Geomechanics and Engineering, *Vol. 33, No. 1 (2023) 1-9*
https://doi.org/10.12989/gae.2023.33.1.001

Estimating the unconfined compression strength of low plastic clayey soils using gene-expression programming

Muhammad Naqeeb Nawaz[1], Song-Hun Chong*[1], Muhammad Muneeb Nawaz[2], Safeer Haider[3], Waqas Hassan[2] and Jin-Seop Kim[4]

[1]*Department of Civil Engineering, Sunchon National University, 255, Jungang-ro, Sunchon-si, Jeollanam-do, 57922, Korea*
[2]*NUST Institute of Civil Engineering, National University of Sciences and Technology, Islamabad 44000, Pakistan*
[3]*Department of Civil, Environmental and Architectural Engineering, University of Padua, Italy*
[4]*Radioactive Waste Disposal Research Division, Korea Atomic Energy Research Institute, Daejeon 34057, Korea*

(*Received November 28, 2022, Revised January 24, 2023, Accepted January 26, 2023*)

Abstract. The unconfined compression strength (UCS) of soils is commonly used either before or during the construction of geo-structures. In the pre-design stage, UCS as a mechanical property is obtained through a laboratory test that requires cumbersome procedures and high costs from in-situ sampling and sample preparation. As an alternative way, the empirical model established from limited testing cases is used to economically estimate the UCS. However, many parameters affecting the 1D soil compression response hinder employing the traditional statistical analysis. In this study, gene expression programming (GEP) is adopted to develop a prediction model of UCS with common affecting soil properties. A total of 79 undisturbed soil samples are collected, of which 54 samples are utilized for the generation of a predictive model and 25 samples are used to validate the proposed model. Experimental studies are conducted to measure the unconfined compression strength and basic soil index properties. A performance assessment of the prediction model is carried out using statistical checks including the correlation coefficient (R), the root mean square error (RMSE), the mean absolute error (MAE), the relatively squared error (RSE), and external criteria checks. The prediction model has achieved excellent accuracy with values of R, RMSE, MAE, and RSE of 0.98, 10.01, 7.94, and 0.03, respectively for the training data and 0.92, 19.82, 14.56, and 0.15, respectively for the testing data. From the sensitivity analysis and parametric study, the liquid limit and fine content are found to be the most sensitive parameters whereas the sand content is the least critical parameter.

Keywords: gene expression programming; external criteria; parametric study; sensitivity analysis; statistical check; unconfined compression strength

1. Introduction

Unconfined compressive soil strength (UCS) is commonly used for the analysis and design of geo-structures (e.g., shallow foundations and piles). This is directly measured from laboratory tests. However, in-situ sampling and sample preparation methods are cumbersome and expensive. As an alternate way, empirical models developed from limited datasets are commonly adopted to estimate the UCS. However, these models may not be valid for use in various soils and may be less accurate. Therefore, it is necessary to develop a reliable model involving many parameters that affect the one-dimensional soil compression response.

A recent advancement in artificial intelligence (AI) has led to the development of accurate and reliable models for engineering problems (Fattahi and Hasanipanah 2021, Kwak and Ko 2022, Luat *et al*. 2020, Moayedi *et al*. 2020, Sasmal and Behera 2021). Artificial neural networks (ANN), genetic algorithms (GA), and genetic programming are examples of artificial intelligence (AI) technologies that are based on natural analogs (GP). A problem-solving algorithm used by an ANN mimics the human brain. In comparison to multiple linear regression analysis or nonlinear multivariable regression, the predictive formulas based on ANN models are more precise. However, because of their non-transparent solutions, they are regarded as black boxes. GP is a supervised machine learning method that employs the principles of Darwin's theory of evolution. (John, 1992). A branch of GP known as gene expression programming (GEP) uses a computer program to create a solution to a problem. Using a variety of operators, populations in GEP are chosen based on the fitness function and then presented with a gene. In fact, it can establish reliable predictive models without making any assumptions about the potential structure of functional connection (Mohammadi *et al*. 2020).

Previous studies have used basic index soil properties to predict the UCS of artificial soils using artificial neural network (ANN) and multiple linear regression (MLR) techniques (Gunaydin *et al.* 2010, Sharma and Singh 2018). However, while developing these models, researchers determined the UCS using the disturbed state of soil samples, which does not represent the actual strength behavior of soils. Moreover, ANN is considered a black box and MLR relies on predefined equations and assumes the

*Corresponding author, Ph.D.
E-mail: shchong@scnu.ac.kr

ISSN: 2005-307X (Print), 2092-6219 (Online)

