Mathematical Biosciences and Engineering

MBE, 19(5): 4526–4546.
DOI: 10.3934/mbe.2022209
Received: 04 December 2021
Revised: 23 January 2022
Accepted: 15 February 2022
Published: 04 March 2022

http://www.aimspress.com/journal/MBE

Research article

Prediction of slope stability using Tree Augmented Naive-Bayes classifier: modeling and performance evaluation

Feezan Ahmad[1], Xiao-Wei Tang[1,*], Jiang-Nan Qiu[2,*], Piotr Wróblewski[3,4], Mahmood Ahmad[5,6] and Irfan Jamil[7]

[1] State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China
[2] Faculty of Management and Economics, Dalian University of Technology, Dalian 116024, China
[3] Faculty of Engineering, University of Technology and Economics H. Chodkowska in Warsaw, Jutrzenki 135, 02-231 Warsaw, Poland
[4] Faculty of Mechatronics, Armament and Aerospace of the Military University of Technology, Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland
[5] Department of Civil Engineering, Faculty of Engineering, International Islamic University Malaysia, Jalan Gombak, Selangor 50728, Malaysia
[6] Department of Civil Engineering, University of Engineering and Technology Peshawar (Bannu Campus), Bannu 28100, Pakistan
[7] Department of Civil Engineering, University of Engineering and Technology Peshawar, Peshawar 25120, Pakistan

* **Correspondence:** Email: tangxw@dlut.edu.cn, qiujn@dlut.edu.cn.

Abstract: Predicting slope stability is critical for identifying terrain that is prone to landslides and mitigating the damage caused by landslides. The relationships between factors that determine slope instability are complicated and multi-factorial, so it is sometimes difficult to mathematically characterize slope stability. In this paper, new Tree Augmented Naive-Bayes (TAN) model was developed to predict slope stability subjected to circular failures based on six input factors: cohesion, internal friction angle, pore pressure ratio, slope angle, unit weight, and slope angle. A total 87 slope stability case records obtained from published literature was used to train and test the proposed TAN model. According to the results of the performance indices—accuracy, precision, recall, F-score and Matthews correlation coefficient, the established TAN model was proven to be better at predicting slope stability with acceptable accuracy than other formerly developed empirical models in the

literature. Furthermore, the slope height was revealed as the most sensitive factor in a sensitivity analysis.

Keywords: machine learning algorithm; Tree Augmented Naive-Bayes; sensitivity analysis; slope stability prediction

1. Introduction

Slope collapses are complicated natural disasters with devastating consequences [1]. Every year, such hazards cause significant damage to public and private property, traffic disruptions, and lives lost [2–4]. As a result, slope stability analyses are essential to prevent and mitigate damages, and better tools for slope assessment are desperately needed in the field of civil engineering. The results of the analysis can be used to identify collapse-prone areas. Based on this data, government agencies can gain a better understanding of slope failure occurrences, and the task of providing financial resources to build retaining structures and developing evacuation plans can be completed more efficiently [2].

The accurate prediction of a rock or soil slope's stability is a difficult problem, owing to the slope's dependency on multiple parameters and the difficulty in determining these values [5]. The term factor of safety (*FoS*), is commonly used to describe the stability of slopes. *FoS* is calculated by dividing the resisting forces to the driving forces. The *FoS* is more than one when the resisting forces of a slope are greater than the driving forces; when the resisting forces are less than the driving forces, the *FoS* is less than one and the slope is unstable.

Slope stability analysis and prediction approaches have been the focus of many researchers. These efforts have led to development of a number of different and sophisticated formulations for determining *FoS* and also slope design approaches such as limit equilibrium methods (LEM) [6–9] continuum mechanics-based numerical techniques [10–12], methods based on probabilistic methodologies, such as variational and combination methods [13] and numerical approaches have been widely used as traditional methods for studying slope stability in geotechnical problems. Due of its computational time, LEM simulations may become inadequate. In recent years, however, a number of studies have been conducted to create a number of computational intelligence systems for slope stability analysis.

Data mining approaches have recently proved successful in paving the way for many promising opportunities in slope stability [14–18] and other fields of civil engineering [19–29]. Table 1 lists some representative references for data mining applications for slope stability prediction. The majority of these studies investigated slopes subjected to circular-type failure and stability of these slopes based on geotechnical, geometrical and pore-water pressure parameters. In these studies, data mining approaches based on historical data have been used for two purposes: 1) prediction of slope *FoS*: the output of these models is the *FoS*, and 2) prediction of SS status: the output of proposed models shows the slope's stability or instability. However soft computing techniques have proved successful in predicting SS; the fact that most of these techniques are black boxes. The novelty of this article is the development of a transparent and understandable model for predicting SS in slopes that have experienced circular mode failure. The TAN-based model overcomes the shortcomings of other soft computing techniques by producing transparent and structural model showing the relationship between input and output parameters.

A critical review of existing literature suggests that TAN algorithm implementation in the analysis of geotechnical engineering is scarcely explored. Unlike other soft computing technologies, the TAN

algorithm can produce a model that is simple to understand and interpret. The main contributions of this paper are as follows: 1) a new TAN model is developed to predict the slope stability subjected to circular slope failures; 2) most probable explanation slope sites of unstable is presented; 3) to test the performance of the models proposed, it is applied to field data given in open source literatures; 4) sensitivity analysis is presented owing to know the most sensitive factor; 5) data discretization was conducted to reduce and elucidate the data set, develop the model quickly and easily, and acquire easily interpretable outputs in this study; 6) The difference in the class ratio between the sample and the population i.e., sampling bias in the training and test datasets is almost negligible.

The structure of the paper is as follows: in Section 2, the materials and methods are presented. The construction process of the proposed prediction model is described in Section 3. Section 4 presents results and discussion. Finally, the concluding remarks are presented.

Table 1. Previous references on the prediction of slope stability using soft computing methods.

Reference	ANN	SVM/ARVM	GP/GA	NB	RF	GBM	DT	LR	Auxiliary method
Lu and Rosenbaum [1]	●								
Yang et al. [30]			●						
Sakellariou and Ferentinou [5]	●								
Wang et al. [31]	●								
Samui [32]		●							
Zhao [33]		●							
Choobbasti et al. [34]	●								
Ahangar-Asr et al. [35]			●						LSM
Das et al. [36]	●								
Li, Zhao, and Ru [37]		●							MCS
Dong and Li [38]	●	●		●	●				
Manouchehrian et al. [39]			●						
Zhang et al. [40]		●							
Liu et al. [41]	●								
Xue et al. [42]		●							PSO
Feng et al. [43]				●					
Qi and Tang [44]	●	●			●	●	●	●	FA
Sari et al. [45]		●							
Gao et al. [46]	●								ICA
Yuan and Moayedi [47]	●								PSO, GA
Sari [48]								●	
Zhou et al. [49]	●	●			●	●			

*Note: ANN: artificial neural network; SVM: support vector machine; ARVM: adaptive relevance vector machine; GP: genetic programming; GA: genetic algorithm; NB: naive bayes; RF: random forest; GBM: gradient boosting machine; DT: decision tree; LR: logistic regression; LSM: least square method; MCS: Monte Carlo simulation; PSO: particle swarm optimization; FA: firefly algorithm; ICA: imperialist competitive algorithm.

2. Methodology

2.1. Tree Augmented Naive-Bayes (TAN)

The TAN classifier was presented as an extension of the Naive Bayes classifier. TAN allows the independence assumption by permitting arcs between variables. The impact of variable *Xi* on the class variable is likewise determined by the value of *Xj*, as indicated by an arc from variable *Xi* to variable *Xj*. An example of a TAN is shown in Figure 1. This approach is based on a Chow and Liu algorithm [50] that was proposed earlier. The method is divided into the following five steps.

1) Given the class variable *C*, compute the conditional mutual information. $I(X_i; X_j|C)$, between each pair of variables, $i \neq j$. $I(X_i; X_j|C)$ is defined as follows:

$$I(X_i; X_j|C) = \sum_{x_i,x_j,c_l} P(X_i = x_i, X_j = x_j, C = c_l) \times log \frac{P(X_i=x_i,X_j=x_j|C=c_l)}{P(X_i=x_i|C=c_l)P(X_j=x_j|C=c_l)}, \quad (1)$$

2) When the value of C is known, this function approximates the information provided by *Xj* about *Xi* (and vice versa).

