Asian Soil Research Journal

6(2): 32-46, 2022; Article no.ASRJ.86861
ISSN: 2582-3973

Application of Soft Computing Techniques in Modelling of Soaked and Unsoaked California Bearing Ratio

Akinwamide Joshua Tunbosun [a*], Jacob Odeh Ehiorobo [b], Osuji Sylvester Obinna [b] and Ebuka Nwankwo [b]

[a] Department of Civil Engineering, Faculty of Engineering, The Federal Polytechnic, Ado Ekiti, Ekiti State, Nigeria.
[b] Department of Civil Engineering, Faculty of Engineering, University of Benin, Benin City, Edo State, Nigeria.

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/ASRJ/2022/v6i230128

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/86861

Original Research Article

Received 09 March 2022
Accepted 17 May 2022
Published 23 May 2022

ABSTRACT

In this study, author attempted to establish a correlation between soil physical parameters and California Bearing Ratio of lateritic soils using advanced mathematical techniques such as the Support Vector Machine (SVM), Random Forest (RF), M5 tree, multiple linear regression, and Artificial Neural Network. A total of 480 soil samples were collected and separated into a data set using training and validation of the generated models based on the main soil parameters of Liquid Limit (LL), Plastic Limit (PL), Natural moisture content (NMC), Specific gravity (GS), Fines (F), Gravel, and Sand. The Principal Component Analysis (PCA) was used to minimize the dataset's huge dimension, and the approximate sum of the first four principal components (PC) captured 88 percent of the variability in the response variable with just 12% information loss. The RMSE values of 21.6, 21.23, 295.67, 7.03, 14.54 and 24.43,24.59,326.49,8.63,17.71 are from the MLR, ANN, MS Tree, RF, and SVM models for SCBR and USCBR values, respectively. For SCBR and USCBR, random forest (RF) yielded the lowest values of 7.03 and 8.63, respectively. Similarly, the R values range from 0.1 to 0.94 and 0.01 to 0.92, indicating that the anticipated and real SCBR and USCBR are related. The Random Forest Model for SCBR and USCBR was shown to be the best by the

**Corresponding author: E-mail: akinwamidejoshua2@gmail.com;*

correlation coefficient values, while the MS tree model for SCBR and USCBR was determined to have the lowest coefficient of determination R2. As a result, it can be concluded that Random Forest provided the best Soaked and Unsoaked CBR model based on the dataset, while MS tree provided the poorest model. The model is a valuable tool for evaluating the subsurface indices of a civil engineering site at the preliminary planning stage before final structural design for the substructures, as the anticipated soil parameter values are within permitted accuracy.

Keywords: Compaction characteristics; soaked; unsoaked; california bearing ratio; highway.

1. INTRODUCTION

In today's Nigeria, highways are the primary mode of transportation. They transport over 90% of the country's passenger traffic and 80% of its freight. A well-maintained physical road network in both urban and rural areas is critical for social and economic development. Lane strength is low, and most federal roadways are either double or single lanes [1]. Seventy-five percent of Nigeria's roads are crowded. The majority of roads are of low quality, and road maintenance is underfunded, with only around 25% of maintenance demands being satisfied. This has resulted in the state of the roadway deteriorating and hefty transportation prices for users. Roads are critical civil engineering infrastructure facilities that provide links and access to other aspects of a man's social life in a healthy society. Roadways are designed using soil strength features, as well as the automotive and pedestrian loads they must sustain over their lifetime [2]. To test some of a region's most important geotechnical indices, extensive sample throughout the proposed path would be required, which will take time and money. The attributes of soil strength CBR (California Bearing Ratio) for drenched and unsoaked tires In Nigeria, CBR is the most often used geotechnical parameter for determining the overlay thickness of flexible pavements. Highway engineers encounter issues evaluating the CBR of the soil while measuring the thickness of the Sub-Base and Base-Course layers, and time and budget resources are redistributed during preliminary study or complete project detailed report and collecting of enormous CBR data. In such circumstances, CBR data for any planned civil engineering infrastructure could be estimated using proven models combining CBR and soil index features, which provide efficient and cost-effective solutions. CBR values are frequently required to provide geotechnical solutions for roadway structures, especially during the planning and design stages. Such sub soil evaluation in mapping the variance in their values along the alignment is generally threatened by cost and time [3].

2. LITERATURE REVIEW

Several experts have recently conducted experimental experiments to determine the relationship between CBR and various soil physical characteristics. Ayodele [4] conducted a laboratory investigation on some soils in the South Western state and discovered that CBR and geotechnical indices have a >90% correlation. Dharamveer [5] created statistical formulas. Using five different soils and 100 samples, a solid link was established between the experimented and estimated CBR values. Using triaxle laboratory testing to estimate subgrade soil Moduli, a relationship was developed between MR and soil physical characteristics of cohesive soils and cohesion less (sandy) soil [6]. Multiple Linear Regression (MLR) on CBR was established in relation to the Plasticity Index (PI), MDD, index properties, and OMC [7]. The Plasticity Index is inversely proportional to the CBR values (PI). CBR values drop when PI values rise [8]. CBR and frictional angle can be calculated from fine-grained soils for a variety of soil physical characteristics [3,9]. Single Linear Regression Analysis (SLRA) and MLRA on CBR and geotechnical characteristics were constructed using 33 soil samples obtained from an ongoing road building site. The existence of a link between CBR values and soil physical characteristics was confirmed [10]. Five machine learning tools were chosen to predict soaked and unsoaked CBR from physical soil indices in this study. The goal of this study is to use R and R Studio software to create machine learning models to predict the subsoil properties of non-visited locations using Multiple Regression, RF, ANN, SVM, and MS TREE, and to compare the models' accuracy in prediction by calculating coefficient of determination (R2), MAE, and RMSE in Ekiti – State senatorial districts.

3. MATERIALS AND METHODOLOGY

3.1 Datasets and Data Analysis

The 480 data points used in this study were from established borrowed pits in Ekiti State Senatorial District Zones (ESSDZ) in southern Nigeria. The laboratory test was conducted at the federal polytechnic Ado Ekiti, Ekiti State, Nigeria's Department of Civil Engineering's Material Testing Unit. For this study, R version 4.0.5 and R studio version 1.2.5033 were utilized (R Core Team 202). The Principal Component Analysis (PCA) was used to decrease the dataset's vast number of dimensions.

4. RESULTS AND ANALYSIS

4.1 Soil Index Properties

The results of the soils index properties (NMC, GS, LL, PI and % passing sieve 200) and strength properties (OMC,MDD, Soaked CBR and Unsoaked CBR) of the studied soils samples were presented in Table 1 and 2 respectively.

