Vietnam Journal of Earth Sciences, 46(3), 399-410, https://doi.org/10.15625/2615-9783/20766

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Application of hybrid modeling to predict California bearing ratio of soil

Huong Thi Thanh Ngo*, Quynh-Anh Thi Bui, Nguyen Van Vi, Nguyen Thi Bich Thuy

University of Transport Technology, Hanoi, Vietnam

Received 18 Jannuary 2024; Received in revised form 23 April 2024; Accepted 10 May 2024

ABSTRACT

California Bearing Ratio (CBR) is used to assess bearing capacity, deformation characteristics of roadbed soil, and base layer material in pavement structure. In general, CBR is often determined by laboratory or in-situ tests. However, it is time- and cost-consuming to conduct this experiment because this test requires cumbersome equipment such as a compressor. In this study, two Artificial Intelligence models are developed, including a simple model (Decision Tree Regression, DT) and a hybrid model (AdaBoost - Decision Tree, AB-DT). Using 214 data samples from Van Don - Mong Cai expressway, Vietnam, 10 input variables of the model were considered namely particle composition (content of gravel (X_1), coarse sand (X_2), fine sand (X_3), silt clay (X_4), organic (X_5)), Atterberg limits (Liquid limit (X_6), Plastic limit (X_7), Plastic index (X_8)), and compaction curve (optimum water content (X_9) and maximum dry density (X_{10})). The developed models were evaluated by using a variety of statistical indicators, including coefficient of determination (R^2), Root mean square error (RMSE), and Mean absolute error (MAE). The results show that AB-DT model has higher accuracy than the DT model. Moreover, the SHAP value analysis shows that the variable X_{10} influences the CBR value the most. Thus, it implies that applying the AB-DT model to effectively predict the CBR of the roadbed soil saves time and money for experiments.

Keywords: California Bearing Ratio, AdaBoost, Decision Tree, Artificial Intelligence, Quang Ninh.

1. Introduction

Evaluating soil mechanical parameters is necessary in highway construction (Atkins, 1997). The bearing capacity of the road foundation is usually assessed according to the CBR. It is the ratio (in percent) between the compressive pressure (caused by the compressor) on the test specimen and the compression pressure on the standard specimen corresponding to the exact specified penetration depth (Brown, 1996). The CBR test is widely used in the world to determine the strength and elastic modulus of the foundation, thereby designing the pavement thickness. The CBR test was carried out on a specimen compacted at the optimum moisture content corresponding to the specified compaction method (Ariema et al., 1990; Schaefer et al., 2008). Experiments can be carried out on natural soil in non-immersion and immersion conditions. A soil's CBR value is influenced by several factors, such as particle size, plastic index, water content, void ratio, and specific gravity (Hight et al., 1982; Ampadu, 2007; Mishra et al., 2010). Determining CBR in the laboratory or situ is a time-consuming process that frequently yields

*Corresponding author, Email: huongntt@utt.edu.vn

misleading readings due to sample distortion, testing irresponsibility, and inaccuracy equipment. Moreover, the process requires time, cost, and effort (Alam et al., 2020; Khasawneh et al., 2022). Therefore, a convenient, fast, and reliable CBR prediction is necessary.

Over the last decade, many scholars have used statistical methods and proposed simple and multiple regressions to estimate CBR values. (Rehman et al., 2017; González Farias et al., 2018; Katte et al., 2019; Haupt et al., 2021). Black (1962) predicted CBR from the Plastic and Liquid Index. Johnson and Bhatia (1969) proposed a regression for CBR estimating relied on particle size distribution and plasticity data. Agarwal and Ghanekar (1970) suggested an equation related to CBR and the Atterberg Limit. The CBR estimation gained from statistical approaches has limited generalizability and can only be potentially applicable to local datasets. This is partially because of the small size of the original dataset used for predictive models, the data specificity, the nonlinearity relationship, and the complexity associated with identifying soil characteristics, data dispersion, and particle soil composition. (Black, 1962; 1962).

Machine learning (ML) development for solving real-world problems is receiving global attention in many fields. (Pham et al., 2016; Bui et al., 2022; Hadzima-Nyarko et al., 2022). They have recently been able to deal comprehensively with high nonlinearity and complex problems. (Thanh et al., 2020; Pham et al., 2021; Thai et al., 2022). This is also an effective solution for predicting CBR, as concluded by several researchers. (Venkatasubramanian et al., 2011; Quan et al., 2021; Raja et al., 2022). However, the results of the above studies also show the prediction accuracy paradox. The R^2 value strongly depends on the dataset size; it was reported that the larger dataset gives a smaller R^2 value and vice versa. The larger dataset is more representative and, therefore, expected to be more reliable. It was reported that the hybrid model is often used to improve the prediction accuracy of a small dataset (Kamrul Alam et al., 2024; Ho et al., 2022). Because, in the hybrid model, the optimization algorithm is used to adjust of the based model. Thus, hybrid models were proposed to address these shortcomings in the case of the small dataset. ELM-ANSI (Bardhan et al., 2021), ANN-LMBP (Onyelowe et al., 2023), BP-CG (Raja et al., 2021) in estimating CBR value, it achieved relatively high accuracy (R^2 higher than 0.8). However, in these studies, the influence of input parameters on CBR has not been evaluated. Therefore, developing and applying new, more powerful ML algorithms for enhancing the accuracy of CBR prediction is essential.

