Geomechanics and Engineering, *Vol. 33, No. 2 (2023) 183-194*
https://doi.org/10.12989/gae.2023.33.2.183

Prediction of California bearing ratio (CBR) for coarse- and fine-grained soils using the GMDH-model

Mintae Kim[1a], Seyma Ordu[2b], Ozkan Arslan[3c] and Junyoung Ko[*4]

[1]School of Civil, Environmental, and Architectural Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea
[2]Department of Environmental Engineering, Tekirdag Namik Kemal University, Namık Kemal Mahallesi Kampüs Caddesi No:1, Tekirdağ 59030, Turkey
[3]Department of Electronics and Communication Engineering, Tekirdag Namik Kemal University, Namık Kemal Mahallesi Kampüs Caddesi No:1, Tekirdağ 59030, Turkey
[4]Department of Civil Engineering, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea

(*Received November 28, 2022, Revised April 1, 2023, Accepted April 5, 2023*)

Abstract. This study presents the prediction of the California bearing ratio (CBR) of coarse- and fine-grained soils using artificial intelligence technology. The group method of data handling (GMDH) algorithm, an artificial neural network-based model, was used in the prediction of the CBR values. In the design of the prediction models, various combinations of independent input variables for both coarse- and fine-grained soils have been used. The results obtained from the designed GMDH-type neural networks (GMDH-type NN) were compared with other regression models, such as linear, support vector, and multilayer perception regression methods. The performance of models was evaluated with a regression coefficient (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). The results showed that GMDH-type NN algorithm had higher performance than other regression methods in the prediction of CBR value for coarse- and fine-grained soils. The GMDH model had an R^2 of 0.938, RMSE of 1.87, and MAE of 1.48 for the input variables {G, S, and MDD} in coarse-grained soils. For fine-grained soils, it had an R^2 of 0.829, RMSE of 3.02, and MAE of 2.40, when using the input variables {LL, PI, MDD, and OMC}. The performance evaluations revealed that the GMDH-type NN models were effective in predicting CBR values of both coarse- and fine-grained soils.

Keywords: artificial intelligence technology; California bearing ratio (CBR); group method of data handling (GMDH)

1. Introduction

Artificial intelligence (AI) technology is rapidly developing, due to the development of computers and calculations. Recently, numerous studies have been extensively conducted to apply these technologies in the field of geotechnical engineering (Armaghani *et al.* 2020; Liu *et al.* 2020, Luat *et al.* 2020a, b). Especially, the group method of data handling (GMDH) is known as one of the first deep learning methods (Schmidhuber 2015). The GMDH is a type of neural network algorithm that optimizes the structure and parameters of models for computer-based mathematical modeling of multi-parametric datasets. After the GMDH was proposed by Ivakhnenko (1968), it has been developed and used in various applications such as prediction, data mining, knowledge discovery, complex system modeling, optimization, and pattern recognition (Madala and Ivakhnenko 1994). Li *et al.* (2017) reported that GMDH-type neural networks (GMDH-type NN) had better reliability than other classical prediction methods (e.g., single exponential smooth, double exponential smooth, and back-propagation neural network). In addition, the GMDH-type NN has been introduced in civil engineering and is widely used in geotechnical engineering to make better predictions.

The California bearing ratio (CBR) test is a standard method widely used in geotechnical engineering for measuring the strength of road subgrades and pavement materials, as well as, for road sub-bases, soil dams, and bridge piers (Bowles 1992). The CBR test (American Association of State Highway and Transportation Officials 2021, American Society for Testing and Materials 2021) involves the penetration of a standard piston into the soil, either compacted or uncompacted, at a speed of 1.27 mm/min. The CBR value is calculated as the ratio of the measured pressure and the standard crushed rock bearing capacity. This value provides valuable information about the bearing capacity of the subgrade, and its results can be compared to standard test curves to offer insights into geotechnical engineering (Yildirim and Gunaydin 2011).

In highway construction, geotechnical and transportation engineers use the CBR test to determine the condition of the road subgrade and design the road sub-base and flexible pavement (in terms of material types and layer heights) accordingly. However, geotechnical and transportation engineers have faced many difficulties in

*Corresponding author, Assistant Professor
E-mail: jyko@cnu.ac.kr
[a]Research Professor
[b]Assistant Professor
[c]Assistant Professor

http://www.techno-press.org/?journal=gae&subpage=7
ISSN: 2005-307X (Print), 2092-6219 (Online)

finding the design CBR values. The geotechnical properties of the subgrade may vary along the highway due to its geological origin. Thus, the CBR test is required to be conducted frequently along the road, but determining CBR values is a complex and time-consuming process, so the number of test sites is often limited (Nagaraj and Suresh 2018). Furthermore, performing the CBR test in the laboratory is costly, tedious, and labor-intensive because it takes about a week to evaluate accurate CBR values with the optimum moisture content (OMC), maximum dry density (MDD), and required compaction energy. All of these issues have the potential to slow down project progress and increase project expenses. Therefore, it is essential to forecast the CBR value of the soil with simply ascertainable characteristics in order to get around these challenges (Alam *et al.* 2020).

Prior to the development of deep learning technology in geotechnical engineering, a lot of efforts have been made to predict CBR values. In the past years, many studies have been conducted on the correlation of CBR values with different soil parameters. Agarwal and Ghanekar (1970) tried to develop the equation between CBR values and liquid limit (LL), plastic limit (PL), and plastic index (PI) but could not develop a substantial correlation between these parameters. Instead, they developed the correlation of the CBR value with OMC and LL. This correlation equation is shown below:

$$CBR = 2.0 - 16.0 \times \log(OMC) + 0.07 \times LL \quad (1)$$

The National Cooperative Highway Research Program (NCHRP 2001) proposed equations that can predict the CBR value based on the grain size distribution characteristics and Atterberg limit parameters in both coarse - and fine-grained soils. The best-fitted equation proposed by the NCHRP for the CBR value of coarse-grained soils is as shown below

$$CBR_{coarse} = 28.09(D_{60})^{0.358} \quad (2)$$

where D_{60} = the particle size that corresponds to 60% passing.

The NCHRP recommended equation for the CBR value of fine-grained soils is as shown below

$$CBR_{fine} = \frac{75}{1 + 0.728(wPI)} \quad (3)$$

where w = the percentage of sample passing No. 200 sieve.

After introducing deep learning technology in geotechnical engineering, several studies have been conducted to forecast CBR values through various deep learning methods (Yildirim and Gunaydin 2011, Alawi and Rajab 2013, Harini and Naagesh 2014, Rakaraddi and Gomarsi 2015, Taha *et al.* 2019, Alam *et al.* 2020, Tenpe and Patel 2020, Hao and Pabst 2022). In the early stages, many studies focused on a comparative evaluation of artificial neural network (ANN) models with conventional regression models. Yildirim and Gunaydin (2011) predicted CBR values using simple regression analysis (SRA), multiple regression analysis (MRA), and ANN models with the results of soil sieve analysis, MDD, and OMC. The reliability of the models was demonstrated as to how soil properties affect CBR values with the help of statistical parameters. Similarly, Alawi and Rajab (2013) utilized multiple linear regression (MLR) and ANN models with the results of the Los Angeles (LA) abrasion test, OMC, and MDD. However, Harini and Naagesh (2014) presented that MLR models were no more accurate than ANN models in predicting CBR values in fine-grained soil. Further studies investigated the effect of input parameters on the prediction of CBR values using predictive models as well as validation of the models. Rakaraddi and Gomarsi (2015) found that LL values of soil had a significant impact on finding wet CBR values. Katte *et al.* (2019) used various mathematical formulas to estimate CBR values from the basic properties of soil using single linear regression (SLR) and MLR analysis. In addition, as the number of soil samples increased, better predictions of CBR values could be obtained. Compared to SLR and MLR models, ANN models gave more reliable results in predicting CBR values in saturated conditions (Taha *et al.* 2019). Alam *et al.* (2020) carried out both wet and dry CBR tests and established a correlation between soil index properties (e.g., specific gravity (G_s), coefficient of uniformity (C_u), coefficient of curvature (C_c), LL, PL, PI, OMC, MDD, and CBR values) In their study, the tested CBR values were compared with the predicted CBR values using gene expression programming (GEP), ANN, and Kriging methods. Tenpe and Patel (2020) compared the performance of ANN, random forest (RF), and decision tree (DT) models in predicting CBR values from index properties of soil. Hao and Pabst (2022) used a deep learning method called convolutional neural network (CNN) to predict CBR values from soil images. The results showed that the CNN model had high accuracy in predicting CBR values and outperformed traditional machine learning models.

