ISSN: 2666-2795 Vol. 7 No.1, 2022, Netherland

International Journal of Applied Engineering Research

Prediction of California Bearing Ratio using Gene Expression Programming: A Novel Machine Learning Approach

Shahid Ali Khan, Irshad Ahmad, Irfan Jamil, Mahmood Ahmad, Alamgir Khalil, Beenish Jehan Khan
shahidalikhan414@gmail.com, irspk@yahoo.com, irfanuop@hotmail.com, ahmadm@uetpeshawar.edu.pk, alamgirkhalil@uetpeshawar.edu.pk, beenish@cecos.edu.pk

How to Cite: Shahid Ali Khan, Irshad Ahmad, Irfan Jamil, Mahmood Ahmad, Alamgir Khalil, Beenish Jehan Khan, 2022. Prediction of California Bearing Ratio using Gene Expression Programming: A Novel Machine Learning Approach. International Journal of Applied Engineering Research, 7(1).

Date of Submission: 3rd December 2021 Revised: 15th February 2022 Accepted: 25th March 2022

***Abstract* - One of the critical parameters in designing roads, railways, earthen dams, airport pavements and railways is the bearing capacity of under laying strata that should fulfill the design criteria which in turn is highly dependent on the California Bearing Ratio (CBR). It is a very important in defining the strength of the subgrade of the aforementioned facilities. Laboratory determination of CBR in soaked conditions is very exhaustive, expansive, laborious and takes very long time (4 days) as well as needs a well-equipped experimental program. So, to side up and replace such laboratory experimental testing programs, there is need of efficient and smart computational techniques that automate the CBR determination. Gene Expression Programming (GEP), a well-known soft computing approach is employed in this study to forecast the CBR of granular materials. The GEP model was trained (70%) and tested (30%) using 168 different datasets. Liquid limit (%LL), Plastic Limit (%PL), Plasticity Index (PI), Gravel Content (%), Sand (%), Silt/Clay (%), Optimal Moisture Content (%OMC), Maximum Dry Density (MDD, lb/ft^3) were the input parameters, whereas CBR (%) was the output parameter. Mathematical equation is extracted once the model has been trained and validated. Validation of the model revealed an excellent correlation coefficient (R) of 0.97 and coefficient of determination (R^2) of 0.94. The performance evaluation of GEP model showed that it can be effectively used in the prediction of CBR.**

Index Terms – CBR, Soft computing approach, GEP, Granular materials

INTRODUCTION

keeping in view the aforementioned problem in laboratory finding of CBR, various researchers bridged that gap by developing different predictive models based on the available laboratory datasets. [1] used the algorithm called light gradient boosting machine (LGBM) for developing models for treated expensive soils with the agriculture waste. [2] studied the cohesive soils and developed models for their CBR determination based on Atterberg limits. [3] demonstrated the correlation between CBR and Atterberg limits. However, the developed expression was not well correlated and was only be able to give the approximate preliminary view of the proposed geotechnical parameters. Additionally, many authors just based his studies on simple linear regression models to predict CBR values from index properties of soil which always does not come with the greater accuracy neither having generalized equations [4]–[12]. Keeping into view the above problems with traditional regression analysis, there is a need to develop sophisticated models based on the real time dataset for more reliable prediction of CBR.

Gene expression programming is one such soft computing or machine learning strategy (GEP) which is a new variant of Genetic Programming (GP). GP is a soft computing modelling technique that is based on the Darwinian natural selection concept. GP generates machine code that can be interpreted into a mathematical equation. This makes GP perfect for mathematical modelling. GP and GEP both invented by Cramer [13], are branches of the genetic algorithm (GA), which is regarded as an evolutionary computing algorithmic technique [13], [14]. It is based on concept of Darwin's "survival of the fittest" hypothesis, which does not need making assumptions about the solution structure beforehand [15]. GP's functional process consists of several steps [16] which are; (1) according to the function and terminal settings, create an initial population (2) assess the performance of the generated population using criteria of fitness function and maximum number of generations. Terminate the program if the requirements are met by the performance of the population or maximum number of generations approaches. Otherwise, constantly create a new population using three genetic processes which are reproduction, crossover, and recombination. Create the population until and unless the threshold criteria do not meet.

In this research work, for developing GEP model, the experimental samples dataset was divided into training (70%) and testing (30%) database. The combinations of the dataset for the training and testing are varied so as to establish consistent data division. Figure 1 shows schematic view of how the input data is incorporated into GEP model so as to produce predicted model. The difference between actual and predicted values results in residual errors, which are decreased by optimizing GEP tool until an optimal model is obtained.

