JSTT JOURNAL OF SCIENCE AND TRANSPORT TECHNOLOGY

Journal of Science and Transport Technology
Journal homepage: https://jstt.vn/index.php/en

UNIVERSITY OF TRANSPORT TECHNOLOGY

Estimation of California Bearing Ratio of Soils Using Random Forest based Machine Learning

Dung Quang Vu[1,*], Duc Dam Nguyen[1], Quynh-Anh Thi Bui[1], Duong Kien Trong[1], Indra Prakash[2], Binh Thai Pham[1]

[1]University of Transport Technology, 54 Trieu Khuc, Thanh Xuan, Hanoi, Vietnam

[2]DDG (R) Geological Survey of India, Gandhinagar 382010, India

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2021.en.1.1.45-58

***Corresponding author:**
E-mail address:
dungvq@utt.edu.vn

Received: 13/11/2021
Revised: 06/12/2021
Accepted: 08/12/2021

Abstract: California Bearing Ratio (CBR) is an essential parameter utilized to evaluate the strength of the soil subgrades and base course materials of different types of pavements. In this study, the Machine Learning (ML) approach has been adopted using Random Forest (RF) model to estimate the CBR of the soil based on 10 input parameters such as Plasticity Index (PI), Liquid Limit (LL), Silt Clay content (SC), Fine Sand content (FS), Coarse sand content (CS), Optimum Water Content (OWC), Organic content (O), Plastic Limit (PL), Gravel content (G), and Maximum Dry Density (MDD), which can be easily determined in the laboratory. An experimental database was collected from 214 soil samples, which were classified according to AASHTO M 145(clayey, gravel, sand, silty and clayey soils). The data was divided into 70% training and 30% test data in the model study. Model performance was evaluated using standard statistical measures such as coefficient of determination, correlations, and errors (relative error, MAE, and RMSE). Based on the analysis results shows the RF model is capable of correct prediction of the CBR of the Soil.

Keywords: California Bearing Ratio; Random Forest; Machine Learning.

1. Introduction

California Bearing Ratio (CBR) is a parameter of great importance in geotechnical engineering to evaluate the strength of subgrade material of pavements structures of roadways, railways, and airfields [1]. CBR can be determined in the field as well as in the laboratory. In the field CBR test method, a loading jack is used to force a piston into the soil mass and subgrade material at the test site, and piston load is compared to the depth of penetration to measure the relative strength of in-situ soils and base course material for pavement design. Field CBR equipment is costly and difficult to carry at different locations/ sites. Therefore, laboratory methods are generally employed for the determination of CBR of soil and subgrade material. In the laboratory, CBR is determined by inserting a plunger of standard diameter at a rate of 1.3 mm/min into a compacted soil specimen prepared at Optimum Water Content (OWC) [2]. The CBR values of any soil samples can be estimated either in soaked conditions or in un-soaked conditions. Normally, the CBR values of soaked soil samples are always lower than the values of un-soaked samples. Therefore, the CBR values of soaked samples are generally accepted as a quality estimation of subgrade materials.

However, the soil samples prepared at OWC need to be kept in water saturation condition for 4 days, and the entire exercise of determining the soil CBR is a time- consuming process [2,13] and hence, the construction time is affected significantly by this process. Generally, CBR is to be determined for a number of samples, which is costly and time-consuming. Moreover, there is a possibility of Plastic Limit (PL), Optimum Water Content (OWC), disturbance of (LL), Plasticity Index (PI), and Maximum Dry Density (MDD).

Nowadays, many researchers have applied computer soft techniques as an effective method to predict the desired output data, combining the real-life problems of geotechnical engineering [3,5-8, 10,12,14-24] such as Layer Perceptron Neural Network (MLPN), Support Vector Machine (SVM), Gene Expression Programming (GEP), Artificial Neural Network (ANN), Multi- Group Method of Data Handling (GMDH), etc. (Table 1).

In this study, we have used Random Forest (RF) Machine Learning (ML) model based on 214 soil samples for the estimation of CBR soil samples during laboratory experiments. Therefore, indirect estimation of CBR may be adopted using other physico-mechanical parameters of soils and subgrade materials which can be determined easily in the laboratory such as Gravel content (G), Coarse sand content (CS), Fine Sand content (FS), Silt Clay content (SC), Organic content (O), Liquid Limit based on limited geo-mechanical input parameters which can be easily determined in the laboratory. Matlab software was used for modeling and data visualization.

Table 1. The previous ML study models used in predicting the CBR of soil

Reference	Model(s) used	Prediction accuracy	Number of soil samples
[3]	GEP, ANN	$0.910<R^2<0.918$	151
[4]	ANN	R^2 = 0.80–0.95	124
[5]	ANN	R^2 = 0.928	207
[6]	GMDH, ANN	$0.9703<R^2<0.9783$	158
[7]	GEP, SVM	$0.7930<R^2<0.912$	389
[8]	Particle Swarm Optimization (PSO)	-	134
[9]	Bayesian Programming (BP), Levenberg–Marquardt Backpropagation (LMBP), and Conjugate Gradient (CG)	R^2 = 0.90	129
[10]	ANN, alternating model trees (AMT), Gaussian processes regression (GPR), elastic net regularization regression (ENRR), least median of squares regression (LMSR), lazy Kstar (LKS), M-5 model trees, and random forest (RF)	$0.80< R^2<0.92$	97
[11]	Multivariate adaptive regression splines with piecewise-linear and piecewise-cubic (MARS-L, MARS-C), Gaussian process regression (GPR), and genetic programming (GP)	R^2 = 0.9686, RMSE = 0.0359	312
[12]	ELM coupled-improved PSO (ELM-IPSO), ELM coupled-TAC PSO (ELM-TPSO), and ELM coupled-modified PSO (ELM-MPSO)	$0.81< R^2<0.91$	312

2. Materials and Methods

2.1. Study area

The study area is along 80.23 km route between Km70+108 and Km150+ at the Van Don -

Mong Cai expressway project, Quang Ninh province, Vietnam (Fig 1). The scale of the project: specification meets TCVN 5729:2012 [25] with 4 lanes, design speed is 100 km/h. The width of the roadbed is 24.5 m, the total width of the vehicle lanes is 4x3.5 m, the width of the median strip is 0.5 m, the width of the inner safety strip is 2x0.5 m, and the width of the outer safety strip is 2x0.5 m, the width of the emergency stop lane (reserved platform) is 2x4.0 m.

Soil samples were collected from soil mines along the route (Fig 2) for laboratory testing. The testing works were carried out between, November 8, 2019 to July 1, 2021 of 214 samples. The tests included Particle size analysis, test for determining the Liquid limit, test for determination of Organic content, test for Moisture–Density relations, and test for the CBR.

