

Research Article

Prediction of California Bearing Ratio of Granular Soil by Multivariate Regression and Gene Expression Programming

Mudthir Bakri ¹, Ibrahim Aldhari ², and Mohammed Saleh Alfawzan ¹

¹Department of Civil Engineering, College of Engineering, Qassim University, Unaizah, Qassim, Saudi Arabia

²Department of Civil Engineering, College of Engineering, Qassim University, Buraydah, Qassim, Saudi Arabia

Correspondence should be addressed to Ibrahim Aldhari; 431114313@qu.edu.sa

Received 9 June 2022; Revised 20 July 2022; Accepted 22 July 2022; Published 31 August 2022

Academic Editor: Qian Chen

Copyright © 2022 Mudthir Bakri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This research demonstrates the results of an investigation into the California bearing ratio (CBR) of granular soils from Qassim region, Saudi Arabia, using multilinear regression (MLR), pure quadratic (PQ) models, and gene expression programming (GEP) methods utilized to develop mathematical models for estimating the CBR based on basic soil index properties. In this study, samples were collected from different borrowing pits in the Qassim area. Forty-three samples of soil were taken and transferred to a laboratory for examination. Seven multilinear regressions and seven PQ models were investigated, while four GEP models were made. The selection of each model variable depends on soil indices, grouping into grain size distribution, Atterberg limits, and compaction parameters. The results of this analysis showed that the PQ model had a higher accuracy [coefficient of determination (R^2) = 0.89, root mean square error (RMSE) = 16.006, uncertainty (U95) = 16.17, and reliability = 57%] than the multilinear regression model, which has a lower accuracy model [R^2 = 0.811, RMSE = 20.791, U95 = 15.569, and reliability = 51%]. The best GEP model yields [R^2 = 0.776, RMSE = 22.552, U95 = 15.787, and reliability = 53%]. Furthermore, sensitivity analysis was conducted to distinguish the influences of different input variables on CBR; it was found that fines percentage (F_{200}), maximum dry density (MDD), and optimum moisture content (OMC) are the most influential variables.

1. Introduction

Saudi Arabian road network consists of 73,171 km. Saudi Arabia has a vast and well-developed transportation network. Roads carry approximately 65 percent of freight and 80 percent of passenger transportation [11]. The total number of automobiles in Saudi Arabia exceeds 12 million, and the growth rate of the number of vehicles is approximately 5% per annum over the last five years [20]. Geotechnical engineering is one of the greatest crucial subjects in the initial stages of infrastructure planning and design, owing to the concept that poor geotechnics might consequence in unnecessarily high costs if not addressed properly. The CBR value is a critical parameter in the structural design of pavements. The CBR value may be determined directly in the laboratory using AASHTO T 193 and ASTM D 1883.

Two broad correlations for predicting CBR were also presented in the Guide for Mechanistic and Empirical

design (2001). The models are constructed using empirical parameter data for the following soils; D60 (passing 60% diameter), P200 (percent passing sieve no. 200; US sieve), and PI (plastic index). PI and % passing no. 200 US sieve (or 0.075 mm size sieve) are two parameters that are included in the suggested model for plastic fine-grained soils [15]. The recommended equation is

$$\text{CBR} = \frac{75}{1 + 0.728F_{200}\text{PI}}, \quad (1)$$

where F_{200} is passing no. 200 US sieve (%).

PI is the plasticity index. For the nonplastic coarse-grained soil, the proposed equation is

$$\text{CBR} = 28D_{60}^{0.358}. \quad (2)$$

Determine the CBR value of each soil sample that demands at least four days in the laboratory. Historically, civil engineers have struggled to determine a representative CBR value for pavement design [36]. The CBR value varies according to the kind of soil and its various qualities. Due to the variability in soil engineering features, conducting this experiment with soil samples gathered from a few places cannot be used for the entire road path. To overcome this, a substantial number of specimens must be collected for testing, which causes the method to be costly and time consuming. Black made the first attempt to predict CBR in 1962 [13]; his work depends on remolded soil and its friction angle. It is also shown that the suction of such a remolded soil and its true angle of friction can be inferred from data on the liquid limit, plastic limits, and water content of the soil. Agarwal and Ghanekar predicted the CBR deepening on the grain size distribution as well as the consistency parameters, specifically the PI [2]. These attempts were followed by many subsequent attempts using more advanced methods of prediction, each of these attempts expresses the classification, considerable types of index properties, and unique study area samples [3, 5, 7]. CBR was determined by Yildirim and Gunaydin [35] utilizing soft computing; they used soil index properties and compaction properties of subgrade soil in Turkey as input parameters. Their work revealed a high correlation ($R^2 = 0.80-0.95$) employing soil indices. They recommend using the provided formulas in the early stages of the design process [35]. Kumar (2014) proposed a linear regression model for ML and MI soils on the basis of the variables PI, liquid limit (LL), plastic limit (PL), MDD, and OMC, for ML and CBR and PI have an inverse linear regression relation [32]. Several linear regression models for estimating soaked CBR in clay soils are compared by Ramasubbarao and Siva [26, 26]. Tenpe et al. constructed several mathematical models to predict CBR utilizing a variety of soil parameters as independent variables, with the assistance of the GEP and ANN models. They also conducted comparative research to demonstrate that there is little difference between the findings acquired from GEP and those from ANN [6]. Numerous scholars have presented such models [6, 29, 34], but there is still a knowledge gap concerning the assessment and comparison of their predictive capability. Aleksandra et al. investigated the effects of compaction parameters, namely MDD and OMC on CBR [10].

Interestingly, no correlation formula has been created utilizing the soil of the Qassim area. The region's geological formations are divided into two large landmasses: the Arabian Shield in the west and the Arabian Shelf in the east. The Arabian Shield is made up of igneous and metamorphic rocks, while the Arabian Shelf is made up of sedimentary rocks. The weathering products of these rocks have generated the soils. Predicting the value of the CBR test would shorten the time from 96 hours to minutes, achieving such a goal would reduce the overall cost of projects in various ways. Al-Qassim region is one of the most important logistical areas. Since at present, there is no published work correlating soil properties and CBR in the Qassim area and its environs. So that the purpose of the study is to determine

if GEP and multivariate regression approaches are applicable to estimate CBR of flexible pavement subbases and subbase layers in the Qassim area.

1.1. Samples Collection. Qassim is a region of predominantly desert nature with a multigeological nature, and the so-called Arab Shield is located in the southwestern part of it. It is located in the middle of the Kingdom of Saudi Arabia [4]. Qassim is considered a vital region linking the north with the south and the west with the east, thus heavy carriers pass through its roads. In Saudi Arabia, roads are built in accordance with the Ministry of Transportation [8] (MOT, 1998) regulations. Tables 1 and 2 present the gradation of the subbase layer and the limit of indices. During the building of new roads in Qassim area, samples were taken from various places and borrowed pits. These samples were sieved to find the percentage of aggregate (gravel) and the percentage of aggregate passing through sieve no. 4 (4.75 mm) and were retained on sieve no. 200 (0.075 mm) (sand) and the percentage passing through sieve no. 200 (0.075 mm) (fines). The consistency tests were conducted on fine soil to determine the liquid and plastic limits. The relationship between moisture content and dry density was then determined by the proctor modified test. Then, each sample was made at OMC and at different densities, with the one with the highest dry unit weight being tested to see what its CBR was for each unit weight.

1.2. Experimental Work. In this research, to study the factors which explain CBR characteristics and its prediction, soil samples were gathered from 43 borrowed pits in the Qassim area as shown in Figure 1, with a different number of specimens taken from each location. The test method used to determine the CBR is given by MOT (1998) specifications, in which it complies with MRDTM: 213 ASTM D 1883-AASHTO 193 [31]. The compaction parameters are found in accordance to ASTM D 698-1557, AASHTO T99-T180. The particle size distribution shall follow MRDTM: 204. MRDTM: 208/209 was used to conduct the test of LL, PL, and the PI. These limits were determined on the soil samples having a fine percentage greater than 5%. Table 3 shows that in general the average numbers of fines (silt and clay), sand, and gravel were 15.32%, 28.3%, and 56.421%, respectively. The averages of LL, PL, and PI were 7.8%, 6.18%, and 1.62%, respectively, while the average water content OMC was 7.73%, and the average MDD was 2.15 g/cm³.

AASHTO classification system [(ASTM designation D-3282; AASHTO method M145)] was used to classify the soil samples, and it was found that 34 samples are classified as A-1a and A-1-b, and ten samples were classified as A-2.

Grain size analyses were performed firstly to classify the soil in the studied area and to quantify the number of fines [F_{200}], sand percent [S], and gravel percentage [G]; grain size distribution curve is shown in Figure 2. Then after analysis of all samples, the tests of LLs and plastic limits were conducted; the plasticity chart is shown in Figure 3. Thirty percent of soil samples had a liquid limit larger than zero, which was classed as inorganic clays of low plasticity to

TABLE 1: Ministry of transportation specification limits or gradation of the subbase layer.

