APPLICATION OF SOFT COMPUTING TECHNIQUES IN MODELLING OF SOAKED AND UNSOAKED CALIFORNIA BEARING RATIO

ABSTRACT

In this study, efforts are made to explore some advanced mathematical techniques such as the Support Vector Machine (SVM), Random Forest (RF), M5 tree, multiple linear regression and Artificial Neural Network to establish a correlation between soil physical parameters and California Bearing Ratio of lateritic soils. Four hundred and eighty (480) soil samples were obtained and divided into data set using training and validation of the developed models from the basic soil parameters namely, Liquid Limit (LL) Plastic limit (PL), Natural moisture content (NMC), Specific gravity (GS), Fines (F),Gravel and Sand. In order to reduce the large dimension of the dataset, the Principal Component Analysis (PCA) was implemented and the approximate sum of the first four principal components (PC) Capture 88% of the variability, in the response variable with 12 % loss of information. The performance test carried out on the actual and predicted values from the models using RMSE and coefficient of determination R^2, 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 ranges between 0.1 – 0.94 and 0.01 ─0.92 which established the relationship among the predicted and the actual SCBR and USCBR. 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. It can therefore be concluded that the best model of Soaked and Unsoaked CBR based on the dataset came from Random Forest, while MS tree gave the worst model. From the fore going its imply that the predicted soil parameters values are within permissible accuracy and the model is a useful tool in estimating the subsurface indices of a civil engineering site at the preliminary planning stage before final structural design for the sub structures.

Keywords: Compaction characteristics, Soaked, Unsoaked, California Bearing Ratio, Highway.

1.0 INTRODUCTION

Highways are the main medium of vehicular movement in Nigeria today. They carry about 90 % of the nation's passenger traffic and 80 percent of its freight. Sound physical road network in the city and rural areas is important for social and economic growth. Lane strength is small and most federal highways are double lanes or single (Venkatasubramanian. C & Dhinakaran. G 2011).75 % of all Nigeria road are congested. Most roads are of less quality and road maintenance is poorly funded only about 25 % maintenance needs are met. This has leads to the deteriorating condition of highway and high transport fees for users. Roads are the essential civil engineering infrastructural facilities give link and accessibilities to other social life of a man in a healthy society. Roadways are developed based on soil strength properties, vehicular and pedestrian load required carrying in their lifetime (Srinivasa, et al., 2019). To measure some major geotechnical indices of a particular region, will require huge sampling along the proposed route, which consume time and huge financial resources. The soil strength properties California Bearing Ratio (CBR) for both soaked and unsoaked CBR is the most popular applied geotechnical property for measurement of overlay thickness of flexible pavements in Nigeria. Highway engineers are faced with challenges in evaluating the CBR of the soil while estimating the thickness of Sub-Base and Base –Course layers and during preliminary study or comprehensive project detailed report and collection of huge CBR data is redistricted by time and budget resources. Under such conditions, CBR data for any proposed civil engineering infrastructure could be determined through the established models between CBR and soil index properties as they proffer efficient and economical solutions. CBR values are usually needed to proffer geotechnical solutions to highway structures particularly during planning and design. Resources cost and time normally posts a threat for such sub soil evaluation in mapping the variation in their values along the alignment (Vasu et al., 2017).

2.0 LITERATURE REVIEW

In recent times several researchers have carried out experimental studies to correlate CBR and various soil physical parameters. Ayodele (2009), carried out laboratory investigation on some soils in the South western state and found out that >90% relationship exists between CBR and geotechnical indices. Statistical formulae were developed by Dharamveer, 2011. Using five various soils with 100 samples and established a sound relationship between the experimented and estimated CBR values. Relationship was developed between MR and soil physical parameters of cohesive soils and cohesion less (sandy) soil using triaxle laboratory tests to estimate subgrade soil Moduli (Rahim, 2005). Multiple Linear Regression (MLR) on CBR with respect to Plasticity Index (PI), MDD, index properties, and OMC were established (Vinod P, 2008). The CBR values are inversely proportional to the Plasticity Index (PI). If PI values increase CBR values decreases (Patel, 2010). It was established that CBR and frictional angle can be determined from fine grained soils for varied soil physical properties (Datta, 2011; Magdi 2012). 33 soil samples of expansive soil data collected from the on-going road construction site, it was established that Single Linear Regression Analysis (SLRA) and MLRA on CBR and geotechnical parameters It was affirmed that correlation exists among CBR values and soil physical parameters (Valentine, 2017). In the present work five machine learning tools was selected to predict soaked and unsoaked CBR from their physical soil indices. The present study aims to develop machine learning models to predict the subsoil properties of non-visited locations using Multiple Regression, RF, ANN, SVM and MS TREE using R and R Studio software and to compare the models accuracy in prediction by calculating coefficient of determination (R^2), MAE and RMSE of the models in Ekiti – State senatorial districts.