3) Use the variables as nodes in a complete undirected graph. Assign a weight $I(X_i; X_j|C)$ to each arc connecting X_i and X_j.

4) Create the maximum weighted spanning tree.

5) To convert an undirected tree into a directed tree, select a root variable and determine the direction of all arcs to be outward from it.

6) Connect the classification node C to each Xi using an arc.

The aforementioned technique, produces TANs that optimize the network's log likelihood given the training data and has a time complexity of $O(n^2.N)$, where n is the number of variables and N is the number of data points [51]. With the same computational complexity and robustness, experimental data demonstrated that TANs outperformed Naive Bayes [51].

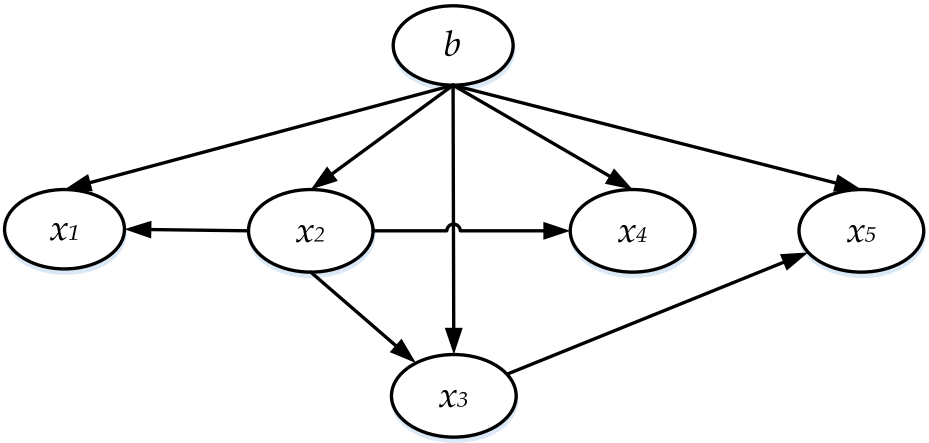

Figure 1. A simple TAN structure.

A sufficient threshold probability for classification must be chosen. Because there are only two classes in this study, a 0.5 threshold is typically used [52]. For example, "Stable" slopes, are defined as P(Stable|X) > 0.5.

2.2. Database description

Slope stability study results suggest that six parameters, {γ, c, ϕ, β, H and ru}, influence circular failure of a slope. They are in line with the parameters that are usually found in the literature [18,40,41,46,49]. Other indicators are theoretically possible, but collecting these would be a significant challenge before they could be used in practice. Slope height (H) and slope angle (β) are geometric properties of a slope that are frequently used to determine slope failure conditions. The slope stability rapidly reduces as the slope height rises. The slope stability decreases as the slope angle increases. Water infiltration reduces the shear strength of the rock and soil owing to softening. Slope stability suffers as a result of all of these changes. The investigations in this research were performed on a dataset encompassing 87 case studies that were investigated for circular critical failure mechanisms and were acquired from various literatures [5,31,39,53]. The input parameters are H, γ, c, ϕ, β and ru, whereas the output parameter is SS status. SS in the database was classified as stable or unstable based on whether or not considerable soil movement was observed on the slope surface. If there is no considerable movement of the soil in the slope surface that affects safety, the slope status is considered stable. Otherwise, a slope's status is unstable. 42 of the 87 database cases are stable, while the rest are unstable. The SS is coded as 0 for unstable slopes and 1 for stable slopes in this study.

ANN, ELM, and SVM algorithms were used in much previous research to predict specific *FoS* values [5,41,54,55], the *FoS* cannot always reflect the actual condition of the slopes, i.e., when the *FoS* exceeds one, the actual condition of the slopes is sometimes considered "unstable" (see for example Case Nos. 7, 17, 19 in Table A1 shown in Appendix). Therefore, rather than predicting specific *FoS* values, in the present study, the dataset was used to establish a relation among these six factors as input and the actual slope condition as output.

Figure 2 depicts the cumulative percentage and frequency distributions for all of the input and output parameters of the mentioned database utilized in the modeling of SS of circular mode failure. The data points of every input parameter are distributed over its range. Table 2 shows minimum (Min) and maximum (Max) values, mean, and standard deviation (SD) of all six input parameters. It's worth noting that each parameter's min and max values establish the ranges in which predictions can be made.

Table 2. Descriptive statistics of the data set.

Parameter	Min	Max	Mean	SD
γ	12	31.3	31.342	4.178
c	0	150	25.080	25.347
ϕ	0	45	27.975	9.983
β	9.792	53	34.628	9.792
H	3.6	511	104.017	132.879
r_u	0	0.5	0.219	0.164

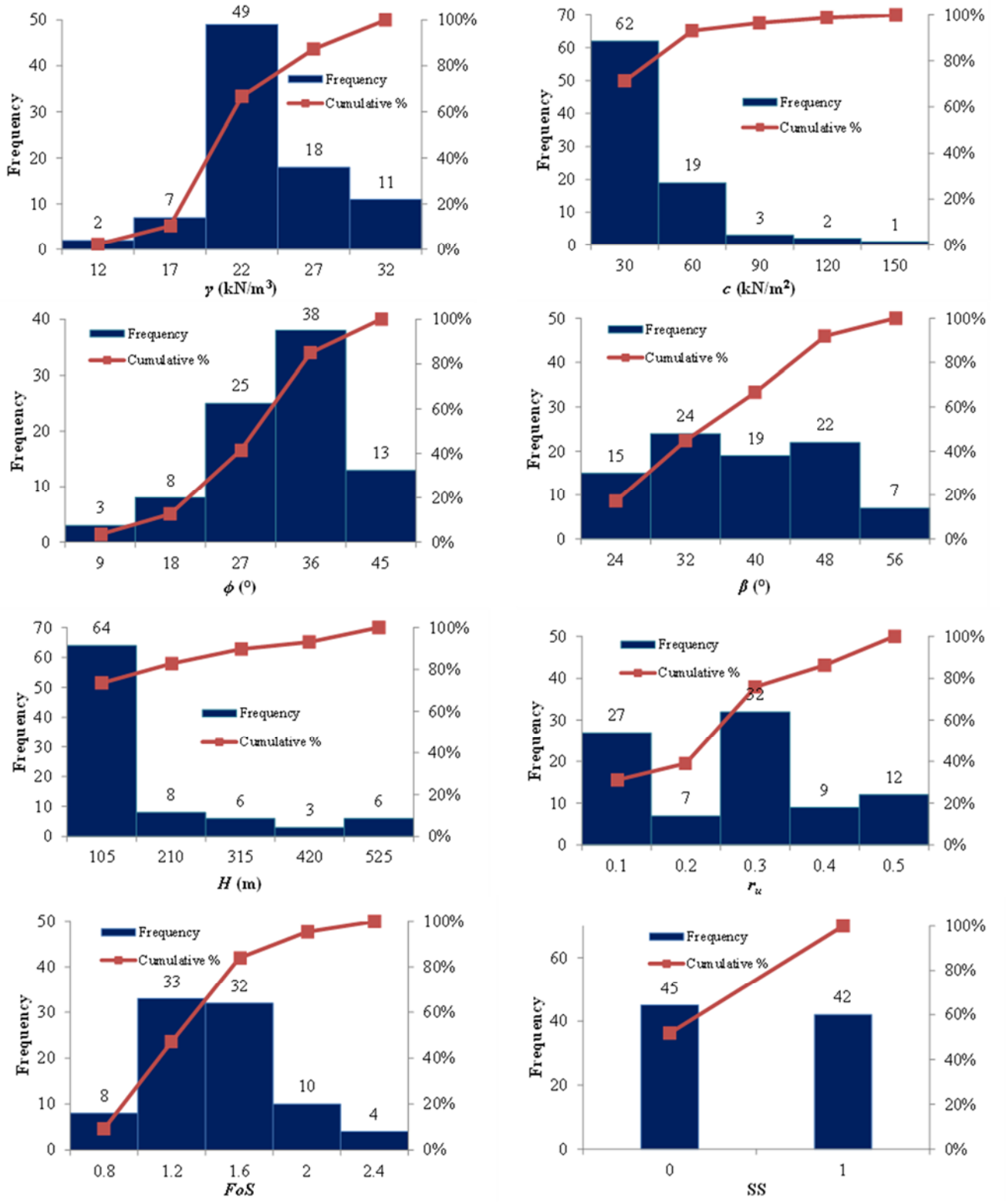

Figure 2. Histograms of the input and output parameters considered in this study.

