SCIENTIFIC REPORTS
natureresearch

Check for updates

OPEN

Prediction of slope failure in open-pit mines using a novel hybrid artificial intelligence model based on decision tree and evolution algorithm

Xuan-Nam Bui[1,2], Hoang Nguyen[3]✉, Yosoon Choi[4]✉, Trung Nguyen-Thoi[5,6], Jian Zhou[7] & Jie Dou[8]

In this study, the objective was to develop a new and highly-accurate artificial intelligence model for slope failure prediction in open-pit mines. For this purpose, the M5Rules algorithm was combined with a genetic algorithm (GA) in a novel hybrid technique, named M5Rules–GA model, for slope stability estimation and analysis and 450-slope observations in an open-pit mine in Vietnam were modeled using the Geo-Studio software based on essential parameters. The factor of safety was used as the model outcome. Artificial neural networks (ANN), support vector regression (SVR), and previously introduced models (such as FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC) were also developed for evaluating the proposed M5Rules–GA model. The evaluation of the model performance involved applying and computing the determination coefficient, variance account for, and root mean square error, as well as a general ranking and color scale. The results confirmed that the proposed M5Rules–GA model is a robust tool for analyzing slope stability. The other investigated models yielded less robust performance under the evaluation metrics.

Slope collapse is a critical hazard in open-pit mines as it can be of any scale, small or large, and directly affect people, equipment, and production processes (Fig. 1). Therefore, analysis and calculation of slope stability are of vital importance for preventing disasters that occur because of such instabilities.

Numerical methods[1,2] and three-dimensional techniques[3] for slope stability analysis were applied to various geological structures. Wei, *et al.*[4] combined the generalized Hoek-Brown and strength reduction method to evaluate the stability of slopes in rock mass. Many types of seismic actions were also investigated and assessed for the stability of slopes in different conditions[5–9]. However, owing to the complexity of geological structures, slope stability is a challenging aspect for large open-pit mining projects[10].

Furthermore, as soil layers exhibit heterogeneous characteristics, geotechnical and geological uncertainties can worsen the poor estimation of slope stability[11,12]. The random finite element method (RFEM) and limit equilibrium method (LEM), as well as other modeling methods based on finite elements and stochastic simulation, were typically applied to calculate slope stability[13–16].

In recent years, more advanced computational techniques have been widely applied in many fields[17–20], particularly in the prediction of landslides including slope stability[21–23]. Artificial intelligence (AI) is a powerful tool capable

[1]Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam. [2]Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, Duc Thang, Bac Tu Liem, Hanoi, Vietnam. [3]Institute of Research and Development, Duy Tan University, Da Nang, 550000, Vietnam. [4]Department of Energy Resources Engineering, Pukyong National University, Busan, 48513, Korea. [5]Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam. [6]Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam. [7]School of Resources and Safety Engineering, Central South University, Changsha, Hunan, 410083, China. [8]Civil and Environmental Engineering, Nagaoka University of Technology, 1603-1, Kami-Tomioka, Nagaoka, Niigata, 940-2188, Japan. ✉e-mail: nguyenhoang23@duytan.edu.vn; energy@pknu.ac.kr

Figure 1. Slope collapse disaster in an open-pit mine in Vietnam.

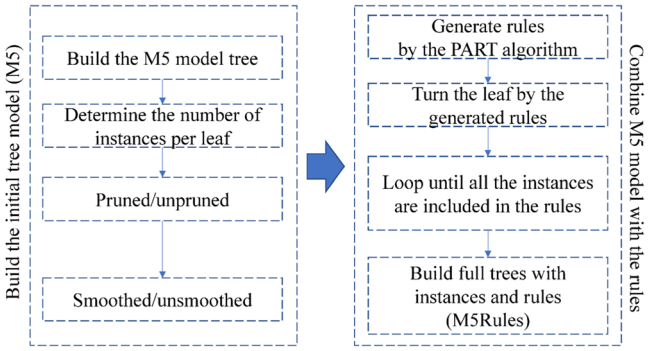


Figure 2. Workflow of the M5Rules model.

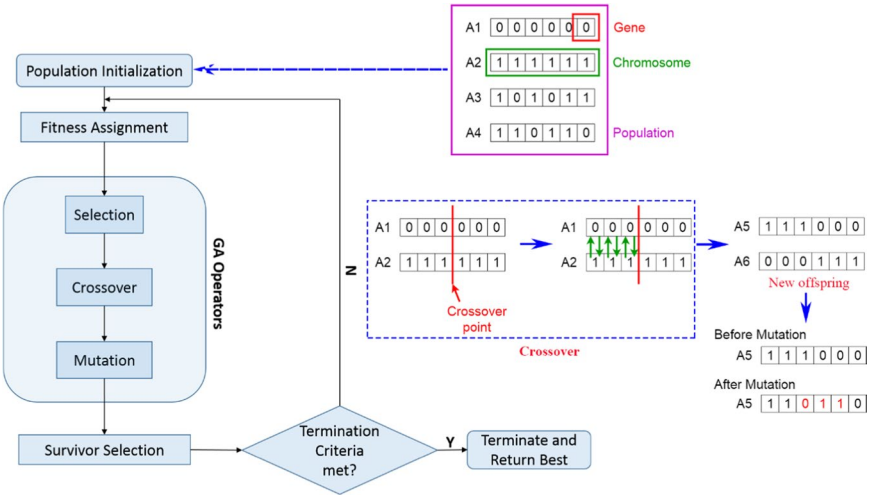


Figure 3. Description of GA with operators.

of replacing traditional methods such as slope stability and landslide[24–26], blast-induced problems (e.g., ground vibration, air over-pressure, fly-rock, rock fragmentation, etc.)[27–32], optimization in mine planning[33,34], and geology and geophysics[35–37]. For slope stability prediction, Qi and Tang[38] developed six different soft computing models based on a meta-heuristic algorithm (i.e., firefly optimization) and machine learning algorithms (i.e., random forest, logistic regression, gradient boosting machine, support vector machine, decision tree, and multilayer perceptron neural network). A promising result was found in their study for predicting slope stability when the area under the receiver operating characteristic curve reached up to 0.967. Sakellariou and Ferentinou[39] introduced an artificial neural network (ANN) model that used geometrical and geotechnical parameters to predict the factor of safety (FOS) based on their database of experiments. In another study, Samui[40] applied support-vector machine (SVM) for

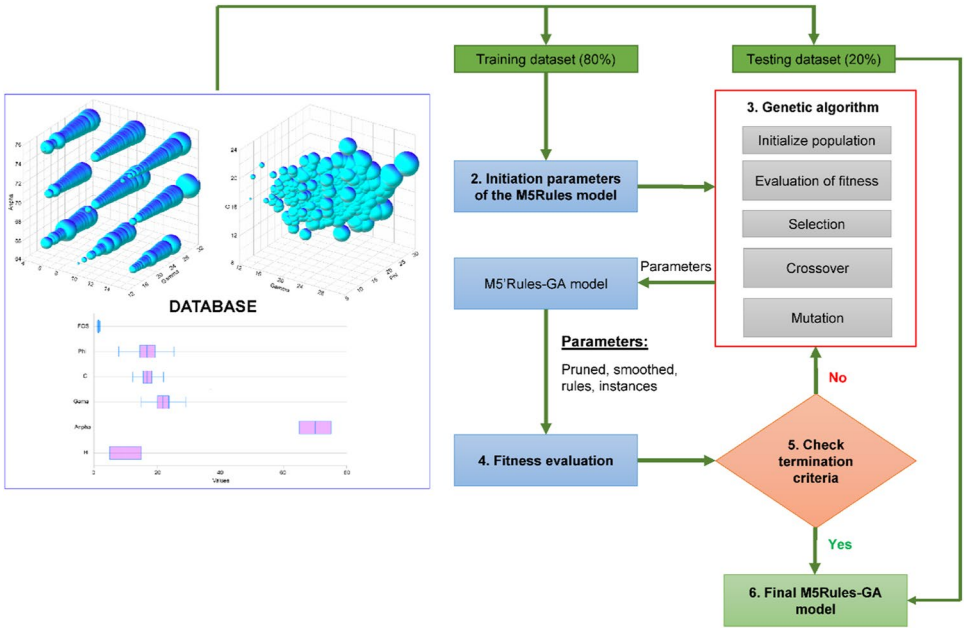


Figure 4. Flow chart of M5Rules–GA model for analyzing slope stability.