The study aims to develop single computational models (regression decision trees) and a hybrid model based on a boosting algorithm (AdaBoost - Decision Tree) to predict CBR. These effective and popular models are successfully applied to many practical problems. (Rätsch et al., 2001; Tso et al., 2007; Chengsheng et al., 2017; Pekel, 2020; Rakhra et al., 2021). This study used a dataset of 214 samples based on 10 input variables related to particle composition, Atterberg limits, and compaction curves. The model accuracy was evaluated by a variety of performance indicators such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and correlation coefficient (R^2). Moreover, the input influence on output parameters is evaluated based on the Shapley Additive explanation (SHAP) value.

2. Data and Method

2.1. Data used

A database was built by using soil testing results obtained at the Van Don - Mong Cai freeway in Quang Ninh province, Vietnam. The route has 4 lanes with a 120 km/h design speed. The route is more than 80 km long, going through many areas with different complex geology, soil type, and status.

The database's reliability greatly influences the predictability of the developed ML model. Trustable data might include significant statistical representative samples and ensure the data distribution complies with the basic principles of static (Maulud et al., 2020). 214 soil specimens taken at the study site were prepared for testing during construction. Particle size, liquid limit, organic content, compaction curve, and CBR tests were conducted.

A previous study surveyed the effect of the ratio between training/testing datasets on the performance of the prediction of the machine learning method. It was indicated that a ratio of 70/30 gave a better prediction accuracy. (Nguyen et al., 2021). Therefore, the database in this study is split into two groups based on uniform distribution, including the training and testing set. The first 70 % trains the DT and AB-DT models, while the second 30 % evaluates predictive accuracy (Nguyen et al., 2021).

2.1.1. CBR (outputs)

CBR relates the bearing capacity of a laboratory or field compacted soil material to the standard crushed stone. CBR test was first introduced and guided by the ASTM and the AASHTO to assess the loading capacity of the roadbed and base or subbase of pavement structure (Atkins, 1997). Specimen humidity and testing load can be varied according to the requirements of each project and standard specification. In this study, CBR tests were performed according to the guidelines of ASTM D1883. The CBR test was performed with a sample saturated with water after 4 immersion days. The compressive loads to produce penetrations of 1 inch and 2 inches are measured. The CBR is then determined by dividing the actual experiment result by the reference based on standard crushed-stone results. Different minimum CBR requirements depend on the road grade (Ariema et al., 1990; Atkins, 1997).

2.1.2. Affecting factors (inputs)

Various parameters influence soil CBR, such as size of particles, soil structure, plastic index, water content, and specific gravity (Ampadu, 2007; Katte et al., 2019; Quan et al., 2021). Samples were fabricated at in situ humidity and dry density, respectively. While the in-situ specific gravity of soils can be predicted relatively accurately, estimating the in-situ water content could be difficult. When water content rises and saturation increases, CBR is significantly reduced on optimal moisture (Abdulnabi et al., 2020). The particle content and distribution make the skeleton of the soil structure (Rehman et al., 2017; Alam et al., 2020; Onyelowe et al., 2023). Therefore, it is closely related to the soil's properties, condition and bearing capacity. In this study, 10 input variables include particle composition (content of gravel (X_1), coarse sand (X_2), fine sand (X_3), silt clay (X_4), organic (X_5)), Atterberg limit (Liquid limit (X_6), Plastic limit (X_7), Plastic index (X_8)), and compaction curve (optimum water content (X_9), maximum dry density (X_{10})) was used. These parameters are detailed by statistical analysis in Table 1. In addition, the frequency distribution graph of each input and output is shown in Fig. 1.