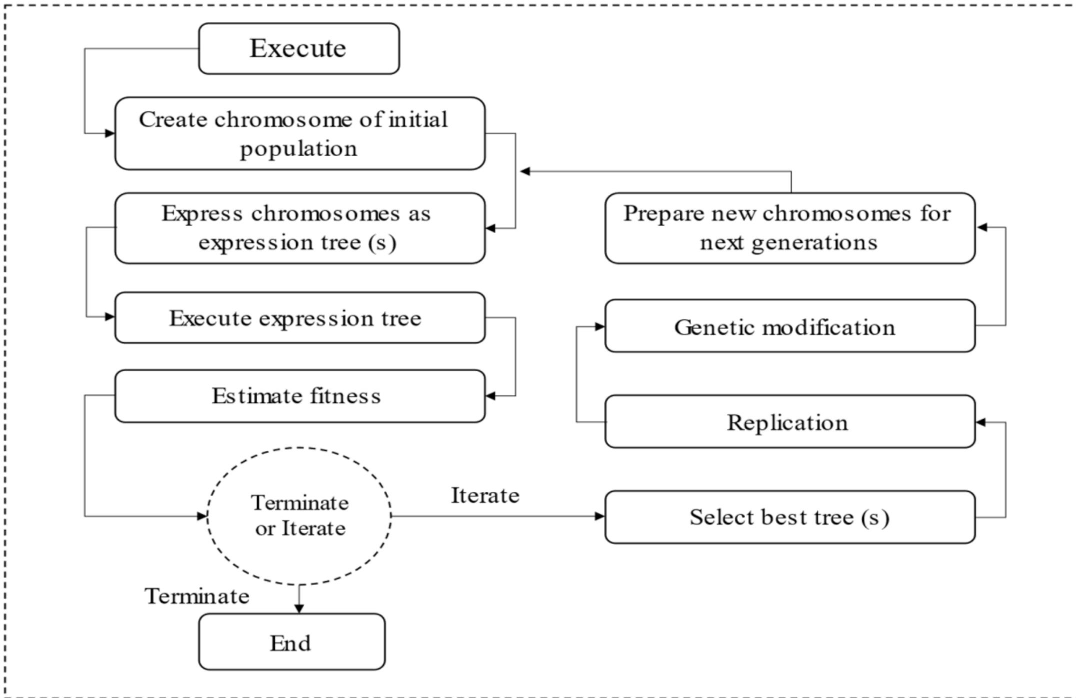

FIGURE 1
SCHEMATIC VIEW OF GEP MODELLING PROCESS

The results and performance of GEP is highly dependent on the size of the sample and variable distribution which can be seen in frequency histograms in Figures (2-10).

PREPARATION OF MATERIAL

Peshawar, Khyber Pakhtunkhwa (KP) Pakistan was chosen as the study region in this study. This study focuses on CBR data of granular material which were collected from large number of boreholes at varying depths throughout the city. The total number of data points collected were 168. Laboratory tests were conducted on the data set, including particle size analysis, Atterberg limits and compaction properties of various soil samples. Gravel content (%), sand content (%), silt/clay content (%), PI (%), MDD (lb/ft^3), OMC (%), and soaked CBR (%) are all retrieved from the laboratory results. Table 1 lists the samples dataset results that were utilized for GEP model development to determine the CBR of soil in soaked conditions.

DATA PROCESSING AND ANALYSIS

Necessary data processing elements like descriptive statistics, Pearson correlation matrix and statistical performance evaluation methods have been employed to the data for proper analysis, explained below.

TABLE 1
SOIL SAMPLES DATABASE

S.No	Predictor/Input Parameters								Output Parameter
	LL (%)	PL (%)	PI (%)	Gravel (%)	Sand (%)	Silt/Clay (%)	OMC (%)	MDD (lb/ft3)	CBR (%)
1	21	17	4	17.1	56.5	26.4	8.1	122.5	10.7
2	24	19	5	74.3	16.9	8.8	6.9	138.1	69.7
3	NP	NP	NP	74.6	22.8	2.6	5.3	135.8	52
4	26	19	7	61.3	6.9	31.8	7.1	129.2	25.9
5	21	18	3	11.8	73.4	14.8	5.3	137.8	51
6	NP	NP	NP	82.4	16.8	0.8	6.5	136.8	56.3
7	20	16	4	71.1	11.2	17.7	12.4	124.2	34.1
8	NP	NP	NP	48.8	37.3	13.9	6.9	131.3	51.2
9	28	12	16	66.5	14.4	19.1	8.7	128	22.4
10	NP	NP	NP	73.1	18.6	8.3	5.5	134.6	51.6
-	-	-	-	-	-	-	-	-	-
-	.	-	.	-	.	-	.	-	.
-	.	-	.	-	.	-	.	-	.
-	.	-	.	-	.	-	.	-	.
-	.	-	.	-	.	-	.	-	.
-	.	-	.	-	.	-	.	-	.
162	24	18	6	73.8	16	10.2	5.6	137.4	53.5
163	20	18	2	57.5	27.7	14.8	6.4	136.7	58.5
164	NP	NP	NP	38.1	23.4	38.5	8.8	132.1	45.7
165	25	19	6	62	14.7	23.3	6.9	131.5	37.3
166	NP	NP	NP	59.9	23.5	16.6	13.8	118.6	31.5
167	NP	NP	NP	73.5	15.8	10.7	5.6	141	69.8
168	NP	NP	NP	78.8	12	9.2	6.6	137.3	59.5

I. Descriptive Statistics and Statistical Visualization

Table 2 presents the descriptive statistics for the input and output parameters. The minimum and maximum ranges for all input and output parameters are shown in this statistical summary. For each parameter, the standard deviation (SD), Kurtosis, and skewness are also presented. A low SD indicates that the majority of the data are close to the mean, whereas a higher SD indicates that the numbers are more distributed. Skewness measures the asymmetry of a real-valued random variable's probability distribution with regard to its average or mean. It can be any of the following: positive, zero, negative, or undefined [17]. Negative values indicate that the tail is prolonged on the left side of the statistical distribution curve (Gravel, MDD, CBR), whereas positive values indicates that the tail is spread on the right side (sand, OMC, PI), as shown in Figures (2-10) of frequency histograms. just like skewness, the shape of a probability distribution can be expressed by Kurtosis [18]. Generally, the Correlation measure of kurtosis is taken as 3 for a particular univariate normal distribution. Platykurtic distributions have kurtosis values less than 3, implying that the distribution contains lesser and produce far fewer extreme outliers than a normal distribution.

II. Pearson Correlation Matrix

Using the statistical analysis tools, IBM SPSS, Pearson's correlation matrix was constructed from the data set, which included eight input and one output variables. A Pearson correlation matrix is a square, symmetric P x P matrix with the ijth element equal to the correlation coefficient between ith and jth variable.

