




Predicting Young's modulus of Indian coal measure rock using multiple regression and artificial neural network

Author(s) ORCID Identifier:

Sayantana Chakraborty:  0000-0003-2131-1249

Rohan Bisai:  0000-0001-8716-0725

Rohit Roy:  0000-0002-6567-1185

Sathish Kumar Palaniappan:  0000-0002-6799-9918

Samir Kumar Pal:  0000-0002-3140-6826

Karanam Uma Maheshwar Rao:  0000-0002-1022-5372

Follow this and additional works at: <https://jsm.gig.eu/journal-of-sustainable-mining>



Part of the [Explosives Engineering Commons](#), [Oil, Gas, and Energy Commons](#), and the [Sustainability Commons](#)

Recommended Citation

Chakraborty, Sayantan; Bisai, Rohan; Roy, Rohit; Palaniappan, Sathish Kumar; Pal, Samir Kumar; and Rao, Karanam Uma Maheshwar () "Predicting Young's modulus of Indian coal measure rock using multiple regression and artificial neural network," *Journal of Sustainable Mining*: Vol. 22 : Iss. 1 , Article 4. Available at: <https://doi.org/10.46873/2300-3960.1373>

This Research Article is brought to you for free and open access by Journal of Sustainable Mining. It has been accepted for inclusion in Journal of Sustainable Mining by an authorized editor of Journal of Sustainable Mining.

Predicting Young's modulus of Indian coal measure rock using multiple regression and artificial neural network

Abstract

Accurate information on Young's modulus (E) is required for simulating rock deformation in mines; on the other hand, it is very cumbersome to obtain in the laboratory and collecting drilled cores in sufficient amounts, especially in the case of soft rocks, is quite impossible. Empirical equations were deduced for E from easily determinable rock properties, and the final model was selected through different statistical strength parameter tests. The generalization of the equation was verified through the normal distribution tests of residues of the equation. R² came to be 0.609 and was validated using an artificial neural network with an improved value of 0.73

Keywords

sandstone; shale; multiple regression; outlier analysis; artificial neural network

Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Authors

Syantana Chakraborty, Rohan Bisai, Rohit Roy, Sathish Kumar Palaniappan, Samir Kumar Pal, and Karanam Uma Maheshwar Rao

Predicting Young's Modulus of Indian Coal Measure Rock Using Multiple Regression and Artificial Neural Network

Sayantana Chakraborty^{a,*}, Sathish Kumar Palaniappan^a, Rohan Bisai^b, Rohit Roy^a, Samir Kumar Pal^a, Karanam Uma Maheshwar Rao^c

^a Indian Institute of Technology, Department of Mining Engineering, Kharagpur, India

^b Techno Main Salt Lake, Department of Civil Engineering, Kolkata, India

^c National Institute of Technology Rourkela, India

Abstract

Accurate information on Young's modulus (E) is required for simulating rock deformation in mines; on the other hand, it is very cumbersome to obtain in the laboratory and collecting drilled cores in sufficient amounts, especially in the case of soft rocks, is quite impossible. Empirical equations were deducted for E from easily determinable rock properties, and the final model was selected through different statistical strength parameter tests. The generalization of the equation was verified through the normal distribution tests of residues of the equation. R^2 came to be 0.609 and was validated using an artificial neural network with an improved value of 0.73.

Keywords: sandstone, shale, multiple regression, outlier analysis, artificial neural network

1. Introduction

The different aspects of rock engineering, like the design and construction of tunnels, underground structures, underground pillar stability of mines, foundations on rocks, rock slope stability etc., have two basic strength parameters to be incorporated. The major strength parameters of the rock are uniaxial compressive strength (UCS) and Young's modulus (E). For RMR classification, the only strength parameter to assess rock material characteristics is UCS [1]. One of the main problems for detecting UCS and Young's modulus from the standard method is to use of sophisticated instruments and destruction methods [2]. Obtaining drilled core insufficient amounts, especially in the case of soft, fragile, or foliated rocks, is quite impossible [3]. The knowledge of the in-situ stresses is required for the quantification of fracture propagation and fracture mapping in the rock mass. Young's modulus is required as an input parameter for numerical studies such as the finite element

method for the estimation of the in-situ stresses. The numerical studies provide a more accurate estimation of the in-situ stresses than the hydraulic fracturing and flat jack method. These studies also validate the measurements taken by hydraulic fracturing and flat jack methods. It is also used in other applications such as drilling and requires avoiding the well bore instability problems. "These parameters have great importance in rock physics applications, viz., onshore and offshore geo-mechanical engineering, tunnelling, dam design, rock drilling and blasting, rock excavation and even for slope stability" [4]. Many researchers have introduced empirical equations derived through either simple or multiple regression analysis to estimate the UCS of various rocks [5–13]. Different studies have been performed using different rock types and properties, leading to different empirical relations for different rocks. Further multiple regression techniques have been used for finding UCS using at least two parameters [14–20]. Young's modulus is essential for the estimation of the in-situ

Received 27 June 2022; revised 1 September 2022; accepted 1 September 2022.
Available online 3 January 2023

* Corresponding author at:
E-mail address: schuckjuviv@gmail.com (S. Chakraborty).

<https://doi.org/10.46873/2300-3960.1373>

2300-3960/© Central Mining Institute, Katowice, Poland. This is an open-access article under the CC-BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

stresses. Prior knowledge of in-situ stresses is required to design hydraulic fracturing. It is also essential in oil and gas industry applications. Moreover, fracture propagation and mapping are strong functions of the values and directions of the in-situ stresses [21]. Artificial neural networks and multiple regressions have been used to predict UCS for travertine sandstone samples [22–24]. Also, the correlation of physic–mechanical properties of rock from seismic waves is investigated to predict UCS and Young's modulus [25–27]. Some researchers have worked on the statistical estimation of UCS from Shore hardness using genetic expression programming [28]. Sometimes, the machine stiffness also comes into play, which leads to erroneous results of Young's modulus whilst measuring from the laboratory machines [29–31]. It has been increasingly common to develop alternatives that would predict the elastic moduli of rocks using theoretical and empirical approaches in order to get around the challenges associated with measuring the elastic characteristics of rocks through laboratory experiments. These methods either correlate the elastic and shear moduli to simple physical characteristics like density, and indirect tensile strength, or other mechanical characteristics like indices, or even to the mineral constitution of the rocks [32]. These properties are determined in the laboratory or documented in the literature. In the following paper, multiple regressions have been used to form an expression for Young's modulus using other easily determinable physic mechanical properties of rocks, viz. density, depth, uniaxial compressive strength, and tensile strength. Trial equations were carried out by running multiple regressions, and a preferred equation containing all the parameters was achieved with a fair coefficient of determination (see Tables 6 and 7).

2. Materials and methods

The physico-mechanical properties of coal measure rocks were selected for an Indian coal mine. Samples were collected from boreholes of the Salanpur block in the Asansol region up to a maximum depth of 440 m. Different coal measure rocks, such as coarse-grained, medium-grained and fine-grained sandstones, shale, carbonaceous shale, sandy shale etc., were subjected to uniaxial compression and indirect tensile tests. The samples were prepared according to the ISRM norms. The rate of stress was held constant at 0.5 MPa/s. The cores were sawed cut into uniaxial and Brazilian test samples according to ISRM norms. The cylindrical

Abbreviations

UCS	uniaxial compressive strength
UTM	universal testing machine
ITS	indirect tensile strength
D	depth
ρ	density

samples were polished before testing. The samples were air-dried to maintain the constant mass. For UCS samples, different precautions were taken. The sides' parallelism was verified, and both surface roughness and perpendicularity stayed within tolerances of 0.01 rad and 0.02 mm, respectively. At 0.5 MPa/s, the loading rate is maintained constant [33]. A total of 264 samples were prepared for the present study. In this paper, the mechanical test of rocks involves the determination of UCS. In this investigation, a universal testing machine (UTM) of INSTRON brand and SATEC series KN type was employed. Every experiment was carried out at room temperature [34].

