ISSN: 2096-3246
Volume 58, Issue 01, 2026

# PREDICTING SOIL CALIFORNIA BEARING RATIO USING SUPERVISED MACHINE LEARNING ALGORITHMS

**Nikita Rahaja [1], Ashok Kumar Gupta [2], and Kushal Kanwar [3]**
[1,2] Department of Civil Engineering, Jaypee University of Information Technology, Waknaghat, Solan 173234, Himachal Pradesh, India
[3] Department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology, Waknaghat, Solan 173234, Himachal Pradesh, India

**Received : 10-11-2025 Revised: 06-03-2026 Accepted: 30-04-2026**

**Abstract**; California Bearing Ratio (CBR) is an important indicator to evaluate base course materials and subgrade soils in pavement systems. This implementation of the machine learning (ML) approach in this research predicts CBR (California bearing ratio) values of the soil based on seven predictors for which values such as maximum dry density, soil classification, optimum moisture content, liquid limit, plastic limit, plastic index and swell can be easily obtained from the laboratory data using random forest (RF), decision tree (DT), linear regression(LR) and artificial neural network(ANN) models. Three hundred fifty-two soil samples that composed an experimental data set were classified in accordance with AASHTO M 145. They were divided into test data (20%) and training data (80%) in this model study. The performance of the models was evaluated with respect to standard statistical metrics like MSE (mean squared error), MAE (mean absolute error), RMSE, coefficient of determination and correlations. Based on these evaluation metrics, the random forest algorithm receives a smaller error and relative fashioned error (R2) values against other algorithms Thus a random forest algorithm developed using the analysis can make an accurate prediction of the soil's CBR.
Keywords: Expansive soil, California Bearing Ratio, Machine learing Algorithms

## 1 Introduction

The California Bearing Ratio (CBR) is widely recognized as a parameter used to determine the strength of subgrade materials in pavement layers constructed for roads, airfields, and railways; it is an important parameter from civil engineering, specifically construction materials and geotechnical engineering. CBR can be measured either in field or laboratory conditions. The test method for CBR is such, the two is you will push down a piston in the soil or subgrade material at your test site using a loading jack. This method is used for assessing the shear strength of in situ soils and base course materials for pavement design. On the other hand, at this stage field CBR testing tools are expensive, heavy and cumbersome to transport from site-to-site. Thus, laboratory testing is applied to determine the CBR of soil and subgrade materials more usually. In laboratory method, a plunger of standard diameter is pressed into a compacted soil sample prepared at optimum moisture content, generally at penetration rate of 1.3 mm/min. The CBR values can be determined from soaked and unsoaked soil samples. Typically, the soaked soil samples give lower CBR values compared to unsoaked ones [1]. Due to this reason, soaked CBR is routinely used for quality evaluation of weak or subgrade material.

Yet, measuring CBR in laboratories require relatively long periods of time and this may cause construction delays. Soil that is remolded at the optimum moisture content usually needs to soak in water for four days prior to performing testing. This condition shows the worst case scenario particularly when rainfall will be continuous 04 days. Specifications of CBR values are often required for large number of soil samples, involving lengthy and expensive laboratory testing [2].
Skilled lab technicians who can perform CBR testing also may not be as accessible, which can further hamper project completion. One proposed approach to tackling this issue is by predicting the CBR values through machine learning which eliminates the requirement for continuously repeating laboratory tests. Machine learning is a subfield of artificial intelligence concerned with computer algorithms that evolve their performance using past experience without being explicitly programmed to improve the next trial [3]. Normally machine learning includes different types of learnings like supervised learning, unsupervised learning, reinforcement and semi supervised. Vermeth (2019) Effects of Feature Selection Algorithms on Prediction Performance of CBR Values Using Supervised Regression Techniques · Zoned soil classification Schemes this study deals with supervised regression techniques for prediction of CBR values. Regression is the procedure through which it attempts to find out the relationship of independent variables with a dependent output variable. More recently, machine learning approaches have also been used to address real geotechnical engineering problems. Artificial neural networks (ANN), multilayer perceptron neural networks (MLP), gene expression programming (GEP), support vector machines (SVM) and the multigroup method of data handling have been used to predict geotechnical output parameters of interest. Related studies for this area are shown in Table 1.
Researchers from these prior studies used different methods for prediction modeling. CBR prediction is dependent on various attributes such as size of sample collected, soil index properties and types of predictors used and also the quality of laboratory testing protocols. However, available literature notes that having a larger sample size does not always coincide with increased prediction accuracy. An example is the study of one SVM model which predicted 98 percent with just 49 samples [5]. Conversely, a different research utilizing 389 samples achieved less prediction accuracy [10]. The current study used a dataset where 252 samples were employed and provided an accuracy of 84 percent by the model. This indicates that the soil formation characteristics differ from previous studies, and further study discusses collection techniques in relation to a specific study area; therefore, predictive models can be constructed specifically for that field site. Further, this study also pinpoints and analyses important features which have the greatest influence on CBR prediction. The research critically aims at inferring the value of CBR by observing the response variable behavior with respect to retained independent variables only.

