Slope Stability Prediction Based on GA-HIDMS-PSO-SVM

Jinbo Zhang[1a], Yubin Lin[1b], Xinhong Li[1c], Yuanhao Chen[1d], Xue Chen[1e], Dacai Chen[1f], Xiaoxiang Chen[2g*]

[1]Economic and Technological Research Institute, State Grid Fujian Electric Power Co., Ltd., Fuzhou 350000, Fujian, China
[2]POWERCHINA Fujian Electric Power Engineering Co., Ltd., Fuzhou 350003, China

[a]45257509@qq.com, [b]18606932711@163.com, [c]1536427587@qq.com, [d]462078820@qq.com, [e]50580111@qq.com, [f]751063221@qq.com, [g]chenxx919@126.com
*Corresponding author's email: chenxx919@126.com

Abstract: Slope instability leads to significant global economic losses annually. To evaluate slope stability **swiftly and precisely**, ensuring the safety of slope engineering. The GA-HIDMS-PSO-SVM algorithm is introduced to develop a slope stability prediction model. Six typical slope parameters, including bulk density, internal friction angle, cohesion, slope angle, slope height, and pore water pressure ratio, were selected as input factors, while the slope state was chosen as the output factor. The training dataset was built using data from 80 real-world engineering projects. The model achieves an accuracy value of 0.958, a recall rate of 0.959, a precision of 0.960, and an F1-score of 0.959, demonstrating exceptional prediction accuracy, robust generalization, and reliable results in identifying slope instability. Combined with engineering examples, it is proved that the evaluation results of slope state are consistent with the actual situation. The results show that GA-HIDMS-PSO-SVM model can be applied to slope stability prediction, which can provide basis for slope design and construction, and has a good application prospect in practical engineering application.

Keywords: Slope stability, Machine learning, GA-HIDMS-PSO-SVM model

1 Introduction

Slope instability is a complicated natural phenomenon, which has caused irreversible natural disasters and economic losses in many countries and regions in the world. Hence, assessing slope stability is essential[1]. Slope stability research, a key focus in geotechnical engineering for over 300 years, is more challenging and risk-prone than other projects. There are many types of slope failure factors, including slope material, shape, stress-strain conditions, geological structure factors, climatic reasons, earthquake, drainage, construction and so on. For the stability of the slope, the safety factor (FOS) is commonly used to evaluate, which is used to represent the overall or local

G. Zhao et al. (eds.), *Proceedings of the 2024 6th International Conference on Civil Architecture and Urban Engineering (ICCAUE 2024)*, Advances in Engineering Research 262,
https://doi.org/10.2991/978-94-6463-688-8_49

stability of the slope. A factor of safety (FOS) of 1 indicates that the slope is in a critically stable state, while the smallest FOS value represents the slope's stability condition. Numerous stability assessment techniques are currently available, including simple assessments, planar failure analysis, limit state criteria, the limit equilibrium method, numerical analysis, hybrid methods, advanced techniques, and both 2D and 3D implementations[2]. Although the stability analysis theory is relatively perfect, in some practical engineering applications, many factors affect the stability of slope, and under the coupling action of these factors, it is still difficult to accurately express the change of slope and forecast the safety factor.

With advancements in computer technology, scholars have employed machine learning algorithms to address the complex nonlinear problem of slope deformation. These methods use existing slope data to analyze the relationship between slope stability and its corresponding effects. Existing studies have demonstrated that methods such as artificial neural network (ANN) [3], gradient propulsion machine (GBM) [4], radii basis function (RBF) [5], support vector machine (SVM) [6], LightGBM algorithm [7], etc., have obtained relatively good prediction results. The Support Vector Machine (SVM) model, grounded in statistical learning theory, replaces the empirical risk minimization principle used in traditional machine learning methods with the structural risk minimization principle. The model has good generalization ability, which overcomes the shortcomings of traditional artificial neural network. Studies indicate that the SVM model performs exceptionally well in accuracy and stability, though its generalization capability is highly sensitive to parameter selection [8]. The primary limitation of the SVM model is its difficulty in capturing key modeling variables [9], significantly restricting its performance.

Therefore, it is imperative to apply optimization algorithm to automatically determine the optimal parameters of support vector machine model. GA-HIDMS-PSO is an advanced hyperparameter optimization algorithm which combines heterogeneous improved dynamic multi-swarm particle swarm optimization (HIDMS-PSO) with genetic algorithm (GA). In essence, GA-HIDMS-PSO is an efficient, parallel and global search method. Moreover, GA-HIDMS-PSO has the characteristics of parallelism and global optimization, and has shown the optimization performance that other search methods do not have in many researches[10, 11].

