ORIGINAL

Hadi Fathipour-Azar ⬚

Data-oriented prediction of rocks' Mohr–Coulomb parameters

Abstract Mohr–Coulomb shear strength parameters, namely cohesion strength and internal friction angle, are the key determinants of intact rock strength. Triaxial tests are generally performed at different stress levels to determine these shear parameters. However, as an alternative to this expensive and time-consuming method, data-oriented computational approach is available to employ and provides a new technical means for geomaterial strength parameters prediction. In this paper, random forest, alternating model tree, and support vector machine techniques are utilized to predict Mohr–Coulomb shear strength parameters by using the published database of the uniaxial compressive strength, uniaxial tensile strength, and different stress conditions in which failure occurs. In this regard, 80% of data (176 samples) are used to train models, while 20% (45 samples) is for testing the developed models. For internal friction angle, coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) are $R^2 > 0.90$, MAE < 1, and RMSE $< 1.6°$ in the training and testing phases for established models. Based on the results, $R^2 > 0.98$, MAE < 1.1, and RMSE < 1.5 MPa are obtained for developed models of cohesion strength prediction in the training and testing phases, which demonstrate the efficiency of proposed approaches in predicting shear strength parameters of sandstone rock. The sensitivity analyses indicate that confining stress has the most significant influence in increasing the prediction accuracy. This work provides a general and robust data-centric intelligent framework for predicting micro-/macro-parameters of geomaterial, which improves shear strength design in the field.

Keywords   Shear strength · Cohesion · Friction angle · Surrogate models · Data-oriented techniques · Sensitivity analysis

1 Introduction

The strength of intact rock mainly depends upon the shear strength parameters. Therefore, determinations of the shear strength properties of the rock are of primary interest in the geomechanics and geoengineering investigation. The common procedure to determine the shear strength parameters of rock is triaxial tests. While the experimental studies need notable expenditure and time to have a well-equipped laboratory with series of experiments, the predictive models are valuable to advance understanding strength parameters of rock under various conditions. Moreover, since complex instruments such as a triaxial test apparatus are used, high technological skills are often needed. Therefore, it is time to establish reliable approaches that are able to predict the geomechanical properties of rocks.

Several scholars have made contributions in the field of predicting shear strength parameters in rocks. Sivakugan et al. [1], Karaman et al. [2], Shen and Jimenez [3], and Moon and Yang [4] investigated the feasibility of compressive and tensile strength data (UCS and UTS values) to determine shear strength parameters of

H. Fathipour-Azar (✉)
Faculty of Mining Engineering, Petroleum and Geophysics, Shahrood University of Technology, Shahrood, Iran
e-mail: fathipour.hadi@gmail.com

rock using mathematical regressions. Shen et al. [5] determined Mohr–Coulomb shear strength parameters from generalized Hoek–Brown criterion. Zhang et al. [6] studied the statistical correlation between shear strength parameters of rock masses using Hoek–Brown (H–B) failure criterion and triaxial test result of intact rocks. As indicated [6–8], the utilization of incorrect $C - \varphi$ correlation coefficient values may result in either underestimation or overestimation of failure probability in risk analysis of geomechanics engineering. Wei et al. [9] estimated the equivalent Mohr–Coulomb parameters based on the Hoek–Brown criterion. They also investigated its applicability in slope analysis.

Recently, soft computing techniques have been used to predict shear strength parameters of rock. Armaghani et al. [10] estimated shear strength parameters (interlocking and friction angle) of shale rock using rock index tests through particle swarm optimization–artificial neural network (PSO–ANN) integrated model. Rock index tests include dry density, point load index, Brazilian tensile strength, ultrasonic velocity, and Schmidt hammer test. They found that the proposed PSO–ANN model could accurately predict shear strength parameters of shale rock. Shen and Jimenez [3] employed genetic programming to estimate the Mohr–Coulomb shear strength parameters of sandstone rock using two strength indices (UCS and UTS) under triaxial stress conditions. Confining stress (σ_3) is taken into account to consider the nonlinearity of the failure envelopes and enhance the effectiveness of estimates. Based on laboratory tests on shale rock specimens [10, 11] studied rock interlocking through two hybrid ANN-based models by considering genetic algorithm (GA) and fuzzy inference system. They indicated that developed hybrid models are reliable. Moreover, the ANFIS predictive model performed slightly better than the GA-ANN method in terms of interlocking estimation. Shao et al. [12] predicted the internal friction angle of shale rocks using two hybrids neural net-based models that integrate ANN with GA and imperialist competitive algorithm (ICA). They considered p-wave velocity, Schmidt hammer, and point load as a system input. Based on performance indices, they presented ICA–ANN as a new method for estimating the internal friction angle of shale rock samples. Although the NN-based method is widely used in this topic, in recent years, there has been a growing interest in the application of data-centric machine learning (ML) paradigms in geomechanics and geoengineering investigations [13–24].

The present study is the first attempt at indirect prediction of Mohr–Coulomb shear strength parameters such as internal friction angle and cohesion strength of sandstone rock employing random forest (RF), alternating model tree (AT), and support vector machine (SVM) with two kernels [radial basis function (RBF) and Pearson VII kernel function (PUK)] methods under triaxial conditions using the uniaxial compressive strength (UCS) and uniaxial tensile strength (UTS). Quantification of potentiality and functionality of proposed ML algorithms are provided in terms of statistical performance indicators. The sensitivity analyses are also carried out to determine the importance of each input parameter in order to predict the Mohr–Coulomb shear strength parameters. The rest of the paper is organized as follows. Section 2 reviews the background knowledge of ML algorithms. Section 3 provides the study data. Section 4 presents the performance measure indices. Section 5 delivers the obtained results and discussion. Section 6 presents sensitivity analysis. Finally, conclusions are represented in Sect. 7.

2 Data-oriented computational methods

2.1 Random forest (RF)

Random forest (RF) is an ensemble tree-based regression method. The RF algorithm consists of regression trees trained using various bootstrap samples (bagging) of the training database. Each tree functions regression on its own, and the final output decision of RF is the average of the outputs of individual regression tree regressors. In the case of bagging, the training dataset contains about two-third of the data from the actual training dataset (InBag). The rest out-of-bag (OOB) training data (one-third out of the bootstrap sampling data) from each tree developed are utilized to assess estimation error and variable significance. The RF regression model is built using two main parameters: the number of trees and the selected features that split the nodes [25]. The flow chart of the RF algorithm is presented in Fig. 1.