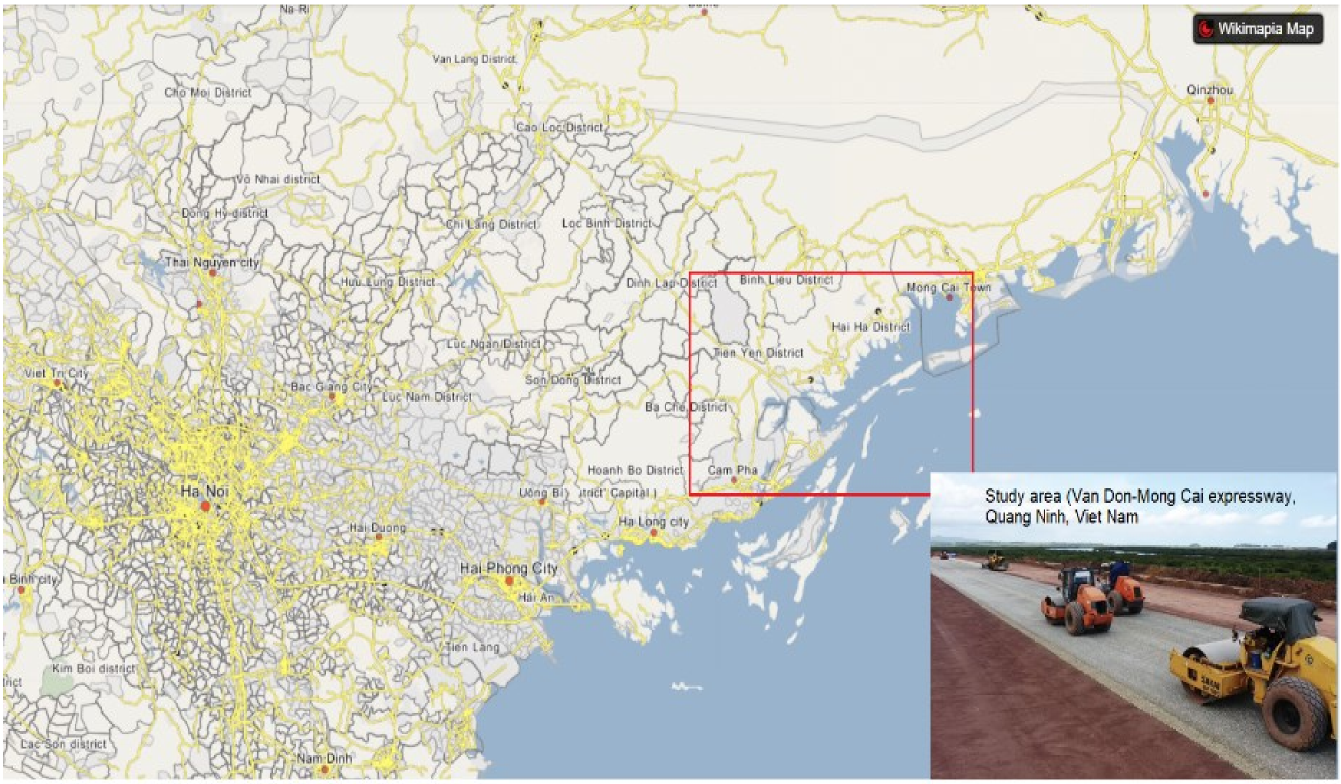


Fig 1. Location of study area

Fig 2. Soil sample collections along road

2.2. Brief description of determination of geo-mechanical properties of soils

The soil compaction process is the densification by mechanical impacts (Fig 3). Soil densification leads to air out of the soil, soil comes to more densification which improves the

strength characteristics of soils, and reduces permeability. Using setup compaction energy, the density of soil varies as a math function of water content. This relationship is called the moisture-density curve, or the compaction curve. Steps to find this curve equation has been standardized and are generally determined by Standard Proctor (ASTM D 698 [26] and AASHTO T 99 [27]) and Modified Proctor (ASTM D 1557 [28] and AASHTO T 180 [29]) tests. These tests can apply to cohesive soils. For cohesionless soils, should be used (ASTM D 4253 [30] and ASTM D 4254 [31] [32]).

Identifying the swelling properties of soils in the road subgrade is a key component of the geotechnical survey. Soil locations distributed at shallow depths beneath the proposed pavement elevation are generally sampled as part of the survey, and their swell potential may be identified in a number of methods. Atterberg's limits are used as a common method for identifying swell potential. The plastic and liquid limits and/or shrinkage limits will usually be performed in the laboratory [1]. Determining Liquid Limit (LL), Plastic Limit (PL), and Plasticity Index (PI) of soils according to AASHTO T 89 [33] and ASTM D 4318 [34]. LL is defined as the moisture content, when it increases than will cause plastic soil to do as a liquid. The PL is defined as the moisture content, when it increases than will cause semi-solid soil to become plastic. The PI is the difference between the liquid limit and the plastic limit (PI = LL - PL). The soils that have a higher PI tend to be predominantly clay, on the other hand, those with a lower PI tend to be predominantly silt [32].

The classification of soils is very important (through laboratory tests) in order to evaluate the vertical and horizontal variability of the subgrade. The classification through laboratory testing also provides information to determine stabilization requirements to improve the strength of the subgrade should additional support be required [1]. We have to determine particle size, particle-size analysis of soils was conducted according to AASHTO T 88 [35] and ASTM D 422 [36].

CBR value is obtained as the average of the ratio of laboratory stress to standard stress for the two penetration depths and expressed as a percentage and referenced to optimum water content and a maximum dry density are determined by a standard compaction test [37]. The CBR value was determined according to AASHTO T 193 [2] (Fig 3).

The organic content was determined according to AASHTO T 267 [38].

Fig 3. Testing work of California Bearing Ratio (CBR) of Laboratory-Compacted Soils

2.3. Data used

2.3.1. CBR values determined in the laboratory

CBR test method was developed by the California Department of Highways in the late 1920s with the purpose to determine properties of the cohesive soil in the subgrade and subbase layers of road pavement sections. The test method was specified by the American Association of State Highway and Transportation Officials [39] and the American Society for Testing and Materials and

Materials [40].

In the United States, some organizations such as Federal Highway Administration (FHWA), Federal Aviation Administration (FAA), and AASHTO, etc. have used CBR values for designing highways, airports, parking lots, and other pavement. Researchers determined the empirical correlation between CBR and resilient modulus and a variety of other engineering soil properties. CBR is not a fundamental material property so that it is unsuitable for application in mechanistic and mechanistic-empirical design procedures. However, it has a long history in pavement design, this test performs relatively easy and inexpensively, and it has reasonably well correlations with more fundamental properties like resilient modulus. Therefore, it continues to be used in practice [1].