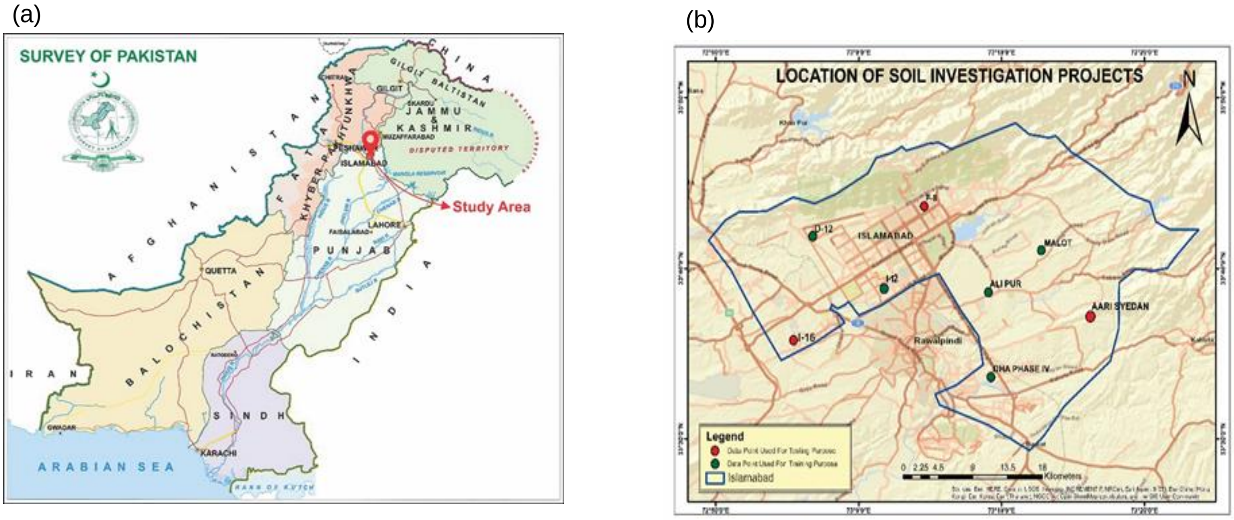


Fig. 1 Location points of eight soil investigation projects located in Islamabad, Pakistan: (a) Location map of Islamabad, Pakistan (Survey of Pakistan 2020) and (b) Location points of soil investigation projects. The points in red color represent projects of 54 training data samples, and those in green color represent projects of 25 testing data samples

normality of the residuals. Similarly, a framework based on computational intelligence approaches was proposed to estimate the UCS of soft ground soils (Narendra *et al.* 2006). However, this framework failed to estimate the UCS when the water-cement ratio was set to zero.

In addition, some previous studies have employed the GEP technique to predict the unconfined compression soil strength from different soil mixtures. The UCS of fly ash-based geopolymers was characterized, and the prediction model established from the experimental results was used to explore the sensitivity of the input parameters (Leong *et al.* 2015, 2018). An empirical model of the UCS of geopolymer concrete was proposed using ground granulated blast-furnace slag (Shahmansouri *et al*. 2020). The UCS of clay stabilization with common cementitious binders was predicted using extensive datasets including basic soil properties, binder types and contents, mixing methods, and curing periods (Pham *et al.* 2022). A UCS prediction model was developed for soils stabilized with cement by-product materials (Abbey *et al.* 2017). Although earlier research studies suggest models to predict the UCS of soils, these models consider either soil admixtures or the disturbed state of the soil, a few fundamental soil parameters, and algorithms that are challenging for practitioners to use. Thus, it is necessary to develop a UCS prediction model based on the undisturbed soil state and common influencing parameters.

This study proposes a GEP-based prediction model for the unconfined compressive soil strength based on common affecting soil properties. First, experimental studies are conducted to measure the unconfined shear strength of soils by using undisturbed soil samples. Basic soil index tests are conducted to obtain the fundamental soil properties required as input parameters. The UCS prediction model is then developed using the GEP technique. The proposed model is validated by performing statistical error checks, external criteria analysis, sensitivity analysis, and parametric analysis.

2. Research methodology

2.1 Sample collection

Site investigation projects are carried out for the construction of commercial and residential buildings located in Islamabad (the capital of Pakistan). Undisturbed soil samples are retrieved from different depths in the boreholes drilled during the soil investigation. Fig. 1 shows the project locations (Survey of Pakistan 2020). A total of 79 soil samples are collected to develop a prediction model.

2.2 Laboratory testing

The unconfined shear strength of the soil is determined using an unconfined compression test (UCT) in the laboratory following ASTM D-2166. In particular, a soil sample with a length to diameter ratio of 2 is axially compressed at a constant strain rate (1 mm/min). The maximum stress value from the stress-strain curve is manually selected as the unconfined shear strength. Soil index properties are obtained as input parameters. Sieve analysis (ASTM D 422) is performed to determine the sand content (particle sizes between 4.75 mm and 0.075 mm) and fine contents (particle size smaller than 0.075). Atterberg limits (ASTM D 4318) are adopted to evaluate the liquid limit (LL) and plastic limit (PL) to characterize soil plasticity (soil volume changes according to the water content). In addition, the specific gravity (G_s) is measured (ASTM D 854).

2.3 Laboratory test results

In the laboratory tests, soil samples are extracted from 1 m (near the surface) to 36.5 m (deep location). The highest frequency is within 15 m. The soil samples mainly consist of low-plastic clay, according to the unified soil classification system.

Table 1 Summary of the statistical analysis of experimental tests performed in this study: (a) training dataset and (b) testing dataset

(a) Training dataset	d [m]	FS [%]	FC [%]	LL [%]	PL [%]	G_s [ ]	UCS [kPa]
Maximum	36.50	59.90	97.12	36.50	19.40	2.74	264.9
Minimum	1.00	1.15	25.77	17.80	3.20	2.59	50.16
Mean	6.42	13.12	76.04	24.99	17.80	2.67	154.5
Standard Deviation	5.52	12.97	22.32	5.10	21.95	0.036	53.21
(b) Testing dataset	**d [m]**	**FS [%]**	**FC [%]**	**LL [%]**	**PL [%]**	**G_s [ ]**	**UCS [kPa]**
Maximum	31.00	84.86	95.60	34.70	20.70	2.74	248.0
Minimum	1.00	3.51	15.14	16.00	7.00	2.59	61.37
Mean	8.67	14.53	76.24	24.33	15.21	2.67	153.0
Standard Deviation	8.11	18.40	21.66	5.24	3.34	0.04	51.91

Fig. 2 presents the distribution of the soil properties obtained from the experimental results. The sand content FS varies between 1.15% and 84.9%, and the fine content FC ranges from 15% to 97%. Most of the samples have FC frequencies exceeding 50%. Respectively, liquid limit LL and plastic limit PL show relatively high values. The specific gravity has a narrow range, and the difference between the maximum and minimum values is less than 0.15. As an output parameter, the UCS varies between 50 and 265 kPa.