## 2 Methodology

### 2.1 Index Properties of Soils

Soil compaction is a method of soil densification that increases the density of soils by mechanical energy through the removal of air voids among soil particles. It more dynamically improves the strength characteristics of soil and reduces permeability. The densification of the soil depends on the effort that has been applied for compaction and on how much water has been added to the soil. The

relationship between water content and density is typically represented by a curve called the compaction curve , moisture density curve, or modified proctor test. There are standard methods for obtaining this curve which is generally generated by performing the standard proctor test, modified proctor test and AASHTO testing procedures.

Standard methods used for the determination of moisture density relation and related soil properties include California Bearing Ratio (CBR) test under AASHTO T 193 [11], swell test under AASHTO T 258 [12], maximum dry density = MDD, optimum moisture content (OMC), OMC value based on imbibition or based on suction using a porous stone, saturated surface dry SSD mass of water absorbed per unit volume: Vv to blue circle show experimental soothing; and empirical ball/ring criterion. With these tests, you can be performed in coarse grained and fine graine soils. Soil swelling potential is an essential component of geotechnical investigation in order to get pavement design criteria. So in these type of studies, the soil samples are generally extracted from shallow depth beneath the expected pavement level, and their swelling behaviour can be quantified through different testing methods.

Atterberg limits are popularly used to evaluate the shrink–swell potential of soil and to find out the plasticity state condition of soil. The tests of shrinkage limit and plastic limit are generally done in the laboratory. Plasticity Index, Liquid Limit and Plastic limit of a soil by AASHTO T 89 and AASHTO T 90 respectively. This is the point where plastic soil starts acting like a liquid when its moisture content is at the second limit, which is referred to as the liquid limit. Plastic Limit – It is the moisture content at which a semi solid soil changes into plastic state. The plasticity index, or PI, is calculated as PI = LL − PL where LL is the liquid limit and PL is the plastic limit.

**2.2 Dataset**

Originally developed by the California Department of Highways in the late 1920s, its purpose is to determine the bearing ratio of cohesive soil particle material that would be used in pavement subgrade and subbase layers. The test was subsequently standardized by organizations including the American Association of State Highway and Transportation Officials and the American Society for Testing and Materials.

Major organisations like the Federal Highway Administration, the Federal Aviation Administration and AASHTO in America have employed CBR values to design and construct roads, airports, parking areas and other pavement structures. Similar relationships were established by researchers between CBR and other engineering soil parameters (resilient modulus) through the empirical functions among them. Since CBR is not a fundamental material property, it is less appropriate for mechanistic or mechanistic empirical design methods than other testing techniques but it still has wide use in practice as has long history associated with its use in pavement construction, relatively easy and inexpensive to conduct test and reasonably good correlation with more fundamental properties such as resilient modulus.

Subgrade materials are often characterized in relation to both their strength and their stiffness. In the US it is CBR, elastic or resilient modulus and modulus of subgrade reaction that are the primary parameters used to characterise subgrade stiffness or strength. However, stiffness is only one of the

ISSN: 2096-3246
Volume 58, Issue 01, 2026

most frequently used parameters for evaluating subgrade, and other aspects should also be taken into account in processes such as swelling possibility in clayey soils. The uniformity of the subgrade also impacts pavement performance. Demanding a perfectly uniform subgrade is unrealistic as soil naturally varies and moisture, temperature and construction activities affect the uniformity of the subgrade. Work by [16] indicated that when the subgrade strength is less than CBR 10, then there may be deflection in a similar manner to the subgrade under traffic loading. This causes an initial deflection, that influences the performance of pavements, firstly in the case of flexible pavements and later in rigid pavements. Statistical significance of the datasets used in this study is shown in Table 2. Figure 1 shows the relationship among those parameters.