The subgrade materials are typically characterized by their strength and stiffness. The United States commonly used three basic subgrade stiffness/strength characterizations: the California Bearing Ratio (CBR), the modulus of subgrade reaction (k), and the elastic (resilient) modulus. Although there are other factors related to evaluating subgrade materials (such as swell in the case of materials that have clay content), however, stiffness is the most common characterization [32].

Further, pavement performance also depends on subgrade uniformity. However, it's difficult to obtain a perfect subgrade due to the inherent variability properties of the soil and the influence of water, temperature, and construction activities. In the United States, the research of [32] has shown that with a subgrade strength of less than a CBR value of 10, the subbase layer will deflect under traffic loadings in the same manner as the subgrade. That deflection then impacts the pavement, initially for flexible pavements, but ultimately rigid pavements as well.

2.3.2. Influencing factors (input parameters)

In this study, we considered 10 influencing factors: Plasticity Index (PI), Liquid Limit (LL), Silt Clay content (SC), Fine Sand content (FS), Coarse sand content (CS), Optimum Water Content (OWC), Organic content (O), Plastic Limit (PL), Gravel content (G), and Maximum Dry Density (MDD) for the estimation of CBR using FR model.

To obtain a high-quality subgrade, a proper understanding of soil characteristics, proper grading, and quality control testing are required. However, pavement design requirements and the level of engineering effort should be consistent with the relative importance, scale, and cost of projects. Therefore, must to has knowledge of subgrade soil's basic engineering characteristics is required for design. These include soil classification, soil density, coefficient of lateral earth pressure, and estimated CBR or resilient modulus. Typical CBR values of different soils relate to soil types are available in the American Concrete Pavement Association, Asphalt Paving Association, State of Ohio, State of Iowa, and Rollings et al. [1, 32, 41].

According to AASHTO M 145 [42] and ASTM D 2487 [43], the classification of soils needs to determine particle-size distribution, liquid limit, and plasticity index. Especially, ASTM D 2487 [43] also takes into account organic content.

The water content and the dry density as well as the texture of the soil affect the CBR. Normally, the CBR test in the laboratory is conducted on test samples that have different moisture, further, water content is likely to be achieved in the field. The difficulty is to determine the stable moisture content to find maximum dry density. Many other countries usually use the 4-day soaked CBR samples for determining CBR values [44].

2.4. Methods used

A database comprising of 10 input variables (G, CS, FS, SC, O, LL, PL, PI, OWC, and MDD) was employed, such that 70% of the data was used for training whereas the remaining 30% was selected for testing. Random Forest model was formulated and its performance was evaluated using prominent statistical indices namely RMSE, MAE, and R^2.

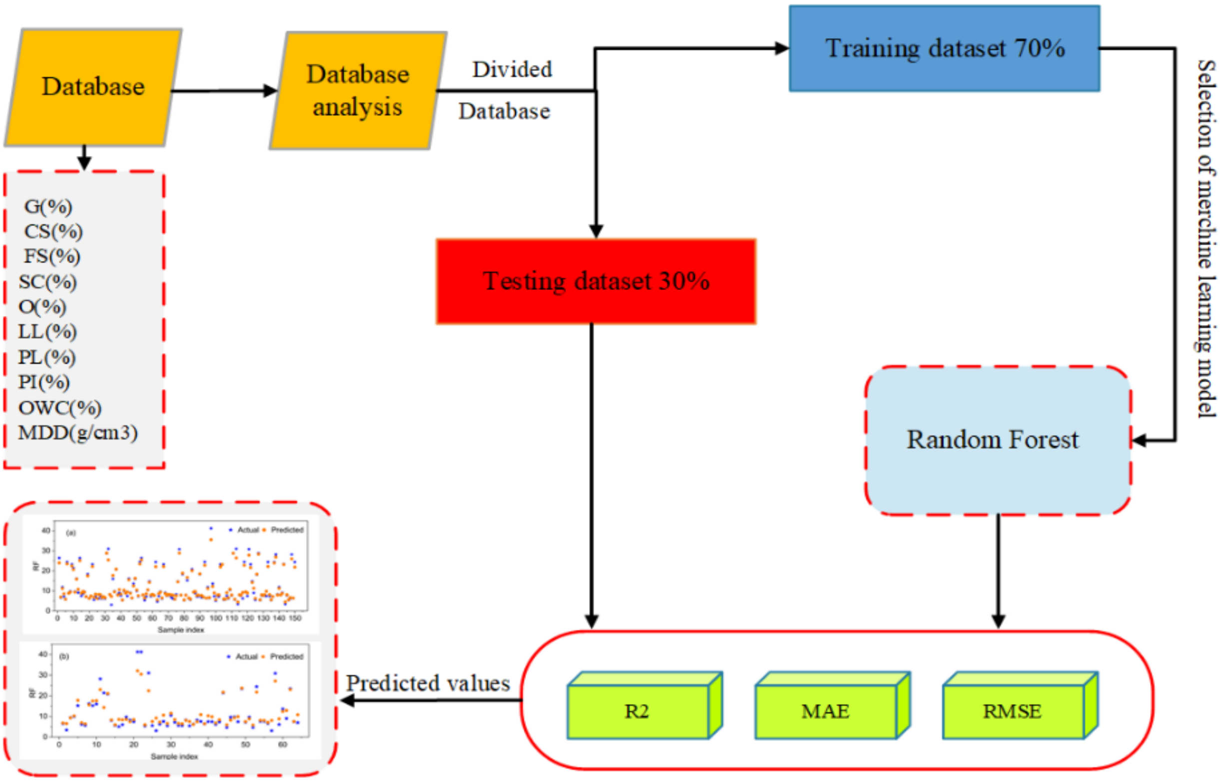


Fig 4. Methodological framework used for this study

Table 2. Statistical analysis of the parameters used in this study

Parameters	Codes	Unit	Minimum	Maximum	Mean	Median	StD
Inputs							
G	X1	(%)	0.00	51.40	22.057	24.750	13.295
CS	X2	(%)	3.00	46.30	24.101	23.700	7.017
FS	X3	(%)	2.50	41.50	9.035	7.250	6.468
SC	X4	(%)	17.87	88.70	44.807	44.550	10.447
O	X5	(%)	0.12	2.94	1.509	1.511	0.372
LL	X6	(%)	2.08	48.45	39.514	39.994	6.173
PL	X7	(%)	1.17	28.49	20.316	20.833	3.067
PI	X8	(%)	0.91	27.48	19.198	18.436	4.078
OWC	X9	(%)	9.30	21.50	14.010	14.273	2.618
MDD	X10	(g/cm^3)	1.672	2.140	1.881	1.871	0.118
Output							
CBR	Y1	(%)	3.09	41.26	11.804	7.953	8.175

*St.D. = Standard Deviation

2.4.1. Random Forest ML model

Random forests [45] are a set of approaches for creating an ensemble of decision trees using a randomized version of the tree induction procedure. The way random forests methods add random perturbations into the induction procedure distinguishes them from other decision trees. The challenge is to inject randomness into arbitrary decision trees while minimizing (x) and retaining a low bias.