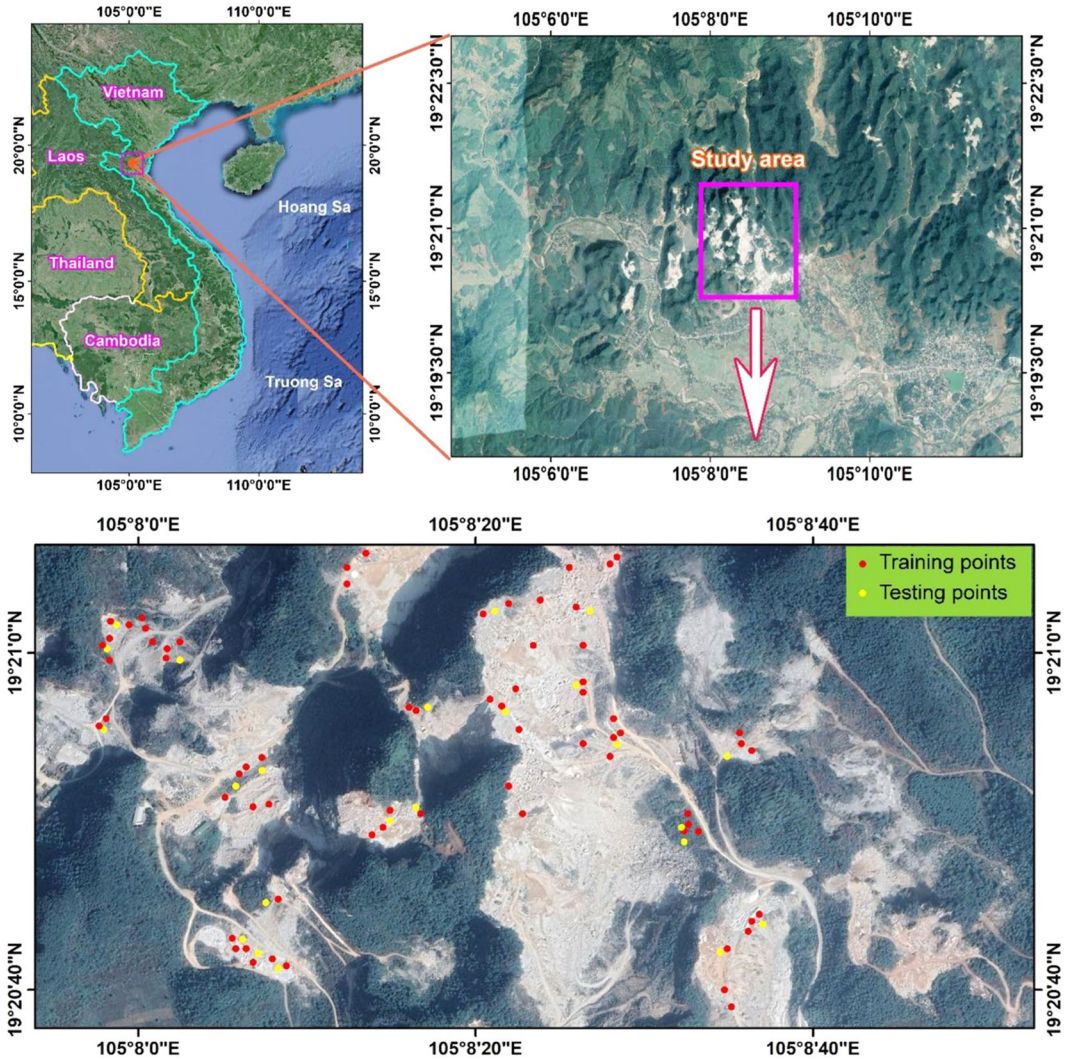
slope stability analysis using a database of practical investigations. Choobbasti, *et al.*[41] also conducted similar work with ANN models. Despite the high appreciation from researchers for the effectiveness of AI techniques in predicting slope stability, experimental data are often insufficient because of time and cost constraints.

To overcome the abovementioned limitations, various simulation software based on RFEM and LEM methods, among others, were introduced (such as OptumG2 and Geo-Studio)[42,43]. Based on these tools, slope stability prediction was analyzed and accurately evaluated for many models. Moreover, scientists have applied a big-data approach to AI for slope stability analysis[44–46]. Chakraborty and Goswami[47] simulated 200 cases with different shear strengths and geometric parameters to evaluate slope stability using ANN and multiple linear regression (MLR). Jellali and Frikha[48] used OptumG2 to generate 30,000 elements and predicted slope stability using the particle swarm optimization (PSO) algorithm with promising results. Mojtahedi, *et al.*[49] applied the Monte Carlo technique using Geo-Studio software with 224 datasets. Saleh[50] applied ANN to a database of 2,180 simulated slope cases using Geo-Studio software. Qi and Tang[38] attempted to develop six soft computing models using optimization approaches based on firefly algorithms (FFA), concluding that an FFA–SVM model was the best. Koopialipoor, *et al.*[51] applied various hybrid AI models, such as ANN-PSO, ANN-ICA (imperialist competitive algorithm), ANN-GA (genetic algorithm), and ANN-ABC (artificial bee colony), based on OptumG2 software analysis results. They found that the ANN-PSO model provided better performance than the other models. Gao, *et al.*[52] successfully developed a promising hybrid model called ICA–ANN based on the combination of ANN and ICA with 400 OptumG2 simulations. Qian, *et al.*[53] also performed similar work for forecasting slope stability based on OptumG2 software.

The review of previous works reveals that AI techniques are widely applied in slope stability prediction and analysis. However, such methods are not applied in all areas/regions. Furthermore, many AI models and techniques are yet to be investigated. To promote continued improvements in safety, development of knowledge, and enhancing slope stability, prediction performance in other areas is necessary. Therefore, a novel hybrid model, namely M5Rules–GA, for predicting slope stability (i.e., FOS) using a genetic algorithm (GA) and M5Rules was proposed and investigated in this study. It is worth mentioning that the team of authors developed the M5Rules-GA model for predicting the energy efficiency of buildings (i.e., heating load) with high accuracy[54]. However, it is not taking into account to predict and evaluate the stability of slopes. Furthermore, the performance, as well as the parameters of the M5Rules-GA model, are different based on different databases. Therefore, the M5Rules-GA model was investigated to predict slope stability herein and it is considered as a novel model in this field. Besides, several ANNs, support-vector regression (SVR), and previously introduced slope stability prediction models (such as FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC) were also implemented for a comprehensive comparison of the proposed M5Rules–GA model. There were 450 simulations of open-pit mine slopes in Vietnam, as a database for predicting slope stability.

Background: M5Rules and GA

In this study, eight AI techniques were used to develop slope failure predictive models, including ANN, SVR, M5Rules, PSO, FFA, ICA, ABC, and GA. However, the details for ANN, PSO, FFA, ICA, ABC, and SVR techniques were presented in many previous works[19,55–60]. Therefore, these details are not included in the present study. This section highlights the background of M5Rules and GA for developing the new hybrid M5Rules–GA model.



Legend: This figure was created in the ArcGIS version 10.5 and Google earth free.
(link: https://earth.google.com/web/@4.44767136,-67.63045051,2279.14659005a,21941688.60396624d,35y,-0h,0t,0r)

Figure 5. Study area and its landscape via Google Earth.

M5Rules. M5Rules is well-known as an enhanced model of the M5 model with rules[61]. It is a type of decision tree algorithm in machine learning that can be applied for regression and classification problems[62,63]. In regression, M5Rules is based on a combination of regression tree models[64]. In addition, the partial and regression tree (PART) algorithm is applied to generate the rules for the M5 tree model[65]. These rules can improve the performance of the M5 tree model rather than the regression tree models, and it is called M5Rules. In M5Rules, the tree can be developed based on the four following options: pruned/unpruned tree; smoothed/unsmoothed predictions; build regression tree/rule; and define the minimum number of instances per leaf[66]. The workflow for the development of the M5Rules model is shown in Fig. 2.

Genetic algorithm. Meta-heuristic algorithms are well-known as robust algorithms for optimization problems. Among those, the GA has one of the dominant performances in optimization based on the theory of Darwin[67,68]. Four steps are conducted in GA for finding an optimal function: genetics, mutation, natural selection, and crossover. Before implementing an optimization of an objective function, GA needs an initial population and their fitness are calculated[69,70]. Note that the number of population of individuals is generated heuristically or randomly[71]. In GA, the quality of the population can be improved by the selection operator. Subsequently, two individuals are generated with higher fitness using the crossover operator. Mutation operators can create a new generation with better performance in the population by randomly modifying some genes[72]. It is worth mentioning that replacement strategies can be applied to replace the current generation by newly generated offsprings. There are two main types of replacement: generational and steady-state replacements[73]. In addition, other related replacement strategies, such as elitism, delete n-last, delete n, random replacement, weak parent replacement, and both parents replacement, can be applied for replacement of generation in GA[74]. The structure and the framework of the GA are simulated in Fig. 3.