2.2 Alternating model tree (AT)

Decision tree techniques are commonly used approaches for classification and regression problems. Alternating model tree (AT) is a new algorithm based on ensemble learning method whose strategy is to iteratively correct
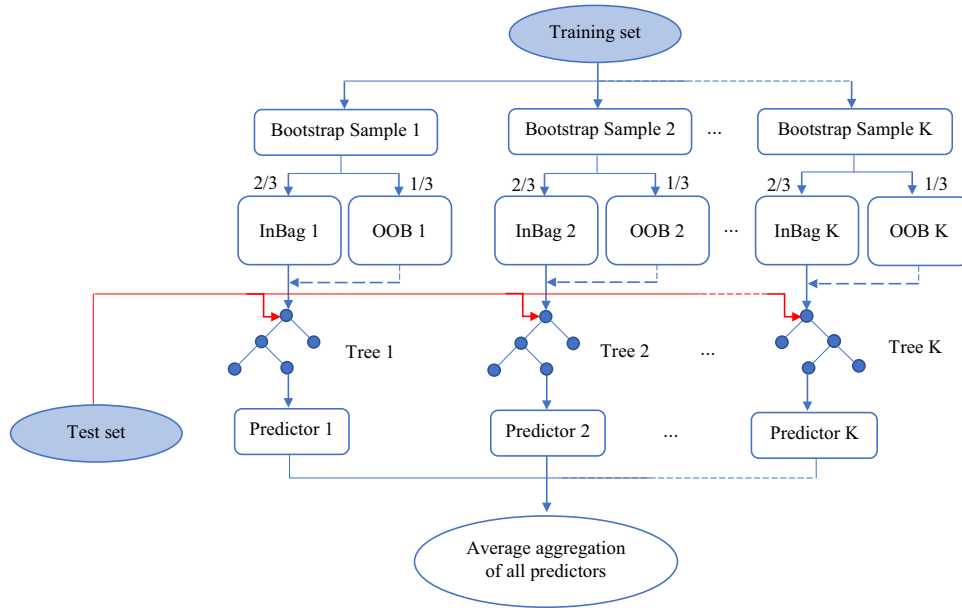
Fig. 1 Flow chart of RF algorithm

the predictions of an additive model by using the residual errors it makes on the training data as target [26]. The layout of the AT is a single tree. AT performs regression with two different nodes, namely splitter and prediction nodes. The splitter node splits numeric attributes at the median value. An estimator node that produces a model response using a linear regression method. The AT model is composed of k base models of the form of $F_k(\vec{x}) = \sum_{j=1}^{k} f_j(\vec{x})$ that are fitted in a forward stage-wise additive regression manner to minimize a squared errors measure, i.e., $\sum_{i=1}^{n} (F_k(\vec{x}_i) - y_i)^2$ through n training samples $(\vec{x}_i, y_i)$. The target variable is the residual of the fit at the $k - 1$ stage for every iteration k. The approach uses a shrinkage parameter λ, which is a multiplicative factor in the range (0, 1] that is applied to the predictions of the base model before they are included in the additive model. It dampens the values towards the mean prediction and aids in model regularization.

2.3 Support vector machine (SVM)

Support vector machine (SVM) is based on statistical learning theory [27]. The SVM implements the structural risk minimization principle. Let (x, y) be a training dataset, where $x \in R^n$ stands for an input vector with n components and y represents the corresponding output. In the regression model of SVM, the goal is to determine a function $f(x)$ whose deviation from each target y_i is at most ε for all training data. The estimating function can be represented by the following equation:

$$f(x) = \omega.\varphi(x) + b \tag{1}$$

where ω is the weight vector, b is the bias term, and $\varphi(x)$ denotes a set of nonlinear transformations. The coefficients ω and b are calculated by minimizing the following regularized risk function:

$$\frac{1}{2}||\omega||^2 + C \sum_{i=1}^{n} (\xi_i, \xi_i^*)$$

$$\text{Subject to} : \begin{cases} \omega.\varphi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ y_i - \omega.\varphi(x_i) - b \le \varepsilon + \xi_i \\ \xi_i, \xi_i^* \ge 0, i = 1, 2, .., n \end{cases} \tag{2}$$

Table 1 Statistics analysis of the datasets

	UCS (MPa)	UTS (MPa)	σ_3 (MPa)	Friction angle (°)	Cohesion (MPa)
Mean	83.835	5.644	29.342	44.169	20.222
Standard deviation	45.319	4.237	20.356	4.889	9.439
Minimum	41.47	1.75	1.47	33.6	8.14
Maximum	196.97	16.15	100	56.02	49.62

where C is a regulation parameter. ξ_i and ξ_i^* are positive slack variables. The primal Lagrangian form is used to solve this constrained optimization problem [27, 28], and finally written as follows:

$$f(x) = \sum_{i=1}^{n} (\beta_i - \beta_i^*) . K(x_i, x_j) + b \tag{3}$$

where β_i and β_i^* are the Lagrange multipliers and $K(x_i, x_j)$ is the kernel function. It is necessary to choose suitable covariance or kernel function since it directly affects predictive efficiency. In the study, two different (radial basis kernel and Pearson VII kernel function (PUK)) commonly used kernel functions are selected for SVM development.

$$K_{\text{RBF}}(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2} \tag{4}$$

$$K_{\text{PUK}}(x_i, x_j) = \left(1 / \left[1 + \left(2\sqrt{||x_i - x_j||^2} \sqrt{2^{(1/\omega)} - 1} / l \right)^2 \right]^{\omega} \right) \tag{5}$$

where γ, l, and ω are the kernels parameters (also known as hyper-parameters). $||x_i - x_j||$ is the Euclidean distance between points x_i and x_j. Besides, the data standardization is calculated through the following equation:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{6}$$

where x_i, μ, and σ are the original data, mean, and standard deviation of the data, respectively.

3 Data used for the study

The comprehensive database utilized for developing proposed ML approaches is the same used by Shen and Jimenez [3] for genetic programming analysis derived from the 32 groups of sandstone triaxial test data available in RocData presented by Rocscience [29] company. Groups with some possibly implausible triaxial data and less than three points are excluded from the regression analysis. Therefore, 23 groups with 221 datasets are used for ML analysis.

Data division is the first step in constructing a model layout for prediction research. The data were divided randomly. The randomized data are divided into training (80%) and testing (20%) datasets. Training dataset is utilized for developing and learning the ML techniques, while testing dataset is used for testing the techniques based on performance criteria which are described in the next section.

