ORIGINAL ARTICLE

Machine-learning method applied to provide the best predictive model for rock mass deformability modulus (E_m)

Enayatallah Emami Meybodi[1] · Ali DastBaravarde[2] · Syed Khaliq Hussain[1] · Somayeh Karimdost[1]

Abstract

Rock mass classification systems are used to estimate deformability modulus. This study offers a relationship to estimate deformability modulus using geomechanical rock parameters such as rock quality designation (RQD), Uniaxial Compressive strength test (UCS), joint condition (J_c), joint spacing (J_s), dynamic modulus (E_{dyn}), and groundwater condition (W_c). Dilatometer tests (DMTs) were performed on 88 samples taken from Khersan dam-II. Then, the collected data were analyzed using machine learning methods (MLM) and statistical methods. The analyses were performed using different types of MLM, i.e., Multivariate Non-Linear Regression (MNLR), Support Vector Regression (SVR), and Copula-based methods. These analyses revealed that the SVR model predicted the values of rock mass deformability modulus (E_m) with the determination coefficient (R^2) of 0.868. In comparison, the Copula-based and MNLR power methods predicted E_m with an R^2 of 0.851 and 0.866, respectively. The cross-validation (CV) technique shows that SVR is the best-designed model for predicting E_m. Also, MLMs provided an alternative and effective way to predict E_m. Overall, the statistical analysis was an efficient tool for selecting the best parameters and correlations of data for E_m.

Keywords Deformability modulus · Machine learning methods · Rock mass · Statistical methods

Introduction

The deformability modulus (E_m) is among the essential properties of intact rock, affecting the large-scale engineering behavior of rock mass. The stress-to-strain ratio during loading a rock mass, both in elastic and inelastic behavior, is known as rock mass E_m (Sousa and Grossmann 2022). E_m is commonly used as an input parameter in empirical and numerical relationships to determine the engineering behavior of rock mass (Saedi et al. 2019). This parameter provides the basis for structural design and stability analyses (Li et al.

2020; Wu et al. 2020). E_m is a factor in the design of rock structures, indicating the deformation properties of the rock mass (Aksoy et al. 2022). This parameter is determined by field and laboratory tests, such as dilatometric tests (DMTs), radial loading, and plate-jacking method. Unpredictable conditions of formation, high cost, operational problems, and time-consuming are the factors that expose some problems when conducting these tests (Ko et al. 2016). Using the standard procedures for E_m estimation in in-situ tests is very difficult. These difficulties can be overcome using regression models that directly measure E_m (Aboutaleb et al. 2018; Ceryan et al. 2021). Researchers have investigated the effect and relation of E_m with rock discontinuity (Wu et al. 2020). The E_m of rock mass is a complex and sensitive parameter that is measured using highly advanced methods and techniques. In the past two decades, machine learning methods (MLMs) have been widely used in geotechnical engineering (Abdallah 2019; Aengchuan and Phruksaphanrat 2018), tunneling and engineering geology (Ghorbani et al. 2021; Marcher et al. 2020), and for rock material and composite material. These methods have high efficiency in data processing and do not require conducting time-consuming, expensive in-situ experiments (Sun et al. 2020a;

✉ Enayatallah Emami Meybodi
  en.emami@yazd.ac.ir

  Ali DastBaravarde
  dastbaravarde@yazd.ac.ir

  Syed Khaliq Hussain
  syed.khaliq93@gmail.com

  Somayeh Karimdost
  sm_karimdoost@yahoo.com

1 Department of Geology, Yazd University, Yazd, Iran

2 Department of Statistics, Yazd University, Yazd, Iran

Xu et al. 2022). MLMs can develop a relationship between complex and high-dimensional variables (Zhang et al. 2020). Various MLM has been used in geoengineering (Zhao et al. 2018) to predict seepage behavior (Xiao and Zhao 2019), rock mechanics (Huang et al. 2021; Xia et al. 2019) slope stability (Đurić et al. 2019; Huang et al. 2022), the mechanical and physical properties of rocks (Salimi et al. 2019), and interpolation and stratification of multilayer soil property profile (Zhao and Wang 2020). In this respect, Sun et al. (2020b) showed that the support vector regression (SVR) method provides the best predictive model for determining rough rock fractures. Alemdag et al. (2015) used statistical and artificial neural network (ANN) methods to estimate the E_m. In another study, the effect and application of E_m on the fracture rock mass were estimated using the equivalent discrete fracture network (E-DFN) under uniaxial compression (Ma et al. 2020). Hasanipanah et al. (2022) applied the Levenberg–Marquardt algorithm (LMA), conjugate gradient (SCG), and Bayesian regularization (BR) algorithm in the cascaded forward neural network (CFNN) MLM to determine E_m. In another effort, four ANNs were built to determine the influences of overburden stress on E_m (Tokgozoglu et al. 2021). Fathipour-Azar (2022) applied various machine learning (ML) models to estimate triaxial rock mass strength. Several machine learning (ML) models, such as SVR, Gaussian process regression (GPR), and ANN, have been designed to predict the elastic modulus of magmatic rock (Ceryan et al. 2021). Meng et al. (2018) applied the back-propagation neural network technique to find the effective fracture propagation zone in gypsum rock. The indirect estimation is the most cost-effect technique for E_m prediction. In this respect, Fattahi (2021) developed a relevance vector regression (RVR) model for the indirect prediction of E_m. (Babanouri and Fattahi 2018) used SVR to construct a constitutive model for rock fracture to estimate the shear and peak displacement.

The present research aims to extend the use of geomechanical parameters in E_m estimation. A significance index of geo-mechanical parameters is assigned to each parameter's effect and relation with E_m. This index represents the influence of a single parameter among the entire parameters network on E_m. Statistical analyses were established to determine the best parameters for the E_m. Finally, multiple MLMs were conducted with SVR, Copula method, and multivariate nonlinear regression (MNLR) to validate the reliability of the best design model for E_m. The Copula method used in this study can exhibit the structure dependence between two or more random variables. Five dependent variables and one independent variable relationship were determined in this case. The present study is the first effort to use the Copula method to forecast E_m regarding the complications in the relationship of variables dependency. In general, this study aims to provide optimal predictive

models where the correlation between the input variables is not an unfavorable parameter. Therefore, the K-fold Cross-validation (CV) method was used to determine the models' forecasting ability. The results confirmed the capability of predictive models using these input parameters.

Study area and data collection

Khersan Dam-II is located about 2 km southwest of Lordegan city in Chaharmahal and Bakhtiari Province (Iran) on the Khersan River. The coordinates of the study area are 50 36° E and 31 25° N in the southwestern part of Zagros. Access to the dam location is possible by a 2-km long asphalted road from Lordegan through the villages of Qale-e-Madresh and Abchnar. Abchnar village is almost 6 km long from the village of Shamalk on the right bank of the river and the dam axis. The stratigraphy of the Khersan Dam-II site consists of four low-upward-folding rock units, i.e., Asmari (lower, middle, and upper), Gachsaran, Aghajari, and Quaternary deposits, including slopes, crumbs, and river uplifts. Almost all structures of the dam site except the surface power plant are located in the upper Asmari. These structures consist of thick to medium-layered limestone, with regular layering and a small percentage of thin-layer limestone to marl lime, are shown in the geological map of the dam site in (Fig. 1).