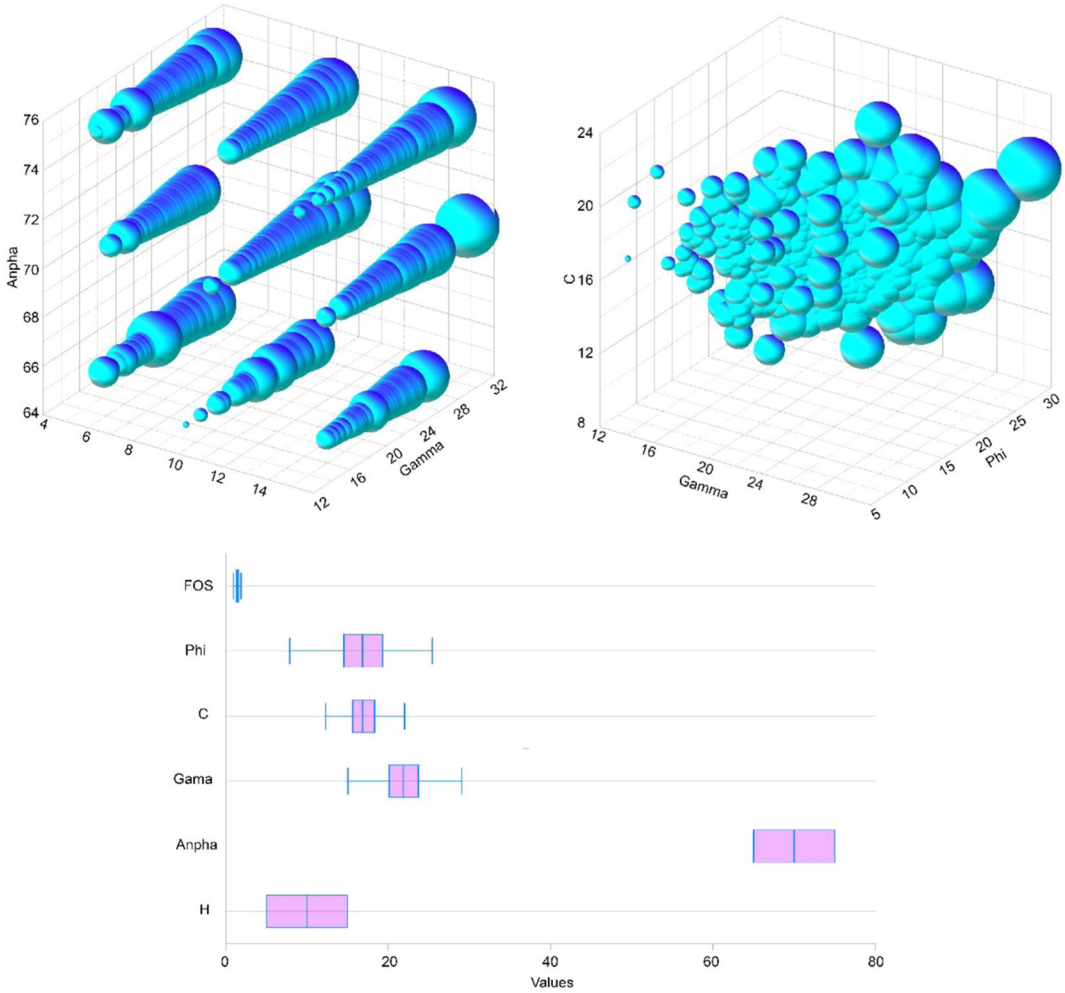


Figure 6. Visualization of slope stability database.

Features	H	α	γ	C	φ	FOS
Min.	5	65	12.83	11.22	7.3	0.76
1st Qu.	5	65	20.09	15.6	14.5	1.29
Median	10	70	21.85	16.85	16.8	1.41
Mean	10	70	21.87	16.96	16.8	1.42
3rd Qu.	15	75	23.71	18.3	19.3	1.54
Max.	15	75	31.53	23.34	27.8	1.98

Table 1. Summary of the features of inputs and output.

In GA, the cycle of operators (i.e., selection, crossover, and mutation) is employed and looped. To end the algorithm, two stopping conditions can be applied as follows:

1. Tcn the structure of the chromosome.
2. The change in fitness from newly generated offsprings is less than a specified constant.

Framework of M5Rules–GA model for slope stability analysis. In this section, the M5Rules–GA model, which is the slope stability prediction model proposed in this study, is presented and highlighted. The performance of the M5Rules model is determined by its parameters. Pruning and smoothing tasks can be applied during the development of the M5 tree model. Additionally, rules and the number of instances per leaf are important parameters affecting the performance of the M5Rules model. Therefore, the GA is applied to optimize the parameters of the M5Rules model: *pruned, smoothed*, *rules*, and *the number of instances per leaf*. Note that the PART algorithm generates rules for the M5 tree model. With each round, GA searches the M5Rules model

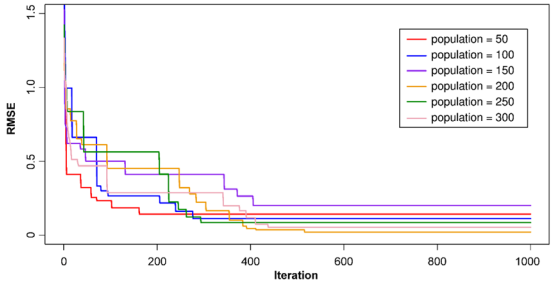


Figure 7. Optimizing the M5Rules model with the GA procedure.

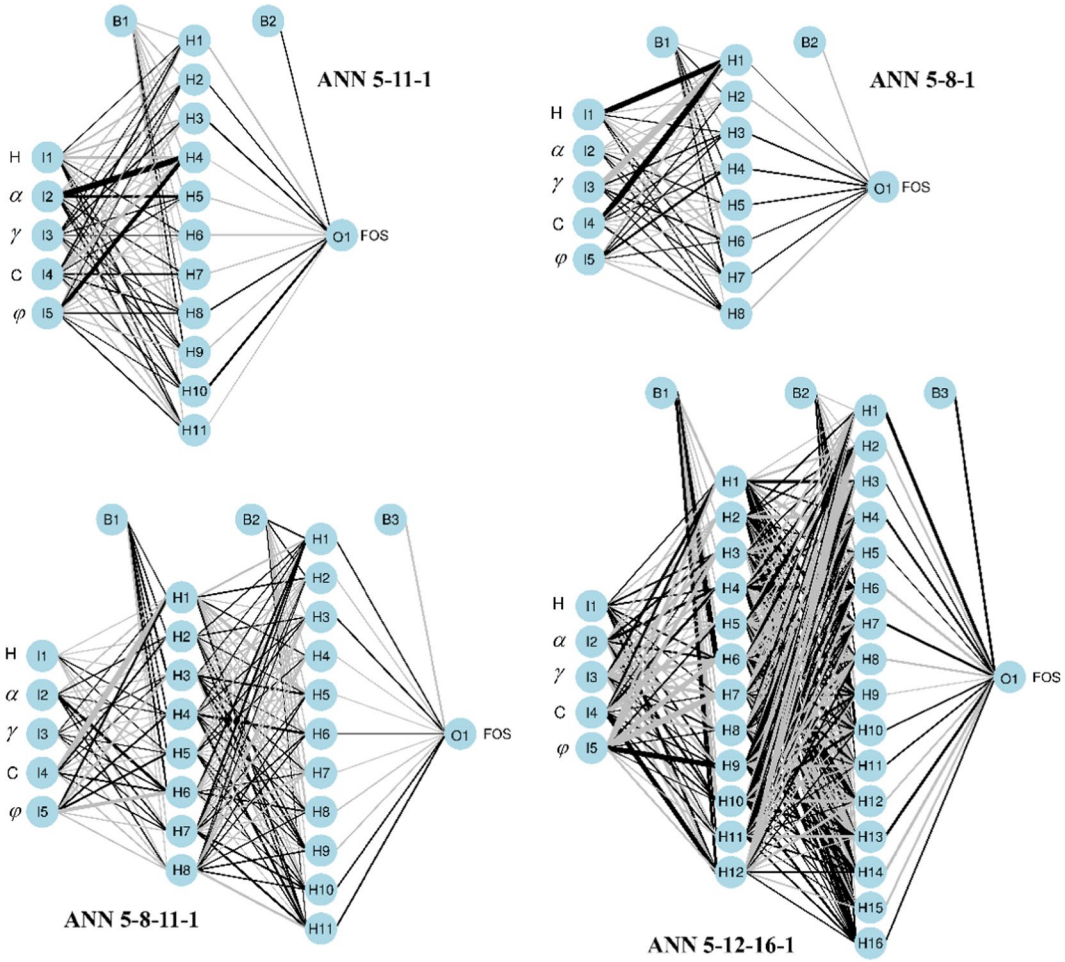


Figure 8. ANN models for analyzing slope stability in this study.

parameter values with the M5Rules model performance evaluated using a fitness function (i.e., root mean square error (RMSE)). The search process is performed until the optimal M5Rules model is found, i.e., the model with the most extreme fitness function value. The framework of the M5Rules–GA model is illustrated in Fig. 4.

Statistical criteria. To evaluate the accuracy and error of the developed models, RMSE, determination coefficient (R^2), variance account for (VAF), and color intensity were applied based on the measured and predicted values on both training and testing phases. They were calculated according to Eqs. (1–3).

