Predicting California Bearing Ratio Using Machine Learning to Model Soil Behavior for Road Construction in Tshimoyapula, Botswana

Botlhe B. Pule[1*], Jerome A. Yendaw[1], Rodrigo S. Jamisola Jr[2], Onalethata Saubi[1]
[1]Dept. of Mining and Geological
[2]Dept. of of Mechanical, Industrial and Energy Engineering
Botswana International University of Science and Technology (BIUST)
Palapye, Botswana
botlhe.pule@studentmail.biust.ac.bw[*], *yendawj@biust.ac.bw, jamisolar@biust.ac.bw,*
onalethata.saubi@studentmail.biust.ac.bw

Abstract*— The present research work is carried out to predict California Bearing Ratio (CBR) values of soils from the Tshimoyapula area near Serowe in the Central District of Botswana based on the index properties using machine learning techniques. The CBR test is a very important and common test performed on soils to assess the stiffness modulus and shear strength of subgrade materials to determine the thickness of overlaying layers in pavement design. It is an expensive and time-consuming test in addition to difficulty in keeping the sample in the desired condition. The construction industry is one of the least digitized in the world, and using Artificial Intelligence could help achieve profitability, efficiency, safety, and security. Machine learning techniques, namely, Regression Analysis (RA) and Artificial Neural Network (ANN) were developed with different configurations using various laboratory soil properties comprising Liquid Limit (LL), Plastic Limit (PL), Plastic Index (PI), Maximum Dry Density (MDD), and Optimum Moisture Content (OMC) of 200 soil samples that laboratory CBR test was performed on. The index properties were used as input parameters for different models with the CBR as output. Results indicate a good correlation between the input parameters and the output. Artificial Neural Networks showed the least error and the highest accuracy followed by Linear Regression among the models.

***Keywords*— California bearing ratio, index properties, machine learning, artificial neural network, linear regression.**

I. I INTRODUCTION

In the construction of railroads, dams, buildings, and roads there is a lot of testing and analyses of soils to determine their suitability for the project. Variability in the soil conditions from one location to another makes it difficult to predict the behavior of soil. As a result, soil conditions at every point must be completely investigated for proper design. In Botswana, most of the road networks consist of flexible pavements which are made up of different layers namely sub-grade, sub-base, and surface layers [1]. The design and performance of the pavement substantially depend on the strength of the subgrade material based on the California Bearing Ratio (CBR) test on soil samples. The subgrade is the most bottom layer that serves as a foundation of road pavement and the wheel load from pavement surfaces is ultimately transferred to the subgrade. The CBR test is an empirical method of design of flexible pavement and can be defined as the ratio of resistance to penetration of a material to the penetration resistance of a standard crushed stone base material under controlled density and moisture content [2]. The test was originally developed in the 1920s by the California State Highway Department and later incorporated by the Army Corps of Engineers for the design of flexible pavements. Thereafter, the CBR test became so globally popular that it has been incorporated into many international standards like AASHTO T193 and ASTM D-1883, D-449. The significance of the CBR test emerged from the following two factors [3];

1. For almost all pavement design charts, unbound materials are characterized in terms of their CBR values when they are compacted in pavement layers.
2. The CBR value has been correlated with some fundamental properties of soils, such as plasticity indices, grain-size distribution, bearing capacity, modulus of subgrade, shear strength, density, and moisture content.

CBR test being crucial to construction projects is still laborious and time-consuming. Furthermore, the results occasionally aren't accurate due to the poor skill of the technicians testing the soil or the extreme erratic behaviour of the soil [4]. An alternative method which involves the use of the Nuclear Gauge for the measurement of density and moisture content. Is, however, both expensive and requires

ISSN: 2521-2293

special precautions against radioactive material. Other techniques, such as Vane Shear Test, Cone Penetration Test, Unified Compression, and Texas Triaxial Test have been reported in the literature to correlate well with the CBR test, but couldn't replace it due to either some inherent shortcomings of the tests or their limitations to laboratory applications [5].