Exploratory operations were carried out in the first phase of studies of the Khersan Dam and Power Plant Design. These operations included drilling boreholes and galleries and conducting laboratory tests, including the Uniaxial Compressive Strength test (UCS) and field dilatometry test. Rock quality designation (RQD), Joint condition (J_c), joint spacing (J_s), and groundwater condition (W_c) data were obtained to determine E_m through the rock mass classification (RMR) system. This project used a flexible dilatometer and volume change measurement method to determine the E_m. Thus, within wells with a diameter of 1 mm and a predetermined depth, the dilatometer is sent as a metallic cylinder with rubber cladding. Afterward, compressed air with stepwise pressure values of 1, 2, and 3 MPa was entered into the space between the middle part and the rubber cladding in a borehole. In the meantime, borehole wall deformability values were recorded at each step. Of the 149 dilatometric test data that determine the E_m in the left and right flanks of the dam, only 88 of the dilatometric test results matched with other data from exploration boreholes and field surveys. The results of laboratory and field experiments are divided into two quantitative and qualitative forms (Table 1). In this section, experiments are conducted, and the collected results are corrected to increase the validity of probabilistic models. Next, the outlier values are removed from the final estimation calculations.
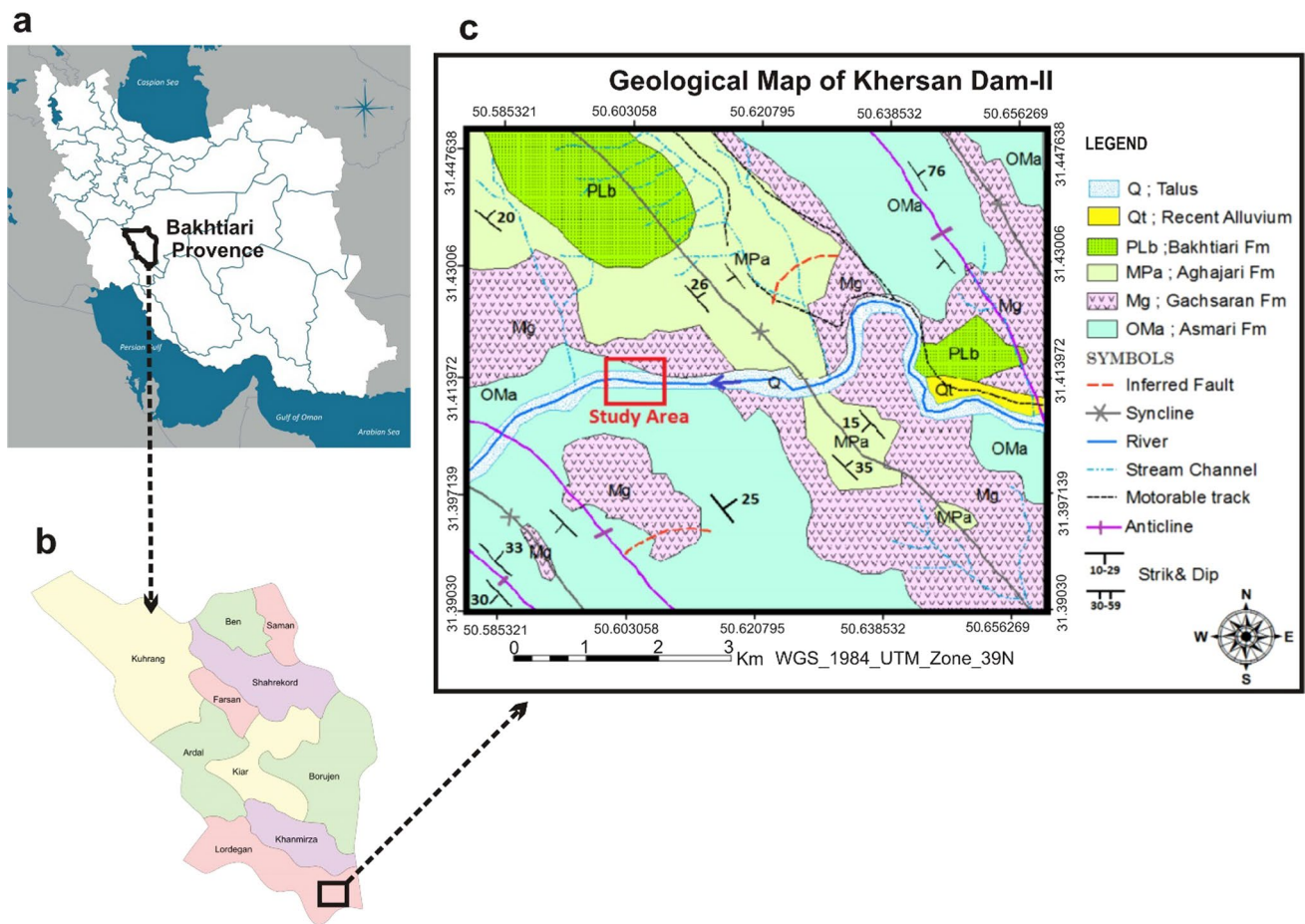
Fig. 1 **a** The location of Chaharmahal and Bakhtiari Province (Iran), **b** the location of Lordegan in the Bakhtiari Province, and **c** the Geological map of the site location of Khersan Dam-II

Table 1 Khersan Dam-II site Laboratory and Field Parameters data

Standard deviation	Average	Max	Min	Number	The parameters studied	
90.2	41.5	00.12	68.0	88	Modulus of deformation (GPa)	Dilatometer test
75.3	49.7	45.19	24.1	88	Dynamic modulus of elasticity (GPa)	Laboratory tests
73.30	34.100	230.00	59.00	88	Uniaxial compressive strength (MPa)	
7.24	17.62	100.00	19.00	88	RQD (%)	Field experiments
22.6	71.19	00.30	00.9	88	Joint Conditions	
13.5	04.12	20	5	88	Joint spacing (m)	

Methodology

Statistical methods

Today, statistical analysis is among the most commonly used tools in engineering geology, rock mechanics, civil engineering, hydrology, and petroleum engineering. In this section, we offer a relation for estimating E_m using geomechanical parameters of rock. R software was used for estimating E_m, regarding its high efficiency and ease of use for complex calculations (Charbel and Hassan 2021). One of the features of this software is to perform computational

operations such as multivariate regression, principal component analysis, and function analysis (Górecki and Smaga 2019). For this purpose, the statistical data of dilatometric experiments at the Khersan dam-II site were analyzed using R software, and standard statistical indices were used to select the appropriate regression models. The data in R software were first identified by sensitivity analysis, followed using various estimation methods to estimate this modulus.

Machine learning methods (MLMs)

In this paper, E_m is predicted using (1) the log-linear approach, (2) the SVR approach, and (3) the Copula approach. The previous section mentioned that log-linear predictive modeling could lead to better results than linear models. Recently, MLMs have offered new possibilities for solving complex problems related to regression issues. In addition, machine learning regression methods directly use data to train a model without knowledge of the predictor-target relationship (Abdallah 2019). The SVR method has good stability and generalization abilities to solve complex and high-dimensional problems (Kang and Li 2016). There are two types of support vector machine (SVM), namely (1) the function fitting performed by SVR and (2) pattern identification used by support vector classification (Kavaklioglu 2011). SVR is widely used in engineering geology and geotechnical engineering, such as for rock fracture and Blast-Induced Rock Movement (Sun et al. 2020b; Yu et al. 2020). The significance of the Copula method is handling multivariate analysis and dependence modeling. In addition, it can decompose the multivariate joint distribution function into all of its marginal (univariate) distribution functions and a Copula function (Qian et al. 2020). Copula theory plays an important role in hydrology problems, geotechnical reliability analysis, engineering geology, rock mechanics, and wind engineering. Therefore, it has received increasing interest regarding its high accuracy compared with other methods (Abdollahi et al. 2019; Gaidai et al. 2019; Ismail et al. 2018; Zhou et al. 2019). In this study, MNLR, SVR, and Copulas method were applied to develop prediction models to study the influence of the geomechanical parameters such as the UCS, RQD, joint condition (J_c), joint spacing (J_s), dynamic modulus (E_{dyn}), and groundwater condition (W_c)

on E_m. These analyses were performed using R-software. Statistical indices were used to evaluate the performance of prediction models and compared the predicted values with the actual values. These indices include root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Equations (1) and (2) give expressions of MAE and RMSE, respectively.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| \left(Y_i - \acute{Y}_i \right) \right| \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \acute{Y}_i \right)^2}. \tag{2}$$

