An Interpretable Probabilistic Assessment Method for Slope Stability Based on PSO-LightGBMLSS-SHAP

Jinbo Zhang[1a], Yuanhao Chen[1b], Yan Yan[1c], Xiaofei Ruan[1d], Xiaoxiang Chen[2e*], Lexuan Cao[1f], Jianxin Gong[1g]

[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]462078820@qq.com, [c]605759602@qq.com, [d]xiaofei_ruan@163.com, [e*]chenxx919@126.com, [f]546214990@qq.com, [g]406480705@qq.com
[*]Corresponding author

Abstract. This study introduces a probabilistic and interpretable method for assessing slope stability using Particle Swarm Optimization (PSO), LightGBMLSS, and Shapley Additive Explanations (SHAP). The method models the probability distribution of the slope safety factor through the LightGBMLSS model, with hyperparameters optimized by PSO, and SHAP used for result interpretation. By quantifying uncertainty in slope stability, it offers reliable support for slope prevention and risk management. Validation on 100 cases shows high-precision probabilistic predictions with good interpretability, highlighting its effectiveness for practical use.

Keywords: Slope stability, Particle Swarm Optimization (PSO), LightGBMLSS, SHAP, Interpretability, Probabilistic prediction

1 Introduction

Slope stability is a critical issue in various sectors such as mining, water conservancy, construction, and transportation [1]. Its assessment is essential for engineering risk management and the development of preventive measures [2]. Currently, slope stability assessment methods include methods based on physical rules, and methods based on statistical theory or machine learning [3,4,5]. The physical rule method provides a solid theoretical foundation and is suitable for quantitative analysis on a small scale, but may be overly simplified when facing complex environments [6]. Statistical methods can reveal trends and relationships in data, but may not fully capture physical mechanisms [7]. Compared with the previous two, the methods primarily based on statistical principles or machine learning have gained wide recognition in slope stability assessment and landslide prediction due to their ability to handle complex nonlinear problems and account for the influence of different factors on slope stability [3].

© The Author(s) 2025
Y. Zhang et al. (eds.), *Proceedings of the 2024 10th International Conference on Architectural, Civil and Hydraulic Engineering (ICACHE 2024)*, Advances in Engineering Research 259,
https://doi.org/10.2991/978-94-6463-658-1_88

Although machine learning methods have achieved considerable success in slope stability evaluation and prediction, they still have the following limitations: (1) Inability to accurately reflect and quantify the inherent randomness in slope stability evaluation. The slope system is a highly complex nonlinear system, influenced by numerous internal and external factors, such as rainfall, earthquakes, etc. This results in strong randomness in slope stability [8]. (2) Machine learning is a black-box model, which makes it challenging to objectively reflect and quantify the contribution of different influencing factors to slope stability, resulting in a lack of interpretability.

This study aims to develop a probabilistic and interpretable slope stability assessment framework that combines the strengths of PSO, LightGBMLSS, and SHAP. PSO is employed to optimize the hyperparameters of the LightGBMLSS model, which is expected to obtain high-precision slope stability prediction results. LightGBMLSS, which integrates LightGBM with Generalized Additive Models for Location, Scale, and Shape (GAMLSS), allows for effective probabilistic modeling of the factor of safety (FoS) of the slopes, which can accurately capture the inherent uncertainty in slope systems to handle the problem of inherent randomness. Finally, SHAP is used to explain the model's predictions and provide insights into the contribution of individual features to the overall slope stability assessment, which solves the problem of black box models. The proposed method not only provides accurate probabilistic predictions of slope stability but also offers interpretable insights that can be used by engineers and decision-makers for effective slope prevention design and risk management.