Therefore, in this study, GA-HIDMS-PSO algorithm is used to optimize the parameters of support vector machine, and a correlation prediction model of GA-HidmS-PSO-SVM is constructed. Based on the prediction model, six factors including cohesion, bulk density, internal friction Angle, slope Angle, pore water pressure ratio and slope height were selected as the prediction indexes of slope stability. The predictive performance of the model was evaluated using area under subject operating characteristic curves (AUC), F1 scores, accuracy, precision, recall, and confusion matrices. This study aims to provide a new method for slope stability prediction, and can provide valuable reference value and suggestions for similar slope control measures.

2 GA-HIDMS-PSO-SVM Model

2.1 SVM Principle

Support vector Machine (SVM) is a general learning model developed on the basis of statistical learning theory and structural risk minimization principle. SVM shows strong generalization ability when dealing with small samples and nonlinear and high-dimensional pattern recognition problems. Let the sample training set be $T = \{(x_i, y_i)|i = 1,2,\cdots,N\}$, where $x_i \in R^d, y_i \in \{-1,1\}$, segment the sample by looking for hyperplanes. The fundamental principle of SVM segmentation is to maximize the margin between classes, which is expressed as a convex quadratic programming problem. The objective is to maximize the classification boundary.

Solving the optimal classification hyperplane problem is often translated into solving the following optimization problems:

$$\begin{cases} \min\ (\frac{1}{2}w^2 + C\sum_{i=1}^{N}\varepsilon_i) \\ s.t. y\ (wx + b) \geq 1 - \varepsilon_i, i = 1,2,\cdots,N \end{cases} \tag{1}$$

Where, w is the normal vector of the hyperplane; b is offset; C is a penalty parameter that is used to balance the tradeoff between maximizing classification margin and minimizing classification errors, thereby affecting the performance of the support vector machine (SVM). The variable ε_i is used as the relaxation parameter to estimate the number of misclassified samples.

Lagrange multiplier α_i and α_i^* was introduced to transform the objective function into a dual problem:

$$\begin{gathered} min\left\{\frac{1}{2}\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) - \sum_{i=1}^{n}\alpha_i(y_i - \delta) + \sum_{i=1}^{n}\alpha_i^*(y_i + \delta)\right\} \\ s.t.\ \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{gathered} \tag{2}$$

Where $k(x_i, x_j)$ is a kernel function where α_i is a Lagrange multiplier. The common types of kernel functions are polynomial kernel function, linear kernel function, sigmoid function and radial basis kernel function (RBF). The kernel function can be expressed as follows:

$$k(x_i, x_j) = \varphi(x_i)\varphi(x_j) \tag{3}$$

Where $\varphi(x_i)$ is a nonlinear mapping of sample set in high dimensional space. According to equations (2) and (3), the nonlinear support vector machine is expressed as:

$$f(x) = sgn(\sum_{i=1}^{N} \alpha_i y_i k(x_i, x_j) + b) \tag{4}$$

2.2 GA-HIDMS-PSO Principle

GA-HIDMS-PSO is a meta-heuristic optimization algorithm by combining the complex particle swarm optimization algorithm (PSO) with genetic algorithm (GA). The hybrid model can make full use of the heterogeneous characteristics of HIDMS-PSO and the evolutionary characteristics of genetic algorithm. So, in this particular framework of GA-HIDMS-PSO, HIDMS-PSO serves as the primary search mechanism. Genetic algorithm can be used as an auxiliary method to improve the performance of homogeneous and heterogeneous subgroups and reduce the loss of diversity. After iteratively running through these two methods, the computational time of HIDMS-PSO, as the main search strategy, is longer than that of genetic algorithm. HIDMS-PSO generates initial solutions of genetic algorithms from homogeneous and heterogeneous subpopulations. The final solution generated in the genetic algorithm replaces the previously calculated solution in the HIDMS-PSO, starting a new round of searching, and introducing more different particles to guide the population. Please refer to the previous literature for detailed information. Based on a, the penalty factor (c) and RBF kernel function parameter (g) of the SVM model were optimized, and the GA-HIDMS-PSO-SVM prediction model was constructed, as shown in Figure 1.