Results

Results of statistical analysis

Descriptive statistics are applied to summarize data in an organized manner to describe the univariate analysis of data and the various relationship between them. Descriptive statistics is an essential part of data analysis that provides the foundation of inferential statistics. Tables 2 and 3 provide univariate descriptive statistics of the data. Here, Table 2 summarizes the descriptive statistics for the continuous data (i.e., E_m, J_c, UCS, RQD, and E_{dyn}). Also, Table 3 presents a frequency table for the non-continuous data (i.e., J_s and W_c). The univariate analyses show that E_m and J_c are almost symmetric, but UCS, E_{dyn}, and RQD are asymmetric (i.e., highly right-skewed, right-skewed, and left-skewed, respectively). An extreme asymmetry is observed in W_c. In addition, RQD

Table 3 Descriptive statistics for non-continuous data

J_s		W_c	
Category	Frequency	Category	Frequency
5	12	0.01	4
8	15	4	7
10	26	7	7
15	16	10	22
20	19	15	48

Table 2 Descriptive statistics for continuous data

	n	Min	Max	Mean	Median	SD	Skew	Kurtosis
E_m (GPa)	88	0.68	12.00	5.59	5.80	2.79	0.06	− 0.29
J_c	88	9.00	30.00	19.71	20.00	6.22	0.07	− 1.12
UCS (Mpa)	88	59.00	230.00	100.34	93.50	30.74	1.52	3.81
RQD (%)	88	19.00	100.00	62.17	66.00	24.71	− 0.14	− 1.41
E_{dyn} (GPa)	88	1.24	19.45	7.5	7.32	3.75	0.31	0.06

and J_c are heavy-tailed relatives of a normal distribution, but UCS is light-tailed.

Figure 2 presents the relationships between the considered data. The upper panel shows the Pearson correlation between the continuous variables and the box plots for the combinations between the continuous and the non-continuous variables. In addition, the lower panel shows the scatter plots with fitting smooth nonlinear curves between the continuous variables and the dot plots for the combinations between the continuous and the non-continuous variables. Finally, the diagonal panel represents the density plots of the continuous variables and the bar plots of the non-continuous variables. Figure 3 shows a heatmap of the correlation matrix of the dataset containing the data and the natural logarithm of continuous data. The Pearson correlation can only reflect a linear relationship of variables but ignores other

relationship types. Hence, the natural logarithm of continuous variables was considered because of the possibility of exponential nonlinear relationships. Figure 4 illustrates a heatmap of Spearman's correlation coefficient matrix of the data. Spearman correlation can determine whether there is a monotonic association between variables. Also, Spearman's correlation coefficient is invariant to monotone transformations, such as the natural logarithm transform. Analyzing the relationships between the data show a nonlinear relationship between the data. The logarithm of E_m has a stronger relationship with predictor variables than E_m. Hence, we considered the log-linear technique to model the data.

Next, we used the best subsets method to identify the best-fitting models. During regression modeling, eliminating unessential variables will make the model easier to interpret, less susceptible to data overfitting, and more
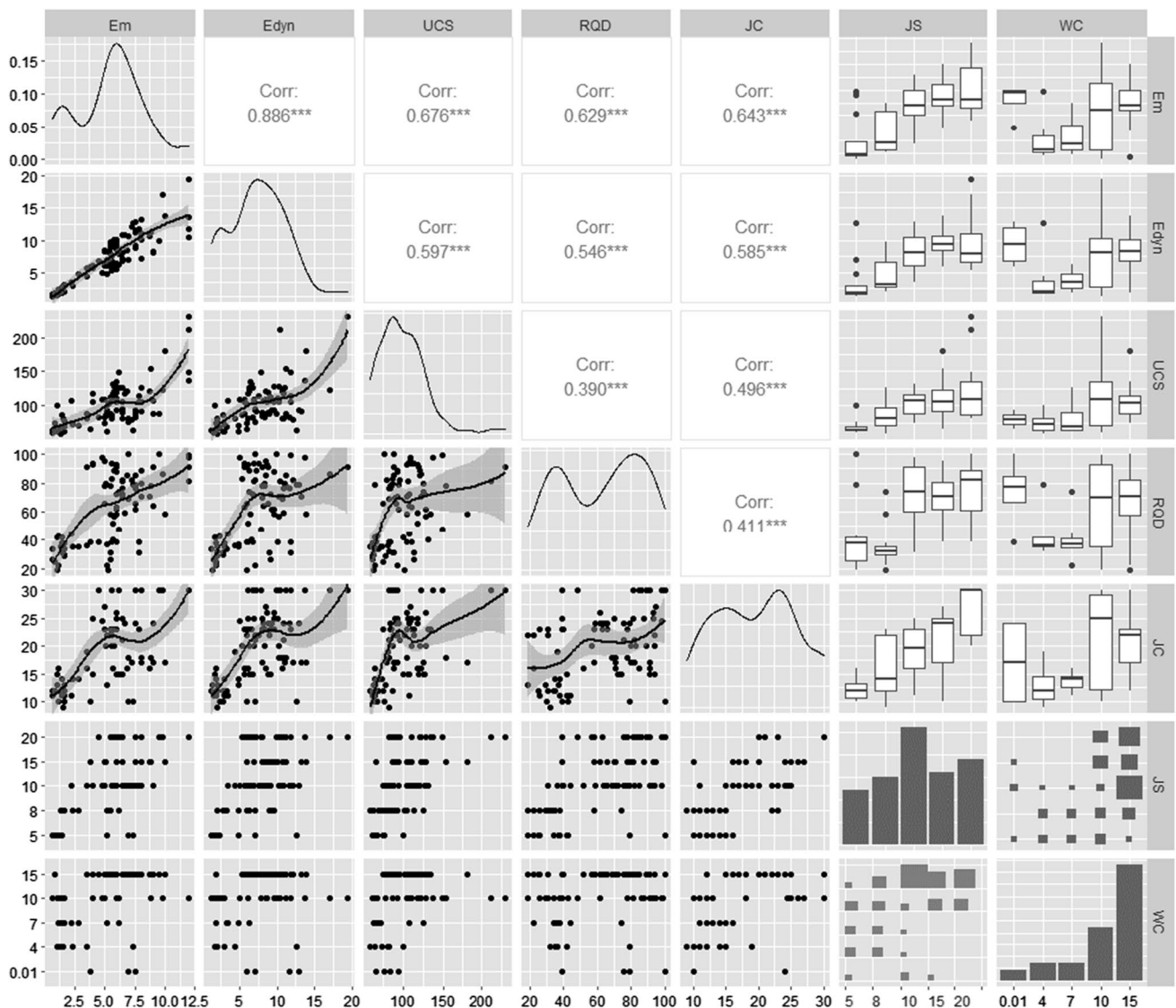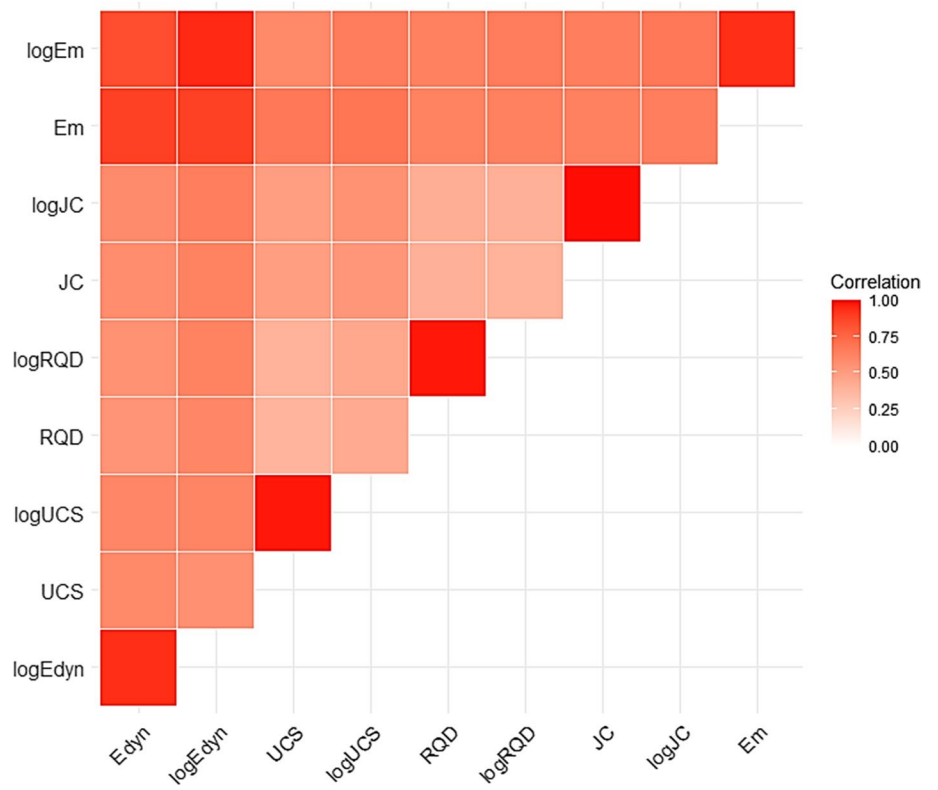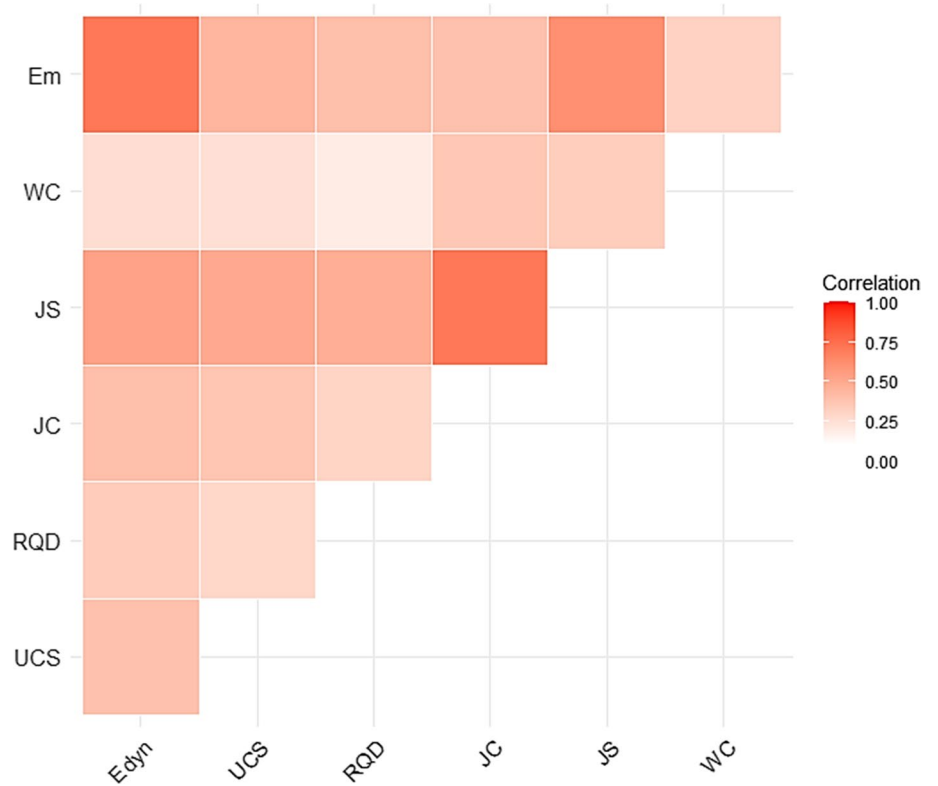

