Application of Artificial Neural Networks in Predicting Subbase CBR Values Using Soil Indices Data

Shakir Al-Busultan¹,², Gafel Kareem Aswed²,³, Raid R. A. Almuhanna²,⁴, Sajjad E. Rasheed²,⁵

¹ Assist. Prof. Dr, Faculty member, University of Kerbala, Iraq, email: s.al-busultan@uokerbala.edu.iq
² Member of Highway and Transportation Research Association, Iraq
³ Lecturer, Faculty member, University of Kerbala, Iraq, email: gafelkareem@gmail.com
⁴ Lect. Dr, Faculty member, University of Kerbala, Iraq, email: raidalmuhanna@yahoo.com
⁵ Asst. Lecturer, External Faculty member, University of Kerbala, Iraq, email: sejjad90@gmail.com

Abstract. Subbase strength characteristics is one of the main inputs of pavement design, and such strength characteristics are normally represented by indices such as resilient modulus, dynamic modulus, and California Bearing Ratio (CBR), with the latter being a widely used index among pavement and geotechnical engineers. This paper examines the capability of Artificial Neural Networks (ANN) to develop a correlation between subbase CBR and primary soil data, which could help with estimating CBR for prediction purposes and with identifying the significance of each index with regard to subbase strength. Data were sampled from different areas in Karbala, Iraq, and a total of 358 subbase samples were used for model training and validation. The results showed that the proposed ANN model could successfully predict the CBR value using soil index data. Additionally, a sensitivity analysis was conducted to determine the importance of each contributing factor, and within the boundaries of the local subbase characteristics, the test results indicated that soluble salts were the most effective factor among soil parameters with an importance percentage of 39.46%, while the Plasticity Index (PI) was the least important factor, with a percentage of 2.06%. Based on the validity and quality of subbase soil tests, using ANN to predict CBR value may offer a suitable replacement for lengthy and expensive laboratory testing based on validated data for materials supplied from Karbala quarries.

Keywords: ANN; CBR; Neural Power; soil index; SPSS; subbase

1. Introduction

The strength characteristic of the subgrade and unbound materials is one of the most important components in flexible pavement design. The California Bearing Ratio (CBR) is the most common test characterising subgrade and unbounded pavement materials, being the ratio (expressed as a percentage) of the stress needed to penetrate a soil mass with a 50 mm diameter plunger at a rate of 1.25 mm/min to the stress needed for corresponding penetration of a standard material (normally defined as crushed stone[1]). Generally, the ratio is calculated at penetrations of 2.5 and 5 mm; while normally the ratio at 2.5 is considered, when that at 5 mm is reliably greater, the ratio at 5
mm is considered. The corrected load values (at 2.5, 5 mm penetration) are then extracted from the load penetration curve, and the CBR is determined using the following formula [2]

\[
CBR = \frac{\text{Applied stress}}{\text{Standard stress}} \times 100\% \quad \ldots\ldots(1)
\]

The CBR test was originally developed by the California Division of Highways in 1929 in an attempt to eliminate some field loading tests deficiencies and to provide a quick method for comparing the local subbase and base materials available for reinforcing subgrade [3]. The CBR test can be implemented on disturbed or undisturbed soil samples, and the samples can be dry or soaked in water. The CBR test on the standard material (crushed stone) produces a CBR value of 100, known as the Standard Stress. To date, of the available unbound material strength indices such as resilient modulus and dynamic modulus, CBR remains the most widely known and used by paving and geotechnical engineers and designers. Many highway agencies also rely on CBR to predict other soil indices. Thus, many researchers [4-11] have studied the effects of types and characteristics of soil on the values of CBR, examining characteristics such as plasticity index, liquid limit, and direct shear. Simultaneously, some research work has focused on estimating CBR from these other characteristics or indices using traditional modelling techniques.

Artificial Neural Networks (ANN) represent a simplified model of the human brain, featuring a complex communication network that consists of hundreds of simple processing units wired together. Neural Networks thus manage information in similar ways to the human brain. A Neural Network is comprised of numerous interconnected neurons which simultaneously work to solve a specific problem; Neural Networks thus cannot be set to achieve an exact task, and they learn by example; these thus have to be select ed carefully to train the network in a correct mode [12]. A multi-layer perceptron (MLP) format is considered the most well-organised modelling style. An MLP is comprised of input and output layers with one or more hidden layers. MLPs are beneficial in research because of their ability to solve problems stochastically [13].

