Evaluation of Highway Collapse Hazard Based on Rough Set and Support Vector Machine

Hujun He (hsj2010@chd.edu.cn)
Chang'an University https://orcid.org/0000-0003-2116-0161
Guorong Quan
Chang'an University
Haolei Zhu
Chang'an University
Wei Li
Chang'an University
Rui Xing
Chang'an University
Yichen Zhao
Chang'an University

Research Article

Keywords: highway collapse, hazard evaluation, rough set, support vector machine

Posted Date: January 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1248387/v1

Evaluation of highway collapse hazard based on rough set and support vector machine

Hujun He[1,2,*], Guorong Quan[1], Haolei Zhu[1], Wei Li[1], Rui Xing[1],Yichen Zhao[1]

[1]School of Earth Science and Resources, Chang'an University, Xi'an 710054, China

[2]Key Laboratory of Western Mineral Resources and Geological Engineering, Ministry of Education, Xi'an 710054, China

*Correspondence should be addressed to Hujun He; hsj2010@chd.edu.cn

Abstract: The prediction of possibility and risk classification of collapse is an important issue in the process of highway construction in mountain area. Based on the principle of rough set and support vector machine, a landslide hazard prediction model was established. First of all, according to field investigation, an evaluation index system and a sample set of evaluation index data were established, the rough set decision table was constructed by preprocessing the original data based on the function classification of standard evaluation index, and then, the influence indexes of the collapse activity were reduced by rough set theory, and the main 9 indexes affecting the collapse activity as the key discriminant factors of support vector machine model, namely slope shape of slope, aspect of slope, slope of slope, height of slope, exposed structural face, stratum lithology, relationship between weakness face and free face, vegetation cover rate and weathering degree of rock were extracted. Then, taking the data of 13 post earthquake collapses provided by the author etc in Yingxiu-Wolong highway of Hanchuan County as training samples, the optimal model parameters were analyzed and calculated. When the penalty parameter C is 8 and the kernel parameter g is 0.5, the correct rate of cross-validation is 100%, and the model is optimal. At last, 4 other landslide data were tested, the discriminant results of the test sample data were compared with the results obtained by uncertainty measure and distance discriminant analysis. The results show that the discriminant results of the test sample data by RS-SVM were consistent with the results obtained by uncertainty measure and distance discriminant analysis, the accurate rate is 100%. The collapse hazard analysis model based on rough set and support vector machine can reduce the computation while ensuring the accuracy of evaluation, and better solve the small sample and nonlinear problems, can provide certain a good idea for collapse hazard evaluation in the future.

Key words: highway collapse; hazard evaluation; rough set; support vector machine

1. Introduction

Collapse is a geological phenomenon in which the rock and soil mass on a steep slope suddenly separates from the parent body under the action of gravity and other external forces to fall, roll and accumulates in a valley (or slope foot) (Yue 2008; Liu 2010; Li 2011; Qiu 2012). With the deepening of the reform and opening-up and the implementation of the western development strategy, highway, railway and other projects continue to extend to the mountain area, high and steep slopes are more and more, so slope collapse and rockfall disasters are also increasing. In particular, the high and steep slopes after an earthquake, which are often unfavourable for earthquake resistance, are prone to collapse and rockfall disasters. For example, the 5.12 earthquake (2008 Wenchuan earthquake) triggered a large number of collapses, seriously damaging the transportation infrastructure such as highways etc., has greatly affected the disaster area people's production and life and earthquake relief work. Therefore, it will provide strong support for regional geological disaster assessment and sustainable development of geological environment to carry out the evaluation and prediction of highway collapse geological disaster by information, quantification and science.

Because of the numerous and uncertainty factors that affect the collapse activity, many methods for predicting and evaluating the collapse geological hazards have emerged. For example, Liu (2010) applied the analytic hierarchy process (AHP) and the fuzzy comprehensive evaluation method to evaluate the risk of collapse disaster, Zhang et al. (2009) used the factor weighted summation model of the improved analytic hierarchy process to evaluate the sensitivity of landslides induced by the earthquake of Beichuan County, Xue et al. (2011) proposed the risk evaluation model of collapse disaster based on extension theory and fuzzy theory, Gao et al. (2006) constructed a landslide collapse risk assessment model based on GIS and information quantity model, He et al, (2013, 2017) established a comprehensive evaluation model of collapse hazard based on uncertainty measure and an information entropy and distance discriminant analysis model for predicting the grade of collapse hazard, Liu (2016) used Newmark displacement calculation model and probability method etc. to evaluated the risk of landslide induced by volcanic eruption of Changbai mountain pool in the sky. Broeckx et al. (2018), Greco and Sorriso-Valvo (2013), Mandal and Mondal (2019), Yang et al. (2021) used logistic regression model to evaluate and predict the susceptibility of landslides in the study area. Feng (2019) used the weight of evidence model, information model and logistic regression model to evaluate the susceptibility of landslides in Shilou-Jixian section of the middle and lower reaches of the Yellow River based on ArcGIS platform etc..

