Forecasting Factor of Safety of Slopes Stability Using Several Machine Learning Techniques

Arsalan Mahmoodzadeh (✉ arsalan.mahmoodzadeh@uoh.edu.iq)
University of Halabja

Mokhtar Mohammadi
Lebanese French University

Research Article

Keywords: Slope stability, Factor on safety, Machine learning, PLAXIS, Feature selection.

Posted Date: August 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-669928/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Forecasting Factor of Safety of Slopes Stability using Several Machine Learning Techniques

Arsalan Mahmoodzadeh*, Mokhtar Mohammadi

Department of Building and Construction Engineering, University of Halabja, Halabja, Kurdistan Region, Iraq
Department of Information Technology, Lebanese French University, Erbil, Iraq

Email: arsalan.mahmoodzadeh@uoh.edu.iq
Email: mukhtar@lfu.edu.krd

*Corresponding author (A. Mahmoodzadeh)

Abstract

Because of the disasters associated with slope failure, the analysis and forecasting of slope stability for geotechnical engineers are crucial. In this work, in order to forecast the factor of safety (FOS) of the slopes, six machine learning (ML) techniques of Gaussian process regression (GPR), support vector regression (SVR), decision trees (DT), long-short term memory (LSTM), deep neural networks (DNN), and K-nearest neighbors (KNN) were performed. A total of 327 slope cases in Iran with various geometric and shear strength parameters analyzed by PLAXIS software to evaluate their FOS, were employed in the models. The K-fold (K=5) cross-validation (CV) method was applied to evaluate the performance of models' prediction. Finally, all the models produced acceptable results and almost close to each other. However, the GPR model with $R^2 = 0.8139$, RMSE = 0.160893, and MAPE = 7.209772%, was the most accurate model to predict slope stability. Also, the backward selection method was applied to evaluate the contribution of each parameter in the prediction problem. The results showed that all the features considered in this study have significant contributions to slope stability. However, features $\phi$ (friction angle) and $\gamma$ (unit weight) were the most effective and least effective parameters on slope stability, respectively.

Keywords: Slope stability, Factor on safety, Machine learning, PLAXIS, Feature selection.

1. Introduction

In geotechnical engineering, the analysis of slope stability is of importance. Economics has developed over time and, with that progress, the number of slopes is growing. As a result, landslides affected by slope instability have been among the three most catastrophic natural disasters in the world, besides volcanoes and earthquakes. In order to minimize or avoid landslide failure, it is significant to understand the importance of slope stability analysis. It is complicated to anticipate the stability of slopes due to the complexity of them because they are not dependent on specified data or a single parameter, but they are mostly dependent on a variety of variables, such as slope angle, slope height, soil properties, and seasonally changing soil water content (Sakellariou and Ferentinou 2005; Duncan 2000; Suman et al. 2016).

While there is still much to be done to predict and evaluate slope stability, considerable research has been performed across several parts of the world. Several useful models have been implemented in the previous studies by different researchers to analyze and predict slopes' failures (Lin et al. 2018). These approaches can be categorized into four groups including qualitative analysis (Bye and Bell 2001; Trivedi et al. 2012), quantitative analysis (Thiebes et al. 2013; Thiebes et al. 2014), physical simulation (He et al. 2012; Kim et al. 2016; Sarkar et al. 2012), and field monitoring analysis (Pirone et al. 2015). These researchers have done a remarkable job of tackling slope stability. However, the
problem is that the techniques they built are not optimal and only give specific solutions for certain slopes. These models are not suitable to solve all slope stability problems completely because of several drawbacks (Lin et al. 2018):

a) The methods used are not really appropriate because of the shortcomings. For instance, the limit equilibrium method cannot consider the constitutive relationship of soil, and it needs to search for the most dangerous sliding surface. Besides, the analysis is complicated for multi-layer soils. Therefore, this approach does not accurately elucidate the proper safety and reliability during the stability slopes analysis.

b) The safety factor is not clearly defined. In general, slopes with a factor of safety (FOS) of greater than 1.20 are safe. In spite of this, there have been failed slopes with the FOS greater than 1.2 in real engineering.

c) These methods are affected by human subjective factors.

In recent years, artificial intelligence and soft computing models have been broadly applied in geotechnical applications, e.g., tunnel resources (Mahmoodzadeh and Zare 2016; Mahmoodzadeh et al. 2019; Mahmoodzadeh et al. 2020a; Mahmoodzadeh et al. 2020b; Mahmoodzadeh et al. 2021a), mechanical property (Liu et al. 2013; Liu et al. 2015), tunnel surface settlement (Mahmoodzadeh et al. 2020c), landslide displacement (Youssef et al. 2016), rock bursts (Zhou et al. 2016), tunnel geomechanical properties (Mahmoodzadeh et al. 2021b; Mahmoodzadeh et al. 2021c), sidewall displacement of underground caverns (Mahmoodzadeh et al. 2021d).

In the meantime, several supervised learning approaches have recently been available for slope stability computation. These methods have shown remarkable results (Lin et al., 2018). For instance, Lu and Rosenbaum (2003) were able to merge artificial neural networks and the gray system approach for analyzing the stability of slopes based on the acquired situations’ geotechnical characteristics and historical behavior. Artificial neural networks (ANN) were implemented by Wang et al. (2005) and Das et al. (2011) in order for the stability analyses of slopes. Zhao et al. (2012) investigated the correlation between slope stability and its effect parameters using the relevance vector machine. Liu et al. (2014) evaluated 97 cases using the extreme learning machine method to assess slope stability. Furthermore, through 168 situations, Hoang and Pham (2016) have implemented metaheuristic optimize least squares support vector regression for slope analysis.

Manouchehrian et al. (2014) developed a genetic algorithm to attempt to anticipate the FOS for slopes and have shown that this model is reasonably straightforward and efficient. Xue et al. (2014) suggested particle swarm optimization (PSO) for picking the necessary factors for the support vector machines (SVM) model. The outcomes revealed that the combination of SVM and PSO models is an effective analysis method for predicting slopes' stability. Li et al. (2013) and Samui (2008) similarly evaluated slopes' stability based on SVM and achieved valuable outcomes. In addition, Li et al. (2013) and Samui (2008) have projected slope stability using SVM and gained great outcomes. Wen and Zhang (2014) gathered data from various situations, and 30 of them utilized the testing dataset for training a Random Forest (RF) model that predicts slope stability. The remaining 12 data sets were used to assess the qualified prediction models' efficiency and noticed that testing the model enhanced its efficiency. Khajehzadeh et al. (2012) and Raihan et al. (2013) used a gravitational search algorithm (GSA) to solve global optimization problems and minimize the FOS in the study of slope stability. Shi et al. (2010) trained the Bayes model with the gathered data of slope situations and observed that the estimation outcomes were precisely the same as the real situations. The model may be used to accurately evaluate slope stability, achieving high predictive performance and low misclassification with a low error rate.

The mentioned intelligent models explain slopes’ failures, but they are not adequate to solve them entirely. A particular technique could be suitable for some situations but not appropriate for others, and no distinction between intelligent modeling approaches has been observed in the prediction of slope stabilities. Therefore, the models’
credibility and precision are not well established. Furthermore, cross-validation (CV) is an efficient and accurate approach to confirm the models' prediction performance (Lin et al. 2018). It is crucial to recommend CV for accuracy verification of optimized models.