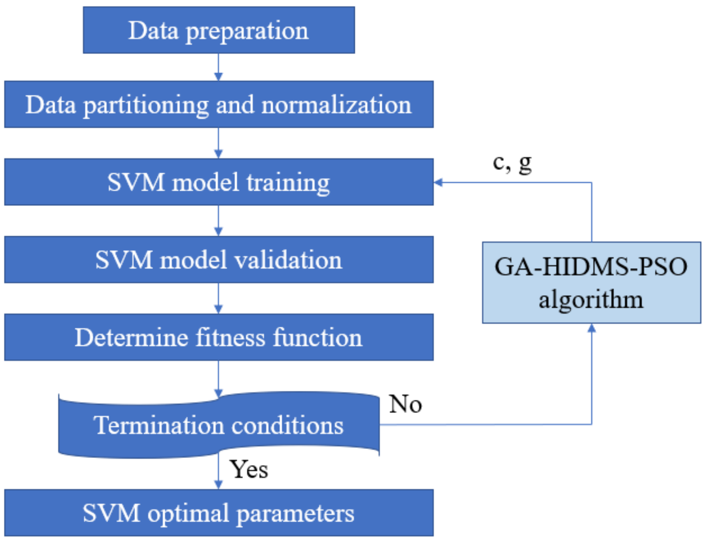

Fig. 1. Flowchart of GA-HIDMS-PSO-SVM model

3 Slope Stability Prediction Model

3.1 Data Preparation

By collecting engineering site data and reviewing related literature, 80 soil slope projects were gathered to evaluate the reliability of the GA-HIDMS-PSO-SVM model. The input parameters of the model were determined to be six factors: bulk density (γ), internal friction angle (φ), cohesion (c), slope angle (β), pore water pressure ratio (r_u) and slope height (H), which are the primary factors influencing slope stability. The output parameter is the slope state, categorized as 1 for "slope instability" and 2 for "slope stability." As illustrated in Table 1, the data exhibit diverse characteristics and dimensions, with considerable variability among the samples. To mitigate dimensional disparities and improve data comparability, normalization of the data is essential. The sample data are normalized to the interval [-1, 1].

Table 1. Partial sample data

γ	c	φ	β	H	r_u	Slope state
18.84	15.32	30	25	10.7	0.38	2
18.84	0	20	20	7.6	0.45	2
21.43	0	20	20	61	0.5	2
19.06	11.71	28	35	21	0.11	2
…	…	…	…	…	…	…
22.4	10	35	45	10	0.4	1
24	0	40	33	8	0.3	2
20	0	24.5	20	8	0.35	2
18	0	30	20	8	0.3	2

3.2 Determination of Evaluation Indicators for Slope Stability

As a binary classification problem, slope stability evaluation needs to ensure the independence of each factor. C A correlation analysis should be performed on the six index factors in the dataset, and highly correlated factors should be excluded to prevent interference between them. Fig. 2 shows the correlation of each factor. The Pearson correlation values were computed using the Origin software platform. If the Pearson correlation value is < 0.2, the factors are uncorrelated; if the value is between 0.2 and 0.4, the correlation is weak; if the value is between 0.4 and 0.7, the correlation is moderate; and if the value exceeds 0.7, the factors exhibit strong collinearity. As shown in Fig. 2, the highest correlation coefficient among all evaluation factors is 0.49, indicating that there is no significant correlation between the factors. The selected factors can be used as training samples for the GA-HIDMS-PSO-SVM model.