However, with the deepening of the research, it is found that the indexes affecting the activity of highway collapse are quite complex, which include the internal characteristics of the collapse body itself, such as elevation, slope direction, slope, stratum lithology, exposed structural face, soil type, etc., there are also external factors such as groundwater, precipitation, rock weathering, earthquake and various human activities that induce collapse disasters. Some of these indexes are redundant and have nothing to do with the evaluation results. When the above-mentioned mathematical theory method was used for evaluation, the attribute importance degree of several indexes was not analyzed in order to optimize the evaluation indexes, especially, as a non-linear, multi-level, fuzzy and complex system, and the conditions of highway collapse are different in different areas, it is difficult to obtain the complete collapse index data. As a pattern recognition method of minimize structural risk based on statistical theory, support vector machine (SVM) can deal with the objective and practical problems such as small sample or finite sample, non-linearity and so on. Therefore, in this paper, rough set theory and support vector machine technology are combined to construct a mathematical model for comprehensive assessment and prediction of highway collapse risk. Based on the data analysis of the collapse of Yingxiu-Wolong highway in Hanchuan County of Sichuan Province after the 5.12 earthquake in Wenchuan County, the spatial dimension of the input information, which is the initial index, is reduced and optimized, find out the key index system that affects the evaluation and forecast of highway collapse risk, and remove the irrelevant index. On this basis, more purposeful and targeted research on this section of the highway has been carried out again to obtain the survey data, and apply the support vector machine model to carry out a comprehensive evaluation of the risk of collapse in this section of the highway. The good results are obtained, it is significant to evaluate and forecast the risk of highway collapse.

2. Rough set theory

Rough set (RS) theory (Zhang et al. 2001; Cao et al. 2007; Cao and Ruan 2009; Lai and Wu 2011; Du 2012) as a scientific study of intelligent computation was put forward by the Polish mathematician Pawlak in 1982, which is a set of mathematical theory methods for expression, learning, induction, etc. of incomplete data, imprecise knowledge. The essence of this theory is attribute reduction. It is well known that when the data in the knowledge expression system (information system) is collected at random, there is general redundancy. Decision table is a knowledge expression system with conditional attribute and decision attribute. A knowledge representation system can be expressed as $S=(U,A,V,f)$, where, U is a finite set of objects, $A=C \cup D$ is a finite set of attributes, C and D is a set of conditional attributes and a set of decision attributes, respectively, V is a domain of attributes A, $f: U \times A \rightarrow V$ is an information function, refers to the property value of each object. In the knowledge expression system, the importance degree of attributes is different. In condition of keeping the classification ability of knowledge base unchanged, according to the relevance in the set of conditional attributes, the goal of attribute reduction is to find some important condition attributes, which make the classification of decision attributes consistent and uniform.

Reduction is usually not unique, and all or minimum reduction in finding attributes has been proved to be an NP (non-deterministic polynomial)-hard problem. At present, the rough set attribute reduction algorithms mainly include Johnson greedy algorithm, exhaustive algorithm, attribute importance heuristic algorithm, genetic algorithm, dynamic reduction, concept lattice and so on. The exhaustion algorithm is to simplify the resolution function derived from the resolution matrix by the absorption law, and make it a minimum disjunctive normal form to obtain the reduction of the data attribute set. Considering that the exhaustion algorithm is only suitable for small data sets, and all reduction can be obtained despite the complexity of the algorithm, so the paper adopts the exhaustion algorithm to simplify the evaluation index of highway collapse risk grade.

