees trained independently. It was introduced by Breiman in 2001 and combines multiple decision trees used for classification and

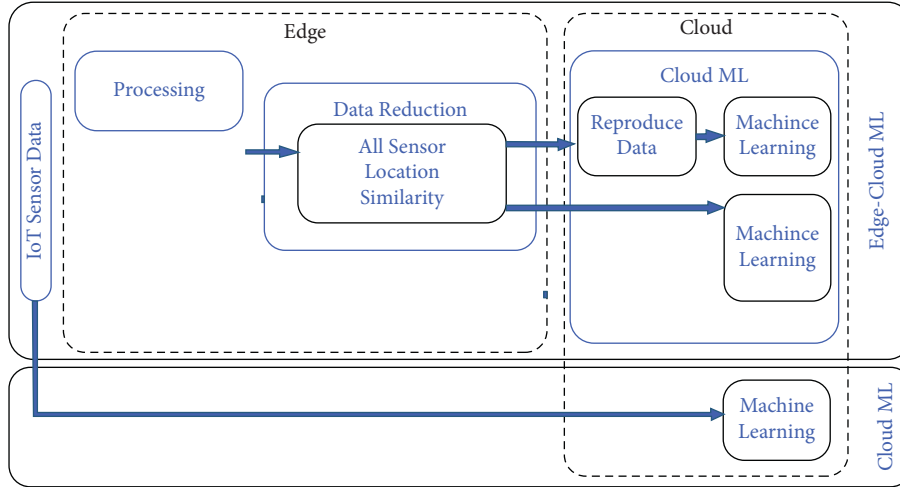


FIGURE 3: Edge-cloud computing framework for machine learning.

prediction. It has obvious effect in classification and regression problems [15, 16]. The RF can be defined as a set of random trees (decision trees). The basic method for classification problems is based on training each decision tree alone, and the final result is estimated by considering the results obtained by each decision tree.

The random forest algorithm works as follows:

- (1) Resample the original data and repeat it several times.
- (2) In each resampling process, a group of disaster-pregnant factors are randomly selected as the eigenvalues.
- (3) The resampling and the corresponding eigenvalue of the disaster-pregnant factor are estimated to obtain the decision tree set.
- (4) Aggregate the estimated decision tree set in order to obtain a single decision tree.

Therefore, the basic notion of the RF procedure is to generate multiple decision trees on a random subset [17]. In fact, the performance of the suggested RF method predominantly depends on the amount of decision trees (Ntree), as well as the candidate features that are enclosed in the subset (mtry) [18]. It should be kept in mind that larger Ntree values may potentially increase modeling time, while the smaller Ntree values may cause prediction errors. The RF model can summarize and minimize the risk of overfitting without any pruning process. The training process involves creating many different boot samples from the original data set, one-third of which is excluded from the process as test cases, and based on this test case to estimate unbiased test error, known as out-of-bag-error, which represents the predictive ability of the RF model [19]. For the purpose of classification, the RF model uses the high variance between individual trees. This is achieved by voting each tree as a class member and allocating the corresponding class value according to the public vote. Furthermore, the RF classifier is more accurate and robust than a single classifier, because it has many advantages; for example: (i) it can handle large

databases relatively very effectively, and (ii) it offers a way to calculate the proximity between pairs of cases used to locate outliers, etc. [20, 21].

The RF algorithm also uses the Gini index as the attribute selection metric to measure the purity of attributes and classes. Assuming that the sample R corresponding to the characteristic index in the data preprocessing set R^* contains J categories, then its Gini index is given by the following equation [22]:

$$\text{gini}(T) = 1 - \sum_{j=1}^N p_j^2, \quad (1)$$

where p_j is the probability of the j^{th} sample. After one segmentation, the set R^* is divided into m parts $\{N_1, N_2, \dots, N_m\}$. Then, the segmented Gini index $\text{gini}_{\text{split}}(T)$ is given by

$$\text{gini}_{\text{split}}(T) = \frac{N_1}{N} \text{gini}(T_1) + \dots + \frac{N_m}{N} \text{gini}(T_m). \quad (2)$$

The final $\text{gini}_{\text{split}}(T)$ is the Gini index corresponding to each feature sample, and its set is set as $G = \{g_1, g_2, \dots, g_j\}$. The weight corresponding to each feature index is illustrated using

$$\theta_j = \frac{g_j}{\sum_{i=1}^j g_i}. \quad (3)$$

Finally, the weight set is $\theta = \{\theta_1, \theta_2, \dots, \theta_j\}$.

