Integration of grey analysis with artificial neural network for classification of slope failure

Ashanira Mat Deris[1], *Badariah* Solemon[2*], and *Rohayu* Che Omar[2]

[1]Fakulti Teknologi Kejuruteraaan Kelautan dan Informatik, Universiti Malaysia Terengganu 21030 Kuala Nerus, Terengganu
[2]Institute of Energy Infrastructure, Universiti Tenaga Nasional, Jln IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia

Abstract. With the advent of technology and the introduction of computational intelligent methods, the prediction of slope failure using the machine learning (ML) approach is rapidly growing for the past few decades. This study employs an "artificial neural network" (ANN) to predict the slope failures based on historical circular slope cases. Using the feed-forward back-propagation algorithm with a multilayer perceptron network, ANN is a powerful ML method capable of predicting the complex model of slope cases. However, the prediction result of ANN can be improved by integrating the statistical analysis method, namely grey relational analysis (GRA), to the ANN model. GRA is capable of identifying the influencing factors of the input data based on the correlation level of the reference sequence and comparability sequence of the dataset. This statistical machine learning model can analyze the slope data and eliminate the unnecessary data samples to improve the prediction performance. Grey relational analysis-artificial neural network (GRANN) prediction model was developed based on six slope factors: unit weight, friction angle, cohesion, pore pressure ratio, slope height, and slope angle, with the factor of safety (FOS) as the output factor. The prediction results were analyzed based on accuracy percentage and receiver operating characteristic (ROC) values. It shows that the GRANN model has outperformed the ANN model by giving 99% accuracy and 0.999 ROC value, compared with 91% and 0.929.

1 Introduction

Landslide is a common geological hazard that occurs worldwide every year. Slope failure is the most contributed factor to the landslide, triggered by excessive rainfall, earthquake, or human intervention such as deforestation and construction. Slope failures can be described as a movement of soil or rock downward the earth's gravity [1]. Furthermore, this natural disaster can cause significant damage to the environment and properties and contribute to the human loss of life. To mitigate the impacts and damages caused by the landslides, the prediction of slope failure is a great consideration to which researchers have been investigated progressively for the past decades. Nevertheless, the prediction of slope failure remains a challenging task due to its complexity and uncertainties of geological factors and unbalance data samples. Traditional methods such as the limit equilibrium method (LEM), finite equilibrium method (FEM) are found to be computationally efficient, as reported by Ray [2]. However, these methods fail to have a comprehensive understanding of the slope's behavior due to the inherent drawbacks. Furthermore, the method often criticized its intensive computational power required and lacked computational efficiency for small probability levels [3]. Thus, with the introduction of the machine learning (ML) approach that is capable of modeling complex problems, the prediction of slope failure has been rapidly growing for the past few decades. ML approach is a powerful intelligent method capable of serving as a replacement for the traditional methods to predict slope failures.

ML approaches such as artificial neural network (ANN), decision tree (DT), Naïve Bayes (NB), support vector machine (SVM), Random Forest (RF), Logistic Regression (LR), etc. have been well studied and attracted interest among researchers [4-9]. ANN method is proposed based on the ML algorithm is capable of learning the relationship between the input factors and the FOS value from the historical slope cases. The previous ANN studies found that this intelligent computational tool can tackle complex problems in the geotechnical engineering field. Chakraborty and Goswami developed an ANN model to predict FOS value for slope stability analysis in Jorabat-Shillong Expressway, India [4]. Initially, 200 slope cases were analyzed based on four conventional methods: "Bishops, Morgenstern and Price, Janbu and Fellenius method," and the obtained FOS value was used to develop prediction models. The result found that ANN gives the correlation value of 99.63% compared to multilinear regression with 96.14%. Ray et al. have developed ANN prediction models to evaluate FOS of residual soil slope in Shiwalik, Himalaya. Eleven slope factors were investigated, and the factors that have a higher correlation coefficient were chosen as the input factors for the development of ANN prediction models