Previous studies have explored multiple models to improve predictions of CBR values, yet none have emerged as the standout choice. All tested models have demonstrated sufficient predictive capability compared to actual CBR values; however, it was important to designate models with soil properties related to appropriate conditions as an independent variable. Leveraging the power of AI technology to develop new models using GMDH-type NN, a nonlinear regression method, holds the potential for increased efficiency in predicting CBR values. As this artificial neural network-based method enables examining the relationship between dependent and independent variables, the most significant feature of the GMDH algorithm, a self-organizing network model, is the automatic selection of the necessary input variables for the formation of the polynomial equation in modeling. Therefore, the objective of this study is to predict CBR values for coarse- and fine-grained soils using the GMDH-type NN. Due to different soil structures, prediction models were separately designed with different input variables for coarse- and fine-grained soils. The results of the GMDH-type NN were compared with those of linear regression (LR), support vector regression (SVR), and multi-perception regression (MLP), and the performance of each method was evaluated using the correlation coefficient (R^2), root-mean-square error (RMSE), and mean absolute error

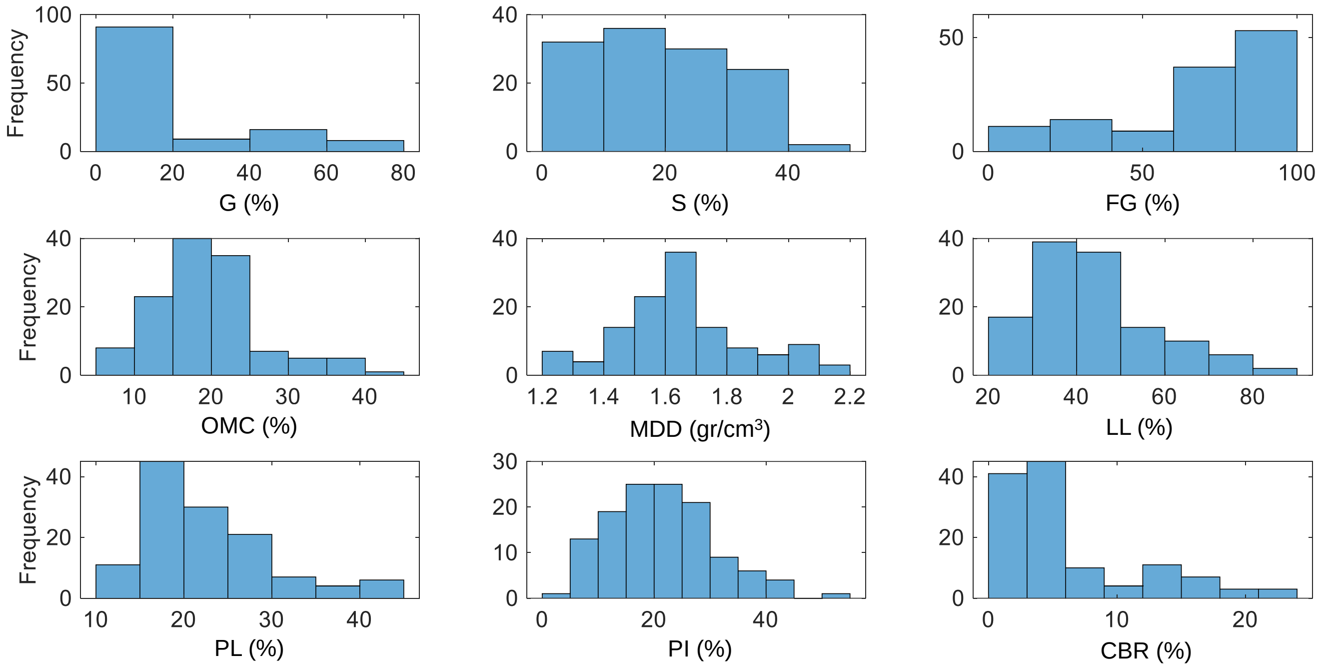

Fig. 1 The frequency histograms of the soil parameters and CBR

Table 1 Description statics of the dataset

	Minimum	Maximum	Mean	Std. Dev.
G (%)	0	78	13.05	21.51
S (%)	0.9	49	18.50	11.13
FG (%)	10	99.10	68.44	26.98
OMC (%)	7.20	40.20	19.51	6.78
MDD (gr/cm^3)	1.21	2.18	1.65	0.20
LL (%)	20	89	43.02	13.78
PL (%)	11	43	22.33	7.29
PI (%)	3	52	20.69	9.44
CBR (%)	0	23	6.15	5.44

(MAE). Moreover, the developed polynomial equations with various combinations of independent variables were presented in order to predict CBR values for coarse- and fine-grained soils.

2. Dataset and procedures

A total of 124 datasets were collected from previous studies and used in this study (Satyanarayana Reddy and Pavani 2006, Yildirim 2009, Yildirim and Gunaydin 2011, Venkatasubramanian and Dhinakaran 2011, Ramasubbarao and Sankar 2013, Rakaraddi and Gomarsi 2015, Nagaraj and Suresh 2018, Khatri *et al.* 2019, Ravichandra *et al.* 2019, Tenpe and Patel 2020, Montgomery *et al.* 2021). The dataset containing 124 data is divided into training (80%) and testing (20%) sets. The performance of the prediction models was evaluated on the testing set. The dataset consists of gravel contents (G), sand contents (S), fine grain ratio (FG), OMC, MDD, LL, PL, PI, and CBR values. Table 1 shows the statistical values of the soil parameters and CBR values that make up the dataset content.

For coarse-grained soils, G, S, FG, OMC, and MDD were used as independent variables, and LL, PL, PI, OMC, and MDD were chosen as input parameters for fine-grained soils. Moreover, various combinations of independent variables for both coarse- and fine-grained soils were used in the development of models for the prediction of CBR values.

Fig. 1 presents the frequency histograms showing the distribution of the parameters. Also, the correlation between CBR values in the dataset and soil parameters is shown in Fig. 2. This figure shows that the correlation between CBR values and parameters plays a major role in obtaining mathematical prediction models. That is, a high positive or negative correlation with the CBR value indicates that designing the prediction model with these input variables gives more accurate and reliable results. As shown in Fig. 2,