2 Methodology

The proposed methodology consists of four main components: data collection and processing, LightGBMLSS model construction, PSO optimization, and SHAP interpretability analysis. The flowchart of the proposed method is shown in Fig.1:

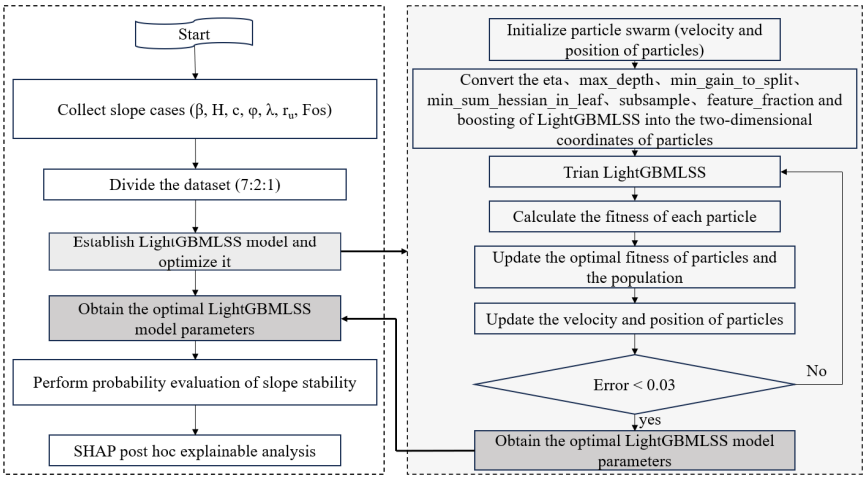

Fig. 1. PSO-LightGBMLSS-SHAP framework

2.1 Data Collection and Processing

In this study, slope data samples were collected, including factors such as unit weight, cohesion, internal friction angle, slope angle, slope height, and pore pressure ratio. The target variable was the FoS. The stability of slopes will decrease with the increase in slope height and angle [9]. The cohesion and internal friction angle of geotechnical materials significantly influence slope stability [7]. The shear strength reduction and inversion techniques can repeatedly adjust cohesion and internal friction angle to obtain a series of experimental safety factors, thereby determining the actual stability state of the slope [10]. In engineering practice, water is the main factor inducing slope instability, so pore pressure is selected as an external triggering factor [11]. To prepare for model training, the collected data were normalized as Eq.(1), then divided into training, validation, and testing datasets in a ratio of 7:2:1.

$$X'_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{1}$$

This article uses PSO algorithm to optimize the hyperparameters of LightGBMLSS model. PSO uses swarm intelligence for search, simulating the motion of particles in the search space, and can find optimized solutions in fewer iterations [12]. Grid search tests all possible combinations one by one in the parameter space, which is computationally expensive, especially in high-dimensional spaces where search efficiency is low [13]. Compared with grid search, PSO demonstrates better global search capability in various hyperparameter optimization tasks, effectively avoiding local optima and being more flexible in high-dimensional parameter space [14].

2.2 LightGBMLSS Model Construction

The LightGBMLSS model was utilized to effectively model the probability distribution of the target variable, specifically the FoS of slopes. The integration of LightGBM and GAMLSS enables the modeling of uncertainty inherent in the slope stability assessment, thereby facilitating probabilistic regression and prediction.

The LightGBMLSS model comprises two core components: (1) determining the appropriate likelihood or loss function, and (2) identifying the optimal distribution of the target variable. The log-likelihood/loss function serves to evaluate the prediction errors at each iteration, with measures to penalize the addition of tree leaves to avoid overfitting. The regularization parameter and gradients (first-order derivatives) and Hessians (second-order derivatives) are utilized to update the best-fit distribution for the target variable. The model's core objective is to model the distribution of the target variable, represented as:

$$y_i \sim D\begin{pmatrix} h_1(\theta_{\mathrm{i}1}) = \delta_{i1} \\ h_2(\theta_{\mathrm{i}2}) = \delta_{i2} \\ \vdots \\ h_{\mathrm{K}}(\theta_{\mathrm{iK}}) = \delta_{iK} \end{pmatrix} \tag{2}$$

where y_i represents the target variable, D is the distribution, θ_K represents the distribution parameters, and $h_K(\cdot)$ denotes the relationship between the distribution parameters and the predictors, defined as:

