Civil and Environmental Research www.iiste.org
ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online)
Vol.14, No.3, 2022

Estimation of Some Geotechnical Indices of Soils using Machine Learning Techniques

Joshua Akinwamide[1]* Odeh Jacob [.2] Sylvester Osuji [2] Nwankwo Ebuka [2] Ngozi Kayode Ojo [2]
1. Department of Civil Engineering, Faculty of Engineering, The Federal Polytechnic Ado Ekiti, Ekiti State,Nigeria
2.Department of Civil Engineering, Faculty of Engineering University of Benin ,Benin City, Edo State, Nigeria
*Email:akinwamidejoshua2@gmail.com

Abstract
Adoption of a good estimation model for the prediction of sub soils properties before the commencement of a construction project, or at the preliminary stage of project planning is highly imperative. This will mitigate the most unexpected costs incurred during construction which are mostly geotechnical in nature. This research aims to use Machine Learning ML tools such as Multiple Linear Regression (MLR) Artificial Neural Network(ANN),Support Vector Machine(SVM), Random Forest(RF) andM5 Tree (M5P) in geotechnical Engineering with a view to correlate Optimum Moisture Content(OMC), Maximum Dry Density(MDD) and Soaked California Bearing Ratio(SCBR) and Unsoaked California Bearing Ratio (USCBR) from the measured index properties. The results from index properties classified the soils of the study area as A-2-4, A-2-6, A-2-7 and A-7-5 for Ekiti Central Senatorial Districts (ECSD) and A-2-4, A-2-5, A-2-6, A-2-7, A-4, A-5, A-6 and A-7-5 for Ekiti South Senatorial Districts (ESSD) while Ekiti Northern Senatorial Districts (ENSD) were classified as A-2-4, A-2-5, A-2-6, A-2-7, A-6 and A-7-6. Conversely. The strengths of the developed Machine Learning models have been examined in terms of regression coefficient (R^2) and Root Mean Square Error (RMSE) values. It is found that all the five ML models predict OMC %, MDD, SCBR and USCBR close to the experimental value. However, the prediction of OMC %, MDD, SCBR and USCBR by RF is found better than other ML models deployed in this research.
DOI: 10.7176/CER/14-3-04
Publication date: May 31st 2022

1. Introduction

Empirical correlations are frequently applied in Geotechnical Engineering to asses various engineering parameters of soils. Correlations are generally derived with the help of statistical methods using data from extensive laboratory or field testing. Least Square Regression(LSR), Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), K.NearestNeighbor's (KNN) and M5 model trees (M5P) are some of the types of machines learning (ML) techniques currently used for predicting geotechnical indices. These techniques learnfromdatacasespresentedtothemtocapturethefunctionalrelationshipamongthedataeven if the fundamental relationships are unknown or the physical meaning is tough to explain. This contrasts with most traditional empirical and statistical methods, which need prior information about the nature of the relationships among the data. ML is thus well suited to model the complex performance of most Geotechnical Engineering materials, which, by their very nature, exhibit extreme erraticism. This modeling possibility ,as well as the ability to learn from experience, has given ML methods superiority over most traditional modeling techniques since there is no need for making assumptions about what could be the primary rules that govern the problem in hand. These methods have been widely applied to tackle various civil engineering problems by different authors: (Goh1995; Saini et al.,2007; Siddique et al.,2011; Pal etal.,2012; Puri et al.,2015; Singh et al., 2016; Anbazhagan et al., 2016; Prasadetal.,2017; Singhetal.,2017). Most of these geotechnical properties are evaluated in the laboratory and some are estimated in the field. Their calculation requires a specific laboratory equipment, an experienced geotechnical engineer with a team of skilled technicians. Thus, determination of these properties is costly and time consuming. Also, soil is a highly erratic material as its performance is based on the processes due to which it is formed. Hence, correlations developed for one region may not be applicable for the other. This as certains the need to develop region-based correlations to predict geotechnical properties. Experimental affinity measure essential area of the Geotechnical Engineering where it has been applied as a solution to many challenges, interpreting various situation and prediction of the initial unknown data based on other measured parameters during the preliminary geotechnical assessment (Ameratunga et al., 2016; Dysli et.al., 2013; Michelet al.,2013).Prediction of soil engineering properties from their index and state parameters using Machine Learning (ML) approach is not new. Earlier empirical correlations have been developed and the compaction parameters(Ajayi et al., 2010),permeability (Boroumand et al., 2005) unconfined compressive strength(Gunaydin e tal.,2010) (Kalkanetal.,2009),angle of shearing resistance (Kayadelen et al., 2009),shear strength (Goktepe et al., 2008), (JainRajeev et al., 2010), (Korayen et al.,1996), (Sivrikaya, 2009) bearing capacity(Nejad et al.,2009), resilient modulus(Zaman et al.,2010), of soils have been related to their index properties such as void ratio, particle sized 10(size corresponding to 10%

finer),% finer than 425micron, liquid limit, plasticity index etc.The CBR is also been related to some of the index properties (Kin Mak, 2006; Linveh 1989; Stephens 1990; Taskiran, 2010; Yildirim et al., 2011). Some of the correlations, which are found in the literature, areas follows. Black(1962) has given the graph betweens oil indices Plasticity Index(PI),Liquidity Index(LI) and the CBR, which is applicable for saturated clays. Bhatia and Johnson (1969) have correlated CBR with suitability index, which is a function of plasticity and gradation of soil. Agrawal and Ghanekar (1970) have proposed the relation in the form of an equation: CBR=2.0-16.0*log(OMC)+0.07*LL..........(1)where, OMC is the standard Proctor moisture content infraction and LL is the liquid limit value of the soil. Regression A N analysis is a statistical tool which could be used to predict the correlation between two or more variables ,It includes various methods for modeling and analyzing different variables and finally fitting alinear or nonlinear equation, Artificial Neural Networks (ANNs) are artificial intelligence which try to imitate the human brain and nervous system (Alshayeb, et al., 2013). Past researchers have made comparative studies between artificial neuron network and multiple linear regression in the field of Geotechnical Engineering; for example,(Harinietal.,2014)compared them for prediction of California Bearing Ratio(CBR)of fine-grained soils;(Boadu et al.,2013; Siddiqui, et al., 2014) tried to predict geotechnical indices from electrical measurements using both models and compared their results. Measurement of the compaction characteristics and California Bearing Ratio (CBR) of soil in the laboratory is neither time-efficient nor cost-efficient. The growing need for a predictive model as an alternative to laboratory testing was the impetus for motivation in this project. In a geotechnical engineering project, an accurate prediction of the maximum dry density (MDD), optimum moisture content (OMC), CBR for soaked and Unsoaked will not only save time, but will also help reduce the costs, cut down on the use of resources, and lessen the required human labor. The key purpose of this research was to develop an advanced mathematical model that can explain the relationship between the physical properties of fine-grained soil and each of its compaction properties. Furthermore, a comparative study was undertaken to determine the models that produced the best results. The analyses were focused primarily on artificial neural networks, support vector machines, Random Forest, MS Tree along with multiple linear regression models, to make the comparative study more meaningful and, at the same time, more intrigue.