In this work, six machine learning (ML) methods were used to forecast the slopes' FOS: Gaussian process regression (GPR), support vector regression (SVR), decision trees (DT), long-short term memory (LSTM), deep neural networks (DNN), and K-nearest neighbors (KNN). PLAXIS software is used to measure the FOS of 327 slope cases in Iran, using input parameters such as unit weight ($\gamma$), cohesion ($c$), friction angle ($\phi$), slope angle ($\alpha$), slope height ($H$), and pore pressure ratio ($r_u$). These datasets are employed in the prediction models. The prediction output of the models is evaluated using the K-fold (K=5) CV process.

The analyzed FOSs using PLAXIS software are compared to the forecasting values made by the ML models. Finally, the best prediction model has been specified. Also, in order to show the effect of each of the input parameters on the SOF of a slope, the backward selection method is used. In this way, the most important parameters affecting the stability of slopes are identified. Fig. 1 presents the complete flowchart of the analysis.

2. Database and statistical analysis

In Fig. 2, a slope's geometry and the effective parameters on slopes' stability are shown. It should be noted that, in this study, the external factors generating the failure, such as earthquake effects and other human-made factors, have not been considered.

In Table 1, various studies were performed to predict the circular mode of slope stability utilizing soft computational methods are presented. In most of these studies, six factors of $\gamma$, $c$, $\phi$, $\alpha$, $H$, and $r_u$ have been considered effective slope stability factors. Therefore, according to the literature and data availability, these six factors are considered in this study.

In order to evaluate the efficiency of the prediction models applied in this work for slope stability, a database including 344 slope cases analyzed for circular failure mechanisms was employed. Circular failure is a prevalent type of failure of slopes in soil and more fractured rock areas. To determine the value of FOS, all the slopes were studied using the finite element software of PLAXIS. These analytical outcomes are compared with the prediction results of the ML models.

In this study, data analyses of bivariate correlation and basic descriptive statics are used so that a deeper understanding of the database employed can be achieved. From the statistical point of view, when the training dataset factors are in larger variance ranges, a more applied predictive model can be developed. Also, in the developed model, the dependent variable(s) are generally predicted by interpolating the dataset. Thus, it is crucial to define an effective range of variance for the independent variable(s). Hence, to define the variance set of the parameters, simple descriptive statistical analysis is performed on the database (Manouchehrian et al. 2014). Such a common statistical tool as boxplot is used for presenting groups of datasets via their five-number summaries containing: the lower whisker, lower quartile, median, upper quartile, and upper whisker. This boxplot can also detect any available outlier in the dataset. For each parameter in the initial dataset, the boxplots are presented in Fig. 3. For each correlation, the tops of the box (25 percentage) and the bottoms of the box (75 percentage) are elucidated. There are interquartile ranges which are the spaces between the bottoms and tops. The sample median is the line that runs through the center of each box. For three boxplots of $\gamma$ and $c$, the median is roughly in the middle of the box, while the others are not, indicating sample skewness.

The whiskers are defined by the continuation of lines above and below the boxes. The whiskers are spread from the inter-quartile range's ends to the whisker range's farthest data. Referring to Fig. 3, it can be observed that the length
linking the higher whisker and the higher quartile differs from the distance between the lower whisker and the lower quartile for H, c, and r factors; however, these distances are roughly the same for γ, φ, and α factors. In addition, it also illustrated the allocation of the data. Beyond the whisker length, there are outliers, which are the other observations. Several outlying findings are presented in the H and c parameters, as depicted in Fig. 3. To minimize the chance of misleading for statistical interpretations derived from datasets including outliers, for much more investigations, 55 outliers need to be deducted from the dataset. The medians of the c and the γ are around 10 kN/m² and 20 kN/m³ respectively. As per the study of Liu et al. (2014), Zhang et al. (2014), and Lin et al. (2018), It could be acquired that these slopes are mostly soil slopes or heavily fractured rock. The very large values of cohesion, which are outlier cases, are perhaps for rock slopes. The outcome of Zhang et al. (2014) study explained that the improvements in models’ performance could be achieved with the removal of these cases. Therefore, these outliers (17 slope cases) should be eliminated, and 327 slope cases remained. Table 2 shows the outcomes of the simple descriptive statistical analysis achieved of the reducing datasets.

For the 327 datasets considered in this study, there are two slope stability modes of failure (168 cases) and stable (159 cases). The two modes of slope stability cases employed in this research are relatively balanced, as shown in Fig. 4.

In the soft computing problems, to assess their general efficiency, prediction models’ performance is necessary to be verified on new datasets. Therefore, the entire dataset could be subdivided into two training and testing sets. The prediction model is trained, and the hyper-parameters are tuned using the training datasets. The testing datasets are utilized separately to measure the overall efficiency of the prediction of models.

This article applies the K-fold CV (K=5) method to classify the database into two groups: training and testing. The work method is such that whole datasets are divided into five equal categories so that each time one category is used as training sets, and the other four categories are used as testing sets. In the K-fold CV method, all data will be experienced in both training and testing sets. Therefore, using this method will give more credibility to the predicted results. Furthermore, all the datasets were normalized before modeling in the ranges [-1, 1] in order to eliminate variations in input and output parameters. However, standardization of the dataset is not efficient, and the results are unfavorable due to an excessive distribution of the data.

Some statistical indices, such as coefficient of determination ($R^2$), root mean square error (RMSE) and mean absolute percentage error (MAPE), are used to determine the models’ precision. The following formulas are used for calculating these indices (Eqs. 1-4).

$$R^2 = 1 - \frac{\text{sum squared regression (SSR)}}{\text{sum of squares total (SST)}} \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - y'_i| \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y'_i)^2} \quad (3)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y'_i}{y_i} \right| \times 100\% \quad (4)$$
where $y_i$ is the real value, $y'_i$ is the forecasted value, $\bar{y}_i$ is the mean of $y_i$, $\bar{y}'_i$ is the mean of $y'_i$, and $n$ is the number of test datasets.

### 3. ML prediction models of slope stability and results analysis

Despite the merits of various ML approaches, according to the No-Free-Lunch (NFL) theorem, there is no ML model to solve all engineering problems as the best method successfully. Therefore, researchers have tried to evaluate the efficiency of various ML approaches for solving various optimization. (LSTM), (DNN), (KNN), (GPR), (SVR), and (DT) are six separate ML and deep learning (DL) models with different features and capabilities that we use as NFL theorem. However, the key features of the aforementioned models, which motivate us to use them, is as follows:

- **Regression Analysis**

  Regression analysis is a statistical technique for predicting a continuous target variable from one or more independent variables. Rather than variables that have been modified by experimentation, regression analysis is typically used for naturally occurring variables. There are several different forms of regression, so how do we select the one to use once we have chosen to use regression analysis?

  **We chose GPR because:**

  - GPR directly captures the model uncertainty. In regression, for example, GPR provides a distribution for the prediction value rather than a single value as the prediction. In neural networks, this ambiguity is not explicitly captured.