2.1. Determination of compressive strength

UCS is the compressive stress at which the sample fails were calculated using the following formula. For the determination of uniaxial compression strength, an INSTRON, SATEC series KN model, universal testing machine (UTM), is used in this study. All the experiments are performed at normal room temperature. UCS is the compressive stress at which the sample fails and is calculated using the following formula:

$$UCS = \frac{P_{\max}}{(\pi D_1^2)/4} \quad (1)$$

where, P_{\max} is the load at failure, and D_1 is the diameter of the sample.

2.2. Determination of indirect tensile strength

BTS of the sample is calculated by dividing the maximum load carried by the sample during the test by the contact area of the sample. BTS is calculated as described below:

$$ITS = \frac{2P'_{\max}}{\pi D_1 L} \quad (2)$$

where, P_{\max} is the load at failure during the test, D_1 is the diameter of the sample, and L is axial length of the sample.

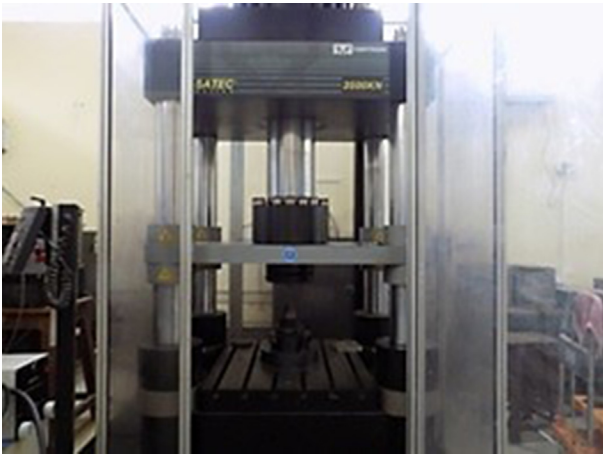


Fig. 1. Universal testing machine (Instron).

The experiments were carried out in INSTRON of 3500 KN capacity in the Rock Mechanics Laboratory of IIT, Kharagpur. The instrument is shown in Fig. 1.

3. Results and discussion

The statistical relationship (particularly the mean and variance) between a dependent variable as a function of one or more independent (or control) variables may be determined empirically through regression analysis. Multiple linear regressions also find many applications in engineering. It correlates two or more independent (or control) variables with a dependent variable. The equation comes in the form:

$$E = (x_1, x_2, x_3, x_4) = k + ax_1 + bx_2 + cx_3 + dx_4 \quad (3)$$

3.1. Multiple regression analysis

As increasing parameters strengthen the correlation, multiple regression analysis is inevitable to carry out with the study. The software SPSS 25 has been used for this purpose. First, the histograms (Fig. 2) of the data of the different parameters, viz. UCS, ITS, density (ρ), and depth (D), are drawn to depict the spatial distribution of the data about the mean. All variables are quantitative and interval types. There is also less co-linearity amongst each other, i.e., the value of the Pearson coefficients is less than 1 when each of UCS, ITS, depth and density are correlated [35]. The Pearson coefficient is the standardized coefficient denoting the strength of the statistical relationship between two parameters. It is given by:

$$r = \frac{\text{cov}(x, y)}{(N - 1)s_x s_y} \quad (4)$$

where cov_{xy} is the covariance of the data set, s_x and s_y are the standard deviations of each of the x and y data sets, where x is the independent and y are the dependent data set. In this study, the correlation strength is determined by the coefficient of determination, i.e., the square of the Pearson coefficient (R^2). The primary assumptions of multiple regressions are henceforth met for the above data. Forty-two (42) trial equations are derived to obtain a relatively coherent relation between Young's modulus, E (dependent variable) and the predictors. The log-transformed E is used for regression. In general, multiple regression provides better results if the variable data are normally distributed. In this present case, the data of Young's modulus is not normally distributed. So, to improve the result, the log-transformed Young's modulus has been used. The normality distribution is shown in Fig. 2. The normal curve of log-transformed Young's modulus is more uniformly distributed around the dataset than the normal curve of Young's modulus data. Different combinations of UCS, ITS, density, and depth are used in the analysis to reach an equation that would predict the value of Young's modulus with the different types of physico-mechanical properties of rock.

The general form of the linear equation comes to be

$$E = k + a\text{UCS} + b\text{ITS} + cD + d\rho \quad (5)$$

where, a , b , c , d are the regression coefficients, E is Young's modulus of the rock (in GPa), UCS is the uniaxial compressive strength of the rock (in MPa), ITS is the indirect tensile strength of the rock (in MPa), ρ is the density of rock (in KN/m^3), and D is the depth of occurrence in m. Outlier analysis has been done to avert the geological disturbances in the lithology of the boreholes. The main aim of finding the correlation of Young's modulus is to find a simple equation to make it easy to handle for any geotechnicians. So, different types of trial equations have been used to find a suitable equation that would be at par with the requirements for a coherent equation. The models then underwent different statistical parameters tests. Three types of empirical equations are considered for the estimation of Young's modulus. They are linear, log-linear and mixed linear equations (i.e., logarithmic and linear parameters). The equations are selected on the basis of the highest value of the correlation