* Corresponding author: badariah@uniten.edu.my

[10]. The models were develop based on a multi-layer feed forward back-propagation network (FFBPN) with the optimal number of hidden neurons. The prediction results were evaluated using coefficient of determination (R2), root mean square error (RMSE), residual error and variance account. The author concluded that ANN could give a good performance and a straightforward, reliable, and valid ML approach for slope prediction. ANN also proven has outperformed other techniques. Bharati et al. applied ANN to predict the FOS value of dragline dump slope using 216 simulated dragline dump slope [11]. The prediction models were developed based on input parameters, including slope height, slope angle, and coal-rib height. The prediction performances were investigated based on the R and RMSE value. The prediction result of ANN was then compared with MLR, and it was found that ANN gives higher accuracy than the MLR model. It should be noted that although the previous study discussed above show their significance, there is still a problem that needs to be adequately conveyed, such as the significant relationship between each of the influencing parameters has not been extensively explored. Hence, this study attempts to integrate an ML method with the statistical method to improve the prediction performance.

A countless number of slope failure prediction models have been investigated and developed using physics and statistical-based modeling. Generally, physics-based models developed on creep theory depict soil and rock's fundamental relationship, while statistical models developed based on regression analysis of slope failure's historical occurrence [12]. However, physics-based models need further observations on-site and laboratory tests to identify mechanical and physical parameters and limit their broad applications. In contrast, the statistical model does not need the determination of physical parameters. It uses various statistical analysis methods to establish the response relationships between the slope failure occurrences and their associated contributing factors. Consequently, the statistical model has increased its popularity among researchers due to its simplicity and ease of implementation. Statistical analyses such as regression, grey system, and autoregressive-integrated moving average (ARIMA) have been widely applied in slope failure prediction. The integration of statistical analysis and the ML approach has been increasing in popularity in modeling landslide susceptibility. This approach attracts interest to address the challenge of landslide modeling for large regions, especially when the geotechnical data may not be adequate to implement the physics-based method [13].

The current study's objective is to develop an integration prediction model named GRANN to predict slope failure utilized a case study conducted by Sah et al. [14]. This model is the integration of machine learning, ANN, and statistical approach, GRA. The GRA is considered as a feature selection method, whereas the ANN is the classifier. The proposed model is expected to investigate the significant relationship between FOS and its influencing factors and eliminate the least influencing factor to improve the prediction performance. The result of GRANN was then compared with a single ANN model to observe its performance for slope failure prediction.

2 Methodology

2.1 Grey Relational Analysis

Professor Deng Julong has proposed an analysis method in Grey System theory which is called GRA [15]. In the grey theory, a level of information is between the black (unknown) and white (known), which means some of the information, is known, and some of the information is unknown. According to Lu and Rosenbaum, the possible examples of "grey" conditions in slope stability include ground quality characteristics that may refer to fuzzy values such as good, fair, poor, or slope fuzzy dampness values of damp, wet, and dry [16]. These slope's fuzzy conditions make it complex and challenging to measure the exact values of slope factors. Furthermore, the relationship between input factor and output factor may also be complicated to establish. Grey System provides various methods to deal with "grey" information or knowledge and thus can help predict the slope instability slope, considering its inherent subjectivity and ambiguity. GRA, one of the techniques mainly applied in Grey Systems, is used to illustrate how each of the input factors influences the output factor, FOS. This influence is denoted by grey relational grade (GRG) value. An input factor that gives a higher GRG value indicates a more significant relationship to the FOS. In comparison, a factor that offers a lower GRG value indicates a less significant relationship to the output factor. Figure 1 shows the simplified steps of GRA.