2.2.2 Influencing Factors

This research also used the random forest model to predict CBR based on seven important input factors These include soil classification such as A-7-5, A-2-4, A-2-7, A -2 -6 and A 7; Liquid limit (LL); Water ratio optimal (OMC), plastic limit (PL); Maximum dry density and Plasticity index. The dependencies include the California Bearing Ratio and swell. The subgrade material of high quality depends on the detailed knowledge of their properties, their grading, skilled laboratory technicians and experienced geotechnical engineers along with modern quality control testing. On the other hand, the standards for pavement design and engineering investigation should be commensurate with the significance, scale, lifespan and investment cost of a project. Thus, a somewhat elementary recognition of subgrade properties is required for pavement design. This includes properties of soil such as the type, density and coefficient of lateral earth pressure (K0) with permeability, internal friction angle $\phi$, cohesion and estimated CBR or resilient modulus, data sets 352 records and 8 CBR related attributes were retrieved. The data gathered required preprocessing to improve the results of a model before it could be developed. Data preprocessing includes cleaning, transformation, integration and attribute selection [18]. There was no need to put time into data filling, as data from the DANA consulting laboratory had no missing values. Although this data transformation was required due to the categorical nature of soil classification values. To represent these categorical values, corresponding numerical values were assigned which could therefore be used in the chosen machine learning methods and algorithms. Table 1 presents the attributes employed in this research work.

The distribution over the attributes used for the analysis is depicted in figure 2. Figure where the density plot of all variables in the dataset is shown with smooth curves for each variable against a histogram of counts. According to this plot, the no skewness from liquid limit, plastic limit, maximum dry density and optimum moisture content; indicating that the mean < median. For CBR, the distribution is right skewed due to a greater mean compared to the median. It is clear from the box plot of attribute distributions that outliers are present in plastic limit, maximum dry density, optimum moisture content and CBR. This means that certain values are higher than the packed range of the data. The green lines inside the box plot are the median values of parameters. The percentage of CBR values from outer dataset as shown in Figure 1 it can be seen that most data point fell in between value ranges from 0.5 to 5.

2.3 Methods Used

ISSN: 2096-3246
Volume 58, Issue 01, 2026

The prediction model was developed using the Python programming language in Jupyter Notebook environment for this experimental study. For example, we split our dataset into two sets, 80 percent of our data is used for training purposes and 20 percent of the data treatments were used for testing. Supervised machine learning algorithms (e.g., random forest, decision tree and linear regression) were chosen according to the data linearity in consideration of sample size and number of input parameters. In this study, a deep learning method: artificial neural network (ANN) technique was also used as another tool to predict the CBR value.

Using the data of soil sample collected in lab for study area these techniques were implemented. We compared the prediction performance among each method and chose the one with highest prediction accuracy as the best approach. Supervised learning is a worked development between input variables and the tar – get output variable.

Table 1: Statistical value of the data in the study.

| | SC (%) | LL (%) | PL (%) | PI (%) | MDD (kg/m$^3$) | OMC (%) | $^{Swell}$CBR (%) | (%) |
|---|---|---|---|---|---|---|---|---|
| Max. | 34 | 80.1 | 39.29 | 36.57 | 14 | 29.5 | 10.23 | 104 |
| Min. | 39 | 34 | 24.78 | 8.53 | 1.03 | 1.6 | 0.15 | 0.14 |
| Avg. | 33.57 | 51.1 | 29.52 | 24.14 | 1.0 | 19.03 | 2.16 | 11.17 |
| Mean | 33.57 | 51.1 | 29.52 | 24.14 | 1.0 | 19.03 | 2.16 | 11.17 |
| Skew | −0.17 | −0.08 | 0.78 | −0.12 | 0.15 | 0.15 | 0.13 | 1.53 |
| Kurt | −1.74 | −0.59 | 3.94 | −0.75 | 224.55 | 1.82 | −1.03 | 2.9 |
| Var | 0.15 | 86.37 | 7.64 | 55.12 | 1.0 | 16.55 | 5.17 | 254.03 |
| Std | 0.4 | 7.82 | 2.1 | 6.07 | 1.0 | 3.72 | 1.03 | 13.85 |

where SC soil classification, LL liquid limit, PL plastic limit, PI plasticity index, MDD maximum dry density, OMC optimum moisture content, CBR California bearing ratio.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

