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

Prediction of California Bearing Ratio (CBR) of Stabilized Expansive Soils with Agricultural and Industrial Waste Using Light Gradient Boosting Machine

Van Quan Tran[1,*], Hai Quan Do[2]
[1]University of Transport Technology, Hanoi 100000, Vietnam
[2]Center for Structures and Materials, Viettel Aerospace Institute - Viettel Group, Lot D26, Cau Giay New Urban Area, Yen Hoa ward, Cau Giay District, Hanoi, Vietnam

Article info
Type of article:
Original research paper

DOI:
https://doi.org/10.58845/jstt.utt.2021.en.1.1.1-8

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

Received: 27/09/2021
Revised: 17/10/2021
Accepted: 20/10/2021

Abstract: Using agricultural and industrial waste such as bagasse ash, groundnut shell ash and coal ash in stabilizing expansive soils are used as a subgrade material to reduce harmful impaction of swelling/shrinkage of expansive soils, reduce construction costs. It is also a solution for environmental protection. California Bearing Ratio (CBR) is an important criterion to evaluate the application technique of stabilized expansive soil such as road construction, building construction, highway construction, airport construction, etc. Using the traditional method such as experimental methods or empirical approach, the estimation of CBR of stabilized expansive soils is costly, time consuming for the experiment or low accuracy for empirical method. In this investigation, open-source code of Machine Learning technique Light Gradient Boosting Machine algorithm is introduced to predict the CBR. In order to build model, data of 207 experimental samples was synthesized from the literature to create a database. The database consists of 6 input variables (ash content, ash type, liquid limit LL, plastic limit PL, optimum moisture content OMC and maximum dry density MDD) to obtain output variable CBR. The results show that the LightGBM model can successfully predict the CBR of stabilized expansive soils with high accuracy. The ash content is the most important input factor for CBR prediction using LightGBM model. In order of importanc input factor affecting CBR prediction are ash content, MDD, ash type, OMC, LL, PL.

Keywords: Stabilized expansive soil, Machine learning, Light Gradient Boosting, California Bearing Ratio (CBR), Agricultural/Industrial waste.

1. Introduction

Swelling/Shrinkage of expansive soils causes mechanical deterioration of the subgrade where the variation of water content takes place. Therefore, the strong swelling/shrinkage occurs, that will induce the instability of subgrade structures which affects the safety of construction. Stabilizing expansive soils is the appropriate technique in limiting the negative effects of swelling/shrinkage of expansive soils. Cementitious materials are often selected for the stabilized soil process to improve the mechanical

properties of the expansive soils. In addition, using cementitious materials derived from agricultural and industrial waste such as bagasse ash, groundnut shell ash and coal ash contributes both in environmental protection and sustainable development.

To evaluate the mechanical properties such as stiffness modulus and shear strength of expansive soils after stabilization process of the subgrade of construction project such as road foundation, airport foundation, etc., California Bearing Ratio (CBR) is often used. CBR is an indirect measurement where the CBR value is the ratio between the strength value of the subgrade material and the strength of the standard crushed rock. In fact, the different soil samples need to be collected and compacted to determine the Optimal Moisture Content (OMC) and Maximal Dry Density (MDD) in experimental measurement of CBR. In next step, these samples are then further soaked in water for four days before the CBR determination are carried out. The process of determining the CBR index takes about a week. Therefore, the number of samples to be determined is high for the large project area that will require a long time as well as high cost. The extended time leads to an increase in the project cost. To overcome this situation, the CBR index can be estimated from easily identifiable parameters of soil such as Atterberg limits, effective compaction process (OMC, MDD). A number of studies have been conducted to provide empirical equations to determine CBR. Black [1] introduced an empirical relation between CBR and plasticity index (PI). CBR can be empirically estimated from liquid limit (LL) and PI [2]. More complex, different empirical correlation equations between CBR and LL, plastic limit PL, PI and effective compaction were also established [3], [4]. However, these equations were given with a small number of experimental samples, so the general and accuracy of these equations can be increased.