Sieve size designation	MOT specification limits (% passing)	
	Upper limit	Lower limit
2"	100	100
1 1/2" 1"	100	90
1"	85	55
3/4"	80	50
3/8"	70	40
No. 4	60	30
No. 10	50	20
No. 40	30	10
No. 200	15	0

TABLE 2: MOT specifications of soil indices for materials of subbase layer.

Sand equivalent	Plasticity index	Abrasion loss	CBR
25 min.	6 max.	50 max.	50 min.

medium plasticity and inorganic silt of low compressibility. The majority of the samples depicted fall above line A and below line U. The existence of soil samples parallel to the line A on the plasticity chart shows that they have the same geological origin [19, 30]. Table 3 provides a full statistical summary of the soil. To see the distribution of the parameters more clearly, frequency histograms were created for each parameter in Figure 4.

1.3. Multivariate Regression. Single and multivariate regression analysis are well-known predictive techniques that are frequently used to construct models for predicting desired output parameters based on specified input factors such as the physical and mechanical characteristics of rocks [17]. Single regression studies examine the connections between input and output variables using a linear, logarithmic power or exponential function. Multivariate regression studies are advantageous when several input parameters are present in complicated connections, and the common form of a multivariate regression equation is

$$CBR = a + \sum_{i=1}^n bG_{pi} + \varepsilon, \quad (3)$$

where CBR is California bearing ratio (dependent variable). G_{pi} is soil index or mechanical (independent variable), n is the number of variables, b_i is the coefficient of regression, a is the constant, and ε is the error estimate.

CBR will be used as the dependent variable, while the soil parameters in Table 4 will be used as independent variables individually, or they will be used as groups of parameters depending on the relationship between these data and the correlation matrix as shown in Table 5. In addition, the selection of input variables of multilinear equations depends on grouping the soil indices into three categories; the first one is the grain size distribution, in which the parameters are gravel percent, sand percent, and fines percent. The second

category is Atterberg limits, which are LL, PL, and PI. The third category is "compaction variables," which include "MDD" and optimum water content. This will help to find the highest correlation parameters with CBR.

1.4. Multinonlinear Regression. Nonlinear regression estimates the dynamic connection between the dependent variable and the function. Nonlinear multivariable regression estimates the interaction between independent and dependent variables. There are various modeling techniques and regression approaches for specific aspects, especially when data are limited [17].

It is possible that the greatest association between the answer and independent variables may be explained by step-by-step regression. This algorithmic technique selects the proper model subsets to filter forward or backward. First, choose a permanent model and use the model words before optimizing fitness. That is, step-by-step regression is a forward selection procedure that examines the significance of all prior variables. For variables whose partial square sums do not meet the minimal requirement, the retroactive exclusion procedure is improved and variables are excluded one at a time until the minimum condition is met. Step-by-step regression requires more computations than forward or backward detailed calculations, but it yields better results [33]. Scheffe proposed a second-order polynomial for MNR analysis of mechanical characteristics [21]. MNLR in PQ forms has been studied.

1.5. Gene Expression Programming (GEP). GEP, which is an algorithm based on genetic algorithm (GA) and genetic programming (GP), was first proposed by Ferreira [27]. The GEP technique is an integration of two evolutionary strategies that are broadly acknowledged in the scientific community. The first approach is called the GA methodology, and it simplifies complex relationships by depicting them with linear structures called chromosomes that are of a fixed length. In addition, the second method is known as GP, and it makes use of the expression tree (ET) configurations that may take on a wide range of forms and dimensions [22]. Developing a program code recorded in linear chromosomes of fixed length, it exhibits some characteristics akin to biological evolution. GEP's primary goal is to create a mathematical function that fits a collection of data supplied to the GEP model. The GEP method performs symbolic regression on the mathematical equation using the majority of GA's genetic operators. GEP and GA have some distinctions. Individuals are shown in GA as symbolic strings of fixed length (chromosomes), but they are depicted in GP as nonlinear entities of varying sizes and forms (parse trees), whereas in GEP, each individual is represented by a fixed-length string that is used to make ETs of different sizes and shapes [9]. The primary GEP process is presented in Figure 5. Five variables are employed in GEP analysis: function set, terminal set, fitness function, control parameters, and stop condition. After encoding the issue into the candidate solution and specifying the fitness function, a randomly generated population of viable people (chromosomes) is

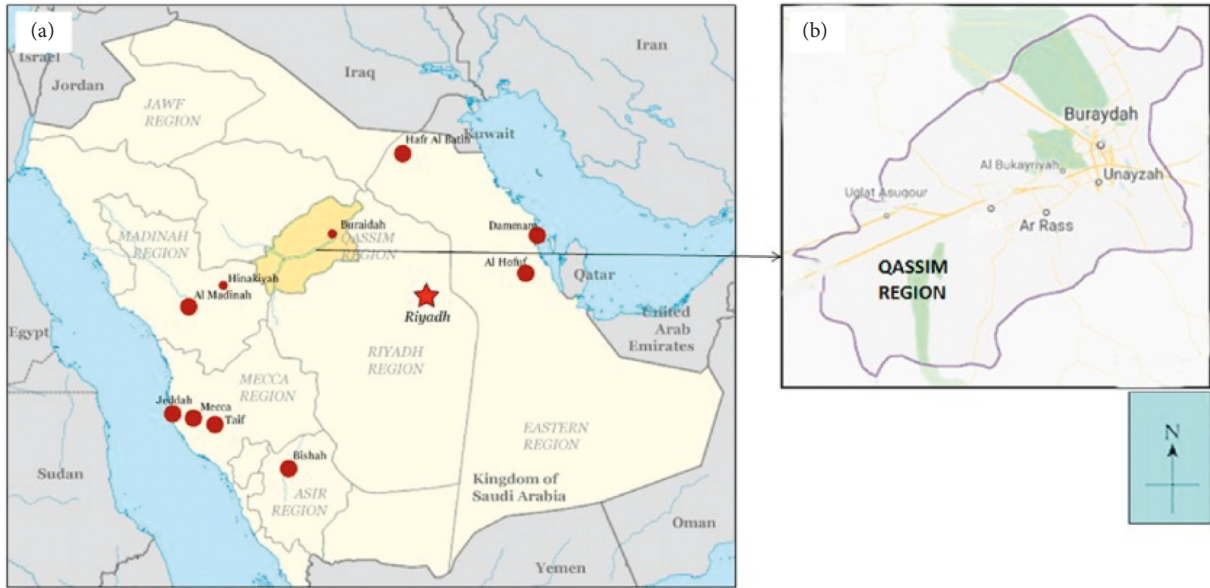


FIGURE 1: Location of samples.

TABLE 3: Descriptive statistics analysis of the tested samples.

	Gravel %	Fines %	Sand %	Liquid limit	Plastic limit	Plasticity index	Maximum dry density	Optimum moisture content	CBR %
Mean	56.421	15.327	28.3	7.8	6.181	1.623	2.15	7.73	89.45
Std. error of mean	3.028	1.054	2.35	1.841	1.48	0.377	0.0124	0.288	7.37
Std. deviation	19.856	6.91	15.43	12.07	9.7	2.473	0.08134	1.89	48.32
Variance	394.27	47.81	238.17	145.8	94.06	6.12	0.07	3.591	2334.77
Median	63.4	14	19.1	0	0	0	2.157	7	60
Minimum	10.70	5.3	12.63	0.00	0.00	0.00	1.91	5.60	21
Maximum	80.37	28.3	75.8	35.20	28.70	6.80	2.24	16.50	172

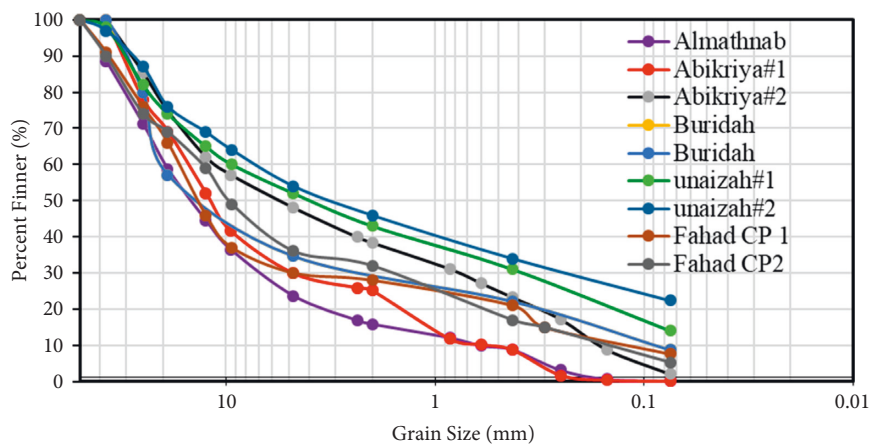


FIGURE 2: Representative grain size distribution of soil samples.

constructed and then translated into an ET matching to a mathematical equation. Following that, the projected values are compared to the actual ones, and the fitness score for each chromosome is calculated. The algorithm is ended when a suitable fitness score is reached. If this is not the case,

some chromosomes are randomly selected and then mutated to generate new generations. This method is repeated until a suitable fitness score is produced, at which point noncoding chromosomal analysis is undertaken to determine the optimal solution to the problem [14].

