

# Application of Machine Learning Techniques in Modelling of Soaked and Unsoaked California Bearing Ratio

**Akinwamide Joshua Tunbosun**

Department of Civil Engineering, Faculty of Engineering  
The Federal Polytechnic  
Ado Ekiti, Ekiti State, Nigeria  
akinwamidejoshua2@gmail.com

**Jacob Odeh Ehiorobo, Ph.D. Osuji SylvesterObinna. Ph. D. and Ebuka Nwankwo. Ph.D.**

Department of Civil Engineering, Faculty of Engineering  
University of Benin  
Benin City, Edo State, Nigeria  
jacchi@uniben.edu, osujisyl@yahoo.com, ebuka.nwankwo@uniben.edu

## Abstract

Machine learning techniques such as the Support Vector Machine (SVM), Random Forest (RF), M5 tree, Multiple linear regression and Artificial Neural Network was adopted to correlate soil physical parameters and California Bearing Ratio (CBR) of soils for Soaked (SCBR) and Unsoaked (USCBR). Four hundred and eighty (480) soil samples were obtained and divided into data set using training and validation of the developed models from some basic soil parameters. Principal Component Analysis (PCA) was implemented to reduce the large dimension of the data set and the actual and predicted values from the models using Root Mean Square Error (RMSE) and coefficient of determination  $R^2$ , it showed RMSE as 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 showed 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.

**Keywords:** Modelling, Machine Learning, Soaked California Bearing Ratio, Unsoaked California Bearing Ratio, Highway

## 1. 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, and Dhinakaran 2011) 75 % of all Nigeria road are congested. Most roads are of less quality and road maintenance are poorly funded only about 25 % maintenance needs are met. This has led to the deteriorating condition of highway and high transport fees for users. Roads are the essential civil engineering infrastructural facilities which 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 to carry in their lifetime (Srinivasa Reddy 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 restricted 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 post a threat for such sub soil evaluation in mapping the variation in their values along the alignment (Vasude et al. 2017). 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 cohesionless (sandy) soil using triaxle laboratory tests to estimate subgrade soil Moduli (Rahim2005). Multiple Linear Regression (MLR) on CBR with respect to Plasticity Index (PI), Maximum Dry Density (MDD), index properties, and Optimum Moisture Content (OMC) were established. (Vinod, 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. Materials and Methods

#### 3.1 Datasets and data analysis

The 480 dataset applied in this research were 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 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 regression 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.