2.3. Correlation analysis

To check the suitability of the TANs applied in this study, correlation coefficients (ρ) are determined to verify the strength of the relationship between the various factors (see Table 3). Given a pair of random variables (m,n) the formula for ρ is:

$$\rho(m,n) = \frac{cov(m,n)}{\sigma_m \sigma_n}, \tag{2}$$

where cov is the covariance, σm is the standard deviation of m and σn is the standard deviation of n. Values of $|\rho| > 0.8$ testify a strong correlation between m and n, values between 0.3 and 0.8 a moderate correlation, whereas values of $|\rho| < 0.30$ testify a weak correlation [56]. According to Song et al. [57], a correlation is considered "strong" if $|\rho| > 0.8$. According to Table 3, H, α, c, ϕ, β, γ, and ru are correlated in order of moderate to weakest. So, none of the parameters was deleted for developing the slope stability predicting model. From Table 3, the maximum absolute value of correlation coefficient is found to be 0.639 and no "strong" correlation exists between the different pairs of factors.

Table 3. Correlation coefficients between various factors.

Parameter	γ	c	ϕ	β	H	r_u
γ	1					
C	0.359	1				
ϕ	0.512	0.298	1			
B	0.496	0.429	0.639	1		
H	0.638	0.257	0.382	0.416	1	
r_u	0.059	-0.10	0.066	-0.00	-0.05	1

2.4. Data discretization

Data discretization is a method of converting continuous data into discrete data with a set of intervals. Several reasons are there to discretize data, the most significant of which are as follows [19]: i) simplifying the dataset; ii) make modeling simple and quick; iii) getting outputs that are simple to understand and iv) only discrete data can be employed in the statistical method. When no prior knowledge from domain experts is available, discretization methods are employed to discretize continuous variables. In the literature, there are various discretization methods. Among them are: equal frequency discretization [58], information-preserving discretization [59], error-based discretization [60], entropy-based discretization [61], and the one-rule discretization [62]. The six parameters employed are continuous. All six parameters in this study are continuous, and the data was partitioned into intervals with nearly the same number of cases using the equal frequency binning algorithm from the Waikato Environment for Knowledge Analysis (WEKA) software package. Table 4 shows the state intervals as well as their definitions.

Table 4. Intervals of input parameter values and their related states.

Parameter	Intervals/States		
γ	(12,18.92)/Low	(18.92,22.2) Medium	(22.2,31.3)/High
c	(0,11.985)/Low	(11.985,29.7)/Medium	(29.7,150)/High
ϕ	(0,25.5/Low	(25.5,34)/Medium	(34,45)/High
β	(9.792,29.6)/Low	(29.6,40.5)/Medium	(40.5,53)/High
H	(3.6,20.5)/Low	(20.5,89.25)/Medium	(89.25,511)/High
r_u	0/Dry	(0,0.5)/Wet	-
SS	0/Failed	1/Stable	-

2.5. Model evaluation criteria

To evaluate and compare the proposed model's performance to that of existing models in the literature, a number of measures were used: accuracy (*Acc*), precision (*Prec*), recall (*Rec*), *F-score*, and Matthews correlation coefficient (*Mcc*). The percentage of successfully identified samples to the total number of samples is called *Acc*. The *Prec* assesses the accuracy of predictions for a particular class (stable or unstable), whereas the *Rec* measures the accuracy of predictions only taking into account predicted values. The correlation coefficient between predicted and actual is measured by *Mcc*. The weighted harmonic mean of precision and recall is the *F-score*. All of the measures used are based on the confusion matrix.

Table 5. Confusion matrix for slope stability classification.

Actual condition	Predicted condition Stable (1)	Unstable (0)
Stable (1)	True Positive (TP)	False Negative (FN)
Unstable (0)	False Positive (FP)	True Negative (TN)

The confusion matrix is shown in Table 5, where true positive (TP) refers the number of correctly predicted stable slopes and true negative (TN) defines the number of correctly predicted unstable slopes. False positive (FP) reflects the number of unstable slopes that were wrongly predicted, whereas false negative (FN) represents the number of stable slopes that were incorrectly predicted. The mathematical equations of the performance metrics are given below respectively.

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}, \tag{3}$$

$$Prec = \frac{TP}{TP+FP} \ or \ \frac{TN}{TN+FN}, \tag{4}$$

$$Rec = \frac{TP}{TP+FN} \ or \ \frac{TN}{TN+FP}, \tag{5}$$

$$F-score = \frac{2\times Prec\times Rec}{Prec+Rec}, \tag{6}$$

$$Mcc = \frac{TP\times TN-FN\times FP}{\sqrt{(TP+FP)(TN+FP)(TN+FN)(TP+FN)}}. \tag{7}$$

Figure 3 depicts the overall flowchart of the TAN prediction model development procedure based

on the preceding description.

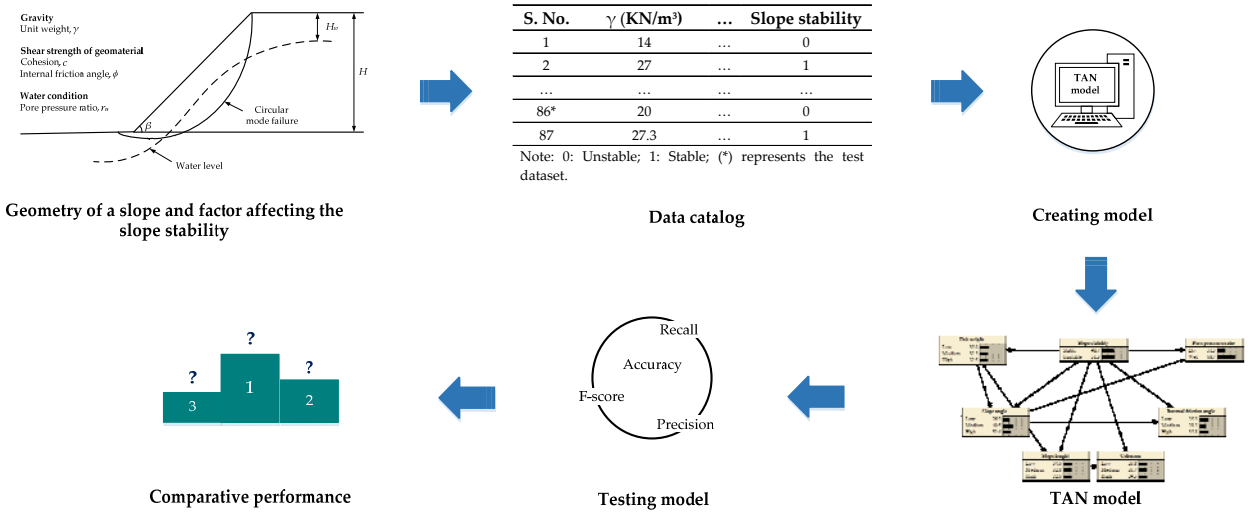

Figure 3. Flow methodology for slope stability prediction using TAN classifier.

3. Development of proposed model

The TAN algorithm was employed in this paper to develop a Bayesian belief network as presented in Figure 4. The TAN algorithm is used to create a network with 7 nodes and numerous lines. The lines linking the nodes represent the relationships between the variables, and the nodes represent the variables. Figure 4 depicts the hierarchical interactions of influencing and being affected by others among various slope stability parameters. The interaction of variables like "slope geometry", "geomaterial shear strength", and "water condition" results in slope stability, all of which are fully captured by the Bayesian belief network structure. In a Bayesian belief network, the interactional hierarchical relationship of variables can fully encompass the actual situation of slope stability.