The dataset is separated into training and test datasets. The training data is used to train and select the optimal prediction model. The selected model is evaluated using a testing dataset. As suggested in previous studies, the standard practice is using 70% of the dataset for training and 30% for testing (Goharzay *et al.* 2017, Mohammadi *et al.* 2020). In this study, the dataset contains 54 training data points (68%) and 25 testing data points (32%). Table 1 summarizes the statistical analysis of the training and testing datasets.

3. Development of prediction model

3.1 Gene expression programming

GEP is an advanced version of genetic programming (GP) proposed by Ferreira (2006). It offers clear and straightforward mathematical equations without making any prior assumptions about the structure of a statistical model. It is composed of two components: expression trees (ETs) of various sizes and shapes and a linear chromosome with a fixed length. GEP evolved genes (sub-expression trees) are represented by tree-like structures that are linked by a simple function. Genes, chromosomes, and expression trees are regarded as the main components of the GEP. The GEP model is expressed by ETs derived from sub-ETs (genes). Numerous chromosomes are found in each gene, and each chromosome has the potential to serve as an input variable, constant value, or function. Each GEP gene is composed of a head and a tail. The head of a gene is comprised of mathematical functions, whereas its tail contains terminal symbols, such as constants and input variables. The main operators, such as selection, mutation, transposition, and crossover, tend to develop the next generation with a better fitness.

Table 2 Setting parameters for the GEP algorithm

Setting parameters	UCS model setting
Genes	4
Chromosomes	150
Head size	10
Set of functions	+, −, ×, ÷
Linking function	+
Numerical constants	
Constant per gene	10
Data type	Floating number
Upper bound	+10
Lower bound	-10
Genetic operators	
Mutation rate	0.00138
Inversion rate	0.00546
IS transposition rate	0.00546
RIS transportation rate	0.00546
One-point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene recombination rate	0.00277
Gene transposition rate	0.00277

Fig. 3 shows the general stepwise process involved in the GEP algorithm. To set up the GEP model, it is necessary to define the arithmetic operators as a function set; constants as a terminal set; statistical checks as a fitness function; number of genes, chromosomes and head size as control variables; and a termination condition. Then, GEP randomly generates chromosomes of the initial population to convert them into ETs with the use of strings by combining function sets and terminals. Subsequently, the fitness function evaluates the predicted outputs. If the proposed model satisfies fitness evaluation criteria, then the process is terminated. Otherwise, genetics is modified by evolving chromosomes or genes through genetic operators. The iteration process is continued until the predicted output meets the criteria (Iqbal *et al.* 2020).

3.2 Prediction modeling result

In this study, we adopt the GeneXpro Tools software (version 5.0) to simulate the GEP model to establish a prediction equation for UCS. To ensure the generalization and robustness of the prediction model, the numerous trial runs and error approaches are adopted to define the specific setting parameters involved in the GEP algorithm. The initial population size (number of chromosomes) affects the running time of the program. The head size controls the complexity of a model, whereas the number of genes determines the number of expression trees (ETs). The optimal combination of the setting parameters is selected through multiple initial runs. The specific parameters used in this study are listed in Table 2.