II. LITERATURE REVIEW

With the development of technology throughout the years, Geotechnical Engineering processes nevertheless continue to be the least digitized and the use of Artificial Intelligence (AI) can push the industry ahead into the future. Most geotechnical engineering procedures use empirical correlations derived with the assistance of statistical methods using antique laboratory or field testing to assess the engineering properties of soils. Most geotechnical parameters which include relative density, compression index (Cu), and Atterberg Limits are determined within the laboratory and some are estimated in the field with some assumptions. Their calculations require specific laboratory equipment, and an experienced geotechnical engineer with a crew of skilled technicians [6]. Integrating AI into these methods through machine learning (ML) algorithms could help attain profitability, efficiency, safety, and accuracy. ML aims to develop methods that can automatically detect patterns in data, after which the uncovered patterns are used to predict future data or others of interest. With this goal, ML is therefore closely related to the fields of statistics but differs slightly in terms of emphasis. Machine learning algorithms can deal with non-linear and plastic issues of soils effectively and avoid the weakness that can be caused by traditional methods [7]. ML is thus well suited to model complex performances of most geotechnical engineering, which by its very nature, exhibits extreme erraticism [6].

Alternative methods such as Plate load test and Dynamic cone penetration test are not achieving the desired outcome, it inspired earlier researchers to develop various predictive models based on the existing data. The literature review showed how different correlations for estimating CBR values were developed and revealed that most of the empirical correlations could not be attained in many cases because most of the empirical equations neither had a high degree of accuracy nor had any generalized solution and the use of these equations in further studies yielded unacceptable results. A summary of previous studies based on the empirical modelling of CBR with parameters used is presented in Table 1. In the age of technology, modern-day researchers have resorted to the use of Artificial Intelligence (AI) as a potential alternative to predict the desired output. [8] further compiled previous studies on AI-based modelling of soil CBR as summarized in Table 2. The developed models resulted in maximum accuracies of R^2=0.9 with lowest values being R^2=0.78. The data sets used varied with different models, with one model having a small set of 20

TABLE I. PREVIOUS EMPIRICAL CORRELATIONS ON CBR VALUES

References	Correlations	No. of samples	R^2
[9]	$Y = 28.79 - 0.61(OMC) + 0.44(PL) + 0.48(PI) - 0.82(\%sand)$	165	0.78
[10]	$Y = -0.1024(PI) + 6.1596$	105	0.94
[11]	$Y = -0.275(LL) + 0.118(PL) + 0.033F + 5.106G$	7	0.96
[12]	$Y = 1.045SPT + 1.653$	15	0.77
[13]	$Y = 198.63 - 3.78(OMC) - 73.37(MDD) + 0.34D_{60} + 1.64D_{30}$	8	0.97

TABLE II. PREVIOUS AI BASED MODELLING OF CBR OF SOILS

References	Models	No. of Samples
Taskiran [19]	ANN, GEP	151
Varghese et al. [20]	ANN	145
Kumar et al. [21]	ANN (GRNN, MLPN)	60
Sabat [22]	ANN, SVM	49
Yildirim and Gunaydin [23]	ANN	124

samples and the other having a large data set of as many as 389 samples.

There are different machine learning algorithms which include Linear Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Decision Trees, Artificial Neural Network, and so on. which can be used depending on the intended outcome and type of input. Each algorithm has its own advantageous and disadvantageous fields when it is utilized to deal with problems [7].

III. METHODOLOGY

The data used in this study was obtained from tests on 200 soil samples from different sites in the Tshimoyapula area of Central District of Botswana for the Mabeleapudi-Tshimoyapula-Serule road project from 2018 to 2019. The soil samples were tested to determine the liquid limit (LL), plasticity index (PI), and, compaction propertie characteristics (Maximum Dry Density (MDD) and Optimum Moisture Content (OMC)) of each sample. The range of value for some of the more significant parameters are given in Table 3. The results obtained were fed into the Machine Learning algorithms with the index properties as input parameters and the CBR value as the output and the correlations for different inputs against the CBR determined.

ISSN: 2521-2293

TABLE III. SUMMARY OF INPUT VALUES

Input Parameters	No. of Samples	Max	Mean	Min
CBR at 100% compaction	200	76	11.4	1.9
Maximum Dry Density (Kg/m^3)	200	2264	1938.03	1619
Optimum Moisture Content (%)	200	19	10.15	4.7
Liquid Limit (%)	200	53	26.73	16
Plastic Index (%)	200	27	10.77	2.3