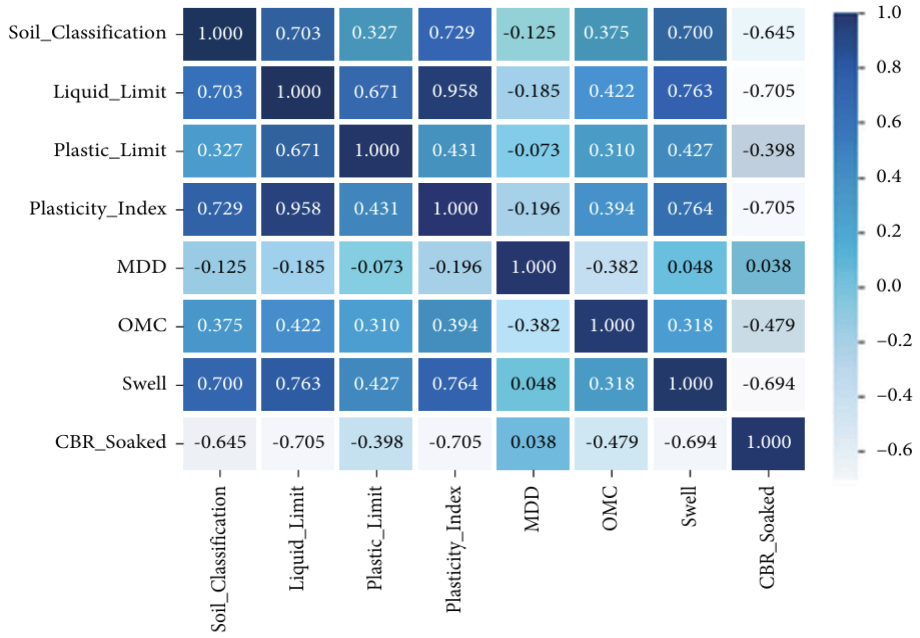



FIGURE 1: The correlation of parameters.

Supervised learning, which is one of the major tasks, also includes regression. Four evaluation metrics were utilized to assess the performance of the selected methods in this study: mean squared error (MSE), root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination ($R^2$). Figure 2 provides an overview of the workflow of the study.

2.3.1 Random Forest ML Model

Random forest is an ensemble machine learning algorithm that creates multiple decision trees using a randomised version of the tree induction process. While using a classic decision tree, the random forest impulse puts randomness into action in model building. This helps create distinct trees and makes the prediction accuracy of the model better.

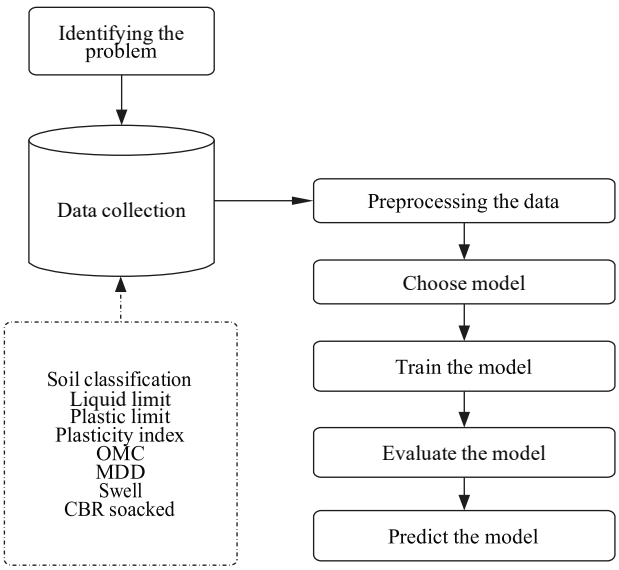



Figure 2: Flowchart of the study.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

The random forest model picks split points close to the top of the tree which are almost as effective as the optimal splits and expands them, at which point it uses ID3 induction as normal.

Breiman was one of the pioneering researchers to mathematically and empirically demonstrate, in his seminal 1996 work, that ensembles of various versions of a predictor can substantially improve the reliability of models. He investigated and proved that the mean model to models (Lm) trained on the training set L where m = 1,…,M) has an expected generalisation error lower than by using other choice of models [21]. The form of (L_m) is multiple copies of (L), where N examples ((x, y)) are selected randomly with replacement from (L). While (|L| = |L_m| = N), through bootstrap replication, we observe that at least 37 percent of the ((x,y)) pairs from (L) survive on average. This happens because even after (N) random selections with replacement it is still quite likely that some observations would never be chosen.

where L is the Training DataSet that consists only about 67 percent of observations subsampling, which will increase bias when it is too small. E.g., model accuracies may become worse, and the increase in bias may be too large for the corresponding anticipated reduction in variance to balance. This can weaken the overall model performance.

Well, bagging is useful in a much wider range of scenarios than this, because it can be used to enhance various different model types (not only decision trees). Breiman used a combination of bagging and random variable selection for each node in the original random forest paper. The integration of these two approaches, plus some controlled chance: yielded one of the best general purpose machine learning algorithms, being successful through many kinds of problems. While boosting and arcing algorithms are forms of competitive techniques aimed at reducing bias, random forests is predominantly geared towards minimizing prediction error.