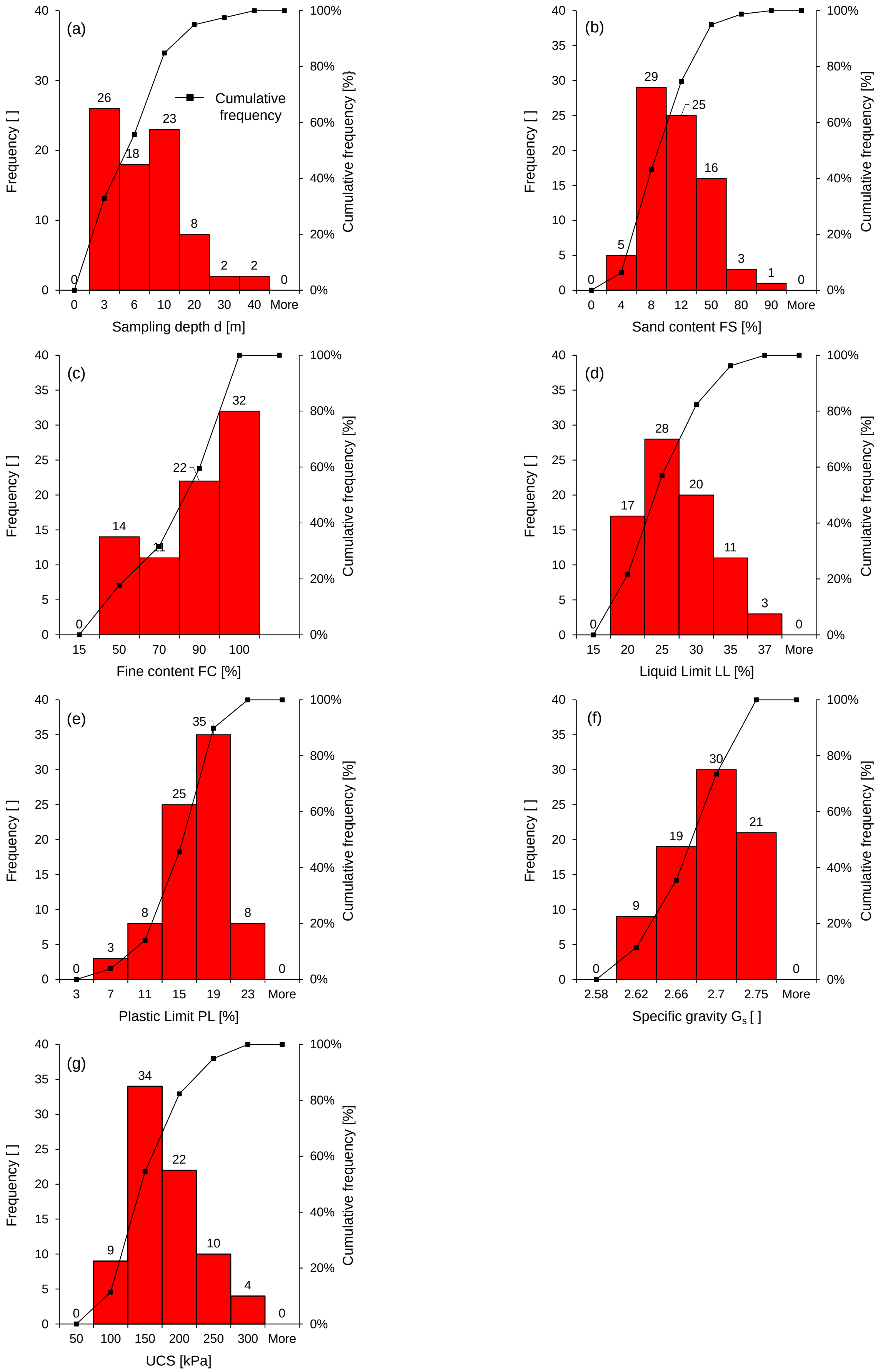


Fig. 2 Statistical distribution of data obtained from the laboratory testing of 79 soil samples: (a) sampling depth, (b) sand co ntent, (c) fine content, (d) liquid limit, (e) plastic limit, (f) specific gravity and (g) unconfined compression soil strength

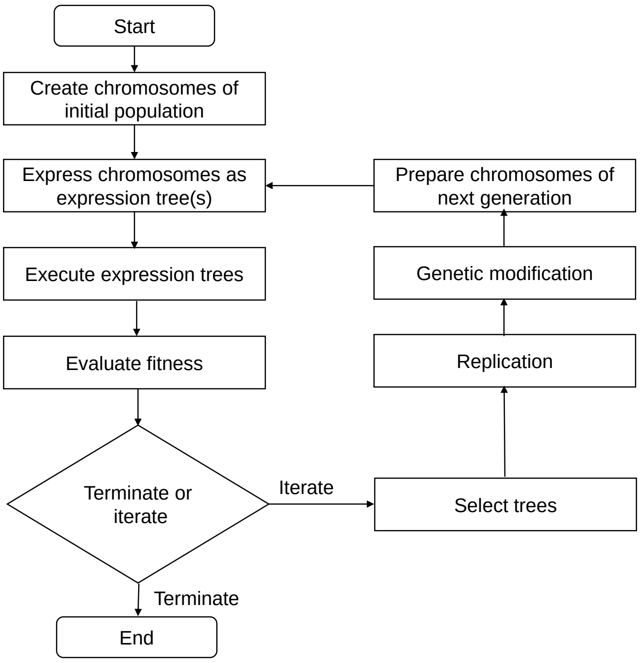


Fig. 3 Generalized step-by-step procedure of developing a prediction model based on the GEP algorithm

The correlation coefficient in simple linear regression analysis is commonly applied to determine the performance of a predicted model. However, because R is sensitive to the multiplication and division of an output to a constant, numerous other statistical checking parameters are utilized. The correlation coefficient (R), root mean square error (RMSE), relative root mean square error (RRMSE), mean absolute error (MAE), relative standard error (RSE), and performance index (ρ) are calculated for model validation. Table 3 lists the mathematical expressions for the error functions.

Fig. 4 shows the output of the GEP model for the UCS prediction model using the training data. Four expression trees (sub-ET 1, sub-ET 2, sub-ET 3, and sub-ET 4) are encoded to formulate the mathematical equations, as provided by Eqs. (1)-(5). According to the Karwa language principle, each expression tree is read from left to right and from top to bottom. It is observed that the prediction model involves all influencing input parameters.

4. Analysis and results

4.1 Performance assessment of the GEP model

The performance assessment of the prediction model is performed using statistical checks. Fig. 5 shows the comparisons between the experimental UCS_{exp} and predicted UCS_{pred} for the training and testing datasets. A 1:1 slope line and ± 5% error lines are added to interpret the acceptable errors. The data points exist within these error ranges. This implies that the experimental and predicted data have a low error.

$$UCS = A + B + C + D \tag{1}$$

Table 3 Error evaluation of a prediction model for both the training and testing datasets. n denotes the number of datasets, UCS_i^{exp} is the ith experimental UCS, UCS_i^{pred} is the ith GEP-predicted UCS, $\overline{UCS_i^{exp}}$ and $\overline{UCS_i^{pred}}$ are the mean values of the experimental and predicted UCS

Description	Statistical parameter
Correlation coefficient	$R = \dfrac{\sum_{i=1}^{n}\left(UCS_i^{exp} - \overline{UCS_i^{exp}}\right)\left(UCS_i^{pred} - \overline{UCS_i^{pred}}\right)}{\sqrt{\sum_{i=1}^{n}\left(UCS_i^{exp}-\overline{UCS_i^{exp}}\right)^2 \times \sum_{i=1}^{n}\left(UCS_i^{pred}-\overline{UCS_i^{pred}}\right)^2}}$
Root mean square error	$RMSE = \sqrt{\dfrac{\sum_{i=1}^{n}\left(UCS_i^{exp} - UCS_i^{pred}\right)^2}{n}}$
Relative root mean square error	$RRMSE = \dfrac{1}{\left\lvert\overline{UCS_i^{exp}}\right\rvert}\sqrt{\dfrac{\sum_{i=1}^{n}\left(UCS_i^{exp} - UCS_i^{pred}\right)^2}{n}}$
Mean absolute error	$MAE = \dfrac{\sum_{i=1}^{n}\left\lvert UCS_i^{exp} - UCS_i^{pred}\right\rvert}{n}$
Relative standard error	$RSE = \dfrac{\sum_{i=1}^{n}\left(UCS_i^{pred} - UCS_i^{exp}\right)^2}{\sum_{i=1}^{n}\left(\overline{UCS^{exp}} - UCS_i^{exp}\right)^2}$
Performance index	$\rho = \dfrac{RRMSE}{1 + R}$

$$A = \frac{LL}{FS - PL - 0.06d - 5.68} \tag{2}$$

$$B = FC - \left[\frac{FC \times (PL - FS + 0.47)}{4.17 \times (FS - PL - 1)}\right] \tag{3}$$

$$C = (FS + 6.41G_s - 25.59) + \left(\frac{1}{4.05}\right)(PL + d + 4.05LL) \tag{4}$$

$$D = 6.06LL + \frac{17.11}{d} - 150.96 \tag{5}$$

where UCS is the unconfined compression strength [kPa], d is the sampling depth [m], FS is the sand content [%], FC is the fine content [%], LL is the liquid limit [%], PL is the plastic limit [%], and G_s is the specific gravity [ ].