Fig. 2 The Scatterplot matrix of the data

Fig. 3 Heatmap of Pearson's correlation coefficients of the continuous data and their natural logarithm

Fig. 4 Heatmap of Spearman's correlation coefficients for E_m

generalizable. Best subsets regression is an exploratory model-building regression analysis that compares all possible models using a specified set of predictors and displays the best-fitting models. The result of the best subsets regression analysis is summarized in Table 4. We used predicted R-square (Pred.R-sq), estimated error of prediction (MSEP), and Sawa's Bayesian Information Criteria (SBIC) to identify the best-predicting models. Models with larger Pred.R-sq values or smaller MSEP or SBIC have better predictive ability. In the log-linear approach, the results of the best subset regression analysis are proposed by modeling E_m based on $\log E_{dyn}$, E_{dyn}, J_s, UCS, W_c, RQD, and J_c, in the order of their appearance. Table 4 shows that the model in the 6th row (containing predictors $\log E_{dyn}$, E_{dyn}, J_s, UCS, W_c, and RQD) and the model in the 7th row (containing predictors $\log E_{dyn}$, E_{dyn}, J_s, UCS, W_c, RQD, and J_c) are the best candidates for predicting models of E_m, in the order of their appearance.

Results of MLMs for rock mass deformability modulus (E_m)

Result of Copulas method

The Copula method was implemented using R-software to determine the E_m. As can be seen from Fig. 5, $R^2 = 0.85$ for E_m in the Copula method. E_m is a dependent variable, and the other four variables (i.e., J_c, UCS, RQD, and E_{dyn}) are independent variables. In Copula MLM for the E_m, 1 for deformability modulus (E_m), 2, 3, 4, and 5 for joint condition (J_c), Uniaxial Compressive Strength (UCS), rock quality designation (RQD), and dynamic modulus (E_{dyn}), respectively, have been used. As a result, four Copula method trees were developed. In Tree 1, pair of parameters is prepared with a ratio of one parameter and another. The pair of parameters, such as deformability modulus and dynamic modulus, provides the best result with a par value of 1.76 and a τ value of 0.69 are presented in Table 5. The graphical models of all

Copula trees are presented in Fig. 6. In Tree 1, the pair of E_m with UCS provides the second-best result with a Tau (τ) value of 0.46 (Fig. 7). Moreover, the pairs of E_m with J_c and RQD give the τ value of 0.41 and 0.41, respectively (Fig. 7). In Copula Tree 2, the pair of parameters was prepared with a ratio of one parameter and two others. In Tree 3, the pair of parameters was prepared with a ratio of two parameters and the other two parameters. Finally, in Tree 4, pair of parameters was prepared with a ratio of two parameters and the other three. It is noteworthy that all these Copula trees provided an independent relationship between the parameters are presented in Table 6.

SVR use to predicate deformability modulus

In this study, E_m was also predicted using the SVR method. The model was constructed by the e1071 package, which is the first and most intuitive package in R software. The UCS, RQD, J_c, J_s, E_{dyn}, and W_c were selected as the model's inputs. The SVR model was constructed based on the kernel function type of the Radius Basis Function (RBF). The kernel function can transform data from a nonlinear to a linear form. The SVR results are presented in Table 9. As can be seen, the models can accurately predict E_m with $R^2 = 0.868$ (Fig. 8). Overall, three SVR models were developed for E_m are summarized in Table 7.