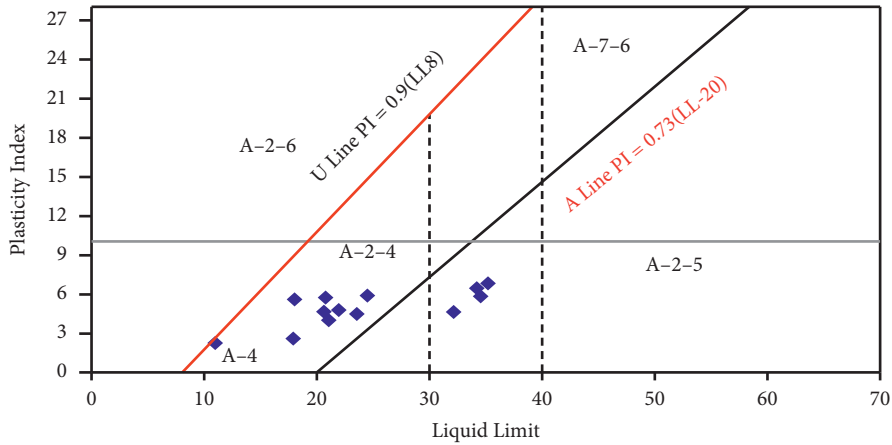


FIGURE 3: Plasticity diagram showing sample results.

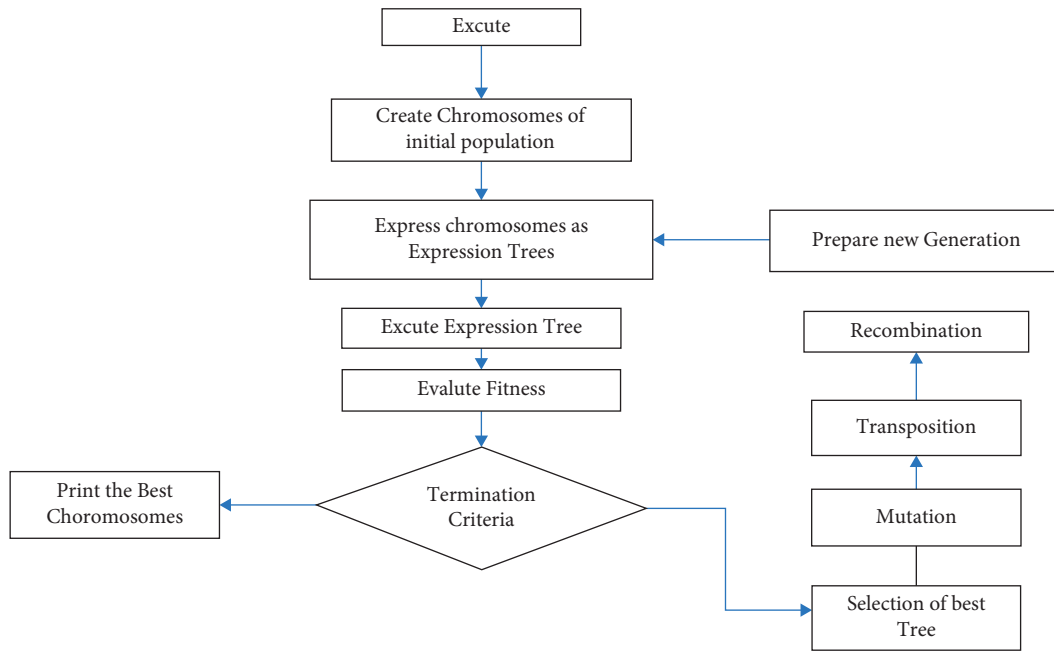


FIGURE 4: GEP algorithm.

1.6. *GEP Elements.* The ET and chromosomes are the two most important components of the GEP. The chromosome is made up of one or more genes and it is represented by a mathematical equation on the genome. To code any mathematical equation, two bilingual and conclusive languages named Karva languages (the language of genes and the language of ET) are utilized, both of which are bilingual and conclusive. This property of the GEP is important for inferring the genotype with more precision than it would otherwise be possible. The GEP genes are divided into two sections, which are referred to as the head and tail. It is necessary to use the head of a gene, which contains certain mathematical functions, variables, and constants, for the encoding of a particular function. However, the tail of a gene that contains just variables and constants is used as supplemental terminal symbols in addition to the main terminal symbols. These symbols are used in the event that the terminal symbols in the head are insufficient to express a

particular function. Many other functions, such as fundamental arithmetic operators, trigonometric functions, and any other mathematical or user-defined functions, may be utilized in the head of a gene.

1.7. *GEP Operators.* The major operators in GEP are selection, mutation, transposition, and crossover (recombination), which are all based on natural selection. By using these operators, the chromosomes may be adjusted to improve the fitness score of the following generation for survival. The operator rates, which are specified at the outset of the model construction, indicate a particular chance of a chromosome appearing in the model. In general, a mutation rate of 0.001 to 0.1 is suggested, with the lower limit being 0.001. The transposition operator and the crossover operator, on the other hand, are recommended to be 0.1 and 0.4, respectively [24].

TABLE 4: Summary of sample results.

No.	Gravel %	Fines %	Sand %	LL	PL	PI	MDD g/cm ³	OMC %	CBR %
1	72	7.6	20.4	0	0	0	2.185	6.4	123
2	48	20.6	31.4	23.5	19	4.5	2.12	8.4	47
3	68	5.3	26.7	0	0	0	2.15	8.6	114
4	73	7.9	19.1	0	0	0	2.19	6.8	113
5	49.9	21.2	28.9	20.8	15.1	5.7	2.157	8.2	123
6	72.1	9.2	18.7	0	0	0	2.236	6.9	172
7	63.4	20.1	16.5	20.6	16	4.6	2.105	8.4	80
8	43.2	13.4	43.4	0	0	0	2.14	6.7	54
9	75.39	7.5	17.11	0	0	0	2.235	6.6	150
10	15.5	26.9	57.6	0	0	0	2.092	6.5	27
11	40.3	23.2	36.5	0	0	0	2.135	7.3	43.8
12	26.9	19.7	53.4	0	0	0	2.156	7	45.5
13	80.37	7	12.63	0	0	0	2.237	6.8	155
14	62.7	19.8	17.5	34.5	28.7	5.8	2.144	7.3	60
15	10.7	13.5	75.8	0	0	0	2.042	7.7	35
16	76.11	9.4	14.49	0	0	0	2.236	6.9	170
17	75.39	7.5	17.11	0	0	0	2.235	6.6	150
18	62.8	21	16.2	0	0	0	2.128	6.8	51
19	75.65	7.2	17.15	0	0	0	2.23	6.6	150
20	59.9	21.5	18.6	0	0	0	2.147	8.2	45
21	65.7	17.3	17	24.5	18.6	5.9	2.166	8.3	51
22	43.2	13.4	43.4	0	0	0	2.184	6.6	54
23	40.7	15.4	43.9	0	0	0	2.205	6.7	60
24	73.11	9.6	17.29	0	0	0	2.225	7.1	120
25	13.6	24.8	61.6	17.9	15.3	2.6	2.149	7.2	21
26	72.69	8.4	18.91	10.9	8.7	2.2	2.231	6.9	154
27	49.9	21.2	28.9	20.8	15.1	5.7	2.157	8.2	51
28	73.3	8.8	17.9	0	0	0	2.237	5.6	143
29	76.64	9.1	16.26	0	0	0	2.236	6.5	153
30	61.6	19.5	18.9	22	17.2	4.8	2.115	8.8	56
31	71.1	14	14.9	32.1	27.5	4.6	2.156	7.9	48
32	63.8	9.1	27.1	0	0	0	2.236	6.9	170
33	45.1	28.3	26.6	34.2	27.8	6.4	2.125	9.3	49
34	63.4	20.1	16.5	20.6	16	4.6	2.105	8.4	80
35	35.2	25.9	38.92	35.2	28.4	6.8	1.99	10.5	37
36	25.9	24.2	49.9	0	0	0	1.914	16.5	48
37	45.3	19.4	35.3	0	0	0	2.05	10.5	54
38	76.41	6.9	16.69	0	0	0	2.24	6.2	130
39	22.9	24.9	52.2	0	0	0	1.93	11.8	58
40	37.5	21.5	41	18	12.4	5.6	2.05	9.2	33
41	70.95	10.8	18.25	0	0	0	2.23	6	126
42	70.5	8.5	21	0	0	0	2.21	6.6	123
43	76.3	8.5	15.2	0	0	0	2.206	6	119

TABLE 5: Correlations matrix.