Random forest technique was used to model the soil test data along the section of Mekane Eyesus to Simada town road in this study. The California Bearing Ratio was predicted using the proposed method thereby minimizing soil testing time and cost in this specific area. The prediction approach can even be applied to other infrastructural/road construction projects in the same study area.

2.4.2 Decision Tree

A decision tree is a bottom-up supervised learning model that finds local data regions based on a splitting process iterated at different data levels. It contains internal decision nodes and terminal leaf nodes. You can use this technique in case of classification or regression task. Building a regression tree follows almost the same process as building a classification tree. The primary difference is that, for regression trees, the classification impurity measure used to split the data in a decision tree is replaced with an appropriate regression direction.

When the error at a node is small enough, i.e., Em < θr type 1 ends up creating a leaf node storing gm by creating piecewise constant approximations along with discontinuities for certain inputs in leaf boundaries. If the error is acceptable then it reaches node m else the data reaching them are divided into further branches so as to minimize the total error in all resulting branches.

Table 2: Results of the Study

| Algorithm | MAE | RMSE | MSE | *R*2 |
|---|---|---|---|---|

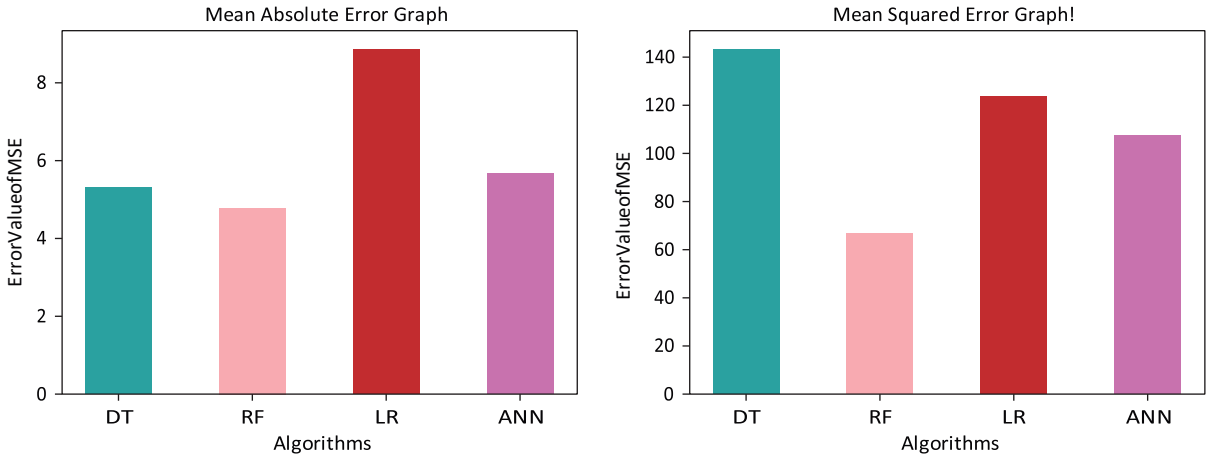
ISSN: 2096-3246
Volume 58, Issue 01, 2026

| | | | | |
|---|---|---|---|---|
| RF | 4.8 | 8.13 | 66.24 | 0.84 |
| DT | 5.3 | 11.92 | 143.3 | 0.66 |
| LR | 8.9 | 11.08 | 122.8 | 0.53 |
| ANN | 5.69 | 10.36 | 107.46 | 0.67 |

2.4.3 Linear Regression

Linear regression is only applicable when the form of the model with regards to regression parameters is linear. Regression analysis checks how certain predictor variables relate with a response variable. Predictor variables ($x_1$, $x_2$,...,xp) (sometimes called independent variables or explanatory variables or control variables or regresors) The response variable is also known as the dependent variable, explained variable or predicted variable and is typically denoted by y.

Regression types — there are 3 in general. The first is simple linear regression, which we can use to model a two-variable relationship between one independent variable x and one dependent variable y while the second type is multiple linear regression where there only exists one dependent variable and multiple independent variables. The response variable is a linear function of the model parameters with more than one independent variable. The third type of regression is done when the dependent variable does not have a linear relation with the independent variable, that is, the parameters of regression do not follow a simple linear pattern [22].