2.0. Materials and Methods

2.1. Location of the Study area

The location where sample materials for this study was obtained is as shown in Fig 1. Soils Samples were obtained from 480 points across the three Senatorial Districts namely : Ekiti North , Ekiti South and Ekiti Central as shown in the legend of figure 1

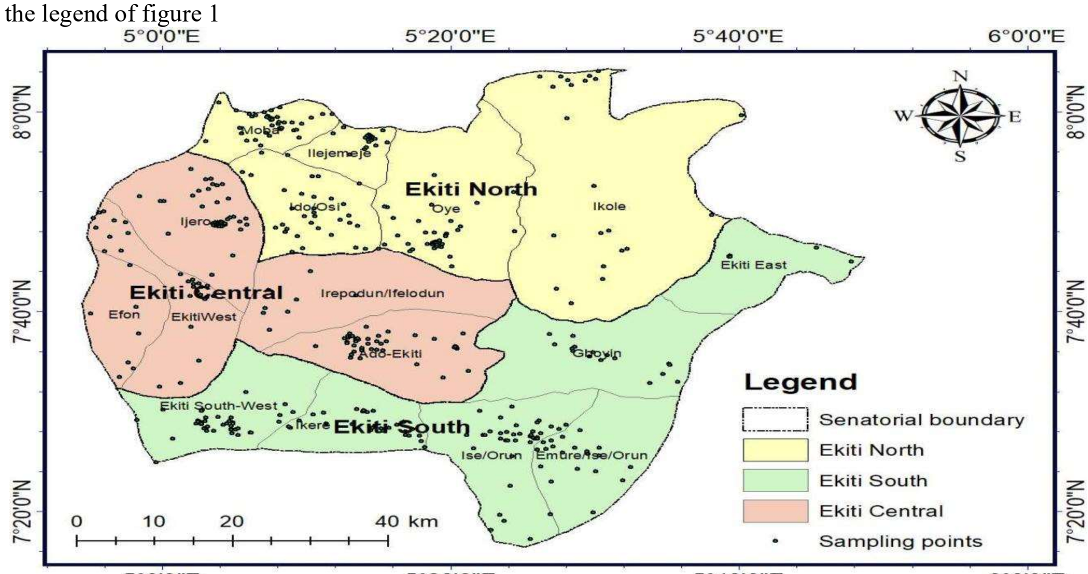

Figure 1. Locationofthestudyarea– the three senatorial districts,EkitiStatesouth western Nigeria.

2.3. Research Materials

The materials used in this study includes: global positioning system device (GPS), other materials include, field notebook and data sheets, sample bags, soils sample, sample labels, trowel, spade and scoop.

2.4. Field Procedure, Sampling and Testing Methods

These involved site recognizances of the study area and proper observation of the study areas and subsequent collection of samples for various laboratory analyses. 480 disturbed soil samples were obtained from borrowed pits found within the three Senatorial Districts of Ekiti State South Western Nigeria attheaveragedepthof2m from the ground surface and analyzed for various index and Soil strength (compaction and California Bearing Ratio) properties. Index properties and soil strength tests were analyzed in The Federal Polytechnic Ado-Ekiti (FPA) geotechnical laboratory in compliance with the methods proposed by British Standard BS1377 (1990).

2.5. Machine Learning implementation methodology

The flow chart in Figure 3.6 present the Machine Learning implementation flow chart describing the key steps for the implementation of Machine Learning working mechanism

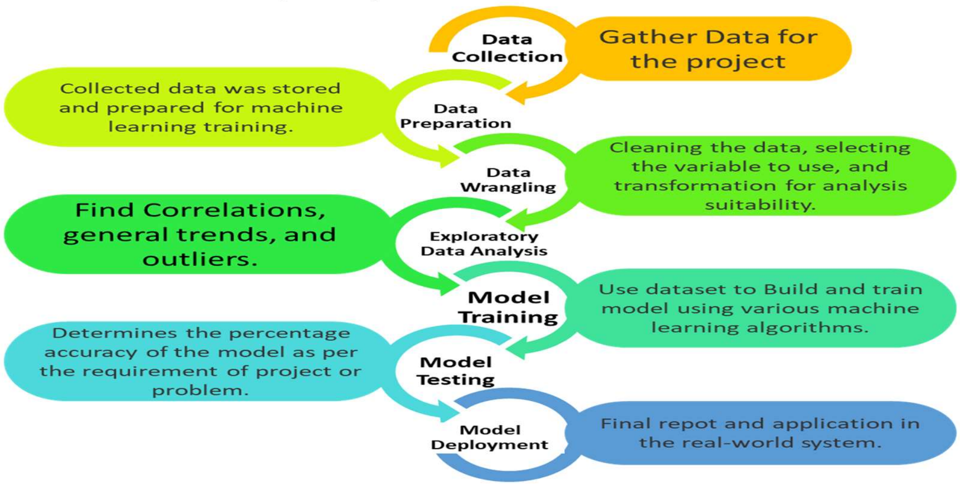

Figure 2.Flow chart describing the key steps for the implementation of Machine Learning working mechanism

2.6. Data Collection for Machine Learning tools (ML) Treatment

The databank was divided into three parts: training (70%),Testing(30%) and validation).The training data-set was used in order to train the machine learning models, the validation data-set was used to stop the learning process and all testing data-set were used to assess the Machine Learning(ML) models performance after completion of the training process. Each data-set consists of the factors that affect the output parameter staking into account the five variable that will be selected as input to develop the machine learning models, the variables were Consistency limits symbolized by LL(%), PL(%) NMC,G Sand Percentage passing 75-micron sieve(%)for the prediction of both compactions Parameters (OMC and MDD),SCBR and UCBR. The consistency limits (LL) (%) and PL(%)) were selected as input parameters for the four variables of MDD, OMC, SCBR and UCBR as outputs layer. This division has been used successfully and reported by Shahin, et al. (2002) in the literatures. The available data was divided into their sub-sets, the input and output data were pre-processed and normalized between -1.0and1.0. The Correlation Coefficient R and the Root Mean Square Error (RMSE) were used to measure the performance of the predictions since the Correlation coefficient is a key function to establish a relative relationship between the expected and the observed data (Shahin et al., 2008). Smith, (1986), this was established by plotting the experimented and predicted values on vertical and horizontal axis respectively (Piñeiro et al., 2008) for the developed equations to measure their individual efficiency.

2.7. Principal Component Analysis (PCA)

The principal components are the linear combinations of the original variables that account for the variance in the data. The maximum number of components extracted always equals the number of variables. The eigenvectors, which are comprised of coefficients used to calculate the principal component scores. The coefficients indicate the relative weight of each variable in the component. Principal Component Analysis is based on only independent variables. So, the dependent variable was removed from the dataset.

3.0. Results and Analysis

From the analysis of index properties the Central Senatorial Districts soils were classified into four classes as clay of low compressibility (CL) and clay of high compressibility (CH) according to (USCS,1986) while AASHTO classification system classifies as A-2-4, A-2-6, A-2-7 and A-7-6 with subgrade rating of Excellent to good and Fair to poor respectively. 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 while 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.