TABLE 2
STATISTICAL PARAMETERS OF INPUT AND OUTPUT VARIABLES

Descriptive Statistics	Predictor Variables								Response Variable
	LL	PL	PI	Gavel	Sand	Silt/Clay	OMC	MDD	CBR
Minimum	0.00	0.00	0.00	7.50	0.00	0.00	4.50	112.30	5.60
Maximum	32.00	23.00	16.00	100.00	80.60	65.10	13.80	142.30	86.90
Sum	1903.00	1522.00	381.00	10210.10	4012.15	2577.75	1212.20	22282.30	8025.90
Mean	11.33	9.06	2.27	60.77	23.88	15.34	7.22	132.63	47.77
Median	0.00	0.00	0.00	64.45	21.05	13.40	6.80	134.50	47.45
SD	11.85	9.38	3.10	18.64	14.62	10.54	1.99	6.18	18.20
Skewness	0.16	0.09	1.55	-0.91	1.52	1.27	1.17	-0.92	-0.04
Kurtosis	-1.83	-1.95	2.50	0.63	2.64	2.82	1.13	0.24	-0.68

The correlation of variables with each other is one that's why, the diagonal members are always one. Thus, the correlations between soil input parameters and output soil are represented qualitatively in the right-hand nine columns of this correlation matrix as shown in Table 3. Correlation factors vary from -1 to 1, with 0 denoting no correlation and ±1 denoting stronger correlation. A positive number indicates a linear correlation between two variables such that the increase or decrease of one variable causes the other variable to be linearly increase or decrease at the same time.

TABLE 3
PEARSON CORRELATION MATRIX BETWEEN SOIL INPUT AND OUTPUT PARAMETERS

	LL	PL	PI	Gravel	Sand	Fines	OMC	MDD	CBR
LL	1								
PL	0.984	1							
PI	0.846	0.739	1						
Gravel	-0.001	-0.026	0.074	1					
Sand	-0.174	-0.138	-0.250	-0.826	1				
Silt/Clay	0.244	0.237	0.215	-0.624	0.074	1			
OMC	0.325	0.293	0.356	-0.245	0.005	0.426	1		
MDD	-0.208	-0.175	-0.268	0.319	-0.147	-0.360	-0.824	1	
CBR	-0.393	-0.354	-0.434	0.435	-0.239	-0.439	-0.690	0.885	1