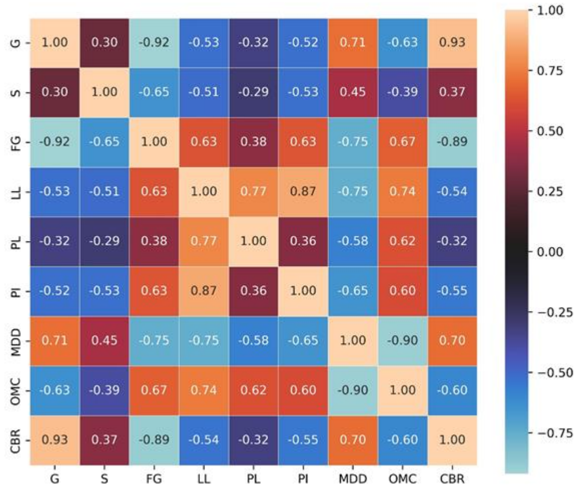

Fig. 2 The correlation between CBR values and soil parameters

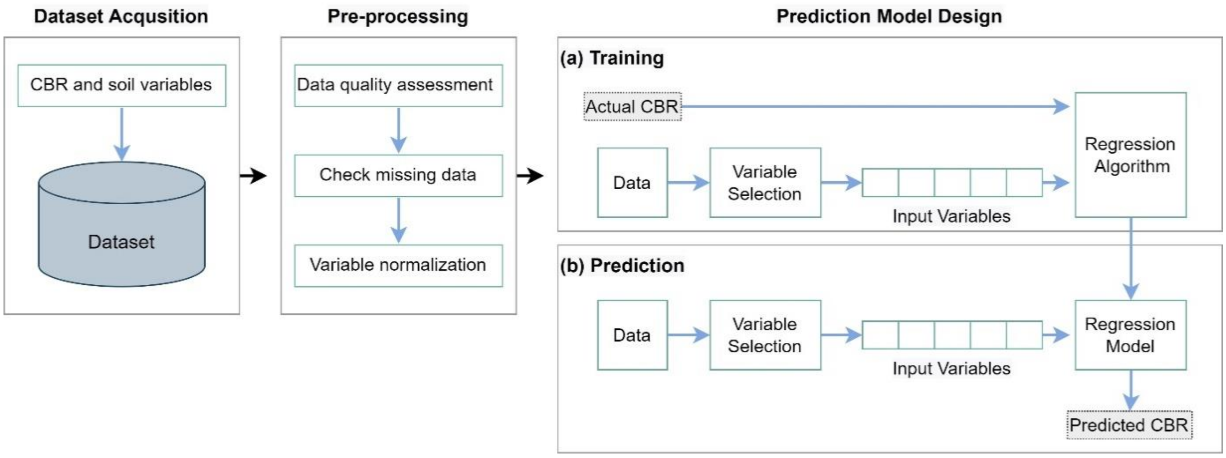

Fig. 3 Flowchart of the proposed regression model for the prediction of CBR

the CBR values had the highest correlation of 0.93 with the G parameter and the lowest correlation of -0.32 with the PL parameter. Moreover, the CBR value had the highest correlation of -0.89 with the FG parameter in the opposite direction.

Fig. 3 presents the methodology and flowchart of the proposed regression model for the prediction of CBR in this study. In prediction procedures, the pre-processing phase was the first step for the dataset. During this phase, data normalization was initially required, and then the independent variables were determined. After selecting the independent variables, high-performance prediction models were obtained by using performance metrics. Here, different combinations of variables were used as inputs in regression algorithms. Then, using the trained model, the most appropriate regression model for CBR estimation was obtained as a result of the test process.

3. Analysis methods

The regression method plays a critical role in the prediction of CBR value. Input parameters, weight coefficient, and regression function type affect the prediction success. Therefore, the GMDH-type NN models with various combinations of input parameters were compared with various regression methods (e.g., LR, SVR, and MLP) in the CBR prediction. The various machine learning-based models that enable finding a relationship between CBR values and input parameters in this study are given below.

3.1 Group Method Data Handling type neural network (GMDH-type NN)

The GMDH-type NN is one of the best models of prediction for problems with complex structures. In MLR and ANN models, all independent variables in the input layer are used in an individual cell in the hidden layer; however, using all inputs in all cells in the network structure may cause overtraining and performance degradation. Difficulties and deficiencies arise in adjusting the bias and weight coefficient, especially when dealing with small-sized data sets. Thus, the GMDH-type NN, which is a self-organizing network model that behaves according to input data, is preferred because all inputs and cells are in all layers (Kondo 1998, Farlow 2020).

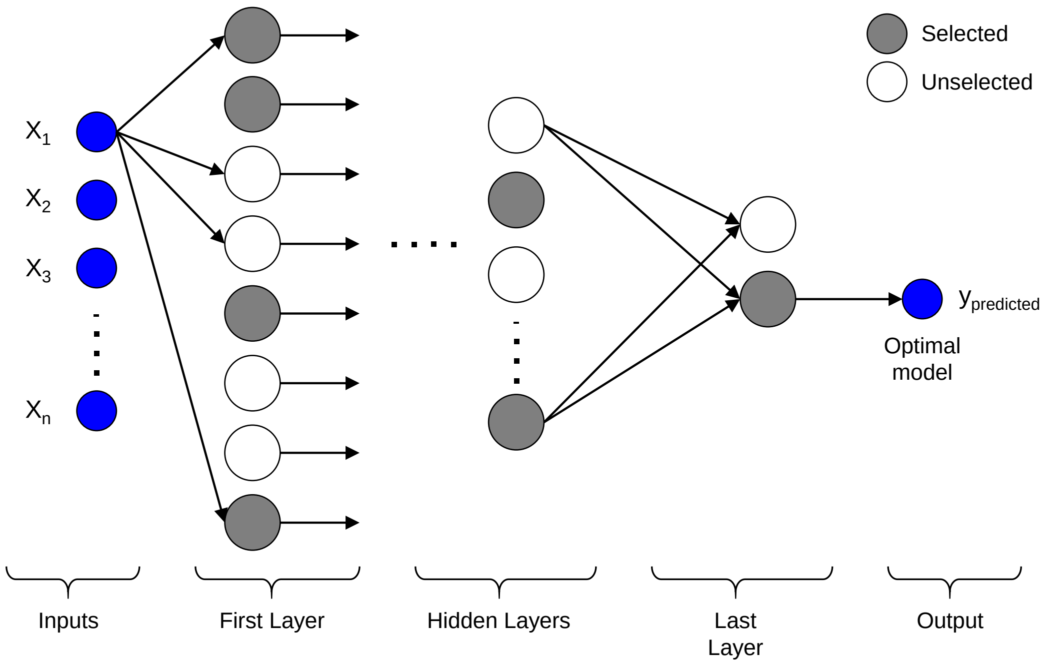

Fig. 4 Schematic structure of GMDH-type NN model (after Kim *et al.* 2022)

The GMDH-type NN is a multilayered structure and uses only the cells that can provide the most effective and accurate results. Each layer consists of independent cells that are used in pairs. Fig. 4 depicts the schematic structure of GMDH-type NN model.

This network model uses a quadratic polynomial function as the activation function. The cells in all layers work independently of each other, and only the outputs that minimize the error rate are preferred among the outputs of the previous layer. Thus, instead of using the cells in all layers, a multilayer neural network model consisting of the optimal cells is created (Elbaz *et al.* 2021).

The GMDH-type is used as a model that matches the given input vector, $X = (x_{i1}, x_{i2}, \cdots, x_{in})$ to the predicted output. The output, $\bar{y}$, is predicted from the inlet vector for M pieces of results obtained for the data pairs in the multi-input, s, and the network model is shown as follows (Oh and Pedrycz 2002, Ardakani and Kordnaeij 2019)

$$\bar{y_i} = \bar{f}(x_{i1}, x_{i2}, \cdots, x_{in}) \;\; i = 1,2,\cdots,M \tag{4}$$

The least-squares method is applied between the actual outputs y_i and the predicted outputs $\bar{y_i}$ to determine the GMDH model. This process is called self-organization of the models, and the cells, where the errors calculated using the least-squares method are minimized, are selected

$$\sum_{i=1}^{M}(\bar{f}(x_{i1}, x_{i2}, \cdots, x_{in}) - y_i)^2 \rightarrow \text{minimum} \tag{5}$$

GMDH-type neural networks are defined as the relationship between input and output parameters expressed in the form of a gradually complicated Kolmogorov–Gabor polynomial function. The Kolmogorov–Gabor function, which is expressed as a nonlinear function form, is defined as below (Oh and Pedrycz 2002)

$$\bar{y} = \alpha_0 + \sum_{i=0}^{n}\alpha_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\alpha_{ij}x_i x_j x_k + \cdots \tag{6}$$

Here, α shows the quadratic polynomial coefficients and $(i, j, k) \in (1, 2, \ldots, n)$.

The Kolmogorov–Gabor polynomial, which generally gives the form of a nonlinear polynomial, is expressed in the form of a quadratic (second-degree) polynomial containing only two variables as follows

$$\bar{y} = G(x_i, x_j) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_j + \alpha_3 x_i x_j + \alpha_4 x_i^2 + \alpha_5 x_j^2 \tag{7}$$

The GMDH-type NN predicts the output for each set of x_i and x_j, and it is used to predict the coefficients to minimize α_i the mean square error (MSE) between the predicted and actual output. With M to show the total number of data, the MSE between the predicted and actual output is minimized as follows

$$E = \frac{\sum_{i=1}^{M}(\bar{y_i} - y_i)^2}{M} \tag{8}$$

In the GMDH-type NN algorithm, the regression structure is established using the polynomial form given in Eq. (4) to obtain actual output data y_i, $i = 1,2,\ldots,M$ of all binomial probability of the independent variables for n piece of inputs as a total. In this network model, after calculating the coefficient vector of the quadratic polynomial, a selection criterion, the objective function (*OF*), is used to eliminate the cells that increase the error rate, and *OF* is expressed as follows

$$OF = \frac{1}{n}\sum_{i=1}^{n}(y_{pre} - y_{mea})^2 \tag{9}$$

Here, y_{pre} and y_{mea} show the predicted and actual outputs, respectively, and n represents the total number of data.