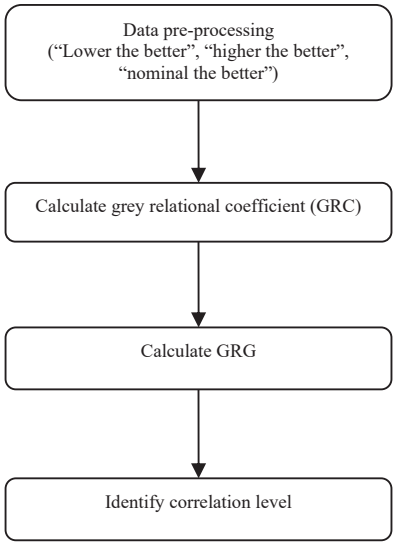


Fig.1. Simplified GRA steps.

Firstly, the data needs to be normalized in range [0,1] based on the three data characteristic namely "higher the better", "lower the better", and "nominal the better". This study applies "lower the better", as described in Eq. (1) [17].

$$x_i{}^*(k) = \frac{\max x_i{}^0(k) - x_i{}^*(k)}{\max x_i^0(k) - \min x_i^0(k)} \quad (1)$$

Where $x_i{}^*(k)$ is the current data to be normalized, $\max x_i^0(k)$ is the maximum value of the data, $\min x_i^0(k)$ is the minimum value of the data and x^0 is the desired target value. The relationship between the ideal and normalized experimental results is expressed GRC value, which is calculated as in Eq. (2):

$$\xi_i(k) = \frac{\Delta\min + \zeta.\Delta max}{\Delta_{0i}(k) + \zeta.\Delta max} \quad (2)$$

Where ζ is a distinguishing coefficient which is in range of [0.1], and generally, ζ = 0.5 is used [18]. $\Delta_{0i}(k)$ is the deviation sequence which can be calculated as follows:

$$\Delta_{0i}(k) = \left\| x_0^*(k) - x_i^*(k) \right\| \quad (3)$$

$$\Delta\min = \min_{\forall j \in i} \min_{\forall k} \left\| x_0^*(k) - x_j^*(k) \right\| \quad (4)$$

$$\Delta\max = \max_{\forall j \in i} \max_{\forall k} \left\| x_0^*(k) - x_j^*(k) \right\| \quad (5)$$

Where $x_j^*(k)$ is the comparability sequence and $x_0^*(k)$ is the reference sequence and. After the GRC is determined, the value of GRG is calculated based on the formula in Eq. (6) :

$$GRG(y_i) = \frac{1}{n} \sum_{k=1}^{n} \xi_i(k) \quad (6)$$

GRG value shows the correlation level among the reference and the comparability sequence. The reference sequence refers to the output data, which is the value of FOS and the comparability sequence refers to the input factors of the slope. The value of GRG is in the range [0, 1] and is equal to 1 if the two sequences are identical. Typically, the input factor with a GRG value of less than 0.6 is considered a less influencing factor to the output. The higher GRG value shows that it influences the output value [19].

2.2 Artificial Neural Network (ANN)

ANN is a supervised ML method that imitates biological neurons in the human brain. ANN is a powerful tool to solve a complex problem, consists of artificial neurons connected for transmitting information in the neurons via the tendons. The most frequently used ANN architecture is multi-layer perceptron (MLP) consists of the input, hidden and output layer, connected to each other through weight and bias [20]. Each layer contains a number of neurons. The neuron in each input factors is multiplied with the corresponding weight to produce the product, as in Eq.(7):

$$net = \sum_{i=0}^{n} w_i x_i + \mathrm{b} \quad (7)$$

where n is the number of inputs, x_i is the input factor, w_i is the corresponding weight and b is bias. Then, the sum is processed using a nonlinear activation function. Note that if the sum is exceeds the threshold, then the neuron is activated. The output y_i as in Eq. (8):

$$y_i = f(net) \quad (8)$$

where f is the activation function. The optimal network architecture can be selected based on trial and error basis [21]. Number of hidden layer and the neurons for hidden layer is selected based on the problem itself. Figure 2 shows the architecture of MLP network.