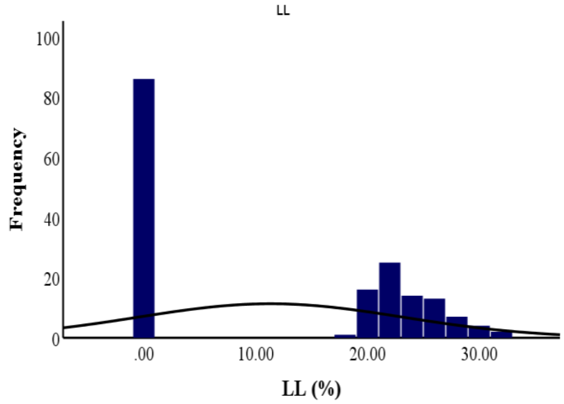

FIGURE 2
FREQUENCY HISTOGRAMS OF THE LL

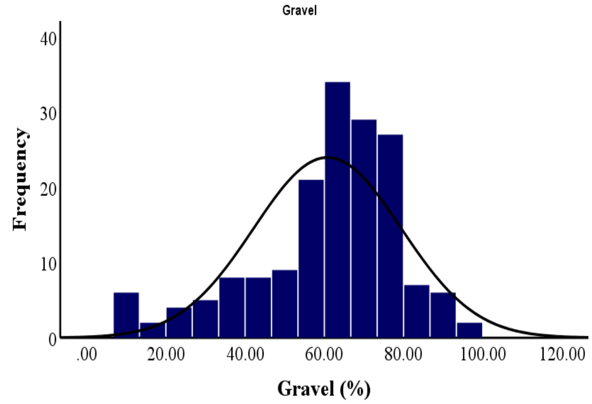

FIGURE 5
FREQUENCY HISTOGRAMS OF THE GRAVEL

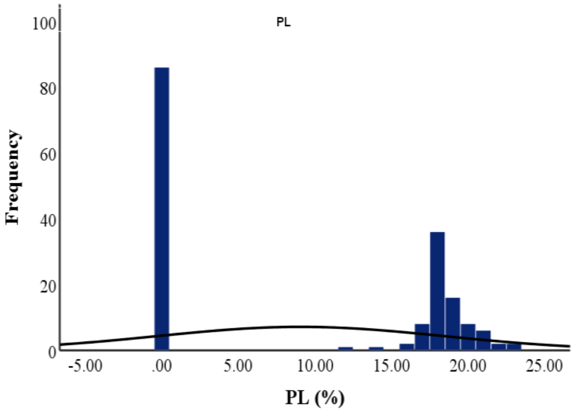

FIGURE 3
FREQUENCY HISTOGRAMS OF THE PL

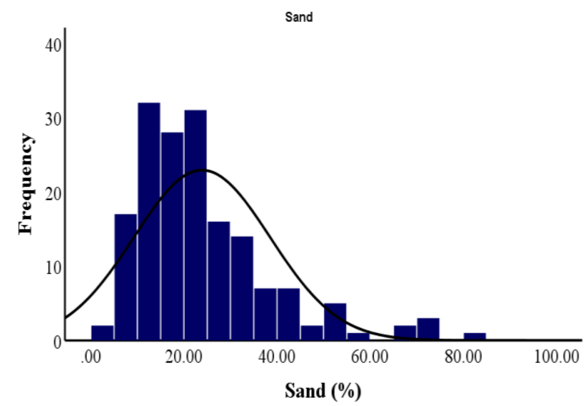

FIGURE 6
FREQUENCY HISTOGRAMS OF THE SAND

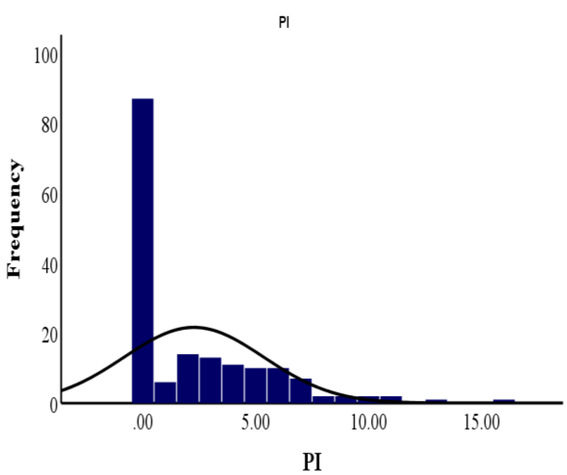

FIGURE 4
FREQUENCY HISTOGRAMS OF THE PI

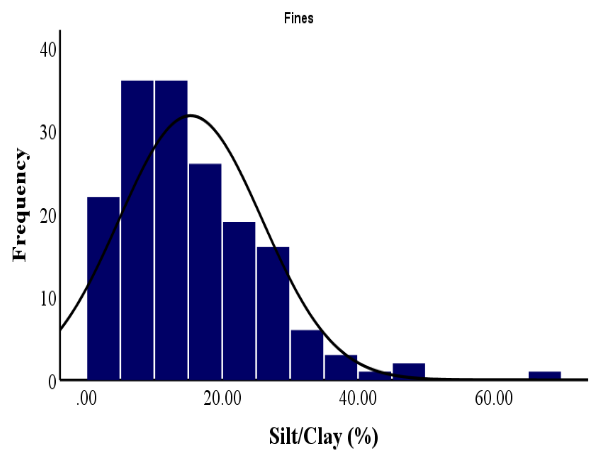

FIGURE 7
FREQUENCY HISTOGRAMS OF THE FINES

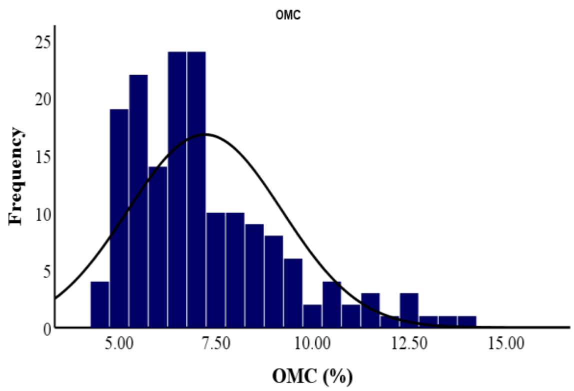

FIGURE 8
FREQUENCY HISTOGRAMS OF THE OMC

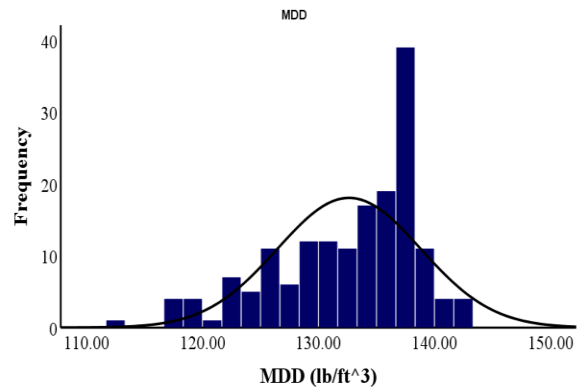

FIGURE 9
FREQUENCY HISTOGRAMS OF THE MDD

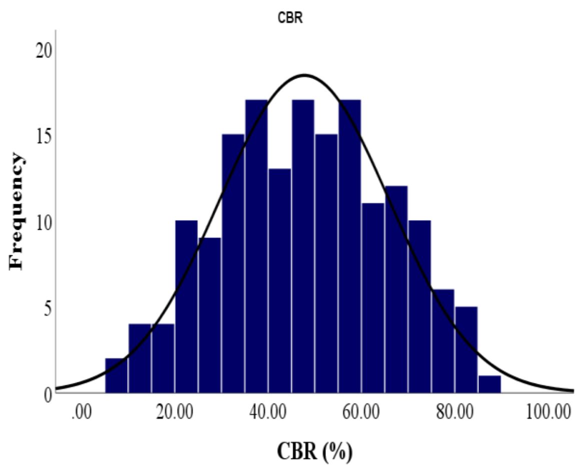

FIGURE 10
FREQUENCY HISTOGRAMS OF THE CBR

III. Performance Evaluation of Model

Correlation coefficient (R), coefficient of determination (R2), Root squared error (RSE), Nash-Sutcliffe efficiency (NSE), root mean square logarithmic error (RMSLE) and relative root mean square error (RRMSE) were used to assess the statistical performance of the GEP models in both in training and testing phase [19], [20]. The following equations describe the above statistical performance indicators.

$$R = \frac{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})(y_{p,i}-\bar{y}_{p,i})}{\sqrt{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2\ (y_{p,i}-\bar{y}_{p,i})^2}}$$

$$R = \frac{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})(y_{p,i}-\bar{y}_{p,i})}{\sqrt{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2\ (y_{p,i}-\bar{y}_{p,i})^2}}$$ (1)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{t,i}-y_{p,i})^2}{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{t,i}-y_{p,i})^2}{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2}$$ (2)

$$RSE = \frac{\sum_{i=1}^{n}(y_{p,i}-y_{t,i})^2}{\sum_{i=1}^{n}(\bar{y}_{t,i}-y_{t,i})^2} RSE = \frac{\sum_{i=1}^{n}(y_{p,i}-y_{t,i})^2}{\sum_{i=1}^{n}(\bar{y}_{t,i}-y_{t,i})^2}$$

(3)

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_{t,i}-y_{p,i})^2}{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_{t,i}-y_{p,i})^2}{\sum_{i=1}^{n}(y_{t,i}-\bar{y}_{t,i})^2}$$ (4)

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[log(y_{p,i}+1) - log(y_{t,i}+1)]^2}$$

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[log(y_{p,i}+1) - log(y_{t,i}+1)]^2}$$ (5)

Where $y_{t,i}$ and $y_{p,i}$ are the i^{th} target (experimental) and predicted variable, $\bar{y}_{t,i}\bar{y}_{t,i}$ and $\bar{y}_{p,i}\bar{y}_{p,i}$ represents average of the target and predicted variables and "n" stands for the total number of data samples. The relationship between experimental (target) and model-predicted outcomes is assessed by the "R" statistical parameter. The value of "R" greater than 0.8 depicts the strong correlation between the target and model-predicted outcomes [21]. On the other hand, "R" is no as much sensitive to the division and multiplication [22]. As a result, unbiased assessment and performance, R2 was also employed. R2 values close to one represents strong prediction of the model. The NSE must be greater than 0.65 for the model to be effective [23]. The RRMSE and RMSLE stands out prominent among the statistical measures which are also measured. Concluding that the models with lower error statistical measures (RSE, RMSLE, RMSLE) and higher NSE and correlation statistical measures (R and R2) perform better.