4.2 Classification of Soils within the Senatorial Districts

The results of Central Senatorial Districts classified the soils into four classes as clay of low compressibility (CL) clay of high compressibility (CH) according to [11] and A-2-4, A-2-6, A-2-7 and A-7-6 for AASHTO classification system. From the foregoing, the soils classified some as low plasticity, Sandy gravelly clay, clayey soils and others as medium compressibility soils which agree with [12] findings. and Southern Senatorial Districts samples were classified into Eight as A-2-4, A-2-5, A-2-6, A-2-7, A-4, A-5, A-6 and A-7-5 which describe soils in the study area as Clay gravelly sand silty clay materials according to USCS while Ekiti State Northern Senatorial Districts were classified into Six classes thus A-2-4, A-2-5, A-2-6,A-2-7, A-6 and A-7-6 respectively. Many of the zones had a very high percentage finer than 0.075 fractions that is > 35% which classifies the soils in the study area as clay of high compressibility (CH) and silty gravelly soils for Southern, Northern and Central senatorial districts respectively. the percentage of fines (% passing sieve 200) has significant effect on the performance of the base / sub-base materials excess fines will result in the reduction in maximum dry density and increases the susceptibility to weakening from water infiltration or ingress (Garg 2009). The results showed that the number of fines found in the soil samples are quite similar . The “well graded” curve represents a non-uniform soil with a wide range of particle sizes that are evenly distributed. The Fine Content varies between 15 to 99 %, 2.9 to 65 % and 2.00 to 67 %. The wide variation observed in this results might be due to the reason opined by various previous researchers that grain size distribution data are extremely varied and erratic [13]. The reason for this are enumerated in details by different researchers [14,15]. When soils are manipulated, their engineering properties vary a lot . Pre -testing drying causes variations in some parameters of soils and this behavior is always attache to the dehydration of the colloidal hydrated oxides occurring in these soils . Densification of a well-graded soil causes the smaller particles to move into the voids between the larger particles [16]. As the voids in the soil are reduced in the soil samples, the density and strength of the soil may increase. In contrast, poorly graded or uniform soils are composed of a narrow range of particle sizes. Akinola and Obasi , [16] reported that the type of clay that formed is not only a function of the nature of parent rock but also the intensity of weathering and the length of time during which it occurred. The fine-grained fractions dominate the composition of the studied soils, in comparison with coarse content fractions particularly for samples from southern and northern districts. Based on the Unified Soil Classification System [17] soils with percentage of fines that range between 15 -35% are said to be very clayey. In the present study, the fines fall above the specified range which is well pronounced in the southern and northern districts of the studied area.

Table 1. Soil index for Ekiti Senatorial district

Index properties	Division of the districts and ranges of values		
	Southern	**Northern**	**Central**
NMC (%)	1.4 to 29.47 %	1.9 to 34 %	1.1 to 33 %
GS	1.87 to 2.60,	1.9 to 2.58	2.17 to 2.90
LL (%)	32.5 -63	35 – 68	21- 55
PI (%)	8.5 -25	11 -28	7.5 -22
%Pass Sieve 200	15 to 55 %,	2.9 to 45 %	2.00 to 37 %.

Table 2. Summary of strength properties

Srength properties	Division of the districts and range s of values		
	Southern	Northern	Central
OMC (%)	11.6 – 31.6	5.06 -23.1	7.6 – 21.8
MDD (kg/m3)	1320 – 2080	1000 -2182	1940 – 2426
CBR soaked (%)	1.75 – 55	6.2 – 60 %	35-79
CBR unsoaked (%)	16- 69.5	14 – 75	45- 95

4.3 California Bearing Ratio (CBR)

The California Bearing Ratio (CBR) parameter help in the design of sub grade in the flexible pavement design. The results varied between 16- 89.5 % and 1.75 – 77 % for southern districts unsoaked and soaked CBR values while northern districts values varied between 14 – 95 % and 6.2 – 90 % for unsoaked and soaked CBR values and Central districts values varied between 9- 95 % and 5-69 % as presented in Table 2 the above analysis showed a high influence of soaking in the results obtained as the CBR values for the 24 hours-soaked samples were much lower compared to the unsoaked sample. A high reduction in CBR values after soaking indicates that the soil is very sensitive to changes in the moisture content. Hence, good drainage facilities are to be provided if these soils are to be used for any construction purpose that will mitigate loss of strength as suggested by Osuji and Akinwamide, [12]. The above analysis shows that materials within Southern and Central districts are quite suitable materials for Base, sub -base materials. Soils from Northern districts can be used as Sub base, Subgrade and earth fill material during construction work.

4.4 Measurement of Interrelationship among the Predictors

It is statistically assumed that there should be no any noticeable connections between the input variables when using multiple linear regressions in statistical analysis. Fig. 1 showed the interconnections that exist between the pairs of the independent variables. Gravel, Fines, Sand, Fines, LL and PL are highly correlated. This leads to multicollinearity issue. It may be erroneous if the model is predicted based on this dataset. Principal Component Analysis (PCA) was introduced to solve the multicollinearity problem as shown in fig. 2. There is no significant relationship among the predictors. This serves as a good foundation for multiple linear regression analysis.

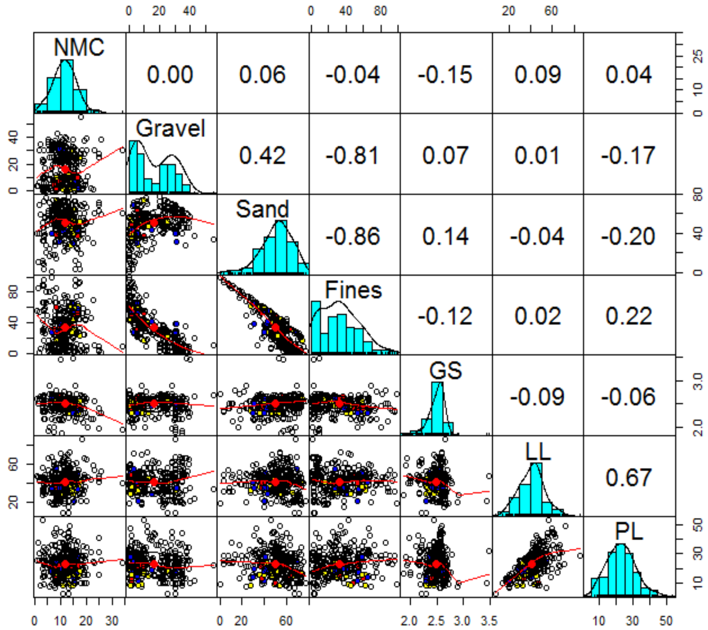

Fig. 1. Scatter matrix of interrelationship among the predictors

Table 3a. Eigen vectors from the PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
NMC	-0.0106	-0.1877	-0.7041	0.6618	-0.1735	0.0291	0.0022
Gravel	-0.4944	-0.1851	0.0097	-0.2370	-0.6940	0.0826	0.4202
Sand	-0.5243	-0.1413	-0.0184	0.1055	0.6744	0.0381	0.4872
Fines	0.6049	0.1916	0.0104	0.0630	-0.0483	-0.0704	0.7655
GS	-0.1361	0.1102	0.6698	0.6996	-0.1694	-0.0515	-0.0021
LL	0.1431	-0.6968	0.1244	-0.0292	0.0128	-0.6910	-0.0007
PL	0.2751	-0.6133	0.1991	0.0233	0.0471	0.7112	0.0000