  - By choosing different kernel functions, we can add prior knowledge and requirements about the shape of the model while using GPR. For example, based on the answers to the following questions, we may choose different priors. Is the model smooth? Is it sparse? Should it be able to change drastically? Should it be differentiable? This capability gives researchers flexible models, which can be fitted to various kinds of datasets.

  **We chose SVR because:**

  The use of kernels, sparse solutions, and Vapnik-Chervonenkis (VC) regulation of the margin and number of support vectors distinguish SVR. One of the key benefits of SVR is that its computational complexity is independent of the input space's dimensionality. It performs lower computation compared to other regression techniques. It also has good generalization, high prediction precision, and is resistant to outliers.

  **We chose DT because:**

  DTs are a form of supervision learning algorithm that partitions the sample multiple times based on specific questions about the sample. For prediction problems, these can play a significant role. They are relatively simple to comprehend and very operative. DTs reflect a series of decisions with varying chances of occurring. Through the use of this technique, the identification of the most substantial variables and the relation between two or more variables can be provided. In our problem, we have various variables, which are related to each other, so we select DT as one of comparing models. In other words, a decision tree has the advantage of forcing the analysis of all possible outcomes and tracing each direction to a conclusion. It performs a thorough analysis of the implications of each branch and defines the required decision nodes.

  **Key advantages:**

  - There is no need to preprocess the data.
  - Data distribution assumptions are not required.
  - It can effectively handle collinearity.
- The prediction in a comprehensible manner can be obtained in DT.

- **We chose KNN because:**

  In this paper, KNN were used as a non-parametric approach for prediction. It is one of the easiest ML approaches that have been recently applied. It is a model of slow learning with local approximation. We use this model, considering the following terms:

  The key Advantages:

  - It is a straightforward machine learning model and easy to use.
  - It can tune only a few hyperparameters.

  **Hyperparameters:**

  - There are two main hyperparameters in KNN: K value & distance function.
  - K value: the number of neighbors to include in the KNN algorithm. The value of k should be adjusted according to the validation error.
  - Distance function: The most commonly used similarity function is Euclidean distance. There are several alternatives including the Manhattan distance, Hamming distance, and Minkowski distance.

  **Assumptions:**

  - A clear understanding of the input domain is required.
  - A sample size should feasibly be moderate (due to space and time constraints).
  - Prior to training, outliers and collinearity should be tackled.

  **Comparison with other models:**

  - The main difference between KNN and other models is that KNN needs a lot of real-time computation.

- **We chose DNN because:**

  The substantial advantage of DNN over its predecessor is that it automatically identifies essential features without the need for human intervention. In this case, we have no idea about the dependency of feature; therefore, we use DNN as one of the comparators to decrease the dimensionality of datasets and see the effects of the automatic dimensionality reduction in the results.

- **We chose LSTM because:**

  The data of our problem has a time dependency feature. i.e., the variables are dependent on each other in terms of time. Therefore, we decide to use RNN-based models to consider this time dependency. When we switch from RNN to LSTM (Long Short-Term Memory), we introduce ever more control knobs that regulate the flow and mixing of inputs according to qualified Weights. As a result, LSTM provides us with the most control and, consequently, better results. However, this leads to an increase in complexity and operating costs. Consequently, in order to obtain the best accuracy, we choose LSTM as a replacement for conventional RNN.

  It is very significant to comprehend the data so that the right algorithm for the right problem can be chosen. In some algorithms, only small sample sets are required, while others can work with tons and tons of samples. Some certain algorithms desire numerical input while others work with categorial. Because our datasets are large, we propose work with the LSTM Algorithm.

In the analysis of slope stability, GPR, SVR, and DT modeling are applied using MATLAB Software 2018. The LSTM, DNN, and KNN algorithms are introduced in Anaconda version 3.6, and it a free, open-source application for scientific
computation in Python and R, which aims at simplifying package management and execution. These six predictive models have been introduced to determine the best predictive model for slope stability. In the following subsections, details on the ML prediction models and their prediction results are presented.

### 3.1 GPR

A Gaussian procedure (GP) is a gathering $F$ of arbitrary factors $F_{x_1}, F_{x_2}, \ldots$ for which any finite subset of the factors has a joint multivariate Gaussian conveyance. Components $x$ of a set $X$ is used to list the variables. For any finite length vector of lists $x = [x_1, x_2, \ldots, x_\ell]^T$, we have a comparing vector $F_x = [F_{x_1}, F_{x_2}, \ldots, F_{x_\ell}]^T$ of factors that has a multivariate Gaussian (or ordinary) distribution (Eq. 5) (Mahmoodzadeh et al., 2020a):

$$F_x \sim \mathcal{N} \{ \mu(x), k(x,x) \}$$

where the components of $\mu(x)$ are given by an earlier mean capacity $\mu(x_i)$, and $k$ is the portion work. The portion uses two files $x_i$ and $x_j$ that provides the covariance between their comparing factors $F_{x_i}$ and $F_{x_j}$. Given vectors of lists $x_i$ and $x_j$, $k$ returns the framework of covariances between all sets of factors where the first in the pair originates from $F_{x_i}$ and the second from $F_{x_j}$. Each $F_{x_i}$ is barely Gaussian, with a mean of $\mu(x_i)$ and difference of $k(x_i, x_i)$ (Mahmoodzadeh et al., 2020b).

Assume there is a capacity $f(x)$ that would want to upgrade. In addition to that, suppose that $f$ could not be watched legitimately, yet that an arbitrary variable $F_x$ can be seen that is listed by the same space as $f$ and whose normal esteem is $f$, i.e., $\forall x \in X, E[F_x] = f(x)$. Particularly, it is accepted that the earlier conviction about the capacity $f$ complies with a Gaussian procedure with earlier mean $\mu$ and part $k$. Furthermore, assume that $F_x$ is a perception of $f(x)$ that has been tainted by zero-mean, i.i.d. Gaussian clamor, i.e., $F_x = f(x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Consequently, $f(x)$ is a shrouded variable whose back appropriation and could be able to derive in the wake of watching tests of $F_x$ at different areas in the space. The following subtraction is called Gaussian procedure relapse (Mahmoodzadeh et al., 2020c).

Give $x$ a chance to be the arrangement of perceptions focuses and $F_x$ be the subsequent genuine esteemed perceptions. This required to process the back appropriation of some new point $\hat{x} \in X$. The appropriation will be Gaussian with mean and difference (Eqs. 6 and 7).

$$\mu(\hat{x} | x) = \mu(x) + k(\hat{x}, x)k(x,x)^{-1} (F_x - \mu(x))$$

$$\sigma^2(\hat{x} | x) = k(\hat{x}, \hat{x}) - k(\hat{x}, x)k(x,x)^{-1} k(x, \hat{x})$$

In the regression learner app embedded in MATLAB software 2018, four different models for GPR include squared exponential, rational quadratic, exponential, and Matern 5/2. After modeling by this program, the model type with the most accurate results is taken into account. Also, the optimization mode is considered in the app, so that the app itself optimizes the amount and type of hyper-parameters of the GPR model. The optimized type and value of the GPR hyper-parameters produced by the regression learner app are presented in Table 3.