	CBR	Gr	F ₂₀₀	S	LL	PL	PI	MDD	OMC
CBR	1	0.773	-0.82	-0.623	-0.42	-0.42	-0.403	0.716	-0.437
Gr	0.773	1	-0.738	-0.953	-0.121	-0.118	-0.127	0.727	-0.46
F ₂₀₀	-0.82	-0.738	1	0.499	0.524	0.516	0.536	-0.741	0.565
S	-0.623	-0.953	0.499	1	-0.081	-0.081	-0.078	-0.6	0.337
LL	-0.42	-0.121	0.524	-0.081	1	0.998	0.968	-0.282	0.26
PL	-0.42	-0.118	0.516	-0.081	0.998	1	0.951	-0.275	0.252
PI	-0.403	-0.127	0.536	-0.078	0.968	0.951	1	-0.296	0.28
MDD	0.716	0.727	-0.741	-0.6	-0.282	-0.275	-0.296	1	-0.844
OMC	-0.437	-0.46	0.565	0.337	0.26	0.252	0.28	-0.844	1

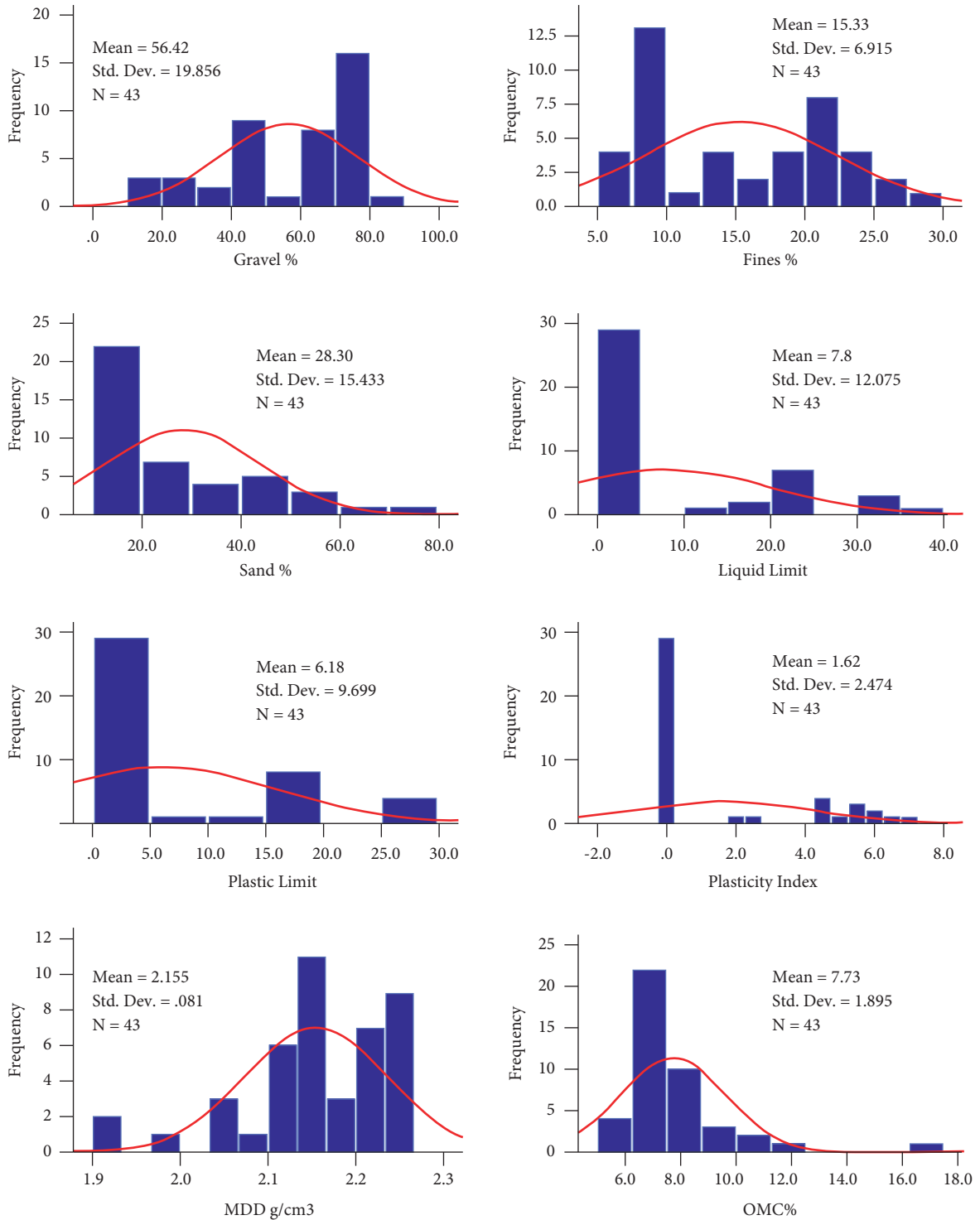


FIGURE 5: Frequency histograms of the index, MDD, and OMC.

1.8. Model Performance. To examine the performance of the prediction models, the RMSE, the coefficient of determination (R^2), as well as mean absolute percentage error

(MAPE), the scatter index (SI) [23], Bias, and index of agreement (IOA) are being used as shown in (4)–(9).

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^N (\text{CBR}) (\text{CBR}_m - \text{CBR}_p)^2}{N}}, \\
 R^2 &= 1 - \frac{\sum_{i=1}^N (\text{CBR}_m - \text{CBR}_p)^2}{\sum_{i=1}^N (\text{CBR}_p)^2}, \\
 \text{MAPE} &= \frac{1}{N} \sum_{i=1}^N \left| \frac{\text{CBR}_m - \text{CBR}_p}{\text{CBR}_p} \right| \times 100, \\
 \text{SI} &= \frac{\sqrt{1/N \sum_{i=1}^N ((\text{CBR}_{p_i} - \overline{\text{CBR}_p}) + (\text{CBR}_{m_i} - \overline{\text{CBR}_m}))^2}}{(1/N) \sum_{i=1}^N \text{CBR}_{m_i}}, \\
 \text{Bias} &= \frac{\sum_{i=1}^N (\text{CBR}_p - \text{CBR}_m)}{N}, \\
 \text{IOA} &= 1 - \frac{\sum_{i=1}^N (\text{CBR}_{p_i} - \text{CBR}_{m_i})^2}{\sum_{i=1}^N (|\text{CBR}_{p_i} - \overline{\text{CBR}_m}| + (\text{CBR}_{m_i} - \overline{\text{CBR}_m}))^2}.
 \end{aligned} \tag{4}$$

n , CBR_m , and CBR_p indicate the total number of data measured experimentally, and the developed predicted value, respectively.

In addition, the Durbin–Watson statistic was used to determine the existence of multicollinearity [1]. The Durbin–Watson statistic has a range of values between 0 and 4, and the midpoint (that is, 2) indicates that there is no association between the input variables. Therefore, a value between 1.5 and 2.5 is sufficient for obtaining models that are unaffected by multicollinearity. This section covers mathematical expressions and models. Each model's R^2 and P -value (p) are described in the Tables 6 and 7.

$$\text{DW} = \frac{\sum_{t=2}^T (e_t - e_{t-1})}{\sum_{t=1}^T e_t^2}, \tag{5}$$

whereas DW is the Durbin–Watson test and e_t and e_{t-1} least-squares regression residuals.

1.9. Uncertainty and Reliability. The main purpose of uncertainty analysis is to narrow the predicted range in which the real value of an experiment's result resides. This estimated range is expressed as an interval and is referred to as the uncertainty interval. It may be approximated using the calculated errors for the measuring method of the experiment in question [28]. U95 is one of the uncertainty analysis algorithms used to calculate the uncertainty interval. The value of U95 relating to a specific experiment result may be read as follows: if the provided experiment is performed again, the real value of the experiment's outcome will fall

within the specified uncertainty range around 95 times out of every 100 trials. Furthermore, the value of U95 is provided by

$$\text{U95} = \frac{1.96}{N} \sqrt{\sum_{i=1}^N (\text{CBR}_{m_i} - \overline{\text{CBR}_m})^2 + \sum_{i=1}^N (\text{CBR}_{m_i} - \text{CBR}_{p_i})^2}. \tag{6}$$

1.10. Relative Absolute Error and Reliability. The relative absolute error is a relative statistic that compares a predictive model's performance to that of a simple model. The predictive model's performance is defined as the total absolute difference between the actual and anticipated values (i.e., the error). The basic model's performance is defined as the total absolute difference between the realized value and the average of all realized values [25]. In other words, the relative absolute error determines if a model outperforms just forecasting the average (i.e., the simple model). The relative absolute error formula is

$$\text{RAE} = \frac{\sum_{i=1}^N |\text{CBR}_{m_i} - \text{CBR}_{p_i}|}{\sum_{i=1}^N |\text{CBR}_{m_i} - \overline{\text{CBR}_m}|}. \tag{7}$$

The meaning of the relative absolute error is straightforward: if the RAE is less than one, the model outperforms the basic model. The relative absolute error for a perfect model is 0. RAE should ideally be as near to zero as feasible.