3.2. Influence Factors of Debris Flow Susceptibility. The initiation of debris flow is caused by many factors such as precipitation, topography, geomorphology, and human factors. In this paper, the selection of debris flow factors is mainly considered in the above aspects. The following factors affecting the development of debris flow are selected: river basin area, average slope of the river basin, shape coefficient, channel length, longitudinal shrinking slope of the main gully, relative height difference of the river basin, rainfall, vegetation coverage rate, and the peak flow of the

TABLE 1: The actual values of each factor in the study area.

Number	River basin area (km ²)	Average slope of the river basin (°)	Shape factor	Channel length (m)	Longitudinal shrinking slope of the main gully (‰)	Relative height difference (m)	Rainfall (mm)	Vegetation coverage (%)	Peak flow (m ³ /s)	Strata lithology
1	0.527	19.61	1.1	857	164.53	349	439.462	15.4401	4.39	Siltstone
2	0.324	15.64	1.18	917	186.48	280	439.604	21.7935	2.85	Siltstone
3	0.786	16.98	1.23	1321	146.86	336	439.34	19.8264	6.97	Siltstone
4	1.794	16.79	1.17	2714	64.11	339	439.489	22.7433	17.93	Siltstone
5	0.623	19.72	1.13	1024	87.89	229	439.205	37.3686	5.53	Siltstone
6	0.351	19.69	1.30	484	161.16	255	439.319	24.7348	3.27	Siltstone
7	0.721	18.35	1.29	1225	130.61	342	439.359	21.793	8.80	Siltstone
8	0.532	15.83	1.09	1432	201.12	371	439.155	19.9189	4.60	Siltstone
9	0.585	18.30	1.07	1328	164.16	374	439.11	19.8484	4.06	Siltstone
10	0.211	17.60	1.31	580	193.10	180	438.935	35.3991	1.33	Siltstone
11	0.226	19.09	1.21	526	62.74	167	438.993	36.2723	2.27	Siltstone
12	0.452	15.55	1.22	1005	150.25	285	438.972	29.2496	4.02	Siltstone

river basin [23]. The actual values are detailed and given in Table 1.

Although the scope of the study area is small and the rainfall is basically the same, in order to ensure the integrity of the factor selection, the stratigraphic lithology is still listed in Table 1. The clear water flow of each river basin is calculated by the debris flow clear water flow formula, which is given by

$$Q_b = 0.278Fri, \quad (4)$$

where Q_b represents the clear water flow in the region (m³/s); F represents the river basin area (km²); i is the production flow coefficient and its value is assumed as $i=0.9$; and r represents hourly surface rainfall (mm/h).

The critical rainfall value of debris flow within 24 hours in Shanxi Province is about 30 mm [24]. According to the characteristics of Huyugou climate and the analysis of rainfall in Taiyuan City, it is concluded that the daily rainfall should be approximately 120 mm/d when Huyu gully triggers severe rainstorm [25, 26].

The calculation formula of peak flow Q_c in debris flow basin is given by

$$Q_c = (1 + \varphi)Q_bD_c, \quad (5)$$

where φ represents the sediment coefficient of the basin; and D_c represents the blockage coefficient in the basin.

Thus, the peak flow in each river basin of debris flow can be obtained.

According to the geological hazard risk assessment standard and related research results [27, 28], the factors are divided into four levels: high (IV), middle (III), low (II), and very low (I), and the classification results are substituted into the random forest method to calculate the weight. The grading standards and weight calculation results are shown in Table 2, and the grading results are shown in Figure 4.

3.3. Results of Debris Flow Gully Susceptibility. As shown in Figure 5, after calculation, the zoning results are as follows: (i) No. 4 and No. 7 are high-prone debris flow branch gullies;

(ii) No. 1, No. 3, and other seven branch gullies are high-prone debris flow branch gullies; and (iii) No. 2, No. 10, and other three branch gullies are low-prone debris flow branch gullies.

4. Risk Assessment of Debris Flow in Huyugou Based on Mass Flow

According to the evaluation results of debris flow susceptibility, area 4, area 7, and main gully in the high-susceptibility area are selected for evaluation.

4.1. Simulation Parameter Value and Working Condition Design

4.1.1. Unit Weight of the Debris Flow. The determination of unit weight can be roughly divided into three methods, namely: (i) field investigation method, (ii) morphological investigation method, and (iii) standard look-up table method. The debris flow severity used in this numerical simulation is mainly determined by field investigation method that can be mathematically expressed as follows:

$$\gamma_c = \frac{G_c}{V}, \quad (6)$$

where γ_c is heavy debris flow fluid (t/m³); G_c is slurry quality (t); and V is the mud volume (m³).