Testing data are within the range of specified training dataset of shear strength parameters during the development of models. Out of 221 datasets, 176 datasets are used for the training and the remaining 45 datasets are used for testing the generalization capability of the models. In the dataset, the input variables are UCS, UTS, and σ_3. Therefore, the effect of the σ_3, in which the shear failure occurs is also considered in models. The output variables are cohesion strength and internal friction angle. The statistical analysis of the input and output parameters for sandstone rock is given in Table 1 and Fig. 2. In Fig. 2, the diagonal components represent the histograms and density functions of variables. The lower panels show scatter plots for each pair of variables. Each group of sandstone triaxial test data is represented with a different colour. The upper panels in the matrix indicate the correlation coefficients for all data.
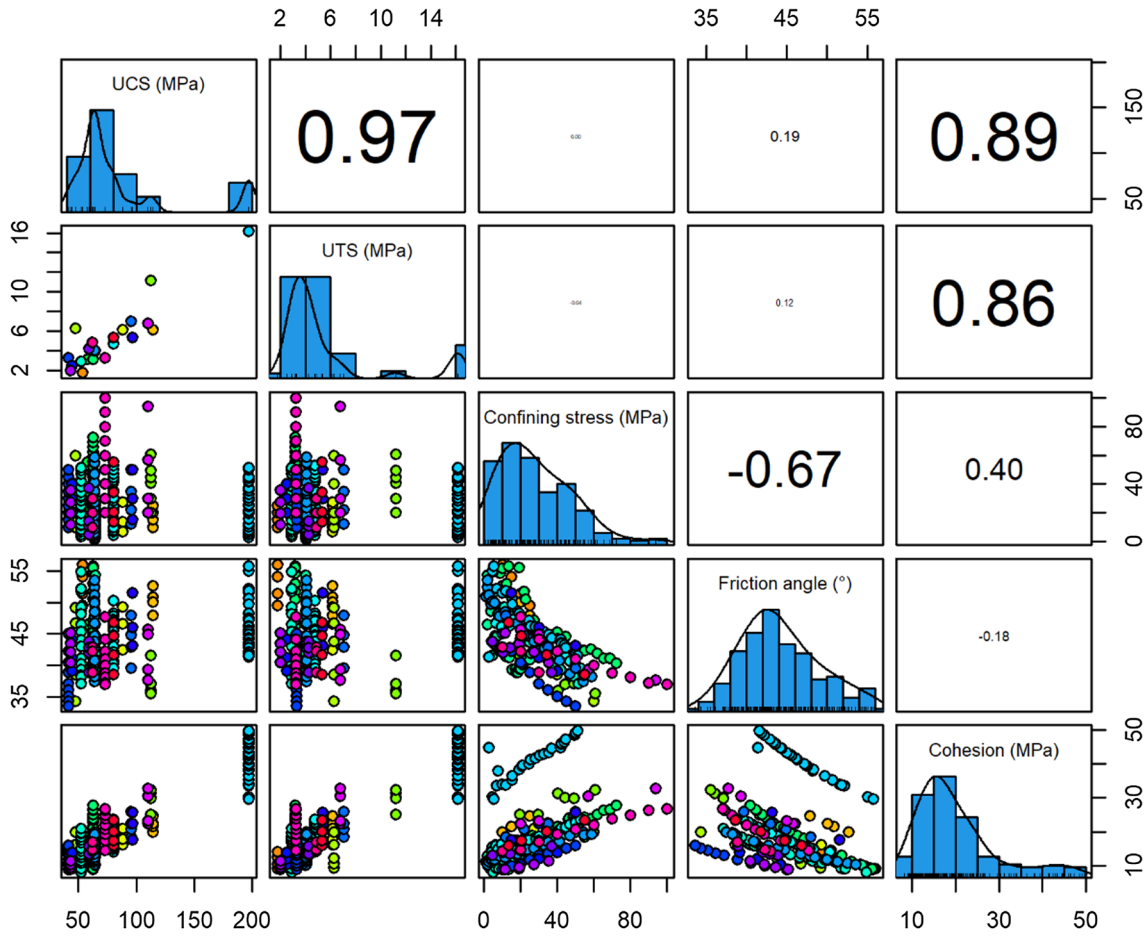
Fig. 2 Scatterplot matrix of datasets

4 Performance evaluation measures

In this study, three well-known performance indicators including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) are employed to evaluate the performance of predictive models of RF, AT, and SVM.

R^2 is used as a measure of the goodness of fit of a statistical model. A higher value of R^2 indicates a better fit for the model. The RMSE is utilized as a measure of the sample standard deviation of the difference between predicted and observed shear strength parameters values, and MAE is employed to measure the closeness of the prediction to the eventual shear strength parameters values. The smaller values of RMSE and MAE indicate better prediction with zero values for these two criteria implying a perfect prediction. The formulation of these indices is expressed by the following equations:

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)(y_i - \overline{y})\right]^2}{\sum_{i=1}^{n}\left(y_i^* - \overline{y^*}\right)^2 \sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{7}$$

$$\text{MAE} = \sum_{i=1}^{n}\left|y_i^* - y_i\right|/n \tag{8}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n}\left(y_i^* - y_i\right)^2/n} \tag{9}$$

where y_i^*, y_i, $\overline{y^*}$, and $\overline{y}$ indicate the predicted, calculated, mean value of predicted and calculated shear strength parameters values, respectively. Also, n is the number of data.

Table 2 Main optimized parameters of developed models

Model	Friction angle	Cohesion
RF	Number of trees = 130, features = 2	Number of trees = 100, features = 2
AT	Iteration = 34	Iteration = 23
SVM-RBF	$C = 11$, $\gamma = 7.8$, $\varepsilon = 0.001$	$C = 22$, $\gamma = 2.9$, $\varepsilon = 0.001$
SVM-PUK	$C = 25$, $\omega = 0.25$, $\sigma = 0.5$, $\varepsilon = 0.001$	$C = 9$, $\omega = 0.2$, $\sigma = 1.5$, $\varepsilon = 0.001$

Table 3 The evaluation indicators of R^2, MAE, and RMSE statistics of RF, AT, and SVM models for the prediction of friction angle and cohesion