3.2 Linear regression (LR)

Linear regression modelling is one of the simplest regression methods to predict a parameter (Montgomery *et al.* 2021). It is often preferred owing to its simple and useful mathematical structure. This regression method provides the mathematical equation of the target parameter to be predicted using a slope and intercept value. The linear regression expresses the relationship between the target parameter CBR and the input parameters as follows

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n \tag{10}$$

Here, Y shows the output, x_i, x_2, …, x_n input parameters and a_i, a_2, …, a_n the coefficient of the model.

3.3 Support vector regression (SVR)

The SVR method works based on the principle of minimizing the errors in the borderline by separation in the dataset. The SVR creates an optimal hyperplane between the data points and provides curve fitting with the maximum number of data (Smola and Scholkopf 2004, Awad and Khanna 2015). This regression model, which is obtained by training the data, typically uses a radial basis as the kernel function.

3.4 Multilayer perception regression (MLP)

MLP method is one of the methods used in the early stages of deep learning. Thus, the MLP consisting of multiple sensors is also called a deep ANN. MLP neural networks consist of an input layer to which input parameters are applied, an output layer that predicts the input, and an arbitrary number of hidden layers that are a computational tool between the input and output layer (Murtagh 1991). MLP regression method is typically used in supervised learning applications. In this method, which is based on the training of input and output parameters, a model that provides a correlation between input and output is learned. In the training phase, the parameters and weight coefficient of the model minimizing the error are obtained.

3.5 Performance evaluation criteria

The performance of regression models is evaluated by determining the error rate of the predictions from the model. Moreover, the suitability of the regression line to the data set is also used as a criterion in performance evaluation. In the evaluation of the regression models, the correlation coefficient (R^2), root-mean-square error (RMSE), and mean absolute error (MAE) of the predicted and actual target parameters are calculated. The R^2 value is between 0 and 1, and a greater value shows that there is a better fit between the predicted and actual values. In addition, lower values of RMSE and MAE indicate better reliability. R^2, RMSE, and MAE are valuable criteria to determine the best fit of the model for the dependent variables. The R^2, RMSE, and MAE are mathematically expressed as follows

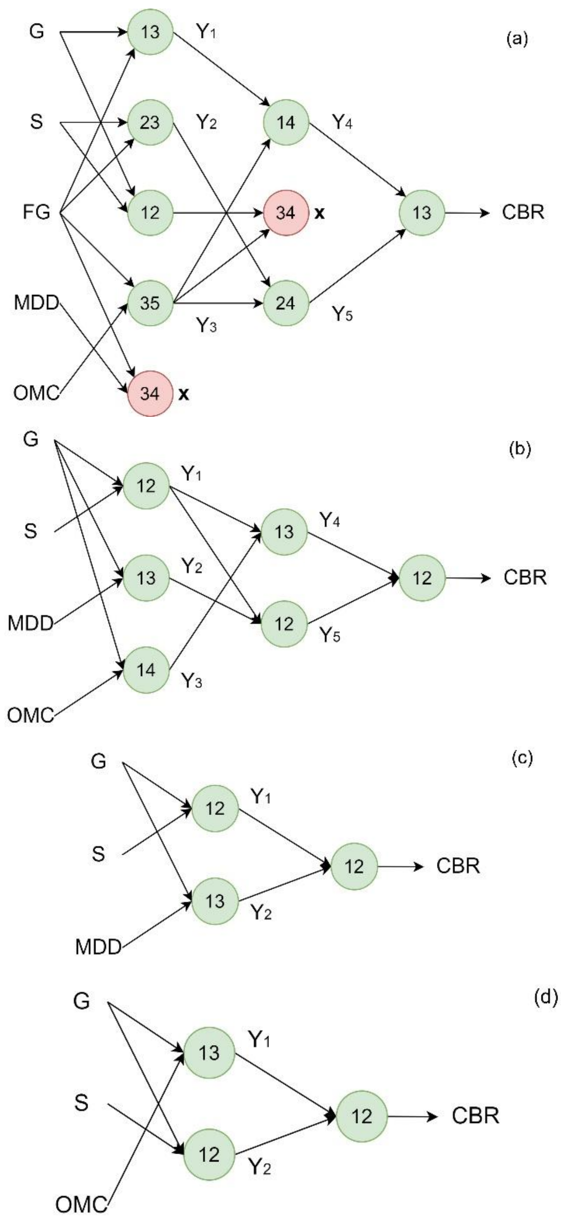

Fig. 5 GMDH-type NN model for coarse-grained soils with different combinations of input parameters: (a) G, S, FG, MDD, and OMC, (b) G, S, MDD, and OMC, (c) G, S, and MDD, and (d) G, S, and OMC

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_{mea} - y_{pre})^2}{\sum_{i=1}^{N}(y_{mea} - y_m)^2} \tag{11}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_{mea} - y_{pre})^2}{N}} \tag{12}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_{mea} - y_{pre}\right| \tag{13}$$

Here, y_{mea}, y_{pre}, and y_m show the average of actual output, predicted output, and actual output, respectively, and N represents the total number of data.

Table 2 Results of performance evaluation for the CBR prediction models for coarse-grained soil

Input parameters	Method											
	LR			SVR			MLP			GMDH-type NN		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
G, S, FG, MDD, OMC	0.89	1.98	1.46	0.85	2.27	1.67	0.89	1.98	1.63	0.941	1.85	1.52
G, S, MDD, OMC	0.89	1.98	1.46	0.84	2.35	1.70	0.88	2.01	1.63	0.937	1.88	1.51
G, S, MDD	0.89	2.00	1.46	0.84	2.34	1.70	0.88	2.03	1.62	0.938	1.87	1.48
G, S, OMC	0.88	2.02	1.47	0.84	2.34	1.64	0.90	1.95	1.58	0.938	1.87	1.50