A. CALIFORNIA BEARING RATIO

In a CBR test, the penetration is measured by applying the bearing load on the sample using a standard plunger of 50 mm at a rate of 1.25 mm/min. The CBR is expressed as a percentage of the actual load causing the penetration of 2 mm and 5 mm to the standard load on the crushed California Limestone at the same depth. A load penetration curve is then drawn. The load values on the standard crushed stone are shown in Table 4 above. A CBR value equal to 3% is for tilled farmland, while CBR equal to 4.75% for turf as moist clay and moist sand may have a CBR value of 10%. And in high-quality crushed rock, the CBR value is around 80% compared to standard material for this crushed stone which has a value of 100%. The CBR test equation is given as;

$$CBR = \frac{P}{Ps} \times 100 \qquad (1)$$

Where:

P = Pressure applied for site soils (N/mm^2)

Ps = Pressure achieved at the same penetration for California crushed limestone (N/mm^2)

1) 1INPUT PARAMETERS

Soils are classified with different engineering properties which affect the behavior of soil under different conditions. Some of these properties which were used as input parameters are;

- Liquid Limit: The liquid limit (LL) is the water content at which a soil changes from plastic to liquid behavior. At this limit, the soil possesses a small value of shear strength, losing its ability to flow as a liquid. In other words, the liquid limit is the minimum moisture content at which the soil tends to flow as a liquid.
- Plasticity Index: The plasticity Index (PI) is the range of water content within which the soil exhibits plastic properties, that is, it is the difference between liquid and plastic limits. Plasticity Index (PI) = Liquid Limit (LL) -Plastic Limit (PL).
- Maximum Dry Density (MDD): Is the dry density of soil corresponding to the optimum moisture content during compaction.
- Optimum Moisture Content (OMC): Is the water content at which the soil attains maximum dry density.

TABLE IV. LOAD PENETRATION DATA FOR CRUSHED ROCK

Penetration depth (mm)	Load (KN)
2	11.5
2.5	13.24
4	17.6
5	19.96
6	22.2
8	26.3
10	30.3
12	33.5

.

B. MULTIVARIATE REGRESSION ANALYSIS

Multivariate regression was used to assess how linearly connected the different independent variables and different dependent variables were to one another. Due to the correlation between the variables, the relationship is considered to be linear [14]. Multivariate regression was applied to the dataset to forecast the behaviour of the response variable (CBR) based on its related predictor factors (LL, PI, MDD, OMC). The main form of MVRA is expressed mathematically in (2):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad (2)$$

In which, β1, β2, ..., βn are the coefficients of the regression model; β0 is a constant value; X1, X2, Xn are the independent variables and Y is the dependent variable.

C. ARTIFICIAL NEURAL NETWORK

ANNs which falls under the category of supervised learning, a branch of machine learning was used to model both input and output data. Artificial neurons are a group of interconnected nodes that form the basis of an ANN and aim to simulate the operation of biological brain neurons [15]. Like the synapses in a human brain, each link has the ability to send a signal to neighbouring neurons. An artificial neuron can signal neurons that are connected to it after processing signals that are sent to it. The output of each neuron is calculated using a transfer function of the sum of its inputs, and the signal at a connection is a real value. Typically, weights and a bias that changes as learning progresses are used to connect neurons. The weight alters a connection's signal intensity by increasing or decreasing it [16]. A neuron's threshold, which is provided by an activation function, may be

ISSN: 2521-2293

such that a signal is only transmitted if the total signal crosses it. The rectified linear activation function (ReLU), the log sigmoid, and the tan-sigmoid functions are a few examples of activation functions. Neurons are typically aggregated into layers. Three layers make up a basic neural network: the input layer, the hidden layer, and the output layer [17]. When trained on data that has both inputs and outputs, neural networks are able to identify the underlying patterns in the data [18]. The training of a neural network from a given data is usually conducted by determining the difference between the measured and predicted output. This difference is the error and is computed in the output layer. The network then adjusts its weights and bias according to a learning rule using this error value. Successive adjustments cause the neural network to produce the desired output. This is known as error minimisation, and the network is said to have learnt the data very well, when it produces a small error [17]. The equation for the analysis is as follows;

$x_i$represent the inputs, w_{ij} are the weights connecting layer i with layer j, b is the bias weight and n is the number of input units.

IV. RESULTS AND ANALYSIS

Table 5 gives the results of various soil properties from the laboratory experiments on two hundred soil samples which were used for the present investigation. The properties include liquid limit, plasticity index, and compaction characteristics (maximum dry density and optimum moisture content) and CBR conducted at optimum moisture content.