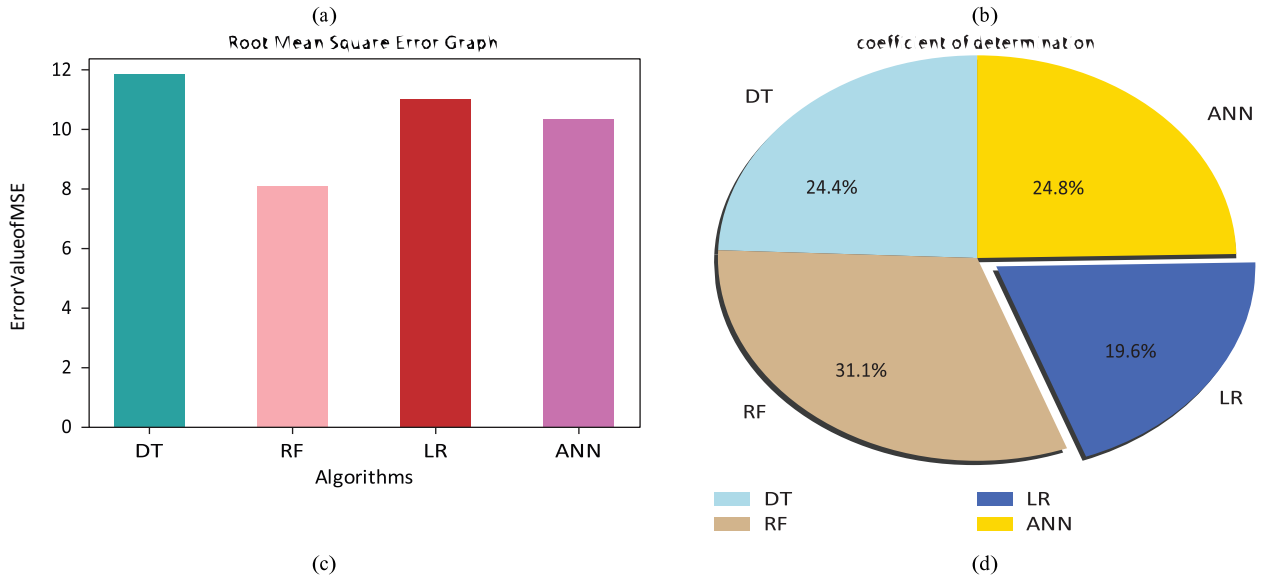



Figure 3: Comparison of the algorithms: (a) mean absolute error, (b) mean squared error, (c) root mean square error, and (d) coefficient of determination.

Xi and yi are the ith experimental and measured outputs, deemed to be a high degree correlation between the actual and estimated values as given when (R2≈1) [23]. The correlation coefficients are standardized such that they is between −1 and +1, where 0 used for absence of a linear or monotonic association. As the coefficient nears an absolute value of 1, it signifies a strong relationship that tends to a straight line. [23]. Secondly, you favour RMSE because RMSE 0 means fewer errors and larger residual errors are treated more gently. And, in other situations when RMSE is not the one to get a higher precision MAE is used because works with both smooth and continuous data. Higher R-values with lower RMSE and MAE values also suggest more reliable model performance and good calibration.

**Results**

Here are the experimental setups that stand in line with this research. The data consists of 252 records and 8 attributes which are separated in to train-test split at along the line of 80%:20%. In this study, - The author predicts the CBR with random forest, decision tree, linear regression and artificial neural network algorithms. To assess the effectiveness of the algorithms, mean squared error, root mean square error, mean absolute error, and relative error are included. Table 4 shows the output of this method.

From the above table, it can be seen that the values of MAE, RMSE and MSE are small for random forest. In other words, CBR of aggregate is the smallest error of these parameters in order to predict this material, a highest value R2 indicates that there is a high correlation between parameters. Figure 6 shows the error value of the algorithms. From the above figure, RF gives lower mean absolute error, mean squared error and root mean square error than others which means a lower rate of prediction errors. In terms of relative error, it also scores the highest value (coefficient of determination). The maximum value of relative error indicates that there is a strongest correlation or most impacted between variables.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

**DISCUSSIOIN**

In general, in this research Random forest, decision tree, linear regression and artificial neural network algorithms are being used on the 80% training and 20% testing data. From this it predicts well RF and minimum error in the value of MAE, MSE and RMSE when compared to others. Furthermore, RF obtains a better score in R2 or coefficient of determination indicating that the relationship between attributes becomes stronger than the other. So there is no way to say that an algorithm is fit for all studies and data. The dataset and the parameters themselves have an impact on obtaining a result. Thus, this purpose was achieved using the best found algorithm for the collected data and identifying which determinant factor is more important in predicting the CBR value. Thus it turns out that random forest is the algorithm chooses at the end for CBR prediction on this study. We have RMSE as a cost function, which computes the difference between the actual target