As shown in Table 3, the field method is used to investigate the density of debris flow. The slurry is mixed at the upstream of channel, middle and lower reaches of the channel, and the exit of the channel in the study area, respectively. Multiple experiments are carried out and the average value is finally obtained. The comprehensive analysis shows that the average unit weight of debris flow in the study area was $\gamma_c = 1.602 \text{ t/m}^3$, and the density is moderate, belonging to rare debris flow. At the same time, according to the morphological investigation method, the fluid and motion characteristics of debris flow are described by the affected villagers [29]. It is concluded that the fluid properties of debris flow should be between dense debris flow and

TABLE 2: Evaluation factor weights and development grades.

Impact factors	High (IV)	Middle (III)	Low (II)	Very low (I)	RF weight
River basin area	0.2~5	5~10	10~100	>100	0.082375
Average slope of the river basin	>32	32~25	25~15	<15	0.090375
Shape coefficient	<1.1	1.1~1.2	1.2~1.3	>1.30	0.105375
Channel length	>10	10~5.0	5~1	<1	0.180375
Longitudinal shrinking slope of the main gully	>200	200~150	150~100	<100	0.103375
Relative elevation difference	>500	500~300	300~100	<100	0.118375
Vegetation coverage	<10	10~30	30~60	>60	0.080375
Peak flow	>10	10~6.85	6.85~3.36	<3.36	0.239375
Value range	(0.75, 1)	(0.5, 0.75]	(0.25, 0.5]	(0, 0.25]	