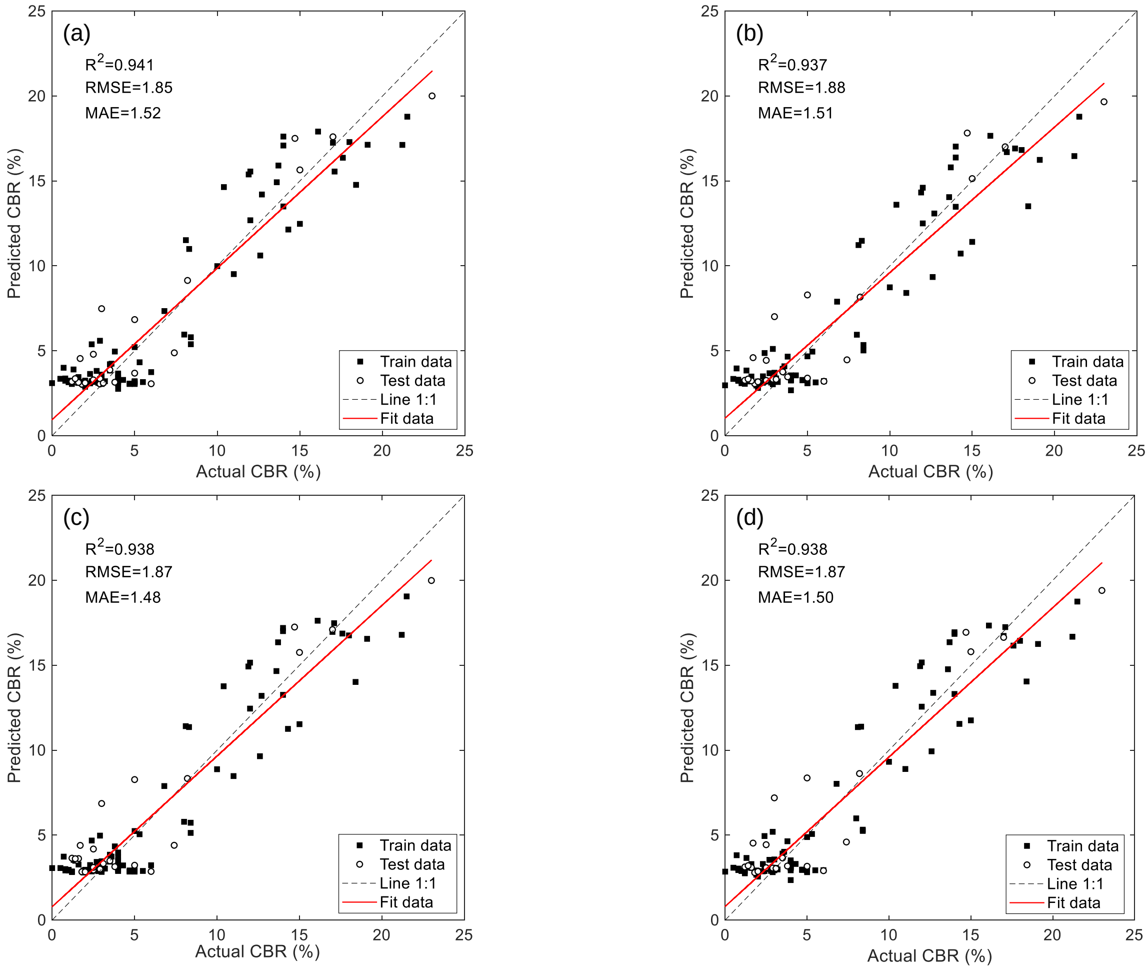

Fig. 6 Comparisons between the actual and predicted values of CBR for coarse-grained soils with different combinations of input parameters: (a) G, S, FG, MDD, and OMC, (b) G, S, MDD, and OMC, (c) G, S, and MDD, and (d) G, S, and OMC

4. Results and discussion

The GMDH-type NN models were created with various numbers of layers and various numbers of neurons in each layer. As computational cost increases as the number of hidden layers and neurons increases, the maximum number of hidden layers was chosen as 5, and the number of neurons as a maximum of 10. Fig. 5 shows the GMDH-type NN models designed for predicting CBR values of coarse-grained soils with different combinations of input parameters.

As a result, the comparisons between the actual and predicted values of CBR parameters for coarse-grained soils are given in Fig. 6 for the GMDH-type NN with different input parameters. In the trials for the testing dataset, the regression coefficients and error rates of the GMDH models were close to each other.

Table 2 shows the results of performance evaluation for LR, SVR, MLP, and GMDH-type NN methods with different independent variables in the prediction of CBR parameters for coarse-grained soils.

As presented in Table 2, the GMDH NN gives the highest regression coefficient (Averaged $R^2 = 0.94$) and the lowest error rates compared with other methods in all conditions for the testing set. The results show that the GMDH method is more effective in using the optimal input parameters and neuron outputs that minimize the error in CBR parameter prediction. As previously mentioned, the most important feature of the GMDH-type NN is the automatic selection of the necessary input variables to create the polynomial equation in modeling. Polynomial equations enable the establishment of the necessary mathematical connection between the input and output variables. Thus, polynomial functions were designed and developed with various input variables to predict the CBR value.

Although there was no significant difference between the models designed with different combinations of input variables considering R^2, RMSE, and MAE, the most influential factors for the prediction of CBR values were determined as G, S, MDD, and OMC parameters for coarse-grained soils based on the correlation between CBR values and other parameters (see Fig. 2). Moreover, FG was found to be the most inefficient variable. The variable pairs formed by the most effective variables were determined as {G, S}, {G, MDD}, and {G, OMC}. The use of these variables yielded polynomial functions for the estimation of the CBR value for coarse-grained soils. The least-squares error method was used for the calculation of the weight coefficients.

The different pairs in the relevant layer determined an error for all combinations in the GMDH-type NN. The variable couple output that generated the minimum error was taken to the next layer. In the prediction model, the cells that reduce the error rate are highlighted in green. These cells transmit the binary variable outputs from the input layer to the next layer. Cells that increase the error rate are highlighted in red and these cells are symbolized with "x" so that the cell outputs are not transmitted to the next layer by the model. Thus, the use of cell output is provided, which reduces the error rate of the prediction model. The polynomial equations of the different couples that make up the least error in each layer were obtained. Pairs of polynomials with given equations were determined using the cells displayed in green and obtained from the models shown Fig. 5. Thus, the polynomial equations of the GMDH models designed for predicting the CBR value in coarse-grained soils are as follows:

{G, S}, {G, FG}, {S, FG}, {FG, OMC} and {FG, MDD} input variable pairs and its polynomial equations

$$Y_1 = 18.95 - 0.015G - 0.3704FG + 0.0006G^2 + 0.0022FG^2 + 0.0029G \cdot FG$$
$$Y_2 = 23.09 - 0.0978S - 0.1800FG + 0.0006S^2 - 0.0002FG^2 - 0.0018S \cdot FG$$
$$Y_3 = 23.98 - 0.4560FG + 0.0799OMC + 0.0021FG^2 - 0.0046OMC^2 + 0.002FG \cdot OMC$$
$$Y_4 = -0.47 + 0.9595Y_1 + 0.2218Y_3 - 0.1878{Y_1}^2 - 0.2016{Y_3}^2 + 0.3815Y_1 \cdot Y_3$$
$$Y_5 = -0.47 + 0.9595Y_2 + 0.2218Y_3 - 0.1878{Y_2}^2 - 0.2016{Y_3}^2 + 0.3815Y_2 \cdot Y_3$$
$$CBR = \left[0.00001 + 0.1051Y_4 - 0.1051Y_5 + 2.7507\,{Y_4}^2 - 2.7523\,{Y_5}^2 + 0.0015\,Y_4 \cdot Y_5\right] \cdot 10^9$$

{G, S}, {G, MDD} and {G, OMC} input variable pairs and its polynomial equations

$$Y_1 = 18.95 - 0.015G - 0.3704FG + 0.0006G^2 + 0.0022FG^2 + 0.0029G \cdot FG$$
$$Y_2 = 23.09 - 0.0978S - 0.1800FG + 0.0006S^2 - 0.0002FG^2 - 0.0018S \cdot FG$$
$$Y_3 = 23.98 - 0.4560FG + 0.0799OMC + 0.0021FG^2 - 0.0046OMC^2 + 0.002FG \cdot OMC$$
$$Y_4 = -0.47 + 0.9595Y_1 + 0.2218Y_3 - 0.1878{Y_1}^2 - 0.2016{Y_3}^2 + 0.3815Y_1 \cdot Y_3$$
$$Y_5 = -0.47 + 0.9595Y_2 + 0.2218Y_3 - 0.1878{Y_2}^2 - 0.2016{Y_3}^2 + 0.3815Y_2 \cdot Y_3$$
$$CBR = \left[0.00001 + 0.1051Y_4 - 0.1051Y_5 + 2.7507{Y_4}^2 - 2.7523{Y_5}^2 + 0.0015Y_4 \cdot Y_5\right] \cdot 10^9$$

{G, S} and {G, MDD} input variable pairs and its polynomial equations

$$Y_1 = 3.29 - 0.15G - 0.076S + 0.0007G^2 + 0.003S^2 + 0.002G \cdot S$$
$$Y_2 = 21.27 - 0.11G - 21.85MDD - 0.0009G^2 + 6.49MDD^2 + 0.089G \cdot MDD$$
$$CBR = -0.26 + 0.37Y_1 + 0.71Y_2 + 0.07{Y_1}^2 + 0.023{Y_2}^2 - 0.11Y_1 \cdot Y_2$$

{G, S} and {G, OMC} input variable pairs and its polynomial equations

$$Y_1 = -0.024 + 0.4G + 0.22OMC - 0.0015G^2 - 0.003OMC^2 - 0.005G \cdot OMC$$
$$Y_2 = 3.11 + 0.18G - 0.053S + 0.0003G^2 + 0.0022S^2 + 0.0016G \cdot S$$
$$CBR = -0.39 + 0.59Y_1 + 0.54Y_2 - 0.054{Y_1}^2 - 0.0034{Y_2}^2 + 0.051Y_1 \cdot Y_2$$

Similarly, several regression analyses were performed with LR, SVR, MLP, and GMDH-type NN for different combinations of input variables to predict CBR values for fine-grained soils. For fine-grained soils, input variables such as LL, PL, PI, MDD, and OMC were employed to predict the CBR value. Fig. 7 shows the GMDH-type NN models designed to predict the CBR value for fine-grained soils with various combinations of input parameters

According to the input variable pairs, the GMDH-type models were created with various numbers of layers and various numbers of neurons in the individual layer. The comparisons between the actual and anticipated values of the CBR parameter in fine-grained soils are presented in Fig. 8 for training and testing sets on the designed GMDH-type models with various combinations of input variables.