3.0 MATERIALS AND METHODOLOGY

3.1 Datasets and Data Analysis

The 480 dataset applied in this research was collected from established borrowed pits within Ekiti State Senatorial District Zones (ESSDZ) South western Nigeria. The laboratory test was performed at the Department of civil engineering t material testing unit, federal polytechnic Ado Ekiti, Ekiti- State Nigeria. R version 4.0.5 and R studio version 1.2.5033 was used for this research work (R Core Team 202). In order to reduce the large dimension of the dataset, the Principal Component Analysis (PCA) was implemented.

3.2 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.

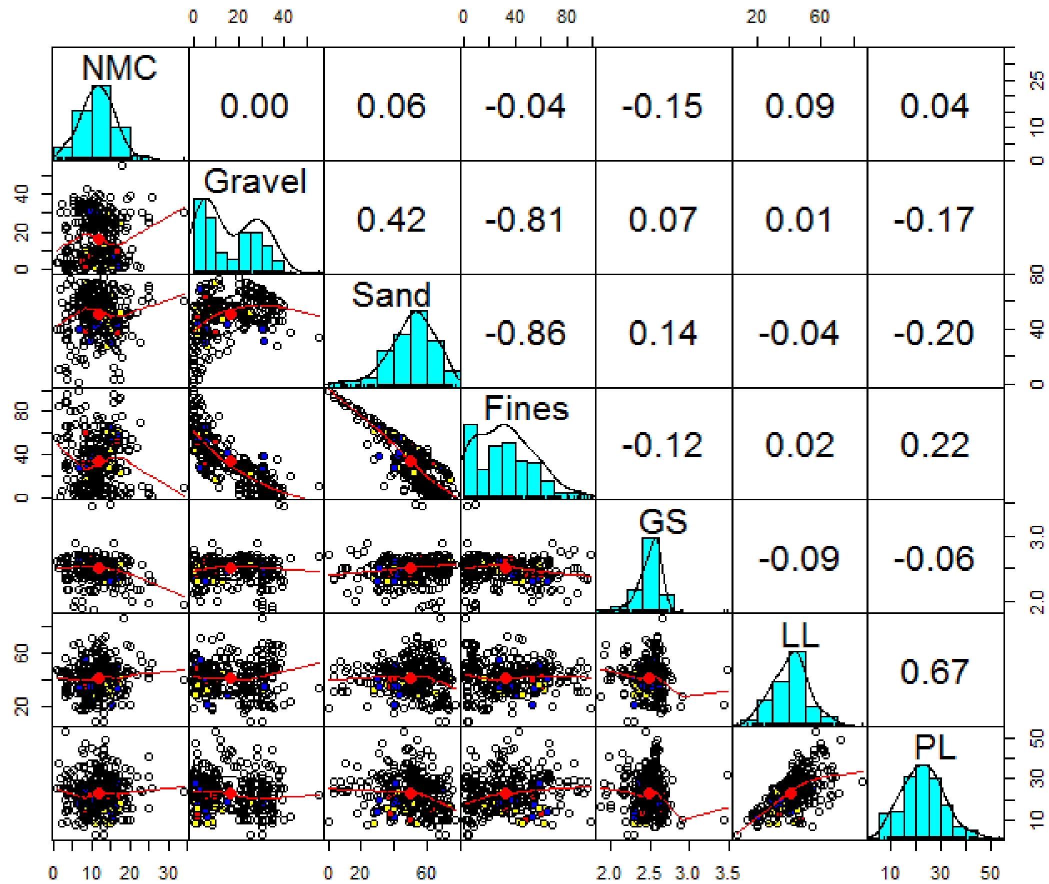

Figure 1: Scatter matrix of interrelationship among the predictors

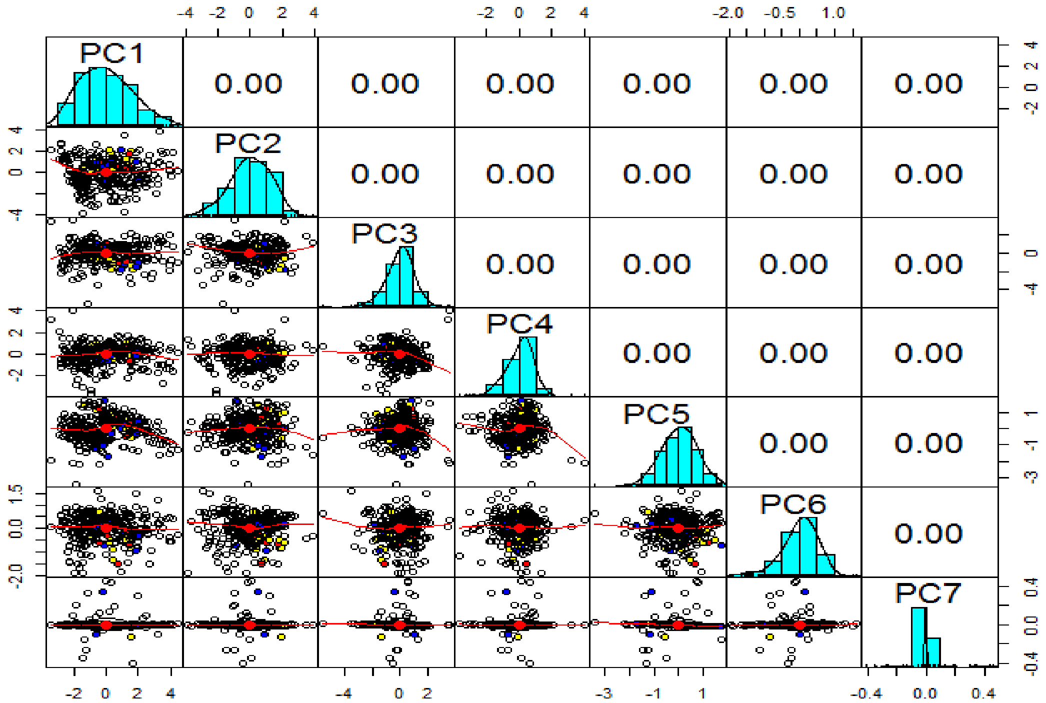

Figure 2: Scatter matrix for no relationship among the predictors

3.2.1 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 1a: 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 1b Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.5960	1.2718	1.0581	0.9198	0.75088	0.54974	0.06340
Proportion of Variance	0.3639	0.2311	0.1599	0.1208	0.08055	0.04317	0.00057
Cumulative Proportion	0.3639	0.5949	0.7549	0.8757	0.95625	0.99943	1.00000

Table 1c 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.