Machine learning (ML) and Artificial Intelligence (AI) techniques have been strongly developed in recent years with advantages such as high accuracy, fast computation time, saving design costs. Especially, the ML model has high generality and accuracy when the model uses large samples in training the model. Therefore, ML models have been applied to solve many problems in civil engineering such as determination of pile bearing capacity [5], [6], unconfined compressive strength of stabilized soil [7], compressive strength of concrete [8], [9], etc. Therefore, the ML models have been developed in determining the CBR of stabilized expansive soils. Taskiran [10] developed the Genetic expression programming (GEP) algorithm to predict the CBR of stabilized expansive soil. The CBR value can be also predicted by Artificial Neural Network (ANN) models [11]. In the development of actual machine learning technique, the accuracy of ML models can be improved. Light Gradient Boosting Machine is a new machine learning technique developed by Microsoft corporation [12] which has been proposed in the present study to determine the CBR of stabilized expansive soils. Model performance of the ML model are evaluated by different criteria such as correlation coefficient R, root mean square error RMSE and mean absolute error MAE.

2. Machine learning approach

2.1. Light Gradient Boosting Machine

Light Gradient Boosting Machine (LightGBM) is an open-source library providing an effective implementation of gradient boosting framework based on tree-based learning algorithms [13]. This algorithm has been designed by Microsoft Corporation since 2016. The algorithm has some advantages such as faster speed of training and high accuracy, reliability with low memory usage to run. The large-scale data in regression problem can be efficiency handled by this algorithm. LightGBM is a relatively new algorithm and easily performed using Python library and list of parameters given in the LightGBM documentation [12].

2.2. Performance evaluation of machine

learning model

In this process, three performance criteria were used namely correlation coefficients R, root mean square error RMSE and mean absolute error MAE to assess the accuracy of LighGBM model [7]:

$$R = \frac{\sum_{j=1}^{N}\left(p_{0,j} - \bar{p}_0\right)\left(p_{t,j} - \bar{p}_t\right)}{\sqrt{\sum_{j=1}^{N}\left(p_{0,j} - \bar{p}_0\right)^2 \sum_{j=1}^{N}\left(p_{t,j} - \bar{p}_t\right)^2}} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(p_{0,j} - p_{t,j}\right)^2} \quad (2)$$

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left(p_{0,j} - p_{t,j}\right) \quad (3)$$

Where: N is the number of data sets, p_0 and $\bar{p}_0$ is the experimental value and average experimental value, p_t and $\bar{p}_0$ is predicted value and average predicted value using LightGBM. R measures the predicted and experimental value association, if the R is closer to 1, the LightGBM model is more accurate. RMSE calculates the square root average difference between the expected values and the experimental values and the difference between the experimental and the predicted values is determined MAE criteria. RMSE and MAE value are closer to 0, the accuracy of the LightGBM is higher.

3. Construction and analysis of database

In this study, the database is built based on the data collection from Rajakumar and Reddy [11], in which 207 experimental samples of stabilized expansive soils are designed with different types of ash (coal ash-type 1, bagasse ash-type 2 and groundnut shell ash-type 3), ash content, Atterberg limits (LL and PL), effective compaction (OMC and MDD). Therefore, LightGBM algorithm uses 6 input variables consisting of: (1) ash type (labelled 1, 2 and 3); (2) ash content (%); (3) LL (%); (4) PL (%), (5) OMC (%) and (6) MDD (g/cm3). CBR (%) is considered only output variable. The whole dataset was randomly divided into two sub-datasets including 70% of whole samples for training LightGBM model corresponding to 145 samples. The remaining samples consisting of 30% of the whole data corresponds to the 62 samples used for testing model. The statistical analysis of database is presented in Table 1.

As mentioned in the above section, three ash types of agricultural waste consisting of coal ash labelled 1, bagasse ash labelled 2 and groundnut shell ash labelled 3 are used for stabilizing expansive soil. With the data distribution shown in Fig 1, the number of samples using coal ash is slightly used more than that using the other ash. The used ash content varies from 0% to 60% by (mean value of 14.03 % and median value of 8.00%). The LL and PL range from 17% to 64% (mean value 39.383% and median value 39.000%) and 12.8 to 30.1% (mean value 20.646% and median value 20.000%), respectively. The OMC and MDD vary from 8.91% to 32.5% (mean value 17.775% and median value 17.020%) and 1.37 g/cm^3 to 1.88 g/cm^3 (mean value 1.615 g/cm^3 and median value 1.620 g/cm^3). Moreover, the data distribution shown in Fig 1 indicates that each input variable seems to weakly correlate with output CBR. Especially, Ash type, PL and OMC seem to not correlate with CBR.