3. Support vector machine

In 1995, Corinna Cortes and Vapnik first proposed the support vector machine (SVM) (Cao and Liang 2002; Li 2007; An 2008; Huang 2009; Nie 2009; Wang 2009; Cao and Zhao 2011; Lai and Wu 2011; Liang et al. 2021), in which supervised learning models for classification and regression analysis are related to related learning algorithms, they can analyze data and identify patterns. Support vector machine allows for the optimal classification of linear and non-linear separable data.

On the problem of optimal classification for linearly separable data, the training sample set $T=\{(x_1,y_1),(x_2,y_2)\cdots,(x_l,y_l)\}$, $x \in R^n$, $y \in \{-1,1\}$ is given to find an optimal hyperplane $(w \cdot x)+b=0$ which satisfies the problem, that is, to find the maximum interval of minimum $\|w\|^2/2$ for the interval of classification $2/\|w\|$, and for the subscript i of $y_i=1$ and $y_i=-1$, there is $(w \cdot x)+b \geq 1$ and $(w \cdot x)+b \leq -1$. Then the optimal classification hyperplane problem is found, that is,

$$\min \frac{1}{2}\|w\|^2$$

$$s.t. \quad y_i[(w \cdot x)+b] \geq 1, i=1,2,\cdots,l \tag{1}$$

Using the Lagrange function, the dual objective function of Eq. (1) is

$$\max W\left[\alpha\right]=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left(x_i \cdot x_j\right)$$

$$s.t. \quad \sum_{i=1}^{l}\alpha_i y_i=0,\alpha_i\geq 0,i=1,2,\cdots,l \tag{2}$$

To find the optimal solution α^* of the quadratic programming problem, and the classification threshold b^*, most samples have $\alpha^*=0$, the corresponding samples for $\alpha^*\neq 0$ are only called support vector, determine the classification results. The optimal classification decision function is

$$f\left(x\right)=\mathrm{sgn}\left\{\left(w \cdot x\right)+b\right\}=\mathrm{sgn}\left(\sum_{i=1}^{n}\alpha_i^* y_i(x_i \cdot x)+b^*\right) \tag{3}$$

When the training set is linear inseparability problem, the relaxation variable $\xi_i,i=1,2,\cdots l$ is introduced, and the optimal classification hyperplane problem is

$$\min\frac{1}{2}\left\|w\right\|^2+C\sum_{i=1}^{l}\xi_i$$

$$s.t. \quad y_i[(w \cdot x)+b]\geq 1-\xi_i,\xi_i\geq 0,i=1,2,\cdots,l \tag{4}$$

Where, C is the penalty parameter. Using the Lagrangefunction, the dual objective function of Eq. (1) is

$$\max_{\alpha}\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j\left(x_i,x_j\right)$$

$$s.t. \quad \mathrm{C}\geq\alpha_i\geq 0,i=1,2,\cdots,l$$

$$y^T\alpha=0 \tag{5}$$

A classification hyperplane $(w^* \cdot x)+b^*=0$ is constructed and the optimal classification decision function is obtained

$$f\left(x\right)=\mathrm{sgn}\left\{\left(w^* \cdot x\right)+b^*\right\} \tag{6}$$

For the nonlinear separable problem, the input vector can be mapped to a high dimensional feature space and an optimal hyperplane can be constructed in the feature space. The optimal hyperplane and decision function are $\left(w \cdot \Phi(x)\right)+b=0$ and $f\left(x\right)=\mathrm{sgn}\left\{\left(w \cdot \Phi(x)\right)+b\right\}$ respectively, and the optimal classification hyperplane problem is

$$\min_{w,b,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i$$

$$s.t. \quad y_i[(w \cdot \Phi(x)) + b] \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \cdots, l \tag{7}$$

The dual optimization problem is also obtained

$$\max W[\alpha] = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j)$$

$$s.t. \quad 0 \leq \alpha_i \leq C, i = 1, 2, \cdots, l \tag{8}$$

Where, $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ is kernel function, in order to realize the linear classification of nonlinear change, the kernel function commonly used are polynomial kernel function (Poly) , Gauss radial basis function (Rbf), multi-layer neural network kernel function (Sigmoid), Fourier kernel function (Fourier) and so on. The optimal solution $\alpha^* = (\alpha_1^*, \cdots, \alpha_n^*)^T$ of the problem is calculated, and then $\omega^* = \sum_{i=1}^{n}\alpha_i^* y_i x_i$, the classification decision function is

$$f(x) = \mathrm{sgn}\left\{\sum_{i=1}^{n}\alpha_i^* y_i K(x_i, x_j) + b^*\right\} \tag{9}$$

