



Effects of binder types and other significant variables on the unconfined compressive strength of chemical-stabilized clayey soil using gene-expression programming

Van-Ngoc Pham^{1,2} · Erwin Oh¹ · Dominic E. L. Ong¹

Received: 13 August 2021 / Accepted: 4 January 2022 / Published online: 24 January 2022
© The Author(s) 2022

Abstract

Soil stabilization is an in situ soil treatment in which soils are mixed with cementitious or other chemical stabilizing agents. Determining the unconfined compressive strength (UCS) of stabilized soil is a principal task in the design and construction of the ground improvement. Hence, this study aims to develop a reliable predictive model for the UCS of clay stabilization with common cementitious binders using the gene-expression programming (GEP) technique. Eleven parameters, including the soil characteristics, the binder types, the binder contents, the mixing method, and the curing period, were considered as the independent variables in the model. The research results show that the selected optimal GEP-based model performs well with an acceptable correlation coefficient ($R = 0.951$) and low errors (e.g., RMSE and MAE). Besides, parametric analyses indicate that the plastic index, the percentage of clay, and the total water content have a negative effect on the UCS of stabilized soil. In contrast, the percentage of silt and sand, the binder types, the binder contents, and the curing time show a positive effect on the strength of stabilized soil. In addition, the strength of stabilized clay could be significantly enhanced by combining cement with slag, lime, or fly ash with a reasonable ratio, or by reducing the natural water content in the soil. The research findings could help engineers choose suitable binder types and cost-effective methods to optimize the UCS of stabilized clay.

Keywords Clay stabilization · UCS model · Chemical binder types · Gene-expression programming · Parametric study

Abbreviations

ANN Artificial neural network
CoV Coefficient of variation
ET Expression tree
FA Fly ash
GEP Gene-expression programming
GP Genetic programming
MAE Mean absolute error

PI Plastic limit
 R Coefficient correlation
RMSE Root mean square error
SD Standard deviation
 t Curing time
UCS Unconfined compressive strength

✉ Van-Ngoc Pham
ngoc.pham2@griffithuni.edu.au; pvngoc@dut.udn.vn
Erwin Oh
y.oh@griffith.edu.au
Dominic E. L. Ong
d.ong@griffith.edu.au

¹ School of Engineering and Built Environment, Griffith University, Gold Coast, Australia

² Faculty of Road and Bridge Engineering, The University of Danang, University of Science and Technology, Danang, Vietnam

1 Introduction

Soft clay accounts for a high proportion of land in coastal areas and river deltas across the world [1–3]. The characteristics of soft clay are high plasticity, high natural water content (even higher than the liquid limit), low shear strength (< 40 kPa), and high compressibility [4]. Such a kind of soft ground is insufficient for supporting the heavy loads; it leads to high settlement and affects the stability of infrastructures [5]. Hence, dealing with soft clay has been challenging for geotechnical engineers [1].

Soil stabilization is one of the most popular and effective soil improvement techniques [2]. Cement, lime, fly ash, and blast-furnace slag are the common additives used to mix with soft clay [6] to improve the workability and compaction characteristics, increase the shear strength of the soil, and reduce the settlement of the ground [7]. Due to such advantages, soil stabilization with chemical binders has been applied widely in many countries [8–10].

It has been reported that some factors could affect the characteristics of soil stabilization, including the soil characteristics, the type of binders, the binder content, the water content, the mixing method, the curing time, and others. Soil characteristics, such as soil type, organic content, grain size distribution, pH, and natural water content, affect the ultimate strength of the soil mixing column [10]. For soft soils, the natural water content is also an important factor. If the water content is higher than the liquid limit, the strength of the stabilized soil will decrease [10]. Moreover, the amount of binder added plays an essential role in developing the final strength of stabilized soil [10]. By raising the amount of the stabilization agent, the UCS of the stabilized soil increases, and the permeability decreases [10–12]. Besides, alternative pozzolans, such as fly ash, lime, and blast-furnace slag, have been used in soil stabilization. These pozzolans can increase the strength and reduce the permeability of treated soil [13, 14]. Furthermore, the UCS of soil–cement material is affected by the mixing method and the curing time [10].