3.1. Measurement of Interrelationship among the Predictors

Principal Component Analysis was adopted to handle multicollinearity issues to avoid any significant relationship among the independent variables. Ahmed, (2016) define Principal component Analysis as an important method in machine due to its two-fold nature. PCA reduces the dimensionality of the dataset, which takes the dimension that encode the most important information. By reducing the number of dimensions, the data utilized less space, thus allowing classification on larger datasets in less time. Further, by taking only salient dimension, PCA project the dataset onto dimensions that hold the most meaning, thus drawing out pattern out pattern in the dataset. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression and is a common technique for finding pattern in data of high dimension. But a major problem in mining scientific data set is that the data often high dimensional. When the number of dimension reaches hundred or even thousands, the computational time for the pattern recognition algorithm can become prohibitive. In many cases there are a large number of features representing the object. One problem is that the computational time for the pattern recognition (Jim. Frost, 2020)

Figure 3a Present a Scatter matrix showing interrelationship among the predictors where the Lower triangles provide scatter plots which help to find a relationship among the variables , the Histograph also reflect the degree of normality ensuring the data obey the MLR assumption that specify normal distribution of variable in a data set .Its evident from NMC, SAND, GS,LL and PL were normally distributed reflecting a dome or bell shape while Gravel and Fines show a skewed shape which is due to soil veracity . The data /value displayed at the upper triangle are represented by the scatter at the lower triangle, the negative values showed an inverse relationship among the variable it has contributed but negative . The lower the values at the upper triangle the more scatter the diagram and vice versa. Gravel and Fines, Sand and Fines are highly correlated. Similarly, LL and PL are also highly correlated which leads to multi co-linearity issues. Predicting the model based on this dataset may be erroneous, this is because it is violation of an assumption in regression that specify that their should not be significant correlation among the variables or should not be significantly correlated. (Jim frost,2020)

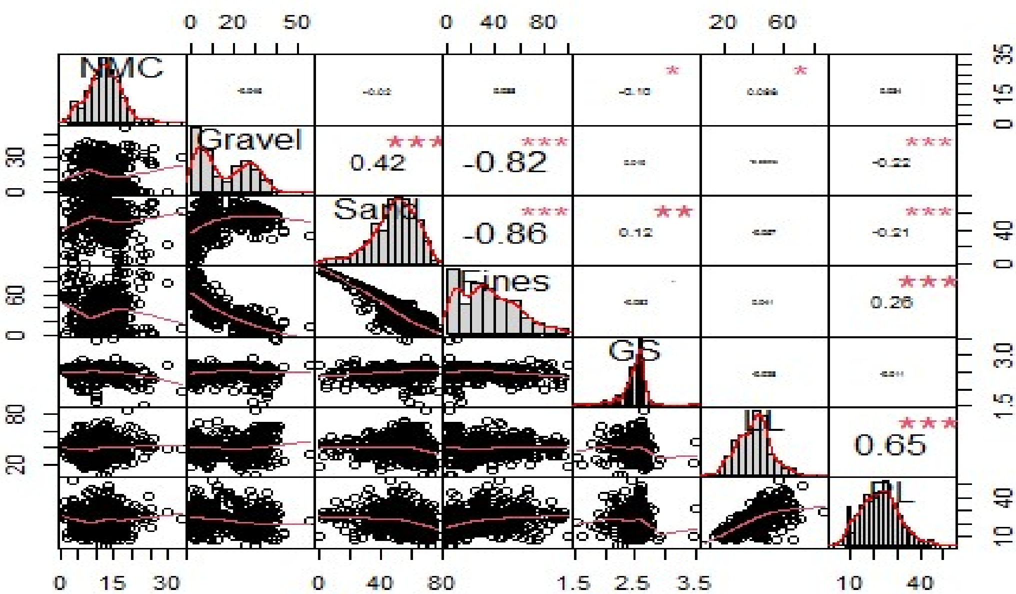

Figure3a: Scattered matrix showing a relationship among the predictors

The scatter matrix presented in fig 3b which show that all the obvious relationship among the input variables is gone, hence there exists no significant relationship among the predictors. This serves as a good foundation for multiple linear regression analysis.

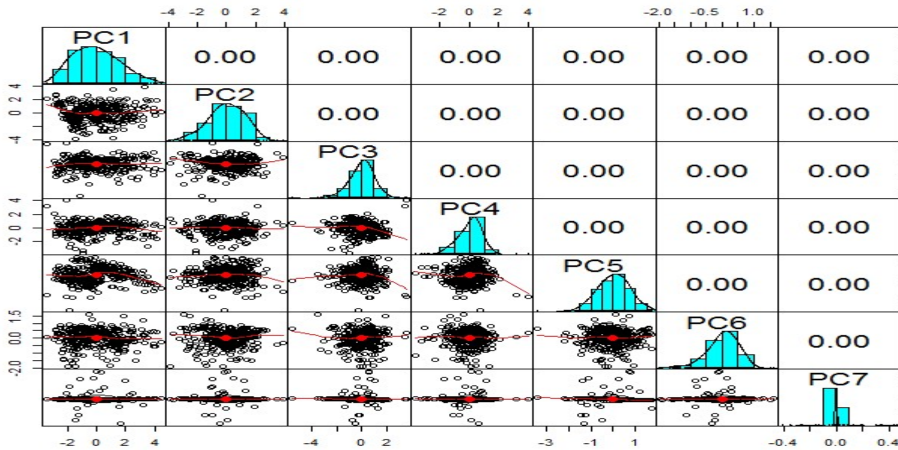

Figure 3b.Scattered matrix showing no relationship among the predictors.

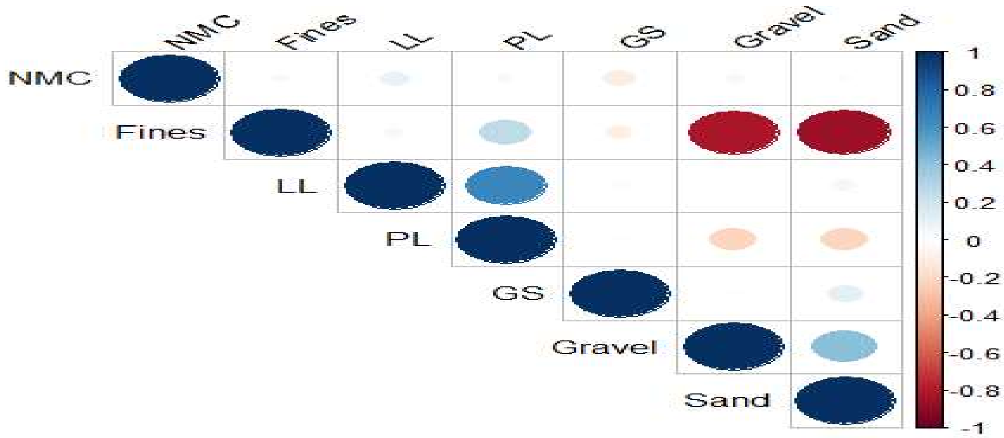

Figure 4.Visualization of correlation matrix using a correlogram

The distribution of each variable as shown on the diagonal as presented in fig 4 . On the bottom of the diagonal, the bivariate scatter plots with a fitted line are displayed. On the top of the diagonal, the value of the correlation plus the significance level as stars. Each significance level is associated to a symbol : p-values(0, 0.001, 0.01, 0.05, 0.1, 1) The variables PI and LL are highly correlated (p-value = 0.65) as presented in Figure 4.10 they should therefore be removed or treated from the model for violation of assumption of non existence of collinearity among predictors while **Sand and Fines** also has a highly correlated (p-value = -0.86) which mean an inverse relationship within the variables as shown in Figure 3a. Positive correlations are displayed in blue and negative correlations in red color. Color intensity and the size of the circle are proportional to the correlation coefficients. In the right side of the correlogram, the legend color shows the correlation. The lower the color intensity, the lower the level of correlation or significant level for each color. It should be noted that those aesterisks values in figure 3a has been treated but they are just displayed to show their effect and for the purpose of explanation.