Table 1. Statistical analysis of database

	Count	Mean	Std	Min	Q25%	Q50%	Q75%	Max	Skw
Ash type	207	1.986	0.815	1.000	1.000	2.000	3.000	3.000	0.027
Ash content (%)	207	14.029	16.650	0.000	4.000	8.000	12.000	60.000	1.593
LL (%)	207	39.383	10.533	17.000	32.200	39.000	46.750	64.000	0.215
PL (%)	207	20.646	4.565	12.800	16.900	20.000	24.050	30.100	0.341
OMC (%)	207	17.775	5.107	8.910	13.915	17.020	21.200	32.500	0.508
MDD (g/cm^3)	207	1.615	0.103	1.370	1.535	1.620	1.680	1.880	0.308
CBR (%)	207	3.775	0.990	1.860	3.055	3.890	4.510	6.460	-0.149

Skw=Skewness; Std=Standard deviation

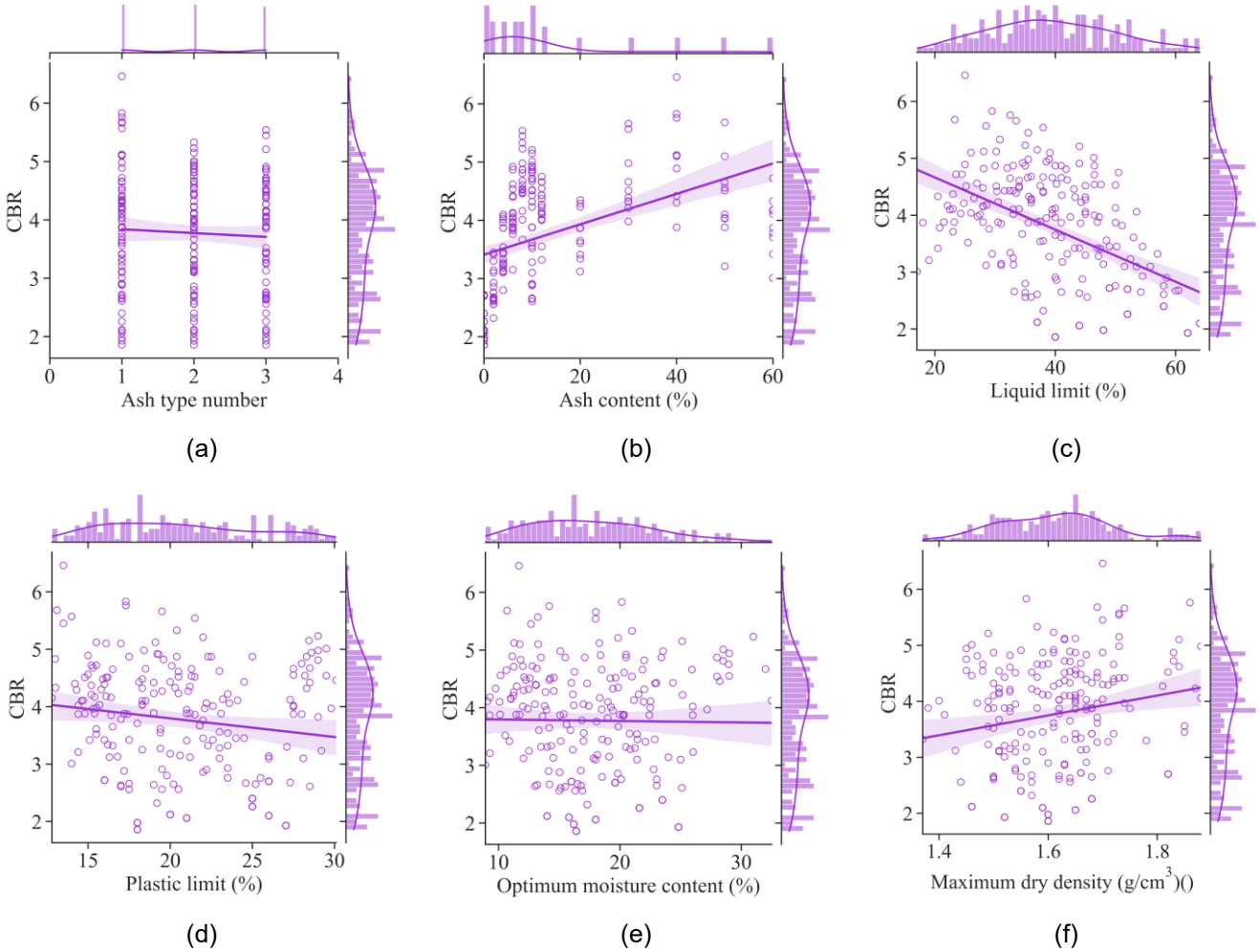

Fig 1. Distribution and correlation line of each input variable and CBR output

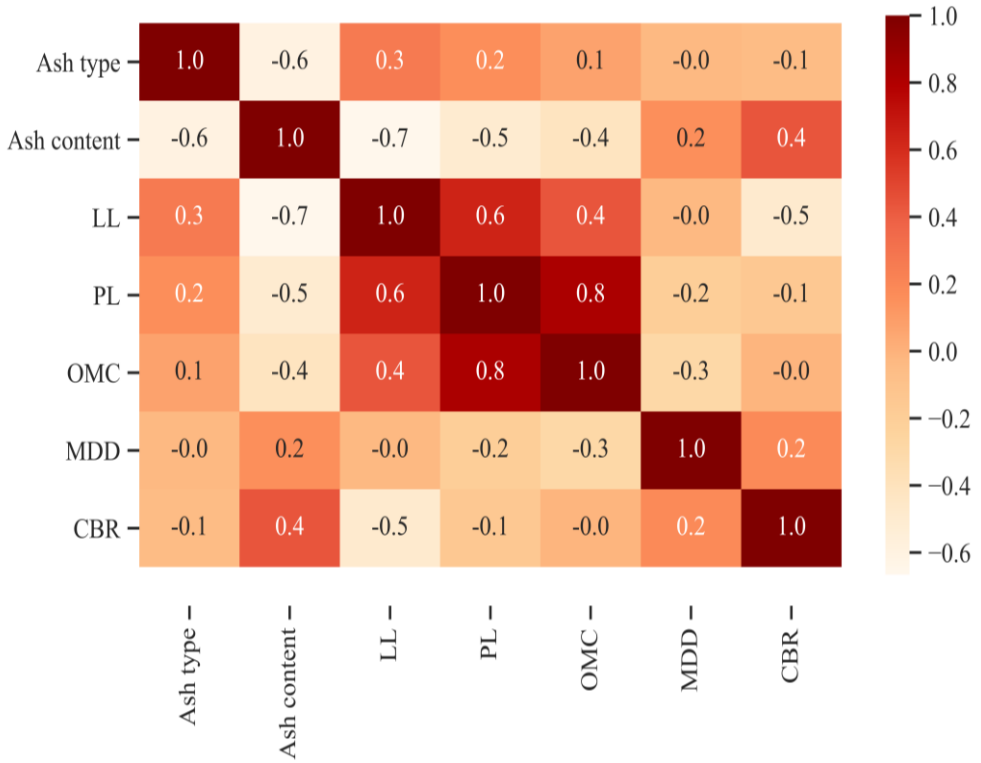

Fig 2. Correlation matrix of input and output variables

In fact, the correlation matrix of input and output variables (Fig 2) shows the ash type, PL and OMC have weak correlation coefficients to e equal to 0.1 and 0, respectively. The correlation between inputs and output, the highest correlation coefficient belongs to LL and CBR, the correlation coefficient is equal to -0.5, it means that higher LL, the CBR decrease. Correlation between 6 input variables, the highest correlation is OMC and PL with the coefficient to be equal to 0.8. However, all correlation coefficients are not high enough to reduce the proposed number of inputs. Moreover, the six input variables can be useful for the feature importance in last section of the paper.

4. Results and discussion

In this study, LightGBM algorithm is performed using the Python programming language. Using the default hyperparameter implemented in Python, the typical prediction results of the LightGBM are assessed in Fig 3 for graphical demonstration. The experimental and predicted CBR of stabilized expansive soils are compared in Fig 3 consisting of the training dataset (Fig 3a) and the testing dataset (Fig 3b). The results show that the predicted CBR of the both training and testing part is in excellent coherent with experimental values. Excellent agreement between experimental and predicted CBR is also indicated by histograms of error prediction for the training dataset (Fig 4a) and testing datasets (Fig 4b). It can be observed that the prediction errors of the training and testing datasets are relatively small. Error values ranges from -0.5 to 0.5%. The error cumulative lines also indicate that about 145 prediction error values vary from -0.5 to 0.5% and 5 prediction values are out of this range for the training part. In testing part, only 6 prediction error values are out of range -0.5 to 0.5%. These error results confirm that the predictive performance of the LightGBM model is excellent algorithm to predict the values of CBR of stabilized expansive soils with agricultural waste including coal ash, bagasse ash, and groundnut shell ash.