Predicating deformability modulus by MNLR

Power Multivariate Nonlinear regression was applied to determine the effect of geomechanical parameters on the E_m. In the MNLR method, E_m is a dependent variable, and UCS, RQD, J_c, J_s, E_{dyn}, and W_c are independent variables. A total of 21 MNLRs were developed for E_m, in which the hybrid MNLR model provides the best results. The hybrid MNLR model is the second-best model among all the machine learning models and provides a relationship, as shown in Eq. (3).

Table 4 Best subsets of regression

Number of predictors	Predictors									Selection criteria		
	$\log E_{dyn}$	E_{dyn}	J_s	UCS	W_c	RQD	J_c	$\log J_c$	logUCS	Pred. R-sq	MSEP	SBIC
1	*									0.899	4.276	− 263.579
2	*	*								0.913	3.622	− 276.911
3	*	*	*							0.916	3.187	− 286.519
4	*	*	*	*						0.924	2.884	− 293.354
5	*	*	*	*	*					0.919	2.740	− 295.999
6	*	*	*	*	*	*				0.921	2.668	− 296.591
7	*	*	*	*	*	*	*			0.921	2.629	− 296.122
8	*	*	*	*	*	*		*	*	0.918	2.637	− 294.367
9	*	*	*	*	*	*	*	*	*	0.916	2.671	− 291.923

Pred. R-sq predicted R-square of the model, *MSEP* estimated error of prediction, assuming multivariate normality, *SBIC* Sawa's Bayesian Information Criteria

Fig. 5 Scatter plots of the predicted E_m from the Copula method

$$y = 0.79 + 0.89\,x$$
$$R^2 = 0.85$$
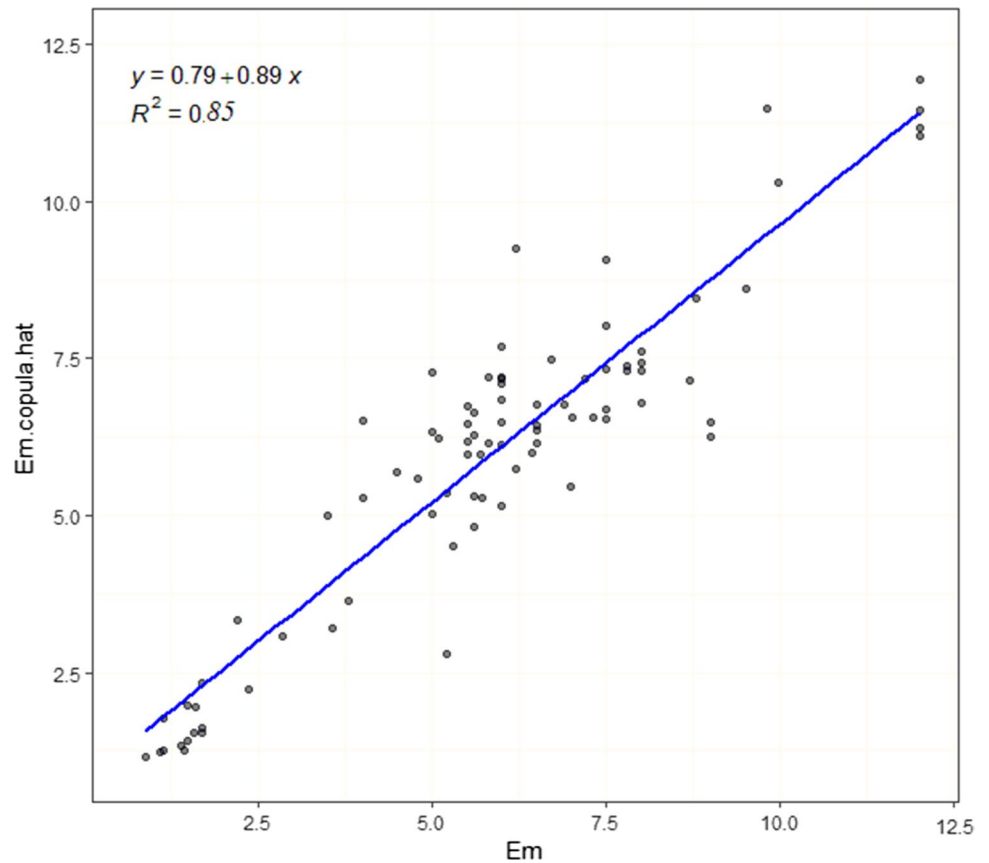
Table 5 The results of the Copula Tree 1 for E_m

Tree No. 1				
Pair	C-vine	Par-1	Par-2	Tau
1,5	BB7	1.76	4.62	0.69
1,4	Survival BB8	2.96	0.88	0.41
1,3	BB7	1.61	0.46	0.46
1,2	BB7	1.57	0.41	0.41

Furthermore, the results of the hybrid MNLR are very near to SVR. The results of nonlinear regression analysis to predict E_m ($R^2 = 0.866$) are presented in Fig. 9.

$$E_m = 0.4675 \times Edyn^{1.3641} \times 0.9220^{Edyn} \times 1.0023^{UCS} \times 1.1262^{I(JS=20)} \times 1.0020^{RQD}. \qquad (3)$$

Comparisons between MLMs

One of the significant challenges in MLMs is to check the model accuracy of unseen data to know whether the designed model performs well. To evaluate the model's accuracy, we need to test it against those data points that were not present during the model training. Using R-programming language, CV is one of the best methods for checking the accuracy of the machine learning model. CV is a standardized technique for testing the performance of a predictive model. In this process, a part of the data set is saved, which will not be used in the model training. When the model is ready, this specific dataset is used for testing purposes. The CV method is divided into two types: (1) non-exhaustive CV (e.g., K-fold CV, holdout method, repeated random sub-sampling validation) and (2) exhaustive CV included (e.g., leave-p-out CV and leave-one-out CV).

The K-fold CV method was used to find the best-designed machine-learning models for E_m. This method is one of the most accurate and reliable methods for testing machine learning models. The CV technique divides the data into equal K subsets (folds). Of these K-folds, one subset is used as a validation set, and the remaining is applied in the training model. The K-fold CV method was used for SVR, Copula Method, and MNLR to find the best MLM for E_m. For these MLMs, fitting models were built by the K-fold CV technique. In the K-fold CV method, 25 models were constructed, including 19 MNLR, 1 Copula method model, and 3 SVR models. Model training was based on 84 observations and tests of