The 5-fold CV results of slope stability predicted by the GPR model are depicted in Fig. 5. As in Fig. 5, the analyzed FOS values are very close to the FOS values predicted by the GPR model. According to Fig. 5, the FOS errors produced by the GPR model are close to zero for most datasets, so that the mean absolute error value is about 0.0842. Such an error in predicting the FOS of a slope is negligible and indicates the GPR model’s high prediction accuracy. To more support these results, other statistical evaluation indices of $R^2$ (Fig. 6), RMSE, and MAPE are obtained equal to 0.8139,
0.160892561, and 7.209771731, respectively. As the database employed in this research, all the outcomes confirm the GPR model’s high prediction accuracy.

### 3.2 SVR

Vapnik (1995) modified his first version model (ε-support vector regression, SVR) by changing the ε-insensitive loss function. This modification permits the SVR model to use the margin idea in the regression process. Margin in the modified model can be described by the summation of the distances of hyperplane from the closest points of two classes. Minimizing errors between the actual training data and the hyperplane are the main target of the SVR. Kernel function idea has introduced by Vapnik (1995) for non-linear SVR. Readers are directed to Vapnik (1995) to understand more about SVR.

To get the FOS predictions using the SVR model, in the MATLAB 2018 software, the regression learner app was embedded and applied. Six model types including cubic, linear, medium Gaussian, quadratic, fine Gaussian, and coarse Gaussian, are provided for the SVR method in MATLAB 2018. By optimizing the mode in the regression learner app, the type and values of the SVR hyper-parameters are obtained as in Table 4.

Fig. 7 shows the FOS predictions made by the SVR model. Compared to the analyzed FOS values in the PLAXIS software, there is acceptable accuracy in the predictions of the SVR model. As it turns out from Fig. 7, the FOS errors are often less than 0.2, so that the MAE value is equal to 0.142. This amount of error can be acceptable and significant in the FOS predictions of a slope. Compared to the analyzed FOS results, other statistical evaluation indices of R², RMSE, and MAPE are obtained by 0.6950 (Fig. 8), 0.20650649, and 11.89200416%, respectively. All of these results suggest that the SVR model should be considered as a somewhat appropriate predictive model for predicting slope stability.

### 3.3 DT

The DT is one of the classifications and regression methods based on the non-parametric survived learning technique. Furthermore, it consists of a set of if-then-else decision rules. The best perdition of the model occurs when the DT goes deeper and deeper to make the best fit with the actual data. There are several advantages of the DT. First, the distribution of explanatory variables does not require assumption. Second, strong relations among independent variables do not affect the DT outcomes. Third, various dependent variables such as survived data, categorical and numerical can be covered by DT. Fourth, this technique comprises the powerful variables and eliminates the least powerful variables which describe the dependent variable. For the DT, it is possible to predict small and large datasets well, even though this technique was initially developed to only well predict large data (Mahmoodzadeh et al., 2020b).

The algorithm of DT can be explained as follow:

1. First, the calculation of the targeted variance is performed.
2. Based on the various attributes, the database is divided into distinct parts, and the variance of each sectioned part is deducted from the variance prior to the division. This can be defined as variance reduction.

Node \( N \) can be defined by the variance reduction as:

\[
I_v(N) = \frac{1}{|S|^2} \sum_{i \in S} \sum_{j \in S} \frac{1}{2} (x_i - x_j)^2 - \left( \frac{1}{|S_t|^2} \sum_{i \in S_t} \sum_{j \in S_t} \frac{1}{2} (x_i - x_j)^2 + \frac{1}{|S_f|^2} \sum_{i \in S_f} \sum_{j \in S_f} \frac{1}{2} (x_i - x_j)^2 \right)
\] (8)

\( S \) is a group of samples that is not separated yet, \( S_t \) is a group of separated samples with true result and \( S_f \) is a group of separated samples with a false result. Without referring to the mean, each of the summands presented above is indeed variance estimates though written in a form. In each summation term in Equation 10, variance estimation is required in such a way the mean is not referred to directly.
The decided node of the attribute is based on the highest VR.

3. Depending on the values of selected attributes, the datasets are separated. If the variance of a part is more than zero, it is separated once more.

4. Keep another trial going until all the data is evaluated.

In the DT approach, three model, such as medium, coarse, and fine, are embedded in the MATLAB 2018. The FOS predictions were performed through these three models and eventually considered the model that provided more precise results. The information about the optimized DT’s hyper-parameter parameters considered in this analysis is provided in Table 5.

In Fig. 9, the FOS results obtained from the DT model are compared with the analyzed ones. As shown in Fig. 9, there are small differences between the predicted outcomes and the analyzed values. The MAE is equal to 0.143. Achieving such an error in predicting the FOS of a slope is good. Other statistical evaluation indices of $R^2 = 0.6830$ (Fig.10), $RMSE = 0.210590831$, and $MAPE = 12.2080868\%$ also indicate the good accuracy in the DT model predictions.

### 3.4 LSTM

Deep learning can be modeled as a typical version of neural networks with several layers. The knowledge found within these deep learning networks is better remembered than traditional neural networks. A recurrent neural network (RNN) is a form of the combination of a loop network in the neural network. Knowledge persistence exists across these loops in the networks. This network gathers inputs and data from prior networks, implements specified procedures and provides the next network output. Some applications may only need current knowledge, while others may also need additional information from the past. A learning lag exists in redundant neural networks in which the gap between the desired input and the requirement is improved. However, LSTMs networks are a special kind of RNN (Mahmoodzadeh et al., 2021b) capable of studying these situations. Such networks are intentionally designed so that, on the permanent networks, the long-term dependency problem can be mitigated.

An essential function of LSTM is recalling knowledge over a long time. It is an ordinary choice since the precision of the model permanently depends on the sum of prior knowledge. The basic LSTM module, defined as the repetitive module, has a particular interaction with four layers with neural networks, as shown in Fig. 11. In the module, there are three gates with activation functions of $\sigma_1$, $\sigma_2$, and $\sigma_3$, and it contains two output activation functions such as $\Phi_1$ and $\Phi_2$, as seen in Fig. 11. To represent the addition and multiplication in the element-wise manner, the $\Sigma$ and $\pi$ symbols are used. A bullet (•) symbol is used to describe concatenating mathematical procedures. The principal element of LSTM is the cell state, in which the current block memory ($S_{t-1}$) is created from the previous block memory. The information flow is then approved immediately down the line. In the network, the first layer ($S_t$) determines the sum of the prior knowledge to flow, and Eq. 9 demonstrates the activity carried out by this layer (Mahmoodzadeh et al., 2021c).

In the network, two layers and their associations influence the mechanism of storing new data in the cell state. A sigmoid layer ($\sigma_2$) that find out the ($i_t$) value to be updated see Eq. 10 and $\Phi_1$tanh layer, which makes a new candidate value ($S'_t$) as seen in Eq. 11. To make a case, both are to be added together with the state. The cell state is lastly replaced with Eq. 12.
\[ c_{ft} = \sigma_1(W_{cf} \cdot [O_{t-1}, x_t] + b_{cf}) \]  
\[ l_t = \sigma_2(W_l \cdot [O_{t-1}, x_t] + b_l) \]  
\[ S_t^* = \tanh(W_S \cdot [O_{t-1}, x_t] + b_S) \]  
\[ S_t = (c_{ft} \times S_{t1}) + (l_t \times S_{t-1}^*) \]  

In order to optimize the LSTM model, this model was tested with different types of parameters and changes in the number of neurons, layers, batch_size, epochs, and so on. Finally, the type/value of parameters of the LSTM were optimized as in Table 6.

