MENDEL
Soft Computing Journal

MODELING CBR VALUE USING RF AND M5P TECHNIQUES

Manju Suthar[1], Praveen Aggarwal[2,✉]

[1]Department of Civil Engineering, Maharaja Agrasen University, Baddi, India
[2]Civil Engineering Department, National Institute of Technology, Kurukshetra, India
manju.manni4@gmail.com, Praveen_agg@hotmail.com[✉]

Abstract

Two modeling techniques namely (i) Random forest (RF) and (ii) M5P model tree are used to model, soaked California bearing ratio (CBR) value of thermal power plant generated stabilized pond ash. Pond ash was stabilized with the help of commercially available lime and industrial waste lime sludge. CBR data generated from exhaustive experimental programme was used in the study. Variations in doses of stabilizer i.e. lime (L) and lime sludge (LS), curing duration (CP) and proctor test results density (MDD) & moisture (OMC) are considered as input variables. Experimentally observed CBR value was used as output variable. Performance of models was measured using standard statistical parameters. Although, both the model's performance in predicting CBR value is satisfactory however from the statistical parameters it is evident that RF method perform better in comparison to M5P model. Sensitivity analyses identify CP as the most influencing factor that affects CBR value of the stabilized pond ash.

1 Introduction

Coal ash is an industrial by-product produced during burning of coal in thermal power plants; its disposal requires huge area. Coal ash produced from thermal power plants are of three types i.e. bottom ash, fly ash and pond ash [1]. Bottom ash is generated in boiler and fly ash in electrostatic precipitators. Pond ash is a mixture of fly ash and bottom ash, pumped into ash ponds in slurry form [2].

As per ASTM C618-08a) [3], two types of ashes are generated upon burning of coal, namely Class C and Class F ash. Calcium content is high in Class C ashes as compared to Class F ashes; as a result, they react with water even without additional binder [4]. It is also observed that ash produced from thermal power plants worldwide is primarily of Class F category [5]. Because of low calcium content, in natural form, Class F ash has poor strength characteristics as a civil engineering construction material. Being available in abundance, many studies have been reported by various researchers to improve its engineering properties through additives like cement, gypsum and lime [6-9].

Industrial waste lime sludge (LS) is generated primarily from sugar mills, paper and fertilizer industry. Annual production of LS in India alone is more than 4.0 million tons [10]. Presently it is dumped in low laying area and its proper disposal is an environmental concern [11-13].

In pavement engineering, California bearing ratio (CBR) value of subgrade is a key parameter that decides the thickness of various pavement layers. Although CBR test can be performed in laboratory as well as in field, however laboratory CBR tests are preferred over field because of better quality control [14]. Performing CBR test is a labour intensive and time-consuming exercise. Also, poor quality control affects the test results badly [15]. Therefore, predictive models can be a valuable alternative tool for quantifying the CBR values of stabilized pond ash.

Modeling techniques, such as artificial neural networks (ANN), are used in predicting the CBR value [16-18]. In spite of an encouraging performance by the neural network approach in predicting CBR, it is noticed that number of inputs are required (such as number of hidden layers and nodes in each hidden layer) and prediction performance is affected by these inputs. It is also observed that over training affects the predictions. In back propagation technique of neural network, local minima also gives misleading results [19].

Considering the application of tree-based regression in many civil engineering application [19], in the present study two such techniques (i) Random forest regression and (ii) M5 model tree-based modeling approaches are used in forecasting the California bearing ratio value of ash. Pond ash used in the present study was obtained from ash pond of a thermal power plant situated in Haryana, India. The pond ash was stabilized with lime

along with lime sludge and cured for different duration. Data set obtained through laboratory investigation was used for modelling purpose.

2 Modeling Techniques

In the present study two machine learning modeling techniques namely (i) Random Forest Regression and (ii) M5 Model Tree are used because of minimal input parameters, ease of use and better results.

2.1 Random Forest Regression

Random Forests (RF) is effective and easy machine learning algorithms for classification and regression analysis. Decision Trees and Decision Tree Learning together comprise a simple and fast way of learning a function that maps input data to output data. One of the biggest advantages of using Decision Trees and Random Forests is the ease in which we can see what features or variables contribute to the classification or regression and their relative importance based on their location depth-wise in the tree. The random forest model has excellent capability of treating numerical values in tabular form or data with definite features. Random forests are able to develop non-linear relationship between the input and target variables. In the regression process, class labels are converted into numerical values [20].