ARTIFICIAL INTELLIGENCE BASED ANALYSIS

I. GEP Model Development

For granular soils, numerous GEP models with varying numbers of genes were developed using a variety of genetic operators such as mutation, transposition, and crossover so as to determine the most effective GEP model. The model which was chosen initially, was composed of two genes and four head sizes (head size, H= 4) with additional connecting functions and run several times. The settings were then changed in stages, raising the number of genes to three, head size (H= 10), the chromosomes were increased to one hundred and fifty, and the weights of function sets. After several iterations, the final predicted model was checked and compared in terms of performance based on statistical performance evaluation methods listed above. Mutation rate, inversion, and recombination features were selected based on previous research [22], [24] and then evaluated to determine their optimal effect. The final mathematical model was obtained after multiple attempts, and the selected parameters are provided in Table 4 which include general information, numerical constants, and genetic operators. Based on the criteria of lesser complexity and best fitness of the mathematical formulation, the final prediction model was chosen. Figure 11 shows the expression tree (ET) for the CBR model in which d0=LL, d1=PL, d2=PI, d3=gravel, d4=sand, d5=silt/clay (fines), d6= OMC, d7=MDD and Y=CBR.

TABLE 4
PARAMETERS SETTING FOR GEP ALGORITHMS.

Parameters	Setting
General	
Training data	118
Testing data	150
Number of genes	3
Number of chromosomes	150
Head size	10
Linking function	Addition
Function set	+, −, ×, ÷, exp, Sqrt, cube root, Average
Numerical Constants	
Constants per gene	8
Type of data	Floating Type
Maximum Complexity	10
Ephemeral Random Constant	[-10, 10]
Genetic Operators	
Mutation	0.00138
Inversion Rate	0.00546
IS transposition rate	0.00546
RIS transposition rate	0.00546
One-point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene recombination rate	0.00277
Uniform recombination	0.00755

The following are the K-expressions and nodal values of gene of the ET of the modelled parameters.

3Rt.*.-.+.d4.c1.+.*.Avg2.Avg2.c6.c7.c4.d7.d6.d6.d0.d2.d7.c2.d3

\+

-.+.-.Sqrt.Avg2.+.Sqrt.d5.-.*.d6.d6.d0.c5.d3.c3.d7.d1.c6.c3.d4

\+

-.*.c3.Avg2.3Rt.d2.Avg2.*-Avg2.c6.c5.c3.d3.d6.d1.c5.d6.d0.d4.c5

Numerical Constants:

Gene 1

c0 = 6.83156834620197
c1 = -80.1336534474177
c2 = 3.65446326906915
c3 = -2.80000218662791
c4 = 3.34846034119694
c5 = -29.3313394573809

c6 = 9.91665102969451
c7 = -7.59378079690794

Gene 2

c0 = 2.16503455610828
c1 = 3.59843745231483
c2 = 1.04953154087954
c3 = 1.24454133660612
c4 = -2.31971190527055
c5 = -21.1443454232799
c6 = 0.261282082583087
c7 = -0.861537522507401

Gene 3

c0 = -9.78966396069216
c1 = -9.40061647389141
c2 = 9.67884853663747
c3 = 13.008926520661
c4 = -2.84483443708609
c5 = 7.014203646825
c6 = -4.44267014006858
c7 = 0.901867732779931

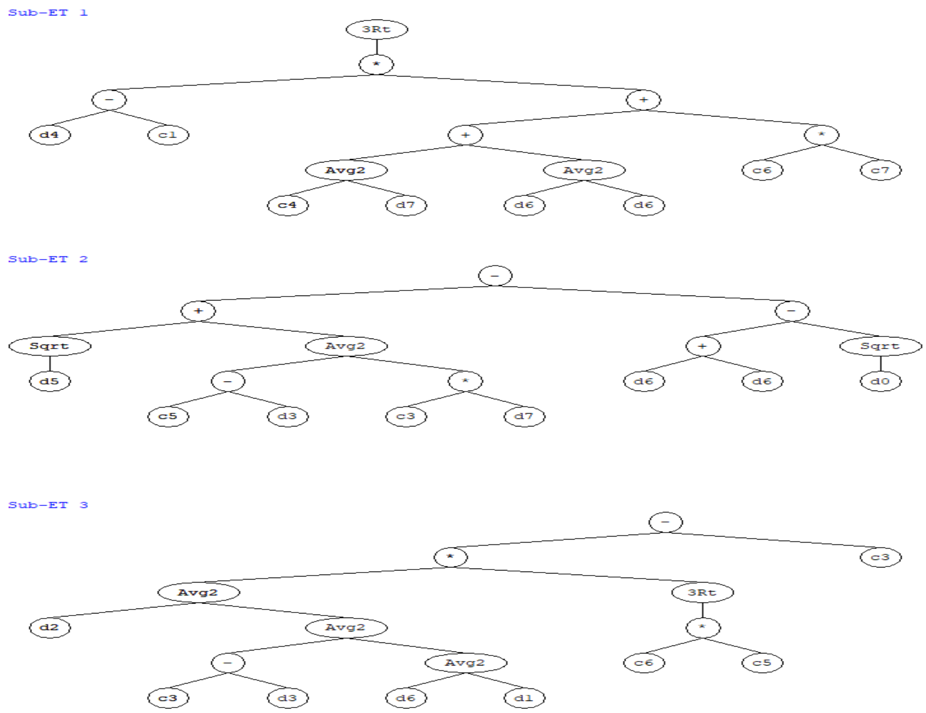

FIGURE 11
EXPRESSION TREE OF THE MODEL DEVELOPED FOR CBR

Following mathematical expression has been derived by decoding the expression tree given in Figure 11.

```
y = gep3Rt(((d[4]+80.13)*((((3.348+d[7])/2.0)+((d[6]+d[6])/2.0))-(9.916*7.6))));
y += ((sqrt(d[5])+(((-21.14-d[3])+(1.24*d[7]))/2.0))-((d[6]+d[6])-sqrt(d[0])));
y += ((((d[2]+(((13-d[3])+((d[6]+d[1])/2.0))/2.0))/2.0)*gep3Rt((-4.44*7)))+13);
```