Fig. 6 The graphical models of all Copula trees for E_m
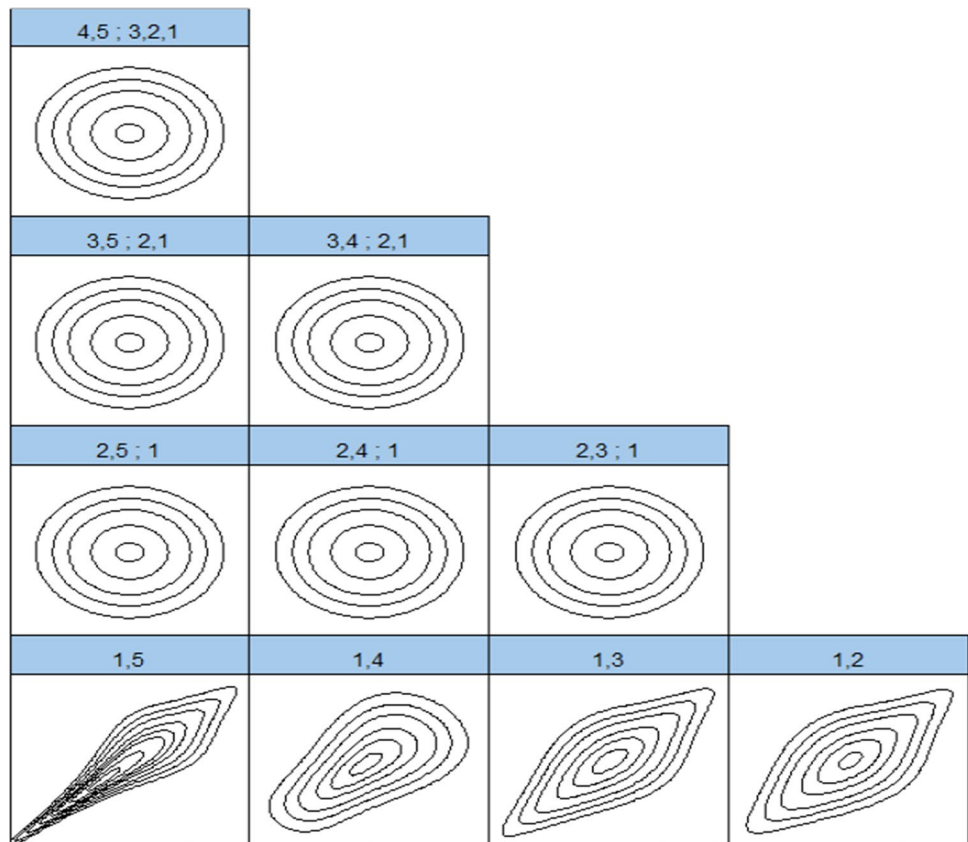
Fig. 7 The results of the Copula Trees 1 and 2 for E_m
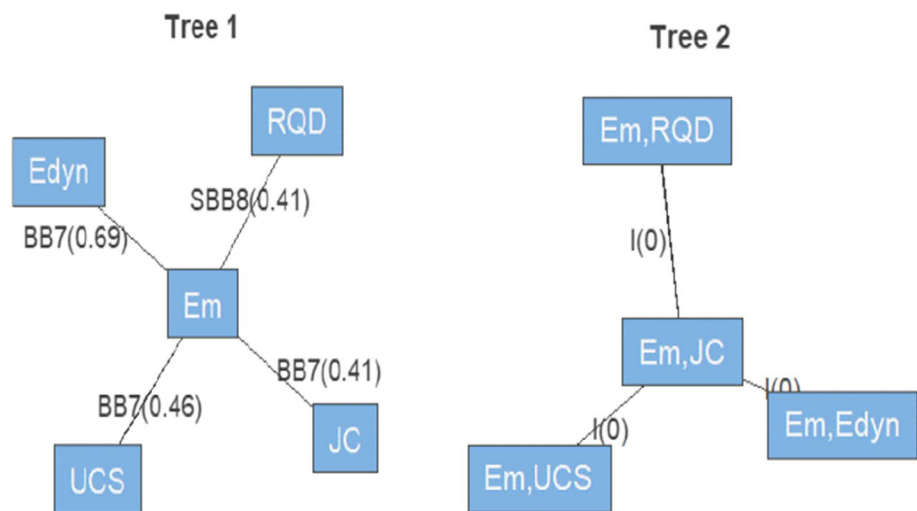
fitting model 4 that were left out randomly (K-fold CV, where $K = 4$). The number of repetitions of the process is also considered 100. Results showed that the SVM Model 25 is the best-designed model for the E_m. However, this model's output is very slightly different from SVR Model 24. Moreover, the hybrid-MNLR Model 17 is the second-best-designed machine learning model for E_m (Fig. 10). The Copula method model did not provide good results for E_m. Besides, the coefficient of variation (ratio of standard deviation to average) of R^2, RMSE, and MAE factors in 100 repetitions of K-fold CV is 0.004, 0.013, and 0.009, respectively, summarized in Table 8.

In the following, we outlined the overall performance of various models presented in this work in predicting the E_m values based on RMSE. Table 9 summarizes the comparison of all statistical indices (MSE, RMSE, and MAE) for different models.

**Table 6** The results of the Copula Trees 2, 3, and 4 for E_m

Tree-2		Tree-3		Tree-4	
Pair	C-vine	Pair	C-vine	Pair	C-vine
2,5;1	Independence	3,5;2,1	Independence	4,5;3,2,1	Independence
2,4;1	Independence	3,4;2,1	Independence		
2,3;1	Independence				

Conclusion

Rock mass deformability is one of the most demanding, sensitive, and complicated subjects in engineering geology,

civil engineering, rock mechanics, and petroleum engineering. In this respect, it is necessary to determine the effect of different geomechanical parameters, e.g., Uniaxial Compressive Strength test (UCS), rock quality designation (RQD), joint condition (J_c), joint spacing (J_s), dynamic modulus (E_{dyn}), and groundwater condition (W_c), on the E_m. For this purpose, different modern and reliable machine learning and statistical methods were used. Statistical analysis is a straightforward technique to select the best parameters and the quantitative analysis of E_m data. These methods are significantly less time-consuming and low-cost than other methods, such as In-situ field tests. The SVR model provided more realistic and correct results for E_m. In comparison, Copula methods for E_m do not provide excellent results due to less data availability. Multivariate Non-Linear Regression (MNLR) is the second-best machine learning model for

Fig. 8 Scatter plots predicted by SVR for E_m

$$y = 0.75 + 0.85\,x$$
$$R^2 = 0.86$$
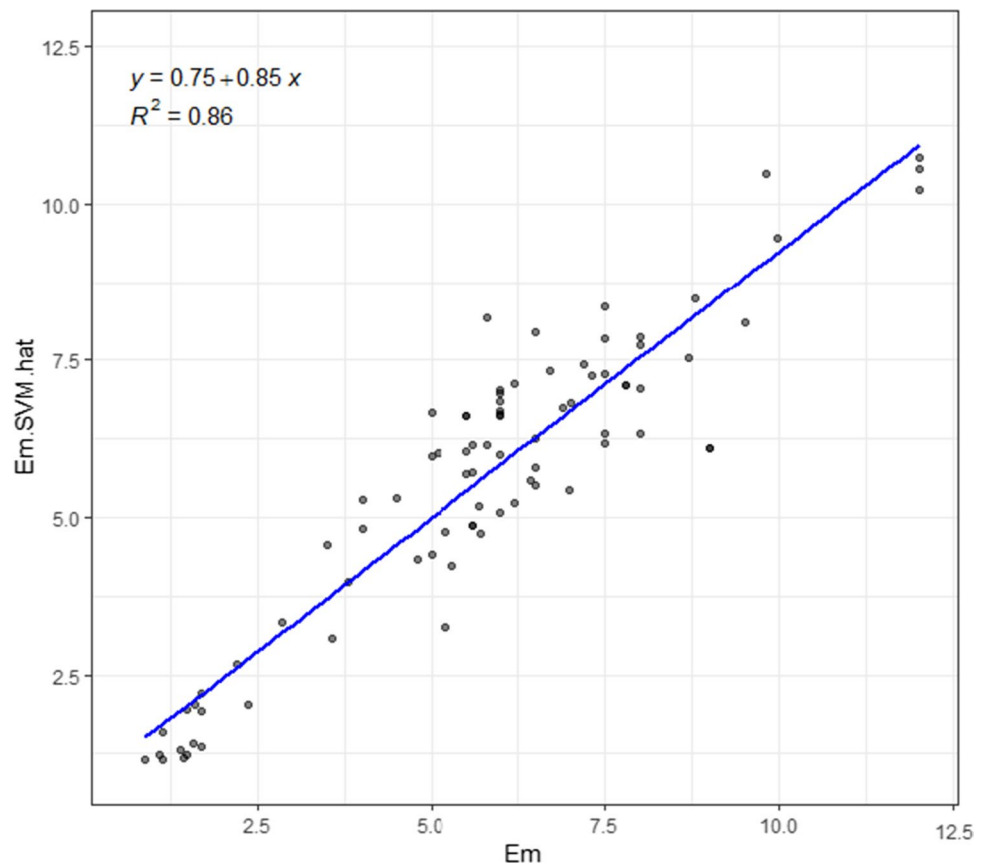
Table 7 Results of SVR model for E_m