The structure of the network optimized LSTM model is shown in Fig. 12.

The FOS results of LSTM are compared with the analyzed FOS results in Figs. 13 and 14. Given that deep learning methods such as LSTM require much more data to learn, it can be said that with the amount of data used in this article, still, it has provided acceptable accuracy. The MAE for the LSTM results is equal to 0.169. Achieving such an error in predicting the FOS of a slope is good. Other statistical evaluation indices of \( R^2 = 0.6616 \) (Fig.14), \( \text{RMSE} = 0.221912033 \) and \( \text{MAPE} = 14.52803317\% \) also indicate good accuracy in the LSTM model predictions.

### 3.5 DNN

In the layered neural network of DNN, a deep feed-forward neural network is available. There is more than one layer of hidden units between its inputs and outputs. The layers contain input layers, which are followed by mid-layers, hidden layers, and finally, the output layer. Consequently, the input, hidden layers, and output layers are all connected to the network’s neighboring layers. It is a feed-forward neural network, which means that there is no forward feeding and that the links between the layers are only one way, forward. Furthermore, DNN is especially well-suited to analyzing raw input data because it can recognize patterns and learn useful features from it without the need for rigorous feature engineering, data pre-processing, or hand-crafted guidelines. Moreover, with the rise in training data, its efficiency further improves. (Mahmoodzadeh et al., 2021b). DNNs have a much wide range of applications, from simple text creation to machine vision tasks, and the early uses of DNN are in automatic translation.

In order to optimize the DNN model, such as the LSTM model, it was tested with different types of parameters and changes in the number of neurons, layers, batch_size, epochs, and so on. Finally, the type/value of parameters of the DNN was optimized as in Table 7.

The network structure of the optimized LSTM model is illustrated in Fig. 15.

The predictions of FOS results of the DNN model are compared with the analyzed FOS results in Figs. 15 and 16. As the LSTM model, the DNN model also needs more data for training, but the amount of data used in this article has provided acceptable accuracy. The MAE for the DNN results is equal to 0.0.118. Achieving such an error in predicting the FOS of a slope is good and acceptable. Other statistical evaluation indices of \( R^2 = 0.7498 \) (Fig.16), \( \text{RMSE} = 0.188416005 \), and \( \text{MAPE} = 9.967075268\% \) also indicate the good accuracy in the DNN model predictions.

### 3.6 KNN

KNN is a simple and well-managed machine learning algorithm that can be applied to classification and regression problems. It is important to apply and comprehend, but it is very difficult to do as the amount of data in use expands considerably. KNN functions through the determination of the ranges between a query and all the examples in the results, choosing the specified number of examples for the query (K), then voting for the most frequent label (in the situation of classifying) or averaging the labels (in the situation of a regression). During the regression and classification process, selecting the correct K for the data is achieved by trying multiple Ks and finding the one that fits
better (Mahmoodzadeh et al., 2020c). To implement the KNN method in this study, Anaconda version 3.6 is used. During the KNN model's predictive analysis, various Ks were evaluated, and the best predictions were obtained with K = 2.

For the KNN model, as shown in Fig. 18, the difference in the predicted FOS values with the analyzed results is low. This MAE = 0.0913 indicates the high accuracy of the predictions made by the KNN model. The results of other statistical evaluation indices of $R^2 = 0.7080$ (Fig. 19), RMSE = 0.205355281, and MAPE = 7.874855748% also indicate the high precision of the KNN model in predicting the slope stability. Thus, considering the database employed in this paper, the KNN model is a significant and considerable method for slope stability prediction.

To determine the best prediction model among the six ML models used in this paper to predict the slope stability, in Fig. 20 and Table 6, a comparison between the results predicted by them has been made. By analyzing and comparing the values of the obtained statistical evaluation indices for each model, it can be concluded that almost all models have provided close results to each other, and there is not much difference between their accuracy. However, the highest accuracy is produced by the GPR model. Therefore, the most acceptable results, which are not very different from the analyzed ones, are provided by the GPR model.

4. Discussion

In this study, the generalization of the suggested Gaussian process regression methodology is discussed. Generalization is a concept used to characterize the model's ability to interact and adapt to new information. Therefore, a model can ingest novel data and predict accurately after practicing with data not used during training. The basis for a model's success and its practical performance is related to its capability to generalize. When a model was so well trained in training data, it cannot be generalized. When new data are given, the model is rendered inaccurate predictions and worthless even if it can accurately predict the training data. A model starts 'memorizing' the training data instead of 'learning'; this is known as overfitting.

Feature selection can be used to avoid the overfitting of the model. In this case, feature selection would minimize the number of features, which decreases the computational complexity of the model. The stepwise approach for choosing an important collection of features from the data sets is used for the total features available.

Stepwise regression methods can be classified into three strategies:

- **The first strategy (forward selection)**, which begins without any predictors in the model, relies on adding more iterative predictors to the model. Simultaneously, it stops when the improvement in the results no longer has a statistically positive impact.

- **The second strategy (backward selection)**, which begins with all predictors in the model, periodically eliminates the lowest contributive predictors; while it stops once you get a model, all its predictors become statistically meaningful.

- **The third strategy (stepwise selection)** is a mixture of forwarding and backward processes. It begins without any predictors and then adds the predictors that contribute most to the outcome sequentially (like backward selection). While adding every new variable, those that no more enhance the model's fit should be removed (like forwarding selection).

In this research, the step ACI [MASS Package] was used, which defines the best design by AIC. The model also has a choice called direction, which takes these values: i) forward (for elimination from forwarding); ii) backward (for elimination from backward); iii) both (sequential replacement, for forward and backward elimination). The best-finished model is recovered. In R, among the most popular search methods for selecting features is stepAIC. For the stepAIC model's values continuously to arrive at the final feature set are attempted to be reduced. In the tables below, the finding listed as follows, three asterisks (*) reflect the highly significant value of p. Therefore, it could reject the
null hypothesis by providing a small p-value for the intercept and path that enables us to create a good relationship between two measured variables (the target and the predictor variables). A p-value around 5% or less would be a good cut-off point for most situations.

In the first step, we fit the model with all predictors and targets. The first step results of feature selection are presented in Table 7. As in Table 7, the effect of all features on slope stability is significant. In this case, we do not delete any of the features, and we stop at this step. According to Table 7, the most effective parameters on slope stability with respect to the p-values are $\phi$ and $c$. Also, the parameter $\gamma$ has the least impact on slope stability.

5. Conclusions

This study proposed six ML models to predict slope stability. 327 datasets analyzed using PLAXIS software including six input parameters of $\gamma$, $c$, $\phi$, $\alpha$, $H$, and $r_u$, and one target of FOS were employed in the models. The 5-fold CV method was used to evaluate the prediction performance of the models.