The slope of the regression line indicates the correlation coefficient R. The proposed model achieves a high R-value (R = 0.98 for the training data and R = 0.96 for the testing data). In addition, the values of RMSE, RRMSE, MAE,

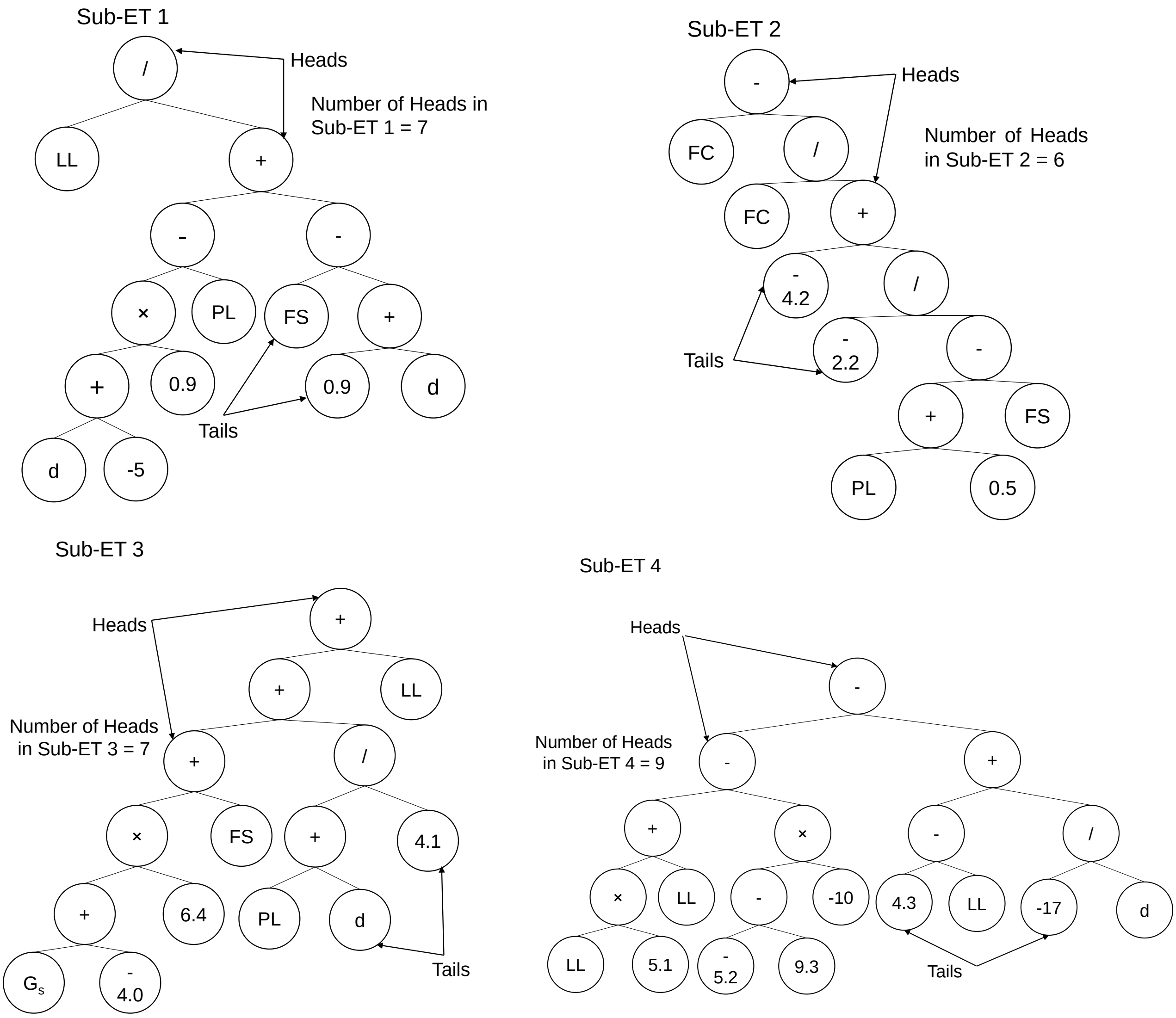


Fig. 4 Expression trees for the GEP model to predict the UCS using the training dataset consisting of 54 soil samples. The ETs are encoded from left to right and top to bottom (rule of Karwa language) to express mathematical equations