$$\delta_K = f_k(x) \tag{3}$$

The predictor function f_k is represented by the LightGBM model. Since the actual distribution of FoS is unknown due to the lack of sufficient data, a mixture distribution approach is adopted to approximate it. The mixture distribution is expressed as follows:

$$d\left(y_i|\theta_i(x_i)\right) = \sum_{m=1}^{M} w_{i,m}(x_i) \cdot d_m\left(y_i|\theta_{i,m}(x_i)\right) \tag{4}$$

where $d(\cdot)$ represents the mixture distribution of the i-th observation, $d_m(\cdot)$ represents the m-th sub-distributions, and $w_{i,m}(\cdot)$ represents the weight assigned to each sub-distribution. The sub-distributions can be simple distributions such as Normal or StudentT distributions, allowing the model to account for the complex uncertainties and variabilities inherent in slope stability.

2.3 SHAP Interpretability Analysis

SHapley Additive exPlanations (SHAP) were employed to perform an interpretability analysis of the LightGBMLSS model's predictions, quantifying the contribution of different input factors to slope stability. The method for calculating SHAP values is also referred to as an additive feature attribution method, which satisfies the additivity of feature contributions. In this method, the predictive model f is treated as an interpretable model g, expressed as:

$$f(x) = g\left(z^{`}\right) = \phi_0 + \sum_{j=1}^{M} \phi_j z_j^{`} \tag{5}$$

where $f(x)$ represents the model's predicted value, M is the dimensionality of the input features, $\phi_j \in R$represents the SHAP value of each feature, $\phi_0$is the initial expected value, and $z' \in \{0,1\}^M$ represents whether the corresponding feature is observed. The SHAP value for a specific feature j in a given instance is computed as follows:

$$\phi_j = \sum_{S \subseteq F \backslash \{x_j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}\left(f_x\left(S \cup \{x_j\}\right) - f_x(S)\right) \tag{6}$$

where F represents the set of all features, S is a subset of F, $f_x(S)$ and $f_x\left(S \cup \{x_j\}\right)$ are the mean values computed based on S and $S \cup \{x_j\}$, $\frac{|S|!(D-|S|-1)!}{D!}$ is those including x_j and those excluding x_j as the difference set, respectively.

2.4 Performance Evaluation Metrics

The point prediction accuracy and probability prediction accuracy of the framework proposed in this article can be evaluated using four indicators: average prediction inter-

val width (MPIW), prediction interval coverage probability (PICP), coefficient of determination (R ^ 2), and root mean square error (RMSE). The formulas for calculating MPIW, PICP, R ^ 2, and RMSE are as follows:

$$\text{MPIW} = \frac{1}{n}\sum_{i=1}^{n}[UP_i^{\alpha} - LO_i^{\alpha}] \tag{7}$$

$$\text{PICP} = \frac{1}{n}\sum_{i=1}^{n} C_i \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(z_i - z^*{}_i)^2}{\sum_{i=1}^{n}(z_i - \bar{z}_i)^2} \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i - z^*{}_i)^2} \tag{10}$$

$$C_i = \begin{cases} 1 & z^*{}_i \in [LO_i^{\alpha}, UP_i^{\alpha}] \\ 0 & z^*{}_i \notin [LO_i^{\alpha}, UP_i^{\alpha}] \end{cases} \tag{11}$$

Where $z^*{}_i$ denotes the predicted value, z_i and $\bar{z}_i$ represent the actual monitoring value and its mean, respectively. n indicates the sample size of the testing dataset, while LO_i^{α} and UP_i^{α} signify the lower and upper bounds of the prediction interval for the i-th slope case at the confidence level of α; C_t is a Boolean variable indicating the number of slope cases that fall within the prediction interval $[LO_t^{\alpha}, UP_t^{\alpha}]$.