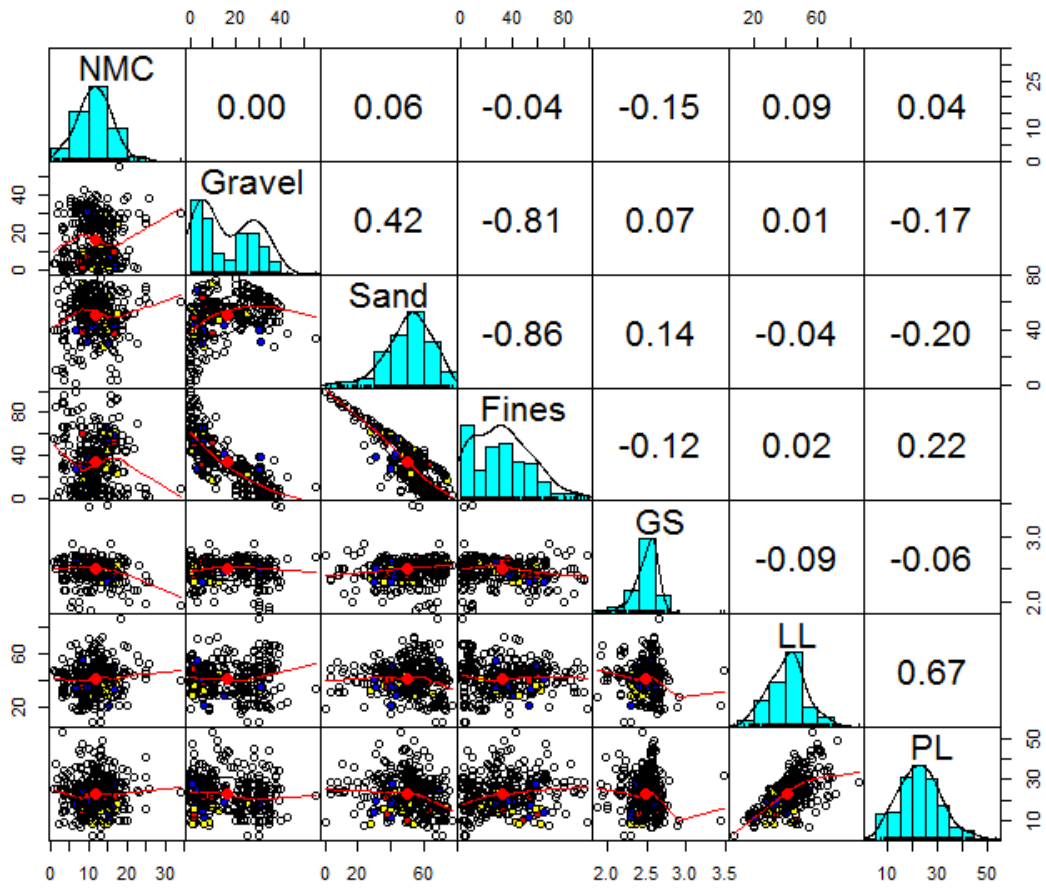


Fig. 1: Scatter matrix of interrelationship among the predictors

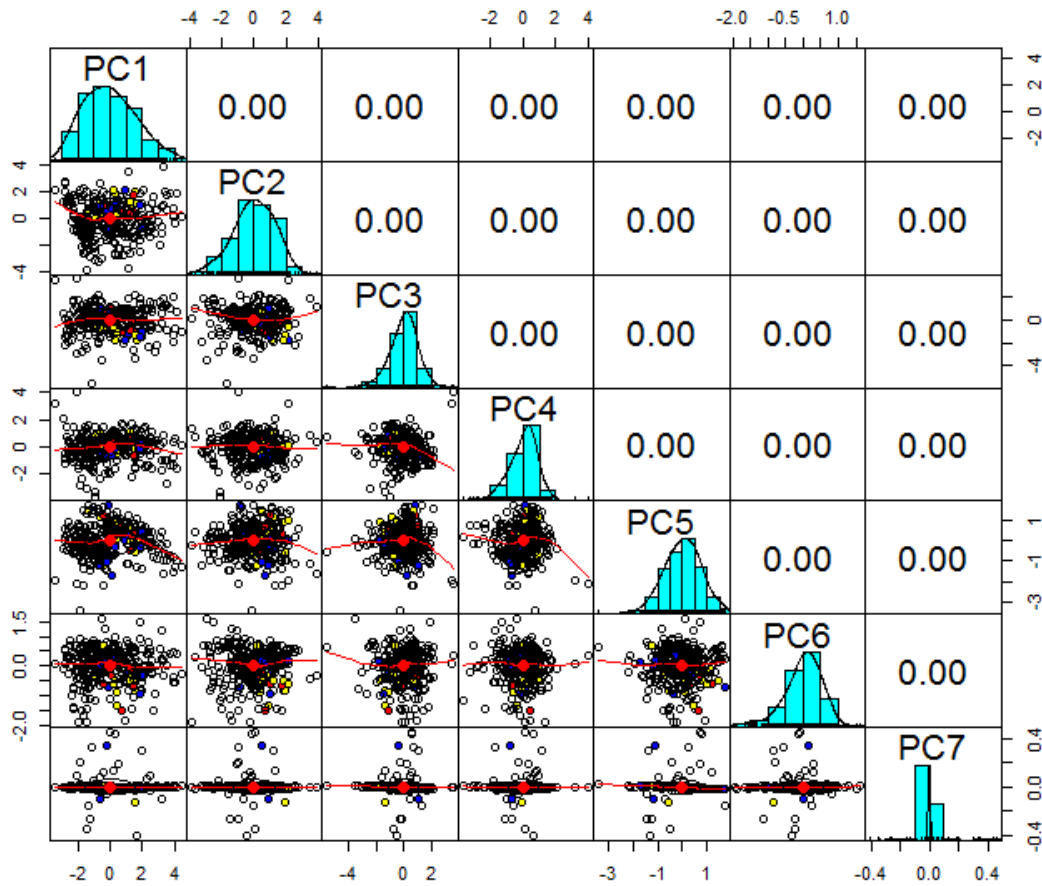


Fig. 2: Scatter matrix for no relationship among the predictors

### 3.2.1 Principal Component Analysis

These are the main major factor that combine with the main data. The maximum number of components extracted I was 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. The eight variable was removed (dependent) from the dataset.

Table 1b: Eigen vectors from the PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
NMC	-	-	-	0.6618	-	0.0291	0.0022
Gravel	0.0106	0.1877	0.7041	-	0.1735	-	-
Sand	-	-	0.0097	-	-	0.0826	0.4202
Fines	0.4944	0.1851	-	0.2370	0.6940	-	-
GS	-	-	-	0.1055	0.6744	0.0381	0.4872
	0.5243	0.1413	0.0184	0.0630	-	-	0.7655
	0.6049	0.1916	0.0104	-	0.0483	0.0704	-
	-	0.1102	0.6698	0.6996	-	-	-
	0.1361	-	-	-	0.1694	0.0515	0.0021

LL	0.1431	-	0.1244	-	0.0128	-	-
		0.6968		0.0292		0.6910	0.0007
PL	0.2751	-	0.1991	0.0233	0.0471	0.7112	0.0000
		0.6133					