Table 3b Importance of components: PC1 PC2 PC3 PC4 PC5 PC6 PC7Standard deviation 1.5960 1.2718 1.0581 0.9198 0.75088 0.54974 0.06340Proportion of Variance 0.3639 0.2311 0.1599 0.1208 0.08055 0.04317 0.00057Cumulative Proportion 0.3639 0.5949 0.7549 0.8757 0.95625 0.99943 1.00000

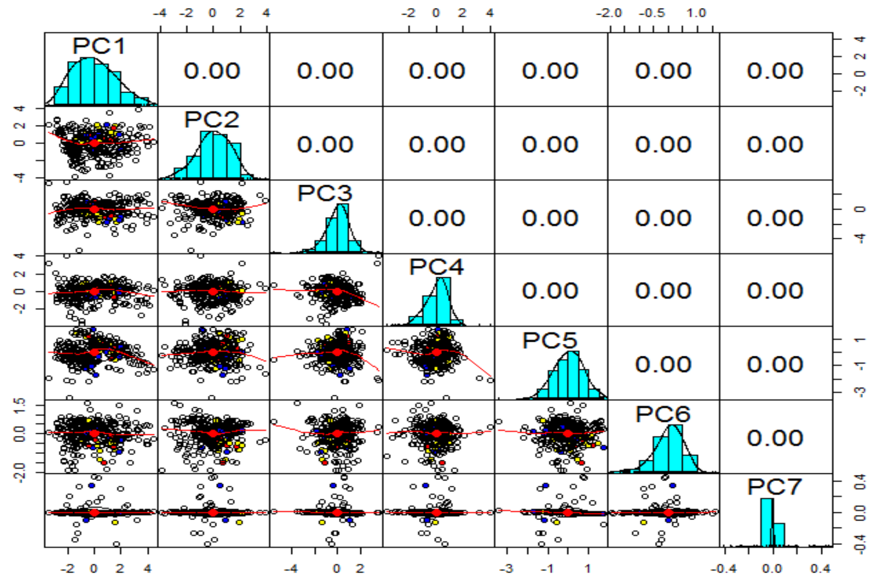

Fig. 2. Scatter matrix for no relationship among the predictors

4.4.1 Principal component analysis

These are the main major factor that combines with the main data. The maximum number of components extracted is usually the same with number of parameters. The eigenvectors, which are comprised of coefficients used to calculate the principal component scores. The coefficients showed the relative weight of each variable in the component. Principal Component Analysis is based on only independent variables. So we removed the eighth variable (dependent) from the dataset.

Table 3b showed the variability of the principal components PCs as 36%, 23%, 16%, and 12% for PC1, PC2, PC3 and PC4 respectively. The approximate sum of the first four principal components (PC) capture 88% of the variability, from the foregoing the first four components capture the majority of the variability, while the remaining components contribute negligible variability. In these results, the marks for the first four principal components can be estimated from the specified data using the coefficients listed under PC1 to PC4 as shown in Table 1a and 1b with figure 3 showing the screen plot and the proportion of variance for selecting the PCA.

5. RESULTS AND DISCUSSION

screeplot

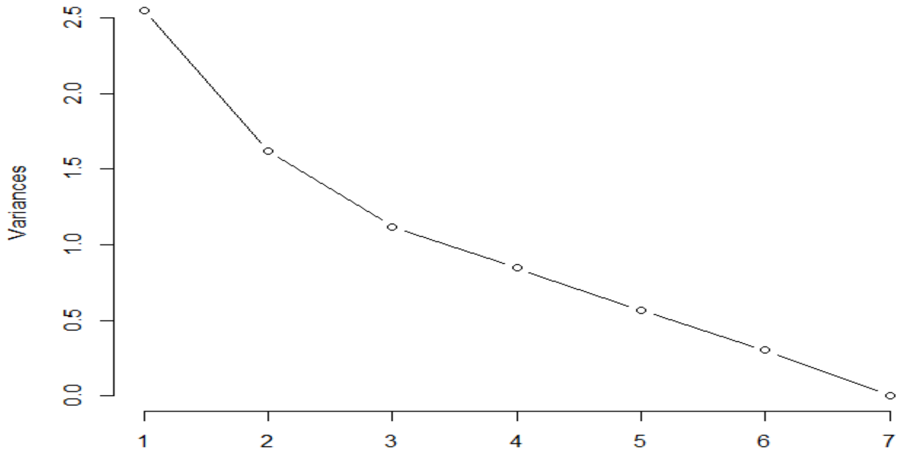

Fig. 3. Screen plot showing the proportion of variance for selecting the PCA

PCA biplot

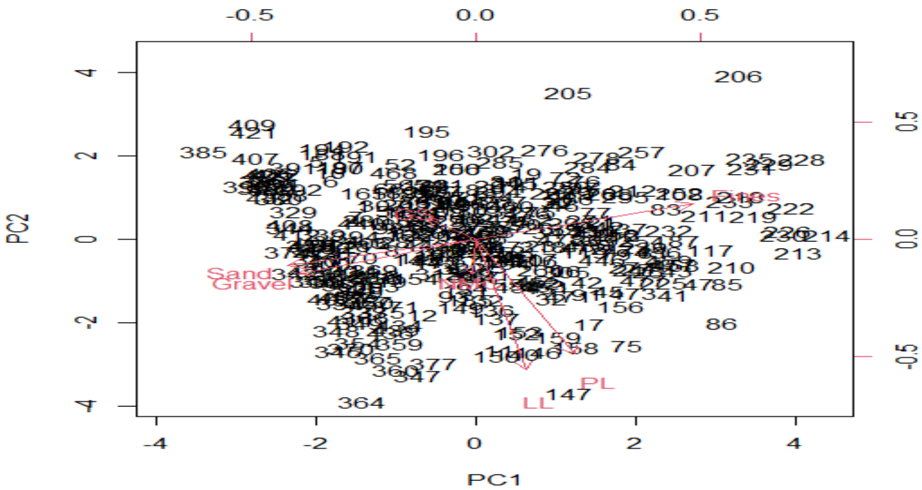

Fig. 4. Bi-plot of the component

5.1 Principal Components Analysis (PCA) Bi-plot

The Bi-plot of the components in Fig. 4 showed that Fine has a high positive relationship with the PC1 while PL and LL have high negative relationship with PC2.