Table 3 shows the performance evaluation results obtained using LR, SVR, MLP, and GMDH-type NN methods for different independent input parameters in the prediction of CBR parameters for fine-grained soils. In the trials for the whole data set, the R^2 values of GMDH-type NN models were calculated in the range of 0.780 – 0.829, and the error rates were similar for all combinations of input variables. Compared to other regression methods in all

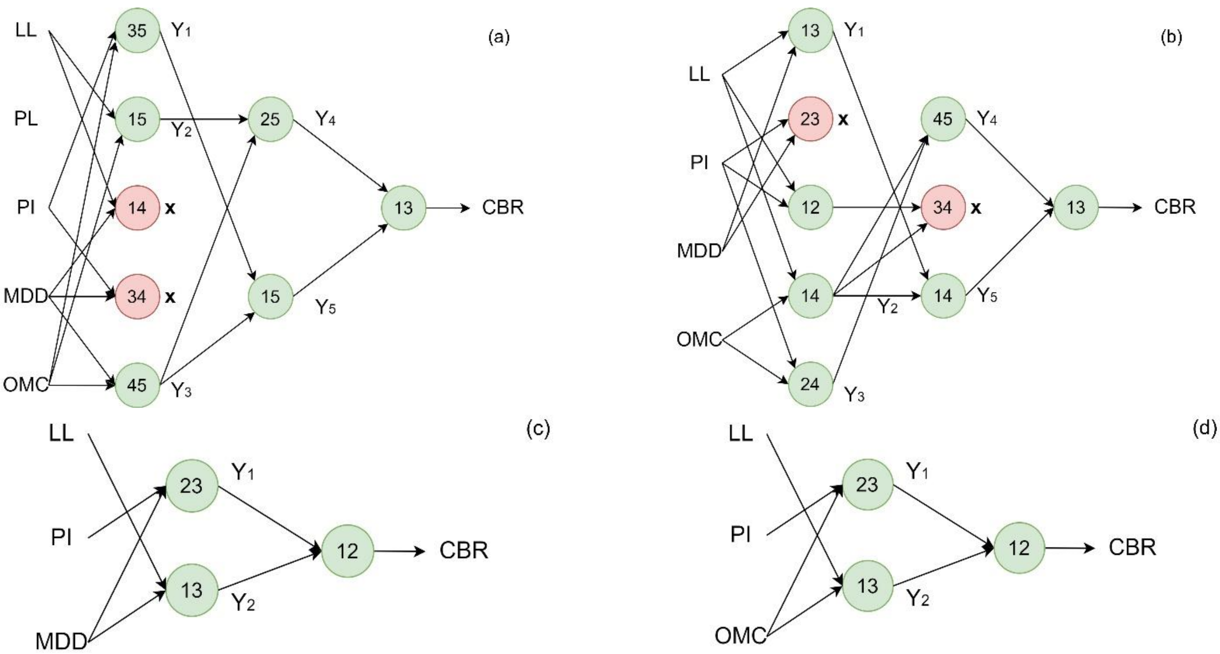

Fig. 7 GMDH-type NN model for fine-grained soils with different combinations of input parameters: (a) LL, PL, PI, MDD, and OMC, (b) LL, PI, MDD, and OMC, (c) LL, PI, and MDD, and (d) LL, PI, and OMC

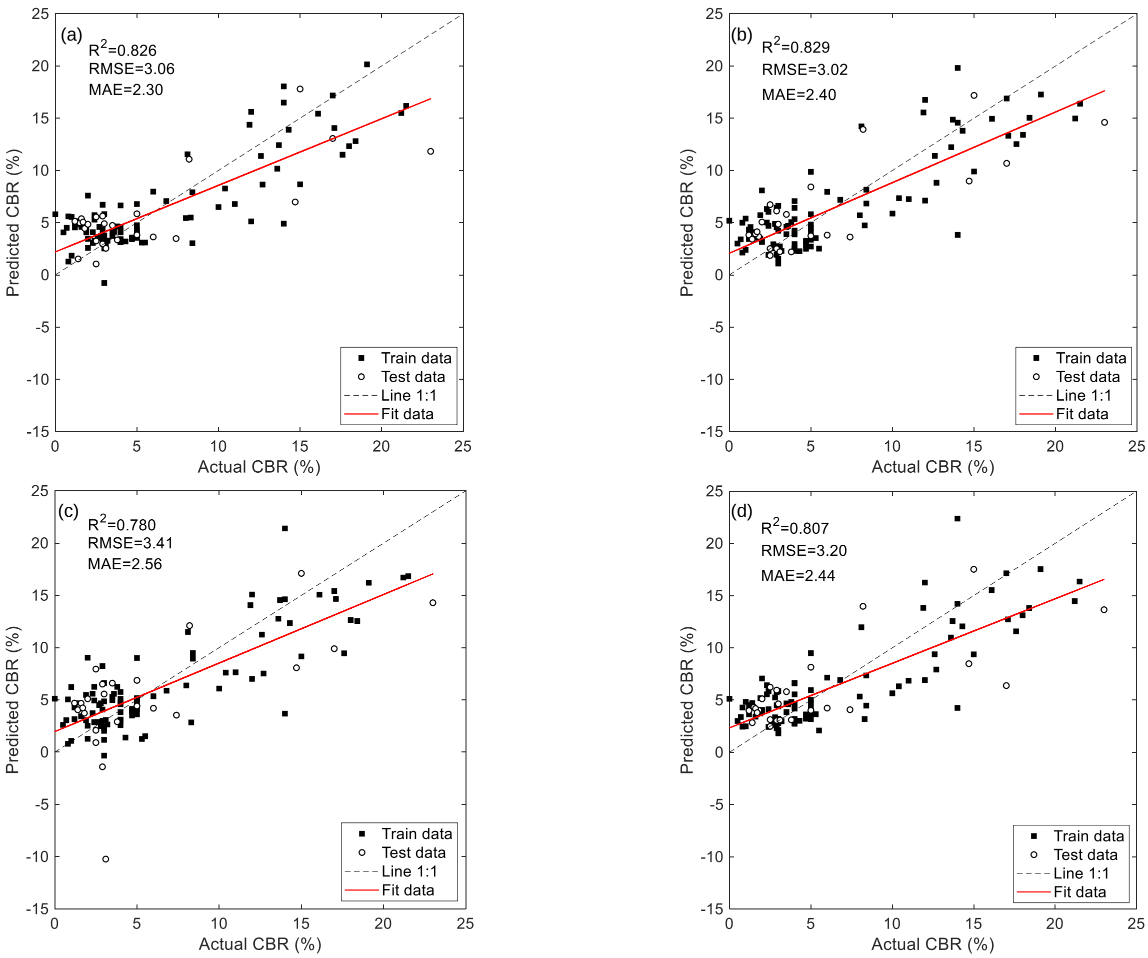

Fig. 8 Comparisons between the actual and predicted values of CBR for fine-grained soils with different combinations of input parameters: (a) LL, PL, PI, MDD, and OMC, (b) LL, PI, MDD, and OMC, (c) LL, PI, and MDD, and (d) LL, PI, and OMC

Table 3 Results of performance evaluation for the CBR prediction models for fine-grained soil

Input parameters	Method											
	LR			SVR			MLP			GMDH-type NN		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
LL,PL,PI,MDD,OMC	0.54	4.00	3.03	0.53	4.06	2.88	0.69	3.28	2.57	0.826	3.06	2.30
LL,PI,MDD,OMC	0.54	4.00	3.04	0.52	4.11	2.90	0.68	3.35	2.64	0.829	3.02	2.40
LL,PI,MDD	0.54	4.02	3.07	0.51	4.14	2.99	0.67	3.39	2.72	0.780	3.41	2.56
LL,PI,OMC	0.41	4.53	3.55	0.46	4.33	3.07	0.65	3.48	2.71	0.807	3.20	2.44

combinations, the GMDH-type NN method had the highest regression coefficient (R^2 = 0.81) and lower error values (RMSE = 3.17 and MAE = 2.43) on average for the testing set. The evaluation results reveal that the GMDH-type NN model has also the most reliable in the prediction of CBR values for fine-grained soils.