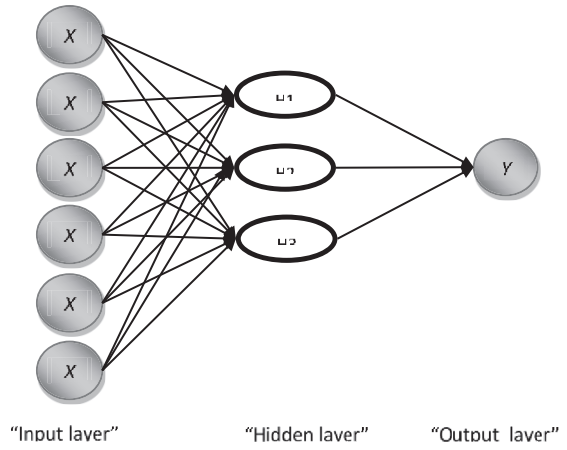


Fig. 2. MLP architecture.

This study applies the ANN prediction model to classify the FOS based on the slope input parameters using a FFBPN. Before the prediction process can be conducted, the slope cases data sets must be divided into training and testing datasets. The data division process is an essential aspect for any ML method where the selection of training and testing datasets will affect the prediction performance. The training dataset should represent the entire data set [22]. The model must be trained appropriately to determine the network's optimal weights and biases [21]. If the training dataset is limited, the network could not learn properly and if too large a training dataset is used, it is difficult to verify the generalization capability. Furthermore, it may lead to overfitting [10]. Hence, it is important to choose the proper training and testing datasets. The current study applies 70% slope cases for training and 30% slope cases for testing.

2.3 Case study

Historical cases of slope cases were extracted from Sah et al., (1994) to developed prediction models. The data consists of 46 cases of the circular slope with six input parameters, namely "unit weight, internal friction angle, cohesion, slope angle, slope height and pore pressure ratio" and FOS as the output factor. Two types of slopes were classififed from the 46 slope cases, which are "stable" and "fail". Figure 3 shows the bar chart of the stable and fail slope cases.

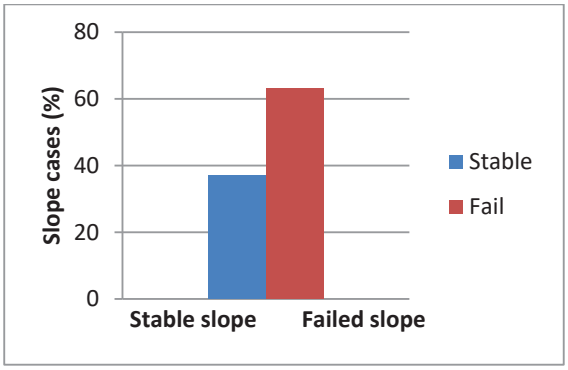


Fig. 3. Stable and failed slope cases.

3 Result and discussion

In GRANN model, the significant factors of six comparability sequences xi namely "unit weight, cohesion, internal friction angle, slope angle and pore pressure ratio" to the reference sequence x0, FOS value is determined in terms of GRG value. Table 1 shows the GRG value for all the input factors of circular slope.

Table 1. GRG values for the input parameters.

Input factors	GRA Analysis	
	GRG	Rank
Unit weight (γ)	0.6388	3
Cohesion, (c)	0.6392	2
Internal friction angle (φ)	0.6227	6
Slope angle (β)	0.6327	5
Slope height (H)	0.6481	1
Pore water pressure ratio (ru)	0.6375	4

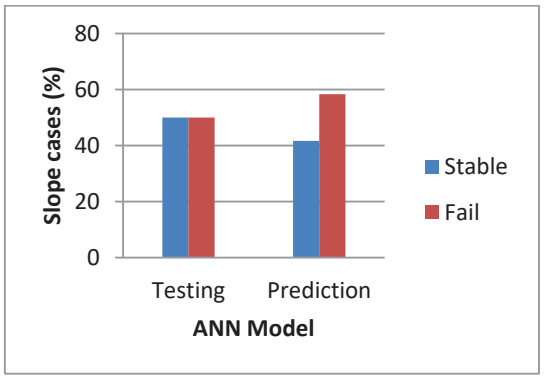
From Table 1, it can be seen that slope height is the most influential factor to the FOS value with the GRG value is 0.6481 and internal friction angle is the least influence factor to the FOS value with the GRG value is 0.6227. Note that the GRG values of the input factors are in range [0.6, 0.7], which is considered as noticeable influenced [19, 23]. Based on the GRG value, the input factor that ranked last will be considered for the elimination from the input data. Elimination the input factor will reduce the input numbers and consequently will reduce the running time of the prediction model [19]. The classification accuracy is calculated as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (9)$$