Results and discussion

screeplot

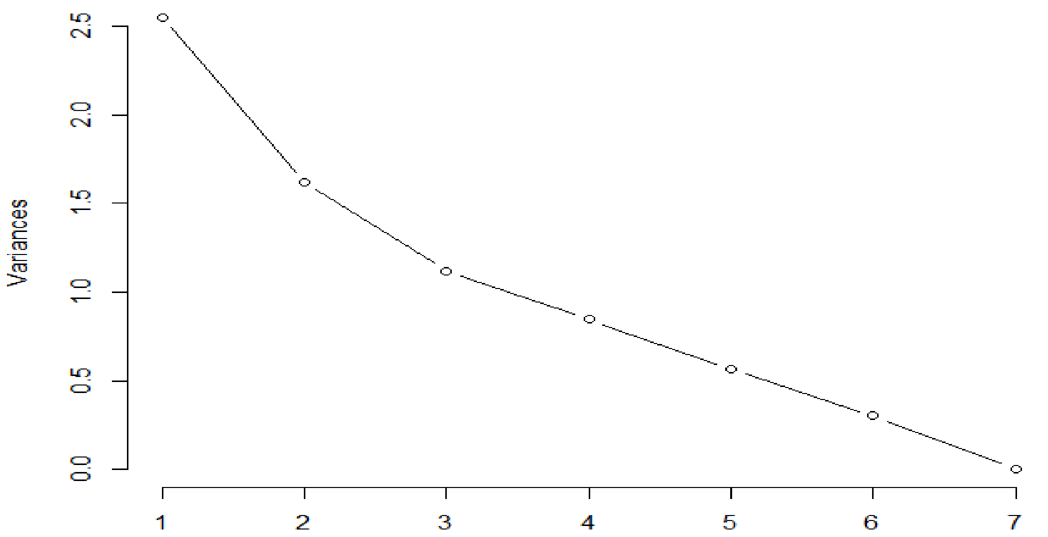

Figure 3: Screen plot showing the proportion of variance for selecting the PCA

PCA biplot

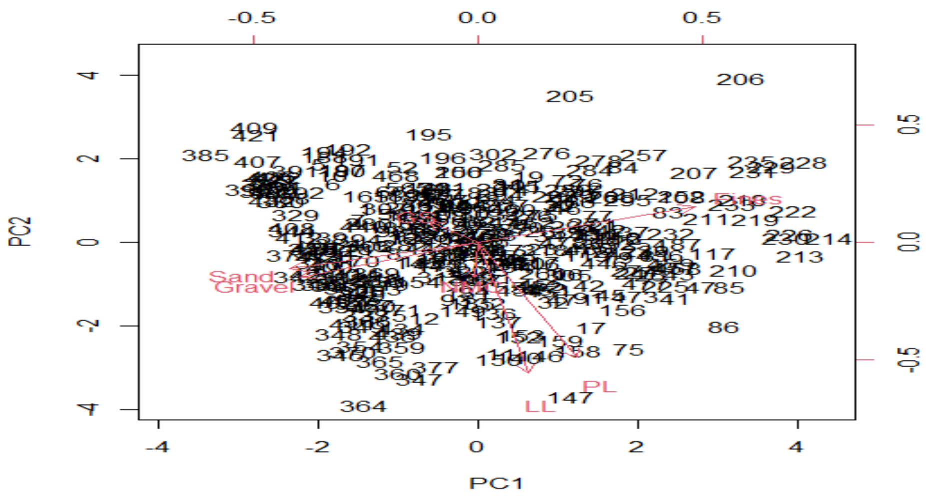

Figure 4: Bi-plot of the component

3.2.2 Principal Components Analysis (PCA) Bi-plot

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

3.2.3 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)

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

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

SCBR = 35.14 - 7.5(PC1) - 3.88(PC2) + 0.13(PC3) - 6.67(PC4) ……………………… (2)

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

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

3.2.4 Detailed estimates and model from 10 hidden layers(neurons) from ANNs

Figure 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.

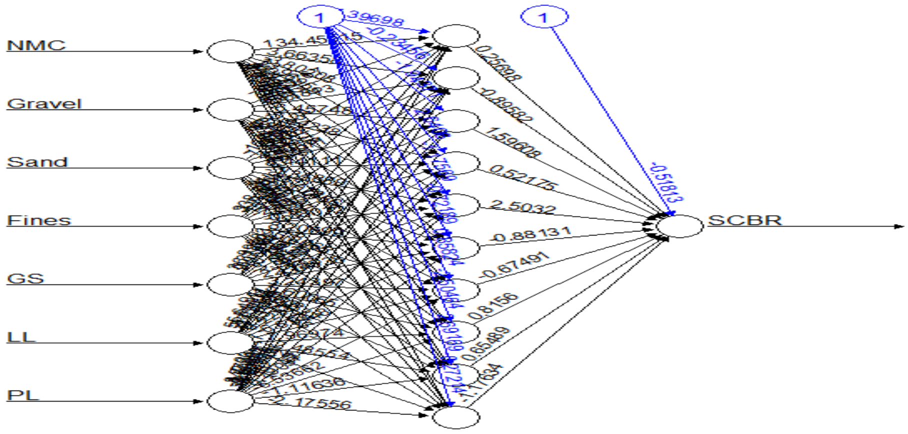

Figure 5: Artificial Neural Networks Net plot for SCBR

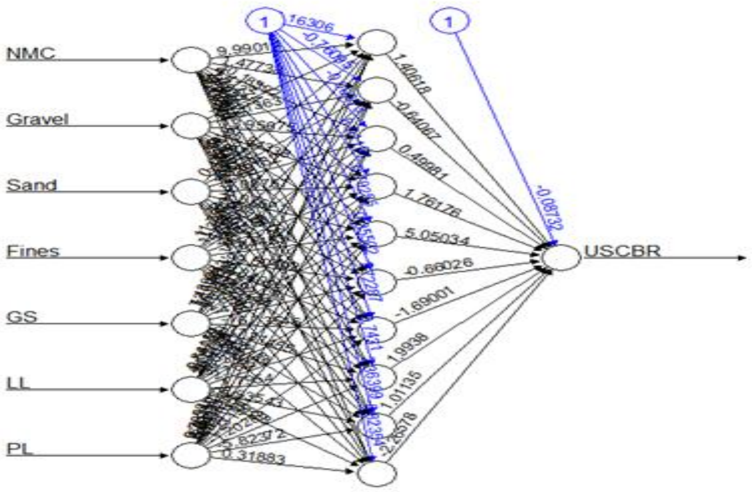

Figure 6: Artificial Neural Networks Net plot for UN-SCBR

3.3Measures 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 (Shahin et al., 2008). (Smith, 1986) 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.