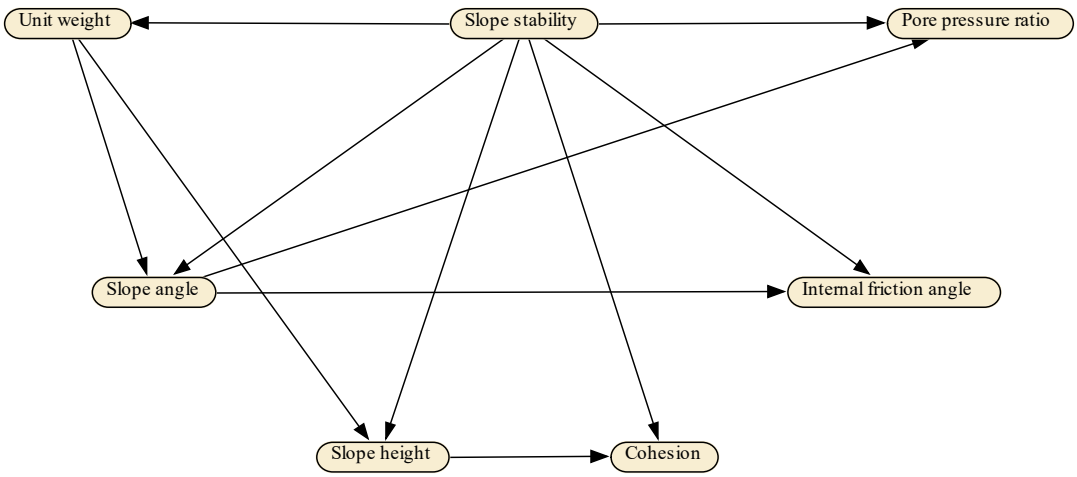

Figure 4. TAN structure of slope stability.

Once a Bayesian belief network topology has been established, parameter learning is carried out to obtain the conditional probability distribution of nodes in Netica. Finally, as illustrated in Figure 5, the Bayesian belief network model for the SS causation analysis can be established.

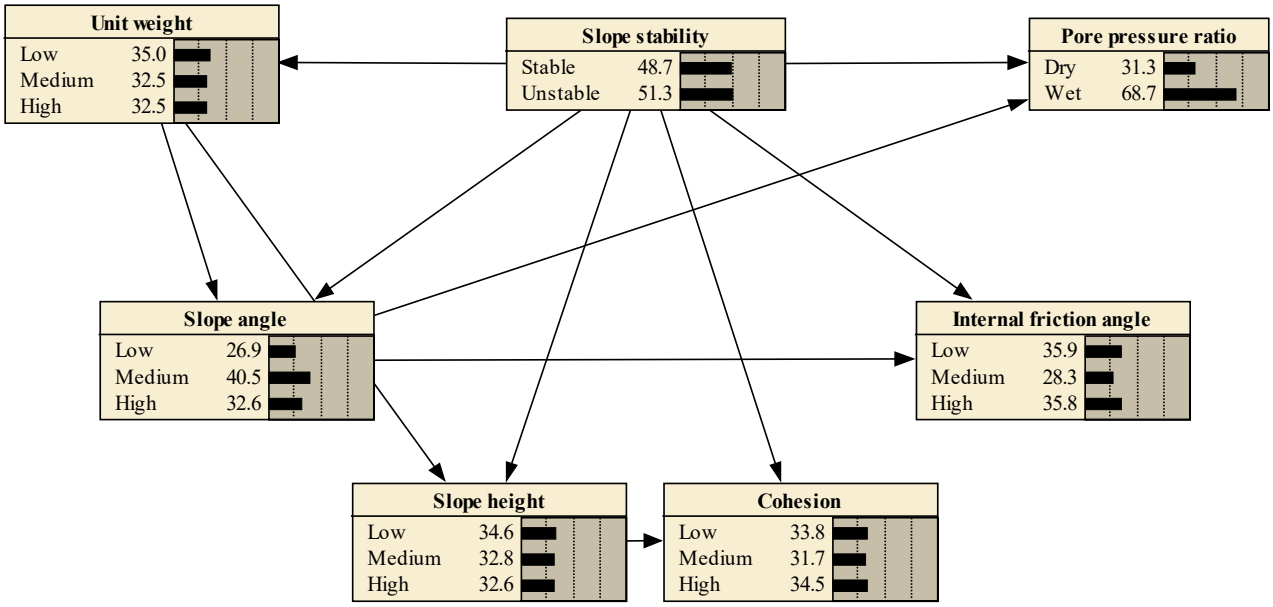

Figure 5. TAN model graphical result.

4. Results and discussion

4.1. TAN model's performance

The TAN model in Section 2.1 was built using historical data from 74 slope stability cases (unstable slope instances (38) and stable slope instances (36)). The sampling bias in the training dataset due to the class ratio of 38:36 for the 74 dataset is 1.05. Table 6 illustrates the training performance results, such as *Acc*, *Prec*, *Rec*, *F-score*, and *Mcc*. The accuracy of the 36 stable slope instances is 0.889 (called *Rec*, which is TP / (TP + FN)). The accuracy is 0.816 for the 38 unstable slope instances (called *Rec*, which is TN / (TN + FP)). The overall Acc is around 85.1 percent, and the *Mcc* is 0.705, both of which are excellent for practical engineering.

Table 6. Confusion matrix and associated TAN classifier training performance.

Actual	Predicted		*Prec*	*Rec*	*F-score*
	Stable	Unstable			
Stable	32	4	0.821	0.889	0.853
Unstable	7	31	0.886	0.816	0.849

*Note: *Mcc* = 0.705, *Acc* = 85.1%.

4.2. Validation of model

In this section, the model's performances are validated using testing dataset that have not been used during the process of model's construction. The significance of validation is to find the capabilities of developed model to be generalized for the conditions that have not been attended during training phase. As mentioned before, the testing dataset consist of 13 slope cases, which are shown in Table 7. It is worthwhile to mention here that the sampling bias in the testing dataset due to the class ratio (i.e., unstable: stable) of 7:6 for the 13 dataset is 1.16. For these 13 cases, the input parameters were fed into developed TAN-based models and the predicted values for SS were obtained. The method of SS prediction for the first case of testing dataset has been schematically indicated in Figure 6. A comparison of predicted and real values of SS for the testing dataset has been given in Table 7. This model has only one unsuccessful prediction case and its overall accuracy is equal to 92.3%. From practical point of view, these results show that the developed TAN classification model is useful and efficient. Finally, to assess the accuracy of developed TAN classification model, it was compared with recently developed soft computing/data mining models in the literature. Table 8 shows the results of this comparison. The confusion matrices were developed using Table 5, and the *Prec*, *Rec*, *F-score*, *Mcc*, and *Acc* were determined using Eqs (3)–(7) in Table 8. As can be seen, the prediction accuracy of proposed TAN classification model is as good as those of other reported techniques. The fundamental advantage of the proposed model is that it may be considered as a "white box" that clearly demonstrates the link between input and output parameters. As a result, users (geotechnical engineers) may use these models to analyze and predict slope stability quickly.

Table 7. Results of test dataset.

No.	γ/kNm^{-3}	c/kPa	ϕ/°	β/°	H/m	r_u	*FoS*	Actual slope stability	Predicted with TAN (P(Stable))
1	21.43	0	20	20	61	0.5	1.03	Unstable	Unstable (1.11%)
2	27	32	33	42.4	289	0.25	1.3	Stable	Stable (77.6%)
3	18.8	25.1	10	25	50	0.2	1.18	Unstable	Unstable (26.8%)
4	14	12	26	30	88	0	1.02	Unstable	Unstable (5.16%)
5	20	10.1	29	34	6	0.3	1.34	Stable	Stable (87.8%)
6	27	35	35	42	359	0.25	1.27	Stable	Stable (77.6%)
7	14.8	0	17	20	50	0	1.13	Unstable	Unstable (32.1%)
8	19.6	12	20	22	12.2	0.405	1.35	Unstable	Stable (51.9%)
9	27.3	31.5	29.7	41	135	0.25	1.245	Stable	Stable (75.7%)
10	20	20	36	45	50	0.25	0.96	Unstable	Unstable (2.65%)
11	28.4	29.4	35	35	100	0	1.78	Stable	Stable (96.6%)
12	24	0	40	33	8	0.3	1.58	Stable	Stable (96.4%)
13	20	0	36	45	50	0.5	0.67	Unstable	Unstable (1.48%)