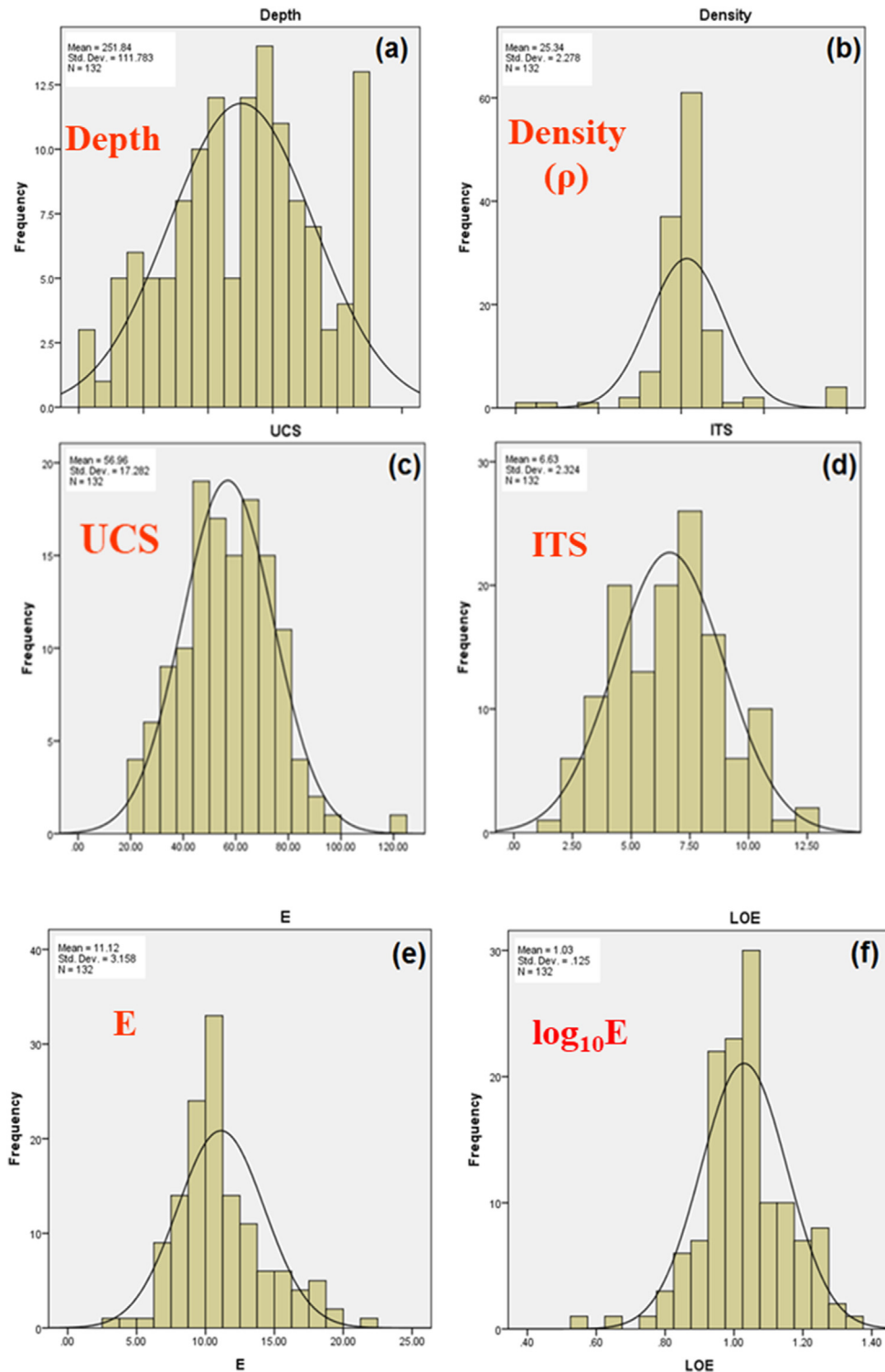


Fig. 2. Histogram of (a) depth, (b) density, (c) UCS, (d) ITS, (e) E and (f) LOE.

coefficient that has been opted and also meets the significant value criteria of 0.05 for all the coefficients of the regression equations, including the constant is opted for the optimum equation. So,

both log-linear and mixed models were considered in the cases. The models were then undergone through different statistical strength parameters tests. Different statistical measures have been

implemented by researchers to compare a model and observations. Some of them are:

- a) fractional bias
- b) geometric mean bias
- c) the normalized mean square error (NMSE).
- d) the geometric variance (GV).
- e) the coefficient correlation (R).

A perfect model would have MG, VG, R and FAC2 = 1.0, and FB and NMSE = 0.0. FB and MG measure systematic errors. NMSE is given by:

$$NMSE = \frac{\overline{(C_o - C_p)^2}}{C_o C_p} \tag{6}$$

FB is given by:

$$FB = 2 \times \left(\frac{\overline{C_o - C_p}}{\overline{C_o + C_p}} \right) \tag{7}$$

MG is based on a logarithmic scale. It is given by:

$$MG = e^{\overline{(\ln C_o - \ln C_p)}} \tag{8}$$

Unpredictable fluctuations generally lead to random errors leading to undesirable results. When measurements are repeated, the arithmetic mean comes out to be null, and the predicted values are scattered against the true values. NMSE and VG are both measures of systematic and random or un-systematic errors [36,37].

$$VG = e^{\overline{(\ln C_o - \ln C_p)^2}} \tag{9}$$

The parameter R represents the degree of co-linearity between the parameters. Estimating the statistical strength of an empirical relationship is a required but not sufficient requirement.

The best regression equation in terms of the highest R² value, which is obtained in linear form, is

$$E = -2.102 + 0.009D + 0.162\rho + 0.133UCS - 0.098ITS \tag{10}$$

The corresponding statistical parameters of the linear regression equation are given in Table 1.

The observed and predicted values of Young's modulus in the linear regression equation are shown in Fig. 3.

The residuals are checked for their normal distribution by histograms. The histogram is shown in Fig. 4.

The best regression equation in terms of the highest R² value which, is obtained in log-linear form, is given by:

$$\log_{10} E = -1.31 - 0.158 \log_{10} D + 0.488 \log_{10} \rho + 0.714 \log_{10} UCS - 0.161 \log_{10} ITS \tag{11}$$

Fig. 5 depicts the graph of the observed and predicted value of E according to (11).

The statistical strength parameter table is given below in Table 2.

The residuals are checked for their normal distribution by histograms. The histogram is shown in Fig. 6.

The most suitable regression model having second best R² value that came out of the independent parameters in the mixed regression form is

$$\log_{10} E = -0.369 + 0.203 \log_{10} D + 0.006\rho - 0.05UCS \tag{12}$$

The statistical strength parameter table is given in Table 3.

Fig. 7 depicts the graph of the observed and predicted value of E according to (12).

The residuals are checked for their normal distribution by histograms. The histogram is shown in Fig. 8.

Hence, the final equation stands to be.

$$\log_{10} E = -0.369 + 0.203 \log_{10} D + 0.006\rho - 0.005UCS \tag{13}$$

3.2. Validation of physico-mechanical properties through the artificial neural network

The Widrow-Hoff learning type of Multilayer Perceptrons (MLP) has been used, and it has connections with adjoining layers. The hidden layers are present between the input and output layers only. This type of overall learning process of the network is backpropagation algorithm which involves the gradient descent method to identify the modified weights corresponding to each hidden input and output layer. The gradient descent method is a type of optimization method which involves the prediction of the minimum value of any parameter under consideration (here, the weighted values corresponding to the layers) to get to the

Table 1. Statistical parameters of the linear regression equation.

Correlation coefficient (R)	Factor of two (FAC2)	Geometric variance (GV)	Geometrical mean bias (MG)	Fractional bias (FB)	Normalized mean square error (NMSE)	Remarks
0.741	0.993554	1.000042	0.993554	1.006488	0.00524	Indicates perfect statistical representation of the population

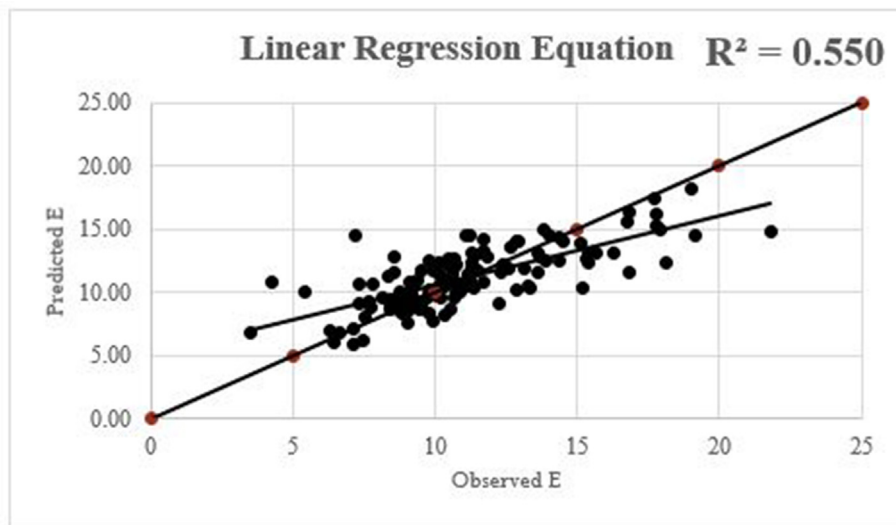


Fig. 3. Graph of the observed and predicted value of the linear equation.