A. REGRESSION ANALYSIS

The various regression analysis between CBR value with respect to the different soil properties are presented in Figures 1 to 4. It shows the linear trend line, which shows the effect of the various soil properties on CBR value. These graphs show the different correlations together with their respective R2 values, with the highest R^2 being maximum dry density at 0.2018 and the lowest being Liquid Limit at 0.0245. Figure 5 shows the sensitivity analysis to support the data. This further shows from the single linear regression, maximum dry density has the greatest effect on the CBR value followed by optimum moisture content, plasticity index and liquid limit in that order. Also, the graphs trendlines show only maximum dry density is proportional to the CBR value and the rest are inversely proportional.

The multiple linear regression analysis was conducted with the results summarized in Table 5. The t-statistic and p-value of multiple linear regression analysis are shown in Table 6. Comparison between the measured CBR values and predicted CBR values from the regression model using a linear relationship shows a model with a correlation coefficient R^2 = 0.249 as in Figure 6.

TABLE V. SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,453867
R Square	0,205995
Adjusted R Square	0,189708
Standard Error	10,15191
Observations	200

TABLE VI. SUMMARY OF MVRA

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-60.69	16.38	-3.70	0.000275
LL	0.137	0.205	0.671	0.503
PI	-0.206	0.216	-0.956	0.340
MDD	0.0371	0.00693	5.35	2.45E-07
OMC	-0.122	0.414	-0.295	0.768

B. ARTIFICIAL NEURAL NETWORK

Table 7 gives a summary of ANN results, in which the model with a Relu Function with a hidden layer with seven number of neurons gives the highest R^2 value of 0.83. The table also shows a model with a Tan-sigmoid Function with a hidden layer with fifteen number of neurons giving R^2 of 0.21. Comparison between the measured CBR values and predicted CBR values from the ANN model using a linear relationship is presented in Figure 7.

TABLE VII. SUMMARY OF ANN RESULTS

Model	Number of Neurons	Activation Function	R2	RMSE
		Relu		
ANN	5		0.69	2.23
	7		0.83	1.24
	13		0.62	3.74
	15		0.51	2.19
		Tan-sigmoid		
	5		0.55	3.69
	7		0.43	3.73
	13		0.34	4.78
	15		0.21	5.01
		Log-sigmoid		
	5		0.33	5.54
	7		0.56	4.36