In the process of regression analysis, randomly drawn 2/3rd of original dataset was used for training purpose and remaining 1/3rd left out data for testing purpose [21]. For induction of tree, one of the pruning method has to be selected out of various approaches. Most common approaches include 'Gini Index' [23] and 'Information Gain Ratio criterion' [22]. Results of tree based algorithms are affected by selection of pruning methods [24]. As per Breiman [20] with the increase in number of trees the generalization error reduces. In RF technique over fitting is not a problem [25]. There are only two user defined variables m and k [20].

2.2 M5 Model Tree

M5P is a reconstruction of Quinlan's M5 algorithm, in this technique trees are induced for regression models. It's a combination of conventional decision tree with the option of linear regression functions at the nodes. Decision tree is produced using divergence metric called Standard Deviation Reduction (SDR). Using linear regression functions tree modal was developed. The process involves pruning, evacuation and substitution of trees. Hence a final tree model is developed.

3 Experimental Investigations

In the present study, CBR tests under soaked condition were performed on pond ash stabilized with lime in varying percentage from 2 to 8% at an increment of 2%. Pond ash was also stabilized using lime plus lime sludge (LS), LS being in 5, 10 and 15%. In order to work out maximum dry density (MDD) and optimum moisture content (OMC) Proctor tests were performed on 17 mixes, as per Indian standards [26]. To find out soaked CBR value, tests were performed on lime and lime stabilized samples, as per IS: 2720 (Part-16)-2002) [27]. In order to study the effect of curing, samples were cured for different duration (7, 28 and 45 days) and soaked for 96 hours before performing the CBR test.

3.1 Data Set

A total of 51, experimentally observed CBR test results, with varying percentage of additives and curing period were used as data set for the modeling purpose. In the present study, random forest regression and M5model were used for estimation of CBR value, using WEKA 3.9 software. In both models, randomly selected 2/3rd of the data were used for the training, whereas 1/3rd was used for testing the models. Various inputs parameters/dependent variables include maximum dry density (MDD), optimum moisture content (OMC), quantity of lime (L) & lime sludge (LS) and curing period (CP). However, CBR value in soaked condition of stabilized pond ash was considered as output parameter.

Statistical parameters which include coefficient of correlation (CC), root relative square error (RRSE), mean absolute error (MAE), root mean square error (RMSE) and relative absolute error (RAE) were identified to compare the results of random forest and M5 model. In both the models, user's defined parameters affect the performance of models. Hence selection of optimal values of these parameters is very important. Statistical characteristics of various experimentally observed data are shown in Table 1. The optimal values of required user defined parameter in both the modeling approaches are given in Table 2.

Table 1: Statistical characteristics of experimentally observed data set used for training and testing purpose

Parameters used		Dataset − Training			
		Minimum	Maximum	Mean	Std. dev.
Input Parameters	MDD	1.175	1.259	1.234	0.024
	OMC	22.8	26.8	24.028	1.14
	L	0	8	4.556	2.645
	LS	0	15	7.083	5.779
	CP	7	45	26.667	15.763
Output parameter	CBR	2.2	63.8	35.806	16.384

Parameters used		Dataset - Testing			
		Minimum	Maximum	Mean	Std. dev.
Input Parameters	MDD	1.194	1.253	1.231	0.017
	OMC	23.1	25.3	24.313	0.621
	L	2	8	5.067	2.12
	LS	0	15	7	5.916
	CP	7	45	26.667	16.088
Output parameter	CBR	14.6	56.3	35.52	14.038