Following Figure 12 shows the comparison between actual and GEP model predicted CBR. Higher values of the performance indicators observed for the proposed model shows that the model prediction is greatly precise and in good agreement with the training data set. A high degree of accuracy is achieved by a model with high “R” values and low RRMSE values. The MAE, RMSE, RSE, and RRMSE values for the proposed model are much lower, while the NSE, R and R^2 values are considerably larger, indicating greater performance of the model.

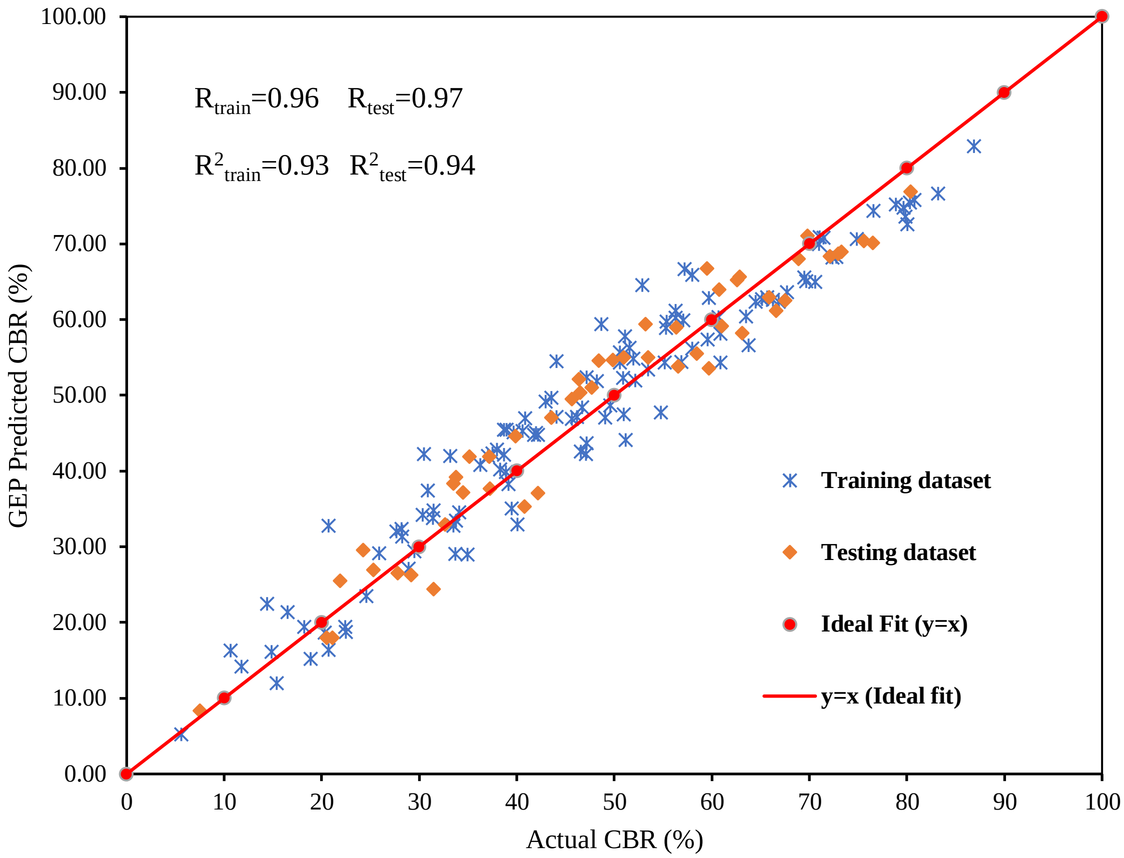

FIGURE 12
REGRESSION PLOTS BETWEEN ACTUAL AND PREDICTED OUTCOMES OF CBR

It can be seen from the graph above that training and testing data points are much closer to the linear fit reflecting the accuracy of the model achieved so that it can further be used in the prediction of CBR with satisfactory results.