3 Case Study

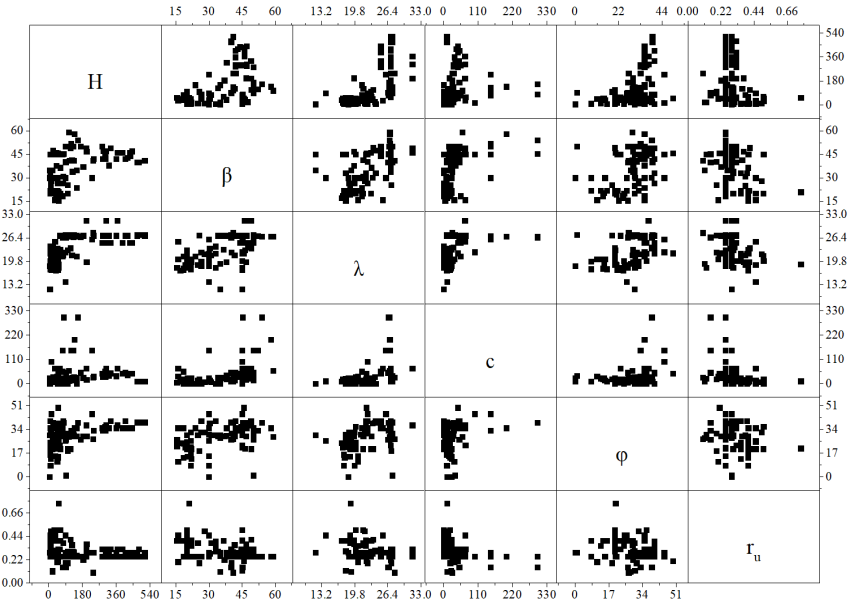

Fig. 2. Slope cases

To demonstrate the advantages of the proposed method, totally 84 slope stability cases are collected and analyzed by the proposed method (Fig. 2). The case contains six factors affecting slope stability: unit weight, cohesion, internal friction angle, slope angle, slope height, and pore pressure ratio, and the target variable is the FoS of slopes. Based on the FoS in the training set, the final mixed distribution and parameters of the stability coefficients in this study are shown in Fig.2. It can be observed that the majority of slope properties follow a log-normal distribution, with no or few outliers in the slope characteristics, indicating that the dataset construction is reasonable.

The hyperparameters of the LightGBMLSS model to be optimized include eta, max_depth, min_gain_to_split, min_sum_hessian_in_leaf, subsample, feature_fraction, and boosting. The initial value ranges for these parameters are [1e-5, 1], [1, 200], [1e-8, 20], [1e-8, 500], [0.2, 1], [0.2, 1], and [categorical, gbdt], respectively. Before performing the PSO-LightGBMLSS, it is necessary to determine the population size, position range, and velocity range, initialize the position and velocity of the particles, set the maximum number of iterations, and initialize the learning factors. Then, the fitness values are compared to update the optimal solution vector of individual particles and the global optimal solution vector of the particle swarm, thereby obtaining the optimal parameters of the LightGBMLSS model. The optimal LightGBMLSS model was used to perform probabilistic predictions of slope stability. The results are shown in Figs. 3 and 4.

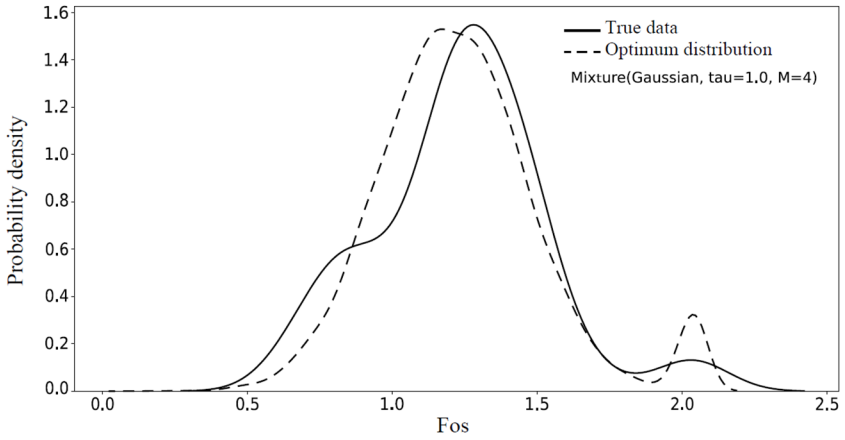

Fig. 3. Optimal estimation for the mixed distribution of the FoS for the case slopes