desired prediction level as close as possible [38,39]. The analyses have been conducted in IBM SPSS 25. The tables indicate the stepwise analysis of the data and the activation functions used in the analysis. The analysis has been done based on the MLP theory of neural networks. The input activation function here is used as a hyperbolic tangent, and the identity function is used as the output activation function. The outputs are as generated in IBM SPSS 25. Table 4 indicates the general process of the neural network. A total of 132 sample sets have been used for the analysis by an artificial neural network, as indicated in Table 4. About 72.7% of the samples

were taken as training data, whereas 27.3% were taken as test data. No data were excluded for either test or training sets. Table 4 shows the different parameters and activation functions implemented in the input and output layers. The independent variables were the components of the input layer. The variables are depth of occurrence of the rock samples (D), the density of the rock samples (ρ), UCS and ITS . The parameters are rescaled using the standardized method. In the hidden layer, a hyperbolic tangent is used as the activation function. One hidden layer is taken under consideration and it consists of five units. The output layer consisted of

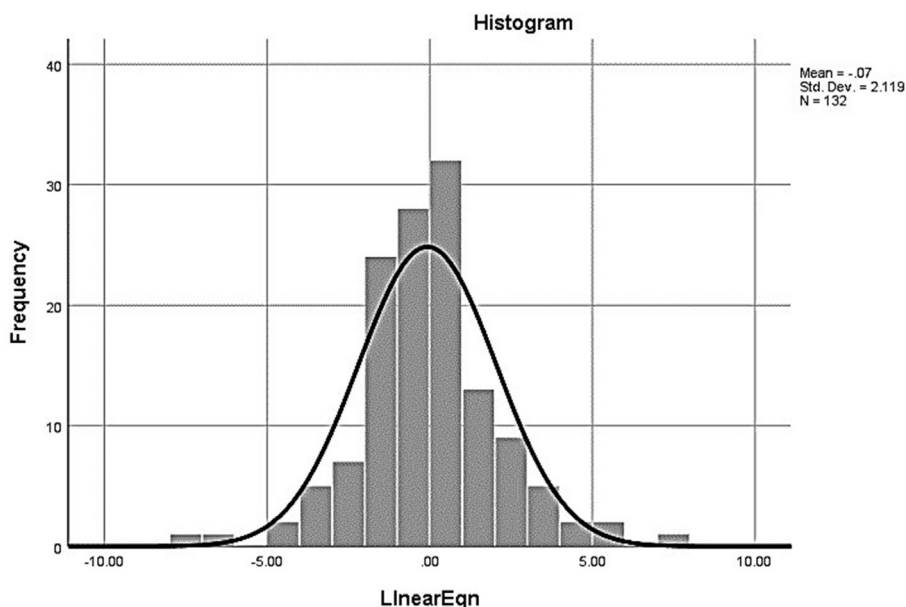


Fig. 4. Histogram showing the normal distribution of the residuals of linear equation.

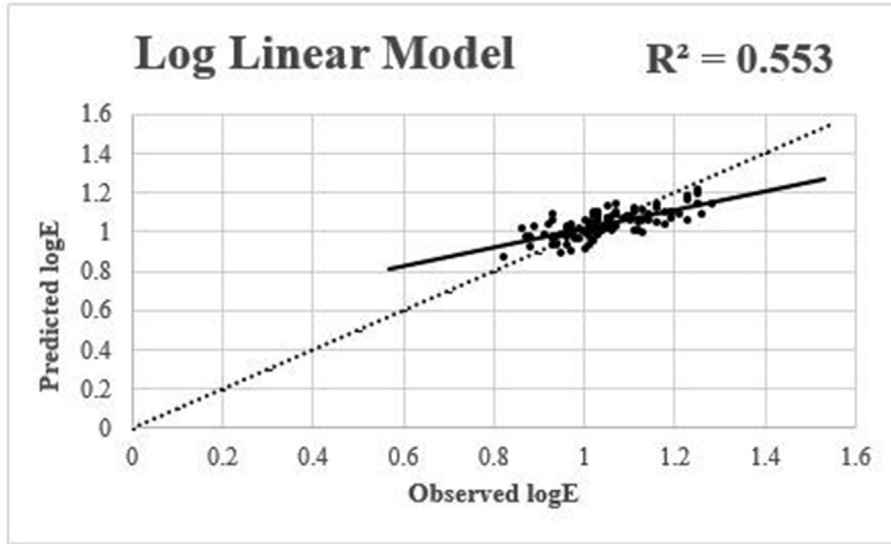


Fig. 5. Graph of the observed and predicted value of logarithmic Young's modulus.

Table 2. Statistical strength parameter table of the log–linear equation.

Correlation coefficient (R)	Factor of two (FAC2)	Geometric variance (GV)	Geometrical mean bias (GM)	Fractional bias (FB)	Normalized mean square error (NMSE)	Remarks
0.743	1.064012	1.0066	1.0000	0.006627	0.00663	Indicates perfect statistical representation of the population

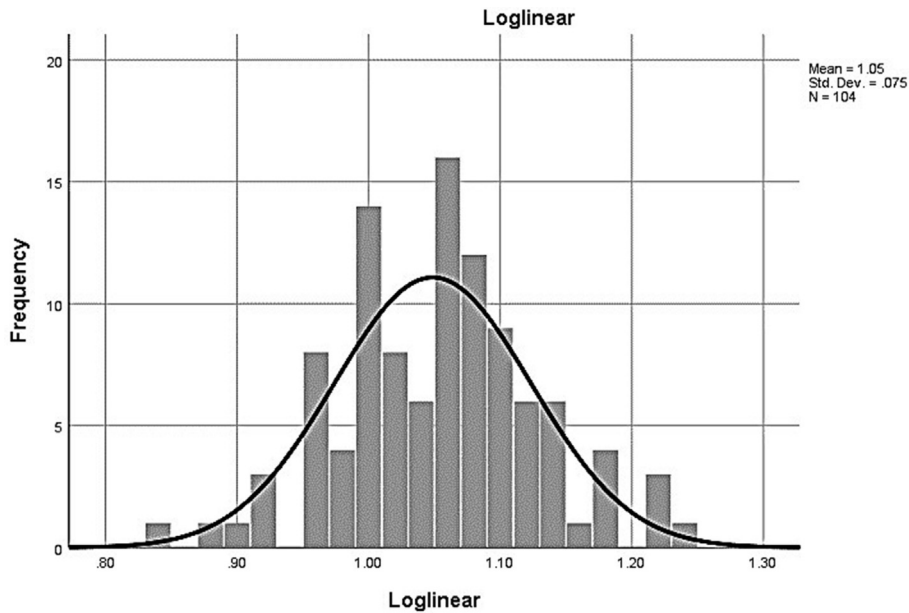


Fig. 6. Histogram showing the normal distribution of the log–linear equation.

Table 3. Statistical strength parameter table in the mixed regression equation.