Model	SVM (Types)	Parameters	Cost	Epsilon
SVM-model-1	eps-regression SVM-Kernel: radial	J_s, W_c, J_c, UCS, RQD, E_{dyn},	7	0.2
SVM-model-2	eps-regression SVM-Kernel: linear	J_s, W_c, E_{dyn}, logE_{dyn}, UCS, logUCS, RQD, logRQD, J_c, logJ_c	14	0.3
SVM-model-3	eps-regression SVM-Kernel: linear	E_{dyn}, log E_{dyn}, UCS, J_{s20}, RQD	4	0.2

Fig. 9 Scatter plots predicted by hybrid MNLR for E_m
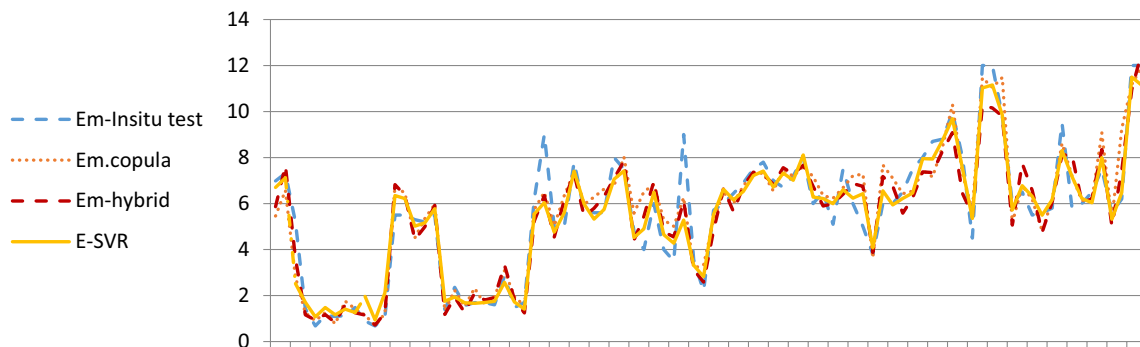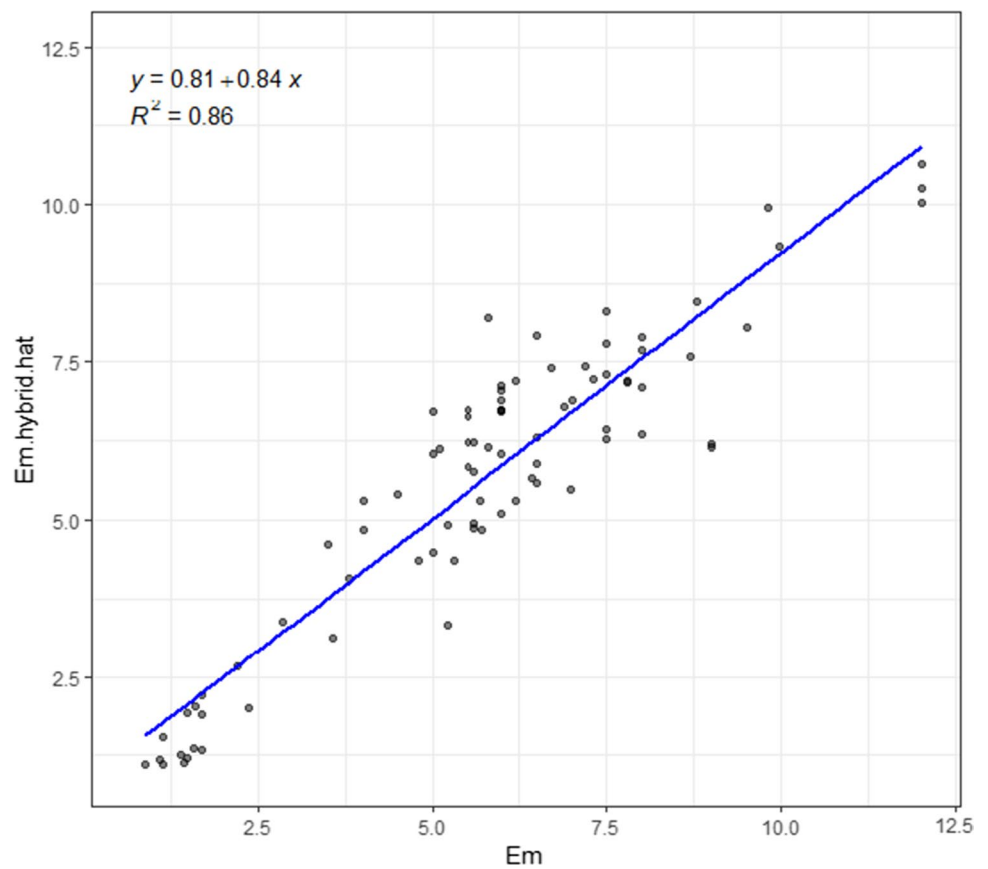
Fig. 10 The comparison of all machine learning models for E_m

Table 8 K-fold cross-validation results of all machine learning models

	M.1	M.2	M.3	M.4	M.5	M.6	M.7	M.8	M.9	M.10	M.11	M.12
R^2	0.584	0.777	0.791	0.785	0.803	0.785	0.784	0.798	0.773	0.796	0.787	0.801
RMSE	2.199	2.146	1.374	1.367	1.321	1.41	1.411	1.345	1.394	1.357	1.36	1.338
MAE	1.624	1.553	0.987	0.978	0.961	0.991	1.005	0.988	1.085	0.977	0.97	0.966

	M.13	M.14	M.15	M.16	M.17	M.18	M.19	M.20	M.21	Copula M.22	SVR1 M.23	SVR2 M.24	SVR3 M.25
R^2	0.782	0.793	0.844	0.857	0.866	0.86	0.826	0.848	0.849	0.851	0.867	0.834	0.868
RMSE	1.42	1.266	1.101	1.051	1.017	1.039	1.179	1.087	1.08	1.074	1.016	1.139	1.014
MAE	1.006	0.969	0.817	0.795	0.8	0.783	0.826	0.841	0.833	0.817	0.79	0.865	0.796

Table 9 Comparison of the best final models

Model	Statistical indices for the model	Characteristics	Total rank
Hybrid-MNLR	$R^2 = 0.866$ RMSE = 1.017 MAE = 0.8	Function: Exponential form	2
Support Vector Regression (SVR)	$R^2 = 0.868$ RMSE = 1.014 MAE = 0.796	SVM-Type: eps-regression SVM-Kernel: radial cost:1 gamma: 0.3333333 epsilon: 0.1 Number of Support Vectors:23	1
Copula Method	$R^2 = 0.851$ RMSE = 1.074 MAE = 0.817	C-vine Copula method	3

predicting E_m. The results showed that E_m directly correlates with a dynamic modulus (E_{dyn}) compared to the UCS, RQD, and J_c. Furthermore, E_m has the weakest relationship with geomechanical parameters such as J_s and W_c. It is noteworthy that both J_s and W_c are classified as non-continuous. Overall, this research effectively predicts the E_m by solving the high-dimensional and nonlinear relationship between the E_m and the mentioned parameters.

Author contributions All authors in this research article contributed very well related to their relevant parts.