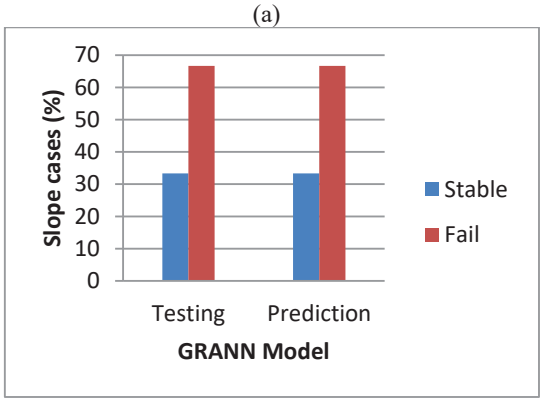
Where TP is "True Positive: the number of correct positive prediction, TN is True Negative: the number of correct negative prediction, FP: False Positive: the number of incorrect positive prediction and FN: False Negative: the number of incorrect negative prediction". Table 2 shows the confusion matrix of ANN and GRANN with the accuracy and ROC value while Figure 4 shows the graphs of ANN and GRANN prediction models.

Table 2. Confusion matrix of ANN and GRANN.

		Actual		Accuracy (%)	ROC
ANN		+	-	91.67	0.9285
Predictive	+	6	1		
	-	0	5		
GRANN				99.99	0.999
Predictive	+	8	0		
	-	0	4		



(a)



(b)

Fig. 4. Graphs of testing and prediction result for ANN (a) and GRANN (b) models.

Based on Table 2, the prediction result shows that for ANN, six stable slope cases are correctly predicted, and one slope case is wrongly predicted, while three failed slope cases are correctly predicted, and no stable slope case is incorrectly predicted. For GRANN, eight stable slope cases are correctly predicted and no stable slope case is incorrectly predicted, while four failed slope cases are correctly predicted, and no failed slope case is incorrectly predicted. ANN prediction model consists of six input parameters, while the GRANN prediction model consists of five input factors in which internal friction angle is eliminated from the input data. The integration of ANN was proven to improve the ANN model's performance by giving the accuracy and ROC value with 99% and 0.999, compared to ANN with 92% and 0.9444, respectively. Figure 4 shows that for the ANN model, the testing dataset gives 50% of stable slope and 50% of the failed slope, while the prediction result shows that 41.67% of stable slope and 58.33% of the failed slope. For the GRANN model, both the testing dataset and prediction result show similarity for stable and failed slopes, which are 33.33% and 66.67%, respectively.

4 Conclusion

This study attempts to develop an integration model of GRA and ANN to predict the slope failure of circular slopes. 46 slope cases were collected from previous study is used to develop prediction model based on ANN

with six input factors; “cohesion, slope height, slope angle, internal friction angle, unit weight and pore pressure ratio”. A three layer FFBPN with 6-3-1 architecture was selected for the prediction model. The dataset is divided into 70% for training and 30% of testing dataset. The prediction result of ANN gives 92% accuracy. To improve the prediction performance, GRA is integrated with ANN during the pre-processing phase where GRA analyze all the input factors. It was found that internal friction angle gives the least significant effect to the FOS value by giving the GRG value 0.6228, hence it will be eliminated from the input factors. The prediction result of GRANN gives 99% accuracy. GRANN show the capability of the model to analyze the slope data and eliminate the unnecessary data samples to improve the prediction performance.