4. Rough set-support vector machine model for risk evaluation of highway collapse

Firstly, the factors affecting the risk of highway collapse are comprehensively analyzed, the key evaluation indexes are selected, the evaluation index system is constructed. The data sample set of evaluation indexes is constructed by data collection and field investigation. According to the grading standards of the evaluation indexes, the RS decision table is constructed by preprocessing the original data, and the RS attribute reduction is used to eliminate the redundant and unimportant attributes, so as to extract the features of highway collapse risk information. Then, training samples and test samples are selected from the data preprocessed by reduction, and the suitable kernel functions are selected and parameters are optimized. Support vector classification is trained with training samples, RS-SVM highway collapse risk assessment model is established to evaluate the risk level of the test evaluation samples. The specific flow of highway collapse risk assessment based on RS-SVM is shown in Fig. 1.

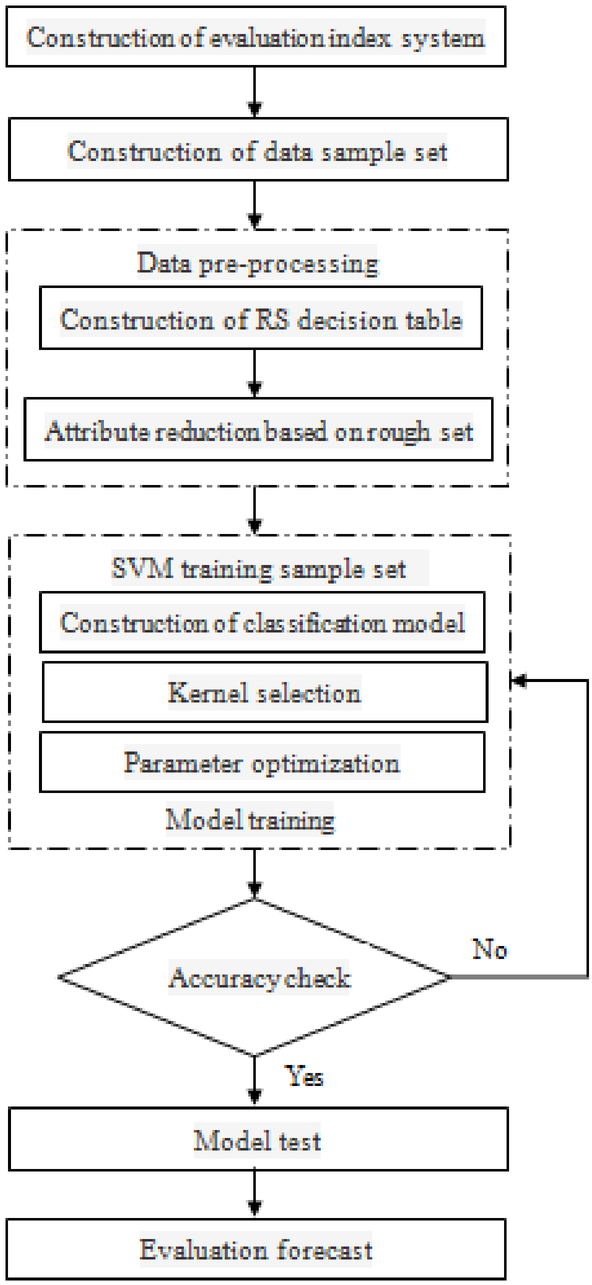

Fig. 1 Flow diagram of highway landslide hazard evaluation based on RS-SVM

4.1. Index system of highway collapse risk assessment

The S303 Yingxiu-Wolong highway in Hanchuan County, Sichuan Province was paved before the May 12, 2008 earthquake. The highway's total length is 45.5 km and it is an important trunk route connecting Yingxiu and Wolong, As the closest highway to the epicenter of the Wenchuan earthquake, the highway is basically routed along the Longmenshan tectonic belt, starts from the Beichuan-Yingxiu fault (the central fault) and passes through the Maoxian-Hanchuan fault (the Houshan fault) in Longmenshan. It has complicated geological conditions, is the earthquake geological disaster most development, the damage is most serious a highway (Fig. 2). In order to comprehensively and objectively analyze and evaluate the geological disaster of highway collapse, the acquisition of information knowledge of highway collapse disaster is the key to complete the collapse risk assessment. The aim is to select the most effective knowledge from the original field geological survey data, eliminate redundant information, reduce the dimension of feature space, and improve the generalization ability of the evaluation system. Therefore, when constructing the index system, in order to achieve the establishment completeness and comprehensiveness of the comprehensive evaluation index system, we must first ensure that the index system has a broad generalization. Therefore, according to the field investigation data of highway collapse, on the basis of comprehensive analysis and study on the evaluation indexes of highway collapse risk at home and abroad, the paper combines with the key factors affecting the risk of highway collapse to determine 4 grades and 15 index items finally (Table 1). The evaluation index system of highway collapse risk is constructed using 15 indexes such as $x_1, x_2, x_3, \ldots, x_{15}$.

Fig. 2 Slope rock mass landslide in 5.12 Wenchuan earthquake destroys a bridge, blocks the road, buries smashes tunnel entrance