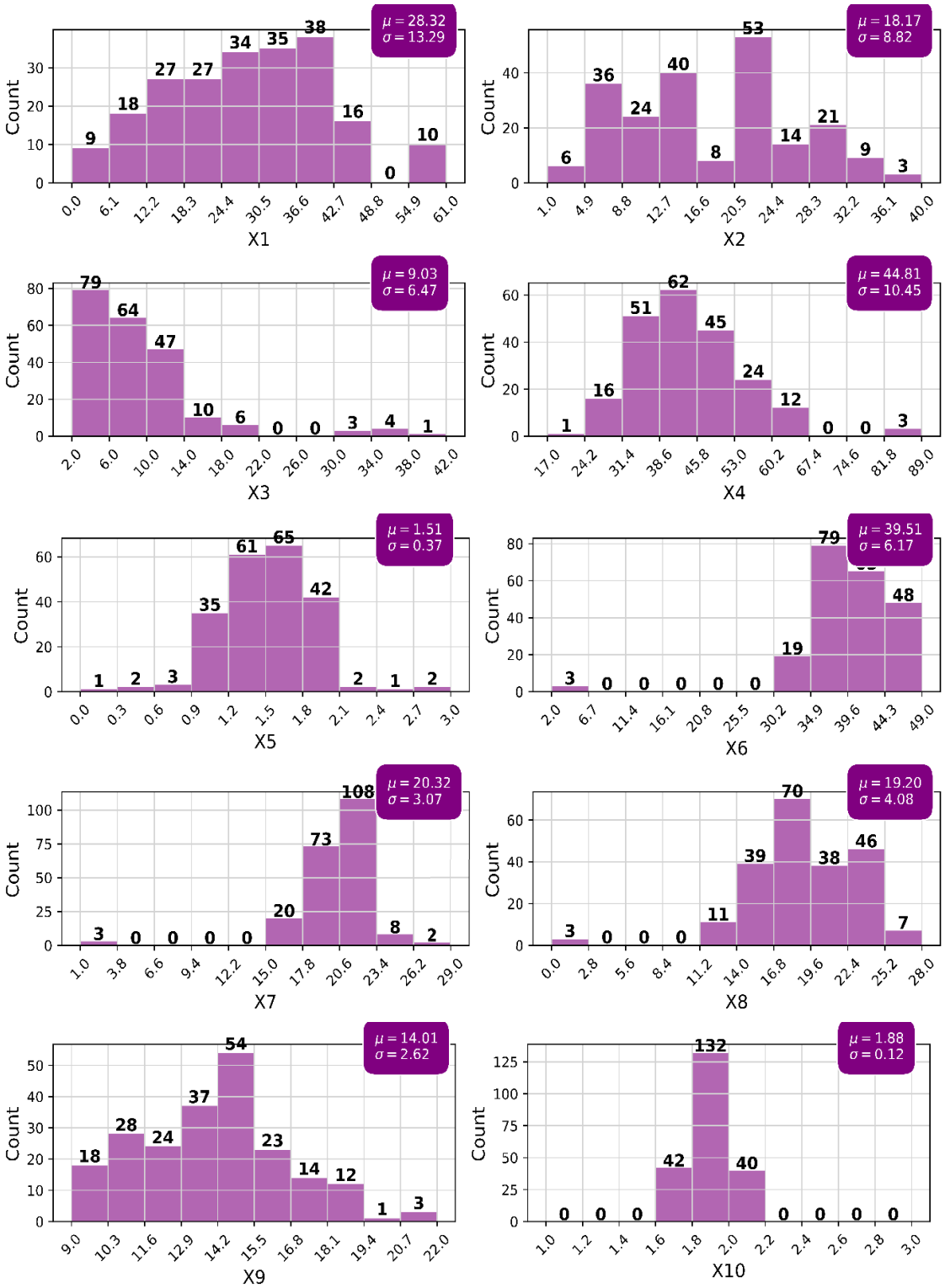

Figure 1. Input statistical analysis in this study

Table 1. Analysis parameters in this study

Parameter	Ab	count	mean	std	min	25%	50%	75%	max
Gravel content	X_1	214	28.32	13.29	0.10	18.28	27.50	37.85	60.82
Coarse sand content	X_2	214	18.17	8.82	1.90	10.65	16.75	24.28	39.60
Fine sand content	X_3	214	9.04	6.47	2.50	4.60	7.25	11.00	41.50
Silt-clay content	X_4	214	44.81	10.45	17.87	37.75	44.55	49.20	88.70
Organic content	X_5	214	1.51	0.37	0.12	1.25	1.51	1.77	2.94
Liquid limit	X_6	214	39.52	6.17	2.08	36.64	39.99	43.51	48.45
Plastic limit	X_7	214	20.32	3.07	1.17	19.29	20.84	21.89	28.49
Plastic index	X_8	214	19.20	4.08	0.91	16.83	18.44	22.32	27.48
Optimum water content	X_9	214	14.01	2.62	9.30	12.19	14.28	15.40	21.50
Maximum dry density	X_{10}	214	1.88	0.12	1.67	1.82	1.87	1.96	2.14
CBR	Y	214	11.80	8.18	3.09	6.47	7.95	15.25	41.26

Table 2 describes the correlation between those inputs (from X_1 to X_{10}) and output (Y). The Pearson correlation coefficient was calculated and noted in each pair of parameters. It can be seen that Y does not have a solid direct correlation with other input parameters. Besides, there are ten input variables, each with a specific influence on the output. As a result, why are all the variables linearly independent of the output? It can be concluded that the variables taken in this study are all independent and are used to build a correlation with Y (dependent variable).

Table 2. Input correlation analysis in this study

Parameter	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y
X1	1.00	-0.29	-0.54	-0.67	-0.14	0.15	-0.05	0.27	-0.27	0.51	0.17
X2	-0.29	1.00	-0.18	-0.36	-0.14	-0.07	0.14	-0.21	-0.25	0.18	0.03
X3	-0.54	-0.18	1.00	0.22	0.11	0.12	0.21	0.03	0.16	-0.29	-0.24
X4	-0.67	-0.36	0.22	1.00	0.23	-0.23	-0.19	-0.21	0.43	-0.59	-0.06
X5	-0.14	-0.14	0.11	0.23	1.00	0.07	0.03	0.08	0.11	-0.19	-0.26
X6	0.15	-0.07	0.12	-0.23	0.07	1.00	0.82	0.90	0.39	-0.23	-0.37
X7	-0.05	0.14	0.21	-0.19	0.03	0.82	1.00	0.48	0.37	-0.19	-0.19
X8	0.27	-0.21	0.03	-0.21	0.08	0.90	0.48	1.00	0.31	-0.20	-0.42
X9	-0.27	-0.25	0.16	0.43	0.11	0.39	0.37	0.31	1.00	-0.77	-0.16
X10	0.51	0.18	-0.29	-0.59	-0.19	-0.23	-0.19	-0.20	-0.77	1.00	0.25
Y	0.17	0.03	-0.24	-0.06	-0.26	-0.37	-0.19	-0.42	-0.16	0.25	1.00

2.2. Methods used

2.2.1. Decision Tree (DT)

DT is an essential and standard ML algorithm for predictive modeling. A DT is a supervised learning algorithm classifies problems (Myles et al., 2004). The algorithm can be used for both categorical and continuous variables. DT is divided into two types: regression tree and classification tree (Quinlan, 1996). In this study, a regression decision tree model is used.