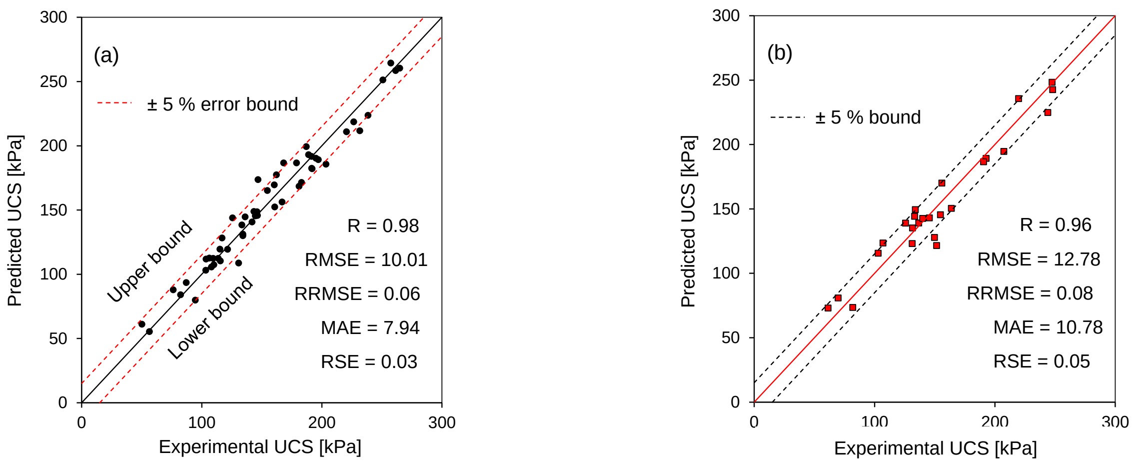


Fig. 5 Statistical analysis for comparison between the experimental data and predicted model: (a) training data (54 soil sample data from five projects) and (b) testing data (25 sample data from three projects). A value of R close to 1 shows higher accuracy. The value of R for the GEP-based UCS model against the training and testing data is 0.98 and 0.96, respectively

Table 4 Statistical parameters for the external validation of the GEP-based model

Statistical parameter	Recommended criteria	GEP-based UCS model
$k = \frac{\sum_{i=1}^{n}\left(UCS_i^{exp} \times UCS_i^{pred}\right)}{\sum_{i=1}^{n}(UCS_i^{exp})^2}$	0.85 < k < 1.15	0.994
$k' = \frac{\sum_{i=1}^{n}\left(UCS_i^{exp} \times UCS_i^{pred}\right)}{\sum_{i=1}^{n}(UCS_i^{pred})^2}$	0.85 < k' < 1.15	1.001
$R_0^2 = 1 - \frac{\sum_{i=1}^{n}\left(UCS_i^{pred} - k' \times UCS_i^{exp}\right)^2}{\sum_{i=1}^{n}\left(UCS_i^{pred} - \overline{UCS^{pred}}\right)^2}$	Close to 1	0.959
$R_1^2 = 1 - \frac{\sum_{i=1}^{n}\left(UCS_i^{pred} - k \times UCS_i^{pred}\right)^2}{\sum_{i=1}^{n}\left(UCS_i^{exp} - \overline{UCS^{exp}}\right)^2}$	Close to 1	0.960
$m = \frac{R^2 - R_0^2}{R^2}$	m < 0.1	0.0008
$n = \frac{R^2 - R_1^2}{R^2}$	n < 0.1	0.0024

RSE, and p are 10.01, 0.06, 7.94, 0.03, and 0.03, respectively for the training dataset and 12.78, 0.08, 10.78, 0.05, and 0.04, respectively for the testing dataset. Additionally, the external criteria are checked for the external validation model. It is recommended that the slopes of the regression lines (k or k') starting at the origin be close to 1.0, and R_0^1 and R_1^2 (squared correlation coefficients) be close to 1.0 (Golbraikh and Tropsha 2002). Furthermore, the values of the performance indices (m and n) must be less than 0.1. Table 4 presents the external criteria checks.

The results demonstrate that the proposed model satisfies the recommended criteria. Indeed, both statistical and external criteria checks indicate that the proposed model is efficiently trained with a high degree of accuracy and a low degree of error.

4.2 Sensitivity and parametric studies

The relative contributions of the input variables to the output parameter is explored by conducting a sensitivity analysis (SA). The sensitivity index is expressed as follows

$$SI = \frac{\sum_{i=1}^{n}\left(P_i^{exp} \times UCS_i^{pred}\right)}{\sqrt{\left(\sum_{i=1}^{n}\left(P_i^{exp}\right)^2 \times \sum_{i=1}^{n}\left(UCS_i^{pred}\right)^2\right)}} \tag{6}$$

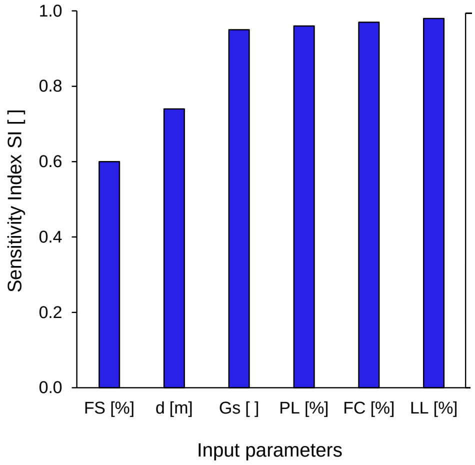


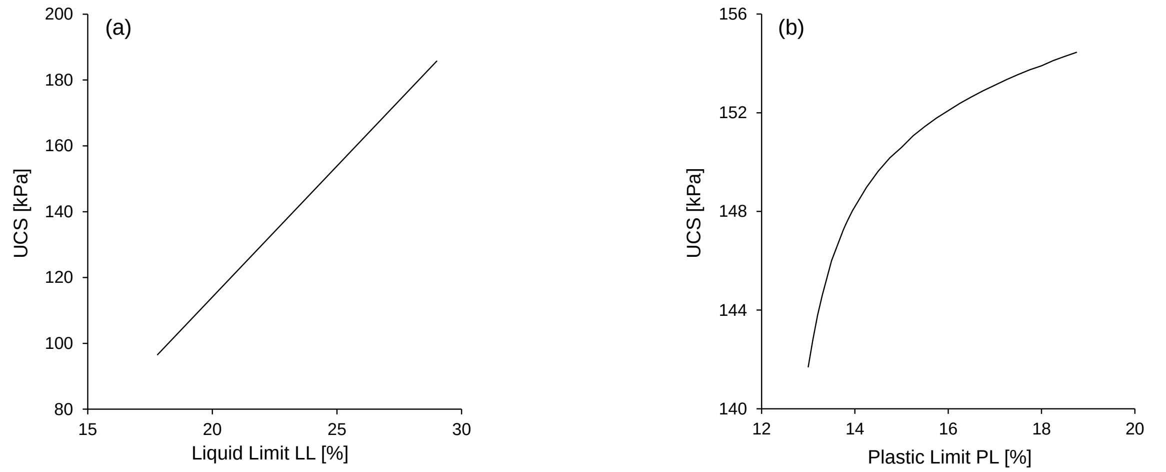
Fig. 6 Sensitivity index of individual input parameters against the UCS. The sensitivity index is obtained by using the input variables and predicted UCS from Eq. (6)