ISSN: 2521-2293

	13		0.38	5.19
	15		0.22	6.9

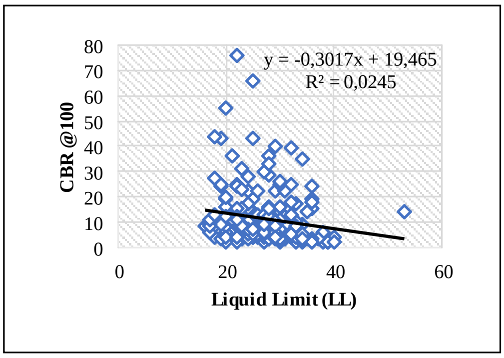

Fig. 1. *Regression model for CBR and LL*

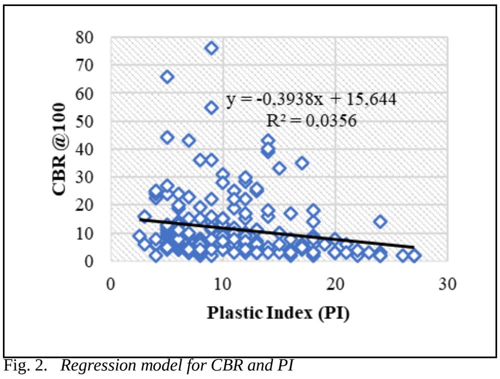

Fig. 2. *Regression model for CBR and PI*

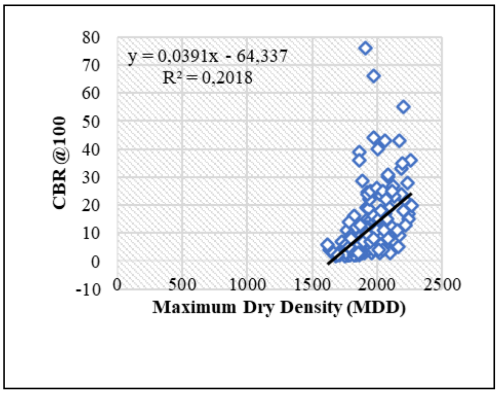

Fig. 3. *Regression model for CBR and MDD*

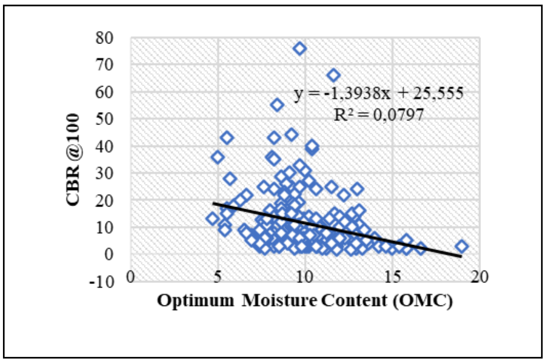

Fig. 4. *Regression model for CBR and OMC*

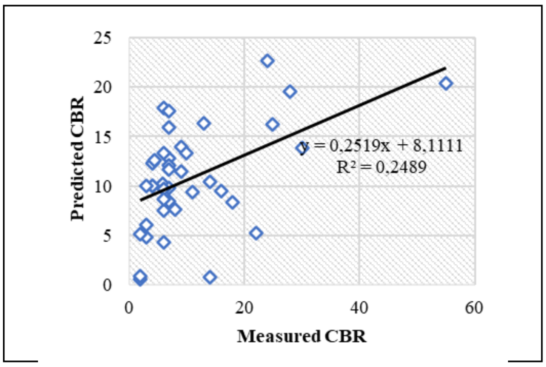

Fig. 5. *MVRA model performance*

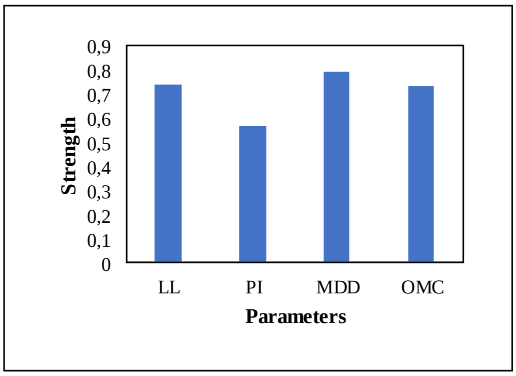

Fig. 6. MVRA model performance

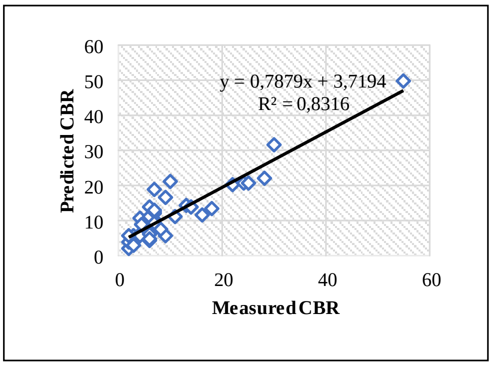

Fig. 7. *ANN Model Performance*

V. CONCLUSION

From the results of the research, the following conclusions are being drawn:

- Based on the above laboratory tests, there isn't any reliable Single linear relationship for predicting CBR value from index properties with relatively low R^2 values.
- The highest coefficient of determination obtained for CBR is 0.2018 with MDD.
- It is observed that CBR values increases with increase in MDD and decreases with increase in LL, PI and OMC.
- The correlation of CBR with LL, PI, MDD and OMC by utilizing MLRA approach gives a relationship with R2 = 0.249
- From the developed correlation, it can be seen that the CBR value is largely dependent on MDD of soil whereas, the effect of the other parameters is lower in comparison.
- The ANN analysis showed that ReLu function model with hidden of seven number of neurons, produced a good relationship with a R^2=0.83
- The results also showed that of the two types of algorithms, ANN has the best output of results

References

[1] O. Bassey Bassey, E. Ambrose, and R. K. Etim "Correlation between CBR Values and IndexProperties of Soils: A Case Study of Ibiono, Oron and Onna in Akwa Ibom State," vol. 7, no. 4, pp. 94–102, 2017,

[2] S. T. Torgano, M. S. Ali, E. T. Yenealem and A. T. Tumato, "CORRELATION BETWEEN CBR VALUES AND INDEX PROPERTIES OF SUBGRADE SOILS: IN THE CASE OF BODITI TOWN," International Journal of Advance Research and Innovative Ideas in Education, vol. 6, no.3, 2020

[3] C. R. Foster and R. G. Ahlvin, "Notes on the Corps of Engineers CBR Design Procedure," Us Army Engineer waterways experiment station, pp. 1-12, 1959.