3.2. Eigen Vector from the Principal Component Analysis (PCA)

The principal components are the linear combinations of the original variables that account for the variance in the data. It is based only on the independent variables, so we removed the response variable from the dataset. The maximum number of components extracted always equals the number of variables, the eigenvectors, which comprised of coefficients used to calculate the principal component scores as presented in Table and . The coefficients indicate the relative weight of each variable in the PCA. The eighth variable is then removed (dependent) from the dataset as shown in Fig.3a and 3b for OMC, MDD ,SCBR and USCBR respectively

Table 1.Eigen vectors from the PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7

Table 2. Eigen Analysis of the Correlation Matrix

	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

The first principal components (PC1) explain the variability around 36%, second (PC2) 23%, third (PC3) 16%, and fourth (PC4) 12%. Summarily, the first four principal components (PC1 to PC4) capture the approximately 88% of the variability which is a majority of the variability in table 1 and 2 respectively. In this case, the resulting Four components score variables are representative of, and can be used in place of the seven original variables with a 12% loss of information, while the remaining components contribute negligible variability. In these results, the scores for the first four principal components can be calculated from the standardized data using the coefficients listed under PC1 to PC4.

3.3. General Predicting models with Principal Components

The general predicting model is as presented in in equation 1 to 4 from the machine learning techniques

$$\mathbf{PC1 = [-0.0105(NMC) - 0.494(Gravel) - 0.52(Sand) + 0.605(Fines) - 0.136(GS) + 0.1431(LL) + 0.2751(PL)} \quad (1)$$

$$\widehat{\mathrm{PC2}} = -0.\mathbf{1877(NMC) - 0.1850(Gravel) - 0.1412(Sand) + 0.1916(Fines) + 0.1102(GS) - 0.6968(LL) - 0.6132(PL)} \quad (2)$$

$$\widehat{\mathrm{PC3}} = -0.\mathbf{7041(NMC) + 0.0097(Gravel) - 0.0184(Sand) + 0.0104(Fines) + 0.6698(GS) + 0.1244(LL) + 0.1991(PL)} \quad (3)$$

$$\mathbf{PC4 = 0[.6618(NMC) - 0.2370(Gravel) + 0.1055(Sand) + 0.0630(Fines) + 0.6995(GS) - 0.0292(LL) + 0.0232(PL)} \quad (4)$$

3.4. The Derived Linear Model from PC1, PC2, PC3 and PC4

The derived Model is given below as theoretical and estimated model in equation (5) to (6) and equation (7) to (8) for MDD, OMC and equation (9) to (12) for CBR soaked and unsoaked respectively, where the first four principal components were adopted since the majority of the information are present in the four components. The resulting four component score variables are representative and can be used in place of the seven original variables with a 12% loss of information.

- **The theoretical model for MDD**

$$\mathbf{MDD = \alpha + \beta_1(PC1) + \beta_2(PC2) + \beta_3(PC3) + \beta_4(PC4) + \epsilon.} \quad (5)$$

- **The estimated model with actual coefficients for MDD**
- **MDD = 1907.25 - 1.13(PC1) + 10.22{PC2}) + 12.48{PC3}) - 35.24{PC4})** (6)
- **The theoretical model for OMC**
- $\mathbf{OMC = \alpha + \beta_1(PC1) + \beta_2(PC2) + \beta_3(PC3) + \beta_4(PC4) + \epsilon}$ **(7)**
- **The estimated model with actual coefficients for MDD**

OMC = 14.23 + 1.01{PC1}) + 0.5{PC2}) - 0.29{PC3}) - 0.24{PC4}) (8)

- **The theoretical model for Soaked California Bearing Ratio (SCBR)**

$\mathbf{SCBR = \alpha + \beta_1(PC1) + \beta_2(PC2) + \beta_3(PC3) + \beta_4(PC4) + \epsilon}$ **(9)**

The estimated model with actual coefficients for Soaked California Bearing Ratio

- **SCBR = 35.14 - 7.5(PC1) - 3.88(PC2) + 0.13(PC3) - 6.67(PC4) (10)**
- **The theoretical model for UN-Soaked California Bearing Ratio (USCBR)**
- $\mathbf{USCBR = \alpha + \beta_1(PC1) + \beta_2(PC2) + \beta_3(PC3) + \beta_4(PC4) + \epsilon}$**(11)**
- **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) (12)**

Where the Alpha (α) represent the interceptand Betas are the regression coefficient measuring the effect of the imput on the output.

3.5. Measures of Accuracy between the Actual and the Predicted Values (Goodness of fit) for OMC , MDD, SCBR and UN-SCBR

Measures of Accuracy between the Actual and the Predicted Values for Compaction characteristics (OMC & MDD) and California Bearung Capacity CBR, soaked and unsoaked is presented in Figure 3 to6 and Table 3 to 6 respectively.

Degree of relationship between experimented and predicted values is taken to be high when the R^2 value is greater than 0.8. Also, RMSE is preferable because larger residual errors are dealt with more sensitively, and RMSE≈0 represents the least errors [Alade et al ,2019]. Furthermore, higher R^2 and lower RMSE values shows a good model results and accurate calibration (Jalal et al., 2021; Azim et al.,2020).

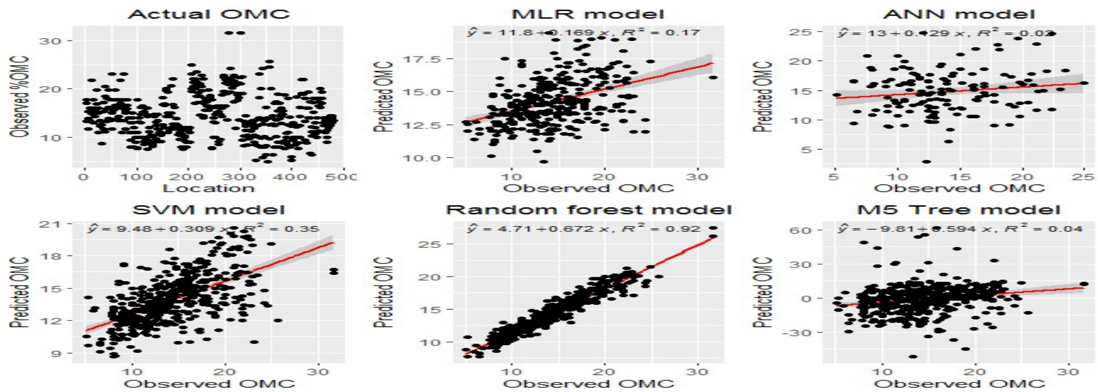

Figure 5.Scattered plots for the performance analysis of the models for predicting Optimum Maximum Moisture Content OMC %