Previously, choosing the binder type and the desired amount of binder has been conducted by creating and testing a thousand trial specimens. The step requires huge effort with high cost and time. Especially, this process has to conduct separately for each project with different input parameters. Therefore, some studies have suggested predictive models for the UCS of stabilized soil based on the common input variables, such as the water to binder ratio, the binder content, and the curing time [15–20].

The normalized empirical models were used to develop the predictive equations based on the experimental data. Abrams' law applied in concrete technology shows that the strength of hardened concrete could be predicted through the water to cement ratio. Liu [21], Horpibulsk [2], and Cong [3] applied the Abrams' law in predicting the strength of soil–cement material based on the ratio of clay water content (including natural water in clay and water in cement slurry) and cement content. Horpibulsk [2], Tsuchida [22], and Yao [1] also suggested new formats of the predictive formulas using the empirical models.

However, according to Narendra [5], there are some limitations in the empirical models. Firstly, these models are usually developed based on several assumptions, simplifications, and approximations. The normalized empirical models are developed based on a small volume of

experimental data points, which could be less accurate and not valid for applying other clay conditions. Furthermore, the empirical formula provides significant errors when applied to similar soil properties. In addition, most predictive formulas consider a few variables, such as the cement content, the water to cement ratio, and the curing time, while the effects of soil characteristics and binder types have not been examined.

It can be seen that an advantage predictive model that meets these requirements has not been found yet. Furthermore, there are limited published studies considering the effect of several types of binder additives on the strength of stabilized clay. As a result, developing a reliable model which could apply to a wide range of clay conditions and consider the effects of common binders by reliable modeling technologies, such as artificial neural network and genetic programming, is beneficial.

Artificial neural network (ANN) is a problem-solving algorithm that simulates the structure of the human brain. In terms of chemical soil stabilization, there have been some studies using the ANN models for predicting the UCS value, such as Das [15], Tinoco [23], Sunitsakul [17], Abbey [18], Ghorbani [19], and Saadat [20]. The predictive formulas developed based on the ANN models are more accurate than nonlinear multivariable regression or multiple regression analyses with high performance [24]. However, ANN is considered to be a “black box” program. The predictive equations are developed based on the complex transfer functions, such as logistic sigmoid and hyperbolic tangent sigmoid functions. As a result, the ANN-based predictive functions are limited in their application as they cannot be used conveniently to calculate the output using the input values [24].

Genetic programming (GP) is a kind of supervised machine learning technique that applies the principles of Darwin's evolution theory [25]. It is another alternative approach to behavior modeling. Gene-expression programming (GEP) is a branch of GP that develops a solution to a problem using a computer program [26], and it is the method that has been used commonly in geotechnical engineering [27]. In GEP, populations are also selected based on fitness function and presented with a gene through several operators [28]. GEP is able to make strong predictive functions without a preliminary assumption about the possible structure of functional connection [29]. The GEP model is a robust, powerful, and accurate predictive tool. In addition, the GEP-based formulas are transparent and more practical than the ANN-based equations. Hence, the proposed predictive equations formulated from the GEP model could be ready to apply in practice. As a result, the GEP technique was used to develop a model for predicting the UCS of soil stabilization based on a comprehensive database gathered from the literature.

The study focuses on clay stabilization with different cementitious binders, including ordinary Portland cement (CEM I), quick lime, fly ash Types C and F, and blast-furnace slag. Both wet and dry mixing methods were considered. The research results apply to determine the UCS in the laboratory condition only. The GEP technique was applied to generate a predictive model. A parametric study was also conducted to examine the effects of each variable, the effects of binder types, the combination of binders, and the total water content on the UCS of stabilized clay.

Within the scope of this study, the effect of the chemical composition of clay, including the content of CaO and SO_4^{2-} , organic content, and pH of the soil, was excluded. The specimen preparation processes, such as the size and shape of the mold, the sample making and curing methods, were assumed to be similar. These parameters could be analyzed in subsequent studies.