This project was funded by UNITEN under Fundamental Research Grant Scheme FRGS/1/2018/STG08/UNITEN/02/1.

References

1. I. Chowdhuri, S. C. Pal, A. Arabameri, P. T. T. Ngo, R. Chakrabortty, S. Malik, P. Roy, Env. Earth Sci. **79**, 20:1−28 (2020)
2. A. Ray, V. Kumar, A. Kumar, R. Rai, M. Khandelwal, T.N. Singh, Natural Hazards **103**, 3: 3523−3540 (2020)
3. S. Y. Liu, L. T. Shao, H. J. Li, Comp. and Geotechnics **63**, 291−298 (2015)
4. A. Chakraborty, D. Goswami, Arabian J. of Geosciences **10**, 17, 1−11.
5. L. Y. Zhou, F. P. Shan, K. Shimizu, T. Imoto, H. Lateh, K.S. Peng, *A comparative study of slope failure prediction using logistic regression, support vector machine and least square support vector machine models*, in AIP Conference Proceedings (Vol. 1870, No. 1, p. 060012), August 2017, Publishing LLC (2017)
6. A. Aditian, T. Kubota, Y, Geomorphology **318**, 101−111 (2018)
7. X. Feng, S. Li, C. Yuan, P. Zeng, Y. Sun, KSCE J. of Civ. Eng. **22**, 3:941−950 (2018)
8. X. N. Bui, H. Nguyen, Y. Choi, T. Nguyen-Thoi, J. Zhou, J. Dou, Sci. reports **(10)**, 1:1−17 (2020)
9. M. Nouri, P. Sihag, F. Salmasi, J. Abraham, Geotechnical and Geological Eng., 1-14 (2021).
10. A. Ray, V. Kumar, A. Kumar, R. Rai, M. Khandelwal, Natural Hazards **103**, 3523−3540 (2020)
11. A. K. Bharati, A. Ray, M. Khandelwal, R. Rai, A. Jaiswal, (2021). Eng. with Comp, 1−9 (2021)
12. Y. Jiang, Q. Xu, Z. Lu, H. Luo, L. Liao, X. Dong, X. (2021). Natural Hazards and Risk **12**, 1:741−762 (2021)
13. J. N. Goetz, A. Brenning, H. Petschko, P. Leopold, P. (2015). Comp. & Geosci. **81**, 1-11 (2015)
14. N. K. Sah, P. R. Sheorey, L. N. Upadhyaya, Inter. J. of Rock Mechanics and Mining Sci. & Geomechanics Abstracts **31**, 47−53 (1994)
15. S. Liu, J. of Grey System **30**, 1:1−4 (2018)
16. P. Lu, M. S. Rosenbaum, Natural Hazards **30**, 3:383−398 (2003)
17. Y. Kuo, T. Yang, G. W. Huang, G. W. Comp. & indust. Eng. **55**, 1:80−93 (2008)
18. A. H. Bademlioglu, A. S. Canbolat, O. Kaynakli, Renewable and Sustainable Energy Reviews **117**, 109483 (2020)
19. R. Sallehuddin, S. M. Shamsuddin, S. Z. M. Hashim, *Forecasting small data set using hybrid cooperative feature selection*, in 2010 12th International Conference on Computer Modelling and Simulation (pp. 80-85), March 2010, IEEE.
20. Z. G. Qian, A. J. Li, W. C. Chen, A. V. Lyamin, J. C. Jiang, Soils & Foundations **59**, 2:556−569 (2019)
21. S. Rukhaiyar, M. Alam, N. Samadhiya, Int. J. Geotech. Eng. **12**, 556−566 (2018)
22. A. Trigila, C. Iadanza, C. Esposito, G. Scarascia-Mugnozza G, *Geomorphology* **249**, 119–136 (2015)
23. A. M. Deris, A. M., Zain, R. Sallehuddin, Meccanica **48**, 8:1937−1945 (2013)