A DT relies on a sequence of rules to predict the class of an object (Czajkowski et al., 2016). Each internal node of the DT denotes an attribute. Connecting the node to its children indicates an individual value for that attribute. Each node in the leaf symbolizes the categorical attribute's predicted value. By calculating the information gain (IG), the DT learns to predict the value of categorical attributes by relying on the training data set to select the root node to split the tree. This process is repeated recursively until no further tree splitting can be performed (Xu et al., 2005).

The training data for a DT is a set formed data: (x, y) = (x_1, x_2, ..., x_i, y) where: y is

called a categorical attribute (also known as a target or dependent variable), and x_1, x_2, ..., x_i is an independent attribute (Myles et al., 2004).

2.2.2. AdaBoost - Decision Tree (AB-DT)

AdaBoost, a boosting algorithm, trains new models based on re-weighting existing data points to help new models focus more on data samples being mis-learned, thereby reducing the loss of the model (Schapire, 2013a). First, initialize the initial weight to be equal (equal to 1/ N) for each data point. At the i^{th} iteration, a new model w_i (weak learner) training is added, and the loss (error) value is calculated. Thereby, the confidence score c_i of the newly trained model is determined. Then, the primary model $W' = W + c_i*w_i$ is updated. Finally, the data points (Incorrectly guessed data points increase the weights, correctly guessed data points decrease the weights) are re-weighted. The loop is done by adding the next model, i+1 (Domingo et al., 2000).

AdaBoost - Decision Tree is a hybrid model with the idea that instead of trying to build a single model (DT model or AdaBoost model), they are combined correctly into an even better model (Solomatine et al., 2004).

2.2.3. SHAP

SHAP is an abbreviation of Shapley's additive explanation, a method of calculating the effect of an attribute (a feature) on the meaning of a target object (Futagami et al., 2021). The concept is that each feature is considered an individual player, and the dataset is considered a team. Every participant contributes differently to the overall success of the team. This solution distributes profits and costs equitably to many players in game theory (Mokhtari et al., 2019). In the ML model, the SHAP value is the mean of the expected marginal contribution of each input variable to the output variable after considering all possible combinations (Mishra et al., 2010). As a result, SHAP uses combinatorial calculations to determine the effect of each attribute on the target object before retraining the model on all possible feature combinations. The mean initial value of a feature's impact on a target object may be employed to assess its significance (Wang et al., 2022).

2.2.4. Validation indicators

Some statistical indicators, including R^2, RMSE, and MAE, were employed in the study to verify the results obtained by the generated ML model. R^2 is the square of the correlation coefficient (R) between the anticipated and actual outcomes, ranging between 0 and 1. A high value of R^2 suggests that the anticipated and actual values are well correlated (Granger et al., 1974). The RMSE is a standard error indicator squared of the mean difference between the developed model's anticipated and actual output, whereas the MAE calculates the mean error between them. In contrast to R^2, lower RMSE and MAE values indicate higher accuracy ML algorithm performance (Willmott et al., 2005). To verify the prediction models, all of the criteria must be met. The formulas to determine these criteria were given in previous publications (Peng et al., 2002; Hair et al., 2012; Barrett, 2007)

3. Result and discussion

The performance of both models (DT and AB-DT) is also assessed by statistical indicators such as R^2, RMSE, and MAE. A comparison of the statistical analysis results for the training and test process is presented in Table 3. The R^2 values of AB-DT are higher than those of the DT model for both training and testing parts. Besides, the values of RMSE and MAE of the AB-DT model are smaller than those of the DT for both training and testing parts. These results indicate that the hybrid model (AB-DT) performs better than the single model (DT). This result is similar to some studies worldwide (Abnoosian

et al., 2023). The AdaBoost algorithm combined with DT has dramatically improved the performance of the DT model in some other problems, as indicated in the previous studies (Schapire, 2013b; Hastie et al., 2009). Thus, applying the AB-DT to predict the CBR of the roadbed soil is practical and feasible.

Table 3. Comparison of the models' performance results

No	Model	Training	Testing
R^2			
1	DT	0.851	0.810
2	AB-DT	0.967	0.934
MAE			
1	DT	2.175	2.346
2	AB-DT	1.250	1.423
RMSE			
1	DT	3.161	3.411
2	AB-DT	1.561	1.739

In the training process, hyperparameters are selected before parameter determination and they are instrumental in finding optimal parameter combinations through Grid Search as shown in Table 4.

Table 4. Hyper-parameters of DT and AB-DT using Grid Search

No	Hyper-parameters	Models	
		DT	AB-DT
1	max_depth	4	4
2	learning_rate	0.2	0.2
3	n_estimators	400	400

Typical AB-DT model results are chosen using the criterion (R^2), which is slightly stricter regarding the model's predictive ability. Figure 2a and 2b show the regression analysis for the training and testing dataset, respectively. Each figure's blue and red diagonals represent the regression lines corresponding to the training and testing datasets. The linear regression lines seem near the data points, confirming the strong correlation between the estimated and actual CBR ratio. The predictors are calculated and expressed for each case: R^2 = 0.967 for training data and R^2 = 0.934 for testing data. The results of R^2 in this study are similar to those obtained in the previous study (Trong et al., 2021).