As shown in the Fig.5, the slope stability assessment results based on the PSO-LightGBMLSS algorithm are generally consistent with the actual FoS, with an average absolute error of 0.125 in the test set, indicating that the PSO-LightGBMLSS model can be practically used for slope stability prediction. For overall prediction, the predicted probability density distributions of all cases in the test set are all close to a Gaussian distribution, with the actual FoS of slopes falling within the 0.05 to 0.95 quantile range of the predicted distribution. In addition, the calculated MPIW value is 0.51 and PICP value is 0.9, indicating that the model has good overall prediction accuracy and

can provide reliable predictions. The results are symmetrically distributed around the predicted mean, with some individual predictions deviating farther from the mean, which is consistent with real-world scenarios. In addition, the calculated RMSE value is 0.151 and the R ^ 2 value is 0.816. This demonstrates that the PSO-LightGBMLSS model can produce high-quality probabilistic predictions, accurately quantifying the inherent uncertainty and randomness in slope stability, and solve the problems of certainty and interpretability in slope stability prediction.

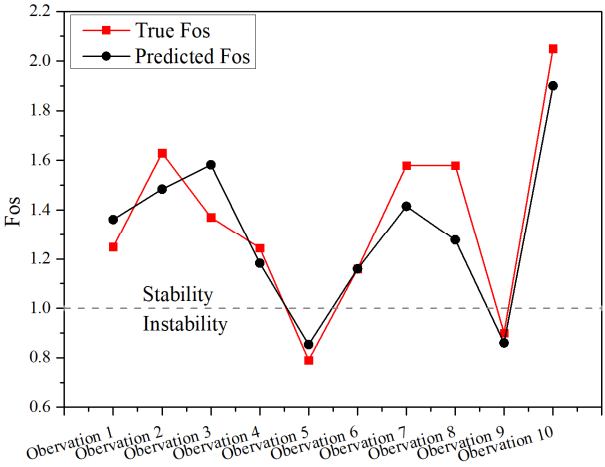

Fig. 4. Test set slope stability prediction results

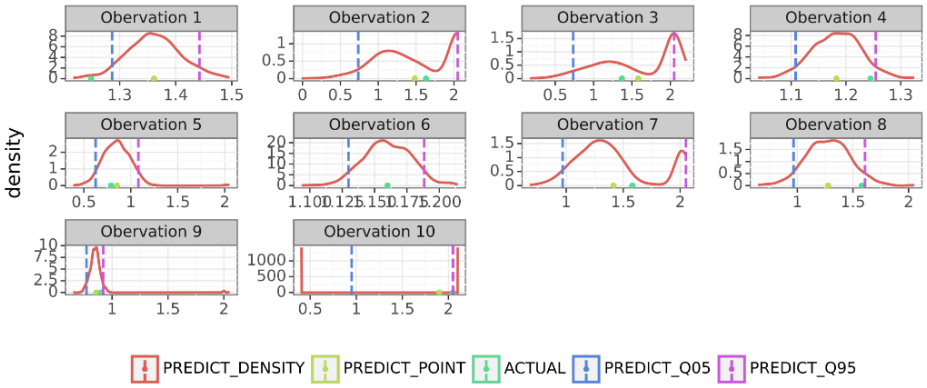

Fig. 5. Probabilistic Prediction Results

Based on the SHAP technique, the contributions of different input factors to the slope stability coefficient were calculated and ranked by feature importance, as shown in Fig. 6. It can be found that slope angle and slope height have the most significant impact on slope stability, followed by cohesion, while unit weight, internal friction angle, and pore pressure ratio have the least impact. Notably, the contribution of pore pressure ratio to the model output is zero, which is related to the specific values of pore

pressure in the selected cases. Among the 84 selected cases, 57 slope cases have a pore pressure ratio of 0.25, accounting for more than 67.8% of the total, which is the primary reason for the zero contribution of the pore pressure ratio. However, in practical engineering, the impact of water on slopes cannot be ignored. When more diverse data is available, the contribution of water to slope stability is expected to change, providing a more accurate representation. So in the later stage, more cases considering the pore water pressure ratio should be introduced to improve the interpretability of the model.