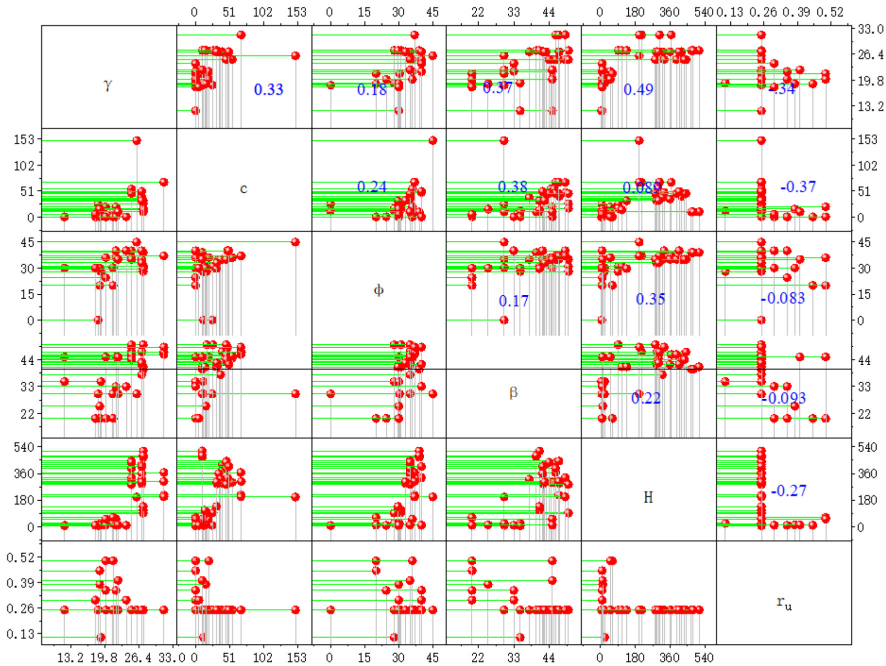

Fig. 2. Evaluation factor and its correlation

3.3 Prediction Results

The prediction of slope stability is actually a problem of supervision classification. This paper builds the model by learning the characteristics of the slope and the corresponding target value of the slope. The test data set must be divided into two sets, one is the training set and the other is called the test set. The content of the training set should be sufficiently representative of the entire data set. If the number of training sets is too small, the predictive model may not be able to properly learn all the cases. If it is too large, it may lead to overfitting of the model. The implication is that overfitting is prone to occur when the proportion of training sets is too large. In this study, the proportion of training set and test set was set to 70% and 30% respectively. With this ratio, it can effectively ensure that the model has enough learning ability, and then it can produce more accurate test results. The data set is randomly divided into two sets: the training set and the test set. The training set is 70% of the main data set, and the test set is 30%. The algorithm is programmed and tested on MATLAB software platform. After the model training is completed, all the samples of the test set can be input into the predicted model to predict the slope stability.

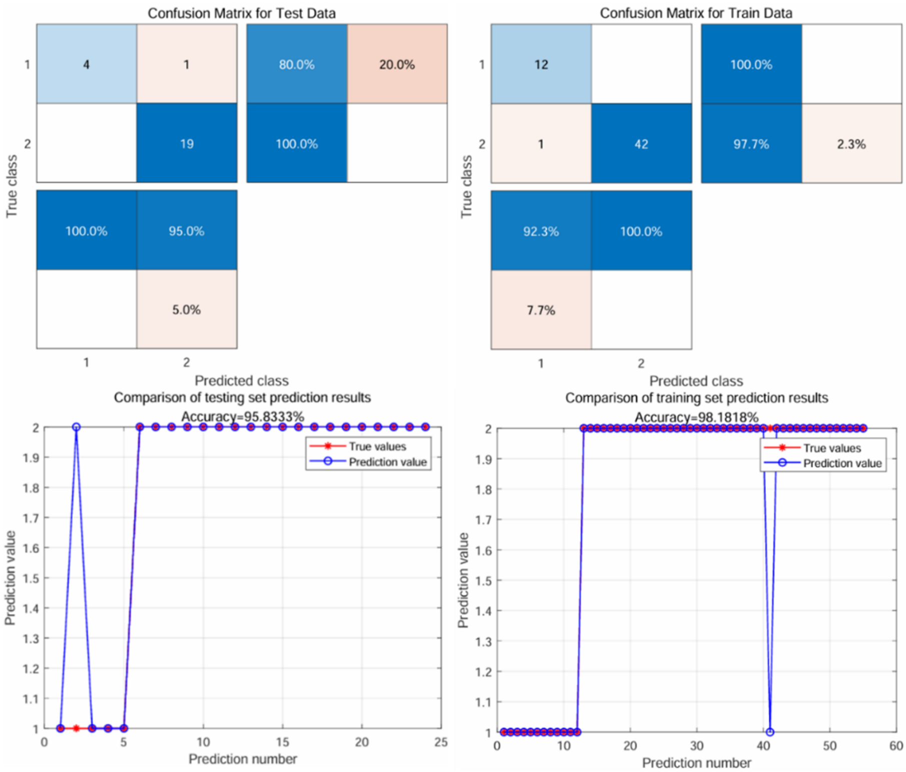

Fig. 3. GA-HIDMS-PSO-SVM model prediction results

The model's prediction results are presented in Fig. 3. Both the training and testing sets demonstrate ideal prediction results, with the predicted stable state closely matching the actual state. This is further supported by the evaluation metrics in Table 2, all of which show an accuracy of 0.95 or higher. These test results show that the model presented in this paper can be used to predict slope stability and has a good prediction result.