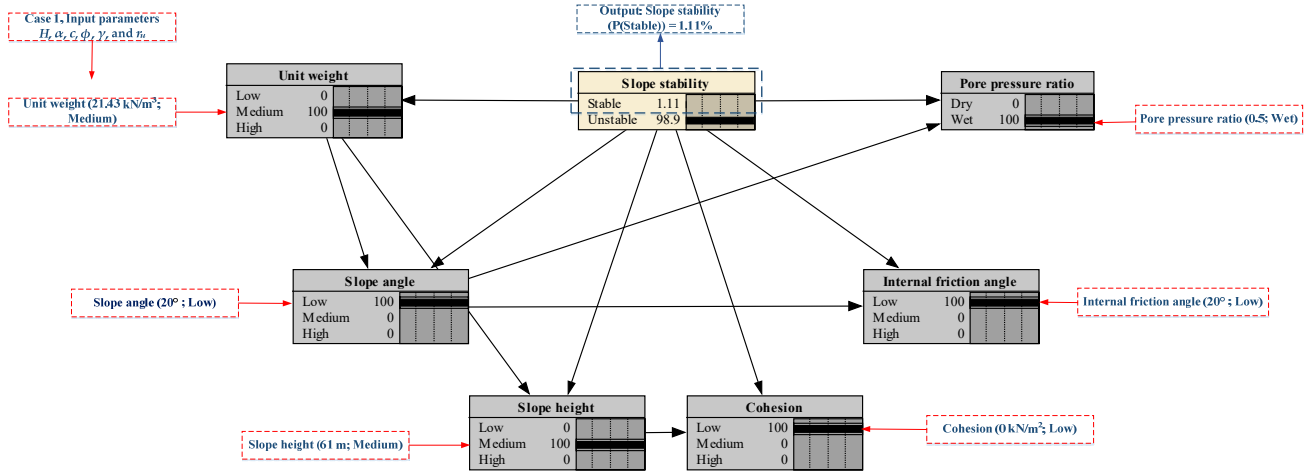

Figure 6. A schematic example of SS prediction.

Table 8. Comparative performance evaluation of the test set.

Model	Actual	Predicted		*Acc* (%)	*Mcc*	*Prec*	*Rec*	*F-score*	Reference
		Stable	Unstable						
SVM	Stable	21	1	73.1	0.541	0.618	0.955	0.750	Zhou et al. [49]
	Unstable	13	17			0.944	0.567	0.708	
ANN	Stable	21	1	82.7	0.684	0.724	0.955	0.824	
	Unstable	8	22			0.957	0.733	0.830	
RF	Stable	21	1	80.8	0.655	0.700	0.955	0.808	
	Unstable	9	21			0.955	0.700	0.808	
GBM	Stable	21	1	86.5	0.746	0.778	0.955	0.857	
	Unstable	6	24			0.960	0.800	0.873	
NB	Stable	7	1	84.6	0.675	0.875	0.875	0.875	Feng et al. [43]
	Unstable	1	4			0.800	0.800	0.800	
RF	Stable	8	1	83.3	0.671	0.800	0.889	0.842	Lin et al. [63]
	Unstable	2	7			0.875	0.778	0.824	
SVM	Stable	4	5	66.7	0.372	0.800	0.444	0.571	
	Unstable	1	8			0.615	0.889	0.727	
NB	Stable	3	6	55.6	0.124	0.600	0.333	0.429	
	Unstable	2	7			0.538	0.778	0.636	
GSA	Stable	8	1	88.9	0.778	0.889	0.889	0.889	
	Unstable	1	8			0.889	0.889	0.889	
TAN	Stable	6	0	92.3	0.857	0.857	1.000	0.923	Present study
	Unstable	1	6			1.000	0.857	0.923	

*Note GSA: Gravitational search algorithm; TAN: tree augmented naive bayes; Mcc: Matthews correlation coefficient.

4.3. Causal inference

System fault diagnostics is also another useful application of the Bayesian belief network. The

bidirectional reasoning technology of the Bayesian belief network can quantify not only the probability of a system failure under combined fault conditions, as well as the posterior probabilities of different components under the system fault condition, allowing users to quickly ascertain the most likely combination that caused system failure. Computational analysis becomes more intuitive and adaptable as a result of this. Consider the "unstable" state in "slope stability" as an example of causal inference. Because the evidence variable is "unstable" in this example, the status probability is 100%. As seen in Figure 7, using Netica's automated updating feature, the probability of "cohesion" state "low" increases significantly from 33.8 to 40.6% after inputting the data. In addition, the probability of "low" in "internal friction angle" increases from 35.9 to 47.2%, reaching the maximum probability. In the absence of additional evidence, this shows that "low" grades of cohesion and internal friction angle are the most likely cause of "unstable" state in slope stability.

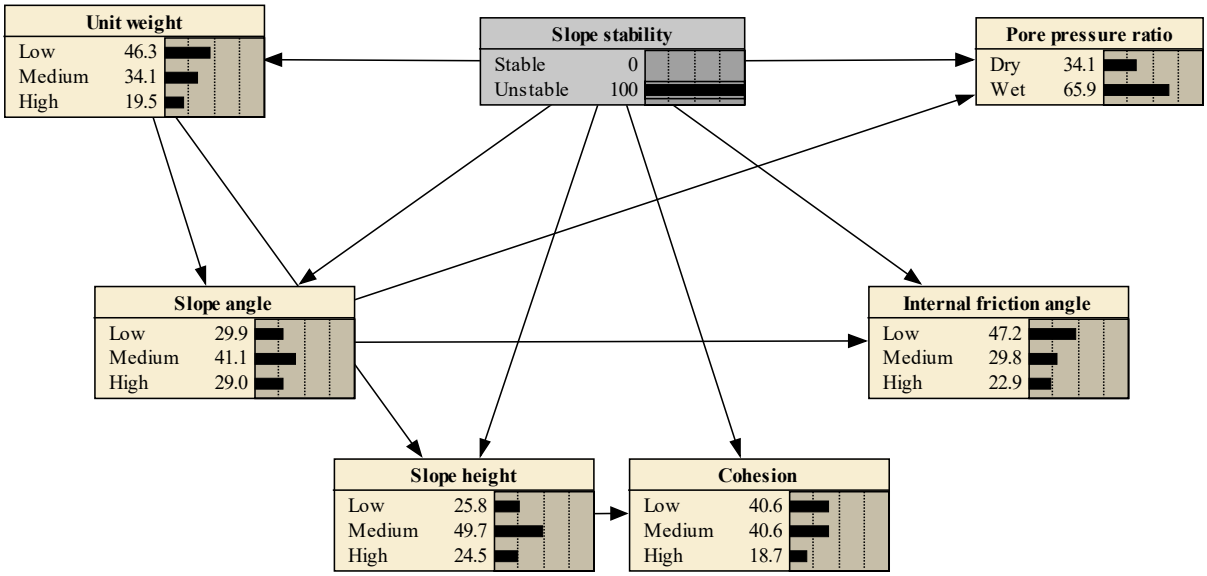

Figure 7. The posterior probability when the evidence variable in slope stability is "unstable".

4.4. Most probable explanation

The TAN model can be used to find the most probable explanations from sets of multiple causes (node states) that are likely to lead to a conclusion; Netica can be used to find the set that is most likely to lead to the result, and the set with the maximum likelihood will be the most probable explanation. Figure 8 depicts the most probable explanation cause (node state) set of "unstable" slope is {pore pressure ratio (ru): wet, slope height (H): low, internal friction angle (ϕ): low, slope angle (β): medium, cohesion (c): medium}.