Correlation coefficient (R)	Factor of two (FA2)	Geometric variance (GV)	Geometrical mean bias (MG)	Fractional bias (FB)	Normalized mean square error (NMSE)	Remarks
0.780	0.98000	0.980000	1.00000	0.0193	0.0004	Indicates perfect statistical representation of the population

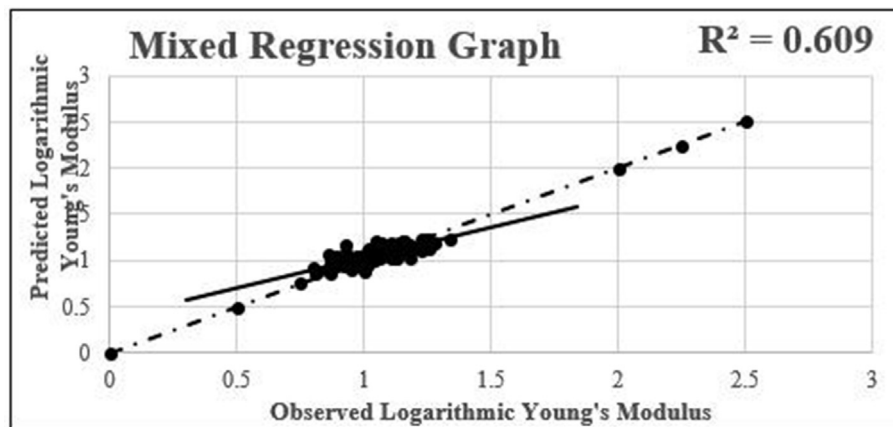


Fig. 7. Graph of the observed and predicted value of E.

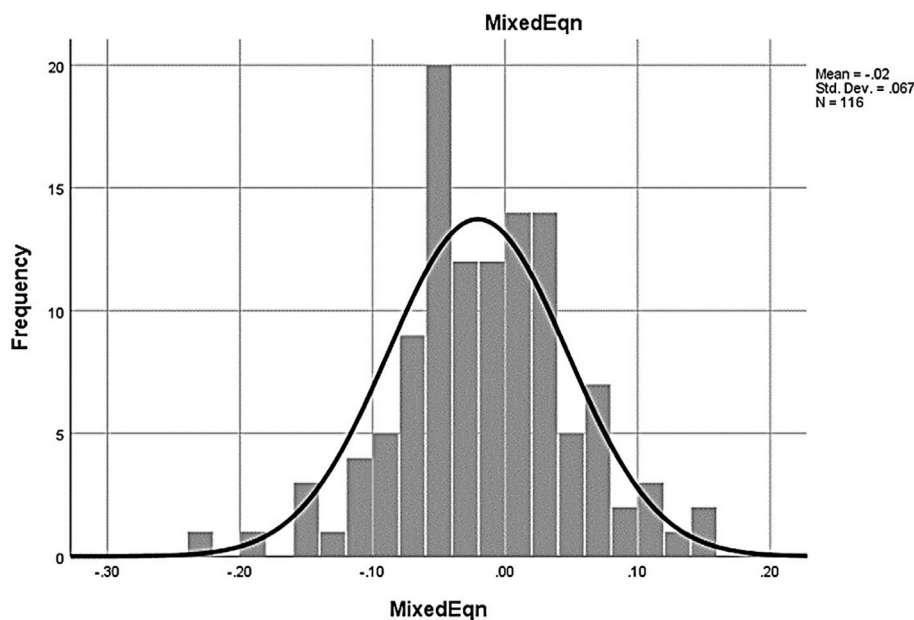


Fig. 8. Histogram showing the normal distribution of the residual of mixed (log and linear) equation.

the dependent variable, i.e., logarithmic transformed Young's modulus (LOE), and hence it had one unit. The identity function was taken as the activation function in the output layer (see Table 5).

Fig. 9 shows the estimated neural network. The input layer consists of four independent units along

with the bias, which generally acts as a constant for the activation function. The hidden layer consisted of five units which, along with the interaction of bias and the activation functions, gives the dependent variable as the predicted output.

Table 6 indicates the training process of the neural network. The sum of the square errors in the training set is high, whereas the relative error is low. On the other hand, in the testing set, the sum of the square error is low, whereas the relative error is high.

Table 7 indicates the bias and weighted values of different parameters. The values were calculated using the activation function, along with the bias, in the input layer, and values hence obtained were used in the output activation layer, thereby giving the output.

Table 4. General process of neural network.

Case processing summary			
		N	Percent
Sample	Training	96	72.7%
	Testing	36	27.3%
Valid		132	100.0%
Excluded		0	
Total		132	

Table 5. Different parameters and activation functions.

Network information			
Input layer	Covariates	1	D
		2	ρ
		3	UCS
		4	ITS
Hidden layer(s)	Number of units	4	
	Rescaling method for covariates		Standardized
	Number of hidden layers		1
	Number of units in hidden layer 1 ^a		5
Output layer	Activation function		Hyperbolic tangent
	Dependent variables	1	LOE
	Number of units		1
	Rescaling method for scale dependents		Standardized
	Activation function		Identity
	Error function		Sum of squares

^a Excluding the bias unit.

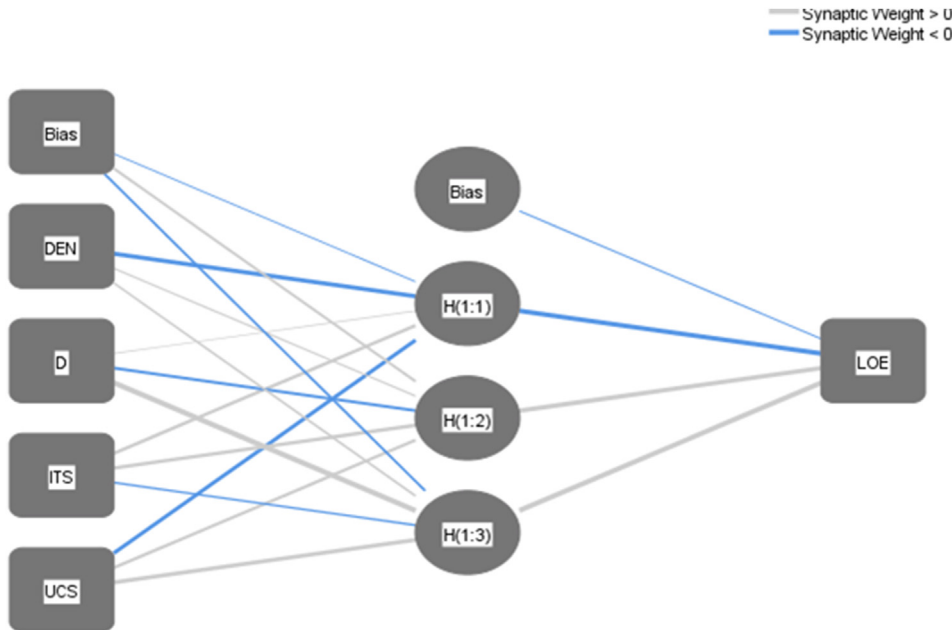


Fig. 9. Estimated neural network.

Table 6. Training process of the neural network.

Model summary		
Training	Sum of squares error	12.316
	Relative error	0.259
	Stopping rule used	1 consecutive step(s) with no decrease in error [§]
	Training time	0:00:00.01
Testing	Sum of squares error	3.668
	Relative error	0.402
Dependent variable: LOE		

Table 7. Bias and weighted values of different parameters.