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{1}$$

Model	Training		Testing	
	RMSE	R^2	RMSE	R^2
M5Rules-GA	0.022	0.985	0.024	0.983
ANN 5-8-1	0.033	0.971	0.031	0.970
ANN 5-11-1	0.033	0.973	0.030	0.975
ANN 5-8-11-1	0.032	0.971	0.027	0.978
ANN 5-12-16-1	0.033	0.970	0.032	0.969
SVR	0.032	0.975	0.030	0.974
FFA-SVR	0.027	0.978	0.029	0.976
ANN-PSO	0.026	0.980	0.026	0.980
ANN-ICA	0.028	0.977	0.029	0.976
ANN-GA	0.029	0.980	0.027	0.979
ANN-ABC	0.030	0.980	0.028	0.978

Table 2. Performance of the slope stability evaluation models.

Model	Performance of the models				Ranking			
	Runtime	RMSE	R^2	VAF	Rank for RMSE	Rank for R^2	Rank for VAF	Total ranking
M5Rules-GA	157.201	0.024	0.983	98.260	11	11	11	33
ANN 5-8-1	31.225	0.031	0.970	97.037	2	2	2	6
ANN 5-11-1	32.033	0.030	0.975	97.280	3	4	3	10
ANN 5-8-11-1	62.779	0.027	0.978	97.707	8	7	8	23
ANN 5-12-16-1	94.542	0.032	0.969	96.840	1	1	1	3
SVR	125.332	0.030	0.974	97.320	3	3	4	10
FFA-SVR	188.382	0.029	0.976	97.517	5	5	5	15
ANN-PSO	217.677	0.026	0.980	97.907	10	10	10	30
ANN-ICA	250.701	0.029	0.976	97.552	5	5	6	16
ANN-GA	282.772	0.027	0.979	97.751	8	9	9	26
ANN-ABC	314.388	0.028	0.978	97.647	7	7	7	21

Table 3. Testing the performance of slope stability predictive models.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)}{\sum_i (y_i - \overline{y})} \tag{2}$$

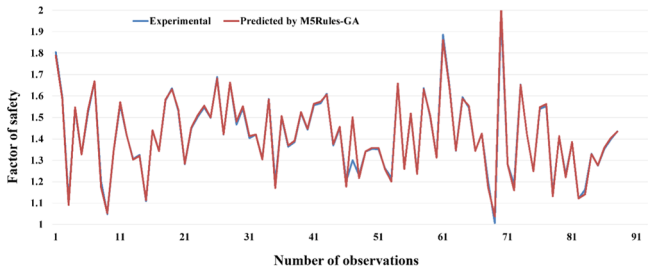
$$\mathrm{VAF} = \left(1 - \frac{\mathrm{var}(y_i - \hat{y}_i)}{\mathrm{var}(y_i)}\right) \times 100 \tag{3}$$

where n represents the number of instances, and $\overline{y}$, y_i, and $\hat{y}_i$ represent the average, measured, and modeled values of the response variable, respectively.

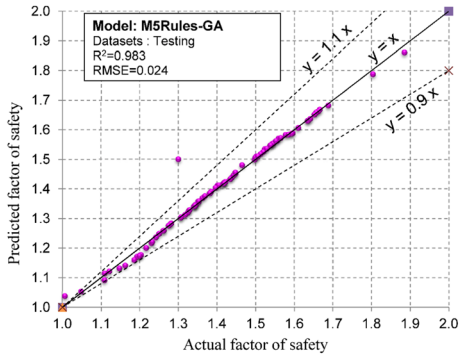
Case study

For assessing the performance of the proposed M5Rules–GA model in practical engineering, a quarry mine in Vietnam was selected as a case study (Fig. 5). The parameters used to predict the stability of the slope included bench height (H), unit weight (γ), cohesion (C), angle of internal friction (φ), and slope angles (α); the FOS was assigned as the output parameter. Properties of the dataset used are detailed in Table 1.

As recommended by previous researchers, γ, α, H,, and C are the most influential parameters that have impacts on the FOS[58,75]. Therefore, these factors were provided to the Geo-Studio (version 2019) for the computations of FOS values. According to Zhou, *et al.*[76], the slope is stable at FOS > 1. However, according to Sakellariou and Ferentinou[39], the slope is stable only at FOS ≥ 1.2. Thus, for safety in mining, FOS was assigned at least 1.2. Therefore, the slopes will be stable if $FOS \geq 1.2$, and fail if $FOS < 1.2$. In this study, 450 simulations were conducted in the laboratory using Geo-Studio software based on the working conditions. The simulation results showed that the slopes of the study site included both stability and failure, i.e., $0.76 \leq FOS \leq 1.98$. The datasets used in this study are visualized in Fig. 6, and summarized in Table 1.

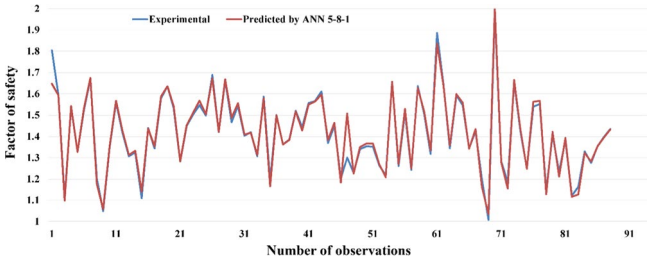


(a)

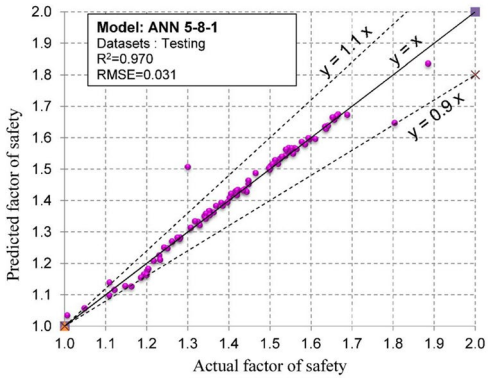


(b)

Figure 9. The accuracy and the converging of the M5Rules-GA model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

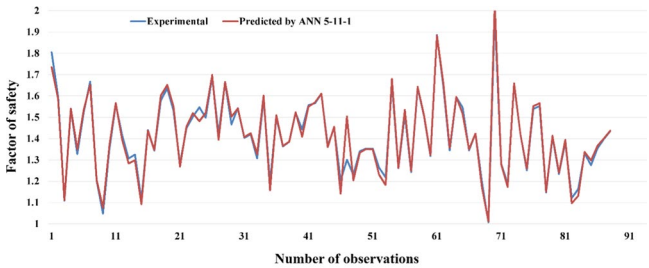


(a)

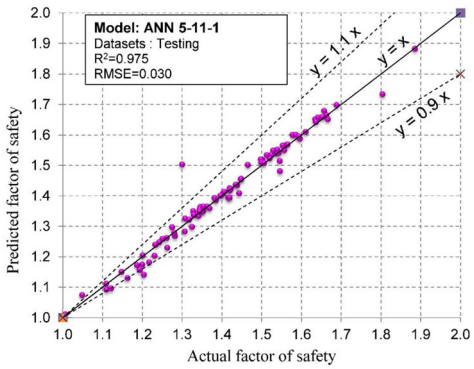


(b)

Figure 10. The accuracy and the converging of the ANN 5-8-1 model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.



(a)



(b)

Figure 11. The accuracy and the converging of the ANN 5-11-1 model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

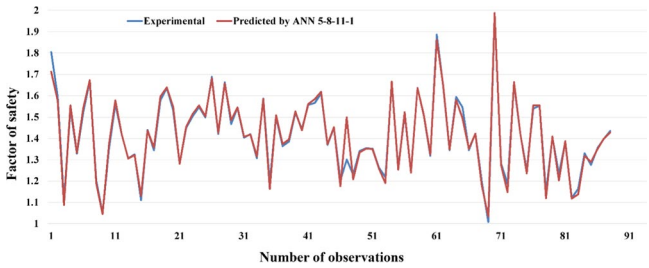
Results and discussions

To develop the models, the FOS database needs to be prepared and normalized. Accordingly, the dataset used should be split into two phases. As recommended in previous studies[56,77,78], 80% of the database was used for the training of the models; the remaining 20% was used to assess the models' performance. The training dataset was randomly selected, and all the abovementioned models were developed based on the same training phase and tested using the same testing phase.