Although ANN has been applied in almost every aspect of civil engineering, only a limited number of studies have been conducted on predicting the CBR value of soils. Taskiran [14] developed ANN and Gene Expression Programming (GEP) models to predict CBR values for fine grained soils; the results showed that both ANN and GEP could predict the relationships between CBR and soil data. Sabat [15] used ANN and Support Vector Machine (SVM) models to forecast CBR values for expansive soils using a soil index. Suthar and Aggarwal [16] developed a multilayer perception-artificial neural network and multiple regression model to estimate CBR values for stabilised pond ash; their results showed that both models could predict CBR values with a high degree of accuracy. Most attempts have been conducted on subgrade soil, however, and few research attempts have been made on using ANN to predict CBR from soil indices for local subbase soils.

2. Study Aim, Objectives, and Limitations

This paper aims to examine the application of ANN in developing CBR models from subbase soil index data. Appropriate objectives were thus drawn up to achieve this aim: identifying the subbase index data available such as percent passing specific sieves according to specifications, processing and analysing the collected data, building the target model using the ANN technique, and identifying the significance of each index in terms of the CBR values. However, some limitations have constrained the research work and the produced model, including the collection of data from a specific laboratory (Karbala Constructional Laboratory), and the use of subbase type B according to Iraqi specifications as collected from Karbala quarries.

3. Soil Database

In order to obtain a correlation between subbase physical properties and CBR, multiple soil samples (358 samples) were collected. The data was collected from soil samples acquired by Karbala Constructional Laboratory, which were tested for gradation, optimum water content (OWC), maximum dry density (MDD), liquid limit (LL), plasticity index (PI), and percentages of SO₃,
Soluble salt, Gypsum, and Organic materials. The subbase samples were tested according to the American Standards for Testing and Materials [17]. Table 1 presents the statistical parameters of the laboratory test results.

| Parameters          | Min.  | Max.  | Average | Standard Deviation |
|---------------------|-------|-------|---------|--------------------|
| Passing no.2        | 79    | 100   | 97.88   | 2.73               |
| Passing no.1        | 76    | 99    | 90.94   | 4.79               |
| Passing no.3/8      | 51    | 91    | 73.48   | 7.12               |
| Passing no.4        | 43    | 81    | 61.06   | 6.91               |
| Passing no.8        | 37    | 71    | 51.98   | 6.37               |
| Passing no.50       | 10.2  | 41    | 18.39   | 3.85               |
| Passing no.200      | 5.1   | 16.3  | 9.48    | 2.01               |
| M.D.D              | 2.17  | 2.33  | 2.24    | 0.03               |
| O.W.C              | 5     | 7.8   | 6.22    | 0.56               |
| L.L                | 18.7  | 35.5  | 26.02   | 3.09               |
| P.I                | 0.066 | 40    | 6.12    | 3.60               |
| SO₃                | 0.0013| 2.47  | 0.40    | 0.42               |
| Soluble salt       | 0.1   | 7.7   | 1.314   | 1.265              |
| Gypsum             | 0.095 | 6.921 | 0.88    | 0.97               |
| Organic            | 0.091 | 0.533 | 0.19    | 0.07               |
| CBR                | 23.5  | 60    | 49.697  | 6.871              |

4. Pre-Processing and Data Division

It was necessary to exclude any possible outliers from the data. As seen in Figure 1, the Box plot for the CBR showed that the records of samples 32, 256, and 292 were outliers; these were thus excluded from data processing.

![Figure 1. Boxplot for the output (CBR)](image)

ANN modelling begins by dividing the available data into subsets. The best division is made by using an iterating procedure based on the default parameters facilitated by IBM SPSS (version 23) to perform nearest neighbour classification to determine the best number of clusters. A two-step algorithm was used to classify 16 inputs into 3 clusters with fair classification quality, as shown in Figure 2. Cluster 1 contains 49.3% of the data, cluster 2 contains 16.4%, and cluster 3 has 34.3%, as shown in Figure 3.
Two samples were selected from each cluster; the first for training and testing, and the second for validation. When a cluster was comprised of only two records, one record was chosen for training set while the other was chosen for validation set; where a cluster had only one record, this record was chosen for the training set.