The regression graphs of both the training part and testing part are presented in Fig 5. It is worth noting that the predictive capability of LightGBM model is high. The performance values are R=0.9473, RMSE=0.3303, MAE=0.2530 and R=0.9385, RMSE=0.3037, MAE=0.2506 for the training and testing parts, respectively. Using MATLAB software, the performance values of ANN model in Rajakumar et Babu [11] are expressed by the correlation coefficient R and the mean square error MSE with the best performance values R=0.9432 and MSE=0.49 (RMSE=0.7000) for the whole dataset. These performances values are lower than that of this investigation consisting of R=0.9452 and RMSE=0.3225 for all dataset (Fig 5c). Moreover, LightGBM is open source of Python programming langue so that this algorithm can be easily approached both by the engineers and researchers.

Therefore, using LightGBM model to predict the CBR of stabilized expansive soils is feasible with high accuracy and user friendly. It could be suited for developing a numerical tool for determining the CBR of stabilized expansive soils for geotechnical engineer.

Fig 6 shows the feature importance analysis of CBR prediction of stabilized expansive soil. The most importance input is the ash content used for stabilizing expansive soil. The first feature ash content is more important 20 times than the second feature MDD (feature importance value 1.5 versus 0.25). The lowest important input is the plastic limit which has the feature importance value to be quite equal to 0. Therefore, this feature can be not taken account for training LightGBM model in predicting CBR of stabilized expansive soils in the future. The liquid limit influence on CBR lower than OMC. Ash type has greater importance than OMC. Overall, the mix design containing ash content and ash type have strong importance on the CBR prediction, the effective compaction (OMC and MDD) influence more importantly than Atterberg limits in predicting the CBR of stabilized expansive soils.