The RMSE plot of the AB-DT model was plotted in Figure 2a and 2b (for training and testing sets, respectively). The errors for the training and test data sets are generally minor. Furthermore, the percentage of observation error in a scope can be done effectively using the cumulative distribution (in red). For instance, for the training data set, the sample rate has an error in the range of [-2; 2] is about 80%. Additionally, for the testing data set, an 80% error between the experimental predicted values in range of [-2; 2] is estimated. The small error percentages in both datasets (RMSE = 1.904 and RMSE = 1.250, respectively) show that the AB-DT model is an excellent choice for rapidly estimating the CBR of soil. Besides, it can be observed that there are two best-fit lines between predicted and actual values for both training and testing parts.

Moreover, the SHAP value method automatically reflects the expected feature contributions. In addition, the impacts of each attribute (input parameter) acting on the prediction are separated into 2 portions: the valuable effect of the attribute itself and its common combined effect and other attributes. Figure 3 depicts the rankings of the attribute influence and the overall impact, as well as the individual attribute impact to the total impact ratio. The provided SHAP value is illustrated by a dot per attribute row for each sample. The SHAP value of each attribute determines the x dot's arrangement, and the dots match up across each attribute row to show the frequency. Color serves as a way to display the attribute's original value. The overview bar chart shows overall attribute importance since the mean absolute SHAP value for all attributes across all models reflects it. The results show that the variables

X_{10} (Maximum dry density) have the most influence, and (X_5) Organic content has the most negligible influence on CBR. This is also consistent with the results found in some previous studies (Patel et al., 2010; Bharath et al., 2021; Lakshmi et al., 2021).

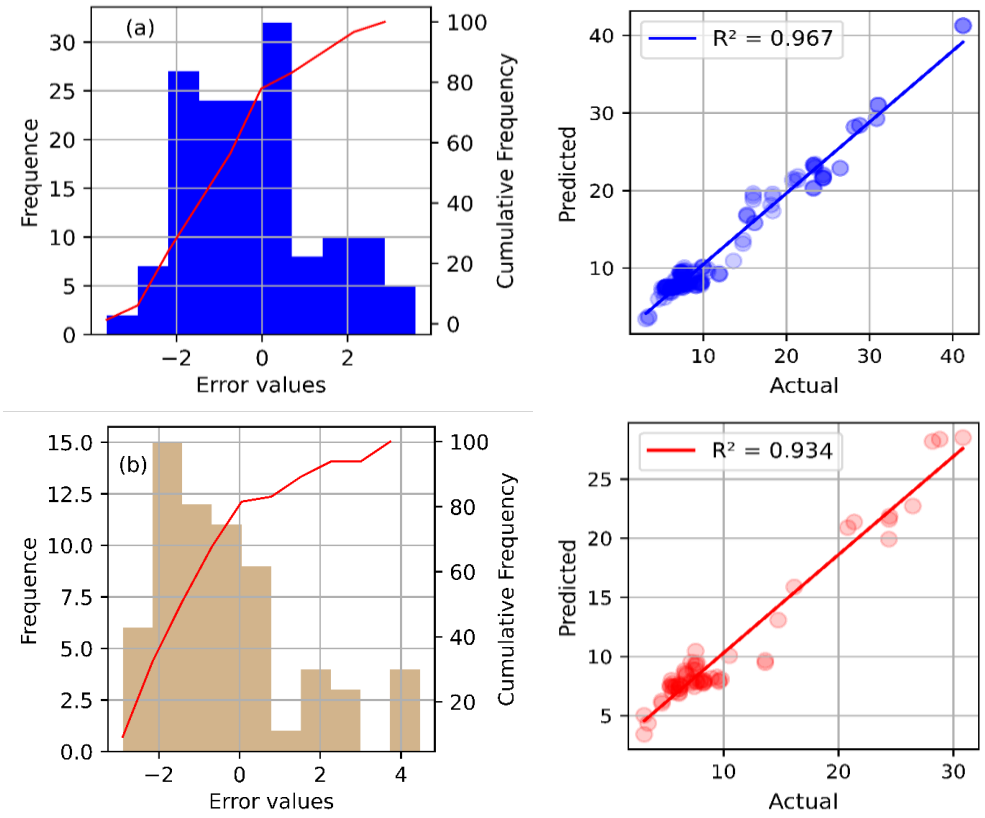

Figure 2. RMSE, R^2 values with (a) training, (b) testing data

4. Conclusions

This study used two ML models, including a simple model (DT) and a hybrid model (AB-DT), to predict the CBR value of roadbed soil. A collected dataset, including 214 CBR test results of roadbed soil of the Van Don - Mong Cai expressway project with 10 input variables, has been used to build models. The results show that the hybrid model (AB-DT) has outstanding reliability compared to the single model (DT) based on the statistical probability assessments. Furthermore, the results of the proposed AB-DT model can effectively and accurately predict CBR values. In addition, it is also possible to quantify the influence of each parameter on CBR through SHAP value. Maximum dry density is the most essential attribute affecting soil's CBR value. The findings of this study can fill the knowledge gap regarding the prediction of CBR value using the machine learning approach. Furthermore, the result of this study is beneficial for practical application, particularly for engineers in pavement structure design and roadbed quality control.

Besides, this study contains some limitations, which should be investigated in future studies. First, this study only used two models: a single model (DT) and a hybrid model (AB-DT). Further studies using different modern and advanced models (including single and hybrid models) should

be done to verify the results obtained in this study. Second, the number of databases in this study is limited, so it is necessary to carry out this study with a larger database to confirm the findings of this study. Finally, to reduce the overfitting problem, cross-validation using K-fold should be conducted in future research.