where P_i^{exp} is the input variable related to the basic soil properties, and UCS_i^{pred} is the response of the predicted UCS model. SI is the sensitivity index. Note that an SI value approaching 0 indicates no significance, whereas an SI value close to 1 shows a strong sensitivity of the parameter.

Fig. 6 shows that FC and the Atterberg limits (LL and PL) are the most sensitive parameters, whereas FS is the least critical parameter for the prediction of UCS. The order of the significance of the parameters is as LL > FC > PL > G_s > d > FS. This implies that careful consideration is required for the practical determination of critical parameters.

Furthermore, a prediction model is used to investigate the effect of each input variable on the UCS. The examined variable is varied within its statistical range, whereas the other input variables are kept constant at their average values. Fig. 7 shows the results of the parametric analysis based on the GEP model. Generally, the soil strength decreases with an increase in the Atterberg limits (Pham *et al.* 2022). However, the soil samples tested in this study are composed of a mixture of clay, sand and silt. Such a mixture with higher Atterberg limits shows greater strength (Karakan *et al.* 2020, Trask and Close 1957). The higher amount of adsorbed water between the clay particles and water enhances the interaction force between the clay interlayers and clay particles; the van der Waals attractive forces between clay particles prevail in a clay mixture (Anandarajah and Chen 1997). In other words, a higher ratio of adsorbed-absorbed water, which enhances the forces that bind the soil particles, also leads to higher Atterberg limits. Consequently, soil strength is increased. Thus, at the given ranges of the liquid and plastic limits, the UCS increases from 90 kPa to 180 kPa and from 142 kPa to 155 kPa, respectively. The specific gravity has little effect on UCS (not presented here).

Fig. 8 shows the impact of the soil particle size on UCS. The percentage of FS is maintained at different levels (10%, 20% and 30%). The fine content is varied to keep the total amount of fine and sand content in the soil below 100%.



(a) (b)

Fig. 7 Variation of the UCS against the (a) liquid limit and and (b) plastic limit. One parameter is varied between the minimum and maximum values while other parameters are kept constant at their average values

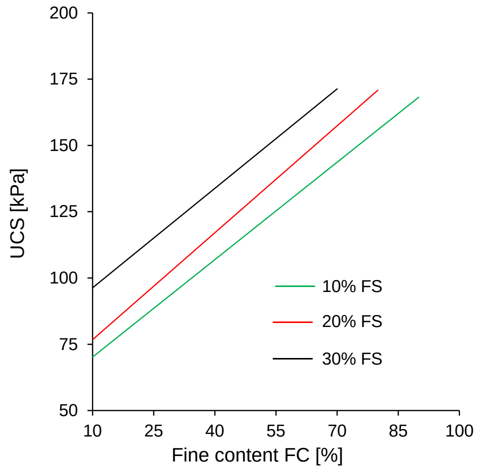


Fig. 8 Effect of sand content (FS) on the UCS. The proportion of sand is kept constant at different levels (10%, 20% and 30%). The total amount of sand and fine content in the soil is 100%. Note that other parameters are kept constant at their mean values

Note that the other parameters are fixed at their average values. At a given sand content, the UCS increases with an increase in the fine content. However, the level of strength gain depends on the sand content (10%, 20% and 30%). This is because, an increase in the sand content enhances the friction between the soil particles, whereas the fine content occupies the micropores created by the sand particles. The overall soil skeleton becomes rigid under uniaxial compression owing to a reduction in the void ratio. Thus, well graded particles generally enhance the strength of the soil.

5. Conclusions

In this study, a high accuracy prediction model is developed for the UCS prediction of soil using GEP. The experimental data is used to development a prediction model. The following conclusions can be drawn from this study:

- The proposed UCS prediction model is based on the simple index properties of soils with high accuracy and low error.
- A 1:1 slope line and ± 5% error lines are used to interpret the acceptable error in the prediction model. The results show that the predicted responses against the experimental data points are within the error ranges.
- Performance assessment of a prediction model is performed using different statistical checks. The values of R, RMSE, RRMSE, MAE, and RSE are 0.98, 10.01, 0.06, 7.94, 0.03, respectively for the training data and 0.96, 12.78, 0.08, 10.78, 0.05, respectively for the testing data. The values of R for both the training and testing datasets exceed 90%, indicating the high accuracy and strong generalization capability of the prediction model.
- The proposed prediction model is also validated using external criteria. The prediction model meets the recommended criteria. This implies that the model is efficiently trained with high accuracy and low error.
- Sensitivity analysis indicates that FC and the Atterberg limits are the most sensitive parameters, whereas FS is the least significant parameter for UCS prediction.
- A parametric study is carried out to determine the effect of individual parameters on the UCS of the soil. The results indicate that higher Atterberg limits and a larger portion of FC tend to increase the UCS.
- The UCS prediction model is appropriate for low plastic clayey soil types, in particular for the dataset domain employed in this investigation.

In further studies, the model proposed herein will be applied to other undisturbed soils and modified by considering other influencing parameters, such as the void ratio, unit weight, and moisture content.