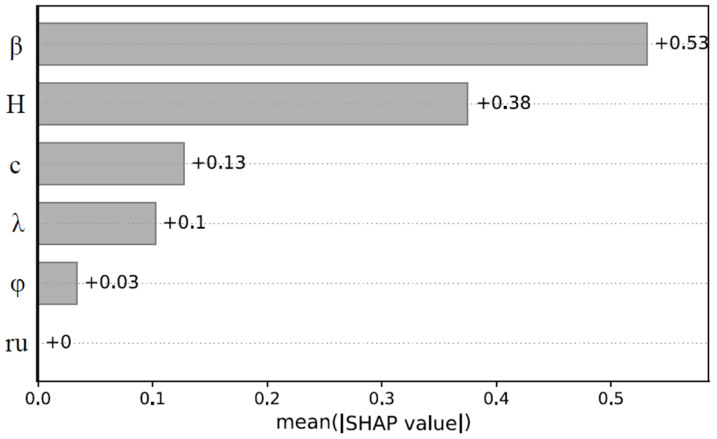

Fig. 6. Importance ranking chart of influencing factors on slope stability

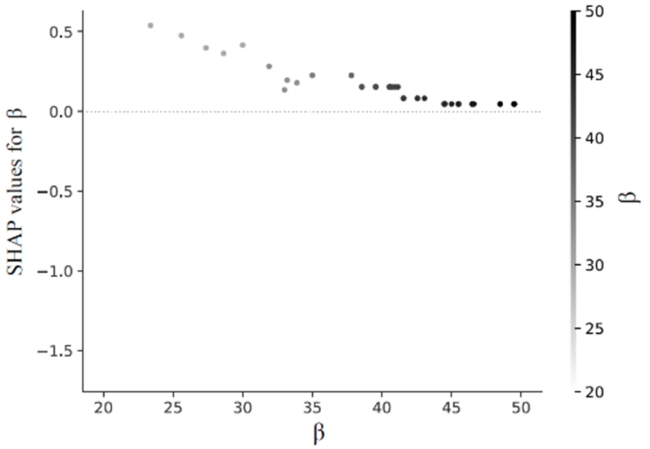

Fig. 7. The first important factor - SHAP dependence plot for slope angle

Additionally, to analyze the detailed contribution of different input features to the model prediction results, SHAP also provides feature dependence plots, as shown in Fig. 7 and 8. It can be observed that as slope angle and slope height increase, the SHAP values of the model decrease, indicating that the greater the slope angle and height, the

lower the slope stability, which is consistent with the actual physical laws governing slope stability. This further demonstrates the accuracy and effectiveness of this study.

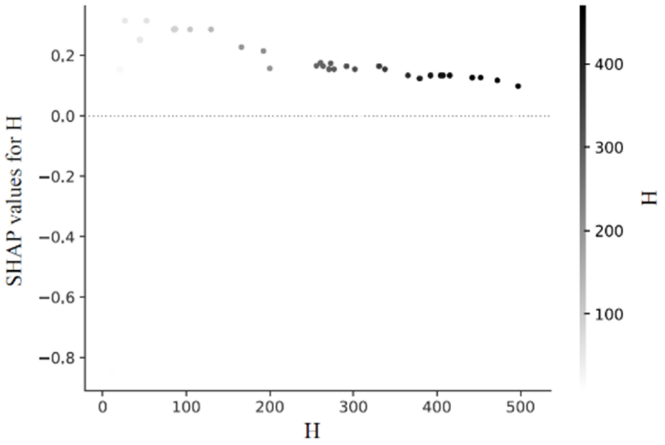

Fig. 8. The second most important factor - SHAP dependence plot for slope height

4 Conclusion

A probabilistic and interpretable slope stability prediction method based on PSO-LightGBMLSS-SHAP is proposed. The LightGBMLSS model is responsible for modeling the distribution of the slope stability coefficient to achieve probabilistic predictions, while the SHAP technique enhances the interpretability of the prediction results, thereby improving the model's transparency and reliability. Training and predictions using 84 slope cases indicate that the PSO-LightGBMLSS-SHAP model can achieve high-precision probabilistic predictions of slope stability and provide reasonable interpretations of the model outputs, thus offering richer decision-making information for slope safety construction and risk management.