The ensemble of decision trees was first introduced by Kwok and Carter [46]. The averaging many decision trees with diverse structures consistently generate better outcomes than any of the ensemble's individual components. However, this method was neither random nor totally automatic: decision trees were created by manually selecting splits towards the top of the tree that was almost as good as the ideal splits, and then expanding them using the conventional ID3 induction mechanism.

Breiman [47] a formerly technical study was one of the first to show, both numerically and practically, that combining numerous versions of a predictor into an ensemble might result in significant improvements in reliability. He observes and demonstrates that the mean model EL{ϕL} ELL has a smaller anticipated generalization error than the model L. As a result, Bagging consists in approximating EL{ϕL} of merging models developed from bootstrap samples [48] Lm (for m = 1, . . . , M) of the training set L. The {Lm} form is a set of L replicas, each of which contains N cases (x, y), picked at random but with replacing from L. Even though |L| = |Lm| = N, the bootstrap replication shows that 37% of the couples (x, y) from L are absent on average. Indeed, after N drawings with replacing, the likelihood of never being chosen is high.

$$\left(1-\frac{1}{N}\right)^N \approx \frac{1}{e} \approx 0.368 \qquad (1)$$

When the training set L is small, however, subsampling 67% of the objects may result in an increase in bias (for example, due to a reduction in model accuracy) that is too great to be balanced by a reduction in variance, resulting in overall lower performance. However, bagging has proven to be a useful method in a variety of situations, with one of its advantages being that it could be used to enhance any type of model, not just decision trees.

Breiman [45] couples Bagging [47] with the random variable selection at each node in his seminal Random Forests (RF) study [49]. Combining both methodologies and modifying randomness results in one of the most effective off-the-shelf machine learning algorithms, which works shockingly well for practically any task. Random Forests are shown to be competitive with boosting [50] and arcing algorithms [51], which also are designed to reduce bias whereas forests emphasize minimizing the error.

2.4.2. Validation indicators

The statistical performance of the developed Random Forest model was assessed by deploying three analytical standard evaluation indices, such as, coefficient of determination (R2), mean absolute error (MAE), and root mean square error (RMSE). These indices are given from Eq. 2 to Eq. 4 [52-55]:

$$R^2 = 1 - \frac{\sum_1^n (X_i - Y_i)^2}{\sum_1^n \left(\bar{Y}_i - Y_i\right)^2} \qquad (2)$$

(worst value = -∞; best value = +1)

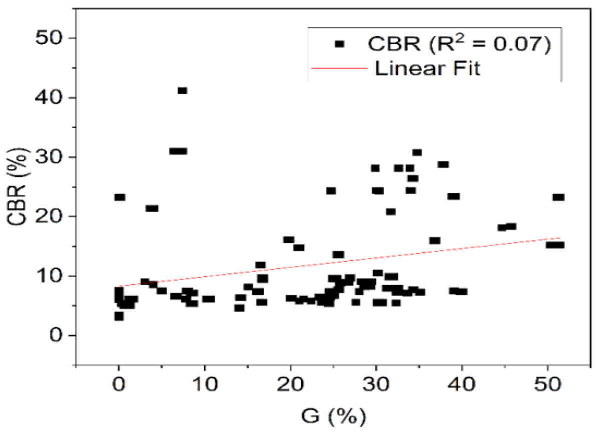
$$MAE = \frac{1}{n}\sum_1^n |X_i - Y_i| \qquad (3)$$

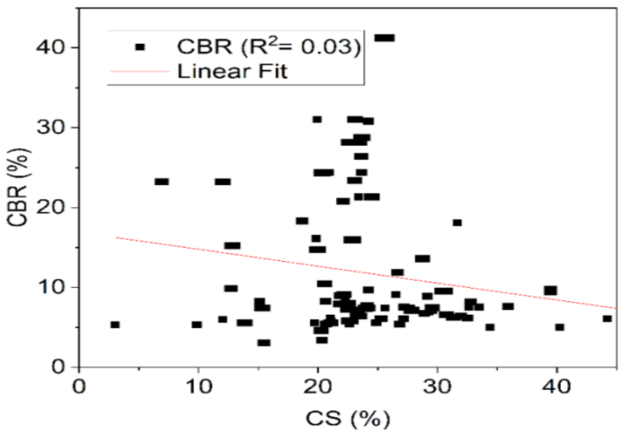
(best value = 0; worst value = +∞)

$$RMSE = \sqrt{\frac{1}{n}\sum_1^n (X_i - Y_i)^2} \qquad (4)$$

(best value = 0; worst value = +∞)

Such that X_i and Y_i are the i^th^ experimental and measured outputs, respectively; $\bar{X}_i$ and $\bar{Y}_i$ are mean of the experimental and measured outputs, respectively, whereas n refers to the total samples.