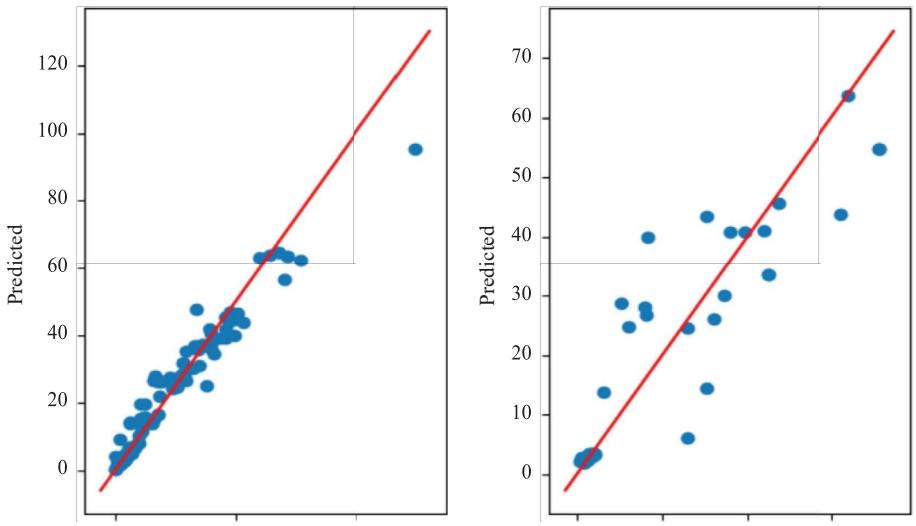



Figure 4: The actual and predicted values of training and testing sets using RF.

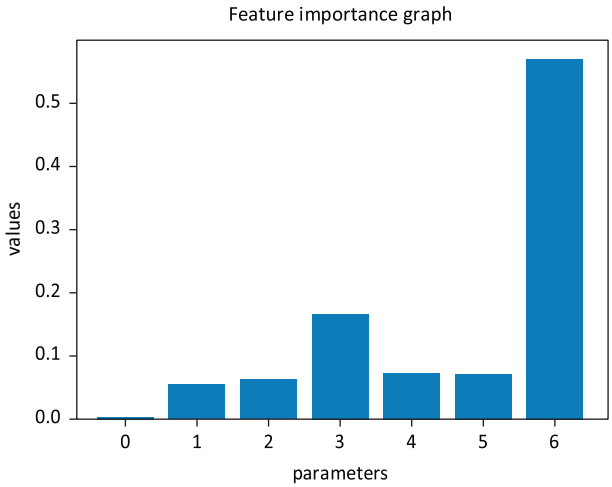



Figure 5 : Feature importance graph. 0 soil classification, 1 liquid limit, 2 plastic limit, 3 plasticity index, 4 maximum dry density 5 optimum moisture content, and 6 swell.

values in the holds-out set and the predictions from your model The comparison RMSE of the

algorithms written in the following Figure 5 As you can see from the figure above, the random forest algorithm gives us better MSE (mean squared error), MAE (mean absolute error) and RMSE (root mean square error) which are smaller than those of other methods. In random forest regressor algorithm, author experimented with the parameters like n_estimators value values as 20,50,100,150 and 200 and max_depth value values as 10,20,30 and 50 in order to get the better result. The variation of n_estimators and max_depth parameters in RMSE metric for random forest is shown in Figure 5. In the table above, actual values are shown in red and predicted values in green. Along with this, Feature Importance was tried to find the most impactful feature from input. From this, the 3 ranked significant attributes used for predicting CBR are swell, PI and MDD shown in Figure10. Swell is the first key predictor of CBR. Beside it, plastic index (PI) and maximum dry density (MDD) are the others important one from top to bottom.

**Conclusions**

The present study consisted of the predictions of soil CBR using various models based on random forest, decision tree, linear regression and artificial neural networks trained in 80% training and 20% test sets. The input variables to the models of the machine learning algorithms are SC, LL, PL, PI, OMC, MDD, swell and CBR. The author attempted to utilize these algorithms and perform the experiment to estimate the CBR amounts for this soil. The major findings of the study are evaluating the methods in this dataset selecting the right algorithm which makes prediction better in this data set. Based on this, R2 value of 0.99 for the RF algorithm while in prediction algorithms The closer to 1 means that there is a stronger relationship between attributes. Furthermore, it produces the least MAE, MSE and RMSE compare to the remaining algorithm. How accurate the prediction specifically is highly depends on the dataset and methods used. Furthermore, in future works the authors can implement RF, DT, LR and ANN models which can be refined using more input data and validate these results against those of various ML models.