A statistical technique known as reliability analysis may be used to evaluate a model's level of general consistency. To be more specific, it determines whether or not a proposed

TABLE 6: Multilinear regression analysis significant models* .

Model id	Parameters	Model	R ²	DW
MLR 1	Ln (Gr), Ln (F ₂₀₀) and Ln (S)	CBR = -60.3 + 10.8 Ln (Gr) - 66.60 Ln (F ₂₀₀) - 17.8 Ln (S)	0.753	1.84738
MLR 2	Ln (Gr), Ln (F ₂₀₀), Ln (S), PL, LL, MDD and OMC	CBR = -835 + 6.76 LL - 8.97 PL + 321 MDD + 9.76 OMC - 15.6 Ln (Gr) - 52.9 Ln (F ₂₀₀) - 32.6 Ln (S)	0.811	1.80383
MLR 3	Gr, F ₂₀₀ , and S	CBR=-1039 + 12.2Gr+ 7.6F ₂₀₀ + 11.4S	0.740	1.82060
MLR 4	MDD and OMC	CBR=-1571 + 717MDD + 14.84OMC	0.610	1.84245
MLR 5	Gr, F ₂₀₀ , S, PL, LL, MDD and OMC	CBR =-1483 + 9.5Gr+ 6.2F ₂₀₀ + 8.9S+ 4.26LL- 5.97PL+ 289MDD+ 8.92OMC	0.792	1.62133
MLR 6	Gr, F ₂₀₀ , S, MDD and OMC	CBR=-1588 + 10.5Gr + 6.5 F ₂₀₀ + 10.0S + 294MDD + 9.25OMC	0.775	1.67606
MLR 7	Ln (Gr), Ln (F ₂₀₀), Ln (S), MDD, and OMC	CBR = -813 - 7.6 Ln (Gr) - 58.4 Ln (F ₂₀₀) - 21.3 Ln (S) + 315 MDD + 9.59 OMC	0.790	1.78415

* Significance whereas $P < 0.01$.

TABLE 7: PQ [multinomial] regression analysis significant models* .

Model id	Parameters	Model	R ²	DW
PQ 1	Ln (Gr), Ln (F ₂₀₀) and Ln (S)	$CBR = -2213 - 972 \text{Ln}(\text{Gr}) - 673 \text{Ln}(F_{200}) - 1041 \text{Ln}(S) - 222 \text{Ln}(\text{Gr})^2 - 106.0 \text{Ln}(F_{200})^2 - 246 \text{Ln}(S)^2$	0.787	1.98443
PQ 2	Ln (Gr), Ln (F ₂₀₀), Ln (S) PL, LL, MDD and OMC	$CBR = 11235 + 6.49 \text{LL} - 8.36 \text{PL} - 10817 \text{MDD} + 15.1 \text{OMC} + 150 \text{Ln}(\text{Gr}) + 50 \text{Ln}(F_{200}) - 80 \text{Ln}(S) + 2616 \text{MDD}^2 - 0.766 \text{OMC}^2 + 53 \text{Ln}(\text{Gr})^2 + 14.5 \text{Ln}(F_{200})^2 - 35.8 \text{Ln}(S)^2$	0.85	1.32803
PQ 3	Gr, F ₂₀₀ and S	$CBR = -638 + 5.2 \text{Gr} + 0.2 F_{200} + 10.8 \text{S} + 0.0366 \text{Gr}^2 + 0.136 F_{200}^2 - 0.0420 \text{S}^2$	0.794	1.85436
PQ 4	MDD and OMC	$CBR = 13434 - 13348 \text{MDD} + 28.3 \text{OMC} + 3286 \text{MDD}^2 - 1.265 \text{OMC}^2$	0.794	1.28769
PQ 5	Gr, F ₂₀₀ , S, PL, LL, MDD, and OMC	$CBR = 9184 + 5.1 \text{Gr} - 4.8 F_{200} + 2.6 \text{S} + 7.86 \text{LL} - 10.0 \text{PL} - 9284 \text{MDD} + 16.6 \text{OMC} - 0.0074 \text{Gr}^2 + 0.211 F_{200}^2 + 0.0145 \text{S}^2 + 2257 \text{MDD}^2 - 0.790 \text{OMC}^2$	0.888	1.50942
PQ 6	Gr, F ₂₀₀ , S, MDD and OMC	$CBR = 9623 + 5.9 \text{Gr} - 3.7 F_{200} + 3.5 \text{S} - 9764 \text{MDD} + 18.0 \text{OMC} - 0.0100 \text{Gr}^2 + 0.187 F_{200}^2 + 0.0103 \text{S}^2 + 2370 \text{MDD}^2 - 0.867 \text{OMC}^2$	0.871	1.64803
PQ 7	Ln (Gr), Ln (F ₂₀₀), Ln (S), MDD, and OMC	$CBR = 12059 + 236 \text{Ln}(\text{Gr}) + 89 \text{Ln}(F_{200}) - 8 \text{Ln}(S) - 11425 \text{MDD} + 14.3 \text{OMC} + 72 \text{Ln}(\text{Gr})^2 + 20.8 \text{Ln}(F_{200})^2 - 22.6 \text{Ln}(S)^2 + 2757 \text{MDD}^2 - 0.7590 \text{OMC}^2$	0.88	1.52370

*Significance whereas P < 0.01.

model reaches a level of performance that is within acceptable limits. A measurement that may be used for the examination of reliability is defined as

$$\text{Reliability} = \frac{100\%}{N} \sum_{i=1}^N k_i. \quad (8)$$

Then after we check to see if RAE_i is greater than δ , if it is, we set k_i equal to 1; otherwise, we set it equal to 0. We define k_i as the number of times the value of RAE is either less than or equal to that of δ . According to Chinese standards, the ideal value of δ is 0.2 [28].

2. Results and Discussion

2.1. Multivariate Models Development and Prediction of CBR. The data were fully analyzed for the relationship between CBR and the parameters, each one individually, followed by a multiple linear regression relationship between CBR and the parameters: grain size, the relationship between CBR and Atterberg limit, and the relationship between the compaction parameter and CBR. With the help of the correlation matrix, a certain parameter will be selected to get the best model representing this data. Each model will be evaluated based on the correlation coefficient of determination (R^2) and its level of significance ($P \leq 0.05$).

The correlation matrix in Table 5 clearly illustrates the nature of the relationship between the CBR and the soil parameters individually. As was expected, the studied area is sandy in nature. Therefore, all Atterberg coefficients appear to have a poor correlation with CBR. On the other hand, the relationship between the granular size distribution parameters, especially F_{200} , appears from the compaction parameters (MDD) such that the highest correlation is related to CBR.

2.2. Multilinear Regression. Several attempts were conducted to find the best combination of parameters that represent the CBR correlation with the soil index and compaction parameters based on the findings from the correlation matrix (Table 5). Out of several models, a possible correlation was made among those who exceeded R^2 of 0.5 and, at the same time, was statistically significant. Multilinear regression models (MLR) are presented herein. Table 6 presents the outcomes of the multilinear regressions. MLR 1 and 3 in Table 6 are controlled by the grain size distribution; those models revealed R^2 of 0.753 and 0.74. Consequently, MLR 4 presents the compaction parameters MDD and OMC as input parameters to predict the corresponding CBR; it revealed R^2 of 0.61. MLR 2, 5, and 7 incorporate the effect of all indices as predictors for CBR; a better performance is revealed from these models where their R^2 is **0.811**, 0.79, and 0.792 consequently. The combination of grain size parameters and compaction parameters in MLR.6 gives R^2 of 0.775. Consistency parameters could be considered with no contribution in MLR.6; this may be attributed to the fact that the soil in this study is nonplastic due to low plasticity soil. The best model of multivariate linear regression model is shown in (8), Table 7.

TABLE 8: GEP parameters obtained for the best three models.

Chromosomes	30
Error type fitness function	RMSE and R^2
Count of the genes	5
Head size	10
Linking function	+
Function set	+, -, ×, /, √
Mutation rate	0.044
One-point recombination rate	0.3
The recombination rate of two points	0.3
Inversion rate and transposition rate	0.1

TABLE 9: Statistical performances of the proposed GEP models.