As described above for coarse-grained soils, polynomial functions for fine-grained soils were developed for the GMDH-type NN and designed with various input variables. For fine-grained soils, LL, PI, MDD, and OMC variables were found as the most powerful predictors for the prediction of CBR value with the GMDH model. Among those variables, MDD had the least impact on the results. The variable pairs formed by the most effective variables were {LL, OMC}, {PI, OMC}, {LL, MDD}, and {PI, MDD}. By using these variables, polynomial functions were determined for the prediction of the CBR value in fine-grained soil conditions. The weight coefficients of these equations were calculated by the least-squares error method, and the low error outputs were used in the next layer. The polynomial equations of the GMDH models designed for the prediction of the CBR value in fine-grained soils are as follows:

{LL, MDD}, {LL, OMC}, {PI, MDD}, {PI, OMC} and {MDD, OMC} input variable pairs and its polynomial equations

$$Y_1 = 31.84 - 0.39PI - 1.71OMC - 0.008PI^2 + 0.018OMC^2 + 0.03PI \cdot OMC$$
$$Y_2 = 36.61 - 0.45LL - 1.55OMC + 0.0009LL^2 + 0.02OMC^2 + 0.01LL \cdot OMC$$
$$Y_3 = -24.95 + 15.06MDD + 1.19OMC + 3.78MDD^2 + 0.003OMC^2 - 0.94MDD \cdot OMC$$
$$Y_4 = -0.3 - 0.67Y_2 + 1.73Y_3 - 0.1{Y_2}^2 - 0.28{Y_3}^2 + 0.39Y_2 \cdot Y_3$$
$$Y_5 = -0.08 - 0.53Y_1 + 1.65Y_3 - 0.28{Y_1}^2 - 0.46{Y_3}^2 + 0.74Y_1 \cdot Y_3$$
$$CBR = 1.21 - 0.53Y_4 + 1.14Y_5 - 0.076{Y_4}^2 - 0.14{Y_5}^2 + 0.24Y_4 \cdot Y_5$$

{LL, PI}, {LL, MDD}, {LL, OMC}, {PI, MDD} and {PI, OMC} input variable pairs and its polynomial equations

$$Y_1 = -84.59 + 1.67LL + 56.99MDD - 0.003LL^2 - 1.98MDD^2 - 0.92LL \cdot MDD$$
$$Y_2 = -79.45 + 2.66LL + 60.39OMC - 0.009LL^2 - 5.54OMC^2 - 1.49LL \cdot OMC$$
$$Y_3 = 33.13 - 0.59PI - 1.57OMC - 0.001PI^2 + 0.018OMC^2 + 0.02PI \cdot OMC$$
$$Y_4 = 1.99 - 0.39Y_2 + 0.84Y_3 - 0.45{Y_2}^2 - 0.49{Y_3}^2 + 0.98Y_2 \cdot Y_3$$
$$Y_5 = 0.2 + 1.23Y_1 - 0.45Y_2 + 0.05{Y_1}^2 + 0.16{Y_2}^2 - 0.19Y_1 \cdot Y_2$$
$$CBR = -1.2 + 0.88Y_4 + 0.45Y_5 + 0.005{Y_4}^2 + 0.026{Y_5}^2 - 0.047Y_4 \cdot Y_5$$

{LL, MDD} and {PI, MDD} input variable pairs and its polynomial equations

$$Y_1 = -67.87 + 2.68PI + 47.57MDD - 0.01PI^2 - 2.25MDD^2 - 1.47PI \cdot MDD$$
$$Y_2 = -57.2 + 1.33LL + 35.08MDD - 0.002LL^2 + 2.62MDD^2 - 0.8LL \cdot MDD$$
$$CBR = -0.82 + 1.41Y_1 - 0.14Y_2 - 0.21{Y_1}^2 - 0.13{Y_2}^2 + 0.33Y_1 \cdot Y_2$$

{LL,OMC} and {PI,OMC} input variable pairs and its polynomial equations

$$Y_1 = 30.71 - 0.54PI - 1.44OMC - 0.003PI^2 + 0.01OMC^2 + 0.02PI \cdot OMC$$
$$Y_2 = 36.3 - 0.51LL - 1.38OMC + 0.002LL^2 + 0.016OMC^2 + 0.009LL \cdot OMC$$
$$CBR = 2.34 + 0.63Y_1 - 0.31Y_2 - 0.58{Y_1}^2 - 0.53{Y_2}^2 + 1.16Y_1 \cdot Y_2$$

In summary, the CBR values of coarse- and fine-grained soils were predicted using the GMDH-type NN and compared with LR, SVR, and MLP methods. In the performance evaluation, the GMDH-type NN algorithm provided the most reliable results in the prediction of CBR values for both coarse- and fine-grained soils. With the GMDH-type NN, the most important parameters were found to be G, S, MDD, and OMC in the CBR prediction for coarse-grained soils. In addition, LL, PI, MDD, and OMC were the most related parameters in the CBR prediction for fine-grained soils. In addition, the required input variables were automatically selected in the development of the GMDH-type NN models; the polynomial equations were presented for both coarse- and fine-grained soils in this study.

5. Conclusions

In this study, various regression models such as LR, SVR, MLP, and GMDH type NN, which is an artificial neural network-based method, were developed for the estimation of CBR values for coarse- and fine-grained soils. In the development of CBR prediction models, independent input variables such as G, S, FG, OMC, and MDD for coarse-grained soils and independent input variables such as LL, PL, PI, OMC, and MDD for fine-grained soils were used. The results show that the GMDH-type NN algorithm had a higher estimation performance than LR, SVR, and MLP methods in the prediction of the CBR value for coarse- and fine-grained soils. The GMDH method had R^2 of 0.938, RMSE of 1.87, and MAE of 1.48 for the input variables {G, S, and MDD} in coarse-grained soils and, it had R^2 of 0.829, RMSE of 3.02 and MAE of 2.40 while the input variables {LL, PI, MDD, and OMC} in fine-grained soils. Furthermore, the polynomial equations of GMDH models were automatically developed with optimal input variables for the estimation of CBR values for both coarse and fine-grained soils. The results show that the GMDH-type NN, which is accepted as the early stage of deep learning, was an effective and successful regression method that could be used to predict the CBR values for both coarse- and fine-grained soils in geotechnical engineering applications.

Acknowledgments

This work was supported by research fund of Chungnam National University and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2022R1C1C1011477).

References

Agarwal, K.B. and Ghanekar, K.D. (1970), "Prediction of CBR from plasticity characteristics of soil", *Proceeding of the 2nd south-east Asian conference on soil engineering*, Singapore, June.

Alam, S.K., Mondal, A. and Shiuly, A. (2020), "Prediction of CBR value of fine grained soils Bengal Basin by genetic expression programming, artificial neural network and krigging method", *J. Geological Soc. India*, **95**(2), 190-196. https://doi.org/10.1007/s12594-020-1409-0.

Alawi, M. and Rajab, M. (2013), "Prediction of California bearing ratio of subbase layer using multiple linear regression models", *Road Mater. Pavement Design*, **14**(1), 211-219. https://doi.org/10.1080/14680629.2012.757557.