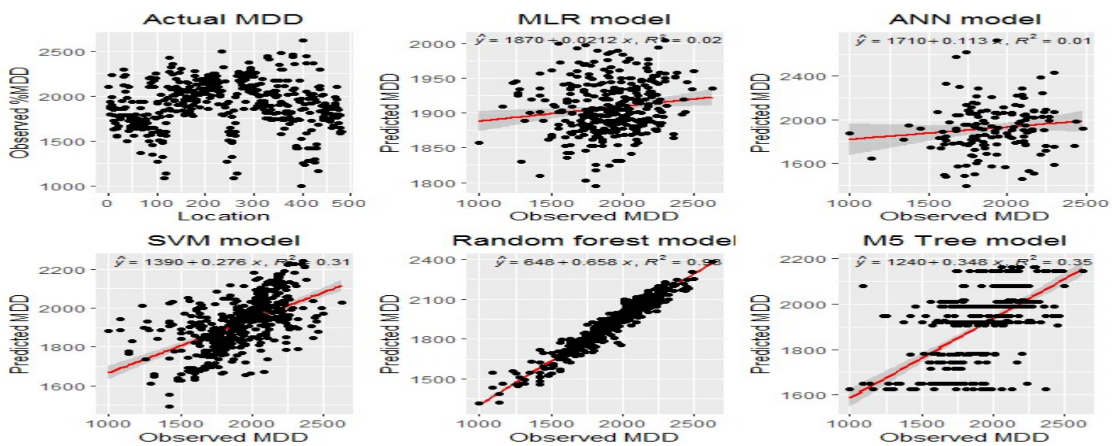

Figure 6.Scattered plots for the performance analysis of the models for predicting Maximum Dry Density (MDD kg/m^3)

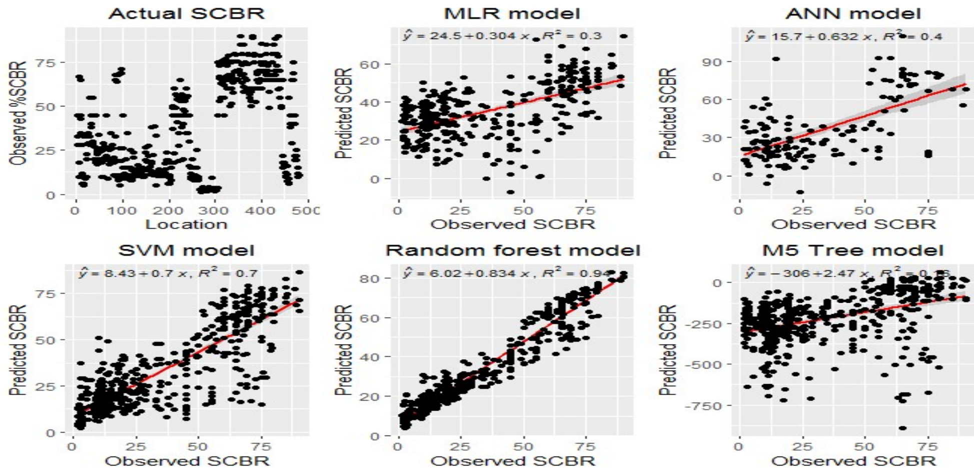

Figure 7.Scattered plots for the performance analysis of the models for predicting Soaked California Bearing Ratio (SCBR)

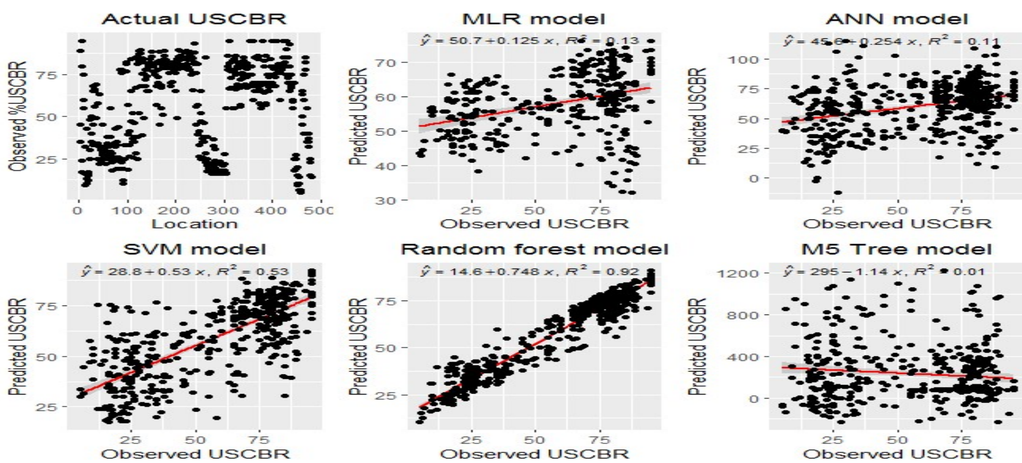

Figure 8.Scattered plots for the performance analysis of the models for predicting UNSoaked California Bearing Ratio (UNSCBR)

Table 3.Measure of accuracy (Goodness of fit) for OMC

Techniques	Soil	Goodness of fit					
	Indices	ME	MAE	MSE	RMSE	R	R^2
MLR	OMC	0	3.15	15.25	3.91	0.41	0.17
ANN	OMC	0.48	3.7	20.69	4.55	0.37	0.14
MSTREE	OMC	0.42	2.58	12.56	3.54	0.59	0.35
R F	OMC	0.01	1.28	2.28	1.67	0.96	0.92
SVM	OMC	-15.6	17.14	405.14	20.13	0.2	0.04

Table 4.Measure of accuracy (Goodness of fit) for MDD

Techniques	Soil	Goodness of fit					
	Indices	ME	MAE	MSE	RMSE	R	R^2
MLR	MDD	0.00	201.29	64435.71	253.84	0.15	0.02
ANN	MDD	0.14	222.73	87286.55	295.44	0.28	0.08
MSTREE	MDD	1046.96	1503.43	360.32	1898.21	0.10	0.01
R F	MDD	-3.45	76.73	10329.04	101.63	0.96	0.93
SVM	MDD	14.62	149.39	47557.59	218.08	0.56	0.31

Table 5.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 6.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

4.1 Comparison of Developed Models using Statistical Parameters such as Coefficient of Correlation (R^2) with previous Authors

The predicted values generated by Random Forest (RF) model seems to move side by side with the actual MDD, OMC, SCBR and UN-SCBR shown in figure 7to 10 respectively, where the coefficient of determination ($R^{2)}$from Random Forest (RF) gave 0.92, 0.93, 0.94 and 0.92 for MDD, OMC, SCBR and UN-SCBR respectively. R^2higher than 0.8 showed that the estimated and observed values are highly correlated (Onyelowe, et al., 2021; Igbal et al., 2021; Jalal et al, 2021; Azim et al., 2020). From the foregoing, its concluded that the strength of the developed models after comparison in terms of regression coefficient (R^2) and Root mean square error (RMSE) values with existing literature as shown in Figure. It is established that all the Machine Learning (ML) model technique predict OMC, MDD and CBR close to the experimental value. Degree of relationship between experimented and predicted values is taken to be high when the R^2 value is greater than 0.8. (Onyelowe et al., 2021).Also, RMSE is preferable because larger residual errors are dealt with more sensitively, and RMSE≈0 represents the least errors (Alade et al., 2019). Furthermore, higher R^2 and lower RMSE and MAE values shows a good model results and accurate calibration (Jalal et al., 2021; Azim et al.,2020). However, the prediction of OMC, MDD and CBR by Random Forest (RF) is found better compared to other technique evaluated. Its noteworthy from figure11 to13 after comparison with previous studies by various researchers that the Random forest results estimated from this research is in agreement with existing work in literature and that Random Forest still stand to be the best model for predicting Compaction characteristics (OMC and MDD) and Soaked and Unsoaked CBR respectively using the basic simple soil index properties. From the foregoing, its concluded that the strength of the developed models after comparison in terms of regression coefficient (R^2) and Root mean square error (RMSE) values. It's established that all the Machine Learning (ML) model technique predict OMC, MDD and CBR close to the experimental value. However, the prediction of OMC, MDD and CBR by Random Forest (RF) is found better compared to other technique evaluated.