Model id	SI	IOA	Bias
GEP1	0.252231	0.958626	-33.5488
GEP2	0.256308	0.933823	-33.7025
GEP3	0.522693	0.984099	-46.5304
GEP4	0.819813	0.928315	-16.7601

$$\begin{aligned} \text{CBR} = & -835 + 6.76\text{LL} - 8.97\text{PL} + 321\text{MDD} - 15.6 \ln(\text{Gr}) \\ & - 32.6 \ln(S) - 52.9 \ln(F_{200}), \end{aligned} \quad (9)$$

where R^2 is **0.81**, and $P < 0.05$.

2.3. PQ Models. The PQ which is none linear generally contains linear variables and constants; it is effective when such a pattern does not appear to be linear, and the relationships of the variables tend to be curvilinear [16]. Table 8 illustrates the PQ models for the prediction of CBR of granular soil. The model's significance (P) value is zero, signifying robust models.

The best model determination coefficient (R^2) is 0.888, which is better than MLR CBR model in (8). The best PQ regression model is shown in (9).

$$\begin{aligned} \text{CBR} = & 9184 + 5.1\text{Gr} - 4.8F_{200} + 2.6S + 7.86\text{LL} - 10.01\text{PL} \\ & - 9284\text{MDD} + 16.6\text{OMC} - 0.0074\text{Gr}^2 \\ & + 0.211F_{200}^2 + .0145S^2 + 2257\text{MDD}^2 - 0.79\text{OMC}^2, \end{aligned} \quad (10)$$

where R^2 is **0.888**, and $P < 0.05$.

2.4. GEP Model Development. The primary objective of developing the GEP model was to provide mathematical functions that might be used to predict the CBR value. Numerous GEP models with differing quantities of input variables were built in this study in order to estimate the CBR. Each model's number of input variables was distinct. The GEP models use the same input parameters as the multivariate models, which are also shown in Table 3. Similar to multivariate models, all GEP data were arranged into a format with seven input parameters, including percentages of gravel (Gr), sand (S), and fines (F_{200}), as well as LL, PI, MDD, and optimal water content (OMC). To create

TABLE 10: Performances of multivariate and GEP models.

Model	MAE	MSE	RMSE	MAPE	R^2	U95	Reliability
MLR1	18.177	563.803	23.745	24.28	0.75	15.940	49
MLR2	16.546	432.261	20.791	24.51	0.81	15.569	51
MLR3	19.208	593.443	24.361	28.12	0.74	16.033	49
MLR4	24.952	890.208	29.836	41.58	0.61	16.833	43
MLR5	17.783	474.159	21.775	27.34	0.79	15.699	49
MLR6	17.618	513.078	22.651	26.83	0.78	15.816	49
MLR7	16.698	478.326	21.871	24.69	0.79	15.702	53
PQ1	15.933	485.939	22.044	18.94	0.79	15.723	55
PQ2	11.560	244.783	15.646	16.75	0.89	15.022	62
PQ3	15.822	470.641	21.694	21.35	0.79	15.678	57
PQ4	16.843	470.530	21.692	26.93	0.79	15.678	51
PQ5	11.490	256.206	16.006	16.17	0.89	15.084	57
PQ6	11.744	293.372	17.128	17.16	0.87	15.174	57
PQ7	12.087	273.340	16.533	18.00	0.88	15.127	62
GEP1	18.251	509.44	22.552	24.79	0.776	15.787	53
GEP2	16.401	523.861	22.888	20.31	0.771	15.828	53
GEP3	16.878	532.636	23.0789	21.195	0.766	16.803	45
GEP4	18.471	510.715	22.599	28.393	0.776	16.766	45

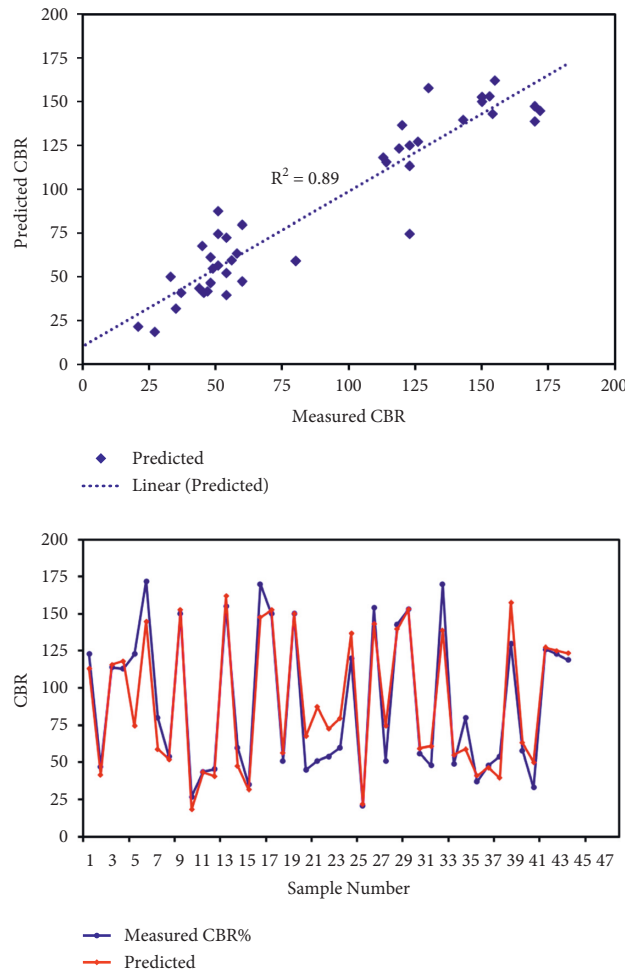


FIGURE 6: Comparison of the estimated PQ model and measured values of CBR.

the best GEP model, the number of chromosomes was altered between 20 and 35, the number of genes between 3 and 7, and the head size between 8 and 13. During the building of

GEP models, it was observed that the same GEP parameters were gathered for the “four best models”; Table 8 lists these parameters.

Results of the best models which have better statistical performance that are presented in (5)–(8), in which the input parameters and mathematical function are included. Model GEP1 has three inputs that include grain size

distribution. Models incorporate the OMC, while model GEP3 includes the parameters compaction and grain size analysis with PI. Model 4 has all indices that can affect the value of CBR.

$$\begin{aligned}
 \text{GEP.1.CBR} &= 18.0824 + 2.073\text{Gr} - F_{200} + \frac{\text{Gr}}{F_{200}} - \frac{\text{Gr} \times F_{200}}{S}, \\
 \text{GEP.2.CBR} &= 21.056 + \frac{4\text{Gr}}{\text{OMC}} + 2\text{OMC} - F_{200} + \frac{7.135\text{Gr}}{F_{200}}, \\
 \text{GEP.3.CBR} &= 3.13 + (\text{Gr} - S) - \text{PI} + 0.334(S - 8.964) - \frac{6.186(1 - \text{Gr})}{F_{200}} + \sqrt{\text{OMC}^2 \times \frac{F_{200}}{\text{Gr}}}, \\
 \text{GEP.4.CBR} &= -148 + 98.8\text{MDD} + 3.9\text{PI} - 0.8603F_{200} + \text{Gr} + \text{OMC} \\
 &\quad - 0.89\text{PL} \times \text{MDD} - \sqrt{\text{MDD} \times F_{200}} + \frac{\text{PI}}{((\text{PL}/(\text{OMC} - 3.882)) - (23.61/F_{200}))}.
 \end{aligned} \tag{11}$$

The IOA, the SI, and Bias are the three indices that are produced in order to carry out a full statistical evaluation of the models that have been offered for the GEP. Table 9 displays the findings of the statistical analysis that was carried out. According to Table 10, GEP 1 and GEP 4 have the highest coefficient of determination (R^2 of 0.776), although GEP4 has the least RMSE with a value of 22.552. According to Table 9, GEP 1 has the lowest SI of 0.252, which indicates that it has better GEP model performance than the other GEP models. Good indexes of agreement can be seen across all GEP models, with values ranging from 0.928 for GEP 4 to 0.984 for GEP 3. In addition, the predictions of CBR that were made available by GEP 1 had a higher reliability (53%) and uncertainty (15.78) in comparison to the predictions that were made available by GEP 4 (45% reliability and 16.766 uncertainty).

2.5. Models Comparison. In this section, an examination, discussion, and comparison of the results were acquired from multilinear, PQ, and GEP models. The coefficient of determination (R^2), RMSE, MAPE, and Durbin–Watson (DW) were employed as statistical verification tools to assess the accuracy of the created models' outputs. Table 10 summarizes the statistical performances of the multivariate and GEP models. According to Table 10, the majority of models exhibit acceptable agreement between the developed model and experimental values. Table 10 further shows that the performance of the majority of models derived by multilinear regression, PQ, and GEP techniques is acceptable in terms of R^2 , RMSE, MAPE, uncertainty, and reliability. This event can be regarded as the models' inputs being appropriately chosen. Table 10 presents the statistical performances of multivariate and GEP models.