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 figure 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.

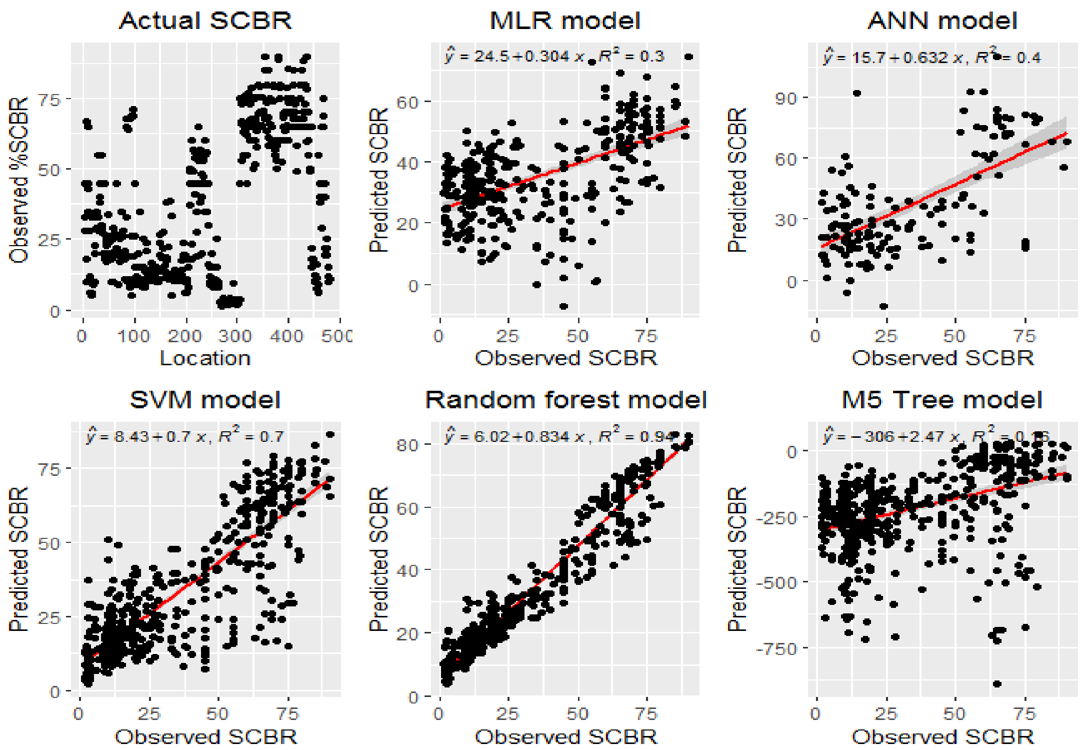

Figure 7: Scatter plots for the predicting performance of the models in terms of coefficients of determination for Soaked California Bearing Ratio (SCBR)