All the models produced acceptable results and almost close to each other. However, the GPR model with $R^2 = 0.8139$, RMSE = 0.160893, and MAPE = 7.209772%, was the most accurate model to predict slope stability.

The backward selection method was used to assess the contribution of each parameter in the prediction problem. The results showed that all the features have significant contributions to slope stability. However, features $\phi$ and $\gamma$ were the most effective and least effective parameters on slope stability, respectively.

Conflict of interest

There is no conflict of interest.

References

Sakellariou MG, Ferentinou MD. A study of slope stability prediction using neural networks. Geotech. Geol. Eng. 23 (2005) 419-445.

Duncan JM. Factors of safety and reliability in geotechnical engineering. J. Geotech. Geoenviron. Eng. 126 (2000) 307-316.

Suman S, Khan SZ, Das SK, Chand SK. Slope stability analysis using artificial intelligence techniques. Natural Hazards 84 (2016) 727-748.

Bye AR, Bell FG. Stability assessment and slope design at Sandsloot open pit, South Africa. Int. J. Rock Mech. Mining Sci. 38 (2001) 449-466.

Trivedi R, Vishal V, Pradhan S, Singh T, Jhanwar J. Slope stability analysis in limestone mines. Int. J. Earth Sci. Eng. 5 (2012) 759-766.

Thiebes B, Bell R, Glade T, Jäger S, Anderson M, Holcombe L. A WebGIS decision-support system for slope stability based on limit-equilibrium modelling. Eng. Geol. 158 (2013) 109-118.

Thiebes, B., Bell, R., Glade, T. et al. Integration of a limit-equilibrium model into a landslide early warning system. Landslides 11, 859–875 (2014). https://doi.org/10.1007/s10346-013-0416-2

He, L., Wu, G. & Wang, H. Study of base friction simulation tests based on a complicated engineered bridge slope. Front. Struct. Civ. Eng. 6, 393–397 (2012). https://doi.org/10.1007/s11709-012-0174-6

Kim JC, Jung H, Kim S, Chung K. Slope based intelligent 3D disaster simulation using physics engine. Wireless Pers. Commun. 86 (2016) 183-199.
Sarkar, K., Singh, T.N. & Verma, A.K. A numerical simulation of landslide-prone slope in Himalayan region—a case study. Arab J Geosci 5, 73–81 (2012). https://doi.org/10.1007/s12517-010-0148-8

Pirone, M., Papa, R., Nicotera, M.V. et al. In situ monitoring of the groundwater field in an unsaturated pyroclastic slope for slope stability evaluation. Landslides 12, 259–276 (2015). https://doi.org/10.1007/s10346-014-0483-z

Liu Z, Shao J, Xu W, Shi C. Estimation of Elasticity of Porous Rock Based on Mineral Composition and Microstructure. Advances in Materials Science and Engineering, vol. 2013, Article ID 512727, 10 pages, 2013. https://doi.org/10.1155/2013/512727

Liu Z, Shao JF, Weiya X, Wu Q. Indirect estimation of unconfined compressive strength of carbonate rocks using extreme learning machine. Acta Geotechnica, Springer Verlag, 2015, 10 (5), pp.651-663. https://hal.archives-ouvertes.fr/hal-01200765

Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S. et al. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856 (2016). https://doi.org/10.1007/s10346-015-0614-1

Zhou J, Li X, Mitri HS. Classification of Rockburst in Underground Projects: Comparison of Ten Supervised Learning Methods. Journal of Computing in Civil Engineering 30 (2016) https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000553

Lu, P. and Rosenbaum, M. S. (2003). Artificial neural networks and grey systems for the prediction of slope stability. Natural Hazards, Vol. 30, No. 3, pp. 383-398, DOI: 10.1023/B:NHAZ.000007168.00673.27.

Zhao, H., Yin, S., Ru, Z., 2012. Relevance vector machine applied to slope stability analysis. Int. J. Numer. Anal. Methods Geomech. 36 (5), 643-652.

Li S, Zhao HB, Ru Z. Slope reliability analysis by updated support vector machine and Monte Carlo simulation. Natural Hazards 65 (2013) 707-722.

Wen T, Zhang B, Prediction model for open-pit coal mine slope stability based on random forest. Sci. Technol. Rev. 32 (2014) 105-109.

Khajehzadeh M, Taha MR, El-Shafie A, Eslami M. A modified gravitational search algorithm for slope stability analysis. Eng. Appl. Artif. Intell. 25 (2012) 1589-1597, Dec. 2012.

Raihan TM, Mohammad K, Mahdiyeh E. A new hybrid algorithm for global optimization and slope stability evaluation. J. Central South Univ. 20 (2013) 3265-3273.

Shi XZ, Zhou J, Zheng W, Hu H, Wang HY. Bayes discriminant analysis method and its application for prediction of slope stability. J. Sichuan Univ., Eng. Sci. Ed. 42 (2010) 63-68.

Feng, X. (2000). Introduction of Intelligent Rock Mechanics, Science Press, Beijing, China, pp. 239-241.

Li, X. (2004). Comparative studies of artificial neural networks and adaptive Neuro-Fuzzy inference system based approach for the circular sliding slopes stability analysis, Master Thesis, University of South China, Hengyang, Hunan, China.

Huang, Z., Cui, J., and Liu, H. (2004). Chaotic neural network method for slope stability prediction. Chinese Journal of Rock Mechanics and Engineering, Vol. 22, pp. 015.

Wang, H., Xu, W., and Xu, R. (2005). Slope stability evaluation using back propagation neural networks. Engineering Geology, Vol. 80, No. 3, pp. 302-315, DOI: 10.1016/j.enggeo.2005.06.005.
Samui, P. (2008). Slope stability analysis: A support vector machine approach. Environmental Geology, Vol. 56, No. 2, pp. 255-267, DOI: 10.1007/s00254-007-1161-4.

Zhao, H.-b. (2008). Slope reliability analysis using a support vector machine. Computers and Geotechnics, Vol. 35, No. 3, pp. 459-467, DOI: 10.1016/j.compgeo.2007.08.002.

Choobasti, A., Farrokhzad, F., and Barari, A. (2009). Prediction of slope stability using artificial neural network (case study: Noabad, Mazandaran, Iran). Arabian Journal of Geosciences, Vol. 2, No. 4, pp. 311-319, DOI: 10.1007/s12517-009-0035-3.

Das, S. K., Biswal, R. K., Sivakugan, N., and Das, B. (2011). “Classification of slopes and prediction of factor of safety using differential evolution neural networks.” Environmental Earth Sciences, Vol. 64, No. 1, pp. 201-210, DOI: 10.1007/s12665-012-1900-1.

Erzin, Y. and Cetin, T. (2013). The prediction of the critical factor of safety of homogeneous finite slopes using neural networks and multiple regressions. Computers & Geosciences, Vol. 51, pp. 305-313, DOI: 10.1016/j.cageo.2012.09.003.

Liu, Z., Shao, J., Xu, W., Chen, H., and Zhang, Y. (2014). “An extreme learning machine approach for slope stability evaluation and prediction.” Natural Hazards, Vol. 73, No. 2, pp. 787-804, DOI: 10.1007/s11069-014-1106-7.