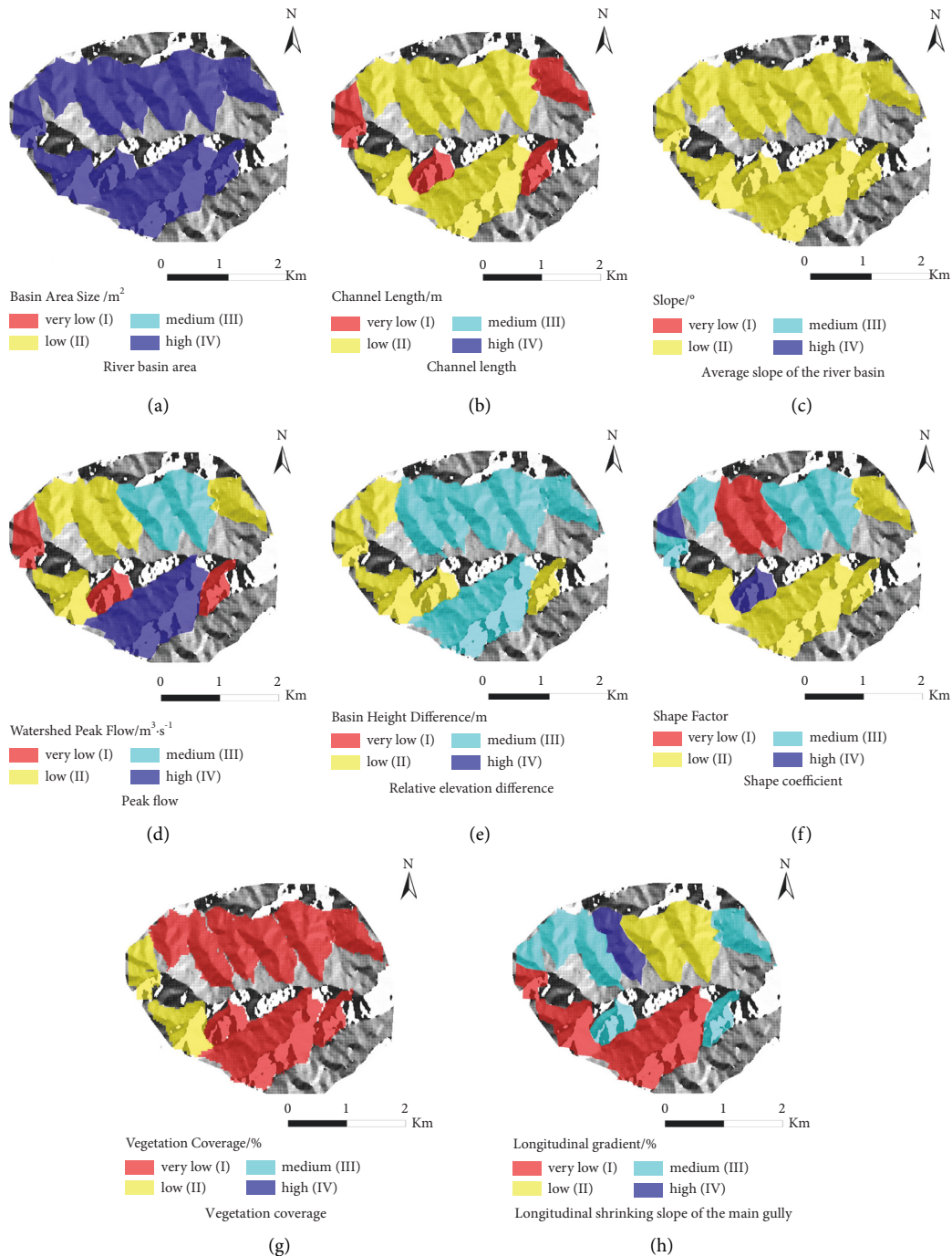


FIGURE 4: Quantitative classification and grading results of factor susceptibility. (a) River basin area; (b) channel length. (c) Average slope of the river basin; (d) peak flow. (e) Relative elevation difference; (f) shape coefficient. (g) Vegetation coverage; (h) longitudinal shrinking slope of the main gully.

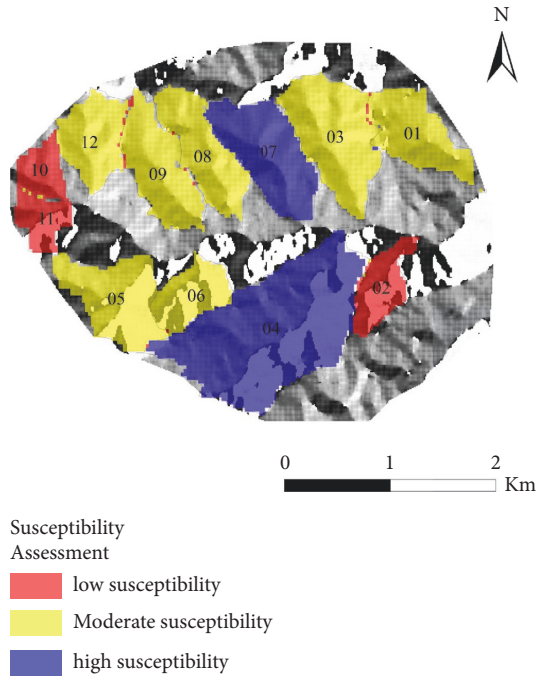


FIGURE 5: Evaluation results of debris flow susceptibility in the study area.

TABLE 3: Calculation of severe field slurry mixing for debris flow in the research area.

Test position	Mud weight (G_c/kg)	Mud volume (V/L)	Severe debris flow ($\gamma_c/t \cdot m^{-3}$)
The upstream of the channel	78.6	50	1.572
The middle and lower reaches of the channel	47.9	30	1.605
The exit of the channel	48.9	30	1.630

dilute debris flow; that is, the unit weight is $1.60 t/m^3$, indicating that the field experiment is accurate.

4.1.2. Debris Flow and Flow Process Line. The clear water flow and debris flow in the river basin have been calculated, respectively, in the above factor calculation, which is not described here. The method used in this simulation is the generalized pentagon theory with high recognition. The method is to take $1/3$ of the complete debris flow time as the node, and the peak flow calculated above is substituted into the boundary point with $1/3$ and $1/4$, respectively, so as to describe the flow process line of debris flow outbreak [30].

4.2. Modeling Results. Figure 6 is the simulation results of the debris flow movement process in the study area under the condition of actual rainfall. Figure 7 shows that the maximum velocity of debris flow is $5.53 m/s \sim 6.41 m/s$ and the maximum mud depth is $5.1 m \sim 6.5 m$ under the condition of major debris flow rainfall in Huyugou in 1996, which is located in the middle and upper part of the gully

accumulation area and the confluence of each channel and the main gully. Note that the measured total risk area is approximately $2.28 \times 10^5 m^2$, the numerical simulation risk area is $1.91 \times 10^5 m^2$, and the accuracy is about 84%.