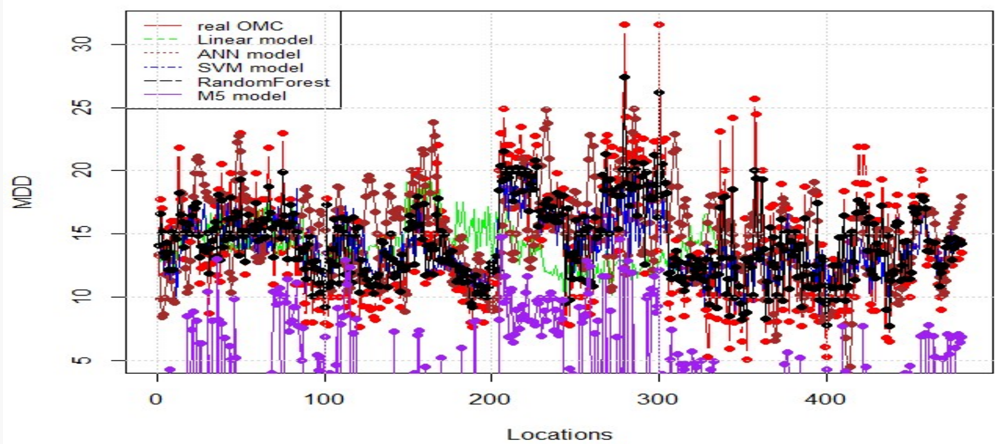

Figure 7. Line plot showing the movement of the observed and the predicted

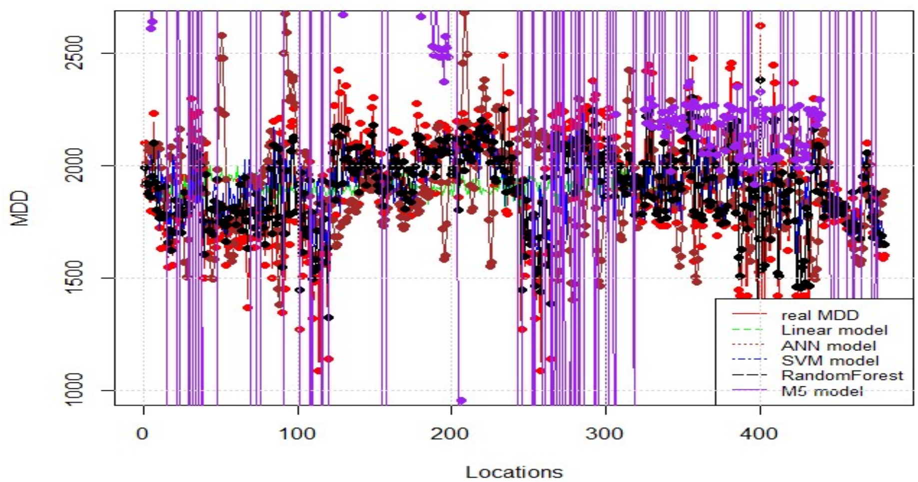

Figure 8.Line plot showing the movement of the observed and the predicted

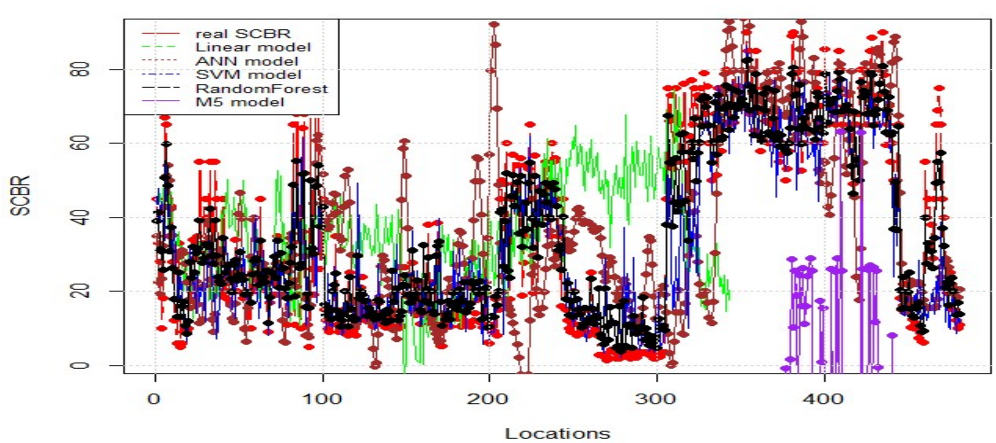

Figure 9.Line plot for the movement of the observed and the predicted SKCBR

Civil and Environmental Research
ISSN 2224-5790 (Paper) ISSN 2225-0514 (Online)
Vol.14, No.3, 2022