Acknowledgments

This work was mainly supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT), No. 2021R1C1C1006003. One of the co-author "Jin-Seop Kim" was partially supported by the Nuclear Research and Development Program of the National Research Foundation of Korea (NRF-2021M2E1A1085193) funded by the Ministry of Science and ICT.

References

Abbey, S.J., Ngambi, S. and Ganjian, E. (2017), "Development of strength models for prediction of unconfined compressive strength of cement/byproduct material improved soils", *Geotech. Test. J.*, **40**(6), 928-935. https://doi.org/10.1520/GTJ20160138.

Anandarajah, A. and Chen, J. (1997), "Van der Waals attractive force between clay particles in water and contaminants", *Soils Found.*, **37**(2), 27-37. https://doi.org/10.3208/sandf.37.2_27.

Fattahi, H. and Hasanipanah, M. (2021), "Predicting the shear strength parameters of rock: A comprehensive intelligent approach", *Geomech. Eng.*, **27**(5), 511-525. https://doi.org/10.12989/gae.2021.27.5.511.

Ferreira, C. (2006), Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Vol. 21). Springer.

Goharzay, M., Noorzad, A., Ardakani, A.M. and Jalal, M. (2017), "A worldwide SPT-based soil liquefaction triggering analysis utilizing gene expression programming and Bayesian probabilistic method", *J. Rock Mech. Geotech. Eng.*, **9**(4), 683-693. https://doi.org/10.1016/j.jrmge.2017.03.011.

Golbraikh, A. and Tropsha, A. (2002), "Beware of q2!", *J. Molecular Graphics Model.*, **20**(4), 269-276. https://doi.org/10.1016/S1093-3263(01)00123-1.

Gunaydin, O., Gokoglu, A. and Fener, M. (2010), "Prediction of artificial soil's unconfined compression strength test using statistical analyses and artificial neural networks", *Adv. Eng. Softw.*, **41**(9), 1115-1123. https://doi.org/10.1016/j.advengsoft.2010.06.008.

Iqbal, M.F., Liu, Q., Azim, I., Zhu, X., Yang, J., Javed, M.F. and Rauf, M. (2020), "Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming", *J. Hazard. Mater.*, **384**, 121322. https://doi.org/10.1016/j.jhazmat.2019.121322.

Karakan, E., Shimobe, S. and Sezer, A. (2020), "Effect of clay fraction and mineralogy on fall cone results of clay–sand mixtures", *Eng. Geol.*, **279**, 105887. https://doi.org/10.1016/j.enggeo.2020.105887.

Koza, J.R. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, Mass., MIT Press.

Kwak, N.S. and Ko, T.Y. (2022), "Machine learning-based regression analysis for estimating Cerchar abrasivity index", *Geomech. Eng.*, **29**(3), 219-228. https://doi.org/10.12989/gae.2022.29.3.219.

Leong, H.Y., Ong, D.E.L., Sanjayan, J.G. and Nazari, A. (2015), "A genetic programming predictive model for parametric study of factors affecting strength of geopolymers", *RSC Advances*, **5**(104), 85630-85639. https://doi.org/10.1039/C5RA16286F.

Leong, H.Y., Ong, D.E.L., Sanjayan, J.G., Nazari, A. and Kueh, S.M. (2018), "Effects of significant variables on compressive strength of soil-fly ash geopolymer: variable analytical approach based on neural networks and genetic programming", *J. Mater. Civil Eng.*, **30**(7), 4018129. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002246.

Luat, N.V., Nguyen, V.Q., Lee, S., Woo, S. and Lee, K. (2020), "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.

Moayedi, H., Mosallanezhad, M., Rashid, A.S.A., Jusoh, W.A.W., and Muazu, M.A. (2020), "A systematic review and meta-analysis of artificial neural network application in geotechnical engineering: theory and applications", *Neural Comput. Appl.*, **32**(2), 495-518. https://doi.org/10.1007/s00521-019-04109-9.

Mohammadi, M., Fatemi Aghda, S.M., Talkhablou, M. and Cheshomi, A. (2020), "Prediction of the shear strength parameters from easily-available soil properties by means of multivariate regression and artificial neural network methods", *Geomech. Geoeng.*, 1-13. https://doi.org/10.1080/17486025.2020.1778194.

Narendra, B.S., Sivapullaiah, P.V., Suresh, S. and Omkar, S.N. (2006), "Prediction of unconfined compressive strength of soft grounds using computational intelligence techniques: A comparative study", *Comput. Geotech.*, **33**(3), 196-208. https://doi.org/10.1016/j.compgeo.2006.03.006.

Pham, V.N., Oh, E. and Ong, D.E.L. (2022), "Effects of binder types and other significant variables on the unconfined compressive strength of chemical-stabilized clayey soil using gene-expression programming", *Neural Comput. Appl.*, 1-19. https://doi.org/10.1007/s00521-022-06931-0.

Sasmal, S.K. and Behera, R.N. (2021), "Application of artificial intelligence methods for predicting transient response of foundation", *Geomec. Eng.*, **27**(3), 197-211. https://doi.org/10.12989/gae.2021.27.3.197.

Shahmansouri, A.A., Bengar, H.A. and Ghanbari, S. (2020), "Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method", *J. Build. Eng.*, **31**, 101326. https://doi.org/10.1016/j.jobe.2020.101326.

Sharma, L.K. and Singh, t.N. (2018), "Regression-based models for the prediction of unconfined compressive strength of artificially structured soil", *Eng. Comput.*, **34**(1), 175-186. https://doi.org/10.1007/s00366-017-0528-8.

Survey of Pakistan. (2020), Topographic Map of Pakistan. http://surveyofpakistan.gov.pk.

Trask, P.D. and Close, J.E.H. (1957), "Effect of clay content on strength of soils", *Coastal Engineering Proceedings*, Gainesville, Florida, **1(6)**, 50. https://doi.org/10.9753/icce.v6.50.