4.3. Risk Assessment of Debris Flow. According to the study of Xu [31] in the Shanyang County (Table 4), the hazard zoning of the debris flow in Huyugou in 1996 is carried out, as shown in Figure 5. The results of debris flow hazard evaluation show that the total area of debris flow hazard zone is $1.91 \times 10^5 m^2$, and the high hazard zone accounts for 52.18% of the total area, which is mainly located in the downstream of the main gully and the intersection of the branch gully and the main gully. Furthermore, the area of medium-risk area is $0.69 \times 10^5 m^2$, accounting for 36.14%, and the low-risk area is relatively small. In general, the study area is a relatively dangerous debris flow, which needs strict prevention.

4.4. Machine Learning and Edge-Cloud Results. In this section, we discuss the results of the machine learning techniques and the training and prediction model were supposed to run on different platforms. From the algorithm perspective, we consider two different machine learning algorithms, namely: (i) random forest (RF) and (ii) CNN. Each algorithm runs in two phases: (i) training and (ii) prediction. From the platform perspective, we use different scenarios. In scenario A, we assume that both phases of each algorithm run over the edge. In this case, since the data is stored on the cloud, we assume that the required data is moved to the edge. Once the data is used, it is deleted from the edge server. In scenario B, we assume that both phases run on the cloud. In scenario C, we assume that the training happens on the cloud while the prediction runs over the edge server. We report the timing durations for the training and prediction phases [32]. The results are illustrated in Table 5.

The findings suggest that, for various algorithms, the response time can be significantly decreased (i.e., from 24.64% to 33.24%) using the proposed cloud-edge platform. Furthermore, we also noted approximately 34.68% to 36.98% reduction in the prediction durations. This improvement is possible at some cost of prediction duration. Furthermore, we observed the RF method outperforms the classical CNN approach (i.e., $\sim 25.54\%$), but we believe these outcomes will change in line with the amount of data.

5. Discussion and Model Accuracy

In this section, we briefly discuss the findings of this research and accuracy of the machine learning methods. After verification, this paper gets the following conclusions and understanding:

- (1) The debris flow susceptibility of each branch gullies in the study area is mainly controlled by the peak flow rate of the river basin, the length of the main gully, and the relative height difference of the river basin. There are 12 branch gullies, 2 high-prone

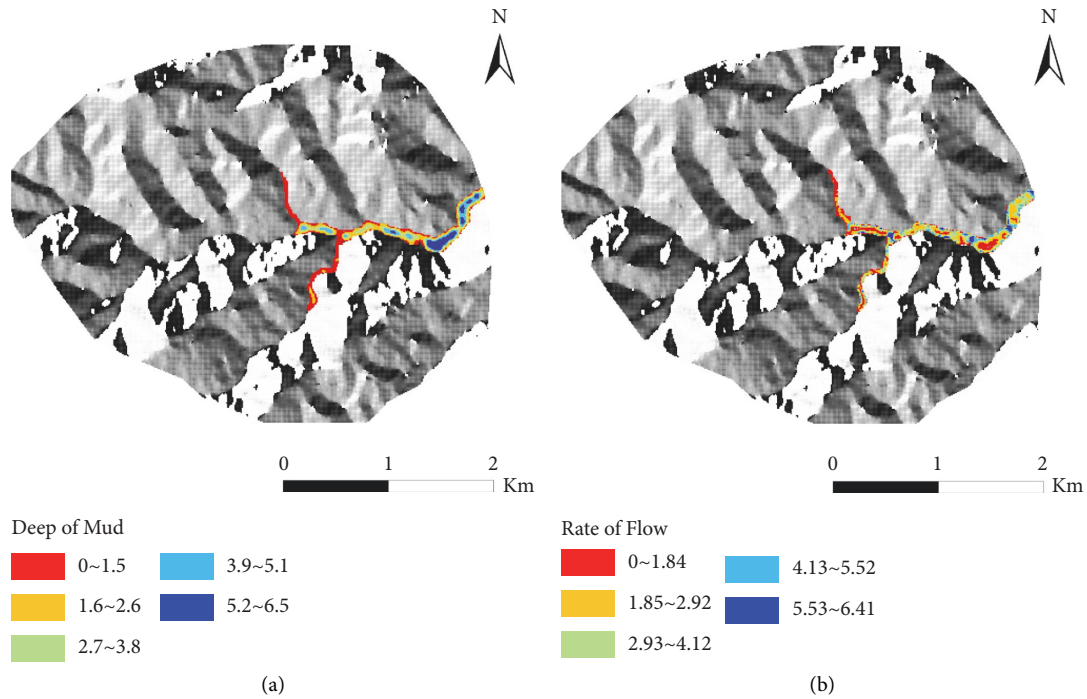


FIGURE 6: Mud depth and velocity of debris flow in the study area.