Table 1 Evaluation factors and grading standard of highway landslide hazard

Evaluation index	Evaluation grades

		I (C_1)	II (C_2)	III (C_3)	IV (C_4)
Topography	(x_1) Slope shape of slope	Fold line slope (4)	Concave slope (3)	Straight slope (2)	Convex slope (1)
	(x_2) Aspect of slope	Sunny slope (4-3)		Shady slope (2-1)	
	(x_3) Slope of slope/°	>60	40-60	20-40	<20
	(x_4) Height of slope/m	>150	100-150	50-100	<50
Geologic structure	(x_5) Meso structure	Development (4)	More development (3)	Less development (2)	No Development (1)
	(x_6) Exposed structural face	Development (4)	More development (3)	Less development (2)	No Development (1)
Stratum	(x_7) Stratum lithology	Loose rock (4)	Cemented chasten (3)	Well cemented half hard (2)	Hard rock (1)
	(x_8) Weakness interlayer	Clear (4)	More clear (3)	Less clear (2)	No clear (1)
	(x_9) Relationship between weakness face and free face	Consequent slope (4)	Oblique crossing (3)	Cross slope (2)	Reverse slope (1)
Climatic and hydrological geological condition	(x_{10}) Mean annual rainfall/ mm/Year	>1500	1000-1500	500-1000	<500
	(x_{11}) Rainfall erosion/m	>0.5	0.3-0.5	0.1-0.3	<0.1
Other factors	(x_{12}) Vegetation cover rate/%	<5	5-15	15-30	>30
	(x_{13}) Weathering degree of rock/%	>30	10-30	5-10	<5
	(x_{14}) Earthquake intensity/°	>8	5-8	3-5	<3
	(x_{15}) Human activity intensity	Big (4)	Bigger (3)	Lesser (2)	Small (1)

In the whole evaluation index system, semi-quantitative method and measured value are used to evaluate the qualitative and quantitative indexes respectively. The classification standards and descriptions are shown in Table 1. The result of comprehensive evaluation is divided into 4 grades, which are expressed by I, II, III, IV respectively. Considering that there are many evaluation indexes in the comprehensive evaluation index system, and the contribution of each evaluation index is different in the whole evaluation, in order to reduce the decentralization of weight and the calculation workload, the redundant influencing factors should be removed firstly. In the paper, rough set theory is used to reduce the whole index system, extract the key influencing factors, and then evaluate the index system based on uncertainty measure.

4.2. Attribute reduction of risk evaluation index of highway collapse

The paper selects 17 collapse points from S303 Yingxiu-Wolong highway as the research object, and obtains the related original index data of collapse points through collecting and field investigation. Based on rough set theory, first of all, the original data should be discretized, in which each index is divided into 4 grades according to the evaluation index system, we can get the rationality two-dimensional information evaluation decision table of the comprehensive evaluation system (Table 2) .

Then, based on the Rosetta data analysis software developed by the scientists and technicians of Warsaw University in Poland and Norwegian University of Science and Technology, the attribute reduction of the decision table is carried out by the exhaustive algorithm, after removing six very small redundant indexes such as x_5, x_8, x_{10}, x_{11}, x_{14} and x_{15}, nine key indexes such as x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{12} and x_{13} are obtained.