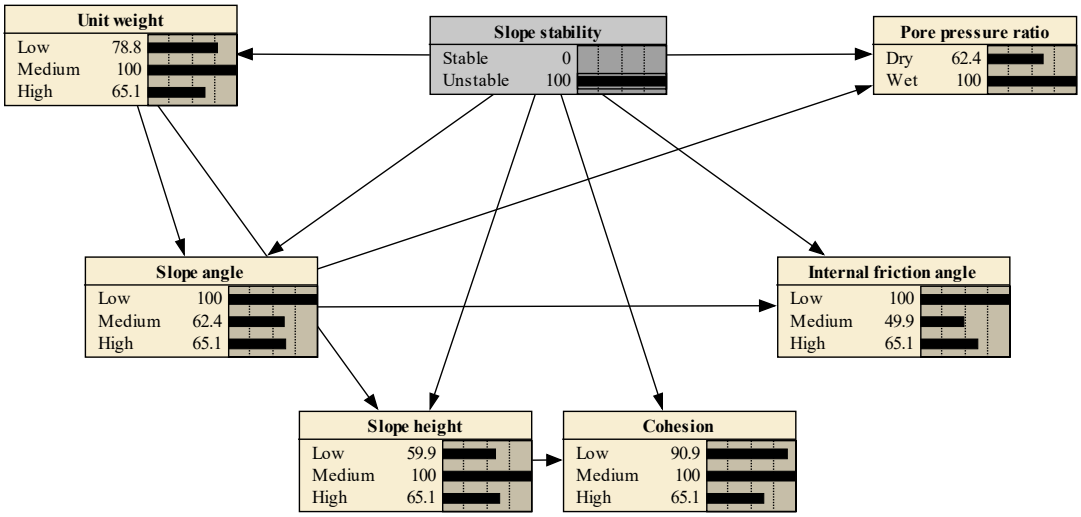

Figure 8. The MPE when the slope stability state is "unstable".

4.5. Sensitivity analysis

To examine the impact of each factor on the slope stability, a sensitivity analysis was performed on six input factors. Mutual information between nodes can reveal whether or not they are interconnected and, if so, how close they are [64]. According to the sensitivity analysis, a basic event with a reasonably large contribution to the probability of a resulting event makes it easier to reduce the probability of these basic events by taking into account effective measures, thereby lowering the probability of a resulting event. For sensitivity analysis, the target node "slope stability" is selected, and the results are displayed in Table 9. Table 9 shows that node "slope height" has the highest mutual info (= 0.10294), that implying the greatest impact on "slope stability", followed by "cohesion" and "unit weight", which have mutual info = 0.08706 and 0.06945, respectively.

Table 9. Sensitivity analysis of "slope stability".

Node	H	c	γ	ϕ	β	r_u
Mutual info	0.10294	0.08706	0.06945	0.06403	0.00589	0.00262
Percent	10.3	8.71	6.95	6.41	0.589	0.263
Variance of beliefs	0.0341049	0.0293855	0.0235044	0.0217402	0.0020344	0.0009074

5. Conclusions

In this study, TAN model was trained and tested using a circular mode failure slope stability database acquired from the literature to predict slope stability based on the input variables such as γ, c, ϕ, β, H and ru. The following are the major findings of this study:

1) The results obtained from TAN modeling suggest that the TAN model has an appropriate capability to accurate prediction of the SS for circular slip failure. The TAN-based model also gives improved performance than other models (i.e., SVM, RF, NB) proposed in literature.

2) Results of sensitivity conclude that the slope height (mutual info = 0.10294) is the main important parameter when the TAN-based model is selected for prediction of SS for circular mode failure for this dataset.
3) The "most probable explanation" set of "unstable" slope is {unit weight (γ): medium, pore pressure ratio (r_u): wet, slope height (H): low, internal friction angle (ϕ): low, slope angle (β): medium, cohesion (c): medium}. This is quite compatible with engineering judgment and well matched.

Follow-up research will look at the rationality of the TAN model as well as other parameters such as rainwater infiltration that could lead to slope instability, in order to develop a more accurate and comprehensive model. Since the TAN model is a probabilistic model, it requires more detailed and extensive basic data to improve its reliability. Furthermore, because the influencing slope stability parameters in reality are greater than that considered in this study, and as the TAN model is also appropriate for the development of a larger and more complex slope stability analysis model, the model can be expanded to a more sophisticated model that takes into account more parameters such as applied seismic acceleration, depth of rock, soil type, and rainfall characteristics.

Acknowledgments

The work was supported by the National Key Research and Development Plan of China under Grant No. 2021YFB2600703.

Conflict of interests

The authors declare no conflict of interest.

References

1. P. Lu, M. Rosenbaum, Artificial neural networks and grey systems for the prediction of slope stability, *Nat. Hazards*, **30** (2003), 383–398. https://doi.org/10.1023/B:NHAZ.0000007168.00673.27
2. M. Y. Cheng, N. D. Hoang, Typhoon-induced slope collapse assessment using a novel bee colony optimized support vector classifier, *Nat. Hazards*, **78** (2015), 1961–1978. https://doi.org/10.1007/s11069-015-1813-8
3. F. Kang, J. Li, Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes, *J. Comput. Civ. Eng.*, **30** (2016). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000514
4. D. T. Bui, B. Pradhan, O. Lofman, I. Revhaug, Ø. B. Dick, Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam, *Nat. Hazards*, **66** (2013), 707–730. https://doi.org/10.1007/s11069-012-0510-0
5. M. G. Sakellariou, M. D. Ferentinou, A study of slope stability prediction using neural networks, *Geotech. Geol. Eng.*, **23** (2005), 419–445. https://doi.org/10.1007/s10706-004-8680-5
6. A. W. Bishop, The use of the slip circle in the stability analysis of slopes, *Geotechnique*, **5** (1955), 7–17. https://doi.org/10.1680/geot.1955.5.1.7
7. N. R. Morgenstern, V. E. Price, The analysis of the stability of general slip surfaces, *Geotechnique*, **15** (1965), 79–93. https://doi.org/10.1680/geot.1965.15.1.79

8. S. K. Sarma, Stability analysis of embankments and slopes, *J. Geotech. Eng. Div.*, **105** (1979), 1511–1524. https://doi.org/10.1061/AJGEB6.0000903
9. E. Spencer, A method of analysis of the stability of embankments assuming parallel inter-slice forces, *Geotechnique*, **17** (1967), 11–26. https://doi.org/10.1680/geot.1967.17.1.11
10. O. C. Zienkiewicz, C. Humpheson, R. W. Lewis, Associated and non-associated visco-plasticity and plasticity in soil mechanics, *Geotechnique*, **25** (1975), 671–689. https://doi.org/10.1680/geot.1975.25.4.671
11. D. V. Griffiths, P. A. Lane, Slope stability analysis by finite elements, *Geotechnique*, **49** (1999), 387–403. https://doi.org/10.1680/geot.1999.49.3.387
12. M. Huang, C. Q. Jia, Strength reduction FEM in stability analysis of soil slopes subjected to transient unsaturated seepage, *Comput. Geotech.*, **36** (2009), 93–101. https://doi.org/10.1016/j.compgeo.2008.03.006
13. R. Baker, *Variational Approach to Slope Stability*, 1977.
14. Y. M. Cheng, L. Li, S. C. Chi, W. B. Wei, Particle swarm optimization algorithm for the location of the critical non-circular failure surface in two-dimensional slope stability analysis, *Comput. Geotech.*, **34** (2007), 92–103. https://doi.org/10.1016/j.compgeo.2006.10.012
15. M. Ercanoglu, C. Gokceoglu, Assessment of landslide susceptibility for a landslide-prone area (north of Yenice, NW Turkey) by fuzzy approach, *Environ. Geol.*, **41** (2002), 720–730. https://doi.org/10.1007/s00254-001-0454-2
16. H. Fattahi, Prediction of slope stability using adaptive neuro-fuzzy inference system based on clustering methods, *J. Min. Environ.*, **8** (2017), 163–177. https://doi.org/10.22044/jme.2016.637
17. W. Gao, Stability analysis of rock slope based on an abstraction ant colony clustering algorithm, *Environ. Earth Sci.*, **73** (2015), 7969–7982. https://doi.org/10.1007/s12665-014-3956-4
18. N. D. Hoang, A. D. Pham, Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: A multinational data analysis, *Expert Syst. Appl.*, **46** (2016), 60–68. https://doi.org/10.1016/j.eswa.2015.10.020
19. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, Evaluation of liquefaction-induced lateral displacement using Bayesian belief networks, *Front. Struct. Civ. Eng.*, **15** (2021), 80–98. https://doi.org/10.1007/s11709-021-0682-3
20. M. Ahmad, J. L. Hu, F. Ahmad, X. W. Tang, M. Amjad, M. J. Iqbal, et al., Supervised learning methods for modeling concrete compressive strength prediction at high temperature, *Materials*, **14** (2021), 1983. https://doi.org/10.3390/ma14081983
21. M. Ahmad, P. Kaminski, P. Olczak, M. Alam, M. J. Iqbal, F. Ahmad, et al., Development of prediction models for shear strength of rockfill material using machine learning techniques, *Appl. Sci.*, **11** (2021), 6167. https://doi.org/10.3390/app11136167
22. M. Ahmad, X. W. Tang, F. Ahmad, Evaluation of liquefaction-induced settlement using random forest and REP tree models: taking pohang earthquake as a case of llustration, in *Natural Hazards-Impacts, Adjustments and Resilience*, Intech Open: London, UK, 2020. https://doi.org/10.5772/intechopen.94274
23. M. Ahmad, J. L. Hu, M. H. Nyarko, F. Ahmad, X. W. Tang, Z. U. Rahman, et al., Rockburst hazard prediction in underground projects using two intelligent classification techniques: A comparative study, *Symmetry*, **13** (2021), 632. https://doi.org/10.3390/sym13040632