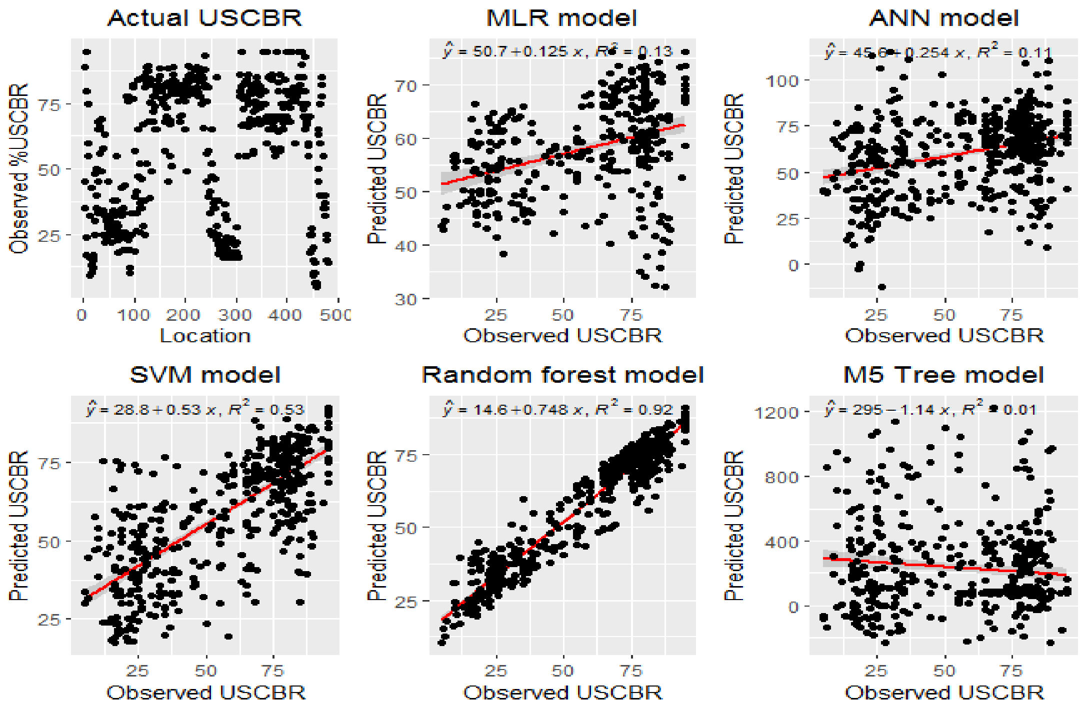

Figure 8: Scatter plots for the predicting performance of the models in terms of coefficients of determination R^2 fo UN-r Soaked California Bearing Ratio (UN-SCBR)

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

Techniques	Soil	Goodness of fit					
	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	Goodness of fit					
	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

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

```
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%)
```