To verify the effectiveness of the proposed model, the optimization algorithm was applied to the ANN model, and these results are shown in Figure 4. The prediction accuracy remains relatively high at 0.916, though it is lower than that of the GA-HIDMS-PSO-SVM model. The robustness of ANN models is relatively low, as confirmed by previous studies. In conclusion, the proposed model delivers satisfactory prediction results.

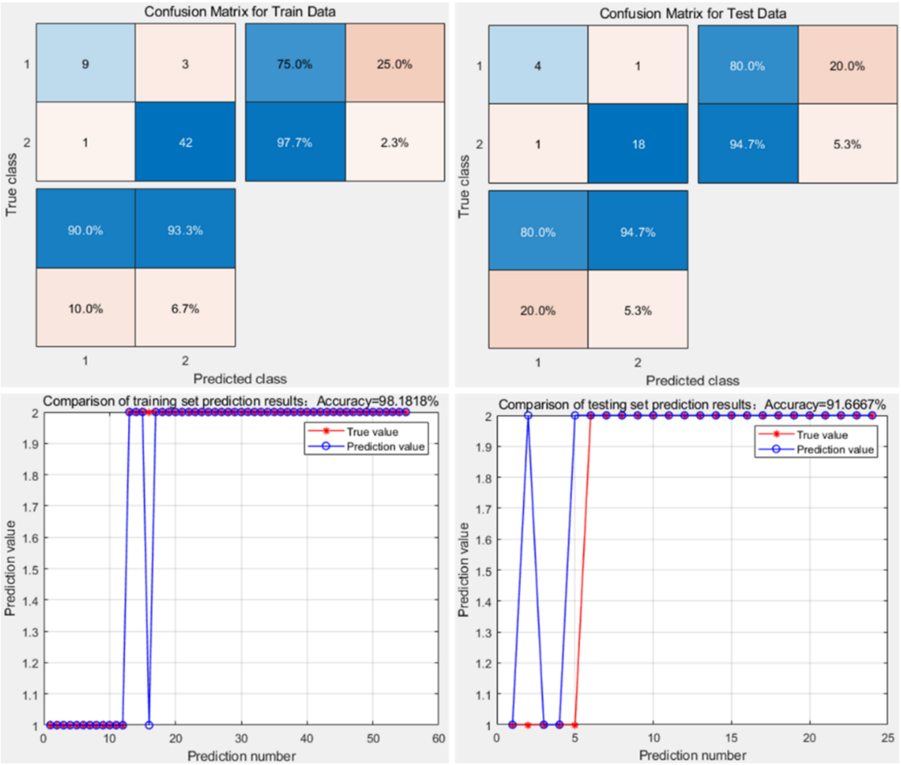

Fig. 4. GA-HIDMS-PSO-ANN model prediction results

Table 2. Prediction performance evaluation results

Model	Accuracy	Recall	Precision	F1-score
GA-HIDMS-PSO-SVM	0.958	0.959	0.960	0.959
GA-HIDMS-PSO-ANN	0.916	0.917	0.918	0.916

3.4 Actual Case Application

In order to further explain the rationality of the prediction performance of this model, it is applied to the classical slope cases recorded in the existing literature. The geometric model parameters and physical and mechanical parameters of slopes under various working conditions are shown in Table 3. The results predicted by this model are shown in Table 4. The proposed model also yields impressive results in actual slope cases, with the exception of an error in predicting the stable state of slope 2 as unstable.

Table 3. Cassic slope case data

No.	γ	c	φ	β	H	r_u	FOS
1	20.6	0.0	30.0	55.0	12.0	0.31	0.76
2	20.6	20.0	36.0	55.0	12.0	0.31	1.12
3	18.7	10.0	31.0	55.0	17.0	0.31	0.81
4	18.7	30.0	30.0	55.0	12.0	0.31	1.18
5	20.6	30.0	39.7	55.0	17.0	0.31	1.23

Table 4. Comparison between the predicted and actual results of the case slope

No.	FOS	True stability	Prediction stability
1	0.76	Instability	Instability
2	1.12	Stability	Instability
3	0.81	Instability	Instability
4	1.18	Stability	Stability
5	1.23	Stability	Stability

Overall, the proposed model effectively distinguishes the stable state of slopes and can be deployed for practical engineering applications. However, the model proposed in this study has the following limitations: (1) The model is primarily applied to homogeneous slopes, and its performance in predicting other types of slopes requires further investigation; (2) This study only considers the impact of groundwater on slopes through the pore pressure ratio, which is not convenient for practical application and lacks comprehensiveness. In future work, the influence of rainfall conditions on slope stability should be incorporated to enhance the practical value of the GA-HIDMS-PSO-SVM model; (3) This study primarily focuses on applying the model to the stability analysis of homogeneous slopes. It does not yet address slope instabilities induced by rainfall, earthquakes, or structural plane-controlled rock slope stability. Once the key factors influencing slope stability are identified, future work can explore the application of the proposed model to these scenarios.