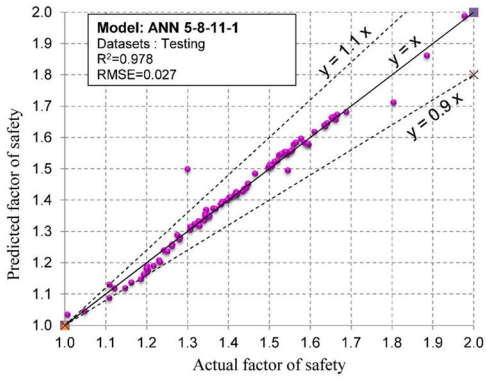
For the development of the M5Rules–GA model, the step-by-step approach shown in Fig. 4 was applied. An initialization of the M5Rules model was developed based on its parameters and the training dataset, as the first step. To improve the model's performance, the 10-fold cross-validation resampling technique was used. Next, the GA's parameters were established as the second step: mutation probability (*Pm*), crossover probability (*Pc*), number variable (*n*), and number of populations (*p*). *Pm*, *Pc*, and *n* were set to 0.1, 0.8, and 5, respectively, and the values for *p* were set to 50, 100, 150, 200, 250, and 300. Additionally, the steady-state replacement method[79] was applied in GA. RMSE was used as the fitness function, according to Eq. 1. The maximum number of iterations was set to 1000 to ensure the finding of the best values of the M5Rules model with the lowest RMSE value (i.e., best fitness value). Figure 7 shows that the M5Rules–GA model reached the best performance with $p = 200$ at the iteration of 412 (RMSE = 0.0218).

For SVR modeling, a kernel function was applied (i.e., radial basis function) with σ and C used as the main parameters for controlling SVR model performance. A 10-fold cross-validation[80] was applied to increase the accuracy while preventing overfitting or underfitting of the SVR model. Furthermore, the Box-Cox transformation technique[81] was applied to reduce the skewness of the data. A trial and error approach with various SVR models was conducted to determine the best SVR model for this study. The best SVR model for analyzing slope stability was found at $\sigma = 0.014$ and $C = 276.385$.

For ANN models, hidden model layers resist definition or explicit explanation. However, according to previous works[28,82–84], ANNs with one or two hidden layer(s) can solve most problems. Therefore, a trial and error approach was conducted to find the best ANN models with one or two hidden layer(s). The min-max scale technique (i.e., [0,1]) was used as a normalization method for the dataset to avoid overfitting of the ANN models. Eventually, four ANN models were established: ANN 5-8-1, ANN 5-11-1, ANN 5-8-11-1, and ANN 5-12-16-1, called ANN 1, ANN 2, ANN 3, and ANN 4, respectively. Their structures can be seen in Fig. 8. In addition, the FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC models introduced by previous researchers were taken into consideration to predict FOS and compare with the developed M5Rules-GA model. To predict the stability of the slope, the FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC models were also developed through two phases: (1) Initializing an ANN model with initial weights and biases; (2) Optimizing the initialization ANN model by the FFA algorithm. Indeed, the weights and biases of the established ANN model were optimized by the FFA algorithm aiming to improve the accuracy of the initialization ANN model. In other words, the role of the pairs ANN and M5Rules, FFA, ABC, PSO, ICA and GA are the same in this study. It should be noted that the Box-Cox transformation technique was applied to preprocess the dataset aiming to prevent overfitting

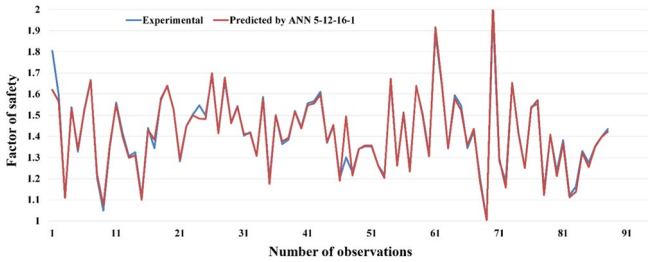


(a)

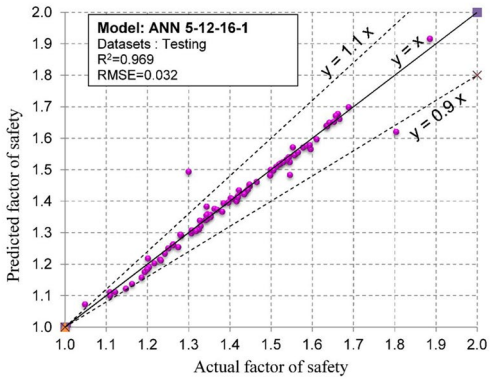


(b)

Figure 12. The accuracy and the converging of the ANN 5-8-11-1 model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

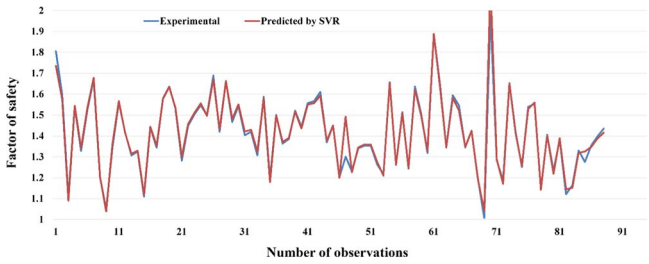


(a)

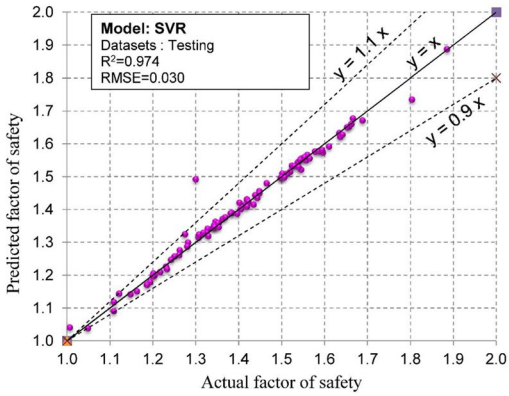


(b)

Figure 13. The accuracy and the converging of the 5-12-16-1 model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.



(a)



(b)

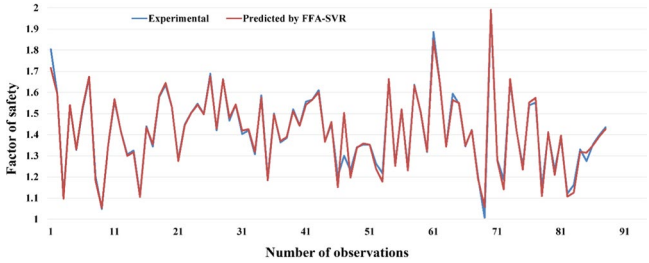
Figure 14. The accuracy and the converging of the SVR model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

of the M5Rules and SVR models. For the ANN models, the MinMax [0,1] technique was used to normalized the dataset for the same purposes as those of M5Rules and SVR models. Ultimately, the performance of the slope stability evaluation models on both training and testing datasets are computed in Table 2.

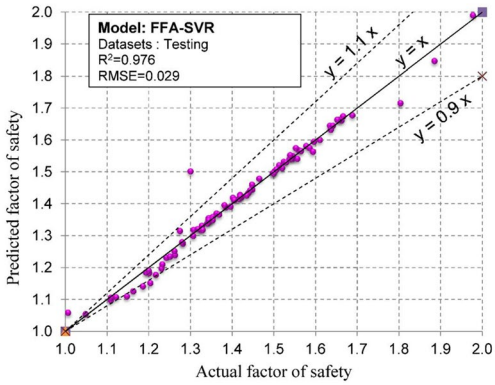
Based on Table 2, it is worth mentioning that all the models performed very well in predicting the slope stability without overfitting. However, it is hard to recognize which model is the best among them. Therefore, once the models were well-established based on the training dataset, their performance should be tested using the testing phase with performance indices (i.e., RMSE, R^2, VAF, a general ranking, and color range). The purpose of using multiple metrics, ranking, and color intensity methods is to recognize the best model in those of the developed models. Furthermore, the testing dataset is taken into account as the new dataset in practical; thus, evaluating the performance of the models on the testing dataset will provide an overview of the reliability of the models in practice. Also, to evaluate the performance of the models through the training time (runtime), the total time of training of the models was calculated in Table 3. It is worth mentioning that the runtime of the models highly depends on the hardware of the computer used. Herein, a workstation computer with the Intel(R) Xeon(R) dual CPU X5675 3.07 GHz, 24 GB RAM, and K5000 VGA (5.0 GB) was used to train the models. The testing results of the predictive models, as well as their ranking, are listed in Table 3.

Based on Table 3, it is clear that the training time of the hybrid models is higher than the single models. This problem is due to the calculation volume of the hybrid models is higher than the individual models with many repetitions. Of those, the training time of the M5Rules-GA model is lowest among the hybrid modes developed (i.e., M5Rules-GA, FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, ANN-ABC) with 157.201 seconds. Whereas, the training time of the ANN-PSO model is higher than those of the M5Rules-GA models even though its accuracy is slightly lower than the M5Rules-GA model. Another hybrid model based on the GA, i.e., ANN-GA, also taken more training time than the M5Rules-GA model (i.e., 282.772 seconds). They showed that the optimization of M5Rules is faster than the ANN model.