Figure 9: M5 pruned model tree using smoothed linear models for SCBR

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

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

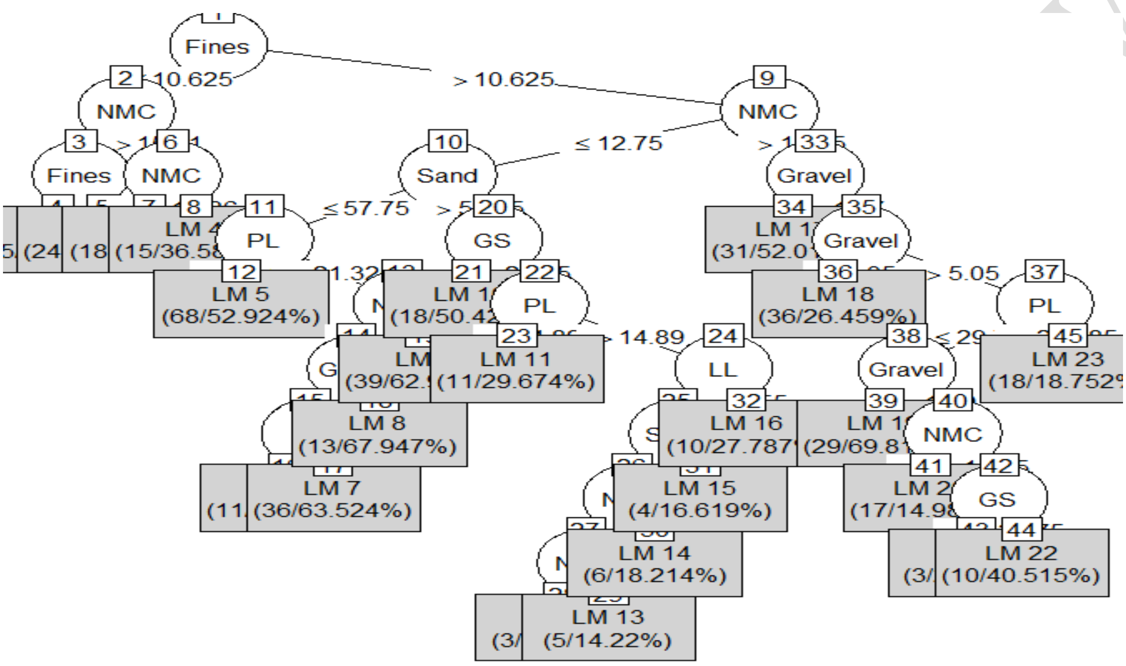

Figure 11: Tree with regression models as leaves for SCBR

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

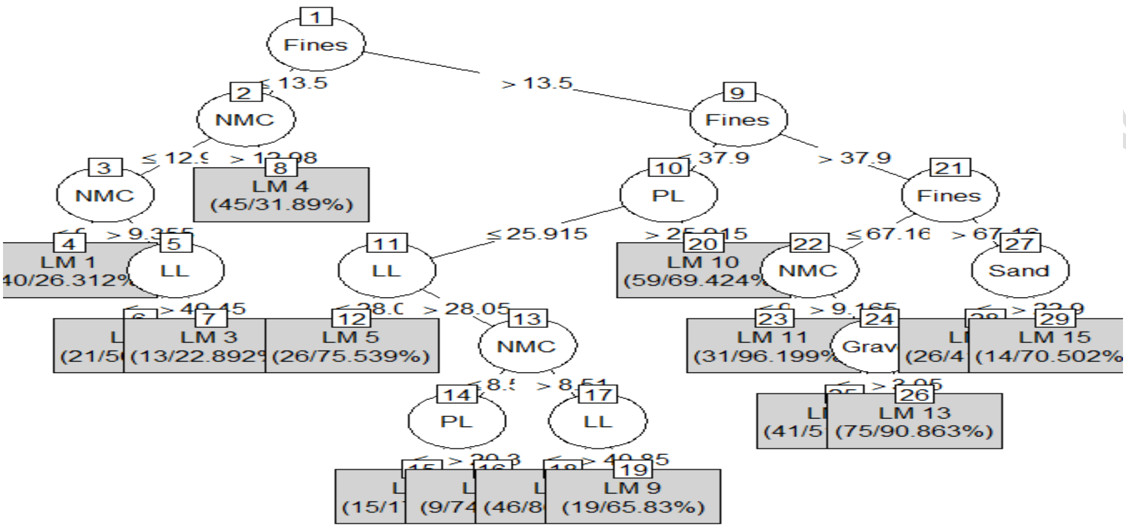

Figure 14: Tree with regression models as leaves USCBR

3.4 Prediction from MS TREE

Figure 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.