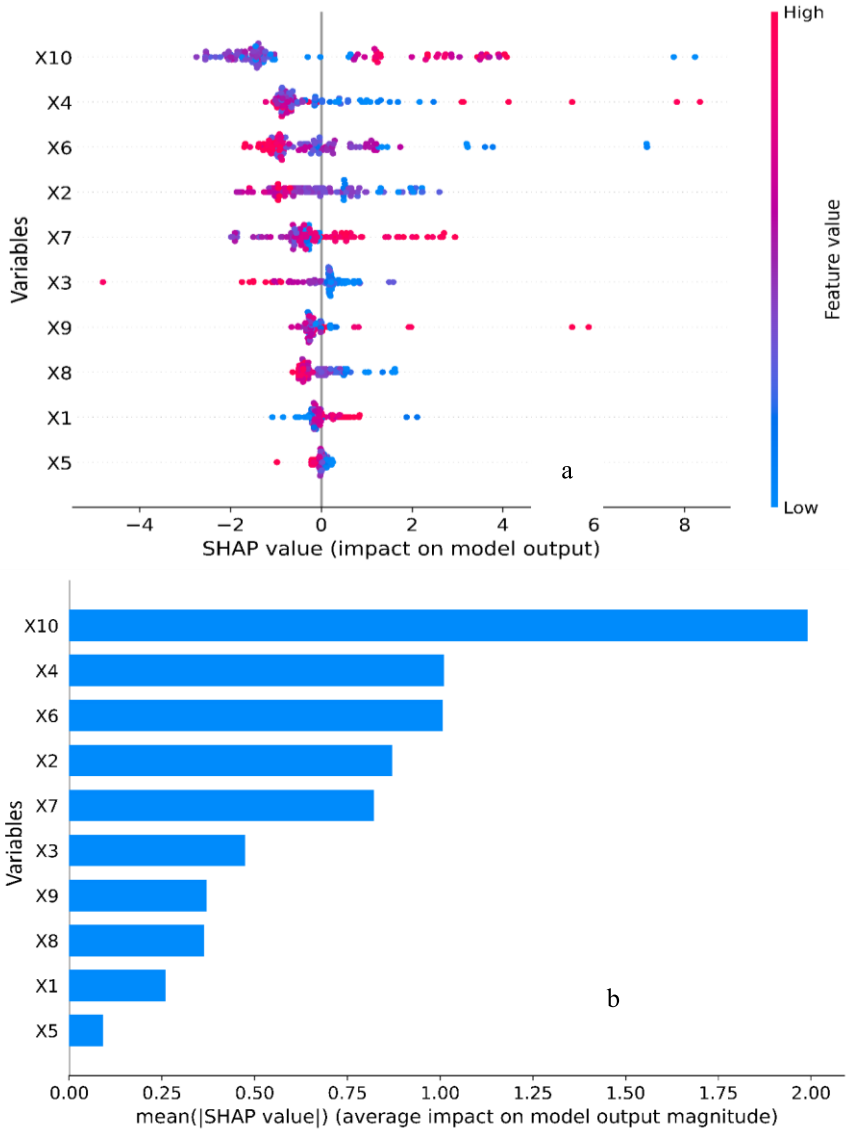

Figure 3. Importance of variables by SHAP: (a) Impact on model output, (b) Average impact on model output magnitude

References

Abdulnabi T.Y., Abdulrazzaq Z.G., 2020. An Estimated Correlation between California Bearing Ratio (CBR) with Some Soil Parameters of Gypseous Silty Sandy Soils. Tikrit Journal of Engineering Sciences, 27(1), 58–64.

Abnoosian K., Farnoosh R., Behzadi M.H., 2023. Prediction of Diabetes Disease Using an Ensemble of Machine Learning Multi-Classifier Models. BMC Bioinformatics, 24(1), 337. Doi: 10.1186/s12859-023-05465-z.

Agarwal K.B., Ghanekar K.D., 1970. Prediction of CBR from Plasticity Characteristics of Soil. Proceeding of 2nd South-East Asian Conference on Soil Engineering, Singapore. June, 11–15.

Alam S.K., Mondal A., Shiuly A., 2020. Prediction of CBR Value of Fine Grained Soils of Bengal Basin by Genetic Expression Programming, Artificial Neural Network and Krigging Method. Journal of the Geological Society of India, 95, 190–96.

Ampadu S.I.K., 2007. A Laboratory Investigation into the Effect of Water Content on the CBR of a Subgrade Soil. Experimental Unsaturated Soil Mechanics, Springer, 137–44.

Ariema F., Butler B.E., 1990. Guide to Earthwork Construction; State of the Art Report 8. Transportation Research Board National Research Council.

Atkins H.N., 1997. Highway Materials, Soils and Concretes, Prentice Hall. New Jersey, USA.

Bardhan A., Gokceoglu C., Burman A., Samui P., Asteris P.G., 2021. Efficient Computational Techniques for Predicting the California Bearing Ratio of Soil in Soaked Conditions. Engineering Geology, 291, 106239.

Barrett P., 2007. Structural Equation Modelling: Adjudging Model Fit. Personality and Individual Differences, 42(5), 815–24.

Bharath A., Manjunatha M., Reshma T.V., Preethi S., 2021. Influence and Correlation of Maximum Dry Density on Soaked & Unsoaked CBR of Soil. Materials Today: Proceedings, 47, 3998–4002.