Regarding the accuracy of the models, a color range can preliminarily provide a performance evaluation of the models. Green and white represent the best and worst performances, respectively. Table 3 showed that the proposed M5Rules–GA model provided the best performance among the developed models in this study. In contrast, the ANN 5-12-16-1 model yielded the worst performance. Considering the accuracy/performance of the proposed M5Rules-GA and previously introduced models (i.e., FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC), it can be seen that the accuracy of the M5Rules-GA model is also higher than the other hybrid models. Indeed, the total ranking of the M5Rules-GA model was 33, whereas the best hybrid model among the FFA-SVR, ANN-PSO, ANN-ICA, ANN-GA, and ANN-ABC models only received a total ranking of 30 (ANN-PSO). It is worth mentioning that the role of the GA is the optimization of parameters of the models (such as M5Rules and ANN). However, we can see that the performance of the M5Rules-GA model is better than the

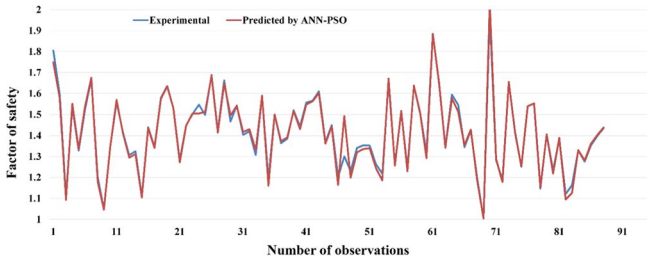


(a)

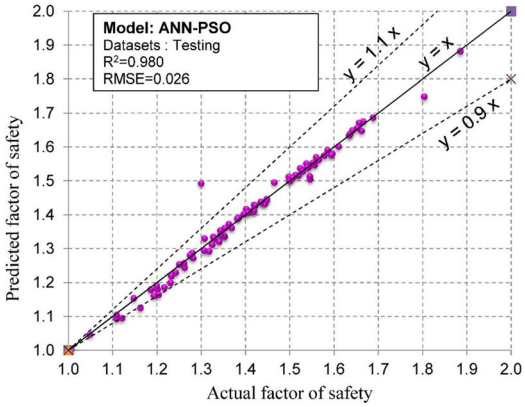


(b)

Figure 15. The accuracy and the converging of the FFA-SVR model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

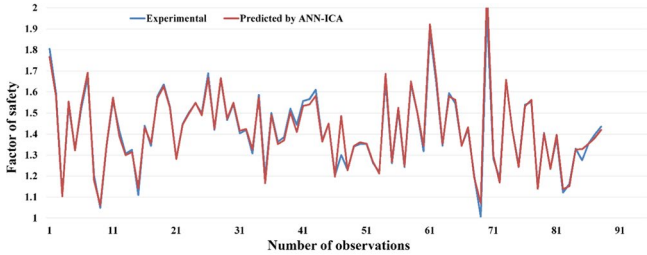


(a)

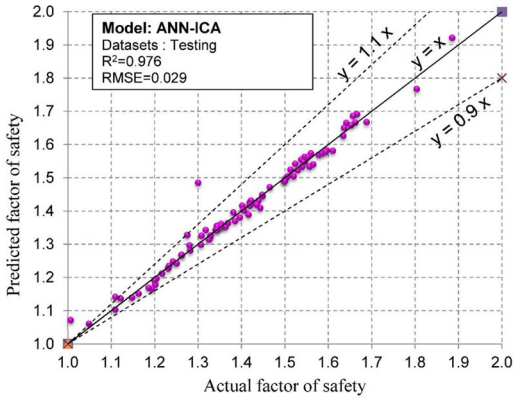


(b)

Figure 16. The accuracy and the converging of the ANN-PSO model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

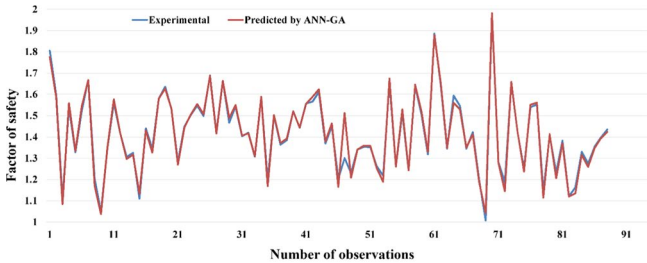


(a)

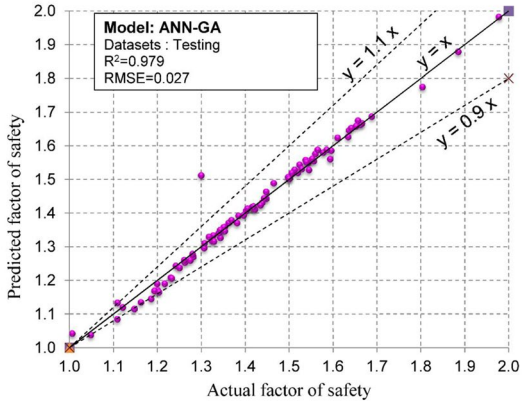


(b)

Figure 17. The accuracy and the converging of the ANN-ICA model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

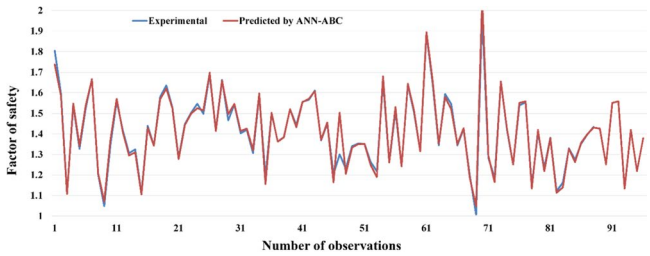


(a)

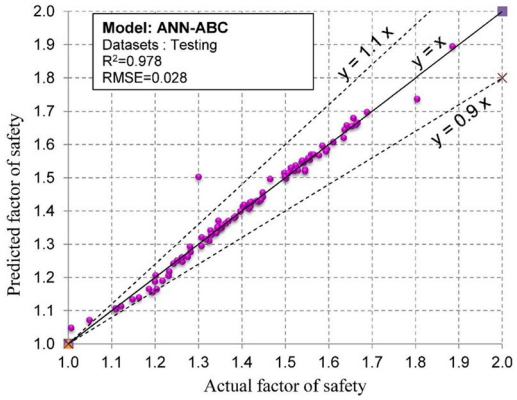


(b)

Figure 18. The accuracy and the converging of the ANN-GA model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.



(a)



(b)

Figure 19. The accuracy and the converging of the ANN-ABC model in predicting FOS. (**a**) Different between the actual and predicted FOS values. (**b**) Correlation analyses of the actual and predicted FOS values.

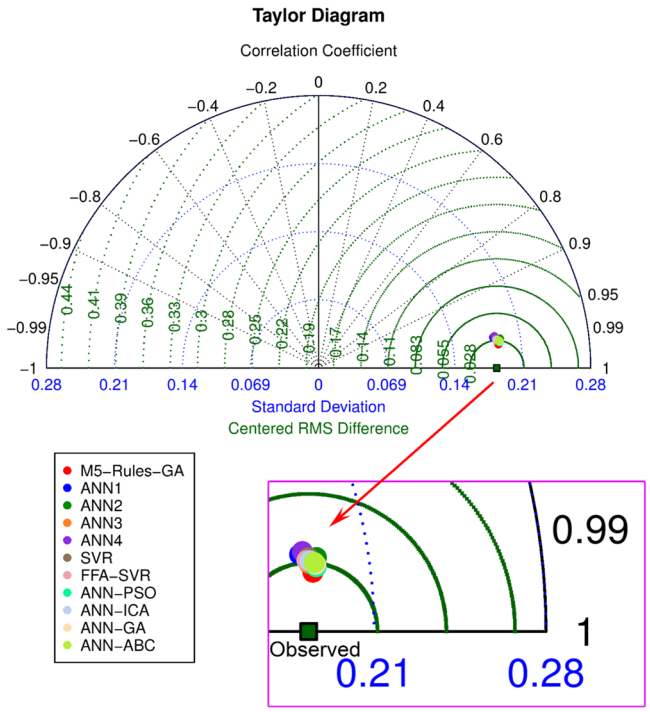


Figure 20. Assessment of the models using Taylor diagram.

ANN-GA model. This finding indicates that the M5Rules model is fitter than the ANN models. The FOS predictions, correlation schemes, and their 90% confidence level of the models are illustrated in Figs. 9–19.

Considering further evaluation criteria, such as standard deviation, centered root mean square (RMS) difference, and correlation coefficient, a Taylor diagram was drawn to visualize and comprehensively assess the developed models, as shown in Fig. 20. As seen in the figure, the proposed M5Rules-GA was on the smallest arcs of standard deviation, centered RMS difference, and correlation coefficient. A closer look at the models shows that the standard deviations of the M5Rules-GA and other models do not differ considerably as they seem to be on the same arc. However, centered RMS difference and correlation coefficient of the proposed M5Rules-GA model were superior.