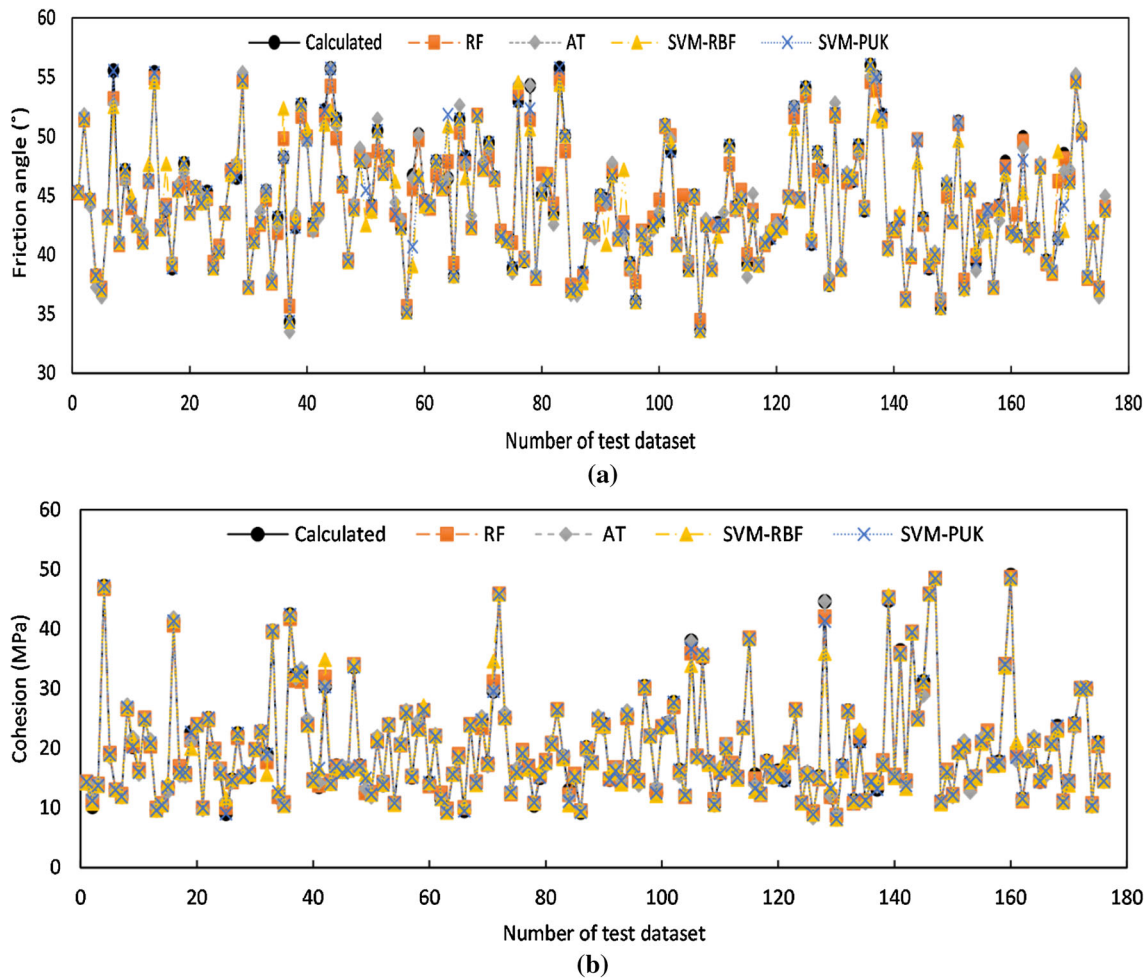| Model	Friction angle						Cohesion					
	Train			Test			Train			Test		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
RF	0.976	0.523	0.81	0.944	0.878	1.203	0.997	0.385	0.548	0.993	0.679	0.942
AT	0.984	0.446	0.627	0.940	0.78	1.128	0.995	0.469	0.624	0.987	0.972	1.324
SVM-RBF	0.908	0.605	1.514	0.924	0.82	1.262	0.983	0.465	1.169	0.983	1.084	1.468
SVM-PUK	0.974	0.164	0.798	0.931	0.999	1.311	0.997	0.145	0.5	0.984	1.088	1.419

5 Results and discussion

In this study, shear strength parameters of sandstone rock are estimated using three data mining techniques. The accuracy of the proposed RF, AT, and SVM models is calculated using the evaluation indices described above. R^2, MAE, and RMSE are used to investigate the performance of the established models by dividing the data into training data and testing data subsets. The optimum values of the hyperparameters of RF, AT, and SVM models are tuned by grid search and trial and error processes. For the RF model, two parameters, namely number of trees and features; for the AT model, the number of iterations; for the SVM-RBF model, three parameters, namely C, γ, and ε; for the SVM-PUK model, four parameters, namely C, ω, σ, and ε, are tuned. These obtained optimum hyperparameters that developed models generalize well are presented in Table 2.

For the three different techniques used, the R^2, MAE, and RMSE statistics are calculated based on the training and testing datasets. The statistical results of different predictive models used in this study are summarized in Table 3. According to these statistical performance evaluation indices, the performance of the proposed models is comparable. As is seen for friction angle prediction, the RMSE value for the AT model is the lowest than that of the other models in both the training and testing sets. In the AT model for friction angle prediction, values of RMSE and MAE are determined to be 0.627 and 0.446, respectively, based on the training dataset and 1.128 and 0.78, respectively, based on the testing dataset. Although MAE of the AT model is higher than that of the SVM-PUK model in the training stage, it is the lowest than that of the other models in the training and testing phases. In addition, the R^2 value in the AT model is the highest than that of the other models in the training dataset ($R^2 = 0.984$). However, this value ($R^2 = 0.94$) is marginally lower than that of the RF model ($R^2 = 0.944$) and higher than that of the SVM (RBF and PUK) model in the testing dataset. This indicates that the AT model performance is better than the other models for the given data of friction angle. Although the SVM-PUK model presents better performance than the SVM-RBF model in the training dataset, the SVM-RBF model shows slightly better performance (smaller MAE and RMSE) than the SVM-PUK model in the testing dataset of friction angle.

As is evident for cohesion prediction, the lowest MAE and RMSE and highest R^2 values among developed models are for SVM-PUK model in the training dataset. These values are MAE = 0.145, RMSE = 0.5, and $R^2 = 0.997$. With a slight difference with the SVM-PUK model, the RF model shows then better performance in the training dataset (MAE = 0.385, RMSE = 0.548, and $R^2 = 0.997$). However, the RF model demonstrates better performance indices with the lowest MAE = 0.679, RMSE = 0.942, and highest $R^2 = 0.993$ than those of the other models in the testing dataset. The SVM-RBF model shows similar performance to that of the SVM-PUK model with the values of the evaluation indices being close to those obtained based on the SVM-PUK model in the testing dataset of cohesion.