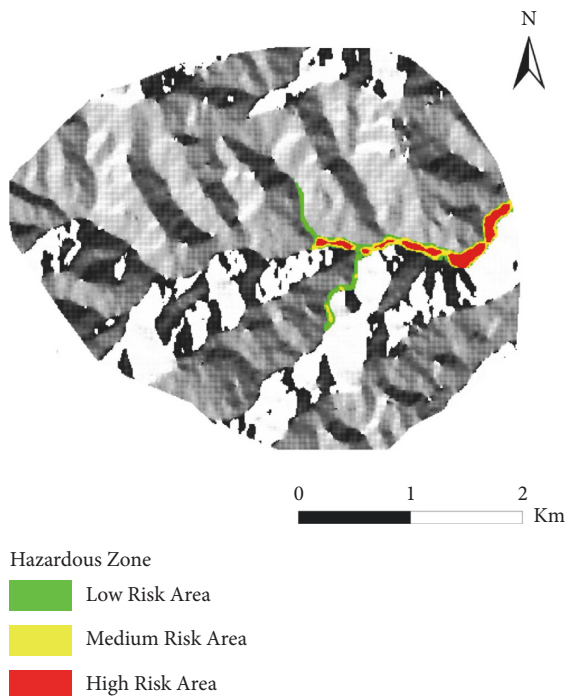


FIGURE 7: Risk zoning map of debris flow in the study area.

TABLE 4: Hazard zoning parameters of the debris flow.

Dangerousness	Depth of mud	Association	Mud depth and flow velocity product
High	$H \geq 3.0$	Or	$VH \geq 3.0$
Middle	$0.5 \leq H < 3.0$	And	$0.5 \leq VH < 3.0$
Low	$0 < H < 0.5$	And	$0 < VH < 0.5$

branch gullies, 7 middle-prone branch gullies, and 3 low-prone branch gullies in the region.

- (2) Through the previous multifactor superposition analysis and parameter calculation, the motion state of the study area is reproduced by numerical simulation. The simulation results show that the mud depth of debris flow at the accumulation of gully mouth and the intersection of gully and main gully in the study area is the largest, about 6.5 m, and the maximum velocity is 6.41 m/s at the middle and lower reaches of the gully and the steep terrain. By testing the goodness of fit of the simulation results, the accuracy is about 84%. The high-risk areas of debris flow in the study area accounted for 52.18%.

The return accuracy of debris flow in the study area under the condition of heavy debris flow rainfall in 1996 is relatively close, which provides corresponding scientific suggestions for the comprehensive evaluation and risk zoning of debris flow in the future. The experimental findings were assessed using different evaluation metrics, i.e., (i) precision or accuracy, (ii) recall rate, (iii) F1-measure, and (iv) IoU. In fact, accuracy is the proportion of correctly forecasted samples to all predicted samples. The recall rate is calculated as the proportion of accurately anticipated positive samples to all real positive samples. Moreover, the F1 score is the harmonic average of recall rate and precisions (accuracy). Finally, the IoU is the crossing of pixels labelled as building in the ground truths and anticipated outcomes and subsequently divided by the union of pixels labelled as building in the ground truths and forecasted outcomes [8]. The following are the calculating formulas:

TABLE 5: Results of the algorithm and implementation methodology

Algorithm	Scenario	Time (seconds)		
		Training	Prediction	Response
RF	A	3532.1	176.44	11.12
	B	3721.69	46.31	388.42
	C	3387.01	61.2	2.74
CNN	A	3679.3	169.84	11.07
	B	3861.42	51.58	402.39
	C	3495.3	62.5	3.68

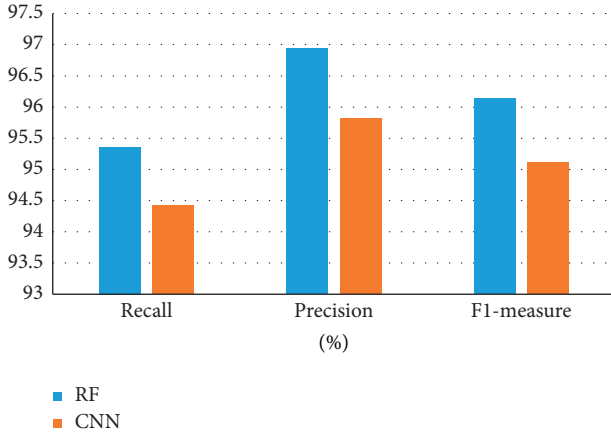


FIGURE 8: Accuracy results of the machine learning approaches (the higher values represent a more accurate approach).