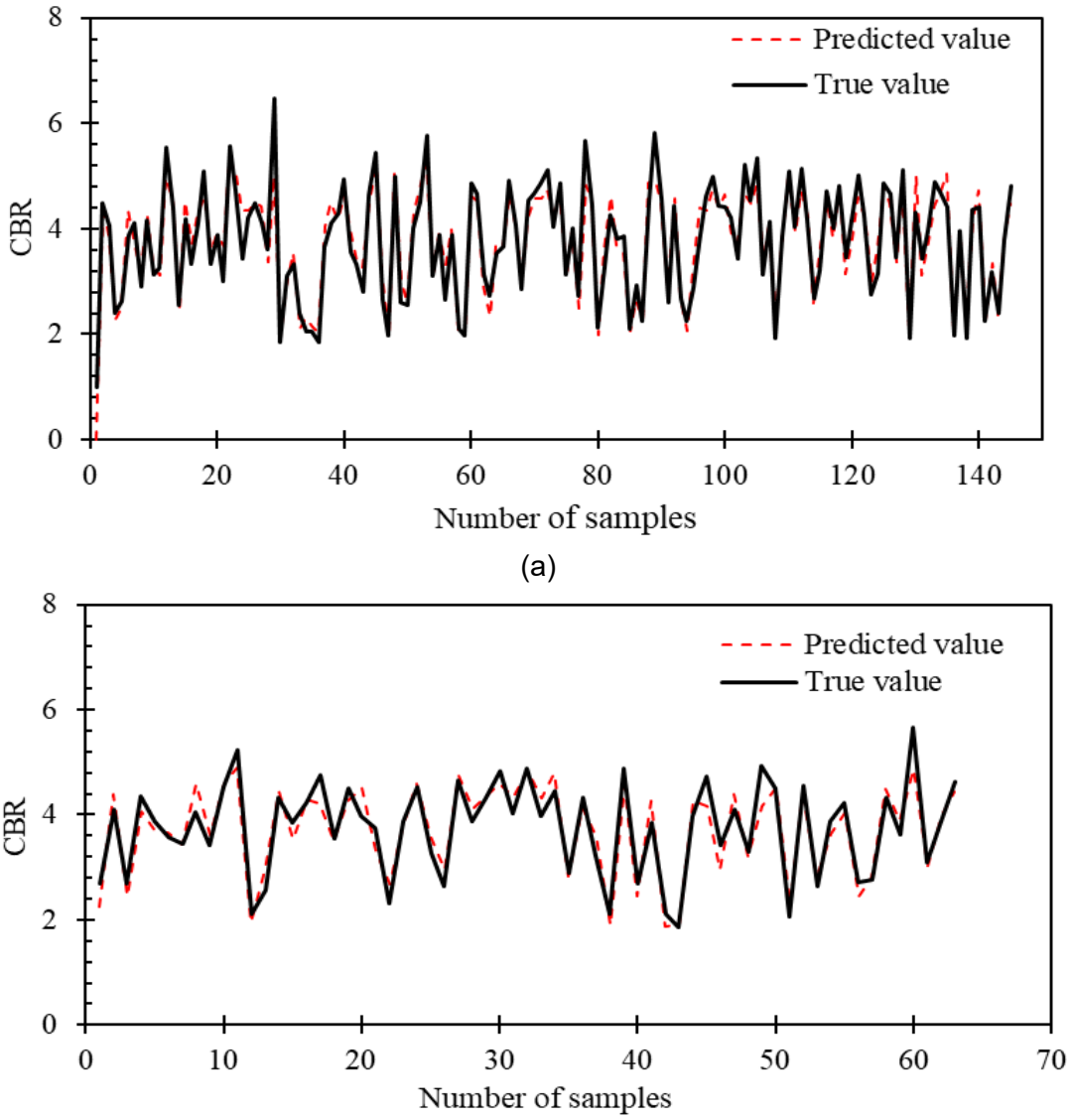

Fig 3. Predicted California bearing ratio (CBR) of stabilized expansive soils LighGBM model (a) Training; (b) Testing

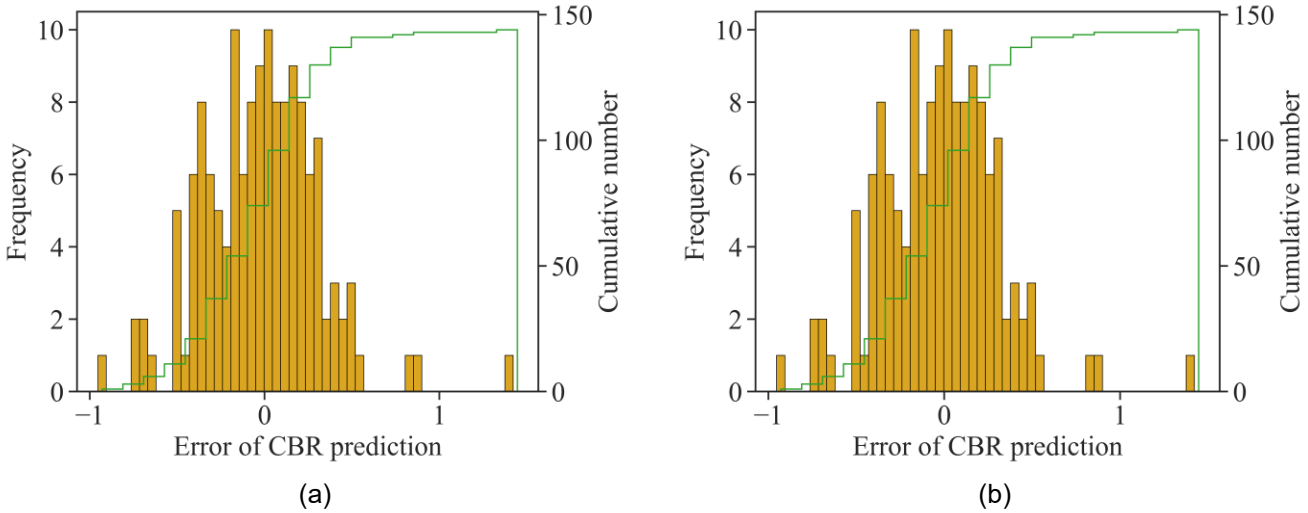

Fig 4. Error value between predicted and experimental CBR of stabilized expansive soils for (a) training part, (b) testing part