5.2 The Derived Linear Model from PC1, PC2, PC3 and PC4

The derived Model is given below as theoretical and estimated model in equation (1) to (2) and equation (3) to (4) for SCBR and USCBR respectively, where the first four principal components were applied. The sum of four component score variables is representative and can be used in place of the seven original variables with a 12% loss of information.
The theoretical model for Soaked California Bearing Ratio (SCBR)

$$\text{SCBR} = \boldsymbol{\alpha} + \boldsymbol{\beta_1}(\text{PC1}) + \boldsymbol{\beta_2}(\text{PC2}) + \boldsymbol{\beta_3}(\text{PC3}) + \boldsymbol{\beta_4}(\text{PC4}) + \boldsymbol{\epsilon} \quad (1)$$

The estimated model with actual coefficients for Soaked California Bearing Ratio (SCBR)

$$\text{SCBR} = 35.14 - 7.5(\text{PC1}) - 3.88(\text{PC2}) + 0.13(\text{PC3}) - 6.67(\text{PC4}) \quad (2)$$

The theoretical model for UN-Soaked California Bearing Ratio (USCBR)

$$\text{USCBR} = \alpha + \beta_1(\text{PC1}) + \beta_2(\text{PC2}) + \beta_3(\text{PC3}) + \beta_4(\text{PC4}) + \epsilon \quad (3)$$

The estimated model with actual coefficients for UN- Soaked California Bearing Ratio

(USCBR USCBR = 57.99 - 5.29 {PC1}) - 1.68 {PC2}) + 2.06 {PC3}) - 2.46 {PC4}) (4)

5.3 Detailed Estimates and Model from 10 Hidden Layers(Neurons) from ANNs

Figs. 5 and 6 showed the results for Soaked CBR (SCBR) and Unsoaked CBR (UNSCBR) respectively, where the values can be read in the Artificial Neural Network plot, which also shows the coefficients of the inputs. They represent the weight of the inputs and their connections in the hidden layers, now we can use the network to make predictions, where the 30 % set aside from the dataset was used for result validation.

5.4 Measures of Accuracy among the Experimented and the Estimated Values (Goodness of fit)

The Correlation Coefficient and coefficient of determination R, R^2 and the Root Mean Square Error (RMSE) are the major yardsticks that are usually adopted to measure the performance of any prediction where the Correlation coefficient and coefficient of determination are the key function to establish a relative relationship between the expected and the observed data [18]. Smith, [19] prepared the following guide to measure R ─/R/≥0.8 Strong correlation, ─0.2< /R/ < 0.8 Correlation exists, /R/≤0.2 Weak correlation and /R/ = 0 No correlation.

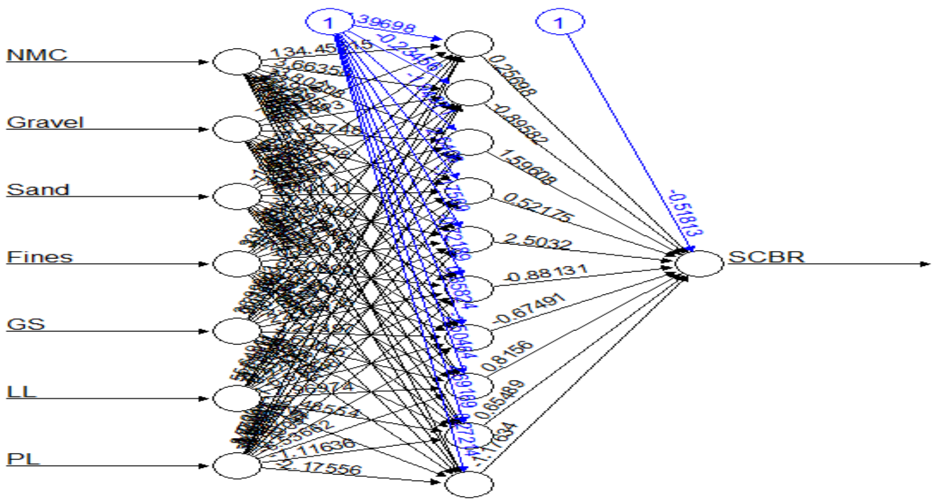

Fig. 5. Artificial Neural Networks Net plot for SCBR

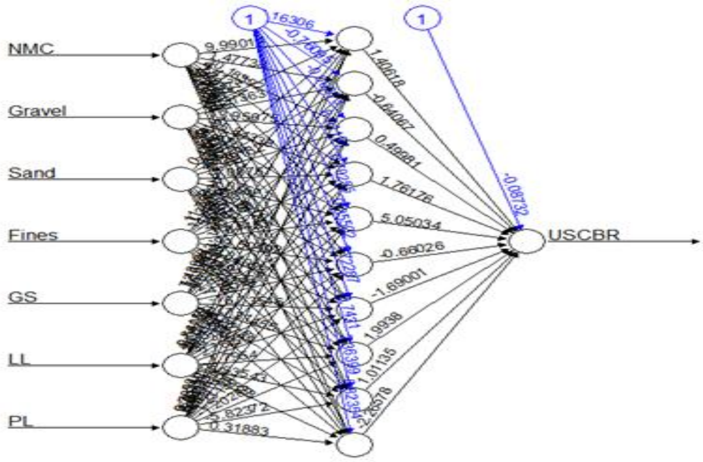

Fig. 6. Artificial Neural Networks Net plot for UN-SCBR

It is obvious from the values of RMSE 21.6, 21.23, 295.67, 7.03, 14.54 and 24.43,24.59,326.49,8.63,17.71 are from; MLR, ANN, MS Tree, RF, and SVM model for SCBR and USCBR values respectively. The least values 7.03 and 8.63 were observed from random forest (RF) for SCBR and USCBR. Similarly, the R values range between 0.1 – 0. 94 and 0.01 –0.92 as reflected in Table 2 to Table 5 and Figs. 7 and 8, which established the relationship among the predicted and the actual SCBR and USCBR using the five machine learning models. The correlation coefficient values deduced the Random Forest Model for SCBR and USCBR as the best, while the model having the least coefficient of determination R^2 is the MS tree model for both SCBR and USCBR respectively.

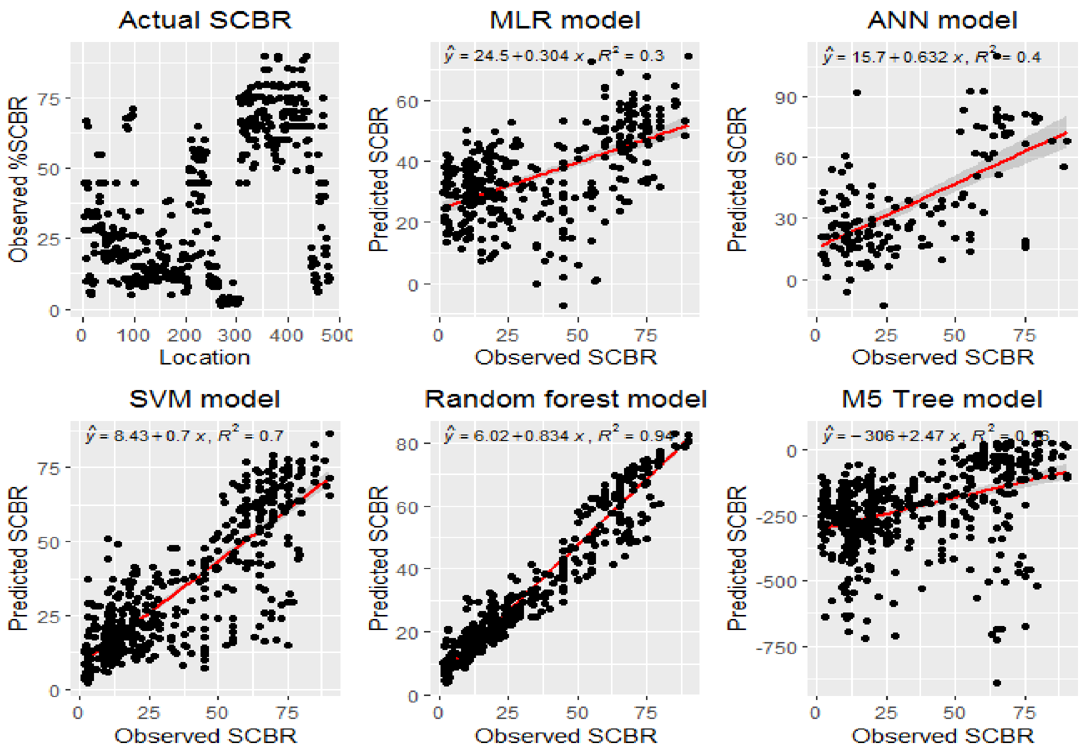

Fig. 7. Scatter plots for the predicting performance of the models in terms of coefficients of determination for Soaked California Bearing Ratio (SCBR)