The calculated shear strength parameters values and their estimated values obtained from the proposed RF, AT, and SVM models for the train and test datasets are plotted in Figs. 3 and 4, respectively. It is clear from

Fig. 3 Comparison of calculated and predicted friction angle using RF, AT, and SVM models in the training dataset: **a** friction angle and **b** cohesion

these graphs that the estimated shear strength values obtained from the proposed models are in good agreement with the calculated values indicating that the models employed accurately predict shear strength parameters. According to Figs. 3 and 4, the performance of the AT model is generally better compared with the RF and SVM (with RBF and PUK) models for the friction angle dataset. However, the RF model generally shows better performance for the cohesion dataset.

In addition, the predicted values of shear strength parameters obtained from the RF, AT, and SVM models along with their corresponding calculated values of shear strength parameters are illustrated in the form of scatter plots in Figs. 5 and 6. Predicted values using developed models are aligned around the 45° line. According to these results, the AT and RF models show promising performances in predicting the given shear strength parameters variations based on the training and testing used datasets. Consistency and agreement between calculated and predicted data show the high capability of these methods in estimating shear strength parameters.

A Taylor diagram [30] is a graphical representation of comparing various model outcomes to measured data. The standard deviation, RMSE, and R between different models and measurements are depicted in this diagram. This diagram is plotted for friction angle and cohesion in Figs. 7 and 8, respectively. The location of each model in the diagram indicates how closely the predicted pattern matches with measurements. According to these figures, due to the distance of developed models points to the measured point, developed models are promising methods for estimating shear strength properties.

Mohr–Coulomb shear strength parameters are important in geomechanics engineering applications indicating the resistance of intact rocks with respect to applied stresses on them. Therefore, predicting the cohesion
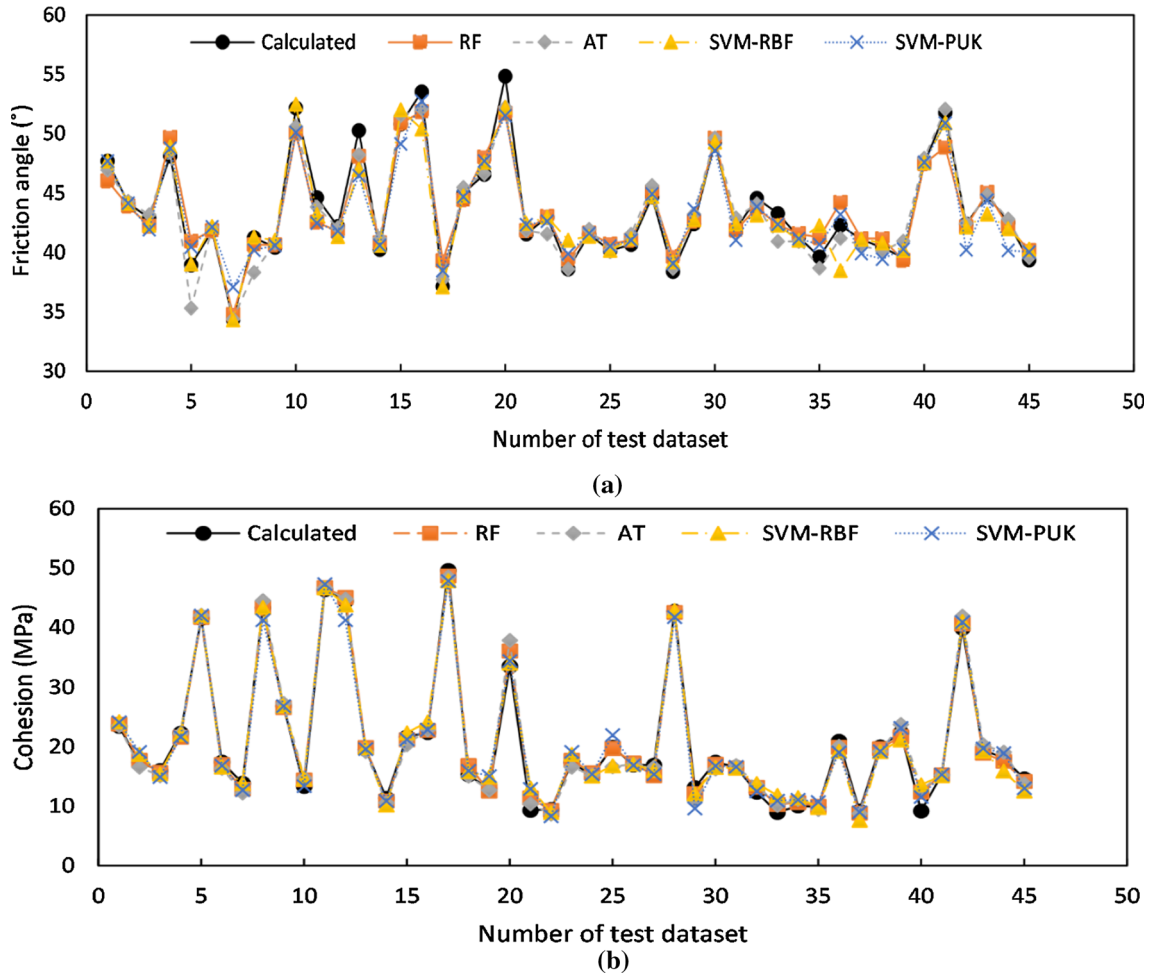
(**a**)

(**b**)

Fig. 4 Comparison of calculated and predicted friction angle using RF, AT, and SVM models in the testing dataset: **a** friction angle and **b** cohesion