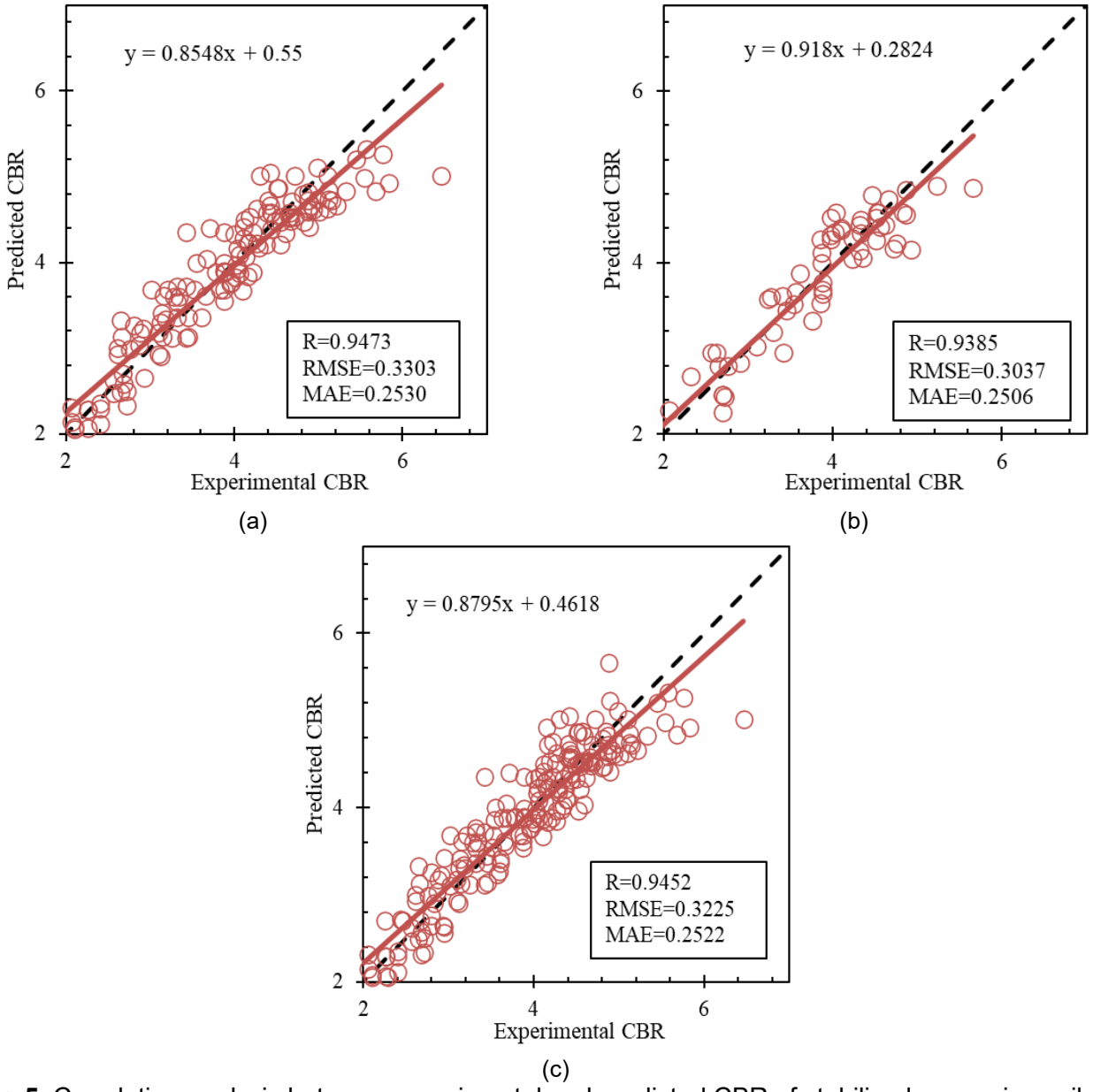

Fig 5. Correlation analysis between experimental and predicted CBR of stabilized expansive soils (a) training dataset, (b) testing dataset and (c) all dataset