4 Conclusion

In this study, the GA-HIDMS-PSO-SVM method is used to take 80 practical engineering cases as the samples of this training set. A high precision slope stability prediction model is established to predict the stability of the slope. The following 6 factors, including bulk density (γ), cohesion (c), slope height (H), internal friction Angle (φ), slope Angle (β), pore water pressure ratio (ru), were selected as input variables, and the stable state of the slope was taken as the output variable of the sample. The area, accuracy, recall, accuracy and f1 scores under the confusion matrix were used to evaluate the predictive performance of the model. Finally, the main conclusions of this paper are as follows:

- The GA-HIDMS-PSO-SVM model was trained using past engineering cases, establishing the response relationship between slope stability and the six factors. The accuracy, recall, precision, and F1-score were 0.958, 0.959, 0.960, and 0.959, respectively, demonstrating the model's high prediction accuracy, strong generalization capability, and reliable performance in identifying slope instability.
- When applied to practical engineering cases, the proposed model's predicted stability state of actual slopes was found to be consistent with the observed conditions. The GA-HIDMS-PSO-SVM model can be utilized in practical engineering to support slope design, construction, and risk management.
- The results of this study also demonstrate that the GA-HIDMS-PSO algorithm optimizes the parameters of the SVM, addressing parameter selection issues, improving classification performance, and ensuring the algorithm's stability.

References

1. H.-J. Park, T.R. West, I.J.E.G. Woo, Probabilistic analysis of rock slope stability and random properties of discontinuity parameters, Interstate Highway 40, Western North Carolina, USA, 79(3-4) (2005) 230-250.
2. A. Chakraborty, D.J.I.J.o.G.E. Goswami, State of the art: Three dimensional (3D) slope-stability analysis, 10(5) (2016) 493-498.
3. M. Azarafza, M. Hajialilue Bonab, R.J.S.R. Derakhshani, A novel empirical classification method for weak rock slope stability analysis, 12(1) (2022) 14744.
4. M. Azarafza, H. Akgün, A. Ghazifard, E. Asghari-Kaljahi, J. Rahnamarad, R.J.I.J.o.D.E. Derakhshani, Discontinuous rock slope stability analysis by limit equilibrium approaches–a review, 14(12) (2021) 1918-1941.
5. J. Zhou, E. Li, S. Yang, M. Wang, X. Shi, S. Yao, H.S.J.S.S. Mitri, Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, 118 (2019) 505-518.
6. W. Wei, X. Li, J. Liu, Y. Zhou, L. Li, J.J.A.S. Zhou, Performance evaluation of hybrid WOA-SVR and HHO-SVR models with various kernels to predict factor of safety for circular failure slope, 11(4) (2021) 1922.
7. J. Wu, X.-Y. Chen, H. Zhang, L.-D. Xiong, H. Lei, S.-H.J.J.o.E.S. Deng, Technology, Hyperparameter optimization for machine learning models based on Bayesian optimization, 17(1) (2019) 26-40.
8. P. Mandal, D. Dey, B.J.J.o.O. Roy, Indoor lighting optimization: A comparative study between grid search optimization and particle swarm optimization, 48 (2019) 429-441.
9. J. Qiao, G. Wang, Z. Yang, X. Luo, J. Chen, K. Li, P.J.S.R. Liu, A hybrid particle swarm optimization algorithm for solving engineering problem, 14(1) (2024) 8357.
10. F. Guzzetti, S. Peruccacci, M. Rossi, C.P.J.L. Stark, The rainfall intensity–duration control of shallow landslides and debris flows: an update, 5 (2008) 3-17.
11. E. Stockton, B. Leshchinsky, Y. Xie, M.J. Olsen, D.J.T.I.G. Leshchinsky, Limit equilibrium stability analysis of layered slopes: A generalized approach, 5 (2018) 366-378.