Conclusions

Based on the results of this study, the proposed M5Rules–GA model provided the best accuracy among all the investigated models for slope stability forecasting. The M5Rules model was substantially enhanced using GA optimization, thereby achieving outstanding performance. We expect M5Rules–GA model to be useful in evaluating and predicting slope stability at mines, thus preventing and minimizing slope collapse damage. In addition, the other models also showed positive results, and they might be considered in other instances. Although the performance of the proposed M5Rules–GA model was interpreted in the context of the present study's dataset, we suggest its broader application to other regions with extended datasets.

Received: 9 December 2019; Accepted: 20 May 2020;
Published online: 18 June 2020

References

1. Vyazmensky, A., Stead, D., Elmo, D. & Moss, A. Numerical analysis of block caving-induced instability in large open pit slopes: a finite element/discrete element approach. *Rock mechanics and rock engineering* **43**, 21–39 (2010).
2. Dyson, A. P. & Tolooiyan, A. Prediction and classification for finite element slope stability analysis by random field comparison. *Computers and Geotechnics* **109**, 117–129 (2019).
3. Wang, L., Sun, D. A., Chen, B. & Li, J. Three-dimensional seismic stability of unsaturated soil slopes using a semi-analytical method. *Computers and Geotechnics* **110**, 296–307 (2019).
4. Wei, Y., Jiaxin, L., Zonghong, L., Wei, W. & Xiaoyun, S. A strength reduction method based on the Generalized Hoek-Brown (GHB) criterion for rock slope stability analysis. *Computers and Geotechnics* **117**, 103240 (2020).
5. He, Y., Liu, Y., Hazarika, H. & Yuan, R. Stability analysis of seismic slopes with tensile strength cut-off. *Computers and Geotechnics* **112**, 245–256 (2019).
6. Wang, Z. *et al*. Stability of inner dump slope and analytical solution based on circular failure: Illustrated with a case study. *Computers and Geotechnics* **117**, 103241 (2020).
7. Li, C., Jiang, P. & Zhou, A. Rigorous solution of slope stability under seismic action. *Computers and Geotechnics* **109**, 99–107 (2019).
8. Yang, M. & Deng, B. Stability study of slope reinforced with piles under steady unsaturated flow conditions. *Computers and Geotechnics* **109**, 89–98 (2019).
9. Zhang, W., Zheng, H., Jiang, F., Wang, Z. & Gao, Y. Stability analysis of soil slope based on a water-soil-coupled and parallelized Smoothed Particle Hydrodynamics model. *Computers and Geotechnics* **108**, 212–225 (2019).
10. Naghadehi, M. Z., Jimenez, R., KhaloKakaie, R. & Jalali, S.-M. E. A new open-pit mine slope instability index defined using the improved rock engineering systems approach. *International Journal of Rock Mechanics and Mining Sciences* **61**, 1–14 (2013).
11. El-Ramly, H., Morgenstern, N. & Cruden, D. Probabilistic slope stability analysis for practice. *Canadian Geotechnical Journal* **39**, 665–683 (2002).
12. Dyson, A. P. & Tolooiyan, A. Probabilistic investigation of RFEM topologies for slope stability analysis. *Computers and Geotechnics* **114**, 103129 (2019).
13. Griffiths, D. & Fenton, G. A. In *Slope stability 2000* 184-193 (2000).
14. Griffiths, D. & Fenton, G. A. Probabilistic slope stability analysis by finite elements. *Journal of Geotechnical and Geoenvironmental Engineering* **130**, 507–518 (2004).
15. Cheng, Y., Lansivaara, T. & Wei, W. Two-dimensional slope stability analysis by limit equilibrium and strength reduction methods. *Computers and Geotechnics* **34**, 137–150 (2007).
16. Liu, S., Shao, L. & Li, H. Slope stability analysis using the limit equilibrium method and two finite element methods. *Computers and Geotechnics* **63**, 291–298 (2015).
17. Zhou, J., Li, X. & Mitri, H. S. Classification of rockburst in underground projects: comparison of ten supervised learning methods. *Journal of Computing in Civil Engineering* **30**, 04016003 (2016).
18. Zhou, J., Li, X. & Shi, X. Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. *Safety science* **50**, 629–644 (2012).
19. Nguyen, H., Choi, Y., Bui, X.-N. & Nguyen-Thoi, T. Predicting Blast-Induced Ground Vibration in Open-Pit Mines Using Vibration Sensors and Support Vector Regression-Based Optimization Algorithms. *Sensors* **20**, 132 (2020).
20. Zhang, H. *et al*. Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm. *Resources Policy* **66**, 101604 (2020).
21. Kim, D. H., Gratchev, I. & Balasubramaniam, A. Determination of joint roughness coefficient (JRC) for slope stability analysis: a case study from the Gold Coast area, Australia. *Landslides* **10**, 657–664, https://doi.org/10.1007/s10346-013-0410-8 (2013).
22. Niu, F., Luo, J., Lin, Z., Fang, J. & Liu, M. Thaw-induced slope failures and stability analyses in permafrost regions of the Qinghai-Tibet Plateau, China. *Landslides* **13**, 55–65, https://doi.org/10.1007/s10346-014-0545-2 (2016).
23. Chen, C.-W. *et al*. Evaluating the susceptibility of landslide landforms in Japan using slope stability analysis: a case study of the 2016 Kumamoto earthquake. *Landslides* **14**, 1793–1801, https://doi.org/10.1007/s10346-017-0872-1 (2017).
24. Müller, A. & Vargas, E. A. Correction to: Stability analysis of a slope under impact of a rock block using the generalized interpolation material point method (GIMP). *Landslides* **16**, 1063–1063, https://doi.org/10.1007/s10346-019-01153-3 (2019).
25. Pradhan, S. P., Panda, S. D., Roul, A. R. & Thakur, M. Insights into the recent Kotropi landslide of August 2017, India: a geological investigation and slope stability analysis. *Landslides*, https://doi.org/10.1007/s10346-019-01186-8 (2019).
26. Zhou, J. *et al*. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Safety Science* **118**, 505–518 (2019).
27. Nguyen, H., Bui, X.-N., Bui, H.-B. & Cuong, D. T. Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study. *Acta Geophysica* **67**, 477–490, https://doi.org/10.1007/s11600-019-00268-4 (2019).
28. Nguyen, H., Bui, X.-N., Bui, H.-B. & Mai, N.-L. A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine, Vietnam. *Neural Computing and Applications*, 1–17, https://doi.org/10.1007/s00521-018-3717-5 (2018).