Table 1 importance of component

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	1.596	1.2718	1.0581	0.919	0.75974	0.54974	0.0634
Proportion of variance	0.3639	0.2311	0.1559	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

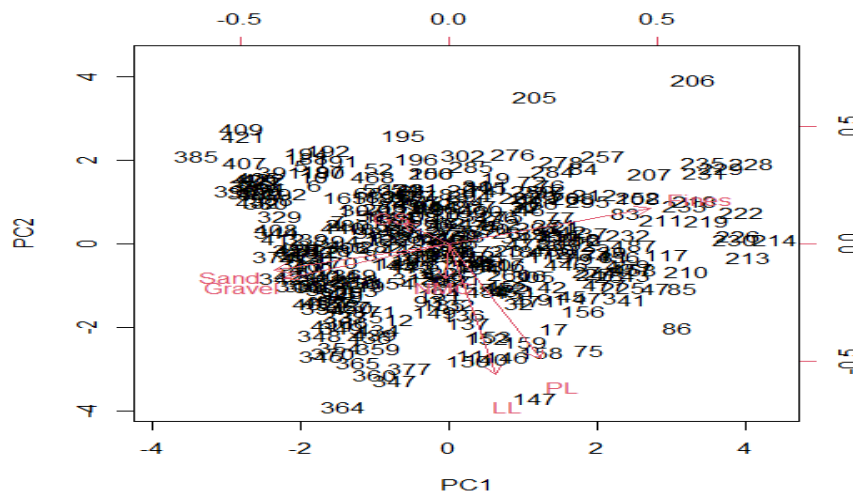


Fig. 3: Bi-plot of the component

### 3.2.2 Principal Components Analysis (PCA) Bi-plot

Figure 3 showed the loading plot which is used to identify which variables have the largest effect on each component. Loadings can range from -1 to 1. Loadings close to -1 or 1 indicate that the variable has strongly influenced the component while Loadings close to 0 indicate that the variable has a weak influence on the component. Evaluating the loadings can also help to characterize each component in terms of the variables. In this case, Fine has a high positive relationship with the PC1 while PL and LL also have high negative relationship with PC2 as shown in fig 3 and models equations .1 to 4 showing general Predicting models with Principal Components.