4.3. Sample set of evaluation model based on rough set and support vector machine

Using the data of 17 collapse points on both sides of S303 Yingxiu-Wolong highway, which is reduced by rough set theory, a sample set based on the support vector machine model is constructed, the first 13 samples are selected as training samples from 17 samples, and the best RS-SVM model is constructed. The other 4 samples (collapse number is W06, W17, W28 and W33) are used as test samples (Table 3) .

Table 3 Evaluation value of evaluation indexes

Collapse number	Evaluation index									Risk grade
	x_1	x_2	x_3	x_4	x_6	x_7	x_9	x_{12}	x_{13}	
BT05	1	3	70	73	3	3	4	6	7	II
BT09	3	1	50	613	2	3	2	18	8	III
BT13	3	3	69	105	4	2	3	20	6	II
BT24	2	3	30	234	2	3	3	25	6	III
BT33	1	1	48	119	2	2	1	25	6	III
BT40	1	3	85	21	1	2	2	35	8	III
BT49	2	1	70	189	4	3	3	8	9	II
BT54	3	3	42	25	0	4	2	20	12	III
BT58	2	3	28	314	3	2	4	25	15	III
BT66	4	1	47	410	1	2	3	20	8	III
BT70	1	3	68	66	2	3	2	25	40	III
BT80	2	1	45	200	2	3	3	20	12	III
YBT03	3	3	42	83	3	4	3	8	15	II
W06	4	3	43	30	0	4	4	25	15	III
W17	4	1	50	162	1	3	3	12	25	II
W28	3	1	40	214	2	3	3	6	28	II
W33	2	3	72	28	1	3	4	6	25	II

In this paper, we use Sklearn package in Python environment to design the support vector machine algorithm

219 **Table 2** Discrete evaluation data tables

	Evaluation index	Collapse data																
		BT05	BT09	BT13	BT24	BT33	BT40	BT49	BT54	BT58	BT66	BT70	BT80	YBT03	W06	W17	W28	W33
Topography	(x_1) Slope shape of slope	1	3	3	2	1	1	2	3	2	4	1	2	3	4	4	3	2
	(x_2) Aspect of slope	3	1	3	3	1	3	1	3	3	1	3	1	3	3	1	1	3
	(x_3) Slope of slope/°	4	3	4	2	3	4	4	3	2	3	4	3	3	3	3	2	4
Geologic structure	(x_4) Height of slope/m	2	4	3	4	3	1	4	1	4	4	2	4	2	1	3	4	1
	(x_5) Meso structure	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	(x_6) Exposed structural face	2	2	3	2	2	1	3	1	2	1	2	2	2	1	1	2	1
Stratum	(x_7) Stratum lithology	3	3	2	3	2	2	3	4	2	2	3	3	4	4	3	3	3
	(x_8) Weakness interlayer	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	(x_9) Relationship between weakness face and free face	4	2	3	3	1	2	3	2	4	3	2	3	3	4	3	3	4
Climatic and hydrological geological condition	(x_{10}) Mean annual rainfall/ mm/Year	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	(x_{11}) Rainfall erosion/m	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Other factors	(x_{12}) Vegetation cover rate/%	3	2	2	2	2	1	3	2	2	2	2	2	3	2	3	3	3
	(x_{13}) Weathering degree of rock/%	2	2	2	2	2	2	2	3	3	2	4	3	3	3	3	3	3
	(x_{14}) Earthquake intensity/°	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	(x_{15}) Human activity intensity	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Decision attribute value		2	3	2	3	3	3	2	3	3	3	3	3	2	3	2	2	2

model. The kernel function uses Gauss radial basis function $K\left(x_i, x_j\right) = exp\left(-\frac{\left|x_i - x_j\right|^2}{2\sigma^2}\right)$ commonly used, for the penalty parameter C and kernel function parameter g in the SVM model, the grid search method is used to get the best parameters of combinatorial optimization. In the paper, the simulation is done in Python environment, and the penalty parameter C =8 and the kernel parameter g =0 are determined by testing the training samples, the correct rate of cross-validation is 100%, and the model is optimal.

In order to verify the correctness and reliability of highway collapse risk discrimination based on the RS-SVM model, the RS-SVM model was used to distinguish training samples BT05 to YBT03. The results are shown in Table 4. All the 13 samples are correctly identified with a misjudgement rate of 0. The other 4 samples are tested according to the RS-SVM model studied well. The results are shown in Table 5. The results obtained by the uncertainty measures and distance discriminant methods are listed in Table 5.