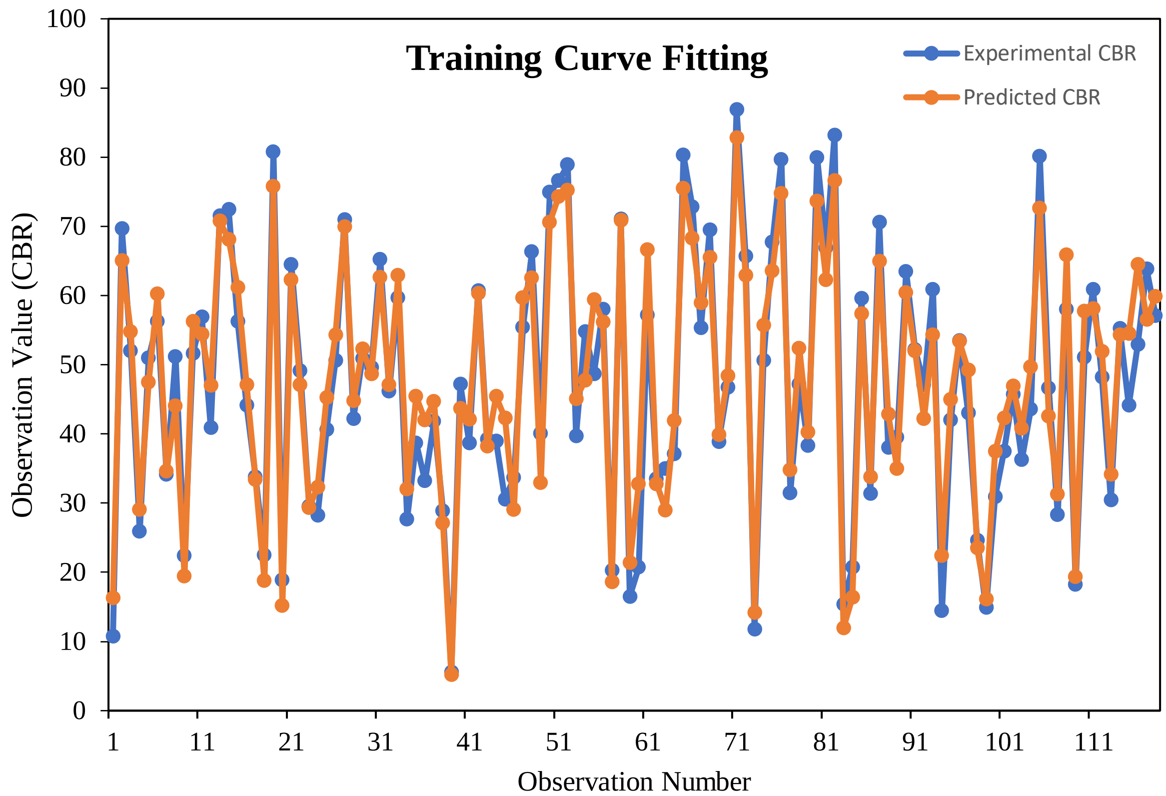

FIGURE 13
VARIATION OF ACTUAL AND PREDICTED RESULTS FOR CBR

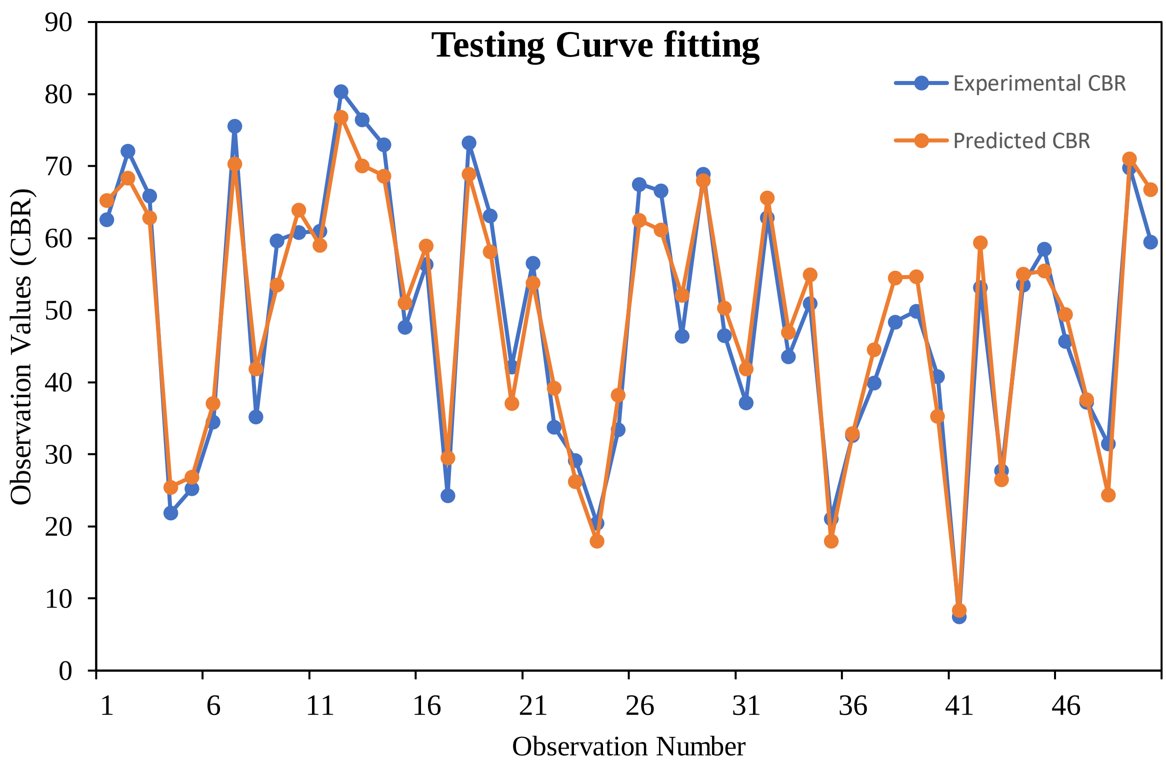

FIGURE 14
VARIATION OF ACTUAL AND PREDICTED RESULTS FOR CBR

TABLE 5
STATISTICAL PERFORMANCE MEASURES FOR GEP CBR MODEL EVALUATION

Data Type	R	R^2	RSE	NSE	RMSLE	RRMSE
Training	0.96	0.93	0.066	0.93	0.015	0.095
Testing	0.97	0.94	0.058	0.94	0.0025	0.086

The above Figure (13 & 14) depicts the comparison of the observed values of CBR from both training and testing phases for all dataset. Graphs shows a very close overlapping for both type of data

CONCLUSIONS

The following conclusions can be drawn from gene expression programming of the California bearing ratio with input parameters of LL, PL, PI, gravel, sand, silt/clay, OMC, and MDD gathered from a series of laboratory exercises yielding a total of 168 data samples:

- The liquid limit, plastic limit, plasticity index, gravel content, sand content, silt/clay content, optimum moisture content, maximum dry density, and California bearing ratio were all determined from laboratory testing of the granular soil samples according to ASTM.
- GEP evolutionary processes were used to simulate the output parameters using the predictors.
- The outcome of the GEP training and testing phase showed a consistent agreement with experimental dataset.
- Gene expression programming, a soft computing technology, can be applied with perfect accuracy for sustainable earthworks and geotechnical laboratory tasks. This is simple to execute when the materials have similar properties to those employed in this research and when the model is proposed with a similar number of predictor parameters.