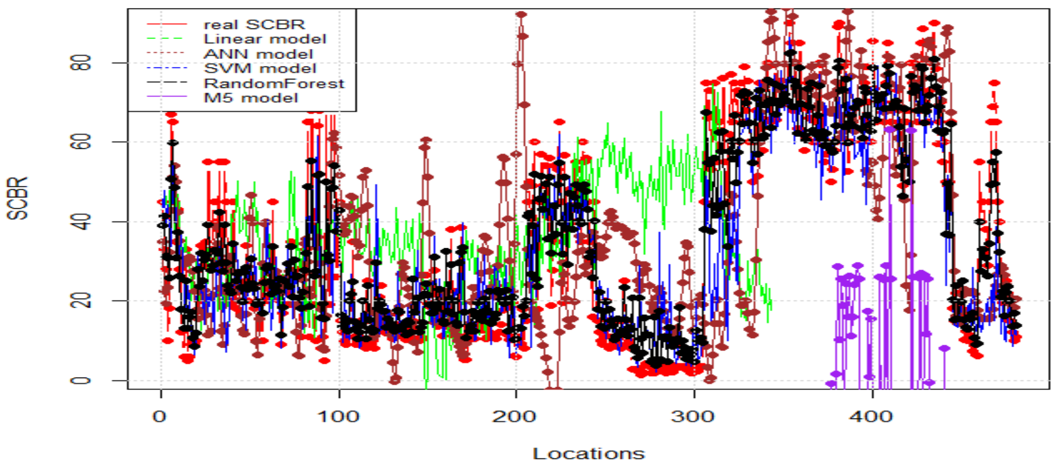

Figure 15: Line plot showing the movement of the observed and the predicted

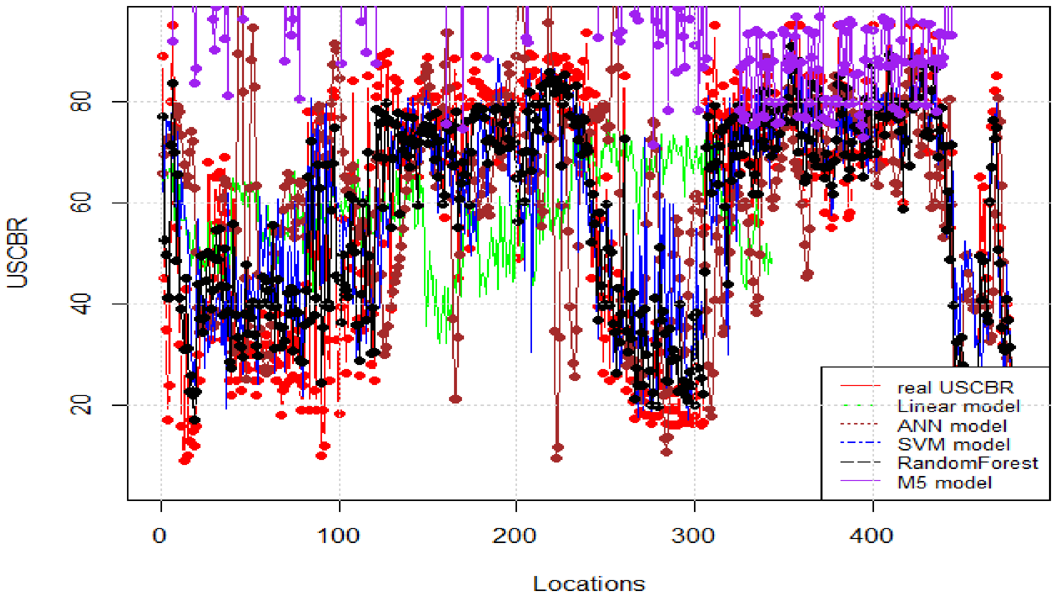

Figure 16: Line plot showing the movement of the observed and the predicted

Figure 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.

4.0 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.

REFERENCES

Ayodele, A.L., Falade, F.A. & Ogedengbe, M.O. (2009). Effect of fines content on some engineering properties of lateritic soil in Ile- Ife, Department of Civil Engineering, Obafemi Awolowo University, Ile-Ife.

Bezeng., Correlation of California Bearing Ratio (CBR) Value with Soil Properties of Road Subgrade Soil, Geotech Geol Eng https://doi.org/10.1007/s10706-018-0604-x

Datta T., Chottopadhyay B.C., (2011) Correlation Between CBR And Index Properties Of Soil, Proceedings of Indian Geotechnical Conference, Kochi (Paper No. A-350).

Dharamveer, S., Reddy, K. S., & Laxmikant, Y. (2011). Moisture and Compaction Based Statistical Model for Estimating CBR of Fine-Grained Subgrade Soils, International Journal of Earth Sciences and Engineering ISSN 0974-5904, Volume 04, No 06 SPL, pp100-103.

Magdi, M. E. Zumrawi. (2012), Prediction of CBR Value from Index Properties of Cohesive Soils, Annual Conference of Postgraduate Studies and Scientific Research, Hall, Khartoum, Proceeding Vol. 1 pp. 111 -117.

Patel, R.S, & Desai M.D (2010). CBR predicted by index properties for alluvial soils of South Gujarat. In: Proceedings of the Indian geotechnical conference, Mumbai, pp 79–82.