Table 4 Training samples of RS-SVM model

Collapse number	Evaluation index									Risk grade			
	x_1	x_2	x_3	x_4	x_6	x_7	x_9	x_{12}	x_{13}	Actual grade	The method of this paper	Uncertainty measure	Distance discriminant method
BT05	1	3	70	73	3	3	4	6	7	II	II	II	II
BT09	3	1	50	613	2	3	2	18	8	III	III	III	III
BT13	3	3	69	105	4	2	3	20	6	II	II	II	II
BT24	2	3	30	234	2	3	3	25	6	III	III	III	III
BT33	1	1	48	119	2	2	1	25	6	III	III	III	III
BT40	1	3	85	21	1	2	2	35	8	III	III	III	III
BT49	2	1	70	189	4	3	3	8	9	III	III	III	III
BT54	3	3	42	25	0	4	2	20	12	III	III	III	III
BT58	2	3	28	314	3	2	4	25	15	III	III	III	III
BT66	4	1	47	410	1	2	3	20	8	III	III	III	III
BT70	1	3	68	66	2	3	2	25	40	III	III	III	III
BT80	2	1	45	200	2	3	3	20	12	III	III	III	III
YBT03	3	3	42	83	3	4	3	8	15	II	II	II	II

Table 5 Test samples of RS-SVM model

Collapse number	Evaluation index									Risk grade			
	x_1	x_2	x_3	x_4	x_6	x_7	x_9	x_{12}	x_{13}	Actual grade	The method of this paper	Uncertainty measure	Distance discriminant method

W06	4	3	43	30	0	4	4	25	15	III	III	III	III
W17	4	1	50	162	1	3	3	12	25	II	II	II	II
W28	3	1	40	214	2	3	3	6	28	II	II	II	II
W33	2	3	72	28	1	3	4	6	25	II	II	II	II

5. Conclusions

(1) In the evaluation and prediction of highway collapse risk, firstly, rough set theory is used to reduce the indexes, and the relatively unimportant indexes are deleted, so as to achieve the goal of index optimization. Then, the 9 key indexes which affect the collapse activity, such as slope shape of slope, aspect of slope, slope of slope, height of slope, exposed structural face, stratum lithology, relationship between weakness face and free face, vegetation cover rate and weathering degree of rock, are extracted to be used as the discriminant factors of the support vector machine classification model.

(2) The actual survey data of 17 collapse points optimized from Yingxiu to Wolong section of the S303 highway are selected as the training and testing samples of the support vector machine, the best parameters of combinatorial optimization are obtained through the training of the training samples, the RS-SVM model is used to evaluate and predict the test samples.

(3) The research results show that the highway collapse discriminant analysis model based on RS-SVM achieves the goal of index optimization, not only can guarantee the accuracy of evaluation, but also reduce the computational load of the model, the learning performance is good, and the prediction accuracy is high, it is an effective method to forecast and evaluate the risk of highway collapse.

Acknowledgements

This work was financially supported by the fund for the Open Fund for the State Key Laboratory of Water Resources Protection and Utilization in Coal Mining (SHJT-16-30.19), and the Special Fund for Basic Scientific Research of Central Colleges (310827151056, 310827153408, 310827173702 and 300102280401)

Data availability statements

All data generated or analyzed during this study are included in this published article.

Conflict of interests

The authors declare that they have no conflicts of interest.

References

An L. P. (2008). *Rough Set Approach to Multi-attribute Decision Analysis*. Science Press, Beijing, China.

Broeckx J., Vanmaercke M., Duchateau R., Poesen J. (2018). A data-based landslide susceptibility map of Africa. *Earth Science Reviews*, 185, 102-121.

Cao X. Y., Liang J. G. (2002). The method of ascertaining attribute weight based on rough sets theory. *Chinese Journal of Management Science*, 10(5), 98-100.

Cao Q. K., Yang Y. L., Yu R. L. (2007). The coal mines safety appraisal based on unascertained set. *Journal of China Coal Society*, 32(2), 181-185.

Cao Q. K., Ruan J. H. (2009). Rationality evaluation on mine ventilation system based on rough set theory and unascertained measure. *Scientific Decision Making*, (5), 88-94.

Cao Q. K., Zhao F. (2011). Risk evaluation of water inrush from coal floor based on fuzzy-support vector machine. *Journal of China Coal Society*, 36(4), 633-637.