### 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 equations (1) to (2) and equations (3) to (4) for SCBR and USCBR respectively, where the first four principal components were applied. The sum of four component score variables are representative and could be used in place of the seven original variables with a 12% loss of information.

The theoretical model for Soaked California Bearing Ratio (SCBR)

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

The estimated model with actual coefficients for Soaked California Bearing Ratio (SCBR)0

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

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

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

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

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

### 3.3 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 (Shahin et al, 2008). (Smith, 1986) prepared the following guide to measure  $R$   $-|R| \geq 0.8$  Strong correlation,  $-0.2 < |R| < 0.8$  Correlation exists,  $|R| \leq 0.2$  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 tables 2 to tables 5 and figures 5 and 6, 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.

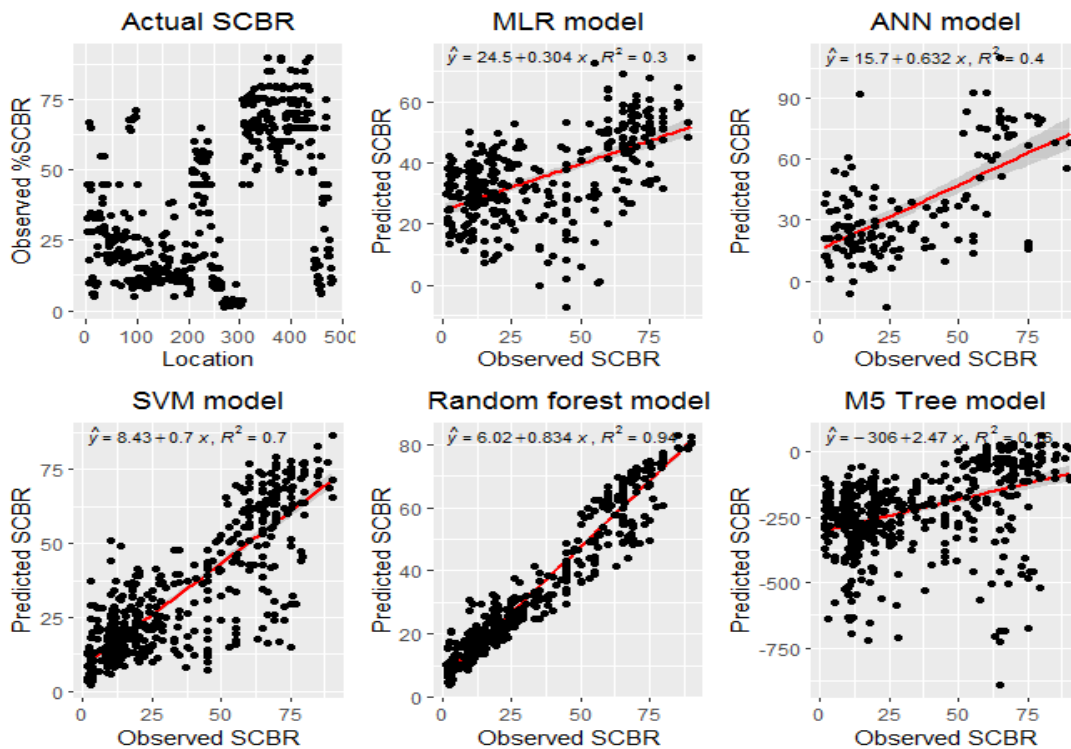


Fig 5: Scatter plots for the predicting performance of the models in terms of coefficients of determination for Soaked California Bearing Ratio (SCBR)