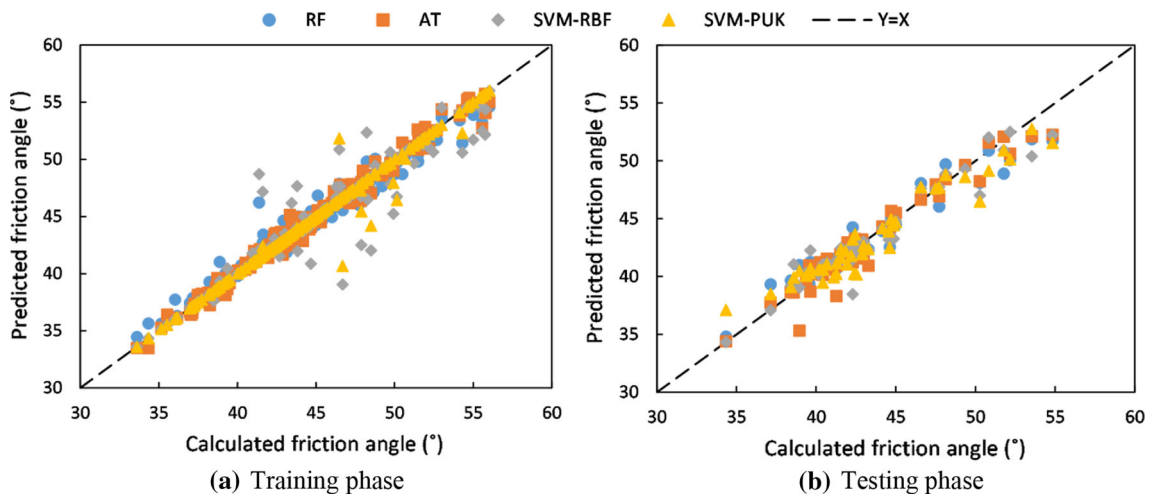(**a**) Training phase   (**b**) Testing phase

Fig. 5 Scatter plot of friction angle values by various developed predictive models for training and testing datasets
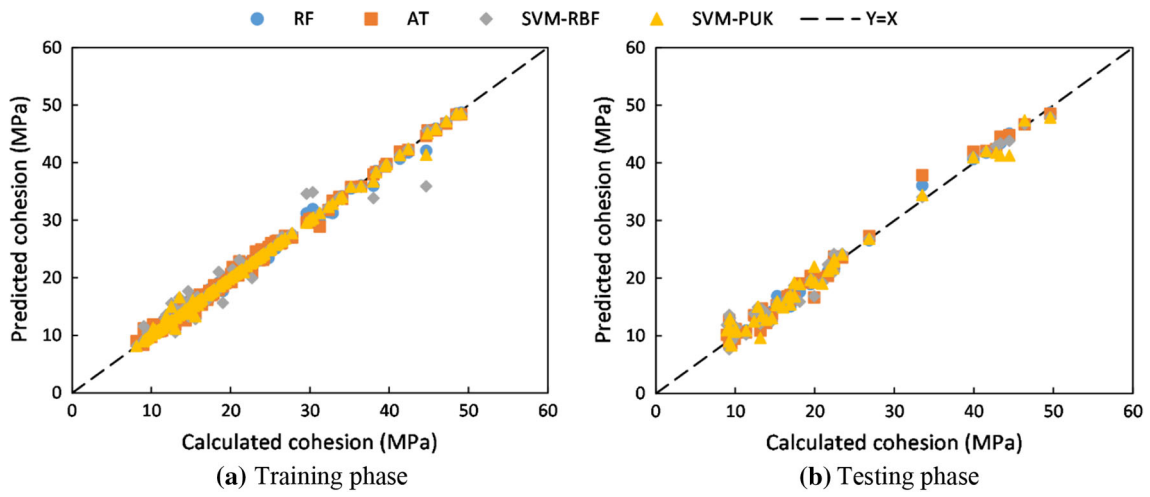
Fig. 6 Scatter plot of friction angle values by various developed predictive models for training and testing datasets
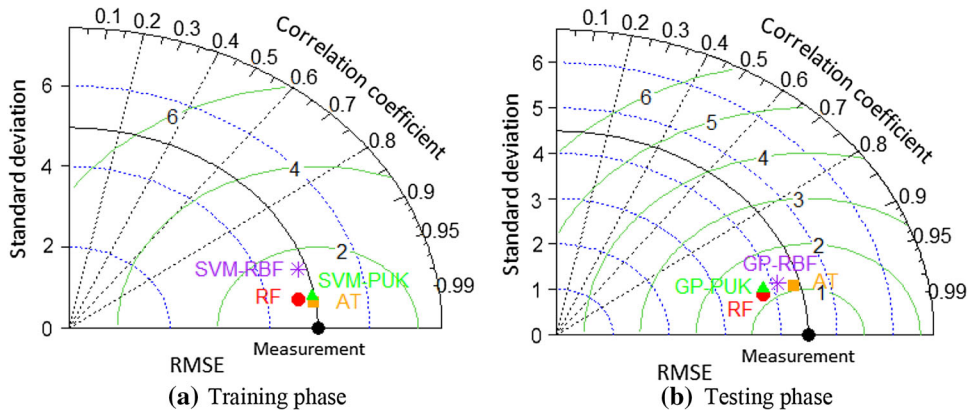

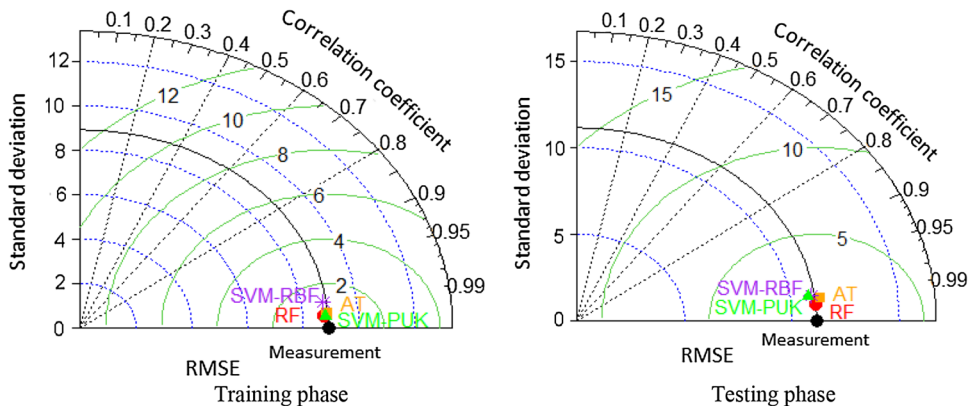**Fig. 7** Comparisons of measured and predicted friction angle by developed models