24. M. Ahmad, X. W. Tang, J. N. Qiu, W. J. Gu, F. Ahmad, A hybrid approach for evaluating CPT-based seismic soil liquefaction potential using Bayesian belief networks, *J. Cent. South Univ.*, **27** (2020), 500–516. https://doi.org/10.1007/s11771-020-4312-3
25. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, Evaluating seismic soil liquefaction potential using bayesian belief network and C4.5 decision tree approaches, *Appl. Sci.*, **9** (2019), 4226. https://doi.org/10.3390/app9204226
26. M. Ahmad, N. A. Shayea, X. W. Tang, A. Jamal, H. M. A. Ahmadi, F. Ahmad, Predicting the pillar stability of underground mines with random trees and C4.5 decision trees, *Appl. Sci.*, **10** (2020), 6486. https://doi.org/10.3390/app10186486
27. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, W. J. Gu, A step forward towards a comprehensive framework for assessing liquefaction land damage vulnerability: Exploration from historical data, *Front. Struct. Civ. Eng.*, **14** (2020), 1476–1491. https://doi.org/10.1007/s11709-020-0670-z
28. M. Ahmad, X. W. Tang, J. N. Qiu, F. Ahmad, W. J. Gu, Application of machine learning algorithms for the evaluation of seismic soil liquefaction potential, *Front. Struct. Civ. Eng.*, **15** (2021), 490–505. https://doi.org/10.1007/s11709-020-0669-5
29. M. Ahmad, F. Ahmad, P. Wroblewski, R. A. A. Mansob, P. Olczak, P. Kamiński, et al., Prediction of ultimate bearing capacity of shallow foundations on cohesionless soils: A gaussian process regression approach, *Appl. Sci.*, **11** (2021), 10317. https://doi.org/10.3390/app112110317
30. C. X. Yang, L. G. Tham, X. T. Feng, Y. J. Wang, P. K. K. Lee, Two-stepped evolutionary algorithm and its application to stability analysis of slopes, *J. Comput. Civ. Eng.*, **18** (2004), 145–153. https://doi.org/10.1061/(ASCE)0887-3801(2004)18:2(145)
31. H. B. Wang, W. Y. Xu, R. C. Xu, Slope stability evaluation using back propagation neural networks, *Eng. Geol.*, **80** (2005), 302–315. https://doi.org/10.1016/j.enggeo.2005.06.005
32. P. Samui, Slope stability analysis: a support vector machine approach, *Environ. Geol.*, **56** (2008), 255. https://doi.org/10.1007/s00254-007-1161-4
33. H. B. Zhao, Slope reliability analysis using a support vector machine, *Comput. Geotech.*, **35** (2008), 459–467. https://doi.org/10.1016/j.compgeo.2007.08.002
34. A. Choobbasti, F. Farrokhzad, A. Barari, Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran), *Arabian J. Geosci.*, **2** (2009), 311–319. https://doi.org/10.1007/s12517-009-0035-3
35. A. Ahangar-Asr, A. Faramarzi, A. A. Javadi, A new approach for prediction of the stability of soil and rock slopes, *Eng. Comput.*, **27** (2010). https://doi.org/10.1108/02644401011073700
36. S. K. Das, R. K. Biswal, N. Sivakugan, B. Das, Classification of slopes and prediction of factor of safety using differential evolution neural networks, *Environ. Earth Sci.*, **64** (2011), 201–210. https://doi.org/10.1007/s12665-010-0839-1
37. S. Li, H. B. Zhao, Z. L. Ru, Slope reliability analysis by updated support vector machine and Monte Carlo simulation, *Nat. Hazards*, **65** (2013), 707–722. https://doi.org/10.1007/s11069-012-0396-x
38. L. Dong, X. Li, Comprehensive models for evaluating rockmass stability based on statistical comparisons of multiple classifiers, *Math. Probl. Eng.*, **2013** (2013). https://doi.org/10.1155/2013/395096
39. A. Manouchehrian, J. Gholamnejad, M. Sharifzadeh, Development of a model for analysis of slope stability for circular mode failure using genetic algorithm, *Environ. Earth Sci.*, **71** (2014), 1267–1277. https://doi.org/10.1007/s12665-013-2531-8

40. Z. Zhang, Z. Liu, L. Zheng, Y. Zhang, Development of an adaptive relevance vector machine approach for slope stability inference, *Neural Comput. Appl.*, **25** (2014), 2025–2035. https://doi.org/10.1007/s00521-014-1690-1
41. Z. Liu, J. Shao, W. Xu, H. Chen, Y. Zhang, An extreme learning machine approach for slope stability evaluation and prediction, *Nat. Hazards*, **73** (2014), 787–804. https://doi.org/10.1007/s11069-014-1106-7
42. X. Xue, X. Yang, X. Chen, Application of a support vector machine for prediction ofslope stability, *Sci. China: Technol. Sci.*, **57** (2014), 2379–2386. https://doi.org/10.1007/s11431-014-5699-6
43. X. Feng, S. Li, C. Yuan, P. Zeng, Y. Sun, Prediction of slope stability using naive Bayes classifier, *KSCE J. Civ. Eng.*, **22** (2018), 941–950. https://doi.org/10.1007/s12205-018-1337-3
44. C. Qi, X. Tang, Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study, *Comput. Ind. Eng.*, **118** (2018), 112–122. https://doi.org/10.1016/j.cie.2018.02.028
45. P. A. Sari, M. Suhatril, N. Osman, M. A. Muazu, H. Dehghani, Y. Sedghi, et al., An intelligent based-model role to simulate the factor of safe slope by support vector regression, *Eng. Comput.*, **35** (2019), 1521–1531. https://doi.org/10.1007/s00366-018-0677-4
46. W. Gao, M. Raftari, A. S. A. Rashid, M. A. Muazu, W. A. W. Jusoh, A predictive model based on an optimized ANN combined with ICA for predicting the stability of slopes, *Eng. Comput.*, **36** (2020), 325–344. https://doi.org/10.1007/s00366-019-00702-7
47. C. Yuan, H. Moayedi, The performance of six neural-evolutionary classification techniques combined with multi-layer perception in two-layered cohesive slope stability analysis and failure recognition, *Eng. Comput.*, **36** (2020), 1705–1714. https://doi.org/10.1007/s00366-019-00791-4
48. M. Sari, Stability analysis of slopes prone to circular failures using logistic regressionin, in *Conference of the Arabian Journal of Geosciences*, (2018), 355–357. https://doi.org/10.1007/978-3-030-01665-4_82
49. J. Zhou, E. Li, S. Yang, M. Wang, X. Shi, S. Yao, et al., Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, *Saf. Sci.*, **118** (2019), 505–518. https://doi.org/10.1016/j.ssci.2019.05.046
50. C. Chow, C. Liu, Approximating discrete probability distributions with dependence trees, *IEEE Trans. Inf. Theory*, **14** (1968), 462–467. https://doi.org/10.1109/TIT.1968.1054142
51. N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, *Mach. Learn.*, **29** (1997), 131–163. https://doi.org/10.1023/A:1007465528199
52. X. Wu, V. Kumar, The top ten algorithms in data mining, CRC Press, 2009.
53. N. Sah, P. Sheorey, L. Upadhyaya, Maximum likelihood estimation of slope stability, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.*, **31** (1994), 47–53. https://doi.org/10.1016/0148-9062(94)92314-0
54. S. Rukhaiyar, M. N. Alam, N. K. Samadhiya, A PSO-ANN hybrid model for predicting factor of safety of slope, *Int. J. Geotech. Eng.*, **12** (2018), 556–566. https://doi.org/10.1080/19386362.2017.1305652
55. X. Xue, Prediction of slope stability based on hybrid PSO and LSSVM, *J. Comput. Civ. Eng.*, **31** (2017), 04016041. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000607
56. T. V. Vuren, Modeling of transport demand – analyzing, calculating, and forecasting transport demand, *Transp. Rev.*, **40** (2020), 115–117. https://doi.org/10.1080/01441647.2019.1635226