Table 2: User defined parameters and there optimal values in the models

Modeling Technique	Parameters
Random forest	K-3, M-1, I-100
M5P model tree	M-4

4 Results and Discussions

Using RF and M5P models, CBR value of pond ash stabilized with two additives i.e. lime and lime sludge are predicted and compared with the experimentally obtained results. Fitness level of predicted CBR values was obtained through statistical parameters which include CC, RMSE, MAE, RAE and RRSE. Statistical parameters of soaked CBR predicted by both modeling technique are shown in Table 3. Figure 1 and 2 present the scattered graph between the observed and predicted values of soaked CBR obtained from RF and M5 model with training and testing dataset respectively. From graphical results, it can be concluded that the predicted CBR value by both the models have good agreement with the experimentally observed CBR value, but M5P model predict negative CBR values against low experimental CBR values. On the basis of statistical parameters (Table 3), random forest model shows slightly better performance because of highest CC and lowest errors values (CC- 0.9786, MAE-3.2646, RMSE-3.6635, RAE-28.3133 and RRSE-27.0129) in comparison to M5 model in predicting the soaked CBR values. Table 4 presents the results of ANOVA Single Factor Test. From the results, it is observed that F-critical is 3.082852 and F-value is 0.002307, which is less than F-critical. P-value (0.997696) is > 0.05, highlighting negligible difference between experimental and predicted values by both the models.

Table 3: Details of statistical parameters over training and testing data through predictive models

Modeling Technique	Training data set				
	CC	MAE	RMSE	RAE (%)	RRSE (%)
Random Forest	0.9969	1.3389	1.6724	10.0582	10.3524
M5P model	0.9640	3.3713	4.2972	25.3273	26.5996

Modeling Technique	Testing data set				
	CC	MAE	RMSE	RAE (%)	RRSE (%)
Random Forest	0.9701	3.2605	4.0796	28.1374	30.0742
M5P model	0.9201	4.3625	5.3639	37.6472	39.5421

Table 4: ANOVA Test Results

Selected Model	P-value	F	F critical	Variation in observed and predicted values
Both models	0.997696	0.002307	3.082852	Negligible (P-value>0.05, F<F-critical,)

4.1 Sensitivity Analysis

Sensitivity analysis is a useful tool to identify significant factors affecting the output. In the present study, different set of training data was created. In each set, at a time, one of the input parameter was eliminated and its influence was determined in terms of statistical parameters with experimental data. Table 5 present the summary of sensitivity analysis. Results from Table 5 suggests curing duration as the most important input variable in predicting the soaked CBR value of stabilized pond ash with used additives, using random forest regression models.

Table 5: Results from sensitivity analysis

Input parameters (Considered)	Input parameter (Eliminated)	RF	
		CC	RMSE
MDD, OMC, L, LS, CP		0.9969	1.6724
MDD, OMC,L, LS	CP	0.8566	8.3551
MDD, OMC,L, CP	LS	0.9942	1.9589
MDD, OMC, LS, CP	L	0.9924	1.9589
MDD, L, LS, CP	OMC	0.9952	1.8160
OMC, L, LS, CP	MDD	0.9969	1.6724