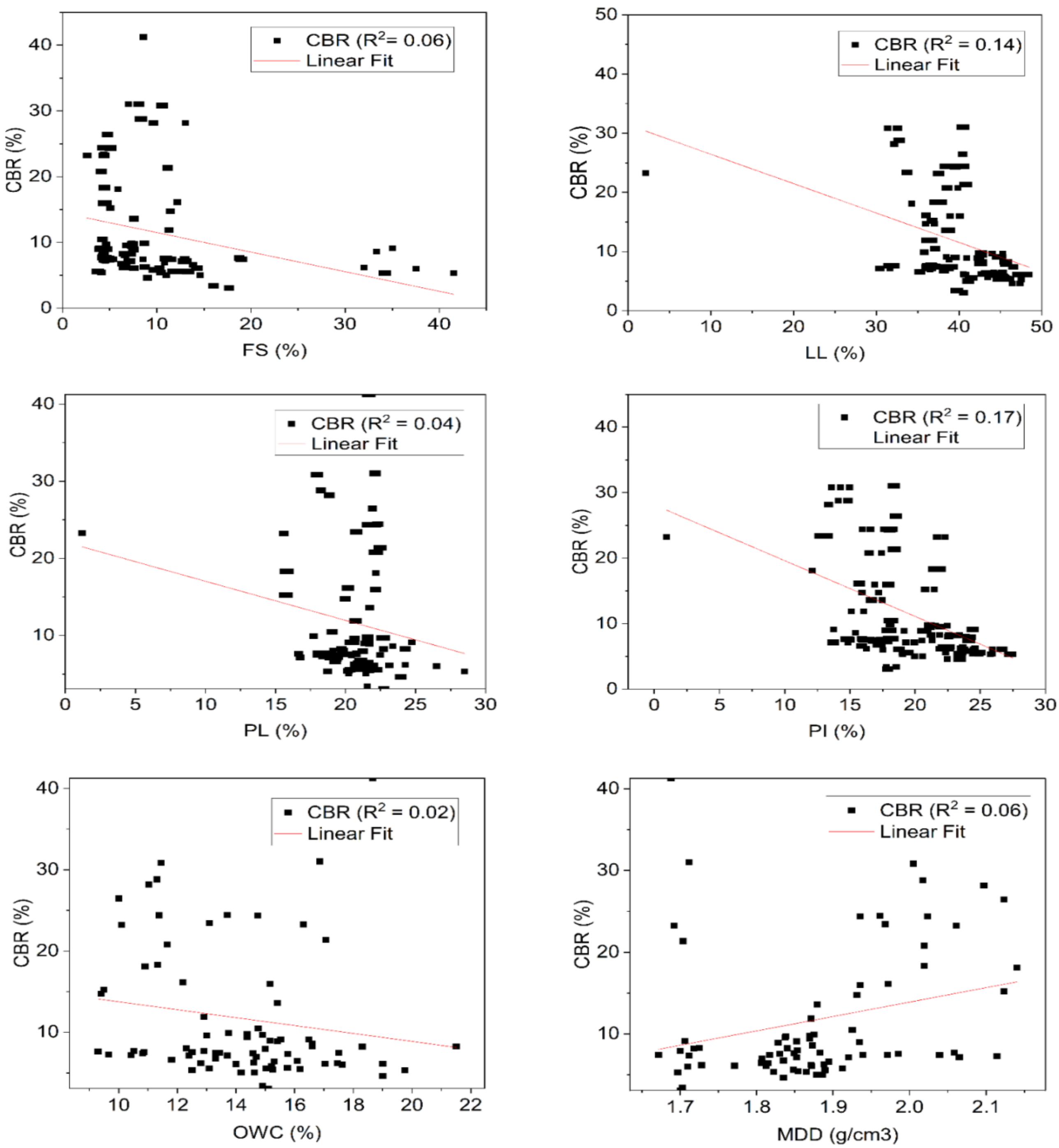


Fig 5. Correlation analysis of input variables and output variable used in this study

The level of correlation between actual and estimated values is considered to be high when the R^2 value is more than 0.8. Second, RMSE is preferable because larger residual errors are dealt with more sensitively, and RMSE≈0 represents the least errors [56]. Conversely, there are cases when RMSE isn't the best option for reaching a greater level of accuracy, and in those cases, MAE is used since it works well with both smooth and continuous data. Furthermore, higher R and lower RMSE and MAE values indicate reliable model performance and accurate calibration [52, 57].

3. Results and discussions

The CBR value for soils will depend on their density, setup moisture content, and moisture content after soaking. Note, the results of laboratory compaction should closely represent the results of field compaction. The first two of these variables must be carefully controlled during the preparation of laboratory samples for testing. Unless it can be ascertained that the soil being tested will not accumulate moisture and be affected by it in the field after construction. The CBR tests should be conducted on soaked samples [32].

The U.S. Army Corps of Engineers (USACE) used the CBR as an efficient evaluation of strength in cohesive soils. The USACE reports, "the un-soaked CBR values are high on the dry side of optimum, but there is a dramatic loss in strength as setup moisture content is increased" [58, 59]. To measure strength, from tests using penetration resistance, Hilf obtained the same results [60]. When soil is in a dry state, it exhibits high strength due to an appreciable inter-particle, the attractive force created by the high curvature of the Menisci between soil particles. However, further moisture increasing greatly reduces this friction strength due to lubrication of the soil particles [32].

The comparison of actual and anticipated results for output 'CBR' for training and testing data is shown in Fig 6 (a and b). The predicted results nearly match the expected values, as can be shown. The R^2 values of 0.98 and 0.92, respectively, reflect the close prediction of the training and testing datasets (Fig 7, Table 3). R^2 greater than 0.8 indicates that the predicted and experimental values are highly correlated [9, 52, 61-64]. MAE interprets values of 0.87 and 2.25 (Fig 8) in the same way, showing the assessment of actual and predicted findings. The forecasted values were drawn on the y-axis and the measured results were put on the x-axis. [52, 61]. The regression line's slope is near to that of the optimum fit (experimental: prediction), indicating a high association. For both the training and testing data, error-values are displayed against the data samples for the created model for prediction of CBR (Fig 8), which indicates a focused fluctuation around the zero-horizontal line. For the training and testing data, the RMSE was 1.43 and 3.96, respectively.