2 Data preparation

The database used in this study was gathered from several experimental studies in the literature. The experimental data related to the UCS of clay stabilization were chosen consistently based on the following criteria. The study examined the chemical-treated clayey soil only, while sandy soil was not considered in this research. Common types of chemical binders used to stabilize soil, including cement (CEM I), quick lime, fly ash (Type F and Type C), and blast-furnace slag, were all selected. The UCS results from the studies with similar sample making and testing standards were chosen.

Table 1 shows the selected data sources available from reliable published journal articles, including Bolton [30], Ge [31], JGS [32], Xiao [33], Naveena [6], Asgari [34],

Correia [35], Oh [36], Tastan [37], Kwan [38], Consoli [39], Kassim [40], and Abbey [18]. The database includes the experimental data for soil stabilization applied for some types of clay in different countries, such as Japan, Singapore, Thailand, Malaysia, India, Australia, the UK, the USA, Portugal, Iran, Brazil, and Taiwan. Finally, approximately 1183 data points were selected for developing a GEP-based formula for estimating the strength of soil stabilization.

The number of independent variables was chosen based on the literature review and several trials. In this research, eleven independent variables considered in the predictive model are:

- Group 1 represents the soil characteristic: the plastic index (PI), the percentage of clay (Clay), the percentage of silt (Silt), the percentage of sand (Sand);
- Group 2 represents the mixing method and curing time: the total water content (Total water), the curing period (Age). The total water content includes the natural water content in the clay and the water used to mix with binders.
- Group 3 represents the binder types and binder content: the lime content (Lime), the cement content (Cement), the fly ash Type F content (FA F), the fly ash Type C content (FA C), and the slag content (Slag).

Table 2(a) and (b) presents the range of variables from the dataset. Table 3 (a) and (b) shows the statistical analysis, including the maximum, minimum, range, mean, standard deviation (SD), and coefficient of variation (CoV) of all variables. The maximum value of the total water content is 265%, while the longest curing period is 360 days. The highest percentages of lime, cement, fly ash Type F, fly ash Type C, and slag used are 20%, 100%, 34.5%, 30%, and 42.5%, respectively. The maximum achievable strength in this study is approximate 6000 kPa.

Table 1 Sources of experiment data from the literature

| Reference | Nation | Clay type | Binder type | Dataset |
|--------------|------------------------|-------------|-------------------------|---------|
| Bolton [30] | Australia and Thailand | Clay | Lime, cement | 57 |
| Ge [31] | Taiwan | Kaolinite | Cement, slag | 76 |
| JGS [32] | Japan | Clay | Lime | 10 |
| Xiao [33] | Singapore | Marine clay | Cement, fly ash F | 41 |
| Naveena [6] | India | Sandy clay | Lime, cement, fly ash F | 352 |
| Asgari [34] | Iran | Clay | Lime, cement | 135 |
| Correia [35] | Portugal and Japan | Clay | Cement, slag | 15 |
| Oh [36] | Australia | Clay | Lime, cement | 33 |
| Tastan [37] | The USA | Clay | Fly ash F and C | 26 |
| Kwan [38] | Australia | Clay | Cement, fly ash F, slag | 73 |
| Consoli [39] | Brazil | Sandy clay | Lime | 60 |
| Kassim [40] | Malaysia | Clay | Lime | 45 |
| Abbey [18] | The UK | Clay | Cement, fly ash C, slag | 260 |