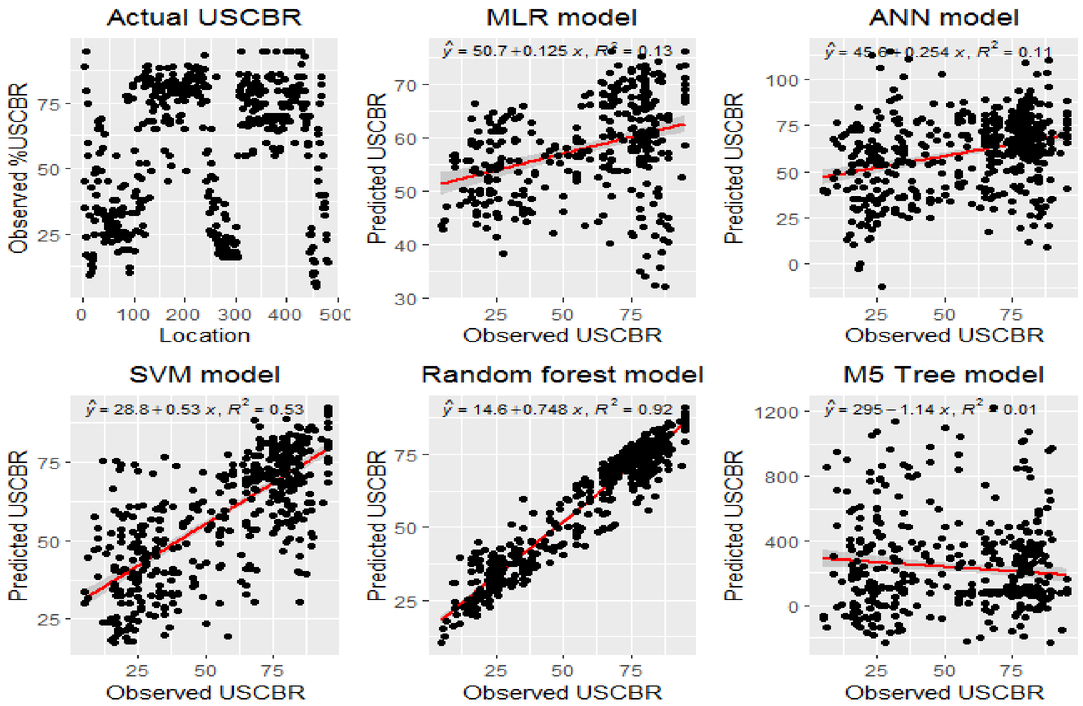

Fig. 8. Scatter plots for the predicting performance of the models in terms of coefficients of determination R^2 fo UN-r Soaked California Bearing Ratio (UN-SCBR)

Table 2. Measure of accuracy (Goodness of fit) for SCBR

Techniques	Soil Indices	ME	MAE	MSE	RMSE	R	R^2
MLR	SCBR	0	18.84	468.28	21.64	0.55	0.30
ANN	SCBR	0.40	16.37	450.70	21.23	0.65	0.43
MSTREE	SCBR	-254.13	2.58	87422.89	295.67	0.40	0.16
R F	SCBR	0.17	254.13	49.37	7.03	0.97	0.94
SVM	SCBR	-2.14	10.01	211.53	14.54	0.83	0.70

Table 3. Measure of accuracy (Goodness of fit) for UN- SCBR

Techniques	Soil Indices	ME	MAE	MSE	RMSE	R	R^2
MLR	UN-SCBR	0.00	20.52	596.68	24.43	0.35	0.13
ANN	UN-SCBR	1.97	18.56	604.73	24.59	0.47	0.22
MSTREE	UN-SCBR	171.24	210.21	106582.4	326.47	0.11	0.01
R F	UN-SCBR	0.07	6.69	74.40	8.63	0.96	0.92
SVM	UN-SCBR	1.58	12.82	313.73	17.71	0.73	0.53

Table 4. Predictions from the five Machine Learning (ML) models For SCBR

ACTUAL SCBR	Pred SCBR MLR	Pred SCBR SVM	Pred SCBR RF	Pred SCBR MS TREE
45	44.5563	43.6177	39.0386	-419.6562
33	44.6977	47.9322	41.3280	-178.3693
28	46.3653	39.2216	31.2479	-469.8580
10	41.8321	44.2962	30.7062	-273.5638
18	43.0993	36.7611	25.6720	-114.6589
67	52.7104	33.7116	50.7859	-102.0693

M5 pruned model tree:(using smoothed linear models)
Fines <= 10.625 :
| NMC <= 15.11 :
| | Fines <= 7.61 : LM1 (55/31.491%)
| | Fines > 7.61 : LM2 (24/24.397%)
| NMC > 15.11 :
| | NMC <= 18.36 : LM3 (18/18.979%)
| | NMC > 18.36 : LM4 (15/36.589%)
Fines > 10.625 :
| NMC <= 12.75 :
| | Sand <= 57.75 :
| | | PL <= 21.325 : LM5 (68/52.924%)
| | | PL > 21.325 :
| | | | NMC <= 10.16 :
| | | | | Gravel <= 20.35 :
| | | | | | PL <= 23.34 : LM6 (11/50.574%)
| | | | | | PL > 23.34 : LM7 (36/63.524%)
| | | | | Gravel > 20.35 : LM8 (13/67.947%)
| | | | NMC > 10.16 : LM9 (39/62.937%)
| | Sand > 57.75 :
| | | GS <= 2.525 : LM10 (18/50.421%)
| | | GS > 2.525 :
| | | | PL <= 14.89 : LM11 (11/29.674%)
| | | | PL > 14.89 :
| | | | | LL <= 41.55 :
| | | | | | Sand <= 70.75 :
| | | | | | | NMC <= 11.2 :
| | | | | | | | NMC <= 8.9 : LM12 (3/16.581%)
| | | | | | | | NMC > 8.9 : LM13 (5/14.22%)
| | | | | | | NMC > 11.2 : LM14 (6/18.214%)
| | | | | | Sand > 70.75 : LM15 (4/16.619%)
| | | | | LL > 41.55 : LM16 (10/27.787%)
| NMC > 12.75 :
| | Gravel <= 1.97 : LM17 (31/52.018%)
| | Gravel > 1.97 :
| | | Gravel <= 5.05 : LM18 (36/26.459%)
| | | Gravel > 5.05 :
| | | | PL <= 29.785 :
| | | | | Gravel <= 11.9 : LM19 (29/69.817%)
| | | | | Gravel > 11.9 :
| | | | | | NMC <= 15.25 : LM20 (17/14.987%)
| | | | | | NMC > 15.25 :
| | | | | | | GS <= 2.375 : LM21 (3/27.138%)
| | | | | | | GS > 2.375 : LM22 (10/40.515%)
| | | | PL > 29.785 : LM23 (18/18.752%)