Using Neural Power version 2.5, a three-layered Multilayer Perceptron (MLP) feed-forward neural network was trained and tested with 323 cases, which represented 90% of the data, with 35 cases (10% of the data) used for validation. A Quick Prop (QP) algorithm with a learning rate of 0.4 and a 0.8 momentum term was chosen to train the network. The trained ANN resulted in 11 nodes in the hidden layer, one output layer, and a hyperbolic tangent function for both, as shown in Figure 4. This network had the lowest prediction error after the testing stage of 3.124 with average \( R = 95.92\% \) and average \( R^2 = 92\% \), as shown in Figure 5. The connection weights in the adopted ANN model and threshold levels are summarised in Table 2.
5. Validity of the ANN Model

The 35 cases not used to train the network were used to validate the developed model. The inputs of the validation set were used to predict the outputs in terms of CBR, and these values were plotted against the measured (observed) values as shown in Figure 6. From this figure, the generalisation
The capability of ANN techniques for all data sets are within the range of data from the trained ANN. This shows that neural nets enable a strong generalisation ability, which means that once they have been suitably trained, they are capable of offering accurate outcomes even for cases they have not encountered previously. In addition, the coefficient of determination R² was 77.73%, suggesting the created model displays an acceptable agreement with the real observations.

![Figure 6. Generalisation of ANN Model](image)

The statistical parameters used to identify the performance of models include:

- Mean Percentage Error (MPE): one of the most important measures of accuracy of a proposed network, this is the mean of the absolute percentage difference between the predicted and the actual values [18] such that
  \[
  MPE = \left\{ \frac{1}{n} \sum_{j=1}^{n} \left( \frac{A - E}{A} \right) \right\} \times 100
  \]  
  where
  - A = actual observation
  - E = predicted value
  - n = total number of tested cases (35 for validation)

- Root Mean Squared Error (RMSE) [19]
  \[
  RMSE = \sqrt{\frac{\sum_{j=1}^{n} (E - A)^2}{n}}
  \]  

- Mean Absolute Percentage Error (MAPE) [20]
MAPE = \left\{ \sum_{i=1}^{n} \frac{|A-E|}{A} \right\} * 100 \quad (4)

- Average accuracy percentage (AA %) [20]
  
  \[ \text{AA\%} = 100\% - \text{MAPE} \quad (5) \]

- The Coefficient of Determination (R²)
- The Coefficient of Correlation (R), a measure used to define the relative correlation and the goodness-of-fit between the predicted and observed data; however, R² portrays how accurately the model outputs match the target value more effectively.

### Table 3. ANN Statistical Measures

| Description | Statistical parameters |
|-------------|------------------------|
| MPE         | 0.0865 %               |
| RMSE        | 4.314165               |
| MAPE        | 6.42824 %              |
| AA%         | 93.5717 %              |
| R²          | 77.73 %                |
| R           | 88.16%                 |

Based on the statistical analysis of the correlation between the observed and predicted CBR, as seen in Table 3, the validity of the produced model is satisfactory. The MPE is less than 5%, the RMSE (which represents the standard deviation of the error of the predicted CBR) is acceptable at approximately 4.3, the MAPE and AA show acceptable accuracy at 6.4% and 93%, respectively, and the model can explain and represent over 77% of the predicted CBR based on the R² value.

### 6. Sensitivity Analysis of ANN Model Inputs

One of the main objectives of the current research is to identify the influence of each soil data index on CBR value. The results suggest that the index of soluble salts ranks highest, with a relative importance of 39.46%, while the Gypsum index is ranked second at 9.83%, and the maximum dry density (MDD) is ranked third at 6.68 %. However, the other indices have less effect, as summarised in Figure 7. Such results are somewhat surprising, especially the limited effects of MDD, LL, PI, and WC, although these may be attributable to limitations in the input data; such data are constrained within very limited range to pass the requirements of the specification for the soil type used here. It is thus highly recommended that a wider testing programme is enacted to obtain a wider range of subbase soil data to develop a more useful model for raw subbase. Nevertheless, the developed model is valid within the aforementioned limits, as stated in Table 1.
7. Conclusions