Gordan, B., Jahed Armaghani, D., Hajihassani, M., and Monjezi, M. (2016). Prediction of seismic slope stability through combination of particle swarm optimization and neural network. Engineering with Computers, Vol. 32, No. 1, pp. 85-97, DOI: 10.1007/s00366-015-0400-7.

Hoang, N.-D. and Pham, A.-D. (2016). Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: A multinational data analysis. Expert Systems with Applications, Vol. 46, pp. 60-68, DOI: 10.1016/j.eswa.2015.10.020.

Suman, S., Khan, S., Das, S., and Chand, S. (2016). Slope stability analysis using artificial intelligence techniques. Natural Hazards, Vol. 84, No. 2, pp. 727-748, DOI: 10.1007/s11069-016-2454-2.

Verma, A., Singh, T., Chauhan, N. K., and Sarkar, K. (2016). A hybrid FEM–ANN approach for slope instability prediction. Journal of The Institution of Engineers (India): Series A, Vol. 97, No. 3, pp. 171-180, DOI: 10.1007/s40030-016-0168-9.

Rukhaiyar, S., Alam, M., and Samadhiya, N. (2017). A PSO-ANN hybrid model for predicting factor of safety of slope. International Journal of Geotechnical Engineering, pp. 1-11, DOI: 10.1080/19386362.2017.1305652.

Xue, X. (2017). Prediction of slope stability based on Hybrid PSO and LSSVM. Journal of Computing in Civil Engineering, Vol. 31, No. 1, pp. 1-10, DOI: 10.1061/(ASCE)CP.1943-5487.0000607.

Feng X, Li S, Yuan C, Zeng P, Sun Y. Prediction of Slope Stability using Naive Bayes Classifier. KSCE Journal of Civil Engineering 22 (2018): 941-950.

Zhou J, Li E, Yang S, Wang M, Shi X, Yao S, Mitri HS, Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories, Safety Science 118 (2019) 505-518. https://doi.org/10.1016/j.ssci.2019.05.046.

Zhang Z, Liu Z, Zheng L, Zhang Y. Development of an adaptive relevance vector machine approach for slope stability inference, Neural Comput. Appl., vol. 25, pp. 2025-2035, Dec. 2014.

Lin Y, Zhou K, Li J. Prediction of Slope Stability Using Four Supervised Learning Methods. In IEEE Access, vol. 6, pp. 31169-31179, 2018, doi: 10.1109/ACCESS.2018.2843787.
Manouchehrian A, Gholamnejad J, Sharifzadeh M. Development of a model for analysis of slope stability for circular mode failure using genetic algorithm,'" Environ. Earth Sci., vol. 71, pp. 1267-1277, Feb. 2014.

Xue, X., Yang, X. & Chen, X. Application of a support vector machine for prediction of slope stability. Sci. China Technol. Sci. 57, 2379–2386 (2014). https://doi.org/10.1007/s11431-014-5699-6

Mahmoodzadeh A, Zare S (2016) Probabilistic prediction of expected Ground conditions and construction time and costs in road tunnels. Journal of Rock Mechanics and Geotechnical Engineering 8:734-745. doi: 10.1016/j.jrmge.2016.07.0 01

Mahmoodzadeh A, Mohammadi M, Daraei A, Rashid T.A, Sherwani A.F.H, Faraj R.H, Darwesh A.M. Updating ground conditions and time-cost scatter-gram in tunnels during excavation, Autom. Constr. 105 (2019) 102822. https://doi.org/10.1016/j.autcon.2019.04.017.

Mahmoodzadeh A, Mohammadi M, Daraei A, Faraj R.H, Omer R.M.D, Sherwani A.F.H. Decision-making in tunneling using artificial intelligence tools. Tunnelling and Underground Space Technology (2020a). https://doi.org/10.1016/j.tust.2020.103514

Mahmoodzadeh A, Mohammadi M, Daraei A, Hama-Ali H.F, Abdullah A.I, Al-Salihi N.K. Forecasting tunnel geology, construction time and costs using machine learning methods. Neural Computing and Applications (2020b). DOI: 10.1007/s00521-020-05006-2

Mahmoodzadeh A, Mohammadi M, Daraei A, Hama-Ali H.F, Al-Salihi N.K, Omer R.M.D (2020c) Forecasting maximum surface settlement caused by urban tunneling. Automation in Construction 120:103375. https://doi.org/10.1016/j.autcon.2020.103375

Mahmoodzadeh A, Mohammadi M, Abdulhamid SN, Ibrahim HH, Hama Ali HF, Salim SG. Dynamic reduction of time and cost uncertainties in tunneling projects. Tunnelling and Underground Space Technology 109 (2021a) 103774. https://doi.org/10.1016/j.tust.2020.103774.

Mahmoodzadeh A, Mohammadi M, Ibrahim HH, Abdulhamid SN, Salim SG, Hama Ali HF, Majeed MK. Artificial intelligence forecasting models of uniaxial compressive strength. Transportation Geotechnics 27 (2021b) 100499. https://doi.org/10.1016/j.trgeo.2020.100499.

Mahmoodzadeh A, Mohammadi M, Hama Ali HF, Abdulhamid SN, Ibrahim HH, Noori KMG. Dynamic prediction models of rock quality designation in tunneling projects. Transportation Geotechnics 27 (2021c) 100497. https://doi.org/10.1016/j.trgeo.2020.100497.

Mahmoodzadeh A, Mohammadi M, Ibrahim HH, Noori KMG, Abdulhamid SN, Hama Ali HF. Forecasting sidewall displacement of underground caverns using machine learning techniques. Automation in Construction 123 (2021d) 103530. https://doi.org/10.1016/j.autcon.2020.103530.

Vapnik, V.N., 1995. The nature of statistical learning theory. New York: Springer.

**Figure captions**

**Fig. 1** Overall procedure of slope stability prediction using ML techniques.

**Fig. 2** A slope geometry and the effective parameters on the stability of slopes.

**Fig. 3** The box graph of six factors.

**Fig. 4** Pie chart of all slope cases.

**Fig. 5** Comparison of the FOS results predicted by the GPR model with the analyzed ones.

**Fig. 6** FOS results produced by the GPR model versus the analyzed results.

**Fig. 7** Comparison of the SVR results with the analyzed ones.
Table captions

Table 1 Earlier studies on the slope stability predictions using soft computing approaches.
Table 2 Statistical features of the remained slope cases.
Table 3 The optimized parameters of the GPR model.
Table 4 The optimized parameters of the SVR model.
Table 5 The optimized model type and hyper-parameters of the DT method.
Table 6 The type/value of the optimized LSTM parameters.
Table 7 The type/value of the optimized DNN parameters.
Table 8 Comparison of the statistical evaluation indices results produced by the prediction models.
Table 9 The first step of feature selection.
Database

- Unit weight
- Pore pressure ratio
- Slope angle
- Slope height
- Cohesion
- Friction angle
- Target: Factor of safety (FOS)

K-fold CV (K=5)

- Training set
- Testing set

Data normalization

ML algorithms

- GPR
- SVR
- DT
- LSTM
- DNN
- KNN

Statistical evaluation indices

- $R^2$
- MAE
- RMSE
- MAPE

Results comparison

- Identify the best prediction model

Feature selection

- Identify the most effective features on the slope stability

**Figure 1**

Water condition:
- $r_u$: Pore pressure ratio

Shear strength of geomaterial:
- $c$: Cohesion
- $\Phi$: Friction angle

Circular slip failure

- $H$: Slope height
- $H_w$: Water level
- $\alpha$: Slope angle

**Figure 2**
Figure 3

Figure 4

Stable: 49%
Failure: 51%
Figure 5

Figure 6

Figure 7
**Figure 8**

![Graph showing predicted FOS vs. analyzed FOS with R² = 0.695.]