Fig. 9. M5 pruned model tree using smoothed linear models for SCBR

Table 5. Predictions from the five Machine Learning (ML) models For UN- SCBR

ACTUAL UN-SCBR	Pred UN-SCBR MLR	Pred UN- SCBR SVM	Pred UN-SCBR RF	Pred UN- SCBR MS TREE
88.9	63.2929	62.1855	71.1080	51.7110
45.0	66.0538	65.9700	52.7521	78.6337
35.0	66.4530	76.6292	49.7509	191.3005
17.0	64.8213	75.5742	41.2342	98.5923
24.0	66.7558	74.3119	41.1535	137.7722
80.0	67.0159	77.7951	71.4316	235.7883

LM num: 1
SCBR = 15.3309+0.0456*Sand+ 0.0436*Fines- 2.312*GS- 0.6755*LL- 0.2307*PL
LM num: 2
SCBR = 7.8575+0.1023*Sand+ 0.0965*Fines+ 0.1216*GS- 0.6118*LL+ 0.0174*PL
LM num: 3
SCBR = 7.5291+0.1483*Sand+0.1048*Fines+0.1216*GS-0.6118*LL-0.1104*PL-0.1444*NMC
LM num: 4
SCBR = 8.4226+0.1065*Gravel+0.1549*Sand+0.1249*Fines+0.1694*GS- 0.6118*LL-0.2339*PL
LM num: 5
SCBR = 5.8789+0.1177*Sand+0.0955*Fines+0.1471*GS-0.6118*LL+ 0.0174*PL
LM num: 6
SCBR = 18.9483-0.253*NMC+0.1731*Gravel+0.1125*Sand+0.1568*Fines-1.4858*LL+0.0218*PL
LM num: 7
SCBR = 10.7033+0.057*Sand+ 0.028*Fines+ 0.0599*GS-0.6875*LL+ 0.0115*PL
LM num: 8
SCBR = 10.4864+0.0316*Sand+0.0098*Fines+0.04*GS-0.2032*LL+0.0052*PL
LM num: 9
SCBR = 10.4523+0.0109*Sand+0.0098*Fines+0.0311*GS-0.2032*LL+0.0052*PL
LM num: 10
SCBR = 17.2629+0.034*Gravel+0.0111*Sand+0.0099*Fines+0.019*GS-0.1918*LL+0.0054*PL
LM num: 11
SCBR = 5.5157+0.0725*NMC+0.0333*Gravel+0.195*Sand+0.0873*Fines+0.1162*GS-0.1918*LL+0.0054*PL
LM num: 12
SCBR = 11.4385+0.021*Gravel+0.0558*Sand+0.0426*Fines+0.06*GS-0.1918*LL+0.0054*PL
LM num: 13
SCBR = 76.7525-0.5466*NMC+0.2958*Gravel-0.0076*Sand-0.0206*Fines-0.1615*GS-14.4483*LL-0.0018*PL
LM num: 14
SCBR = 74.602+0.2958*Gravel-0.0076*Sand-0.0901*Fines-0.1615*GS-14.4483*LL-0.0018*PL
LM num: 15
SCBR = 70.06+0.29*Gravel-0.08*Sand-0.02*Fines-0.16*GS-14.43*LL-0.0018*PL
LM num: 16
SCBR = 69.0281+0.2958*Gravel-0.0076*Sand-0.0206*Fines-0.1615*GS-14.49*LL-0.002*PL
LM num: 17
SCBR =- 15.55+1.001*NMC+0.43*Gravel-0.14*Sand+0.09*Fines+0.12*GS-2.8694*LL+0.68*PL
LM num: 18
SCBR = 13.19+0.76*NMC+0.09*Gravel+0.04*Sand+0.02*Fines-0.04*GS-8.0083*LL+0.23*PL
LM num: 19
SCBR = 141.09+0.09*Gravel-0.14*Sand-0.31*Fines-26.68*GS-19.17*LL+0.45*PL
LM num: 20
SCBR = 69.6+0.31*NMC+0.091*Gravel-0.142*Sand-0.039*Fines-2.83*GS-18.96*LL-0.01*PL
LM num: 21
SCBR = 132.39-0.2*NMC+0.09*Gravel-0.14*Sand-0.12*Fines-22.21*GS-18.96*LL-0.01*PL
LM num: 22
SCBR = 119.5-0.06*NMC+0.09*Gravel-0.14*Sand-0.10*Fines-18.78*GS-18.95*LL-0.01*PL
LM num: 23
SCBR = 57.19+0.01*Gravel-0.107*Sand+0.02*Fines-0.138*GS-14.98*LL-0.01*PL

Fig. 10. Some Linear model provided by MS model tree to predict output data from the dataset for SCBR

5.5 Prediction from MS TREE

Figs. 9 to 14 can easily be used to predict the SCBR and USCBR if the dataset is within the range of independent variable adopted to produce these models. The figure 11 and 14 show the splitting at the nodes, the values at the nodes are the standard deviations. The splitting continues until a reasonable low standard deviation is noticed. After assessing all the possible splits, M5 chooses the one that maximizes the expected error reduction (Taghi Sattari et al., 2010). Division in M5 discontinued when the class variables of all the instances that reach a node vary just slightly, or only a few instances remain.