Parameter estimates		Predicted					Output layer
		Hidden layer 1					logE
Predictor		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	
Input layer	(Bias)	-0.445	-1.412	0.504	0.826	-0.366	
	D	-0.284	-0.260	-0.277	0.019	-0.377	
	ρ	0.589	-1.114	-0.058	0.052	-0.371	
	UCS	-0.877	0.125	-0.457	-0.237	0.139	
	ITS	0.081	0.393	0.204	-0.594	0.013	
Hidden layer	(Bias)						-0.362
	H(1:1)						-0.466
	H(1:2)						-1.589
	H(1:3)						-1.040
	H(1:4)						-0.735
	H(1:5)						0.412

Table 8. Normalized importance value of the variables.

Independent variable importance		
	Importance	Normalized importance
D	0.150	36.1%
ρ	0.317	76.1%
UCS	0.417	100.0%
ITS	0.116	27.8%

Table 8 indicates the normalized importance value of the variables. The importance is shown in the form of a bar diagram in Fig. 10. UCS has the highest contribution to the prediction model as it is directly proportional to Young's modulus. The linear correlation is also stronger for UCS and logarithmic transformed Young's modulus indicating the higher

degree of contribution of UCS to the prediction model.

The scattered plot of the predicted and observed values of logarithmic Young's modulus is shown in Fig. 11.

The scattered plot shows that the sample points are equally dispersed along the normal line. That is, the predicted values are at par with the observed values of the dependent variables. The dispersion of the residuals about the zero line is shown in Fig. 12.

The symmetrical dispersion of the residuals along the zero line of the residual indicates that the sum of the squares of the errors nullifies each other to become minimum for this particular neural network

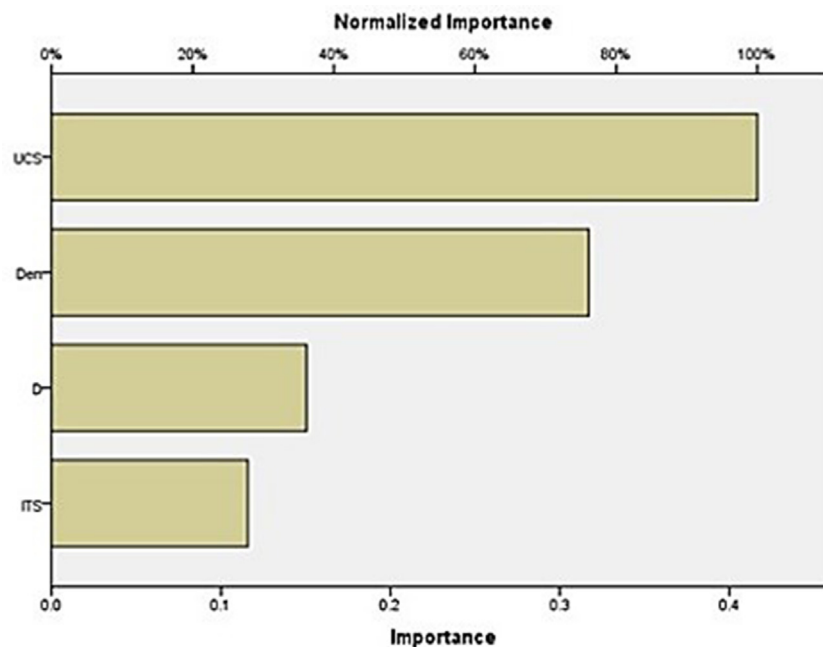


Fig. 10. Normalized importance percentage of the parameters.

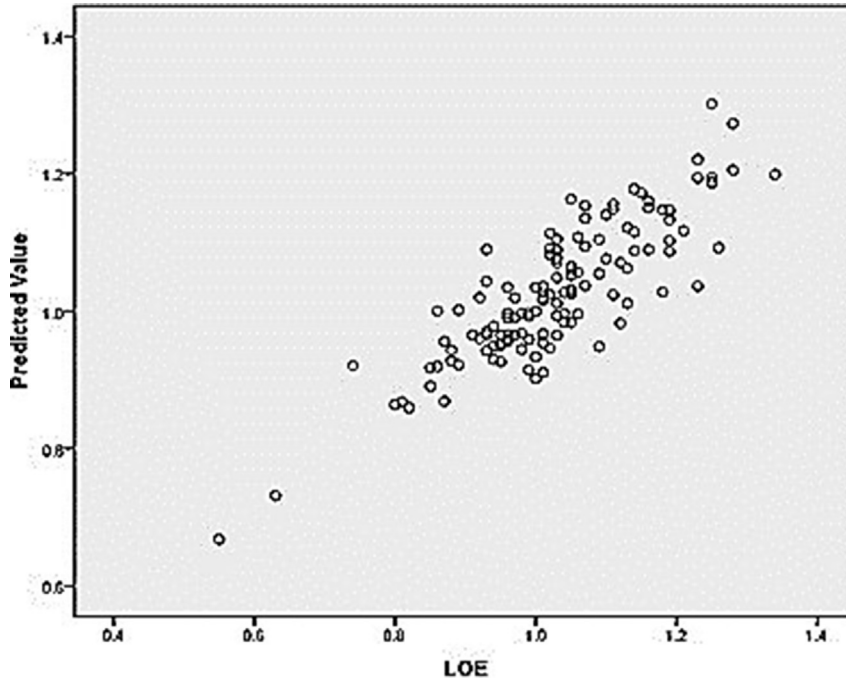


Fig. 11. Scattered plot of the predicted and observed values of logarithmic Young's modulus.

model. The correlation obtained from the multiple regressions after being run through the learning process of the artificial neural network obtained better results which were evident from the improvement of the coefficient of correlation R^2 between the observed and the predicted value of

Young's modulus of intact rocks. The improvement of the results through the artificial neural network analysis leads to validating the obtained relation through multiple regressions. The improved R square value for the predicted logarithmic transformed Young's modulus is shown in Fig. 13.

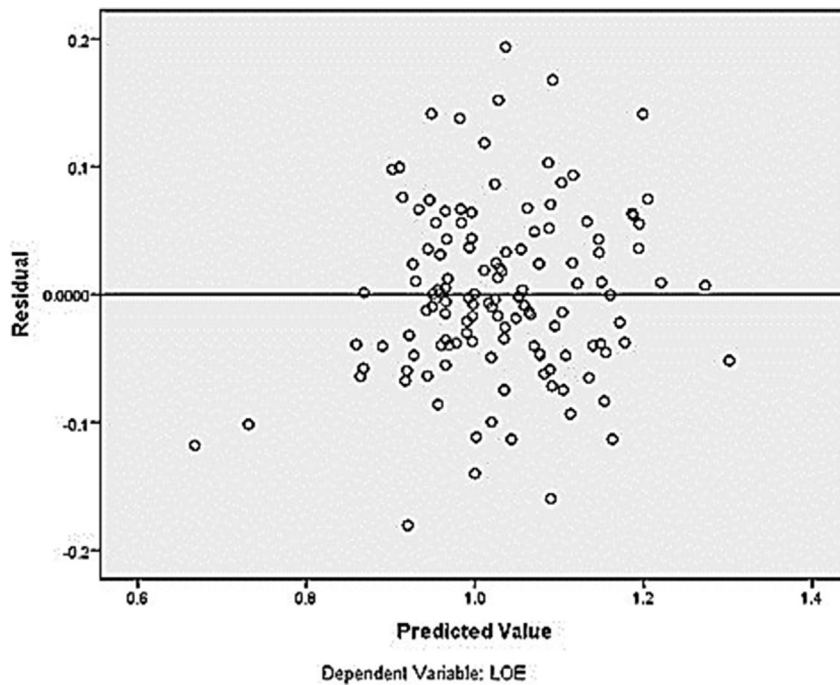


Fig. 12. Dispersion of the residuals about the zero line.