Du Z. S. (2012). Outburst risk evaluation model based on rough set and unascertained measure theory. *Safety in Coal Mines*, 43(10), 7-10.

Feng F. (2019). *Present Situation and Spatial Prediction of Collapse and Landslide in Shilou-Ji County Section of the Middle Reaches of the Yellow River*. Chang'an University, Xi'an, China.

Gao K. C., Cui P., Zhao C. Y., Wei F. Q. (2006). Landslide Hazard Evaluation of Wanzhou Based on GIS Information Value Method In The Three Gorges Reservoir. *Chinese Journal of Rock Mechanics and Engineering*, 25(5), 991-997.

Greco R., Sorriso-Valvo M. (2013). Influence of management of variables, sampling zones and land units on LR analysis for landslide spatial prevision. *Natural hazards and earth system sciences*, 13(9), 2209-2221.

He H. J., Su S. R., Wang X. J., Li P. (2013). Study and application on comprehensive evaluation model of landslide hazard based on uncertainty measure theory. *Journal of Central South University (Science and Technology)*, 44(4), 1564-1570.

He H. J., An L., Liu W., Zhang J. (2017). Prediction Model of Collapse Risk Based on Information Entropy and Distance Discriminant Analysis Method. *Mathematical Problems in Engineering*, (2017), 1-8.

Huang W. Y. (2009). *Study on Coal Gas Early-warning Technology Based on Support Vector Machine and Data Fusion*. China University of Mining and Technology, Xuzhou, China.

Lai H. S., Wu C. F. (2011). Productivity evaluation of standard cultivated land based on rough set and support vector machine. *Journal of Natural Resources*, 26(12), 2141-2154.

Li Y. (2007). Prediction method study on water inrush through coal floor based on SVM. Xi'an Research Institute of China Coal Research Institute, Xi'an, China.

Li S. (2011). *Study on classification and formation mechanism of highway slope landslide disaster after earthquake*. Chang'an University, Xi'an, China.

Liang B., Liu Z., Niu Y. B. (2021). Shearer cutting pattern recognition based on multi-scale fuzzy entropy and support vector machine. Coal Engineering, 53(5), 131-135.

Liu L. (2010). *Study on Landslip Disaster Risk of Chengkun Railway K242-K331*. Southwest Jiaotong University, Chengdu, China.

Liu J. M. (2010). Risk Assessment of Collapse Disaster along Nan-Yan Section of Provincial Road S219. China University of Geosciences, Beijing, China.

Liu X. (2016). Hazard Assessment of Landslide and Collapse induced by Volcanic Eruption in Changbai Mountains. Jilin University, Changchun, China.

Mandal S., Mondal S. (2019). Logistic regression (LR) model and landslide susceptibility: A RS and GIS-based approach. In: Statistical approaches for landslide susceptibility assessment and prediction. Springer, Cham, 107-121.

Nie X. F. (2009). *The Research on the Application of Fuzzy Rough Set and Support Vector Machines in Coal and Gas Outburst Prediction*. Liaoning Technical University, Fushun, China.

Qiu H. J. (2012). *Study on the Regional Landslide Characteristic Analysis and Hazard Assessment: A case study of Ningqiang County*. Northwest University, Xi'an, China.

Wang F. (2009). *Research on Methods of Traffic Flow Forecasting Based on SVM*. Dalian University of Technology, Dalian, China.

Xue K. X. (2011). *Theory and Application of Risk Assessment for Extreme Rainfall-induced Geological Hazards of Mountain Road*. Chongqing University, Chongqing, China.

Yang W. H., Yang Z. Q., Zhao J. L., Liang Y. Q. (2021). Research on the risk of collapse and landslide in Wenchuan earthquake disaster area based on logistic regression model. *Hubei Agricultural Sciences*, 60(6), 42-48, 59.

Yue Y. (2008). *Study on hazard evaluation landslide geological disaster based GIS*. Tongji University, Shanghai, China.

Zhang W. X., Wu W. Z., Liang J. Y. (2001). *Rough Set Theory and Method*. Science Press, Beijing, China.

Zhang J. Q., Fan J. R., Yan D., Guo F. F., Su F. H. (2009). Sensitivity evaluation to the earthquake-induced landslide and collapse: a case study of Beichuan county, Sichuan province. *Journal of Sichuan University (Engineering science edition)*, 41(3), 140-145.