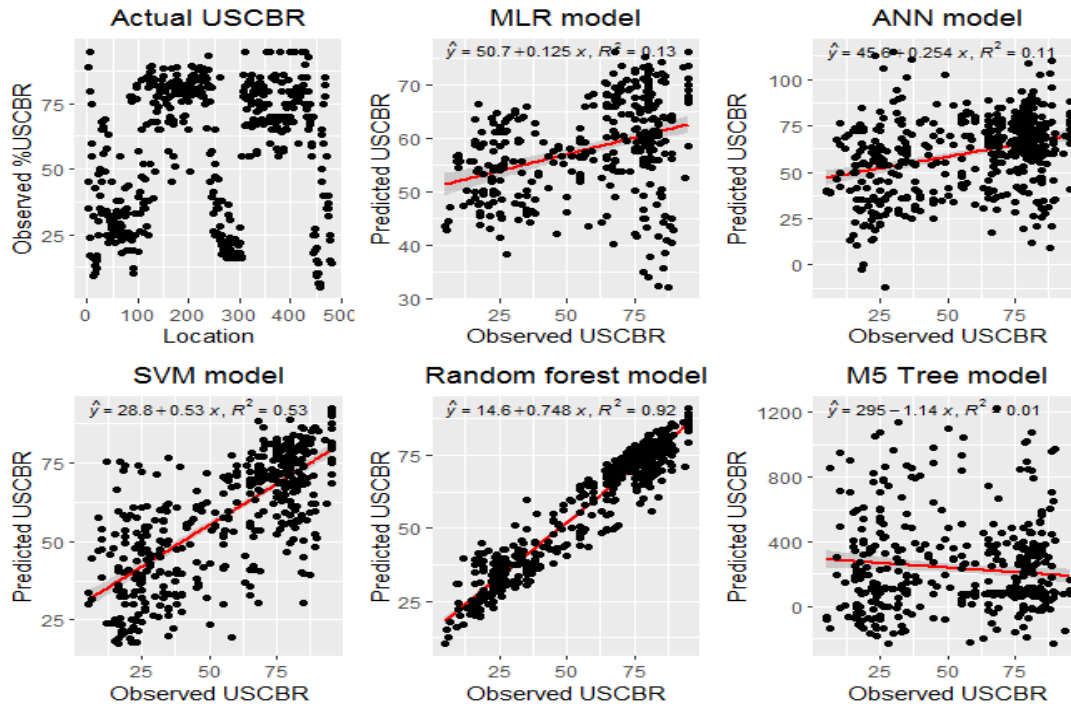


Fig 6: Scatter plots for the predicting performance of the models in terms of coefficients of determination  $R^2$  for UN-r Soaked California Bearing Ratio (UN-SCBR)

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

Techniques	Soil Indices	Goodness of fit					
		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	Goodness of fit					
		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