Black W.P.M., 1962. A Method of Estimating the California Bearing Ratio of Cohesive Soils from Plasticity Data. Geotechnique, 12(4), 271–82.

Brown S.F., 1996. Soil Mechanics in Pavement Engineering. Géotechnique, 46(3), 383–426.

Bui Q.-A.T., Al-Ansari N., Le H.V., Prakash I., Pham B.T., 2022. Hybrid Model: Teaching Learning-Based Optimization of Artificial Neural Network (TLBO-ANN) for the Prediction of Soil Permeability Coefficient. Mathematical Problems in Engineering, 2022, 8938836.

Chengsheng T., Huacheng L., Bing X., 2017. AdaBoost Typical Algorithm and Its Application Research. MATEC Web of Conferences, EDP Sciences, 139, 00222.

Czajkowski M., Kretowski M., 2016. The Role of Decision Tree Representation in Regression Problems-An Evolutionary Perspective. Applied Soft Computing, 48, 458–75.

De Graft-Johnson J.W., Bhatia H.S., Gidigasu D.M., 1969. The Engineering Characteristics of the Laterite Gravels of Ghana. Soil Mech & Fdn Eng Conf Proc/Mexico/.

Domingo C., Watanabe O., 2000. MadaBoost: A Modification of AdaBoost, COLT, 180–89.

Futagami K., Fukazawa Y., Kapoor N., Kito T., 2021. Pairwise Acquisition Prediction with SHAP Value Interpretation, The Journal of Finance and Data Science, 7, 22–44.

González Farias I., Araujo W., Ruiz G., 2018. Prediction of California Bearing Ratio from Index Properties of Soils Using Parametric and Non-Parametric Models. Geotechnical and Geological Engineering, 36(6), 3485–98.

Granger C.W., Newbold P., 1974. Spurious Regressions in Econometrics. Journal of Econometrics, 2(2), 111–20.

Hadzima-Nyarko M., Trinh S.H., 2022. Prediction of Compressive Strength of Concrete at High Heating Conditions by Using Artificial Neural Network-Based Bayesian Regularization. Journal of Science and Transport Technology, 2(1), 9–21.

Hair J.F., Sarstedt M., Ringle C.M., Mena J.A., 2012. An Assessment of the Use of Partial Least Squares Structural Equation Modeling in Marketing Research. Journal of the Academy of Marketing Science, 40(3), 414–33. Doi: 10.1007/s11747-011-0261-6.

Hastie T., Rosset S., Zhu J., Zou H., 2009. Multi-Class Adaboost. Statistics and Its Interface, 2(3), 349–60.

Haupt F.J., Netterberg F., 2021. Prediction of California Bearing Ratio and Compaction Characteristics of Transvaal Soils from Indicator Properties. Journal of the South African Institution of Civil Engineering, 63(2), 47–56.

Hight D.W., Stevens M.G.H., 1982. An Analysis of the California Bearing Ratio Test in Saturated Clays. Geotechnique, 32(4), 315–22.

Ho L.S., Tran V.Q., 2022. Machine Learning Approach for Predicting and Evaluating California Bearing

Ratio of Stabilized Soil Containing Industrial Waste. Journal of Cleaner Production, 370, 133587.

Kamrul Alam S., Shiuly A., 2024. Soft Computing-Based Prediction of CBR Values. Indian Geotechnical Journal, 54(2), 474–88. Doi: 10.1007/s40098-023-00780-x.

Katte V.Y., Mfoyet S.M., Manefouet B., Wouatong A.S.L., Bezeng L.A., 2019. Correlation of California Bearing Ratio (CBR) Value with Soil Properties of Road Subgrade Soil. Geotechnical and Geological Engineering, 37, 217–34.

Khasawneh M.A., Al-Akhrass H.I., Rabab'ah S.R., Al-sugaier A.O., 2022. Prediction of California Bearing Ratio Using Soil Index Properties by Regression and Machine-Learning Techniques. International Journal of Pavement Research and Technology, 1–19.

Lakshmi S.M., Geetha S., Selvakumar M., 2021. Predicting Soaked CBR of SC Subgrade from Dry Density for Light and Heavy Compaction. Materials Today: Proceedings, 45, 1664–70.

Maulud D., Abdulazeez A.M., 2020. A Review on Linear Regression Comprehensive in Machine Learning. Journal of Applied Science and Technology Trends, 1(2), 140–47.

Mishra D., Tutumluer E., Butt A.A., 2010. Quantifying Effects of Particle Shape and Type and Amount of Fines on Unbound Aggregate Performance through Controlled Gradation. Transportation Research Record, 2167(1), 61–71.

Mokhtari K.E., Higdon B.P., Başar A., 2019. Interpreting Financial Time Series with SHAP Values. Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering, 166–72.

Myles A.J., Feudale R.N., Liu Y., Woody N.A., Brown S.D., 2004. An Introduction to Decision Tree Modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275–85.

Nguyen Q.H., Ly H.-B., Ho L.S., Al-Ansari N., Le H.V., Tran V.Q., Prakash I., Pham B.T., 2021. Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering, e4832864. Doi: 10.1155/2021/4832864.

Onyelowe K.C., Effiong J.S., Ebid A.M., 2023. Predicting Subgrade and Subbase California Bearing Ratio (CBR) Failure at Calabar-Itu Highway Using AI (GP, ANN, and EPR) Techniques for Effective Maintenance. Artificial Intelligence and Machine Learning in Smart City Planning. Elsevier, 159–70.