Table 2 Range of variables from input and output data

| Reference | Input variables | | | | | |
|--------------|-----------------|------------|-----------|----------|-----------------|------------|
| | PI (%) | Clay (%) | Silt (%) | Sand (%) | Total water (%) | Age (Days) |
| <i>(a)</i> | | | | | | |
| Bolton [30] | 61–65 | 35–69 | 8–28 | 3–57 | 77.5–105.5 | 7–168 |
| Ge [31] | 53 | 69 | 25 | 6 | 151.2–200 | 3–56 |
| JGS [32] | 47.4–59.4 | 45.5–71.9 | 26.1–44.6 | 1–9.9 | 90.2–160 | 7–28 |
| Xiao [33] | 43 | 45.9 | 47.7 | 6.4 | 120–200 | 7–150 |
| Naveena [6] | 24–78 | 22–31 | 14–27 | 42–64 | 21.3–265 | 7–56 |
| Asgari [34] | 9 | 15 | 39 | 46 | 16.5–20.5 | 7–60 |
| Correia [35] | 28 | 10 | 71 | 19 | 96–177 | 2.5–28 |
| Oh [36] | 38 | 35 | 8 | 57 | 71.5–74.8 | 7–56 |
| Tastan [37] | 30–34 | 15–55 | 10–42 | 3–75 | 31–63 | 7 |
| Kwan [38] | 47 | 25 | 70 | 5 | 100–115 | 3–271 |
| Consoli [39] | 9 | 4.3 | 42 | 53.7 | 14 | 28–360 |
| Kassim [40] | 28–50 | 11–44 | 27.7–68.2 | 0.4–43.2 | 33–121 | 7–56 |
| Abbey [18] | 12–34 | 30–60 | 30–60 | 10 | 36–106 | 7–56 |
| Reference | Input variables | | | | | Output |
| | Lime (%) | Cement (%) | FA F (%) | FA C (%) | Slag (%) | |
| <i>(b)</i> | | | | | | |
| Bolton [30] | 0–15 | 0–15 | 0 | 0 | 0 | 34–1422 |
| Ge [31] | 0 | 2.3–11.3 | 0 | 0 | 12.8–42.5 | 106–4008 |
| JGS [32] | 10–20 | 0 | 0 | 0 | 0 | 250–3720 |
| Xiao [33] | 0 | 13.1–100 | 0–34.5 | 0 | 0 | 199–5840 |
| Naveena [6] | 0–10 | 0–35 | 0–30 | 0 | 0 | 48–1676 |
| Asgari [34] | 0–9 | 0–7 | 0 | 0 | 0 | 230–5360 |
| Correia [35] | 0 | 6.8–20.3 | 0 | 0 | 2.3–6.8 | 90–1995 |
| Oh [36] | 0–15 | 0–15 | 0 | 0 | 0 | 42–445 |
| Tastan [37] | 0 | 0 | 0–30 | 0–30 | 0 | 45–490 |
| Kwan [38] | 0 | 3.8–15 | 0–11 | 0 | 0–11.3 | 14–1572 |
| Consoli [39] | 3–11 | 0 | 0 | 0 | 0 | 200–3520 |
| Kassim [40] | 0–12 | 0 | 0 | 0 | 0 | 25–413 |
| Abbey [18] | 0 | 0–20 | 0 | 0–10 | 0–6.7 | 7–1300 |

PI: plastic index, FA: fly ash, UCS: unconfined compressive strength

It indicates that the database could represent a wide range of input variables. Besides, the high volume of data (1183 data points) brings outstanding results in comparison with previous studies.

3 Model development

3.1 Gene-expression programming

Gene-expression programming (GEP) was developed by Ferreira [41]. It consists of two parts: a linear chromosome of fixed length and parses trees in different sizes and shapes (expression trees—ETs) [41, 42]. It also contains a

terminal set, function set, fitness functions, and termination functions [43]. GEP evolves several genes (sub-ETs) represented as tree-like structures, and they are connected by a linking function.

The main elements of GEP are expression trees (ETs), genes, and chromosomes. The GEP model is expressed by ET, which includes some sub-ETs (genes). Each gene contains several chromosomes, while each chromosome could be an input variable, constant value, or function [44]. Each GEP gene is composed of a head and a tail. The head of the gene contains mathematical functions and terminal symbols, while the tail contains terminal symbols like constant values or variables [45]. Constant values are used to adjust the equations in the model.