Table4 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

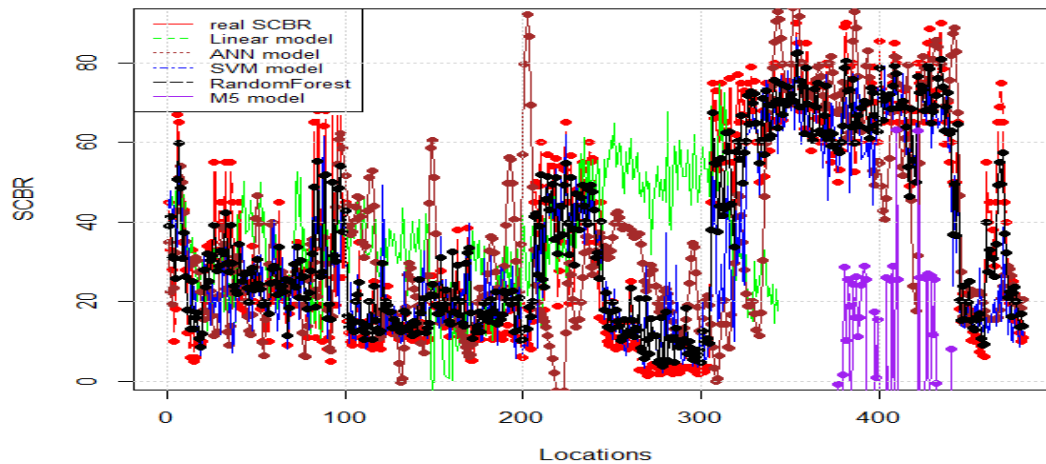


Fig 7: Line plot showing the movement of the observed and the predicted



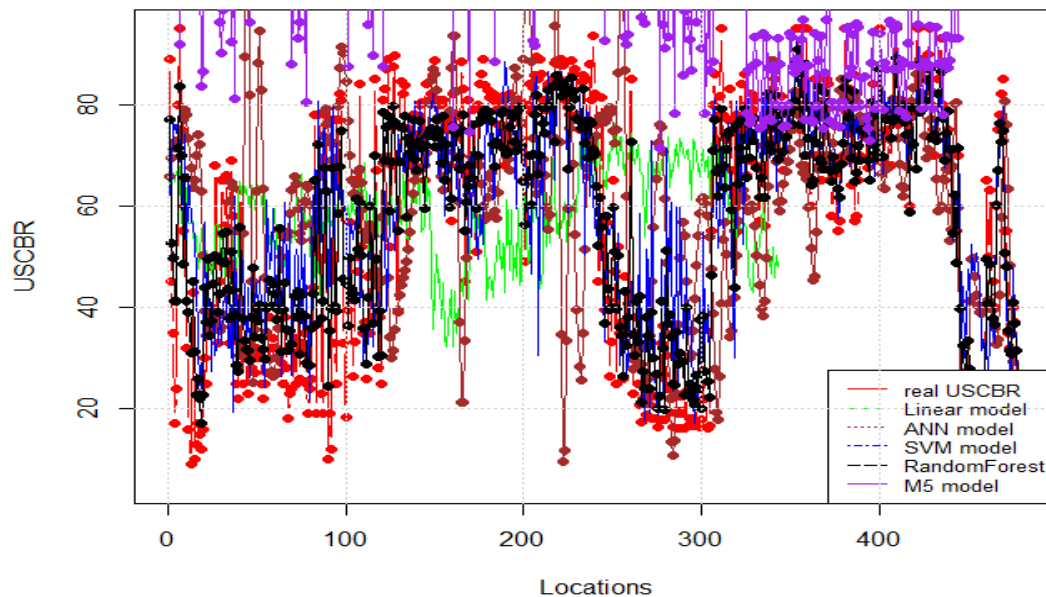


Fig 8: Line plot showing the movement of the observed and the predicted

Figure 7 and 8 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 7 and 8 and table 2, 3, 4 and table 5 respectively, where the coefficient of determination  $R^2$  gave 0.94 and 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. 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. and Ogedengbe M.O., Effect of Fines Content on Some Engineering Properties of Lateritic soil in Ile- Ife, Department of Civil Engineering, Obafemi Awolowo University, Ile-Ife, 2009.
- 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., and Chottopadhyay B.C., Correlation Between CBR And Index Properties Of Soil, Proceedings of Indian Geotechnical Conference, Kochi (Paper No. A-350), 2011.
- Dharamveer, s., Reddy K. S., and Laxmikant, Y., 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, 2011.
- Magdi M. E., and Zumrawi, 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, 2012.

- Patel R.S., and Desai, M.D., CBR predicted by index properties for alluvial soils of South Gujarat. In: Proceedings of the Indian geotechnical conference, Mumbai, pp 79–82, 2010.
- Quinlan, J.R., 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, 1992.
- R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>, 2021.
- Rahim A.M., Subgrade soil index properties to estimate resilient modulus for pavement design *International Journal of Pavement Engineering* Volume 6, 2005 - Issue 3, pp163-169,2005.
- Srinivasa, R., Ruchita., Pankaj, S.,Sathiraju, V. S., 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, 2019.
- Uyanik, G.K., A study on multiple linear regression analysis. *Social and Behavioural sciences* 106 pp 234-240,2013.
- Valentine, Y. K.,Souleyman M.M., and Bertille M., Armand Sylvain Ludovic Wouatong . Lawrence Aleh, 2017.
- Vapnik V., *Statistical Learning Theory*”, Wiley, New York, 1998.
- Vasu,N., and Suresh,P. K., Study of ANN Program for Prediction of California Bearing Ratio of Fine-Grained Soils. *International Journal of Science and Research (IJSR)* ISSN: 2319-7064, 2017.
- Venkatasubramanian, C., and Dhinakaran, G.,Ann model for predicting cbr from index properties of soils. *International journal of civil and structural engineering* Volume 2, No 2 , 2011.
- Vinod, P., Reena, C., Prediction of CBR value of lateritic soils using liquid limit and gradation characteristics data. *Highw Res J IRC* 1(1) pp89–98, 2008.