However, in statistics, the total error performance of a link between two groups may be assessed using coefficient of determination (R) values. According to Benjamin [12], if a

given model provides an R value greater than 0.80, there is a strong correlation between the measured and predicted values for the whole available database of data. In addition, the statistical performance of any model is assessed using error criteria such as RMSE; MAPE, an important criterion; and R value, because a model with a high R^2 value may have a high RMSE or MAPE value. Taking into consideration these criteria, the best multivariate models are MLR2, PQ2, and PQ5. However, the four GEP models have a higher R^2 and a lower RMSE of 22.56. The error magnitudes of these models are sufficient for forecasting the CBR of soils. In addition, Model PQ5 is clearly the best model for both GEP and multivariate models based on all statistical measures. In Figures 6 and 7, the estimated CBR values generated by PQ5, of multivariate and GEP1 are visually contrasted with the observed CBR values. Figures 6 and 7 show that both the multivariate model and the GEP model make good predictions based on the experimental data.

An importance analysis of GEP is also performed to realize to what extent the input variables can affect the output CBR (see Figures 8 and 9). Gravel% and Fines% are found to be the most contributing variables to CBR, followed by Sand%, as shown in model GEP 3 figure. While in model GEP 4, where the contribution of the consistency limit is incorporated into the model, it is found that the consistency parameters are affecting the CBR significantly, which is in agreement with the study performed by El-Ashwah et al. [18].

2.6. Sensitivity Analysis. To determine the quantitatively relative implication of each input variable on the CBR, the MLR, PQ, and GEP models were employed to carry out a sensitivity analysis. The analysis is executed such that one variable of 15(6) and (8) is taken out each time to assess the effect of that variable's inclusion on output. For (5), the results of the analysis depict that F_{200} ($R^2 = 0.723$,

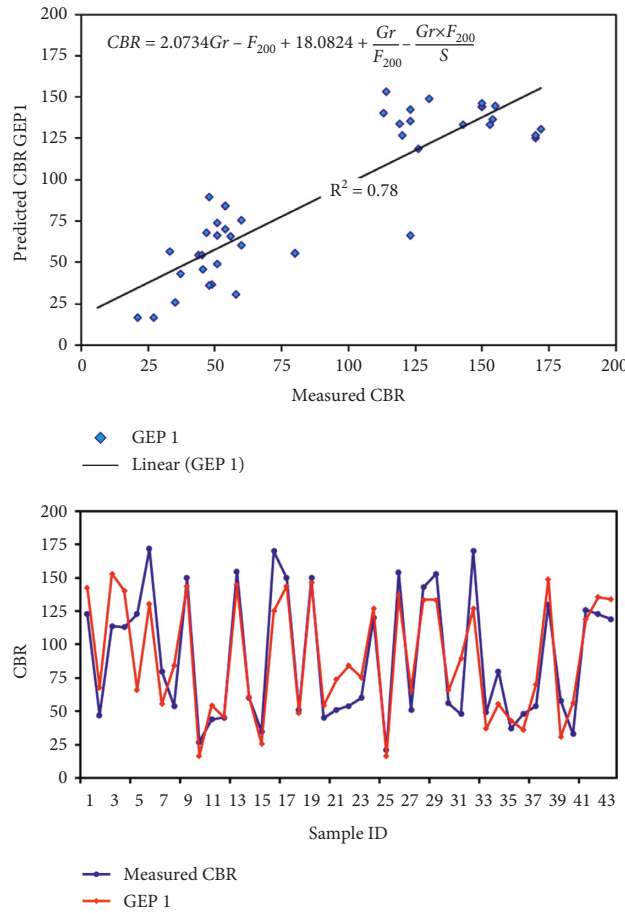


FIGURE 7: Comparison of the estimated GEP1 model and experimental values of CBR.

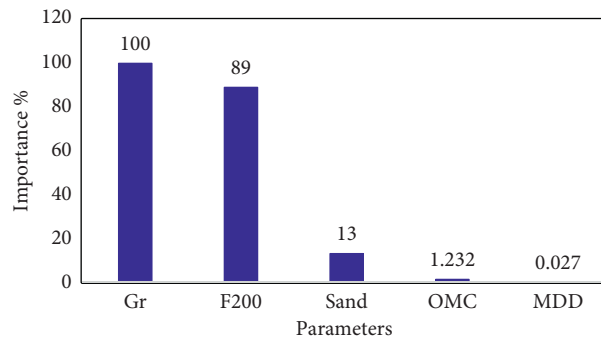


Figure 8: Importance analysis for CBR GEP 3.

RMSE = 24.94, DW = 1.696) is the highest influence variable on the CBR, while GR in which ($R^2 = 0.808$, RMSE = 20.95, DW = 1.718) has the least influence on the CBR equation based on multilinear regression. For (11), the MDD ($R^2 = 0.8162$, RMSE = 20.474, DW = 1.805) is the highest influence variable on the CBR, as well as the Gr ($R^2 = 0.8868$, RMSE = 16.06837, DW = 1.485). In (10), the most significant influencing parameter is OMC, where $R^2 = 0.721$, RMSE = 25.522, DW = 1.767, while PI is the least influencing parameter, where $R^2 = 0.788$, RMSE = 22.38, DW = 1.774. Statistical parameters resulting from sensitivity analysis of the best models are shown in Table 11.

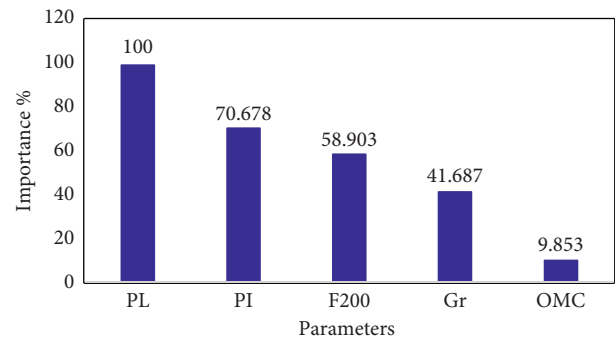


FIGURE 9: Importance analysis for CBR GEP 4.

TABLE 11: Statistical results of the sensitivity analysis.

Best model	Factor missing	RMSE	R^2	DW
MLR	GR	20.94967	0.8075	1.71827
	F ₂₀₀	24.93993	0.7273	1.69628
	S	21.59408	0.7955	1.65002
	PL	21.41879	0.7988	1.69307
	LL	21.33464	0.8004	1.6908
	MDD	22.76174	0.7728	1.81755
	OMC	22.50085	0.778	1.80682
PQ	GR	16.06837	0.8868	1.48538
	F ₂₀₀	16.75109	0.877	1.24835
	S	16.10003	0.8863	1.4504
	PL	16.95178	0.874	1.55601
	LL	16.87631	0.8751	1.53486
	MDD	20.47408	0.8162	1.80475
	OMC	16.45562	0.8813	1.68452
GEP	GR	23.649	0.762	1.609
	F ₂₀₀	23.5512	0.762	1.609
	PI	22.3814	0.788	1.774
	MDD	25.2513	0.7302	1.771
	OMC	25.5221	0.721	1.767

3. Conclusion

This study's primary objective is to investigate the applicability of multivariate regression analysis and gene expression programming (GEP) for predicting CBR. To do this, CBR test data for coarse-grained soils belonging to the A-1 and A-2 soil groups were supplied from various locations in the Qassim area. Seven MLR and PQ nonlinear regressions and four GEP models with different input variables were examined to determine the optimal connection between fundamental soil indices and the parameter CBR. The models' performance was assessed using statistical verification criteria. The models MLR, PQ, and GEP with eight input parameters yielded the greatest results. It can be concluded that the MLR, PQ, and GEP are capable of acquiring the relationship between CBR and basic soil parameters and may be used to predict CBR values of soils. The results of this analysis showed that the PQ model had a higher accuracy [coefficient of determination (R^2) = 0.89, RMSE = 16.006, uncertainty (U95) = 16.17, and reliability = 57%] than the multilinear regression model, which has a lower accuracy [R^2 = 0.811, RMSE = 20.791, U95 = 15.569, and reliability = 51%]. The best GEP model yields [R^2 = 0.776, RMSE = 22.552, U95 = 15.787, and reliability = 53%]. Furthermore, an important analysis revealed that gravel percent and fine percent have the greatest influence on CBR, followed by consistency parameters (LL, PL, PI) and OMC. Given how hard, time-consuming, and expensive CBR soil tests are, it can be emphasized that using MLR, PQ, and GEP models to estimate CBR for granular soils in terms of soil parameters could be a useful method for early stages of material development or as a way to judge the validity of CBR values. Compaction parameters (OMC and MDD) give a weaker correlation with the CBR of granular soils, but when used with gravel and fines percentage, they play a significant role in the determination of the CBR of granular soil in the

Qassim area. A sensitivity analysis was conducted to yield that F₂₀₀ is the most variable factor affecting the multilinear regression model, while MDD is the most variable factor affecting the PQ model, and the optimum moisture content (OMC) significant variable affects the best GEP model. The models used in this work to estimate CBR were based on granular-grained soils with low to nonplastic properties. As a result, it cannot accurately figure out CBR in fine soils or soils of high plasticity.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The researchers would like to thank the Deanship of Scientific Research, Qassim University, for funding the publication of this project.