5 Conclusions

In the present study experimentally observed soaked CBR value of stabilized pond ash is modeled using two modeling approaches (i) Random Forest (ii) M5P. Stabilization was carried out using lime and lime sludge. Fitness of models is evaluated through statistical parameters. Random Forest works better as compared to M5P because of higher CC and lower errors. ANOVA single factor results dictate insignificant difference in predicted and experimental observed value using both the modeling approaches. Sensitivity analysis considers curing duration as the most important input variable that affects the soaked CBR value of lime/lime sludge stabilized, thermal power plant generated pond ash used in the study.
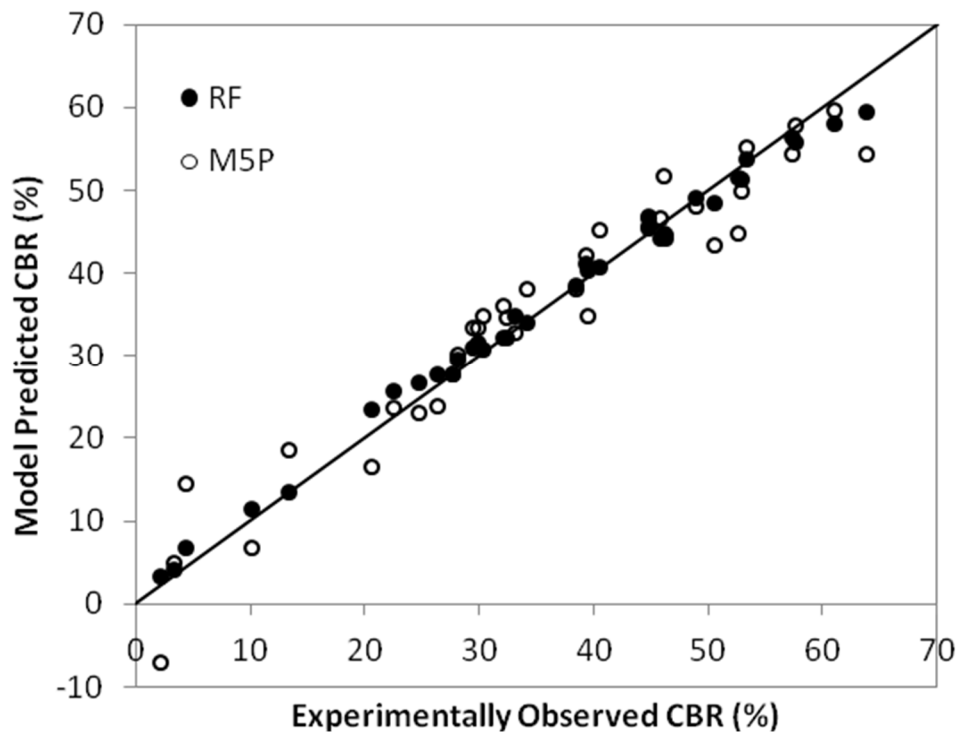
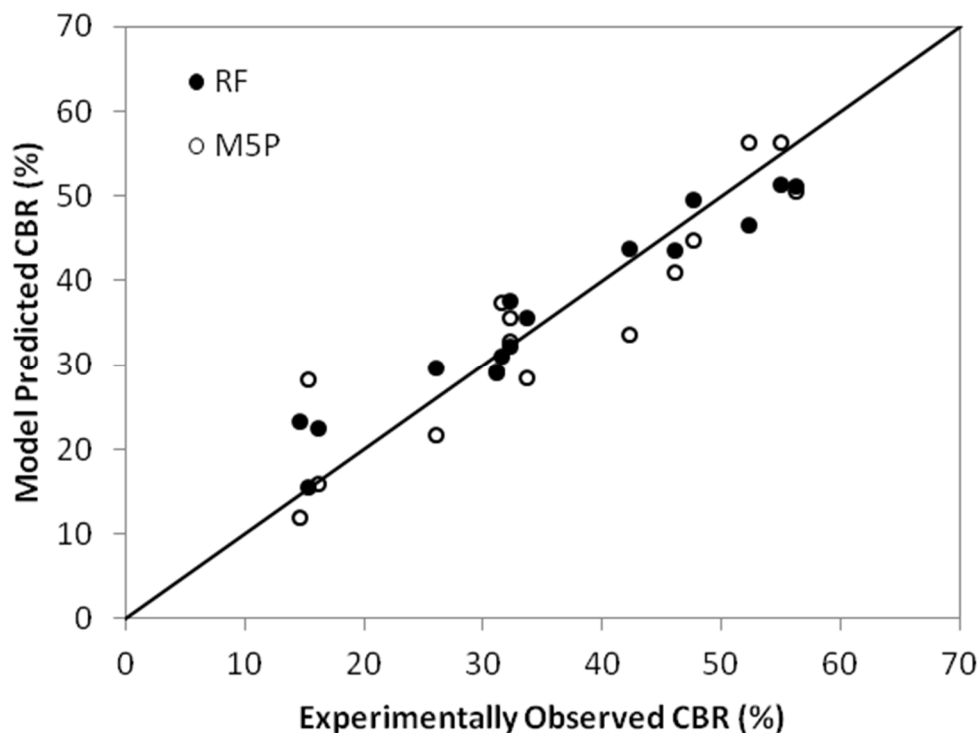Figure 1: Observed vs modeled CBR value using RF and M5P model with training data set

Figure 2: Observed vs modeled CBR value using RF and M5P model with testing data set

References

[1] Bera, A. K., Ghosh A., and Ghosh A. 2007. Compaction characteristics of pond ash. *Journal of Materials in Civil Engineering* 19, 4, pp. 349–357.

[2] Suthar M. and Aggarwal P. 2017. Analysis of heavy metals in pond ash samples from Haryana. In *Proceedings of 29th Research World International Conference*. Las Vegas, USA, 16th-17th March 2017.

[3] ASTM C618-08a. 2008. Specification for fly ash and raw or calcined natural pozzolana for use as material admixture in portland cement concrete. https://www.astm.org

[4] Parsa, J., Munson-McGee, S. H., and Steiner, R. 1996. Stabilization/solidification of hazardous wastes using fly ash. *Journal of Environmental Engineering* 122, 10, pp. 935–940.