Fig. 8 Comparisons of measured and predicted cohesion by developed models
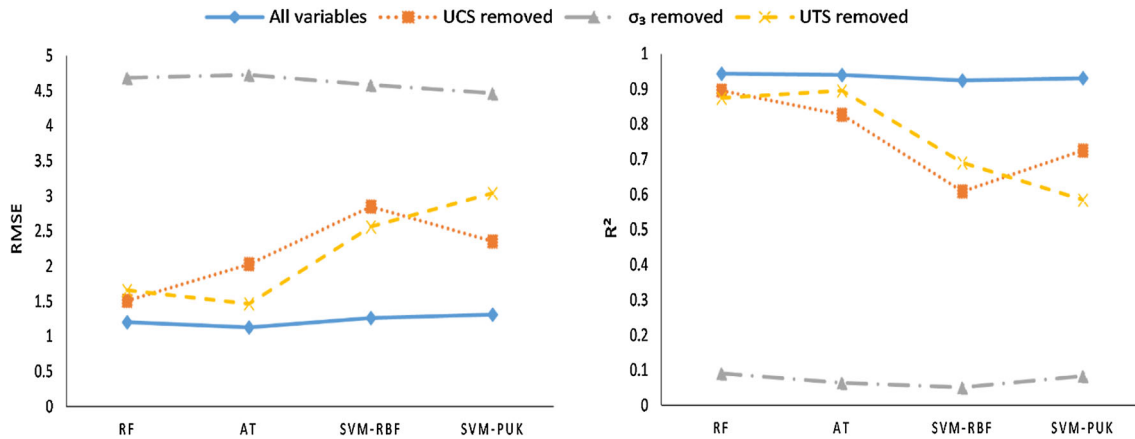
Fig. 9 Influence of removing input variables on the accuracy of different models in the prediction of friction angle parameter

and friction angle as important shear strength parameters is necessary for rock engineering applications such as slope stabilities, tunnelling, and underground cavities. The crucial role of the shear strength of rock in such applications has driven scientists to propose reliable and cost-effective techniques for determining this parameter. Using ML methods with minimal error and high performance plays an important role in providing alternative tools to traditional experiments for estimating desired output. Consequently, comparing the performance of different models gives an insight into the identification of better models for predicting purposes. Other studies have confirmed the applicability and performance of NN-based models [10–12]. This study proposed and compared the accuracy of three data-centric ML techniques, namely RF, AT, and SVM models, in estimating shear strength parameters values. The performance of the models indicates that generally the AT model shows the potential in predicting friction angle, while the RF model shows a promising approach for estimating the cohesion of sandstone rocks. The obtained results demonstrate the high capability of these methods in predicting shear strength parameters.

6 Sensitivity analysis

Sensitivity analysis is performed to assess the effect of each input parameter on the developed models. In terms of the RMSE and R^2, each surrogate model indicates the extent to which the eliminated variable would impact the model accuracy.

Figures 9 and 10 depict the influence of input variables on the friction angle and cohesion predictions, respectively. As can be seen in Fig. 9, σ_3 has the most significant influence in increasing the prediction accuracy. In other words, eliminating σ_3 caused a sharp increase in RMSE values and decrease in R^2 values in all established models for friction angle and cohesion parameters. Therefore, the σ_3 is the most important variable for predicting the shear strength parameters of sandstone rock. The importance of different variables can be displayed as $\sigma_3 > UCS > UTS$ for the cohesion parameter. However, UCS and UTS demonstrate differing influences on the friction angle parameter using established models.

7 Conclusions

The present study explores employing data-oriented ML-based solutions, namely RF, AT, and SVM algorithms, to estimate the shear strength parameters of sandstone rocks using UCS, UTS, and confining stress. The developed surrogate models show good performance in predicting shear strength parameters, with $R^2 > 0.90$, MAE < 1, and RMSE < 1.6° in the training and testing phases for friction angle and $R^2 > 0.98$, MAE < 1.1, and RMSE < 1.5 MPa in the training and testing phases for cohesion prediction. In general, the AT model shows the potential in predicting friction angle, while the RF model shows a promising approach for estimating the cohesion of sandstone rocks. However, there is no certainty that one model will universally outperform other models. Overall, the results prove the viability of estimating shear strength parameters using the proposed techniques. Therefore, establishing faster surrogate models based on a data-driven approach will lead to a
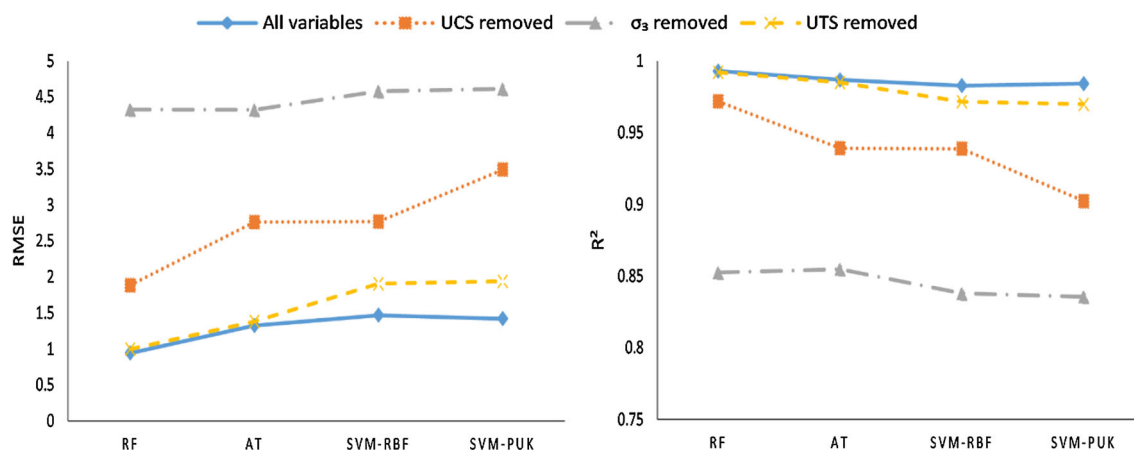
Fig. 10 Influence of removing input variables on the accuracy of different models in the prediction of cohesion parameter

reduction in the cost, effort, and time required for measuring shear strength parameters. Finally, the results of sensitivity analysis show that the σ_3 is the most important influencing input parameter in the estimation of the shear strength parameters for this dataset

Declarations

Conflict of interest The author declares that there is no conflict of interest.