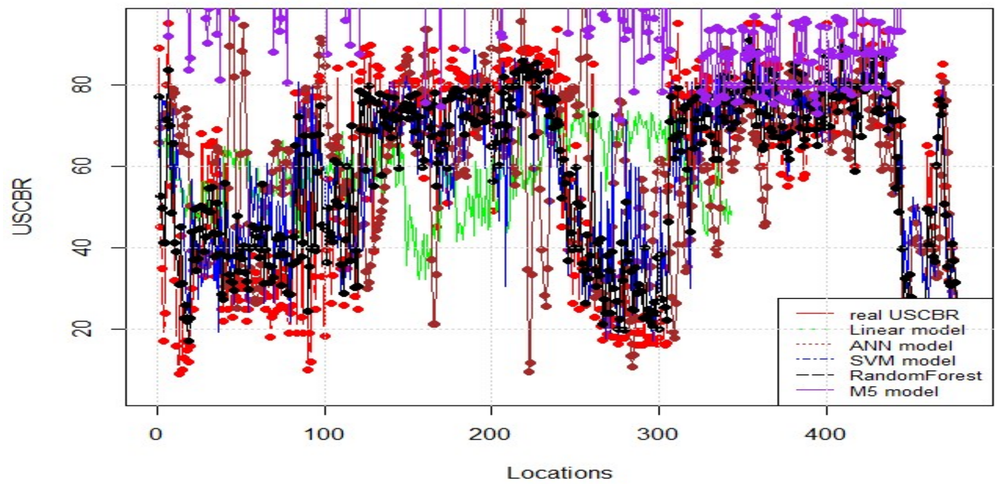

Figure 10.Line plot for the movement of the observed and the predicted USCBR

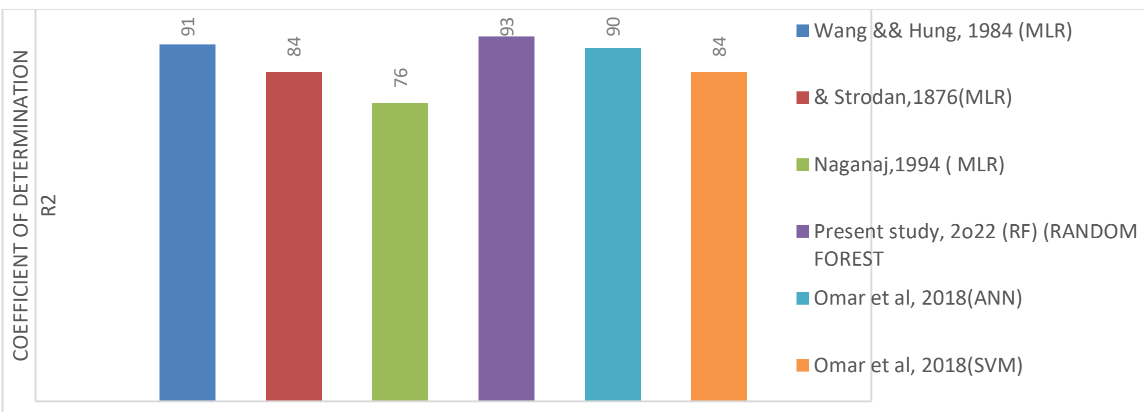

Figure 11.Existing Literature and present Coefficient of determination R^2 with different Machine learning tools for MDD

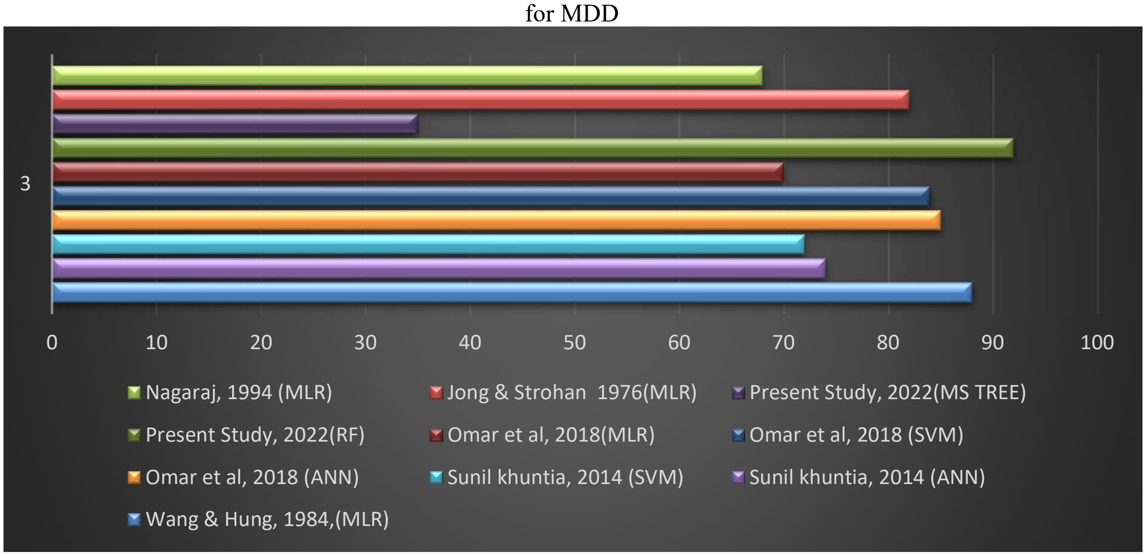

Figure 12.Existing Literature and present Coefficient of determination R^2 with different Machine learning tools for OMC

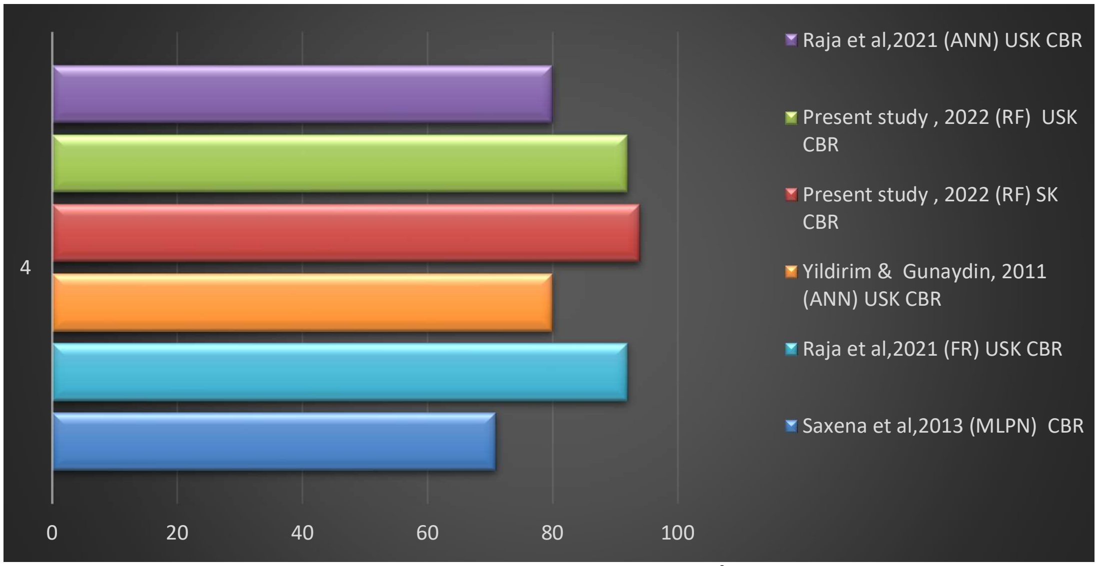

Figure 13.Existing Literature and present Coefficient of determination R^2 with different Machine learning tools for California bearing Ratio (CBR)

4.3 Conclusions

Application of Machine Learning techniques to estimate some geotechnical indices of Ekiti State Senatorial Districts Soils has been examined in this study. A trial pit 1-2 m deep was excavated and 480 samples of undisturbed soils were obtained for laboratory tests. From the results obtained and findings the following conclusion are drawn

1. The soil index properties according to AASHTO and USCS characterized the soils of Central Senatorial Districts into four classes as clay of low compressibility (CL) clay of high compressibility (CH) and A-2-4, A-2-6, A-2-7 and A-7-5. and Southern Senatorial Districts 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 while Ekiti 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 which implies is that Ekiti Central Districts soils has more geomaterials for construction of base course and sub- base courses compared to other districts soils which are more of sub -base and sub-grade materials which may require stabilization to improve their engineering properties
2. The Machine Learning has developed a simplified estimated models for prediction of OMC, MDD, SCBR and USCBR in the study area as shown below:

a. **MDD = 1907.25 - 1.13(PC1) + 10.22{PC2}) + 12.48{PC3}) - 35.24{PC4})**
b. **OMC = 14.23 + 1.01{PC1}) + 0.5{PC2}) - 0.29{PC3}) - 0.24{PC4})**
c. **SCBR = 35.14 - 7.5(PC1) - 3.88(PC2) + 0.13(PC3) - 6.67(PC4)**
d. **USCBR = 57.99 - 5.29 {PC1}) - 1.68 {PC2}) + 2.06 {PC3}) - 2.46 {PC4})**

Acknowledged

The author of this article which to appreciate the authority of the Department of Civil engineering of the school of engineering the Federal polytechnic Ado Ekiti. Ekiti State Nigeria for granting us permission to run the experiment in their material laboratory.. We also thanks all the laboratory staff for giving us necessary support.