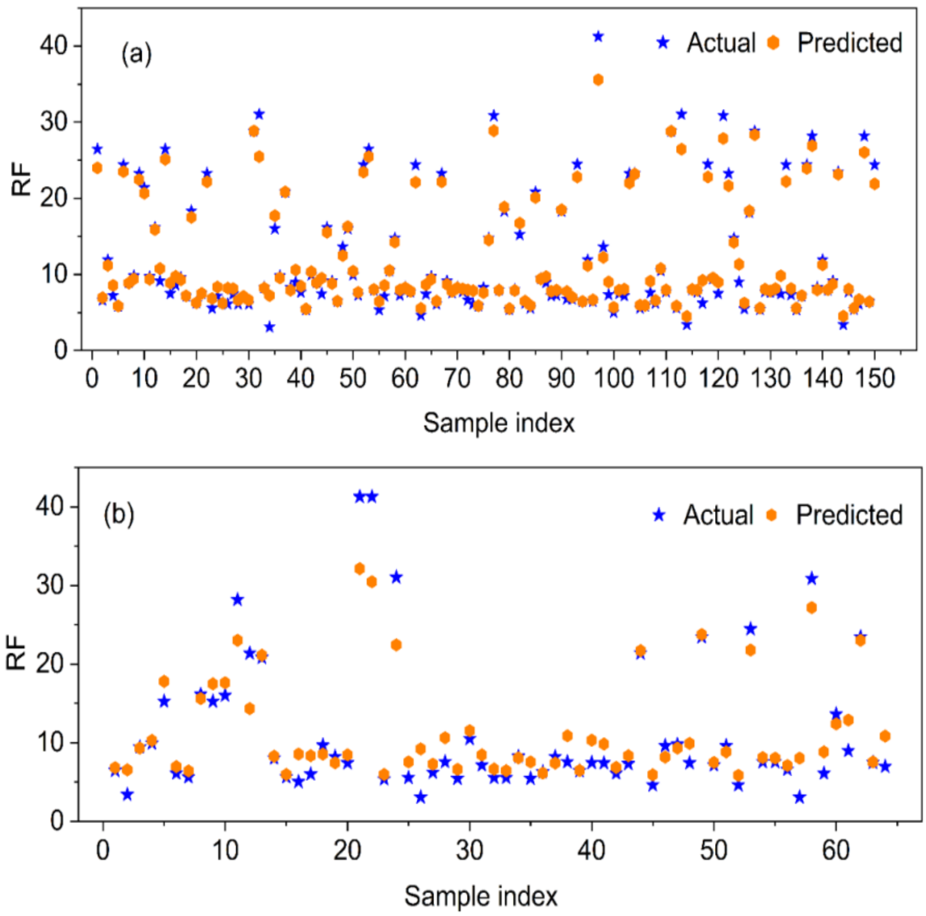


Fig 6. Comparison of the predicted and actual results: (a) training and (b) testing

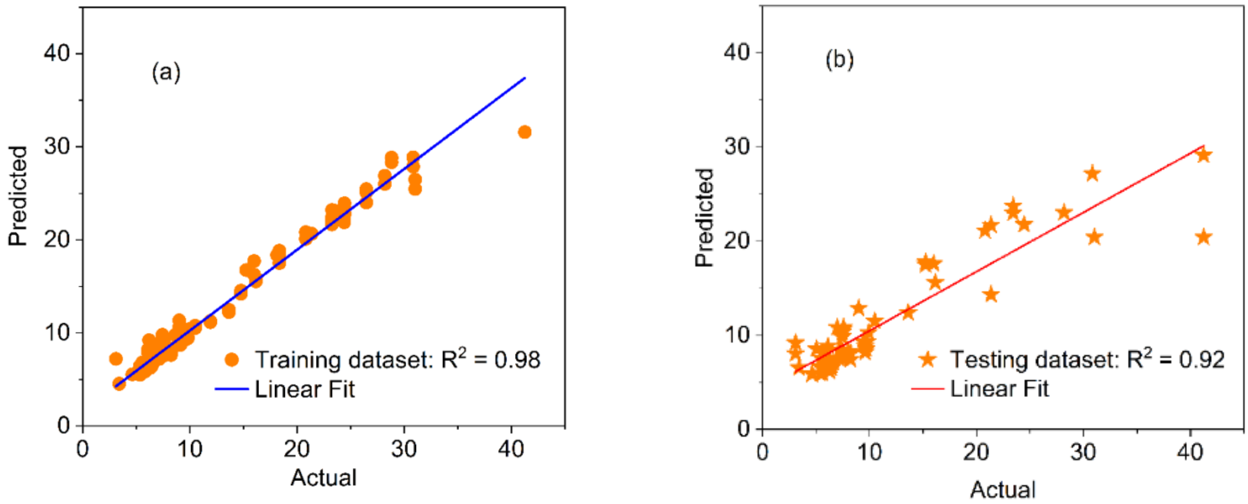


Fig 7. R^2 values of the model: (a) training dataset and (b) testing dataset

Table 3. Model performance

No	Parameters	Datasets	
		Training	Testing
1	R^2	0.98	0.92
2	MAE	0.87	2.25
3	RMSE	1.43	3.96

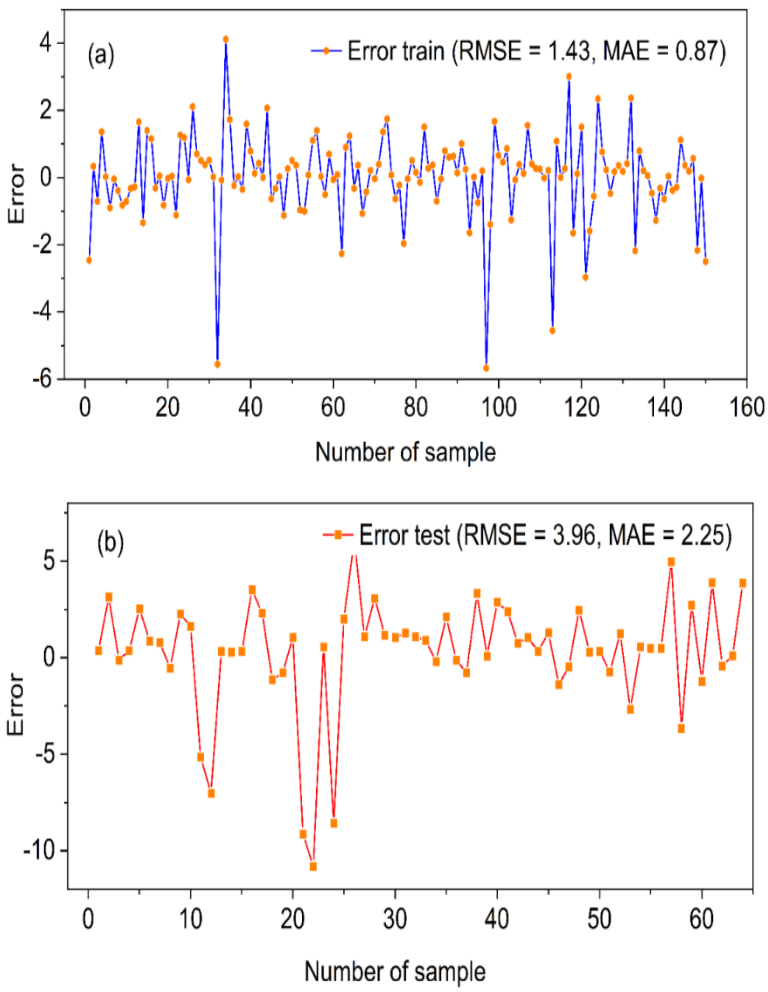


Fig 8. RMSE values of the model: (a) training, (b) testing

4. Conclusions

In this research, the Random Forest model was trained and developed for the prediction of the CBR of soils. The input parameters of the models are G, CS, FS, SC, O, LL, PL, PI, OWC, and MDD. Following are the main conclusions:

1. The constructed Random Forest model is a reliable model in predicting soil CBR based on the obtained R^2 value: 0.98, which is greater than the prediction accuracy of soft computing methodologies available in the literature (Table 1) [7, 9, 11, 65, 66]. The prediction accuracy will depend on the dataset and methodology adopted.