CONFLICT OF INTEREST

The authors state no conflict of interest

REFERENCES

[1] V. Quan and H. Q. Do, "Prediction of California bearing ratio (CBR) of stabilized expansive soils with agricultural and industrial waste using light gradient boosting machine," *J. Sci. Transp. Technol.*, pp. 1–9, 2021.

[2] W. Black, "A method of estimating the California bearing ratio of cohesive soils from plasticity data," *Geotechnique*, vol. 12, no. 4, pp. 271–282, 1962.

[3] K. Agarwal and K. Ghanekar, "Prediction of CBR from plasticity characteristics of soil," 1970, pp. 571–576.

[4] M. Alawi and M. Rajab, "Prediction of California bearing ratio of subbase layer using multiple linear regression models," *Road Mater. Pavement Des.*, vol. 14, no. 1, pp. 211–219, 2013.

[5] W. Araujo and G. Ruiz, "Correlation equations of CBR with index properties of soil in the city of Piura," 2016, pp. 1–7.

[6] S. Bhatt, P. K. Jain, and M. Pradesh, "Prediction of California bearing ratio of soils using artificial neural network," *Am. Int. J. Res. Sci. Technol. Eng. Math.*, vol. 8, no. 2, pp. 156–161, 2014.

[7] Y. Erzin and D. Turkoz, "Use of neural networks for the prediction of the CBR value of some Aegean sands," *Neural Comput. Appl.*, vol. 27, no. 5, pp. 1415–1426, 2016.

[8] I. G. Farias, W. Araujo, and G. Ruiz, "Prediction of California bearing ratio from index properties of soils using parametric and non-parametric models," *Geotech. Geol. Eng.*, vol. 36, no. 6, pp. 3485–3498, 2018.

[9] V. Y. Katte, S. M. Mfoyet, B. Manefouet, A. S. L. Wouatong, and L. A. Bezeng, "Correlation of California bearing ratio (CBR) value with soil properties of road subgrade soil," *Geotech. Geol. Eng.*, vol. 37, no. 1, pp. 217–234, 2019.

[10] K. P. Kumar, R. Nanduri, and N. D. Kumar, "Validation of predicted California bearing ratio values from different correlations," *Am. J. Eng. Res.*, vol. 3, no. 8, pp. 344–352, 2014.

[11] M. Suthar and P. Aggarwal, "Predicting CBR value of stabilized pond ash with lime and lime sludge using ANN and MR models," *Int. J. Geosynth. Ground Eng.*, vol. 4, no. 1, pp. 1–7, 2018.

[12] Tja. Taskiran, "Prediction of California bearing ratio (CBR) of fine grained soils by AI methods," *Adv. Eng. Softw.*, vol. 41, no. 6, pp. 886–892, 2010.

[13] R. S. Faradonbeh, D. J. Armaghani, M. Monjezi, and E. T. Mohamad, "Genetic programming and gene expression programming for flyrock assessment due to mine blasting," *Int. J. Rock Mech. Min. Sci.*, vol. 88, pp. 254–264, 2016.

[14] C. Ferreira, "Gene expression programming in problem solving," in *Soft computing and industry*, Springer, 2002, pp. 635–653.

[15] Z.-L. Cheng, W.-H. Zhou, and A. Garg, "Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree," *Eng. Geol.*, vol. 268, p. 105506, 2020.

[16] A. Garg, A. Garg, and K. Tai, "A multi-gene genetic programming model for estimating stress-dependent soil water retention curves," *Comput. Geosci.*, vol. 18, no. 1, pp. 45–56, 2014.

[17] C. Sharma and C. S. P. Ojha, “Statistical Parameters of Hydrometeorological Variables: Standard Deviation, SNR, Skewness and Kurtosis,” in *Advances in Water Resources Engineering and Management*, Singapore, 2020, pp. 59–70. doi: 10.1007/978-981-13-8181-2_5.

[18] M. K. Cain, Z. Zhang, and K.-H. Yuan, “Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation,” *Behav. Res. Methods*, vol. 49, no. 5, pp. 1716–1735, 2017.

[19] I. O. Alade, M. A. Abd Rahman, and T. A. Saleh, “Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm,” *Sol. Energy*, vol. 183, pp. 74–82, 2019.

[20] D. Mohammadzadeh S, S.-F. Kazemi, A. Mosavi, E. Nasseralshariati, and J. H. Tah, “Prediction of compression index of fine-grained soils using a gene expression programming model,” *Infrastructures*, vol. 4, no. 2, p. 26, 2019.

[21] I. Azim *et al.*, “Semi-analytical model for compressive arch action capacity of RC frame structures,” 2020, vol. 27, pp. 1231–1245.

[22] M. F. Iqbal *et al.*, “Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming,” *J. Hazard. Mater.*, vol. 384, p. 121322, Feb. 2020, doi: 10.1016/j.jhazmat.2019.121322.

[23] F. E. Jalal, Y. Xu, M. Iqbal, B. Jamhiri, and M. F. Javed, “Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms,” *Transp. Geotech.*, vol. 30, p. 100608, 2021.

[24] G. B. Jumaa and A. R. Yousif, “Predicting shear capacity of FRP-reinforced concrete beams without stirrups by artificial neural networks, gene expression programming, and regression analysis,” *Adv. Civ. Eng.*, vol. 2018, 2018.

AUTHORS INFORMATION

Shahid Ali Khan, Department of Civil Engineering, University of Engineering & Technology Peshawar, 25000, Pakistan

Irshad Ahmad, Department of Civil Engineering, University of Engineering & Technology Peshawar, 25000, Pakistan

Irfan Jamil; Department of Civil Engineering, University of Engineering & Technology Peshawar, 25000, Pakistan.

Mahmood Ahmad, Department of Civil Engineering, University of Engineering & Technology Peshawar (Bannu Campus), Bannu 28100, Pakistan.

Alamgir Khalil, Department of Civil Engineering, University of Engineering & Technology Peshawar, 25000, Pakistan.

Beenish Jehan Khan, Department of Civil Engineering, CECOS University of IT and Emerging Sciences, Peshawar, 25000, Pakistan.