57. Y. Song, J. Gong, S. Gao, D. Wang, T. Cui, Y. Li, et al., Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China, *Comput. Geosci.*, **42** (2012), 189–199. https://doi.org/10.1016/j.cageo.2011.09.011
58. J. Dougherty, R. Kohavi, M. Sahami, Supervised and unsupervised discretization of continuous features, *Mach. Learn. Proc.*, (1995), 194–202. https://doi.org/10.1016/B978-1-55860-377-6.50032-3
59. A. Hartemink, *Principled Computational Methods for the Validation and Discovery of Genetic Regulatory Networks*, Ph.D. Dissertation, Massachusetts Institute of Technology, 2001. http://hdl.handle.net/1721.1/8699
60. W. Maass, Efficient agnostic PAC-learning with simple hypothesis, in *Proceedings of the Seventh Annual Conference on Computational Learning Theory*, (1994), 67–75. https://doi.org/10.1145/180139.181016
61. U. Fayyad, K. Irani, Multi-interval discretization of continuous-valued attributes for classification learning, 1993. http://hdl.handle.net/2014/35171
62. R. C. Holte, Very simple classification rules perform well on most commonly used datasets, *Mach. Learn.*, **11** (1993), 63–90. https://doi.org/10.1023/A:1022631118932
63. Y. Lin, K. Zhou, J. Li, Prediction of slope stability using four supervised learning methods, *IEEE Access*, **6** (2018), 31169–31179. https://doi.org/10.1109/ACCESS.2018.2843787
64. J. Cheng, R. Greiner, J. Kelly, D. Bell, W. Liu, Learning Bayesian networks from data: an information-theory based approach, *Artif. Intell.* **137** (2002), 43–90. https://doi.org/10.1016/S0004-3702(02)00191-1

Appendix

Table A1. Dataset used to construct and validate the model.

No.	γ/kNm^{-3}	c/kPa	ϕ/°	β/°	H/m	r_u	FoS	SS
1	14	11.97	26	30	88	0.45	0.625	0
2	27	37.5	35	37.8	320	0.25	1.24	1
3	12	0	30	35	4	0	1.46	1
4	22.4	10	35	45	10	0.4	0.9	0
5	21	35	28	40	12	0.5	1.43	1
6*	20	10.1	29	34	6	0.3	1.34	1
7	27	40	35	47.1	292	0.25	1.15	0
8*	28.4	29.4	35	35	100	0	1.78	1
9*	27.3	31.5	29.7	41	135	0.25	1.245	1
10	22	20	22	20	180	0.1	0.99	0
11	22.4	10	35	30	10	0	2	1
12	27.3	10	39	41	511	0.25	1.434	1
13	19	30	35	35	11	0.2	2	1
14	27.3	10	39	40	470	0.25	1.418	1
15*	14	12	26	30	88	0	1.02	0
16	19.1	10.1	10	25	50	0.4	0.65	0
17	18.7	26.4	15	35	8.2	0	1.11	0

Continued on next page

No.	γ/kNm^{-3}	c/kPa	ϕ/°	β/°	H/m	r_u	FoS	SS
18	20	0	36	45	50	0.25	0.79	0
19	22	20	22	20	180	0	1.12	0
20*	19.6	12	20	22	12.2	0.405	1.35	0
21	16	70	20	40	115	0	1.11	0
22	19	11.7	28	35	21	0.11	1.09	0
23	21	45	25	49	12	0.3	1.53	1
24	20	20	36	45	50	0.5	0.83	0
25	18.8	30	20	30	50	0.1	1.46	1
26*	14.8	0	17	20	50	0	1.13	0
27*	27	35	35	42	359	0.25	1.27	1
28	20	0	24.5	20	8	0.35	1.37	1
29	18	24	30.2	45	20	0.12	1.12	0
30	25	46	36	44.5	299	0.25	1.55	1
31*	27	32	33	42.4	289	0.25	1.3	1
32	22	0	36	45	50	0	0.89	0
33	18.8	20	10	25	50	0.3	0.97	0
34	18.8	25.1	20	30	50	0.2	1.21	0
35	27.3	10	39	40	480	0.25	1.45	1
36	27.3	16.8	28	50	90.5	0.25	1.252	1
37	20	40.1	30	30	15	0.3	1.84	1
38	18.8	14.4	25	20	30.6	0	1.88	1
39	21.5	6.9	30	31	76.8	0.38	1.01	0
40	14	11.97	26	30	88	0	1.02	0
41	26	150	45	50	200	0	1.2	1
42	25	46	35	46	432	0.25	1.23	1
43	18.5	12	0	30	6	0	0.78	0
44	18	45	25	25	14	0.3	2.09	1
45	22.4	100	45	45	15	0.25	1.8	1
46	20.6	16.2	26.5	30	40	0	1.25	0
47	25	46	35	50	284	0.25	1.34	1
48	18.8	20	20	30	50	0.3	1	0
49	21	20	40	40	12	0	1.84	1
50*	18.8	25.1	10	25	50	0.2	1.18	0
51	23.47	0	32	37	214	0	1.08	0
52*	21.43	0	20	20	61	0.5	1.03	0
53	18.5	25	0	30	6	0	1.09	0
54	31.3	68	37	49	200.5	0.25	1.2	0
55	28.4	39.2	38	35	100	0	1.99	1
56	18.8	14.4	25	20	30.6	0.45	1.11	0
57	27.3	14	31	41	110	0.25	1.249	1
58	31.3	68	37	46	366	0.25	1.2	0
59	20	40.1	40	40	10	0.2	2.31	1

Continued on next page

No.	γ/kNm^{-3}	c/kPa	ϕ/°	β/°	H/m	r_u	FoS	SS
60	21.8	8.6	32	28	12.8	0.49	1.03	0
61	18.8	30	10	25	50	0.1	1.4	1
62	18.84	0	20	20	7.62	0.45	1.05	0
63	18.8	10.4	21.3	34	37	0.3	1.29	0
64	20.4	24.9	13	22	10.6	0.35	1.4	1
65	27	32	33	42.6	301	0.25	1.16	0
66	22	0	40	33	8	0.35	1.45	1
67	21.4	10	30.34	30	20	0	1.7	1
68*	20	0	36	45	50	0.5	0.67	0
69	16.5	11.6	0	30	3.6	0	1	0
70	18.8	57.5	20	20	30.6	0	2.04	1
71	12	0	30	45	8	0	0.8	0
72	18	5	30	20	8	0.3	2.05	1
73	18.84	14.36	25	20	30.5	0.45	1.11	0
74	19.1	10.1	20	30	50	0.4	0.65	0
75	25	46	35	47	443	0.25	1.28	1
76	18.8	24.8	21.3	29.2	37	0.5	1.07	0
77	22	20	36	45	50	0	1.02	0
78	25	120	45	53	120	0	1.3	1
79	23	0	20	20	100	0.3	1.2	0
80	20.4	33.5	11	16	45.8	0.2	1.28	0
81	25	46	35	44	435	0.25	1.37	1
82	18.8	15.3	30	25	10.6	0.38	1.63	1
83	21	30	35	40	12	0.4	1.49	1
84*	24	0	40	33	8	0.3	1.58	1
85	14	12	26	30	88	0.45	0.63	0
86*	20	20	36	45	50	0.25	0.96	0
87	27.3	26	31	50	92	0.25	1.246	1

*Note 0: Unstable; 1: Stable; (*) represents the test dataset.