Patel R.S., Desai M.D., 2010. CBR Predicted by Index Properties for Alluvial Soils of South Gujarat. Proceedings of the Indian Geotechnical Conference. Mumbai, 79–82.

Pekel E., Estimation of Soil Moisture Using Decision Tree Regression. Theoretical and Applied Climatology, 139(3–4), 1111–19.

Peng C.-Y.J., Lee K.L., Ingersoll G.M., 2002. An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research, 96(1), 3–14. Doi: 10.1080/00220670209598786.

Pham B.T., Amiri M., Nguyen M.D., Ngo T.Q., Nguyen K.T., Tran H.T., Vu H., Anh B.T.Q., Van Le H., Prakash I., 2021. Estimation of Shear Strength Parameters of Soil Using Optimized Inference Intelligence System. Vietnam J. Earth Sci., 43(2), 189–198. https://doi.org/10.15625/2615-9783/15926

Pham B.T., Pradhan B., Bui D.T., Prakash I., Dholakia M.B., 2016. A Comparative Study of Different Machine Learning Methods for Landslide Susceptibility Assessment: A Case Study of Uttarakhand Area (India). Environmental Modelling & Software, 84, 240–50.

Quan V., Do H.Q., 2021. Prediction of California Bearing Ratio (CBR) of Stabilized Expansive Soils with Agricultural and Industrial Waste Using Light Gradient Boosting Machine. Journal of Science and Transport Technology, 1–9.

Quinlan J.R., 1996. Learning Decision Tree Classifiers. ACM Computing Surveys (CSUR), 28(1), 71–72.

Raja M.N.A., Shukla S.K., 2021. Predicting the Settlement of Geosynthetic-Reinforced Soil Foundations Using Evolutionary Artificial Intelligence Technique. Geotextiles and Geomembranes, 49(5), 1280–93.

Raja M.N.A., Shukla S.K., Khan M.U.A., 2022. An Intelligent Approach for Predicting the Strength of Geosynthetic-Reinforced Subgrade Soil. International Journal of Pavement Engineering, 23(10), 3505–21.

Rakhra M., Soniya P., Tanwar D., Singh P., Bordoloi D., Agarwal P., Takkar S., Jairath K., Verma N., 2021.

Crop Price Prediction Using Random Forest and Decision Tree Regression:-A Review. Materials Today: Proceedings.

Rätsch G., Onoda T., Müller K.-R., 2001. Soft Margins for AdaBoost. Machine Learning, 42, 287–320.

Rehman Z.U., Khalid U., Farooq K., Mujtaba H., 2017. Prediction of CBR Value from Index Properties of Different Soils, Technical Jour. University of Engineering and Technology (UET) Taxila. Pakistan, 22(2), 18–26.

Schaefer V.R., White D.J., Ceylan H., Stevens L.J., 2008. Design Guide for Improved Quality of Roadway Subgrades and Subbases. Iowa Highway Research Board (IHRB Project TR-525), 7, 8–72.

Schapire R.E., 2013a. Explaining Adaboost. Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik, 37–52.

Schapire R.E., 2013. Explaining AdaBoost, in Empirical Inference. B. Schölkopf Z. Luo and V. Vovk, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, accessed April 22, 2024b, from https://link.springer.com/10.1007/978-3-642-41136-6_5, 37–52.

Solomatine D.P., Shrestha D.L., 2004. AdaBoost. RT: A Boosting Algorithm for Regression Problems, 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541). IEEE, 2, 1163–68.

Thai P.B., Nguyen D.D., Thi Q.-A.B., Nguyen M.D., Vu T.T., Prakash I., 2022. Estimation of Load-Bearing Capacity of Bored Piles using Machine Learning Models. Vietnam J. Earth Sci., 44(4), 470–480. https://doi.org/10.15625/2615-9783/17177.

Thanh D.Q., Nguyen D.H., Prakash I., Jaafari A., Nguyen V.-T., Van Phong T., Pham B.T., 2020. GIS Based Frequency Ratio Method for Landslide Susceptibility Mapping at Da Lat City, Lam Dong Province, Vietnam. Vietnam J. Earth Sci., 42(1), 55–66. https://doi.org/10.15625/0866-7187/42/1/14758.

Trong D.K., Pham B.T., Jalal F.E., Iqbal M., Roussis P.C., Mamou A., Ferentinou M., Vu D.Q., Duc Dam N., Tran Q.A., 2021. On Random Subspace Optimization-Based Hybrid Computing Models Predicting the California Bearing Ratio of Soils. Materials, 14(21), 6516.

Tso G.K., Yau K.K., 2007. Predicting Electricity Energy Consumption: A Comparison of Regression Analysis. Decision Tree and Neural Networks. Energy, 32(9), 1761–68.

Venkatasubramanian C., Dhinakaran G., 2011. ANN Model for Predicting CBR from Index Properties of Soils. International Journal of Civil & Structural Engineering, 2(2), 614–20.

Wang D., Thunéll S., Lindberg U., Jiang L., Trygg J., Tysklind M., 2022. Towards Better Process Management in Wastewater Treatment Plants: Process Analytics Based on SHAP Values for Tree-Based Machine Learning Methods. Journal of Environmental Management, 301, 113941.

Willmott C.J., Matsuura K., 2005. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Climate Research, 30(1), 79–82.

Xu M., Watanachaturaporn P., Varshney P.K., Arora M.K., 2005. Decision Tree Regression for Soft Classification of Remote Sensing Data. Remote Sensing of Environment, 97(3), 322–36.