References

1. Sivakugan, N., Das, B.M., Lovisa, J., Patra, C.R.: Determination of c and φ of rocks from indirect tensile strength and uniaxial compression tests. Int. J. Geotech. Eng. **8**(1), 59–65 (2014). https://doi.org/10.1179/1938636213Z.00000000053
2. Karaman, K.A.D.İR., Cihangir, F.E.R.D.İ, Ercikdi, B.A.Y.R.A.M., Kesimal, A.Y.H.A.N., Demirel, S.: Utilization of the Brazilian test for estimating the uniaxial compressive strength and shear strength parameters. J. S. Afr. Inst. Min. Metall. **115**(3), 185–192 (2015)
3. Shen, J., Jimenez, R.: Predicting the shear strength parameters of sandstone using genetic programming. Bull. Eng. Geol. Env. **77**(4), 1647–1662 (2018). https://doi.org/10.1007/s10064-017-1023-6
4. Moon, K., Yang, S.B.: Cohesion and internal friction angle estimated from Brazilian tensile strength and unconfined compressive strength of volcanic rocks in Jeju Island. J. Korean Geotech. Soc. **36**(2), 17–28 (2020). https://doi.org/10.7843/kgs.2020.36.2.17
5. Shen, J., Priest, S.D., Karakus, M.: Determination of Mohr–Coulomb shear strength parameters from generalized Hoek–Brown criterion for slope stability analysis. Rock Mech. Rock Eng. **45**(1), 123–129 (2012). https://doi.org/10.1007/s00603-011-0184-z
6. Zhang, F.P., Li, D.Q., Cao, Z.J., Xiao, T., Zhao, J.: Revisiting statistical correlation between Mohr–Coulomb shear strength parameters of Hoek–Brown rock masses. Tunn. Undergr. Space Technol. **77**, 36–44 (2018). https://doi.org/10.1016/j.tust.2018.03.018
7. Li, D., Chen, Y., Lu, W., Zhou, C.: Stochastic response surface method for reliability analysis of rock slopes involving correlated non-normal variables. Comput. Geotech. **38**(1), 58–68 (2011). https://doi.org/10.1016/j.compgeo.2010.10.006
8. Liu, H., Low, B.K.: System reliability analysis of tunnels reinforced by rockbolts. Tunn. Undergr. Space Technol. **65**, 155–166 (2017). https://doi.org/10.1016/j.tust.2017.03.003
9. Wei, Y., Fu, W., Ye, F.: Estimation of the equivalent Mohr–Coulomb parameters using the Hoek–Brown criterion and its application in slope analysis. Eur. J. Environ. Civ. Eng. (2019). https://doi.org/10.1080/19648189.2018.1538904
10. Armaghani, D.J., Hajihassani, M., Bejarbaneh, B.Y., Marto, A., Mohamad, E.T.: Indirect measure of shale shear strength parameters by means of rock index tests through an optimized artificial neural network. Measurement **55**, 487–498 (2014). https://doi.org/10.1016/j.measurement.2014.06.001
11. Murlidhar, B.R., Ahmed, M., Mavaluru, D., Siddiqi, A.F., Mohamad, E.T.: Prediction of rock interlocking by developing two hybrid models based on GA and fuzzy system. Eng. Comput. **35**(4), 1419–1430 (2019). https://doi.org/10.1007/s00366-018-0672-9
12. Shao, Z., Armaghani, D.J., Bejarbaneh, B.Y., Mu'azu, M.A., Mohamad, E.T.: Estimating the friction angle of black shale core specimens with hybrid-ANN approaches. Measurement **145**, 744–755 (2019). https://doi.org/10.1016/j.measurement.2019.06.007

13. Fathipour-Azar, H.: Machine learning assisted distinct element models calibration: ANFIS, SVM, GPR, and MARS approaches. Acta Geotech. **17**(4), 1207–1217 (2022). https://doi.org/10.1007/s11440-021-01303-9
14. Fathipour-Azar, H.: Data-driven estimation of joint roughness coefficient (JRC). J. Rock Mech. Geotech. Eng. **13**(6), 1428–1437 (2021). https://doi.org/10.1016/j.jrmge.2021.09.003
15. Fathipour-Azar, H.: New interpretable shear strength criterion for rock joints. Acta Geotech. (2022). https://doi.org/10.1007/s11440-021-01442-z
16. Fathipour-Azar, H.: Polyaxial rock failure criteria: insights from explainable and interpretable data driven models. Rock Mech. Rock Eng. **55**(4), 2071–2089 (2022). https://doi.org/10.1007/s00603-021-02758-8
17. Fathipour-Azar, H.: Hybrid machine learning-based triaxial jointed rock mass strength. Environ. Earth Sci. (2022). https://doi.org/10.1007/s12665-022-10253-8
18. Fathipour-Azar, H.: Stacking ensemble machine learning-based shear strength model for rock discontinuity. Geotech. Geol. Eng. (2022). https://doi.org/10.1007/s10706-022-02081-1
19. Fathipour-Azar, H., Saksala, T., Jalali, S.M.E.: Artificial neural networks models for rate of penetration prediction in rock drilling. J. Struct. Mech. **50**(3), 252–255 (2017). https://doi.org/10.23998/rm.64969
20. Fathipour-Azar, H., Wang, J., Jalali, S.M.E., Torabi, S.R.: Numerical modeling of geomaterial fracture using a cohesive crack model in grain-based DEM. Computational Particle Mechanics **7**, 645–654 (2020). https://doi.org/10.1007/s40571-019-00295-4
21. Zhang, W., Phoon, K.K.: Editorial for advances and applications of deep learning and soft computing in geotechnical underground engineering. J. Rock Mech. Geotech. Eng. (2022). https://doi.org/10.1016/j.jrmge.2022.01.001
22. Zhang, W., Wu, C., Zhong, H., Li, Y., Wang, L.: Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. Geosci. Front. **12**(1), 469–477 (2021). https://doi.org/10.1016/j.gsf.2020.03.007
23. Zhang, W., Zhang, R., Wu, C., Goh, A.T.C., Lacasse, S., Liu, Z., Liu, H.: State-of-the-art review of soft computing applications in underground excavations. Geosci. Front. **11**(4), 1095–1106 (2020). https://doi.org/10.1016/j.gsf.2019.12.003
24. Fathipour Azar, H., Torabi, S.R.: Estimating fracture toughness of rock (KIC) using artificial neural networks (ANNS) and linear multivariable regression (LMR) models. In: 5th Iranian Rock Mechanics Conference (2014)
25. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
26. Frank, E., Mayo, M., Kramer, S.: Alternating model trees. In: Proceedings of the 30th annual ACM Symposium on Applied Computing, pp. 871–878 (2015). https://doi.org/10.1145/2695664.2695848
27. Vapnik, V.: The Nature of Statistical Learning. Springer, New York (1995)
28. Vapnik, V., Vapnik, V.: Statistical Learning Theory. Springer, New York (1998)
29. Rocscience: "RocData" (2012). http://www.rocscience.com/products/4/RocData. Accessed 10 Sept 2016
30. Taylor, K.E.: Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. **106**(D7), 7183–7192 (2001). https://doi.org/10.1029/2000JD900719