References

- [1] M. AAbusogi and M. Bakri, "Behaviour of cementitious composites reinforced with polypropylene fibres using restrained eccentric ring test," *Case Studies in Construction Materials*, vol. 16, Article ID e00841, 2022.
- [2] K. Agarwal and K. Ghanekar, "Prediction of CBR from plasticity characteristics of soil," in *Proceedings of the 2nd South-East Asian Conference on Soil Engineering*, Singapore, June 1970.
- [3] H. M. Al-Hashemi and A. H. Bukhary, "Correlation between California bearing ratio (CBR) and angle of repose of granular soil," *Electronic Journal of Geotechnical Engineering*, vol. 21, no. 17, pp. 5655–5660, 2016.
- [4] K. Al-Mutlaq, A. I. Rushdi, and B. Simoneit, "Characteristics and sources of organic matter in desert sand samples from the Riyadh and Al-Qasim areas of Saudi Arabia: preliminary results," *Arab Gulf Journal of Scientific Research*, vol. 20, no. 3, pp. 141–155, 2002.
- [5] T. Al-Refeai and A. Al-Suhaibani, "Prediction of CBR using dynamic cone penetrometer," *Journal of King Saud University-Engineering Sciences*, vol. 9, no. 2, pp. 191–203, 1997.
- [6] S. K. Alam, A. Mondal, and A. Shiuly, "Prediction of CBR value of fine grained soils of Bengal Basin by genetic expression programming, artificial neural network and krigging method," *Journal of the Geological Society of India*, vol. 95, no. 2, pp. 190–196, 2020.
- [7] M. Alawi and M. Rajab, "Prediction of California bearing ratio of subbase layer using multiple linear regression models," *Road Materials and Pavement Design*, vol. 14, no. 1, pp. 211–219, 2013.
- [8] M. Aldagheiri, "The role of the transport road network in the economic development of Saudi Arabia," *WIT Transactions on The Built Environment*, vol. 107, pp. 275–286, 2009.
- [9] F. Aslam, M. A. Elkotb, A. Iqtidar et al., "Compressive strength prediction of rice husk ash using multiphysics

- genetic expression programming,” *Ain Shams Engineering Journal*, vol. 13, no. 3, Article ID 101593, 2022.
- [10] A. Bāk and R. Chmielewski, “The influence of fine fractions content in non-cohesive soils on their compactibility and the CBR value,” *Journal of Civil Engineering and Management*, vol. 25, no. 4, pp. 353–361, 2019.
- [11] M. W. A. A. Bardeesi, *Technical, Economic, and Environmental Considerations for Road Construction and Maintenance Management in*, Saudi Arabia, 2016.
- [12] J. R. Benjamin and C. A. Cornell, *Probability, Statistics, and Decision for Civil Engineers*, Courier Corporation, 2014.
- [13] W. P. M. Black, “A method of estimating the California bearing ratio of cohesive soils from plasticity data,” *Geotechnique*, vol. 12, no. 4, pp. 271–282, 1962.
- [14] F. Cheng, J. J. Cui, Q. J. Wang, and L. Zhang, “A Variable Granularity Search Based Multi-Objective Feature Selection Algorithm for High-Dimensional Data Classification,” *IEEE Transactions on Evolutionary Computation*, p. 1, 2022.
- [15] I. R. Congress, *Guidelines for the design of flexible pavements*, Vol. 37, Indian code of practice, India, 2001.
- [16] H. A. Dahish, M. Bakri, and M. S. Alfawzan, “Predicting the strength of cement mortars containing natural pozzolan and silica fume using multivariate regression analysis,” *International Journal of GEOMATE*, vol. 20, no. 82, pp. 68–76, 2021.
- [17] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, “Gaussian process regression for materials and molecules,” *Chemical Reviews*, vol. 121, no. 16, pp. 10073–10141, 2021.
- [18] A. S. El-Ashwah, E. Mousa, S. M. El-Badawy, and M. A. Abo-Hashema, “Advanced characterization of unbound granular materials for pavement structural design in Egypt,” *International Journal of Pavement Engineering*, vol. 23, no. 2, pp. 476–488, 2022.
- [19] A. El Jazouli, A. Barakat, and R. Khellouk, “Geotechnical studies for Landslide susceptibility in the high basin of the Oum Er Rbia river (Morocco),” *Geology, Ecology, and Landscapes*, vol. 6, no. 1, pp. 40–47, 2022.
- [20] N. A. Khan, N. Jhanjhi, S. N. Brohi, R. S. A. Usmani, and A. Nayyar, “Smart traffic monitoring system using unmanned aerial vehicles (UAVs),” *Computer Communications*, vol. 157, pp. 434–443, 2020.
- [21] L. Ling, G. Li, X. Zhu, and C. Zhang, “R-optimal designs for second-order Scheffé model with qualitative factors,” *AIMS Mathematics*, vol. 7, no. 3, pp. 4540–4551, 2022.
- [22] M. Najafzadeh and G. Oliveto, “More reliable predictions of clear-water scour depth at pile groups by robust artificial intelligence techniques while preserving physical consistency,” *Soft Computing*, vol. 25, no. 7, pp. 5723–5746, 2021.
- [23] M. Najafzadeh and F. Saberi-Movahed, “GMDH-GEP to predict free span expansion rates below pipelines under waves,” *Marine Georesources & Geotechnology*, vol. 37, no. 3, pp. 375–392, 2019.
- [24] K. C. Onyelowe, F. E. Jalal, M. E. Onyia, I. C. Onuoha, and G. U. Alaneme, *Application of Gene Expression Programming to Evaluate Strength Characteristics of Hydrated-Lime-Activated rice Husk Ash-Treated Expansive Soil*, Applied Computational Intelligence and Soft Computing, Egypt, 2021.
- [25] Q. B. Pham, M. Kumar, F. Di Nunno et al., “Groundwater level prediction using machine learning algorithms in a drought-prone area,” *Neural Computing & Applications*, vol. 34, no. 13, pp. 10751–10773, 2022.
- [26] G. Ramasubbarao and S. G. Sankar, “Predicting soaked CBR value of fine grained soils using index and compaction characteristics,” *Jordan Journal of Civil Engineering*, vol. 7, no. 3, pp. 354–360, 2013.
- [27] U. H. Rathod, V. Kulkarni, and U. K. Saha, “On the application of machine learning in savonius wind turbine technology: an estimation of turbine performance using artificial neural network and genetic expression programming,” *Journal of Energy Resources Technology*, vol. 144, no. 6, 2022.
- [28] F. Saberi-Movahed, M. Najafzadeh, and A. Mehrpooya, “Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions,” *Water Resources Management*, vol. 34, no. 2, pp. 529–561, 2020.
- [29] N. B. Shirur and S. G. Hiremath, “Establishing relationship between CBR value and physical properties of soil,” *IOSR Journal of Mechanical and Civil Engineering*, vol. 11, no. 5, pp. 26–30, 2014.
- [30] A. Soltani and B. C. O’Kelly, “Reappraisal of the ASTM/AASHTO standard rolling device method for plastic limit determination of fine-grained soils,” *Geosciences*, vol. 11, no. 6, p. 247, 2021.
- [31] T193, *Standard Method of Test for the California Bearing Ratio*, AASHTO, Washington, DC, 2003.
- [32] D. K. Talukdar, “A study of correlation between California Bearing Ratio (CBR) value with other properties of soil,” *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, no. 1, pp. 559–562, 2014.
- [33] V. Tanoe, S. Henderson, A. Shahirinia, and M. T. Bina, “Bayesian and non-Bayesian regression analysis applied on wind speed data,” *Journal of Renewable and Sustainable Energy*, vol. 13, no. 5, Article ID 053303, 2021.
- [34] C. Venkatasubramanian and G. Dhinakaran, “ANN model for predicting CBR from index properties of soils,” *International Journal of Civil and Structural Engineering*, vol. 2, no. 2, pp. 614–620, 2011.
- [35] B. Yildirim and O. Gunaydin, “Estimation of California bearing ratio by using soft computing systems,” *Expert Systems with Applications*, vol. 38, no. 5, pp. 6381–6391, 2011.
- [36] X. Zhu, X. Meng, and M. Zhang, “Application of multiple criteria decision making methods in construction: a systematic literature review,” *Journal of Civil Engineering and Management*, vol. 27, no. 6, pp. 372–403, 2021.