$$\begin{aligned}
 P_r &= \frac{TP}{(TP + FN)}, \\
 P_p &= \frac{TP}{(TP + FP)}, \\
 F_1 &= 2 \cdot \frac{P_p \cdot P_r}{(P_p + P_r)}, \\
 \text{IoU} &= \frac{TP}{TP + FP + FN},
 \end{aligned} \tag{7}$$

where TP stands for the quantity of correctly taken out pixels, FP for the quantity of incorrectly pulled out pixels, and FN for the amount of lost or misplaced pixels. The accuracy of the RF and CNN methods is shown in Figure 8. We can observe that the RF method is more accurate than the CNN approach in terms of all evaluation metrics.

6. Conclusions and Future Work

Based on the investigation of debris flow disasters in the distribution areas in Shanxi Province, this paper selects the debris flow in the study area as a representative river basin for analysis to explore a relatively reasonable evaluation method of debris flow in the Loess Plateau, especially in Shanxi Province. The method in this paper is mainly based on the weight calculation of the random forest method and

the combination of multifactor superposition and numerical simulation. Through the evaluation of various factors in the river basin, namely, rainfall, topography, and geomorphology, the susceptibility of debris flow in each channel in the region is evaluated, and it is used as the main material source of debris flow. Numerical simulation is combined with the results of multifactor analysis to simulate the movement characteristics of debris flow under this condition and carry out risk zoning. The two complement each other, and the evaluation of debris flow has a more detailed process. The results are more reasonable than a single way.

In the future, we will take into account deep learning techniques that are more suited for mines and the operational monitoring systems, like graph convolutional network (GCN), U-net, and attention networks. But as we saw, not all neurons can be stimulated by the activation function used in this paper, which results in restricted precision and accuracy. As a result, finding the best activation function and improving the model's structure are ongoing research projects. Similar to this, we will look into the effects of the activation functions employed in conjunction with deep learning techniques. To enhance the performance of the suggested system, robust data reduction or aggregation approaches should be looked at.

Data Availability

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also form part of an ongoing study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. C. Huang, "Research on distribution law and influencing factors of major geological hazards in Shanxi," *Energy and Environmental Protection*, vol. 42, no. 1, pp. 87–90, 2020.
- [2] J. J. Yu and Q. Li, "The hazard and prevention of debris flow," *World Non-ferrous Metals*, vol. 1, pp. 204–205, 2019.
- [3] Q. L. Wang, Y. Q. Lv, and M. C. Yuan, "Characteristics and risk assessment of debris flow in Huyugou," *Soil and Water Conservation Bulletin*, vol. 30, no. 6, pp. 219–222, 2010.
- [4] R. P. Qin and G. H. Zhang, "The formation process and development trend of "1996-08-04" debris flow in Huyugou, Taiyuan," *Chinese Journal of Geological Hazards and Prevention*, vol. 15, no. 4, pp. 87–90, 2004.
- [5] H. W. Zhang, *3D Numerical Simulation of Debris Flow Field in Huyugou, Xishan. Master Thesis*, Taiyuan University of Technology, Taiyuan, China, 2013.
- [6] Y. L. Gao, B. C. Li, and J. R. Yang, "Formation conditions and prevention of debris flow in Huyugou, xishan, taiyuan city," *Shanxi Architecture*, vol. 14, pp. 89–90, 2005.
- [7] S. Y. Wen and Y. Li, "Research status and development review of geological disaster prevention and mitigation technology," *Geology and Resources*, vol. 23, no. 3, pp. 296–300, 2014.
- [8] Y. Wang, J. Kong, and H. Zhang, "U-net: a smart application with multidimensional attention network for remote sensing