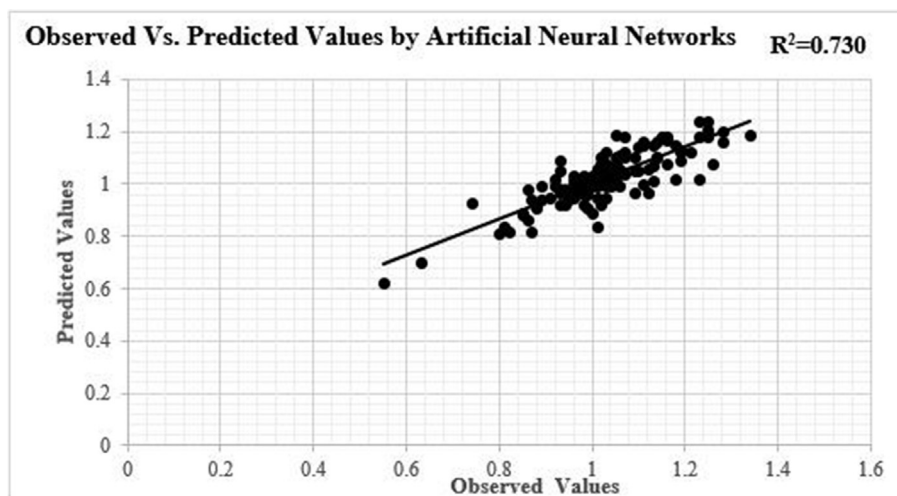


Fig. 13. Scattered plot of the predicted value of the logarithmic Young's modulus and the observed Young's modulus with R^2 value.

Table 9. Comparison of different equations found out by multiple regression and artificial neural networks.

Type of statistical modelling	Correlation coeff. (R)	FAC2	GV	MG	FB	NMSE	Mean of residuals	First determining ratio	Final determining ratio
Linear	0.741	0.993554	1.00004	0.9935	1.00649	0.00524	-0.07	-1981	141
Log linear	0.743	1.064012	1.0066	1.00	0.00627	0.00623	1.05	19,402	119
Log & Linear	0.780	0.980	0.980	1.00	0.0193	0.0004	-0.02	-4,851,762	1950
ANN	0.854	0.9972	1	1.002	0.00272	7.85×10^{-16}	0.0	∞	1,087,898,089,171,970

The comparison of the different equations found through multiple regression and artificial neural network has been made on the same scale of statistical strength parameters and shown in the following Table 9.

The first determining ratio showed two indeterminate values. Therefore, the ratio of correlation coefficients (R) and NMSE has been used as the final determining ratio for the best-fit regression equation. So, from this it has been found that ANN modelling is the best model as compared to the other statistical models such as linear, log linear, log and linear.

4. Conclusions

- An attempt has been made to predict Young's modulus using a functional relationship between UCS, ITS, density, and depth of occurrence of rocks. The results which are obtained are valid for the particular area of the borehole only.
- The study shows that for fair prediction of Young's modulus (E) values, among the above four parameters, a minimum of two parameters are needed.

- Multiple regression was used to find a coherent equation for the prediction of Young's modulus in terms of density, depth of occurrence of rocks, UCS and ITS. In addition to the coefficient of correlation, the fractional bias (FB), geometric variance (GV), normalized mean square error (NMSE) and the factor of predictions within a factor of two of observations (FAC2) were the major statistical performance tests that have also been calculated to check the predictive performance of the model.
- It is suggested that the empirical equations developed using regression analysis in this study give predicted values that are more or less at par with the observed data of the borehole. Further, the data were analyzed through the artificial neural network method. There was a slight improvement in the R^2 value until 0.73. For appropriate generalization of the equations from the sample space over the whole population, the residuals of each of the obtained different types of equations have been checked for the normal distribution. The normal distribution of the residuals overgeneralizes the equation from the given sample space to the population.

- It has also been concluded that an increase in statistical analysis, i.e., through multiple regression to ANN, improves the results and provides more robustness to the equation. It can be a useful tool for field engineers and geotechnicians. Hence, with increasing the sophistication of statistical analysis of the different physico-mechanical properties of rocks, it was found that there is a standard level of relation between Young's modulus and the easily measurable physico-mechanical properties.

Ethical statement

The authors state that the research was conducted according to ethical standards.

Funding body

No funding was required during the study.

Conflict of interest

There is no conflict of interest.

Acknowledgements

The authors express the heartiest gratitude towards the Department of Mining Engineering staff for their kind and patient support towards accomplishing the task.