References

AASHTO (1986)"Standard Specifications for Transportation, Material and Method of Abu-KiefaM.A.1998.

Ahmed, K., Saife Eldin F. O., Rasha Gaffer .M.H., & Somia B. M.(2016*)* Usage of Principal Component Analysis (PCA) in AI Applications *International Journal of Engineering Research & Technology (IJERT)* IJERTV5IS120291, Vol. 5 Issue 12. http://www.ijert.org ISSN: 2278-0181).

Ajayi, F.T., Akande, S.R., Odejide,J.O. & Idowu,B. (2010). Nutritive evaluation of some tropical under utilized grain legume seeds for ruminant's nutrition. J. Am. Sci., 6 (7): 17.

Alshayeb, M., EisaY., & Ahmed, M. (2014),Object-Oriented Class Stability Prediction: A Comparison Between Artificial Neural Network and Support Vector Machine, Arabian Journal for Science & Engineering 39 (11): 7865-7876.

Ameratunga J., Sivakugan, N., & Das, B.M. (2016) Correlations of soil and rock properties in geotechnical engineering, Springer India, India.

Anbazhagan P, A. Uday, S.S.R. Moustafa, & Nassir, S.N.A. (2016) Correlation of densities with shared wave velocities and SPT values." J. Geophys. Engg.13 (3): 320–341.

Bhatia and Johnson (1969). Soil Properties and their Correlations

Black, M.(1962) Models and Metaphors Studies in Language and Philosophy. Madrid Cornell University Press.

Boadu, F.K., Owusu-Nimo, F., Achampong, F., Ampadu, S.I. (2013) Artificial neural network and statistical models for pcredicting the basic geotechnical properties of soils from electrical measurements, Near Surface Geophysics 11(6): 599-612.

Boroumand, A., and Baziar, M.H. (2005) "Determination of compacted clay permeability by artificial neural networks", Ninth International Water Technology Conference, IWTC9, Sharm El-Sheikh, Egypt, pp. 515-526.

British Standards 1377. (1990). Methods of tests for soils for civil engineering purposes, London.

Goh A. T. C. (1995) Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering.

Goktepe, A.B., Selim, A., Gokhan, A., & Tan,O. (2008)"Shear strength estimation of plastic clays with statistical and neural approaches", Building and Environment, vol.43,pp.849–860.

Gunaydin, O., Gokoglu, A., Fener, M. (2010) "Prediction of artificial soil's unconfined compression strength test using statistical analyses and artificial neural networks",Advances in Engineering Software, vol. 41, pp. 1115–1123.

Jain, R., Jain, P. K.,Bhadauria, S.S. (2010)"Computational Approach to Predict Soil Shear strength". International Journal of Engineering Science and Technology,Vol.2 (8),pp.3874-3885

Jim. Frost, (2020): Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models. Singh G., Sachdeva, S.N. & Pal, M.(2016)"M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India". Accident; analysis and prevention 96:108-117.

Kalkan, E.,Akbulut, S., & Tortum, A. (2009) "Prediction ofthe unconfined compressive strength of compacted granular soils by using inference systems*", Journal of Environ Geol,vol.58, pp.1429–1440.*

Kayadelen, C., Gunaydin, O., Fener,M.,& Ozvan,A., (2009) "Modeling of the angle of shearing resistance of soils using soft computing systems", Expert Systems with Applications, vol.36 pp. 11814–11826.

Kin, M. W. (2006) California bearing ratio correlation with soil index properties master of engineering (civil–geotechnics), Faculty of Civil Engineering, University Technology Malaysia.

Korayen, J. (1996) "Prediction of soil shear strength and penetration resistance using some soil properties", Mis. J. Agr. Res, vol.13, no.4, pp. 119-140.

Linveh, M.(1989) Validation of correlations betweena number of penetration test and insitu california bearing ratio test. Transportation Research Record (1219),56–67

Nejad, F.P., Jaksa, M. B., Kakhi, M., & McCabe, B. A., (2009) "Prediction of Pile settlement using artificial neural networks based on standard penetration test data"Journal of Computer and Geotech, volume36, pp. 1125-1133.

Pal,M., Singh, N.K.,&Tiwari,N.K.(2012)" M5 model tree for piers cour prediction using field data set".KSCE *Journal of Civil Engineering*16(6):107-124.

Prasad, H.D., Puri,N., & Jain, A.(2017) "Prediction of compression index of clays using machine learning techniques" Proc. National Conference on Numerical Modeling in Geomechanics, Kurukshetra.

Puri, N. & Jain, A.(2015)"Correlation between California bearing ratio and index properties of silt and clay of low compressibility". Proc. Fifth Indian Young Geotechnical Engineers Conference, Vadodara.

Saini, B., Sehgal, V.K., & Gambhir, M.L.(2007)."Least-cost design of singly and doubly reinforced concrete beam using genetic algorithm optimized artificial neural network based on Levenberg–Marquardt and quasi-Newton back propagation learning techniques". Structural and Multi disciplinary Optimization 34 (3): 243-260.

Shahin, M.A., Jaksa M. B., Maier H.R, (2008) State of the art of artificial neural networks in geotechnical engineering, Electronic Journal of Geotechnical Engineering **8**: 1-26.

Shahin, M.A., Maier, H. R. & Jaksa M.B. (2002) Predicting settlement of shallow foundations using neural networks. J. Geotech. & Geoenv. Eng.

Siddique, R., Aggarwal, P.& Aggarwal, Y.(2011). "Prediction of compressive strength of self- compacting concrete containing bottom ash using artificial neural networks". Advances in Engineering Software42(10):780-786.

Singh G., Sachdeva, S.N. & Pal, M.(2016)"M5 model tree based predictive modeling of road accidents on non-urban section of highways in India". Accident; analysis and prevention96:108-117.

Singh, B., Sihag P., & Singh K. (2017) "Modelling of impact of water quality on infiltration rate of soil by random forest regression". Modeling Earth Systems and Environment (3): 999-1004.

Sivakugan, N., Eckersley J.D. & Li,H. (1998) Settlement predictions using neural networks. Australian Civil Engineering Transactions.

Sivrikaya,O. (2009) “Comparison of Artificial Neural Networks Models with Correlative Works on Undrained Shear Strength”,JournalofEurasianSoilScience,Vol.42,No.13,pp. 1487–1496.

Smith, G. N. (1986) “Probability and statistics in civil engineering: an introduction”, London: Collins.

Stephens, D.J. (1990) “Prediction of the California bearingratio Neural network

Taskiran,T., (2010),“Prediction of California bearing ratio(CBR) of fine grained soils by AI methods”, Advances in Engineering Software, Vol. 41, pp.886–892.

Yildirim, B.& Gunaydin, O. (2011) “Estimation of California bearing ratio by using soft computing systems”, Expert Systems With Applications, Vol.38, pp.6381-6391.

Zaman, M., Solanki, P., Ebrahimi, A. & White,L. (2010)“Neural network modeling of resilient modul us using routine Subgrade soil properties” *Journal of Geomechanics*, ASCE, Vol. 10, No. 1, pp. 1-12, February1.