## Biography

**Engr. Akinwamide Joshua** is a Ph.D. Student in the Department of Civil Engineering University of Benin, Edo-State Nigeria, with a broad knowledge in Geotechnical Engineering/Civil Engineering Materials. He is a lecturer at the Federal Polytechnic, Ado-Ekiti, Ekiti -State Nigeria. He is a Registered Civil Engineer by Council of Regulation of Engineering in Nigeria (COREN). A corporate Member Nigerian Association Technologist in Engineering (NATE) . He holds M. ENG in Geotechnical Engineering. He has over 15years working experience as a Geotechnical engineer and Quality Control and Quality Assurance Specialist.As a researcher he has published over 15 papers in peer-reviewed journals, conferences and workshops at both local and international level

**Engr. Prof. Jacob Odeh Ehiorobo** is a professor of Applied Geomatic, Water Resources and Environmental System Engineering. He is a professor in the Department of Civil Engineering, University of Benin, Edo- State Nigeria. He has held several administrative positions in the university such as Dean of Faculty of School of Environmental, Director of External Linkage for the University and Deputy Vice Chancellor Administration. Upon graduating with BSc, MSc and PhD as a prolific researcher he has published over 100 papers in peer-reviewed journals, conferences and workshops at both local and international level.Engr. Prof. Jacob Odeh Ehiorobo research interest includes: Deformation Surveys and Analysis, Precise engineering surveys, GNSS Positioning and Geodetic Surveys. Remote sensing and GIS Applications in Disaster Monitoring and Control, Water Resources Modelling and Environmental Hazards Analysis, Highway and Transportation including Automatic Vehicle location.

**Engr. Prof. Sylvester Obinna Osuji** is a Professor of Structural engineering in the Department of Civil engineering, University of Benin, Edo-State Nigeria and currently a Visiting Professor on Accumulated Leave at the Department of Civil Engineering, College of Technology, Federal University of Petroleum Resources, Effurun, Delta State, Nigeria.Engr. Prof. Sylvester Obinna Osuji hasheld several administrative position in the University of Benin, Edo- State such as : Head of department of Civil Engineering, Sub -Dean School of Engineering and Deputy Director procurement for the University of Benin. Upon graduating with BSc Civil engineering, M.ENG Structural engineering and PhD in structural Engineering. He has published more than 70 papers in peer-reviewed journals, conferences and workshops at both local and international level.Engr. Prof. Sylvester Obinna Osujiis also a practicing lawyer who holds LLB (Law), (University of Benin, BeninCity, Nigeria) – 2006., B.L., (Nigerian Law School, Abuja) – 2007and L.L.M (Masters of Law) (University of Benin, Benin City, Nigeria) – 2009

**Engr. Dr. Ebuka Nwankwo** is an associate professor in the department of Civil Engineering, faculty of Engineering University of Benin. Dr Nwankwo graduated with a first class honours in Civil Engineering in 2005 from the Federal University of Technology Owerri. After spending some time in the industry, Dr Nwankwo proceeded for his MSc and PhD in Structural Engineering from the Imperial College London. In 2014, he was awarded a PhD from Imperial

College. Dr. Nwankwo has published over 35 articles in peer review journals. Dr Nwankwo has been a visiting researcher at the University of Liverpool. He is COREN registered and member of the Nigerian Society of Engineers (NSE). He has been involved in the training and mentoring of young engineers for NSE. He has a wide range of experience in civil engineering design and project management. He has been called up to give his expert opinions on my projects within and outside Nigeria. Dr Nwankwo has also worked as professional structural engineer in Paris. He has been involved in the design of high-rise structures and geotechnical investigations for civil infrastructure.