Furthermore, the proposed PSO-LightGBMLSS-SHAP method leverages PSO primarily for optimizing the hyperparameters of the LightGBMLSS model, but it can also be replaced by other optimization algorithms, such as grid search or Grey Wolf Optimizer, within the framework of this study.

References

1. Karir, D., Ray, A., Kumar Bharati, A., Chaturvedi, U., Rai, R., and Khandelwal, M. (2022). "Stability prediction of a natural and man-made slope using various machine learning algorithms." Transportation Geotechnics, 34, 100745.
2. Liu, X., and Wang, Y. (2023). "Probabilistic hazard analysis of rainfall-induced landslides at a specific slope considering rainfall uncertainty and soil spatial variability." Computers and Geotechnics, 162, 105706.

3. Mahmoodzadeh, A., Alanazi, A., Hussein Mohammed, A., Hashim Ibrahim, H., Alqahtani, A., Alsubai, S., and Babeker Elhag, A. (2023). "Comprehensive analysis of multiple machine learning techniques for rock slope failure prediction." Journal of Rock Mechanics and Geotechnical Engineering.
4. Bo, Y., Liu, Q., Huang, X., and Pan, Y. (2022). "Real-time hard-rock tunnel prediction model for rock mass classification using CatBoost integrated with Sequential Model-Based Optimization." Tunnelling and Underground Space Technology, 124, 104448.
5. Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P. (2008). "The rainfall intensity–duration control of shallow landslides and debris flows: an update." Landslides, 5(1), 3-17.
6. Stockton, E., Leshchinsky, B., Xie, Y., Olsen, M. J., and Leshchinsky, D. (2018). "Limit Equilibrium Stability Analysis of Layered Slopes: a Generalized Approach." Transportation Infrastructure Geotechnology, 5(4), 366-378.
7. Jiang, S.-H., Huang, J., Griffiths, D. V., and Deng, Z.-P. (2022). "Advances in reliability and risk analyses of slopes in spatially variable soils: A state-of-the-art review." Computers and Geotechnics, 141, 104498.
8. Abbasov, R., Fahs, M., Younes, A., Nowamooz, H., Jørgen Måløy, K., and Toussaint, R. (2024). "Modeling rainfall-induced landslide using the concept of local factor of safety: Uncertainty propagation and sensitivity analysis." Computers and Geotechnics, 167, 106102.
9. Chatterjee, D., and Murali Krishna, A. (2019). "Effect of Slope Angle on the Stability of a Slope Under Rainfall Infiltration." Indian Geotechnical Journal, 49(6), 708-717.
10. Matsui, T., and San, K.-C. (1992). "Finite Element Slope Stability Analysis by Shear Strength Reduction Technique." Soils and Foundations, 32(1), 59-70.
11. Song, X., and Tan, Y. (2021). "Experimental investigation on the influences of rainfall patterns on instability of sandy slopes." Environmental Earth Sciences, 80(24), 803.
12. Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., and Deng, S.-H. (2019). "Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb." Journal of Electronic Science and Technology, 17(1), 26-40.
13. Mandal, P., Dey, D., and Roy, B. (2019). "Indoor lighting optimization: a comparative study between grid search optimization and particle swarm optimization." Journal of Optics, 48(3), 429-441.
14. Qiao, J., Wang, G., Yang, Z., Luo, X., Chen, J., Li, K., and Liu, P. (2024). "A hybrid particle swarm optimization algorithm for solving engineering problem." Scientific Reports, 14(1), 8357.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.