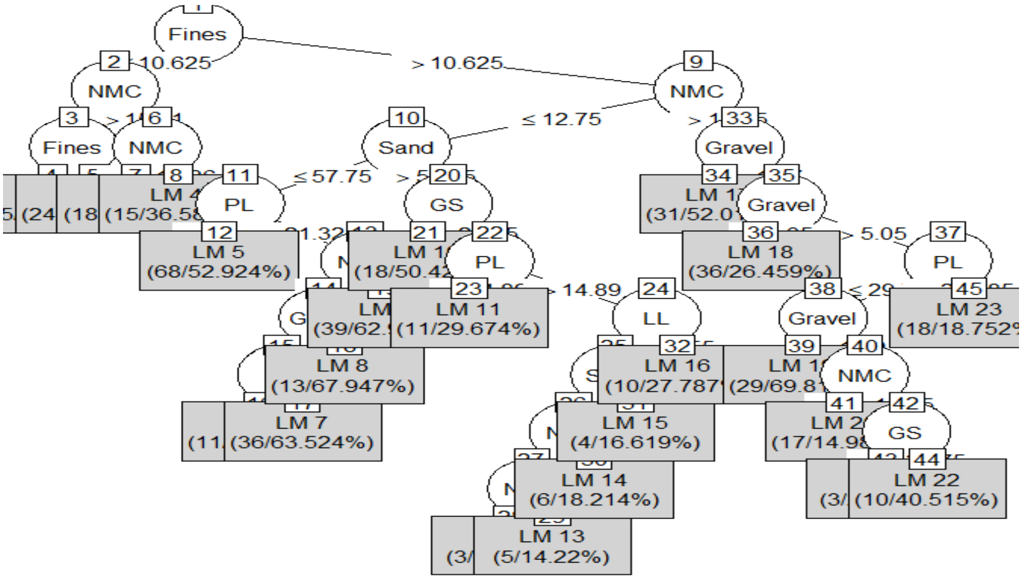

Fig. 11. Tree with regression models as leaves for SCBR

```
Fines <= 13.5 :
| NMC <= 12.98 :
| | NMC <= 9.355 : LM1 (40/26.312%)
| | NMC > 9.355 :
| | | LL <= 49.45 : LM2 (21/50.735%)
| | | LL > 49.45 : LM3 (13/22.892%)
| NMC > 12.98 : LM4 (45/31.89%)
Fines > 13.5 :
| Fines <= 37.9 :
| | PL <= 25.915 :
| | | LL <= 28.05 : LM5 (26/75.539%)
| | | LL > 28.05 :
| | | | NMC <= 8.51 :
| | | | | PL <= 20.3 : LM6 (15/17.729%)
| | | | | PL > 20.3 : LM7 (9/74.831%)
| | | | NMC > 8.51 :
| | | | | LL <= 40.85 : LM8 (46/86.824%)
| | | | | LL > 40.85 : LM9 (19/65.83%)
| | PL > 25.915 : LM10 (59/69.424%)
| Fines > 37.9 :
| | Fines <= 67.16 :
| | | NMC <= 9.165 : LM11 (31/96.199%)
| | | NMC > 9.165 :
| | | | Gravel <= 3.05 : LM12 (41/51.536%)
| | | | Gravel > 3.05 : LM13 (75/90.863%)
| | Fines > 67.16 :
| | | Sand <= 22.9 : LM14 (26/41.311%)
| | | Sand > 22.9 : LM15 (14/70.502%)
```

Fig. 12. Some Linear model provided by MS model tree to predict output data from the dataset for USCBR

LM1

UN-SCBR = 90.19+0.0918*Gravel-0.1848*Fines-0.3582*GS+0.0652*LL-0.0331*PL

LM 2

UN-SCBR = 100.813 + 0.0918*Gravel- 0.2866*Fines- 0.66*GS- 0.0331*PL

LM3

UN-SCBR = 102.926+0.0919*Gravel-0.5494*Fines-0.7411*GS-0.0331*PL

LM4

UN-SCBR = 44.172+0.394*Gravel+0.3734*Sand -0.143*Fines -0.2709*GS -0.0331*PL

LM5

UN-SCBR = -73.34 +0.074*Gravel+0.9*Sand +0.89*Fines +0.75*GS+21.28*LL -0.56*PL

LM6

UN-SCBR =-137.40+0.39*Gravel +1.24*Sand+0.98*Fines+0.92*GS+23.24*LL-0.04*PL

LM7

UN-SCBR = -148.07+0.8*Gravel + 1.07*Sand +0.9*GS+23.23*LL+0.09*PL

LM8

UN-SCBR = -103.79-0.87*NMC+0.07*Gravel+0.9*Sand+1.3162Fines+0.63*GS+18.6*LL-0.5*PL

LM9

UN-SCBR = -21.17- 1.18*NMC+0.07*Gravel+0.6*Sand+0.75*Fines+0.6*GS+17.7*LL-1.6*PL

LM10

UN-SCBR = -54.0+0.129*Gravel+0.333*Sand+29.91*GS+9.02*LL+1.27*PL

LM11

UN-SCBR = 81.149 -1.815*NMC -0.343*Gravel -0.285*Sand -0.245GS +2.07*LL -0.23*PL

LM12

UN-SCBR = 30.22 +0.155*Gravel -0.099*Sand -0.155*Fines 0.09*GS+2.08*LL+0.326*PL

LM13

UN-SCBR = 43.75 -0.155*Gravel -0.099*Sand -0.088*GS+2.077*LL -0.155*PL

LM14

UN-SCBR = 124.216 -0.158*Gravel-0.947*Sand-0.69*Fines+0.079*GS+2.077*LL-0.176*PL

LM15

UN-SCBR = 156.12-1.3*NMC-0.15*Gravel-1.3*Sand-0.93*Fines+0.079*GS+2.079*GLL-0.177*PL

Fig. 13. Some Linear model provided by MS model tree to predict output data from the dataset for USCBR