29. Nguyen, H. Support vector regression approach with different kernel functions for predicting blast-induced ground vibration: a case study in an open-pit coal mine of Vietnam. *SN Applied Sciences* **1**, 283, https://doi.org/10.1007/s42452-019-0295-9 (2019).
30. Nguyen, H., Bui, X.-N. & Moayedi, H. A Comparison of Advanced Computational Models and Experimental Techniques in Predicting Blast-Induced Ground Vibration in Open-pit Coal Mine. *Acta Geophysica Journal*, https://doi.org/10.1007/s11600-019-00304-3 (2019).
31. Nguyen, H., Bui, X.-N., Tran, Q.-H. & Mai, N.-L. A New Soft Computing Model for Estimating and Controlling Blast-Produced Ground Vibration Based on Hierarchical K-Means Clustering and Cubist Algorithms. *Applied Soft Computing*, 1–20 (2019).
32. Zhang, S., Bui, X.-N., Trung, N.-T., Nguyen, H. & Bui, H.-B. Prediction of Rock Size Distribution in Mine Bench Blasting Using a Novel Ant Colony Optimization-Based Boosted Regression Tree Technique. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09603-4 (2019).
33. Djebali, S. *et al*. Milling plan optimization with an emergent problem solving approach. *Computers & Industrial Engineering* **87**, 506–517 (2015).
34. Bui, X.-N., Jaroonpattanapong, P., Nguyen, H., Tran, Q.-H. & Long, N. Q. A novel Hybrid Model for predicting Blast-induced Ground Vibration Based on k-nearest neighbors and particle Swarm optimization. *Scientific Reports* **9**, 1–14 (2019).
35. Essa, K. S. & Elhussein, M. Interpretation of Magnetic Data Through Particle Swarm Optimization: Mineral Exploration Cases Studies. *Natural Resources Research*, https://doi.org/10.1007/s11053-020-09617-3 (2020).
36. Tut Haklidir, F. S. & Haklidir, M. Prediction of Reservoir Temperatures Using Hydrogeochemical Data, Western Anatolia Geothermal Systems (Turkey): A Machine Learning Approach. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09596-0 (2019).
37. Chen, Y., Wu, W. & Zhao, Q. A Bat Algorithm-Based Data-Driven Model for Mineral Prospectivity Mapping. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09589-z (2019).
38. Qi, C. & Tang, X. Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. *Computers & Industrial Engineering* **118**, 112–122 (2018).
39. Sakellariou, M. & Ferentinou, M. A study of slope stability prediction using neural networks. *Geotechnical & Geological Engineering* **23**, 419 (2005).
40. Samui, P. Slope stability analysis: a support vector machine approach. *Environmental Geology* **56**, 255 (2008).
41. Choobbasti, A., Farrokhzad, F. & Barari, A. Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran). *Arabian journal of geosciences* **2**, 311–319 (2009).
42. Karkanaki, A. R., Ganjian, N. & Askari, F. Stability analysis and design of cantilever retaining walls with regard to possible failure mechanisms: an upper bound limit analysis approach. *Geotechnical and Geological Engineering* **35**, 1079–1092 (2017).
43. Arshad, I., Babar, M. M. & Javed, N. Numerical Analysis of Seepage and Slope Stability in an Earthen Dam by Using Geo-Slope Software. *PSM Biological Research* **2**, 13–20 (2017).
44. Luo, Z., Bui, X.-N., Nguyen, H. & Moayedi, H. A novel artificial intelligence technique for analyzing slope stability using PSO-CA model. *Engineering with Computers*, https://doi.org/10.1007/s00366-019-00839-5 (2019).
45. Han, Z. *et al*. Comprehensive analysis of landslide stability and related countermeasures: a case study of the Lanmuxi landslide in China. *Scientific reports* **9**, 1–12 (2019).
46. Carlà, T. *et al*. Perspectives on the prediction of catastrophic slope failures from satellite InSAR. *Scientific reports* **9**, 1–9 (2019).
47. Chakraborty, A. & Goswami, D. Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). *Arabian Journal of Geosciences* **10**, 385 (2017).
48. Jellali, B. & Frikha, W. Constrained particle swarm optimization algorithm applied to slope stability. *International Journal of Geomechanics* **17**, 06017022 (2017).
49. Mojtahedi, S. F. F. *et al*. A novel probabilistic simulation approach for forecasting the safety factor of slopes: a case study. *Engineering with Computers*, 1–10 (2018).
50. Saleh, L. In *IOP Conference Series: Materials Science and Engineering*. 012029 (IOP Publishing).
51. Koopialipoor, M., Jahed Armaghani, D., Hedayat, A., Marto, A. & Gordan, B. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. *Soft Computing*, https://doi.org/10.1007/s00500-018-3253-3 (2018).
52. Gao, W., Raftari, M., Rashid, A. S. A., Mu'azu, M. A. & Jusoh, W. A. W. A predictive model based on an optimized ANN combined with ICA for predicting the stability of slopes. *Engineering with Computers*, https://doi.org/10.1007/s00366-019-00702-7 (2019).
53. Qian, Z., Li, A., Chen, W., Lyamin, A. & Jiang, J. An artificial neural network approach to inhomogeneous soil slope stability predictions based on limit analysis methods. *Soils and Foundations* (2019).
54. Bui, X.-N., Moayedi, H. & Rashid, A. S. A. Developing a predictive method based on optimized M5Rules–GA predicting heating load of an energy-efficient building system. *Engineering with Computers*, https://doi.org/10.1007/s00366-019-00739-8 (2019).
55. Nguyen, H., Drebenstedt, C., Bui, X.-N. & Bui, D. T. Prediction of Blast-Induced Ground Vibration in an Open-Pit Mine by a Novel Hybrid Model Based on Clustering and Artificial Neural Network. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09470-z (2019).
56. Bui, X.-N., Nguyen, H., Le, H. A., Bui, H. B. & Do, N. H. Prediction of Blast-induced Air Over-pressure in Open-Pit Mine: Assessment of Different Artificial Intelligence Techniques. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09461-0 (2019).
57. Gordan, B., Armaghani, D. J., Hajihassani, M. & Monjezi, M. Prediction of seismic slope stability through combination of particle swarm optimization and neural network. *Engineering with Computers* **32**, 85–97 (2016).
58. Koopialipoor, M., Armaghani, D. J., Hedayat, A., Marto, A. & Gordan, B. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. *Soft Computing*, 1–17 (2018).
59. Shang, Y., Nguyen, H., Bui, X.-N., Tran, Q.-H. & Moayedi, H. A Novel Artificial Intelligence Approach to Predict Blast-Induced Ground Vibration in Open-pit Mines Based on the Firefly Algorithm and Artificial Neural Network. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09503-7 (2019).
60. Zhang, X. *et al*. Novel Soft Computing Model for Predicting Blast-Induced Ground Vibration in Open-Pit Mines Based on Particle Swarm Optimization and XGBoost. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09492-7 (2019).
61. Quinlan, J. R. In *5th Australian joint conference on artificial intelligence*. 343–348 (World Scientific).
62. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. Classification and regression trees. Belmont, CA: Wadsworth. *International Group*, 432 (1984).
63. Quinlan, J. R. Simplifying decision trees. *International journal of man-machine studies* **27**, 221–234 (1987).
64. Sattari, M. T. & Sureh, F. S. In *International Civil Engineering and Architecture Conference (ICEARC)*.
65. Frank, E. & Witten, I. H. Generating accurate rule sets without global optimization. (1998).
66. Rouzegari, N., Hassanzadeh, Y. & Sattari, M. T. Using the Hybrid Simulated Annealing-M5 Tree Algorithms to Extract the If-Then Operation Rules in a Single Reservoir. *Water Resources Management* **33**, 3655–3672 (2019).
67. Mitchell, M. *An introduction to genetic algorithms*. (MIT press, 1998).
68. Carr, J. An introduction to genetic algorithms. *Senior Project* **1**, 40 (2014).
69. Raeisi-Vanani, H. *et al*. A Simple Method for Land Grading Computations and its Comparison with Genetic Algorithm (GA) Method. *International Journal of Research Studies in Agricultural Sciences (IJRSAS)* **3**, 26–38 (2017).
70. Goldberg, D. (Addison-Wesley, Reading, 1989).
71. Mirjalili, S. In *Evolutionary algorithms and neural networks* 43–55 (Springer, 2019).

72. Fang, Q., Nguyen, H., Bui, X.-N. & Tran, Q.-H. Estimation of Blast-Induced Air Overpressure in Quarry Mines Using Cubist-Based Genetic Algorithm. *Natural Resources Research*, https://doi.org/10.1007/s11053-019-09575-5 (2019).
73. Rutczyńska-Wdowiak, K. In *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*. 971–975 (IEEE).
74. Sivanandam, S. & Deepa, S. In *Introduction to genetic algorithms* 15–37 (Springer, 2008).
75. Gao, W., Raftari, M., Rashid, A. S. A., Mu'azu, M. A. & Jusoh, W. A. W. A predictive model based on an optimized ANN combined with ICA for predicting the stability of slopes. *Engineering with Computers*, 1–20, https://doi.org/10.1007/s00366-019-00702-7 (2019).
76. Zhou, Y., Cheuk, C. & Tham, L. Numerical modelling of soil nails in loose fill slope under surcharge loading. *Computers and Geotechnics* **36**, 837–850 (2009).
77. Nguyen, H., Bui, X.-N., Tran, Q.-H. & Mai, N.-L. A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical K-means clustering and cubist algorithms. *Applied Soft Computing* **77**, 376–386, https://doi.org/10.1016/j.asoc.2019.01.042 (2019).
78. Moayedi, H. & Rezaei, A. An artificial neural network approach for under-reamed piles subjected to uplift forces in dry sand. *Neural Computing and Applications* https://doi.org/10.1007/s00521-017-2990-z (2017).
79. Lozano, M., Herrera, F. & Cano, J. R. Replacement strategies to preserve useful diversity in steady-state genetic algorithms. *Information Sciences* **178**, 4421–4433 (2008).
80. Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing* **21**, 137–146 (2011).
81. Sakia, R. The Box-Cox transformation technique: a review. *Journal of the Royal Statistical Society: Series D (The Statistician)* **41**, 169–178 (1992).
82. Abbas, A. S. & Asheghi, R. Optimized developed artificial neural network-based models to predict the blast-induced ground vibration. *Innovative Infrastructure Solutions* **3**, 1–10 (2018).
83. Afram, A., Janabi-Sharifi, F., Fung, A. S. & Raahemifar, K. Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy and Buildings* **141**, 96–113 (2017).
84. Muhammad, F. & Ferentinou, M. A holistic open-pit mine slope stability index using Artificial Neural Networks. (2017).

Acknowledgements
This work was supported by the Center for Mining, Electro-Mechanical research, Hanoi University of Mining and Geology (HUMG), Hanoi, Vietnam and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2018R1D1A1A09083947).

Author contributions
Data collection and experimental works: Hoang Nguyen, Xuan-Nam Bui; Writing, discussion, analysis, revision: Hoang Nguyen, Xuan-Nam Bui, Trung-Thoi Nguyen, Yosoon Choi, Jie Dou and Jian Zhou.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to H.N. or Y.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.