American Association of State Highway and Transportation Officials (2021), *T193-13, Standard method of test for the California bearing ratio*. American Association of State Highway and Transportation Officials, Washington, DC, USA.

American Society for Testing and Materials. (2021), *ASTM D1883-21, Standard test method for California bearing ratio (CBR) of laboratory-compacted soils*. ASTM International, West Conshohocken, PA, USA. https://doi.org/10.1520/D1883-21.

Ardakani, A. and Kordnaeij, A. (2019), "Soil compaction parameters prediction using GMDH-type neural network and genetic algorithm", *Eur. J. Environ. Civil Eng.*, **23**(4), 449-462. https://doi.org/10.1080/19648189.2017.1304269.

Armaghani, D.J., Mirzaei, F., Shariati, M., Trung, N.T., Shariati, M. and Trnavac, D. (2020), "ANN-based techniques in predicting cohesion of sandy-soil combined with fiber", *Geomech. Eng.*, **20**(3), 191-205. https://doi.org/10.12989/gae.2020.20.3.191.

Awad, M. and Khanna, R. (2015), *Support vector regression. In Efficient learning machines*. Apress, Berkeley, CA, USA.

Bowles, J.E. (1992), *Engineering properties of soils and their measurements*. McGraw-Hill, Inc., New York, NY, USA.

Elbaz, K., Shen, S.L., Zhou, A., Yin, Z.Y. and Lyu, H.M. (2021), "Prediction of disc cutter life during shield tunneling with AI via the incorporation of a genetic algorithm into a GMDH-type neural network", *Engineering*, **7**(2), 238-251. https://doi.org/10.1016/j.eng.2020.02.016.

Farlow, S.J. (2020), *Self-organizing methods in modeling: GMDH type algorithms*. CRC Press, Inc., Boca Raton, FL, USA.

Hao, S. and Pabst, T. (2022), "Prediction of CBR and resilient modulus of crushed waste rocks using machine learning models", *Acta Geotechnica*, **17**(4), 1383-1402. https://doi.org/10.1007/s11440-022-01472-1.

Harini, H. and Naagesh, S. (2014), "Predicting CBR of fine grained soils by artificial neural network and multiple linear regression", *Int. J. Civil Eng.*, **5**(2), 119-126.

Ivakhnenko, A.G. (1968), "The group method of data handling-a rival of the method of stochastic approximation", *Soviet Automatic Control*, **13**(3), 43-55.

Katte, V.Y., Mfoyet, S.M., Manefouet, B., Wouatong, A.S.L. and Bezeng, L.A. (2019), "Correlation of california bearing ratio (CBR) value with soil properties of road subgrade soil", *Geotech. Geol. Eng.*, **37**, 217-234. https://doi.org/10.1007/s10706-018-0604-x.

Khatri, D.P., Acharya, I.P. and Dhakal, B.B. (2019), "Correlation of California bearing ratio with index properties of sub-grade soil: a case study on Thankot Chitlang road section", *Proceedings of the IOE Graduate Conference*, Kathmandu, Nepal, December.

Kim, M., Okuyucu, O., Ordu, E., Ordu, S., Arslan, O. and Ko, J. (2022), "Prediction of undrained shear strength by the GMDH-type neural network using SPT-value and soil physical properties", *Materials*, **15**(18), 6385. https://doi.org/10.3390/ma15186385.

Kondo, T. (1998), "GMDH neural network algorithm using the heuristic self-organization method and its application to the pattern identification problem", *Proceedings of the 37th SICE Annual Conference. International Session Papers*, Chiba, Japan, July.

Li, R.Y.M., Fong, S. and Chong, K.W.S. (2017), "Forecasting the REITs and stock indices: group method of data handling neural network approach", *Pacific Rim Property Res. J.*, **23**(2), 123-160. https://doi.org/10.1080/14445921.2016.1225149.

Liu, J., Luan, H., Zhang, Y., Sakaguchi, O. and Jiang, Y. (2020), "Prediction of unconfined compressive strength ahead of tunnel face using measurement-while-drilling data based on hybrid genetic algorithm", *Geomech. Eng.*, **22**(1), 81-95. https://doi.org/10.12989/gae.2020.22.1.081.

Luat, N., Lee, K. and Thai, D. (2020a), "Application of artificial neural networks in settlement prediction of shallow foundations on sandy soils", *Geomech. Eng.*, **20**(5), 385-397. https://doi.org/10.12989/gae.2020.20.5.385.

Luat, N., Nguyen, V., Lee, S. and Lee, K. (2020b), "An evolutionary hybrid optimization of MARS model in predicting settlement of shallow foundations on sandy soils", *Geomech. Eng.*, **21**(6), 583-598. https://doi.org/10.12989/gae.2020.21.6.583.

Madala, H.R. and Ivakhnenko, A.G. (1994), *Inductive learning algorithms for complex systems modeling*. CRC Press, Inc., Boca Raton, FL, USA.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2021), *Introduction to linear regression analysis*. John Wiley & Sons, Hoboken, NJ, USA.

Murtagh, F. (1991), “Multilayer perceptrons for classification and regression”, *Neurocomputing*, **2**(5-6), 183-197. https://doi.org/10.1016/0925-2312(91)90023-5.

National Cooperative Highway Research Program (NCHRP). (2001), *Guide for mechanistic and empirical – design for new and rehabilitated pavement structures, final document. In: Appendix CC-1: correlation of CBR values with soil index properties*. Champaign, Illinois, USA.

Nagaraj, H.B. and Suresh, M.R. (2018), “Influence of clay mineralogy on the relationship of CBR of fine-grained soils with their index and engineering properties”, *Transport. Geotech.*, **15**, 29-38. https://doi.org/10.1016/j.trgeo.2018.02.004.

Oh, S.K. and Pedrycz, W. (2002), “The design of self-organizing polynomial neural networks”, *Inform. Sci.*, **141**(3-4), 237-258. https://doi.org/10.1016/S0020-0255(02)00175-5.

Ravichandra, A.H., Shivakumar, K., Vinaykumar, H., Khalid, S. M. and Basavaraj, B. (2019) “Prediction of CBR value by using index properties of soil”, *Int. Res. J. Eng. Technol.*, **6**(7), 3740-3747.

Rakaraddi, P.G. and Gomarsi, V. (2015), “Establishing relationship between CBR with different soil properties”, *Int. J. Res. Eng. Technol.*, **4**(2), 182-188.

Ramasubbarao, G. and Sankar, S.G. (2013), “Predicting soaked CBR value of fine grained soils using index compaction characteristics”, *Jordan J. Civil Eng.*, **7**(3), 354-360.

Satyanarayana Reddy, C.N.V and Pavani, K. (2006), “Mechanically stabilized soils-regression equation for CBR evaluation”, *Proceeding of the Indian geotechnical conference*, Chennai, India, December.

Schmidhuber, J. (2015), “Deep learning in neural networks: An overview”, *Neural Networks*, **61**, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003.

Smola, A. J. and Schölkopf, B. (2004), “A tutorial on support vector regression”, *Statist. Comput.*, **14**(3), 199-222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Taha, S., Gabr, A. and El-Badawy, S. (2019), “Regression and neural network models for California bearing ratio prediction of typical granular materials in Egypt”, *Arabian J. Sci. Eng.*, **44**(10), 8691-8705. https://doi.org/10.1007/s13369-019-03803-z.

Tenpe, A.R. and Patel, A. (2020), “Application of genetic expression programming and artificial neural network for prediction of CBR”, *Road Mater. Pavement Design*, **21**(5), 1183-1200. https://doi.org/10.1080/14680629.2018.1544924.

Yildirim, B. (2009), *Determination of California bearing ratio (CBR) by using regression analyses and artificial neural networks*. Nigde University.

Yildirim, B. and Gunaydin, O. (2011), “Estimation of California bearing ratio by using soft computing systems”, *Exp. Syst. Appl.*, **38**(5), 6381-6391. https://doi.org/10.1016/j.eswa.2010.12.054.

Venkatasubramanian, C. and Dhinakaran, G. (2011), “ANN model for predicting CBR from index properties of soils”, *Int. J. Civil Struct. Eng.*, **2**(2), 614-620.