[5] Ghosh, A. and Subbarao, C. 2006. Tensile strength bearing ratio and slake durability of class F fly ash stabilized with lime and gypsum. *Journal of Materials in Civil Engineering* 18, 1, pp. 18–27.

[6] Ghosh, A. 1997. *Environmental and engineering characteristics of stabilized low lime fly ash.* Ph.D. dissertation, Indian Institute of Technology, Kharagpur, India.

[7] Ghosh, A. and Subbarao, C. 2007. Strength characteristics of class F fly ash modified with lime and gypsum. *Journal of Geotechnical and Geoenvironmental Engineering* 133, 7, pp. 757–766.

[8] Pandian, N. S. 2004. Fly ash characterization with reference to geotechnical applications. *Journal of the Indian Institute of Science* 84, pp. 189–216.

[9] Suthar, R. and Aggarwal, P. 2015. Class-F pond ash a potential highway construction material – a review. *Indian Highways* 43, 8, pp. 23–32.

[10] Sahu, V. and Gayathri, V. 2014. The use of fly ash and lime sludge as partial replacement of cement in mortar. *International Journal of Engineering and Technology Innovation* 4, 1, pp. 30–37.

[11] Battaglia, A., Calace, N., Nardi, E., Petronio, B. M., and Pietroletti, M. 2007. Reduction of Pb and Zn bioavailable forms in metal polluted soils due to paper mill sludge addition: Effects on Pb and Zn transferability to barley. *Bioresource Technology* 98, 16, pp. 2993–2999.

[12] Calace, N, Campisi, T., Iacondini, A., Leoni, M., Petronio, B. M., and Pietroletti, M. 2005. Metal-contaminated soil remediation by means of paper mill sludges addition: chemical and ecotoxicological evaluation. *Environmental Pollution* 136, 3, pp. 485–492.

[13] Mahmood, T. and Elliot, A. 2006. A review of secondary sludge reduction technology for the pulp and paper industry. *Water Research* 40, 11, pp. 2093–2112.

[14] Day, W. R. 2001. *Soil testing manual procedures, classification data, and sampling practices.* McGraw-Hill Professional, USA.

[15] Suthar, M. and Aggarwal, P. 2018. Predicting CBR Value of Stabilized Pond Ash with Lime and Lime Sludge Using ANN and MR Models. *International Journal of Geosynthetics and Ground Engineering* 4, 6, pp. 2–7. doi: https://doi.org/10.1007/s40891-017-0125-3

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

[17] Sabat, A. K. 2015. Prediction of California bearing ratio of a stabilized expansive soil using artificial neural network and support vector machine. *Electronic Journal of Geotechnical Engineering* 20, pp. 981–991.

[18] Erzin, Y. and Y, Turkoz, D. 2016. Use of neural networks for the prediction of the CBR value of some Aegean sands. *Neural Computing and Applications* 27, 5, pp. 1415–1426.

[19] Pal, M., Singh, N. K., and Tiwari, N. K. 2013. Pier scour modelling using random forest regression. *ISH Journal of Hydraulic Engineering* 19(2), pp. 69–75.

[20] Breiman, L. 1999. *Random forests - Random Features.* Technical Report 567, Statistics Department, University of California, Berkeley.

[21] Breiman, L. 1996. Bagging predictors. *Machine Learning* 24, 2, pp. 123–140.

[22] Quinlan, J. R. 1992. Learning with continuous classes. In *Proceedings of Australian Joint Conference on Artificial Intelligence.* World Scientific Press: Singapore, pp. 343–348.

[23] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. 1984. *Classification and Regression Trees.* Chapman and Hall/CRC, Wadsworth, Monterey, CA.

[24] Pal, M, and Mather, P. M. 2003. An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification. *Remote Sensing of Environment* 86, 4, pp. 554–565.

[25] Feller, W. 1968. *An Introduction to Probability Theory and Its Application.* Vol. 1, 3rd ed., Wiley, New York, USA.

[26] Bureau of Indian Standards. 2002. IS 2720-7: Methods of test for soils, Part 7: Determination of water content-dry density relation using light compaction.

[27] Bureau of Indian Standards. 2002. IS 2720-16: IS 2720-16: Methods of test for soils, Part 16: Laboratory determination of CBR.