Quinlan, J.R. (1992). Learning with continuous classes. In: Proc. AI'92 (Fifth Australian Joint Conf. on Artificial Intelligence) (ed. by A. Adams & L. Sterling), Pp 343- 348. World Scientific, Singapore.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R project.org

Rahim, A.M . (2005)., Subgrade soil index properties to estimate resilient modulus for pavement design International Journal of Pavement Engineering Volume 6, 2005 -Issue 3., pp163-169.

Srinivasa, R., Ruchita., Pankaj, S., & Sathiraju, V. S. (2019). Prediction of California Bearing Ratio through Empirical Correlations of Index Properties for Tropical Indian Soils. International Journal of Innovations in Engineering and Technology, Volume 15 Issue 1 December 2019.

Uyanik, G. K. (2013). A study on multiple linear regression analysis. Social and Behavioural sciences 106 Pp 234-240.

Valentine, Y.K. (2017) . Souleyman MoupeMfoyet . Bertille Manefouet. Armand Sylvain Ludovic Wouatong . Lawrence Aleh.

Vapnik, V. (1998). " Statistical Learning Theory", Wiley, New York.

Nagamalli, V., & Kumar S.P. (2017). Study of ANN Program for Prediction of California Bearing Ratio of Fine-Grained Soils. International Journal of Science and Research (IJSR) ISSN: 2319-7064.

Venkatasubramanian, C., & Dhinakaran, G. (2011). Ann model for predicting cbr from indexproperties of soils International journal of civil and structural engineering Volume 2, No 2.

Vinod, P., & Reena, C. (2008) Prediction of CBR value of lateritic soils using liquid limit and gradation characteristics data. Highw Res J IRC 1(1) Pp:89–98.