**References**

[1] S. M. Lakshmi, S. Subramanian, M. Lalithambikhai, A. M. Vela, and M. Ashni, “Evaluation of soaked and unsoaked CBR values of soil based on the compaction characteristics,” *Malaysian Journal of Civil Engineering*, vol. 28, no. 2, 2016.

[2] B. Gunaydin and O. Gunaydin, “Estimation of California bearing ratio by using soft computing systems,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 6381–6391, 2011.

[3] M. Vaquero Barnadas, *Machine Learning Applied to Crime Prediction*, Universitat Politecnica de Catalunya, Barcelona,` Spain, 2016.

[4] S. Taha, S. El-Badawy, A. Gabr, A. Azam, and U. Shahdah, “Modeling of California bearing ratio using basic engineering properties,” in *Proceedings of the 8th International Engineering Conference*, Sharm Al-Sheikh, Egypt, November 2015.

[5] A. K. Sabat, “Prediction of California bearing ratio of a stabilized expansive soil using artificial neural network and support vector machine,” *Electronic Journal of Geotechnical Engineering*, vol. 20, no. 3, pp. 981–991, 2015.

[6] D. Q. Vu, D. D. Nguyen, Q.-A. T. Bui, D. K. Trong, I. Prakash, and B. T. Pham, "Estimation of California bearing ratio of soils using random forest based machine learning," *Journal of Science and Transport Technology*, vol. 1, pp. 48–61, 2021.

[7] D. K. Trong, B. T. Pham, F. E. Jalal et al., "On random subspace optimization-based hybrid computing models predicting the California bearing ratio of soils," *Materials*, vol. 14, no. 21, p. 6516, 2021.

[8] A. Bardhan, C. Gokceoglu, A. Burman, P. Samui, and P. G. Asteris, "Efficient computational techniques for predicting the California bearing ratio of soil in soaked conditions," *Engineering Geology*, vol. 291, Article ID 106239, 2021.

[9] L. S. Ho and V. Q. Tran, "Machine learning approach for predicting and evaluating California bearing ratio of stabilized soil containing industrial waste," *Journal of Cleaner Production*, vol. 370, Article ID 133587, 2022.

[10] A. R. Patel and A. Patel, "Utilization of support vector models and gene expression programming for soil strength modeling," *Arabian Journal for Science and Engineering*, vol. 45, no. 5, pp. 4301–4319, 2020.

[11] T. Taskiran, "Prediction of California bearing ratio (CBR) of fine grained soils by AI methods," *Advances in Engineering Software*, vol. 41, no. 6, pp. 886–892, 2010.

[12] M. Arsyad, I. B. Mochtar, and N. E. Mochtar, "Analysis of settlement of the road with full scale geotextile reinforcement on the very soft soil (case study in tapin regency, south kalimantan)," in *MATEC Web of Conferences*vol. 280, EDP Sciences, Article ID 03012, 2019.

[13] J. Connelly, W. Jensen, and P. Harmon, *Proctor Compaction Testing*, The Constructor Building Ideas, Chennai, India, 2008.

[14] A T89, "Determining the plastic limit and plasticity index of soils," *Standard Specifications for Transportation Materials and Methods of Sampling*, The Constructor Building Ideas, Chennai, India, 2014.

[15] Astm, "3282/AASHTO M 145," *Practice for Classification of Soils and Soil-Aggregate Mixtures for Highway Construction Purposes*, ASTM International, Pennsylvania, USA.

[16] V. R. Schaefer, D. J. White, H. Ceylan, and L. J. Stevens, "Design guide for improved quality of roadway subgrades and subbases," *Iowa Highway Research Board (IHRB Project TR525)*, vol. 7, pp. 8–72, 2008.

[17] H. F. Southgate, *Comparison of Rigid Pavement Thickness Design Systems*, Chennai, India, 1988.

[18] F. Calders and T. Calders, "Data preprocessing techniques for classification without discrimination," *Knowledge and Information Systems*, vol. 33, no. 1, pp. 1–33, 2012.

[19] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 3, pp. 128–138, 2017.

[20] S. W. Kwok and C. Carter, "Multiple decision trees," in *Uncertainty in Artificial Intelligence*, vol. 9, pp. 327–335, Elsevier, 1990.

[21] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

ISSN: 2096-3246
Volume 58, Issue 01, 2026

[22] X. Yan and X. Su, *Linear Regression Analysis: Theory and Computing*, World Scientific, Singapore, 2009.
[23] P. Schober, C. Boer, and L. A. Schwarte, "Correlation Coefficients: Appropriate Use and Interpretation," *Anesthesia and analgesia*, vol. 126, pp. 1763–1768, 2018.