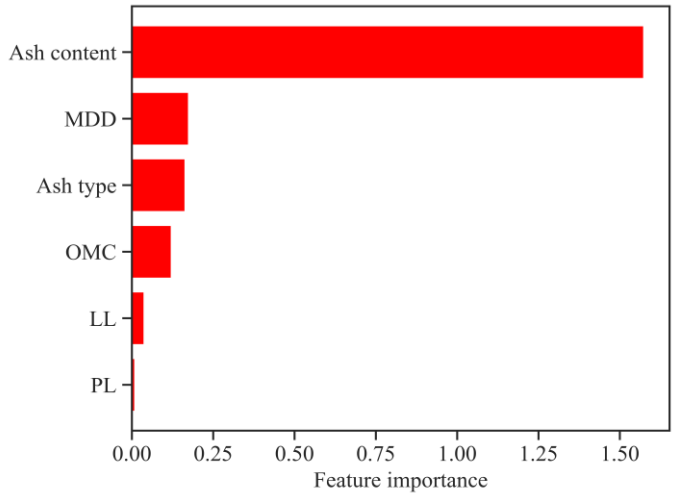

Fig 6. Importance of each input effect on CBR of stabilized expansive soils

5. Conclusion

In this study, the ability of ML techniques to predict the CBR of stabilized expansive soils is presented. The database used for building model was obtained from the literature. To reduce time consumption and cost for performing experiments, a Light Gradient Boosting Machine (LightGBM) model is developed. Using the default hyperparametersthe CBR of stabilized expansive soils can be successfully predicted by the LightGBM model with high accuracy as the performance values are: R (0.9452), RMSE (0.3225) and MAE (0.2522). Therefore, this algorithm is a good technique that can be applied for the estimation of CBR of expansive stabilized soil. The feature affecting the CBR are in order as per importance: ash content, MDD, ash type, OMC, LL, and PL. The input variable PL can be not considered in training LightGBM model for predicting the CBR of stabilized expansive soils in the future. The CBR prediction accuracy of LightGBM model can be improved by using meta heuristic algorithms for determining the hyperparameters of model.

References

[1] W.P.M. Black. (1962). A Method of Estimating the California Bearing Ratio of Cohesive Soils from Plasticity Data. *Géotechnique*, 12(4), 271-282.

[2] J.W. de Graft-Johnson, H.S. Bhatia, and D.M. Gidigasu. The engineering characteristics of the laterite gravels of Ghana. Soil Mech & Fdn Eng Conf Proc, pp 117-128.

[3] R.S. Patel, M.D. Desai. (2010). CBR Predicted by Index Properties for Alluvial Soils of South Gujarat. *Indian Geotechnical Conference,* 79-82.

[4] K. Pal and K. Pal. (2019). Correlation between CBR values and plasticity index of soil for Kolkata region. *International Research Journal of Engineering and Technology (IRJET)*, 6(11), 310-315.

[5] T.A. Pham, H.-B. Ly, V.Q. Tran, L.V. Giap, H.-L.T. Vu, and H.-A.T. Duong. (2020). Prediction of Pile Axial Bearing Capacity Using Artificial Neural Network and Random Forest. *Applied Sciences*, 10(5), 1871.

[6] T.A. Pham, V.Q. Tran, and H.-L.T. Vu. (2021). Evolution of Deep Neural Network Architecture Using Particle Swarm Optimization to Improve the Performance in Determining the Friction Angle of Soil. *Mathematical Problems in Engineering*, 2021, e5570945, 1-17.

[7] V.Q. Tran. (2021). Compressive Strength Prediction of Stabilized Dredged Sediments Using Artificial Neural Network. *Advances in Civil Engineering*, 2021, e6656084, 1-8.

[8] T.-A. Nguyen, H.-B. Ly, H.-V.T. Mai, and V.Q. Tran. (2020). Prediction of Later-Age Concrete Compressive Strength Using Feedforward Neural Network. *Advances in Materials Science and Engineering*, 2020, e9682740, 1-8.

[9] H.-B. Ly, T.-A. Nguyen, H.-V.T. Mai, and V.Q. Tran. (2021). Development of deep neural network model to predict the compressive strength of rubber concrete. *Construction and Building Materials*, 301, 124081.

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

[11] C. Rajakumar and G. Reddy babu. (2021). Experimental study and neural network modelling of expansive sub grade stabilized with industrial waste by-products and geogrid. *Materials Today: Proceedings*, 46(Part1), 131-137.

[12] Welcome to LightGBM's documentation! — LightGBM 3.2.1.99 documentation. https://lightgbm.readthedocs.io/en/latest/ (accessed Sep. 18, 2021).

[13] J.H. Friedman. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378.