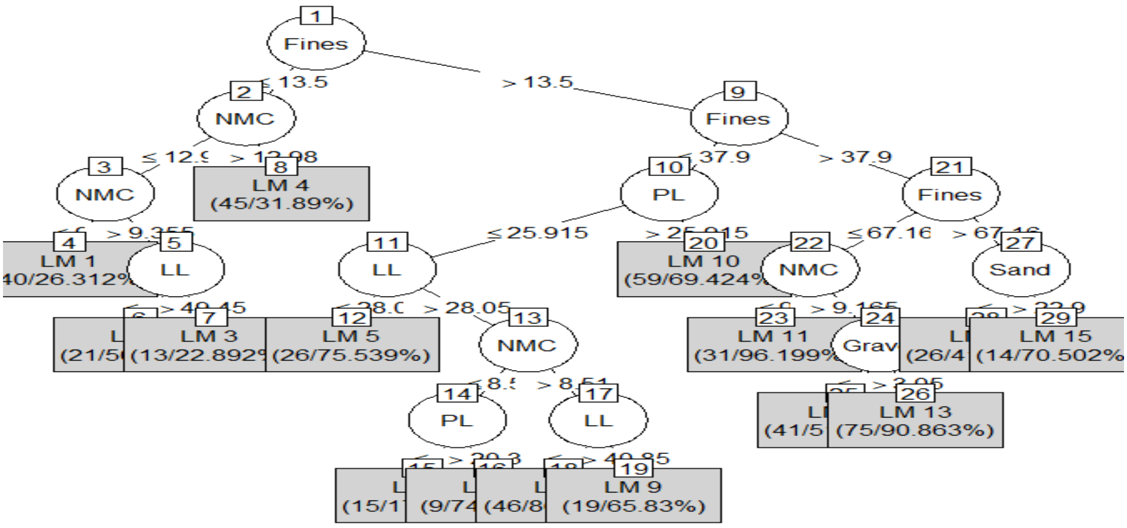

Fig. 14. Tree with regression models as leaves USCBR

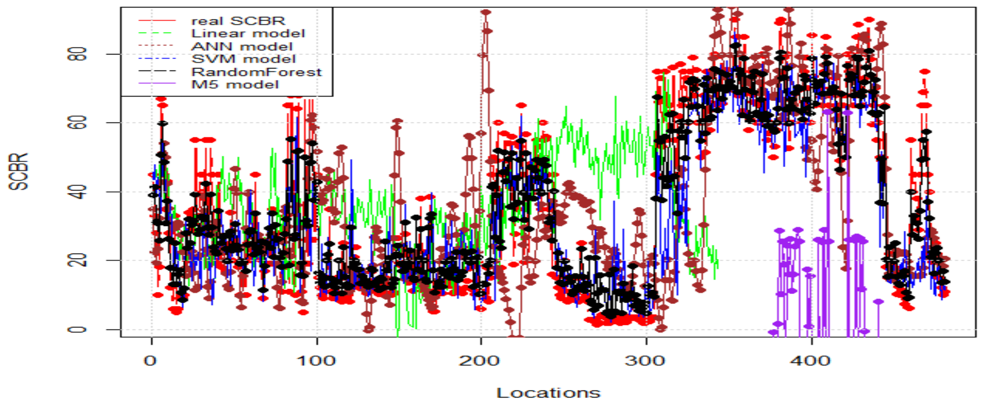

Fig. 15. Line plot showing the movement of the observed and the predicted

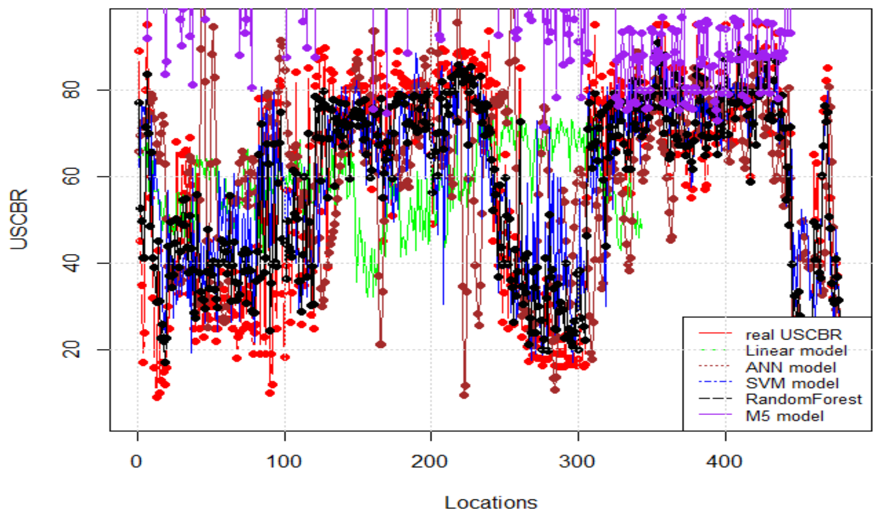

Fig. 16. Line plot showing the movement of the observed and the predicted

Figs. 15 and 16 above showed the predicted values generated by random forest model seems to move side by side with the actual SCBR and USCBR, this suggests a good model and the best among the five applied. while the MS Tree gave a worst performance as shown in figure 15 and 16 and table 2, 3 ,4 and table 5 respectively, where the coefficient of determination R^2 gave 0.94and 0.92 for SCBR and USCBR respectively. From the foregoing the results suggest a good model and of course the best among the five applied. for Random Forest (RF) while MS Tree gave the worst model.

6. CONCLUSIONS

The developed model in the present work relates SCBR and USCBR with some soil physical properties. The results have shown that machine learning techniques has an excellent contribution in the field of geotechnical engineering. From the foregoing support vector machine (SVM) performed better than the MLR. ANNs while M5 tree model exhibits steps of jumped phenomenon in the predicted values of the response variable. However, it is noteworthy that Random Forest came out as the best machine learning techniques for the estimation of SCBR and USCBR in this research work using the correlation and the performance metrics. The results reveal a high correlation coefficient R and could judiciously be used for estimating SCBR and USCBR of a regional soil and gives a very good estimate of SCBR and USCBR without actually performing the test.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Venkatasubramanian C, Dhinakaran G. Ann model for predicting cbr from indexproperties of soils. International Journal of Civil and Structural Engineering. 2011;2(2).
2. Srinivasa R, Ruchita Pankaj S, Sathiraju VS. Prediction of California Bearing Ratio through Empirical Correlations of Index Properties for Tropical Indian Soils. International Journal of Innovations in Engineering and Technology. 2019;15(1).
3. Datta T, Chottopadhyay BC. Correlation Between CBR And Index Properties Of Soil, Proceedings of Indian Geotechnical Conference, Kochi (Paper No. A-350); 2011.
4. Ayodele AL, Falade FA, Ogedengbe MO. Effect of fines content on some engineering properties of lateritic soil in Ile-Ife, Department of Civil Engineering, Obafemi Awolowo University, Ile-Ife; 2009.
5. Dharamveer S, Reddy KS, Laxmikant Y. Moisture and Compaction Based Statistical Model for Estimating CBR of Fine-Grained Subgrade Soils, International Journal of Earth Sciences and Engineering. 2011;04(06):100-103. ISSN 0974-5904,
6. Rahim, AM. Subgrade soil index properties to estimate resilient modulus for pavement design International Journal of Pavement Engineering. 2005;6(3):163-169.
7. Vinod P, Reena C. Prediction of CBR value of lateritic soils using liquid limit and gradation characteristics data. Highw Res J IRC. 2008;1(1):89–98.
8. Patel RS, Desai MD. CBR predicted by index properties for alluvial soils of South Gujarat. In: Proceedings of the Indian geotechnical conference, Mumbai. 2010;79–82.
9. Magdi ME. Zumrawi. Prediction of CBR Value from Index Properties of Cohesive Soils, Annual Conference of Postgraduate Studies and Scientific Research, Hall, Khartoum, Proceeding. 2012;1:111-117.
10. Valentine YK. Souleyman MoupeMfoyet. Bertille Manefouet. Armand Sylvain Ludovic Wouatong . Lawrence Aleh; 2017.
11. AASHTO. “Standard Specifications for Transportation, Material and Method of Abu-Kiefa M. A. 1998; 1986.
12. Osuji OS, Akinwamide JT. Physico-Chemical Properties of Lateritic Soils in Ado-Ekiti, South Western Nigeria. Universal Journal of Environmental Research and Technology. 2018;7(1):10-18
13. Winterkorn HF, Chandrashekharan EC. Laterite soils and their stabilization. Highway Research Board Bull. 1951;44:10-29.
14. Terzaghi K. Design and performance of Sasumua Dam. Inst. Civil Engineers Proc. 9:369-395.67. Topping, J. 1957. Errors of observation and their treatment; 1958.
15. Gidigasu MD. The importance of soil genesis in the engineering classification of Ghana soils. Eng. Geol. 1971;5:117-161.
16. Akinola OO, Obasi RA. Compositional Characteristics and Industrial Potential of The Lateritic Clay Deposit In Ara-IjeroEkiti Areas, Southwestern Nigeria. International Journal of Scientific & Technology Research. 2014;3(8):305 -31.
17. ASTM D. Standard Practice for Description and Identification of soils (Visual -Manual Procedure),’ American Society for Testing and Materials; 2000.
18. Shahin MA, Jaksa MB, Maier HR. State of the art of artificial neural networks in geotechnical engineering, Electronic Journal of Geotechnical Engineering. 2008;8:1-26.
19. Smith GN. Probability and statistics in civil engineering: an introduction. London: Collins; 1986.

Peer-review history:
The peer review history for this paper can be accessed here:
https://www.sdiarticle5.com/review-history/86861