2. The correlation coefficient (R^2) values acquired during the testing process of the Random Forest model were much lower than those obtained during the training process, indicating over-fitting concerns. The Random Forest model showed no signs of over-fitting during the training phase, registering equal coefficient of determination values.

3. As part of future work, further refinement of the RF model can be attempted using more input values and results may be compared with other ML models.

References

[1] B.R. Christopher, C.W. Schwartz, R. Boudreaux, R.R. Berg. (2006). Geotechnical aspects of pavements, United States. *Federal Highway Administration*.

[2] AASHTO. (2017). Code T 193-13, Standard method of test for the California bearing ratio. American Association of State Highway and Transportation Officials.

[3] T. Taskiran. (2010). Prediction of California bearing ratio (CBR) of fine grained soils by AI methods. *Advances in Engineering Software*, 41(6), 886-892.

[4] B. Yildirim, O. Gunaydin. (2011). Estimation of California bearing ratio by using soft computing systems. *Expert Systems with Applications*, 38(5), 6381-6391.

[5] S. Taha, A. Gabr, S. El-Badawy. (2019). Regression and Neural Network Models for California Bearing Ratio Prediction of Typical Granular Materials in Egypt. *Arabian Journal for Science and Engineering*, 44, 8691-8705.

[6] T.F. Kurnaz, Y. Kaya. (2019). Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network. *The European Physical Journal Plus*, 134, 326.

[7] A.R. Tenpe, A. Patel. (2020). Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Materials and Pavement Design*, 21(5), 1183-1200.

[8] T.V. Nagaraju, C.D. Prasad, M.J. Raju. (2020). Prediction of California bearing ratio using particle swarm optimization. *Soft Computing for Problem Solving, Springer*, pp 795-803.

[9] K.C. Onyelowe, M. Iqbal, F.E. Jalal, M.E. Onyia, I.C. Onuoha. (2021). Application of 3-algorithm ANN programming to predict the strength performance of hydrated-lime activated rice husk ash treated soil. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 4, 259-274.

[10] M.N.A. Raja, S.K. Shukla, M.U.A. Khan. (2021). An intelligent approach for predicting the strength of geosynthetic-reinforced subgrade soil. *International Journal of Pavement Engineering*, 23(10), 3505-3521.

[11] A. Bardhan, C. Gokceoglu, A. Burman, P. Samui, P.G. Asteris. (2021). Efficient computational techniques for predicting the California bearing ratio of soil in soaked conditions. *Engineering Geology*, 291, 106239.

[12] A. Bardhan, P. Samui, K. Ghosh, A.H. Gandomi, S. Bhattacharyya. (2021). ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. *Applied Soft Computing*, 110, 107595.

[13] A.R. Tenpe, A. Patel. (2020). Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Materials and Pavement Design*, 21(5), 1183-1200.

[14] I. Yilmaz, O. Kaynar. (2011). Multiple

regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications*, 38(5), 5958-5966.

[15] M. Iqbal, K.C. Onyelowe, F.E. Jalal. (2021). Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash-modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 4, 207-225.

[16] S. Al-Busultan, G.K. Aswed, R.R. Almuhanna, S.E. Rasheed. (2020). Application of artificial neural networks in predicting subbase CBR values using soil indices data. *IOP Conference Series: Materials Science and Engineering, IOP Publishing*, p. 012106.

[17] A.N. Ramakrishna, A.V. Pradeep Kumar, K. Gowda. (2011). Complex CBR (of BC soil-RHA-cement Mix) estimation: made easy by ANN approach [a soft computing technique]. *Advanced Materials Research, Trans Tech Publications*, pp 675-679.

[18] A. Ghorbani, H. Hasanzadehshooiili. (2018). Prediction of UCS and CBR of microsilica-lime stabilized sulfate silty sand using ANN and EPR models; application to the deep soil mixing. *Soils and Foundations*, 58(1), 34-49.

[19] S.A. Kumar, J.P. Kumar, J. Rajeev. (2013). Application of Machine Learning Techniques to Predict Soaked CBR of Remolded Soils. *International Journal of Engineering Research & Technology (IJERT)*, 2(6), 3019-3024.

[20] T.V. Nagaraju, R. Gobinath, P. Awoyera, M.A.H.A. Sayyed. (2021). Prediction of California Bearing Ratio of Subgrade Soils Using Artificial Neural Network Principles. *Communication and Intelligent Systems, Springer*, pp 133-146.

[21] K.L. Timani, R.K. Jain. (2021). Assessment of Soaked California Bearing Ratio of Clay-Gravel Mixtures Using Artificial Neural Network Modeling. *Proceedings of the Indian Geotechnical Conference 2019*, Springer, pp 437-445.

[22] D.T. Bui, V.-H. Nhu, N.-D. Hoang. (2018). Prediction of soil compression coefficient for urban housing project using novel integration machine learning approach of swarm intelligence and multi-layer perceptron neural network. *Advanced Engineering Informatics*, 38, 593-604.

[23] M.R. Islam, A.C. Roy. (2020). Prediction of California bearing ratio of fine-grained soil stabilized with admixtures using soft computing systems. *Journal of Civil Engineering, Science and Technology*, 11(1), 28-44.

[24] J. Hassan, B. Alshameri, F. Iqbal. (2021). Prediction of California Bearing Ratio (CBR) Using Index Soil Properties and Compaction Parameters of Low Plastic Fine-Grained Soil. *Transportation Infrastructure Geotechnology*, 9, 764-776.

[25] TCVN. (2012). Code 5729:2012, Expressway − Specifications for design. Ministry of Science and Technology of Viet Nam, www.most.gov.vn.

[26] ASTM. (2012). Code D698 - 12, Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort (12 400 ft-lbf/ft3 (600 kN-m/m3)), ASTM International, West Conshohocken, PA, 2012, www.astm.org.

[27] AASHTO. (2019). Code T 99, Standard Method of Test for Moisture–Density Relations of Soils Using a 2.5-kg (5.5-lb) Rammer and a 305-mm (12-in.) Drop, American Association of State Highway and Transportation Officials (AASHTO).

[28] ASTM. (2012). Code D1557 - 12, Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Modified Effort (56,000 ft-lbf/ft3 (2,700 kN-m/m3)), ASTM International, West Conshohocken, PA, 2012, www.astm.org.