- images,” *Scientific Programming*, vol. 202211 pages, Article ID 1603273, 2022.
- [9] Y. Zhang, F. He, J. Kong, Y. Zhu, and L. Wang, “Relationship between surface subsidence range and geological mining conditions using numerical simulation and machine learning,” *Scientific Programming*, vol. 2022, Article ID 8720831, 12 pages, 2022.
- [10] J. Kong, Y. Gao, Y. Zhang, H. Lei, Y. Wang, and H. Zhang, “Improved attention mechanism and residual network for remote sensing image scene classification,” *IEEE Access*, vol. 9, no. 2021, Article ID 134800.
- [11] H. Yan, J. Zhang, S. Zhang, and N. Zhou, “Physical modeling of the controlled shaft deformation law during the solid backfill mining of ultra-close coal seams,” *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 5, pp. 3741–3754, 2019.
- [12] F. D. Santana, A. D. Souza, and R. J. Poppi, “Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters,” *Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy*, vol. 191, no. 1, pp. 454–462, 2017.
- [13] J. C. Quiroz, N. Mariun, M. R. Mehrjou, M. Izadi, N. Mison, and M. A. Mohd Radzi, “Fault detection of broken rotor bar in LS-PMSM using random forests,” *Measurement*, vol. 116, pp. 273–280, 2018.
- [14] D. Tsagkrasoulis and G. Montana, “Random forest regression for manifold-valued responses,” *Pattern Recognition Letters*, vol. 101, pp. 6–13, 2018.
- [15] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [16] M. Belgiu and L. Drăguț, “Random forest in remote sensing: a review of applications and future directions,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, 2016.
- [17] C. Nguyen, Y. Wang, and H. N. Nguyen, “Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic,” *Journal of Biomedical Science and Engineering*, vol. 06, no. 05, pp. 551–560, 2013.
- [18] H. Zhang and M. Wang, “Search for the smallest random forest,” *Statistics and Its Interface*, vol. 2, no. 3, pp. 381–388, 2009.
- [19] H. Ebata, A. Yamamoto, Y. Tsuji et al., “Persistent random deformation model of cells crawling on a gel surface,” *Scientific Reports*, vol. 8, no. 1, pp. 5153–5165, 2018.
- [20] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, pp. 93–104, 2012.
- [21] X. Yang, G. Wen, L. Dai, H. Sun, and X. Li, “Ground subsidence and surface cracks evolution from shallow-buried close-distance multi-seam mining: a case study in Bulianta coal mine,” *Rock Mechanics and Rock Engineering*, vol. 52, pp. 2835–2852, 2019.
- [22] Y. B. Wang, T. Sun, and C. Chen, “Application of random forest and fuzzy identification coupling method in river health assessment,” *Jiangsu Water Conservancy*, vol. 5, pp. 25–29, 2019.
- [23] G. Zhang, G. Guo, W. Wei, S. Zhang, and G. Yao, “A new method to control the regional strata movement of super-thick weak cementation overburden in deep mining,” *Stavební obzor-Civil Engineering Journal*, vol. 31, pp. 153–168, 2022.
- [24] B. Y. Tan, “Research on rainstorm debris flow forecast,” *Journal of Soil and Water Conservation*, vol. 4, no. 4, pp. 14–20, 2002.
- [25] X. F. Bai, *Study on Induction Mechanism of Debris Flow in Huyugou*, Taiyuan University of Technology, Shanxi, Taiyuan City, 2011.
- [26] Z. Liang, C. M. Wang, Z. M. Zhang, and K. U. J. Khan, “A comparison of statistical and machine learning methods for debris flow susceptibility mapping,” *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 1887–1907, 2020.
- [27] Gb/T 40112, *Geological Hazard Risk Assessment Specification*, Chinese Standard Press, Beijing, China, 2021.
- [28] X. Y. Hu, W. S. Qin, and Q. Dou, “Debris flow sensitivity analysis based on GIS and random forest model, taking the northern mountainous area of Taonan City, Jilin Province as an example,” *Soil and Water Conservation Bulletin*, vol. 39, no. 5, pp. 204–210, 2019.
- [29] Y. Zhao, X. Meng, T. Qi et al., “AI-based rainfall prediction model for debris flows,” *Engineering Geology*, vol. 296, Article ID 106456, 2022.
- [30] T. J. Liu, S. Q. Sun, Z. Zhao, and X. Zhang, “Effect evaluation of debris flow engineering in Lengzhang gully based on Massflow model,” *Water Conservancy and Hydropower Technology*, vol. 51, no. 10, pp. 195–201, 2020.
- [31] J. W. Xu, *Study on Risk Assessment of Debris Flow in Mountainous Areas. Master Thesis*, Changan University, Xi’an, China, 2014.
- [32] J. Sun, S. Qin, S. Qiao et al., “Exploring the impact of introducing a physical model into statistical methods on the evaluation of regional scale debris flow susceptibility,” *Natural Hazards*, vol. 106, pp. 881–912, 2021.