Declarations

References

Abdallah A (2019) Prediction of the soil water retention curve from basic geotechnical parameters by machine learning techniques. In: International Conference on Inforatmion technology in Geo-Engineering. Springer, pp 383–392

Abdollahi S, Akhoond-Ali AM, Mirabbasi R, Adamowski JF (2019) Probabilistic event based rainfall-runoff modeling using copula functions. Water Resour Manag 33:3799–3814

Aboutaleb S, Behnia M, Bagherpour R, Bluekian B (2018) Using non-destructive tests for estimating uniaxial compressive strength and static Young's modulus of carbonate rocks via some modeling techniques. Bull Eng Geol Environ 77:1717–1728

Aengchuan P, Phruksaphanrat B (2018) Comparison of fuzzy inference system (FIS), FIS with artificial neural networks (FIS+ ANN) and FIS with adaptive neuro-fuzzy inference system (FIS+ ANFIS) for inventory control. J Intell Manuf 29:905–923

Aksoy CO, Aksoy GGU, Yaman HE (2022) The Importance of deformation modulus on design of rocks with numerical modeling. Geomech Geophys Geo-Energy Geo-Resour 8:1–23

Alemdag S, Gurocak Z, Gokceoglu C (2015) A simple regression based approach to estimate deformation modulus of rock masses. J Afr Earth Sci 110:75–80

Babanouri N, Fattahi H (2018) Constitutive modeling of rock fractures by improved support vector regression. Environ Earth Sci 77:1–13

Ceryan N, Ozkat EC, Korkmaz Can N, Ceryan S (2021) Machine learning models to estimate the elastic modulus of weathered magmatic rocks. Environ Earth Sci 80:1–24

Charbel L, Hassan HEH (2021) Mudflow modeling using Flow-R software: case study of Ras Baalbek basin (Lebanon). Geo-Eco-Trop 45:475–486

Đurić U, Marjanović M, Radić Z, Abolmasov B (2019) Machine learning based landslide assessment of the Belgrade metropolitan area: Pixel resolution effects and a cross-scaling concept. Eng Geol 256:23–38

Fathipour-Azar H (2022) Hybrid machine learning-based triaxial jointed rock mass strength. Environ Earth Sci 81:1–11

Fattahi H (2021) Applying optimized relevance vector regression approach for indirect forecasting rock mass deformation modulus. Environ Earth Sci 80:1–10

Gaidai O, Naess A, Xu X, Cheng Y (2019) Improving extreme wind speed prediction based on a short data sample, using a highly correlated long data sample. J Wind Eng Ind Aerodyn 188:102–109

Ghorbani E, Moosavi M, Hossaini MF, Assary M, Golabchi Y (2021) Determination of initial stress state and rock mass deformation modulus at Lavarak HEPP by back analysis using ant colony optimization and multivariable regression analysis. Bull Eng Geol Environ 80:429–442

Górecki T, Smaga Ł (2019) fdANOVA: an R software package for analysis of variance for univariate and multivariate functional data. Comput Stat 34:571–597

Hasanipanah M, Jamei M, Mohammed AS, Amar MN, Hocine O, Khedher KM (2022) Intelligent prediction of rock mass deformation modulus through three optimized cascaded forward neural network models. Earth Sci Inform 15:1–11

Huang M, Hong C, Chen J, Ma C, Li C, Huang Y (2021) Prediction of peak shear strength of rock joints based on back-propagation neural network. Int J Geomech 21:04021085

Huang M, Weng H, Hong C, Xu X, Tao Z, Li C, Huang Y (2022) Novel intelligent approach for the early warning of rainfall-type landslides based on the BRB model. Int J Geomech 22:06022027

Ismail T, Ahmed K, Alamgir M, Kakar MN, Fadzil AB (2018) Bivariate flood frequency analysis using Gumbel copula. MJCE. https://doi.org/10.11113/mjce.v30n2.474

Kang F, Li J (2016) Artificial bee colony algorithm optimized support vector regression for system reliability analysis of slopes. J Comput Civ Eng 30:04015040

Kavaklioglu K (2011) Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Appl Energy 88:368–375

Ko J, Jeong S, Lee JK (2016) Large deformation FE analysis of driven steel pipe piles with soil plugging. Comput Geotech 71:82–97

Li G, Li G, Wang Y, Qi S, Yang J (2020) A rock physics model for estimating elastic properties of upper Ordovician-lower Silurian mudrocks in the Sichuan Basin. China Eng Geol 266:105460

Ma G, Li M, Wang H, Chen Y (2020) Equivalent discrete fracture network method for numerical estimation of deformability in complexly fractured rock masses. Eng Geol 277:105784

Marcher T, Erharter GH, Winkler M (2020) Machine Learning in tunnelling–Capabilities and challenges. Geomech Tunn 13:191–198

Meng T, Bao X, Zhao J, Hu Y (2018) Study of mixed mode fracture toughness and fracture characteristic in gypsum rock under brine saturation. Environ Earth Sci 77:1–25

Qian L, Wang X, Wang Z (2020) Modeling the dependence pattern between two precipitation variables using a coupled copula. Environ Earth Sci 79:1–12

Saedi B, Mohammadi SD, Shahbazi H (2019) Application of fuzzy inference system to predict uniaxial compressive strength and elastic modulus of migmatites. Environ Earth Sci 78:1–14

Salimi A, Rostami J, Moormann C (2019) Application of rock mass classification systems for performance estimation of rock TBMs using regression tree and artificial intelligence algorithms. Tunn Undergr Space Technol 92:103046

Sousa LRE, Grossmann N (2022) Safety and environmental issues in rock engineering, volume 2: Proceedings/Comptes-rendus/Sitzungsberichte/ISRM international symposium, EUROCK'93, Lisbon, 21–24 June 1993, 2 volumes. Taylor & Francis, pp 1119-1125

Sun Y, Li G, Zhang J (2020a) Developing hybrid machine learning models for estimating the unconfined compressive strength of jet grouting composite: a comparative study. Appl Sci 10:1612

Sun Z, Wang L, Zhou J-Q, Wang C (2020b) A new method for determining the hydraulic aperture of rough rock fractures using the support vector regression. Eng Geol 271:105618

Tokgozoglu K, Aladag C, Gokceoglu C (2023) Artificial neural networks to predict deformation modulus of rock masses considering overburden stress. Geomech Geoengin 18:48–64

Wu F, Deng Y, Wu J, Li B, Sha P, Guan S, Zhang K, He K, Liu H, Qiu S (2020) Stress–strain relationship in elastic stage of fractured rock mass. Eng Geol 268:105498

Xia C, Huang M, Qian X, Hong C, Luo Z, Du S (2019) Novel intelligent approach for peak shear strength assessment of rock joints on the basis of the relevance vector machine. Math Probl Eng

Xiao F, Zhao Z (2019) Evaluation of equivalent hydraulic aperture (EHA) for rough rock fractures. Can Geotech J 56:1486–1501

Xu Q, Ouyang C, Jiang T, Yuan X, Fan X, Cheng D (2022) MFFENet and ADANet: a robust deep transfer learning method and its application in high precision and fast cross-scene recognition of earthquake-induced landslides. Landslides 19:1–31

Yu Z, Shi X, Zhou J, Chen X, Miao X, Teng B, Ipangelwa T (2020) Prediction of blast-induced rock movement during bench blasting: use of gray wolf optimizer and support vector regression. Nat Resour Res 29:843–865

Zhang P, Yin Z-Y, Jin Y-F, Chan TH (2020) A novel hybrid surrogate intelligent model for creep index prediction based on particle swarm optimization and random forest. Eng Geol 265:105328

Zhao T, Wang Y (2020) Interpolation and stratification of multilayer soil property profile from sparse measurements using machine learning methods. Eng Geol 265:105430

Zhao T, Hu Y, Wang Y (2018) Statistical interpretation of spatially varying 2D geo-data from sparse measurements using Bayesian compressive sampling. Eng Geol 246:162–175

Zhou X, Zhang G, Hu S, Li J (2019) Optimal estimation of shear strength parameters based on copula theory coupling information diffusion technique. Adv Civ Eng. https://doi.org/10.1155/2019/8738969

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.