References

- [1] Bieniawski ZT. Engineering rock mass classifications. Wiley; 1989.
- [2] Bell FG. The physical and mechanical properties of the fell sandstones Northumberland. England. Eng. Geol. 1978;12: 1–29. [https://doi.org/10.1016/0013-7952\(78\)90002-9](https://doi.org/10.1016/0013-7952(78)90002-9).
- [3] Mishra DA, Basu A. Estimation of uniaxial compressive strength of rock materials by index tests using regression analysis and fuzzy inference system. Eng Geol 2013;160: 54–68. <https://doi.org/10.1016/j.enggeo.2013.04.004>.
- [4] Madhubabu N, Singh PK, Kainthola A, Mahanta B, Tripathy A, Singh TN. Prediction of compressive strength and elastic modulus of carbonate rocks. Measurement 2016; 88:202–13. <https://doi.org/10.1016/j.measurement.2016.03.050>.
- [5] Singh R, Umrao RK, Ahmad M, Ansari MK, Sharma LK, Singh TN. Prediction of geomechanical parameters using soft computing and multiple regression approach. Measurement 2017;99:108–19. <https://doi.org/10.1016/j.measurement.2016.12.023>.
- [6] Armaghani DJ, Amin MFM, Yagiz S, Faradonbeh RS, Abdullah RA. Prediction of the uniaxial compressive strength of sandstone using various modelling techniques. Int J Rock Mech Min Sci 2016;85:174–86. <https://doi.org/10.1016/j.ijrmms.2016.03.018>.
- [7] Singh RN, Hassani FP, Elkington PAS. The application of strength and deformation index testing to the stability assessment of coal measures excavations. In: Proceedings of the 24th US Rock Mechanics Symposium. AEG: Texas A&M University; 1983.
- [8] Cargill JS, Shakoor A. Evaluation of empirical methods for measuring the uniaxial compressive strength of rock. Int J Rock Mech Min Sci 1990;27(6):495–503. [https://doi.org/10.1016/0148-9062\(90\)91001-N](https://doi.org/10.1016/0148-9062(90)91001-N).
- [9] Tugrul A, Zarif IH. Correlation of mineralogical and textural characteristics with engineering properties of selected granitic rocks from Turkey. Eng Geol 1999;51(4):303–17. [https://doi.org/10.1016/S0013-7952\(98\)00071-4](https://doi.org/10.1016/S0013-7952(98)00071-4).
- [10] Sulukcu S, Ulusay R. Evaluation of the block punch index test with particular reference to the size effect, failure mechanism and its effectiveness in predicting rock strength. Int J Rock Mech Min Sci 2001;38(8):1091–111. [https://doi.org/10.1016/S1365-1609\(01\)00079-X](https://doi.org/10.1016/S1365-1609(01)00079-X).
- [11] Kahraman S. Evaluation of simple methods for assessing the uniaxial compressive strength of rock. Int J Rock Mech Min Sci 2001;38(7):981–94. [https://doi.org/10.1016/S1365-1609\(01\)00039-9](https://doi.org/10.1016/S1365-1609(01)00039-9).
- [12] Yasar E, Erdogan Y. Estimation of rock physio mechanical properties using hardness methods. Eng Geol 2004;71(3–4): 281–8. [https://doi.org/10.1016/S0013-7952\(03\)00141-8](https://doi.org/10.1016/S0013-7952(03)00141-8).
- [13] Basu A, Aydin A. Predicting uniaxial compressive strength by point load 379 test: significance of cone penetration. Rock Mech Rock Eng 2006;39:483–90. <https://doi.org/10.1007/s00603-006-0082-y>.
- [14] Sharma PK, Singh TNA. Correlation between P-wave velocity, impact strength index, slake durability index and uniaxial compressive strength. Bull Eng Geol Environ 2008; 67(1):17–22. <https://doi.org/10.1007/s10064-007-0109-y>.
- [15] Meulenkamp F, Grima M Alvarez. Application of neural networks for the prediction of the unconfined compressive strength (UCS) from Equotip hardness. Int J Rock Mech Min Sci 1999;36(1):29–39. [https://doi.org/10.1016/S0148-9062\(98\)00173-9](https://doi.org/10.1016/S0148-9062(98)00173-9).
- [16] Gokceoglu C, Zorlu K. A fuzzy model to predict the unconfined compressive strength and modulus of elasticity of a problematic rock. Eng Appl Artif Intell 2004;17(1):61–72. <https://doi.org/10.1016/j.engappai.2003.11.006>.
- [17] Yagiz S. Predicting uniaxial compressive strength, modulus of elasticity and index properties of rocks using the Schmidt hammer. Bull Eng Geol Environ 2009;68(1):55–63. <https://doi.org/10.1007/s10064-008-0172-z>.
- [18] Dehghan S, Sattari GH, Chehreh CS, Aliabadi MA. Prediction of unconfined compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural. New Min Sci Technol 2010;20(1):41–6. [https://doi.org/10.1016/S1674-5264\(09\)60158-7](https://doi.org/10.1016/S1674-5264(09)60158-7).
- [19] Yesiloglu-Gultekin N, Gokceoglu C, Sezer EA. Prediction of uniaxial compressive strength of granitic rocks by various non-linear tools and comparison of their performances. Int J Rock Mech Min Sci 2013;62:113–22. <https://doi.org/10.1016/j.ijrmms.2013.05.005>.
- [20] Rezaei M, Majdi A, Monjezi M. An intelligent approach to predict unconfined compressive strength of rock surrounding access tunnels in longwall coal mining. Neural Comput Appl 2014;24(1):233–41. <https://doi.org/10.1007/s00521-012-1221-1>.
- [21] Elkhatny S, Mahmoud M, Mohamed I, Abdurraheem A. Development of a new correlation to determine the static Young's modulus. J Pet Explor Prod Technol 2018;8(1):17–30. <https://doi.org/10.1007/s13202-017-0316-4>.
- [22] Cevik A, Sezer EA, Cabalar AF, Gokceoglu C. Modeling of the uniaxial compressive strength of some clay-bearing rocks using neural network. Appl Soft Comput 2011;11(2):2587–94. <https://doi.org/10.1016/j.asoc.2010.10.008>.
- [23] Ozbek A, Unsal M, Dikec A. Estimating uniaxial compressive strength of rocks using genetic expression programming. Journal of Rock Mechanics and Geotechnical Engineering 2013;5(4):325–9. <https://doi.org/10.1016/j.jrmge.2013.05.006>.
- [24] Dehghan S, Sattari GH, Chelgani SC, Aliabadi MA. Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial

- neural networks. *Min Sci Technol* 2010;20(1):41–6. [https://doi.org/10.1016/S1674-5264\(09\)60158-7](https://doi.org/10.1016/S1674-5264(09)60158-7).
- [25] Karakul H, Ulusay R. Empirical correlations for predicting strength properties of rocks from P-wave velocity under different degrees of saturation. *Rock Mech Rock Eng* 2013;46(5):981–99. <https://doi.org/10.1007/s00603-012-0353-8>.
- [26] Khandelwal M, Ranjith PG. Correlating index properties of rocks with P-wave measurements. *J Appl Geophys* 2010; 71(1):1–5. <https://doi.org/10.1016/j.jappgeo.2010.01.007>.
- [27] Kurtuluş C, Sertçelik F, Sertçelik I. Correlating physico-mechanical properties of intact rocks with P-wave velocity. *Acta GeodGeophys* 2016;51(3):571–582N. <https://doi.org/10.1007/s40328-015-0145-1>.
- [28] Závacký M, Štefaňák J, Horák V, Miča L. Statistical estimate of uniaxial compressive strength of rock based on Shore hardness. *Procedia Eng* 2017;191:248–55. <https://doi.org/10.1016/j.proeng.2017.05.178>.
- [29] Małkowski P, Ł Ostrowski. The methodology for the Young modulus derivation for rocks and its value. *Procedia Eng* 2017. <https://doi.org/10.1016/j.proeng.2017.05.164>.
- [30] Labuz JF. The problem of machine stiffness revisited. *Lecture* 1991;18(3):439–42. <https://doi.org/10.1029/91GL00350>.
- [31] Hockett JE, Gillis PP. Mechanical testing machine stiffness: Part I—theory and calculations. *Int J Mech Sci* 1971;13(3): 251–64. [https://doi.org/10.1016/0020-7403\(71\)90007-5](https://doi.org/10.1016/0020-7403(71)90007-5).
- [32] Karacan CO. Elastic and shear moduli of coal measure rocks derived from basic well logs using fractal statistics and radial basis functions. *Int J Rock Mech Min Sci* 2009;46(8):1281–95. <https://doi.org/10.1016/j.ijrmms.2009.04.002>.
- [33] Chakraborty S, Bisai R, Palaniappan SK, Pal SK. Characterization of fracture pattern of Indian coal measure rock under uniaxial compression stress by statistical analysis of fractal dimension of the microcrack orientation. *J Inst Eng (India)* 2022;103:95–106. <https://doi.org/10.1007/s40033-021-00308-8>.
- [34] Bisai R, Palaniappan SK, Pal SK. Influence of individual and combined pre-treatment on the strength properties of granite and sandstone. *Arabian J Geosci* 2020;13:7. <https://doi.org/10.1007/s12517-019-5009-5>.
- [35] Field A. *Discovering statistics using IBM SPSS statistics*. Sage; 2013.
- [36] NIST/SEMATECH e-handbook of statistical methods. Accessed 24/09/2020, <https://www.itl.nist.gov/div898/handbook/eda/section3/contour.htm#:~:text=A%20contour%20plot%20is%20a,where%20that%20z%20value%20occurs>.
- [37] Patryla L, Galerua D. Statistical performances measures—models comparison, French Alternative Energies and Atomic Energy Commission. 2011.
- [38] Brownlee J. *Machine learning mastery*. 2019. <https://machinelearningmastery.com/gradient-descent-for-machine433-learning/>.
- [39] Sayadi A, Monjezi M, Talebi N, Khandelwal MA. Comparative study on the application of various artificial neural networks to simultaneous prediction of rock fragmentation and backbreak. *J Rock Mech Geotech Eng* 2013;5(4):318–24. <https://doi.org/10.1016/j.jrmge.2013.05.007>.