**Figure 9**

![Graph showing errors, analyzed FOS, and DT predictions.]

**Figure 10**

![Graph showing predicted FOS vs. evaluated FOS with R² = 0.683.]

SVR
Linear (SVR)

Linear (DT)

DT
Figure 14

![Graph showing predicted FOS vs. analyzed FOS with R² = 0.6616]

Figure 15

![Diagram illustrating the structure of a neural network with layers and connections]

Figure 16

![Bar graph showing errors and FOS for DNN predictions]
Figure 17

Figure 18

Figure 19
Table 1

| References             | Parameters                  | Soft computing method | Datasets no. |
|------------------------|-----------------------------|-----------------------|--------------|
| Feng, (2000)           | $r_u, H, \alpha, \gamma, \phi, c$ | FFNN                  | 82           |
| Lu and Rosenbaum, (2003)| $H, \alpha, \gamma, \phi, c, r_u$ | BPNN                  | 32           |
| Li, (2004)             | $H, \alpha, \gamma, \phi, c$ | ANN and ANFIS         | 59           |
| Huang et al., (2004)   | $H, \alpha, \gamma, \phi, c, r_u$ | CNN                   | 64           |
| Sakellariou and Ferentinou, (2005) | $H, \alpha, \gamma, \phi, c, r_u$ | BPNN                  | 46           |
| Wang et al., (2005)    | $H, \alpha, \gamma, \phi, c$ | BPNN                  | 27           |
| Samui, (2008)          | $H, \alpha, \gamma, \phi, c, r_u$ | SVM                   | 46           |
| Zhao, (2008)           | $\gamma, \phi, c$          | SVM                   | 10           |
| Choobbasti et al., (2009) | $a_h, X, \phi, \alpha, a_w, c, Y, S_r$ | ANN                  | 36           |
| Das et al., (2011)     | $r_u, H, \alpha, \gamma, \phi, c$ | ANN                   | 46           |
| Erzin and Cetin, (2013)| $\gamma, c, \phi, \alpha, H$ | ANN                   | 675          |
| Liu et al., (2014)     | $r_u, H, \alpha, \gamma, \phi, c$ | ELM                   | 97           |
| Gordan et al., (2016)  | $H, \alpha, \gamma, \phi, c, PGA$ | PSO–ANN               | 699          |
| Hoang and Pham, (2016) | $r_u, H, \alpha, \gamma, \phi, c$ | LS-SVC                | 168          |
| Suman et al., (2016)   | $r_u, H, \alpha, \gamma, \phi, c$ | FNs, MARS, MGGP      | 103          |
| Verma et al., (2016)   | $pp, c, \alpha, \phi$       | ANN                   | 100          |
| Rukhaiyar et al., (2017)| $r_u, H, \alpha, \gamma, \phi, c$ | PSO–ANN               | 83           |
Table 2

| Parameter | Value or type |
|-----------|---------------|
| $\gamma$ (kg/m$^2$) | 327.0000 |
| $c$ (kPa) | 327.0000 |
| $\Phi$ (°) | 327.0000 |
| $\alpha$ (°) | 327.0000 |
| $H$ (m) | 327.0000 |
| $r_u$ | 327.0000 |
| FOS | 327.0000 |

Table 3

| Parameter | Value or type |
|-----------|---------------|
| Kernel function | 'Exponential' |
| Beta | 1.2517 |
| Fit method | Exact Gaussian process regression |
| Sigma | 0.0845 |
| Basis function | 'Constant' |

Table 4

| Parameter | Value or type |
|-----------|---------------|
| Kernel function | 'Medium Gaussian' |
| Solver | 'SMO' |
| Epsilon | 0.0319 |
| Bias | 1.3317 |

Table 5

| Parameter | Value or type |
|-----------|---------------|
| Prune | on |
| Model type | Fine tree |
| Maximum number of splits | 326 |
| Predictor selection | allsplits |
| Minimum leaf size | 4 |
| Split criterion | mse |
| Minimum ParentSize | 10 |

Table 6

| No. of layers | Model type | Neurons no. | Activation function | Optimizer | batch_size | Epochs |
|---------------|------------|-------------|---------------------|-----------|------------|--------|
| 1 input layer | Sequential() | 6 neurons for the input layer | ReLU | Nadam | 10 | 100 |
| 4 hidden layers | | 32 neurons within each hidden layer | | | | |
| 1 output layer | | 1 neuron for the output layer | | | | |
### Table 7

| No. of layers | Model type    | Neurons no.   | Activation function | Optimizer | batch_size | Epochs | Kernel initializer |
|---------------|---------------|---------------|---------------------|-----------|------------|--------|-------------------|
| 1 input layer | Sequential()  | Input layer: 6 neurons | ReLU               | Nadam     | 10         | 100    | Lecun_uniform     |
| 4 hidden layers |              | Hidden layers: 32 neurons |                   |           |            |        |                   |
| 1 output layer |              | Output layer: 1 neuron |                   |           |            |        |                   |

### Table 8

| Method | $R^2$ | MAE   | RMSE  | MAPE [%]     |
|--------|-------|-------|-------|--------------|
| GPR    | 0.8139| 8.42E-02 | 0.160892561 | 7.209771731 |
| SVR    | 0.6950| 1.42E-01 | 0.20650649 | 11.89200416 |
| DT     | 0.6830| 1.43E-01 | 0.210590831 | 12.2080868 |
| LSTM   | 0.6616| 1.69E-01 | 0.221912033 | 14.52803317 |
| DNN    | 0.7498| 1.18E-01 | 0.188416005 | 9.967075268 |
| KNN    | 0.7080| 9.13E-02 | 0.205355281 | 7.874855748 |

### Table 9

| Parameter | Estimate | SE       | tStat | p-value | Significance code |
|-----------|----------|----------|-------|---------|-------------------|
| (Intercept)| 1.3161   | 0.074702 | 17.619| 4.85E-49| ***               |
| $\gamma$  | 0.011828 | 0.0036237| 3.2642| 0.0012166| **                |
| $c$       | 0.023571 | 0.0014477| 16.281| 7.75E-44 | ***               |
| $\phi$    | -0.025905| 0.0014063| -18.421| 3.63E-52 | ***               |
| $\alpha$  | -0.00088078| 0.00012506 | -7.0428 | 1.16E-11 | ***               |
| $\theta$  | 0.0094102 | 0.00084569| 11.127| 1.57E-24 | ***               |
| $r_u$     | -0.72688 | 0.066114 | -10.994| 4.57E-24 | ***               |