[4] T. Abdulnabi and Z. Abdulrazzaq, "An Estimated Correlation between California Bearing Ratio (CBR) with some Soil Parameters of Gypseous Silty Sandy Soils," *Tikrit Journal of Engineering Sciences*, vol. 27, no. 1, pp. 58–64, Mar. 2020.

[5] O. S. B. Al-Amoudi *et al.*, "Clegg Hammer-California-Bearing Ratio Correlations Related papers St abilizat ion of a Saudi calcareous marl soil Met hod and Mechanisms of Soil St abilizat ion Using Elect ric Arc Furnace Dust Clegg Hammer-California-Bearing Ratio Correlations," 2002.

[6] N. Puri, H. D. Prasad, and A. Jain, "Prediction of Geotechnical Parameters Using Machine Learning Techniques," in *Procedia Computer Science*, 2018, vol. 125, pp. 509–517.

[7] Z. Chao, G. Ma, Y. Zhang, Y. Zhu, and H. Hu, "The application of artificial neural network in geotechnical engineering," in *IOP Conference Series: Earth and Environmental Science*, Nov. 2018, vol. 189, no. 2.

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

[9] K. V. S. D. Jayamali, U. P. Nawagamuwa, and U. P. Nawagamuwa, "Empirical Correlations Between CBR and Index Properties for Sri Lankan Soils Determination of Age of MSW through Soil Tests View project Index properties of SL soils and CBR View project Empirical Correlations Between CBR and Index Properties for Sri Lankan Soils," 2015.

[10] P. Sharma, S. VSatyanarayana, S. ReddyN, and S. Venkata Satyanarayana, "Prediction of California Bearing Ratio through Empirical Correlations of Index Properties for Tropical Indian Soils," *International Journal of Innovations in Engineering and Technology*, 2020.

[11] P. G. Rakaraddi and V. Gomarsi, "ESTABLISHING RELATIONSHIP BETWEEN CBR WITH DIFFERENT SOIL PROPERTIES," 2015.

[12] T. Abdulnabi and Z. Abdulrazzaq, "An Estimated Correlation between California Bearing Ratio (CBR) with some Soil Parameters of Gypseous Silty Sandy Soils," *Tikrit Journal of Engineering Sciences*, vol. 27, no. 1, pp. 58–64, Mar. 2020.

[13] V. Chandrakar and R. K. Yadav, ""Study of correlation of CBR value with engineering properties and index properties of coarse grained soil.,"" 2016.

[14] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, "MULTIVARIATE DATA ANALYSIS EIGHTH EDITION," 2019.

[15] W. S. Mcculloch and W. Pitts, "A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY* n," 1990.

[16] S. E. Dreyfus, "Artificial neural networks, back propagation, and the kelley-bryson gradient procedure," *Journal of Guidance, Control, and Dynamics*, vol. 13, no. 5, pp. 926–928, 1990.

[17] P. K. Simpson, Artificial neural systems : Foundations, paradigms, applications, and implementations, New York: Pergamon Press, 1989.

[18] Y. Li, Y. Fu, H. Li, and S. W. Zhang, "The improved training algorithm of back propagation neural network with selfadaptive learning rate," in *Proceedings of the 2009 International Conference on Computational Intelligence and Natural Computing, CINC 2009*, 2009, no. 1, pp. 73–76.

[19] Taskiran, T.,. Prediction of California bearing ratio (CBR) of fine grained soils by AI methods, 2010

[20] Varghese, V.K., Babu, S.S., Bijukumar, R., Cyrus, S., Abraham, B.M., Artificial neural networks: a solution to the ambiguity in prediction of engineering properties of fine-grained soils. Geotech. Geol. Eng. 31, 1187–1205, 2013.

[21] Kumar, S.A., Kumar, J.P., Rajeev, J., Application Of Machine Learning Techniques To Predict Soaked CBR Of Remolded Soils, 2, pp. 3019–3024, 2013

[22] Sabat, A.K., Prediction of California bearing ratio of a stabilized expansive soil using artificial neural network and support vector machine. Electron. J. Geotech. Eng. 20, 981–991, 2015.

[23] . Yildirim, B., Gunaydin, O.,. Estimation of California bearing ratio by using soft computing systems. Expert Syst. Appl. 38, 6381–6391, 2011.