The main objective of this study was the examination of ANN in terms of predicting CBR values based on basic soil tests. To establish this goal, soil samples were collected from different quarries in Karbala and tested to obtain input parameters for the ANN model. The following conclusions can be drawn:

1. Subbase data parameters were trained to achieve the best correlations between CBR and such soil parameters. The resulting ANN model is valid in terms of basic statistical criteria.
2. The ANN model is successful in predicting CBR values as determined by actual CBR tests.
3. Sensitivity analysis is a vital in determining the effectiveness of input parameters, and within the input indices’ ranges, Soluble salts is the most important factor, with an importance percentage of 39.46%, while the plasticity index (PI) is the least important factor at 2.06%.
4. The input parameters ranges play a significant role in sensitivity analysis

References

1. Ayodele, A.L., Some Geotechnical Properties of Selected Sub-Base Materials for Road Construction. 2016. 8: p. 31-39.
2. ASTM, D1883, Standard Test Method for CBR (California Bearing Ratio) of Laboratory-Compacted Soils. 2015, ASTM International, 100 Barr Harbor Drive, USA.
3. Porter, O.J., Development of CBR flexible pavement design for airfields. Development of the original method for highway design. Proc. ASCE,, 1949. 75: p. 11-17.
4. Black, W.P.M., *A method of estimating the California bearing ratio of cohesive soils from plasticity data*. Geotechnique, 1962. 12(4): p. 271–282.

5. Agarwal, K.B. and K. D. Ghanekar, *Prediction of CBR from plasticity characteristics of soil*, in *Proceeding of 2nd south-east Asian conference on engineering*. 1970: Singapore. p. 571–576.

6. Seed, H.B. and P.D. Alba, *Use of SPT and CPT tests for evaluating the liquefaction resistance of sands*. in Use of in situ tests in geotechnical engineering, 1986: p. 281–302.

7. Al-Refieai, T. and A. Al-Suhaibani, *Prediction of CBR using dynamic cone penetrometer*. King Saud UJ Eng Sci, 1997. 9(2): p. 191–204.

8. Reddy, C.N.V.S. and K. Pavani, *Mechanically stabilised soils-regression equation for CBR evaluation*, in *Proceedings of the Indian geotechnical conference*. 2006: Chennai, India. p. 731–734.

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

10. Yildirim, B. and O. Gunaydin, *Estimation of California bearing ratio by using soft computing systems*. Expert Syst. Appl., 2011. 38(5): p. 6381–6391.

11. Alawi, M. and M. Rajab, *Prediction of California bearing ratio of subbase layer using multiple linear regression models*. Road Mater. Pavement Des., 2013. 14(1): p. 211–219.

12. Awodele, O. and O. Jegede, *Neural Networks and Its Application in Engineering*. Proceedings of the 2009 InSITE Conference, 2017: p. 1-13.

13. Rostami, F. and A. Jafarian, *A new artificial neural network structure for solving high-order linear fractional differential equations*. International Journal of Computer Mathematics, 2018. 95: p. 528-539.

14. Taskiran, T., *Prediction of California bearing ratio (CBR) of fine grained soils by AI methods*. Adv. Eng. Softw., 2010. 41(6): p. 886–892.

15. Sabat, A.K., *Prediction of California Bearing Ratio of a Stabilized Expansive Soil using Artificial Neural Network and Support Vector Machine*. Electronic Journal of Geotechnical Engineering, 2015. 20(3): p. 981–991.

16. Suthar, M. and P. Aggarwal, *Predicting CBR Value of Stabilized Pond Ash with Lime and Lime Sludge Using ANN and MR Models*. International Journal of Geosynthetics and Ground Engineering, 2018(1).

17. International, A., *Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort (12 400 ft-lbf/ft3 (600 kN-m/m3)).* 2012: PA, West Conshohocken.

18. Dedecker, A.P., et al., *Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium).* Ecol. Modell, 2004. 174(1–2): p. 161–173.

19. Willmott, C.J. and K. Matsuura, *Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance*. Clim. Res., 2005. 30(1): p. 79–82.

20. Vagias, W.M., *Likert-type Scale Response Anchors*. Clemson International Institute for Tourism., Res. Dev. Dep. 2006: Park. Recreat. Tour. Manag. Clemson Univ.,.