[29] AASHTO. (2020). Code T 180, Standard Method of Test for Moisture–Density Relations of Soils Using a 4.54-kg (10-lb) Rammer and a 457-mm (18-in.) Drop, American Association of

State Highway and Transportation Officials (AASHTO).

[30] ASTM. (2016). Code D4253 - 16e1, Standard Test Methods for Maximum Index Density and Unit Weight of Soils Using a Vibratory Table, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

[31] ASTM. (2016). Code D4254 - 16, Standard Test Methods for Minimum Index Density and Unit Weight of Soils and Calculation of Relative Density, ASTM International, West Conshohocken, PA, 2016, www.astm.org.

[32] V.R. Schaefer, L. Stevens, D. White, H. Ceylan. (2008). Design Guide for Improved Quality of Roadway Subgrades and Subbases. Iowa Highway Research Board, IHRB Project TR-525.

[33] AASHTO. (2017). Code T 89-13, Standard Method of Test for Determining the Liquid Limit of Soils. American Association of State and Highway Transportation Officials.

[34] ASTM. (2017). Code D4318 - 17e1, Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils, ASTM International, West Conshohocken, PA, 2017, www.astm.org.

[35] AASHTO. (2020). Code T 88-20, Standard Method of Test for Particle Size Analysis of Soils, American Association of State and Highway Transportation Officials.

[36] ASTM. (2007). Code D422-63(2007)e2, Standard Test Method for Particle-Size Analysis of Soils (Withdrawn 2016), ASTM International, West Conshohocken, PA, 2007, www.astm.org.

[37] W. Zhou. (2018). Cone Penetrometer. P.T. Bobrowsky, B. Marker (Eds.), Encyclopedia of Engineering Geology, Springer International Publishing, Cham, pp 182-183.

[38] AASHTO. (2018). Code T 267-86 (2018), Standard Method of Test for Determination of Organic Content in Soils by Loss of Ignition, American Association of State and Highway Transportation Officials.

[39] AASHTO. (2017). Code T 193-13 (2017), Standard method of test for the California bearing ratio, American Association of State Highway and Transportation Officials.

[40] ASTM. (2016). Code D1883-16, Standard Test Method for California Bearing Ratio (CBR) of Laboratory-Compacted Soils, American Society for Testing and Materials. https://www.astm.org/Standards/D1883. ASTM International, West Conshohocken, PA, 2016, www.astm.org.

[41] M.P. Rollings, R.S. Rollings. (1996). Geotechnical materials in construction. *McGraw-Hill Professional Publishing*.

[42] AASHTO. (2017). Code M 145-91 (2017), Standard Specification for Classification of Soils and Soil-Aggregate Mixtures for Highway Construction Purposes. American Association of State and Highway Transportation Officials.

[43] ASTM. (2017). Code D2487 - 17e1, Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System), ASTM International, West Conshohocken, PA, 2017, www.astm.org.

[44] S.I.K. Ampadu. (2007). A laboratory investigation into the effect of water content on the CBR of a subgrade soil. *Experimental unsaturated soil mechanics, Springer*, pp 137-144.

[45] L. Breiman. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

[46] S.W. Kwok, C. Carter. (1990). Multiple decision trees. *Machine Intelligence and Pattern Recognition*, 9, 327-335.

[47] L. Breiman, Berkeley. (1994). Bagging predictors (Technical Report 421). UC Berkeley Statistics.

[48] B. Efron. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1), 1-26.

[49] Y. Amit, D. Geman, K. Wilder. (1997). Joint induction of shape features and tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11), 1300-1305.

[50] R. Schapire, Y. Freund. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Second European Conference on Computational Learning*

Theory, pp 23-37.

[51] L. Breiman. (1996). Bias, variance, and arcing classifiers, Techical Report 460. *University of California, Berkeley*.

[52] F.E. Jalal, Y. Xu, M. Iqbal, M.F. Javed, B. Jamhiri. (2021). Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. *Journal of Environmental Management*, 289, 112420.

[53] M. Iqbal, D. Zhang, F.E. Jalal, M.F. Javed. (2021). Computational AI prediction models for residual tensile strength of GFRP bars aged in the alkaline concrete environment. *Ocean Engineering*, 232, 109134.

[54] D. Chicco, M.J. Warrens, G. Jurman. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623.

[55] J.K. Holodinsky, A.Y.X. Yu, M.K. Kapral, P.C. Austin. (2021). Comparing regression modeling strategies for predicting hometime. *BMC Medical Research Methodology*, 21, 138.

[56] I.O. Alade, M.A. Abd Rahman, T.A. Saleh. (2019). Modeling and prediction of the specific heat capacity of Al2 O3/water nanofluids using hybrid genetic algorithm/support vector regression model. *Nano-Structures & Nano-Objects*, 17, 103-111.

[57] I. Azim, J. Yang, M.F. Javed, M.F. Iqbal, Z. Mahmood, F. Wang, Q.-f. Liu. (2020). Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. *Structures*, 25, 212-228.

[58] F. Ariema, B.E. Butler. (1990). Chapter 6 in Guide to Earthwork Construction, State of the Art Report 8. *Transportation Research Board National Research Council*.

[59] H.N. Atkins. (1997). Highway Materials, Soils, and Concretes. *Upper Saddle River, N.J.: Prentice Hall*.

[60] J.W. Hilf. (1956). Foundation Engineering Handbook. *New York: Van Nostrand Reinhold*.

[61] M. Iqbal, D. Zhang, F.E. Jalal, M.F. Javed. (2021). Computational AI prediction models for residual tensile strength of GFRP bars aged in the alkaline concrete environment. *Ocean Engineering*, 232, 109134.

[62] M. Iqbal, K.C. Onyelowe, F.E. Jalal. (2021). Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash-modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques. *Multiscale and Multidisciplinary Modeling, Experiments and Design*, 4, 207-225.

[63] I. Azim, J. Yang, M.F. Javed, M.F. Iqbal, Z. Mahmood, F. Wang, Q.-f. Liu. (2020). Prediction model for compressive arch action capacity of RC frame structures under column removal scenario using gene expression programming. *Structures*, 25, 212-228.

[64] I. Azim, J. Yang, M.F. Iqbal, M.F. Javed, S. Nazar, F. Wang, Q.-f. Liu. (2020). Semi-analytical model for compressive arch action capacity of RC frame structures. *Structures*, 27, 1231-1245.

[65] A. Bardhan, P. Samui, K. Ghosh, A.H. Gandomi, S. Bhattacharyya. (2021). ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. *Applied Soft Computing*, 110, 107595.

[66] T.F. Kurnaz, Y. Kaya. (2019). Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network. *The European Physical Journal Plus*, 134(7), 326.