Table 3 Statistical analysis of input and output data

| Parameter | Input variables | | | | | |
|------------|-----------------|------------|----------|----------|-----------------|------------|
| | PI (%) | Clay (%) | Silt (%) | Sand (%) | Total water (%) | Age (Days) |
| <i>(a)</i> | | | | | | |
| Maximum | 78 | 71.9 | 74.7 | 75 | 265 | 360 |
| Minimum | 5 | 4.3 | 8 | 0.4 | 14 | 2.49 |
| Range | 73 | 67.6 | 66.7 | 74.6 | 252.36 | 357.51 |
| Mean | 38.17 | 30.92 | 38.68 | 30.40 | 93.79 | 35.84 |
| SD | 24.87 | 16.62 | 19.27 | 22.37 | 67.53 | 55.26 |
| CoV | 618.34 | 276.35 | 371.22 | 500.45 | 4560.87 | 3053.18 |
| Parameter | Input variables | | | | | Output |
| | Lime (%) | Cement (%) | FA F (%) | FA C (%) | Slag (%) | UCS (kPa) |
| <i>(b)</i> | | | | | | |
| Maximum | 20 | 100 | 34.5 | 30 | 42.5 | 5840 |
| Minimum | 0 | 0 | 0 | 0 | 0 | 7.0 |
| Range | 20 | 100 | 34.5 | 30 | 42.5 | 5833 |
| Mean | 1.89 | 8.27 | 3.68 | 0.94 | 2.87 | 613.12 |
| SD | 3.52 | 11.78 | 8.00 | 3.16 | 9.15 | 740.4 |
| CoV | 12.41 | 138.70 | 63.94 | 10.00 | 77.8 | 548,503 |

SD: standard deviation and CoV: coefficient of variation

The main operators of GEP are selection, mutation, transposition, and crossover, which are similar to traditional GP. They allow the program to produce the next generation with better fitness scores [46]. The explanation and detail of the GEP structure, as well as its operators, could be accessible through the studies of Soleimani [47], Gandomi [24], and Shahmansouri [28].

For setting up the GEP modeling, it is necessary to define the function set, terminal set, fitness function, control variables, and termination condition for obtaining a solution. GEP then randomly creates an initial population. Chromosomes in that population are converted into an expression tree (ET) by combining terminal and function sets. Next, the fitness function is applied to evaluate each predicted output. If that value does not meet the desired output, chromosomes or genes are evolved through genetic operators (selection, crossover, and mutation) to create new mutagenic generations [48]. That process is stopped when the predicted output meets the desired quality output.

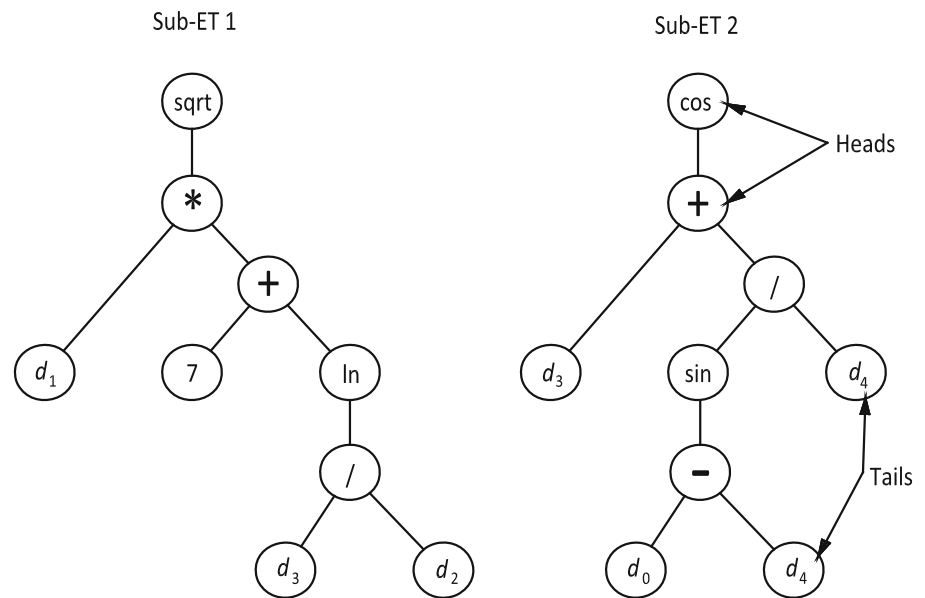
Figure 1 illustrates an example of the GEP solution with two sub-ETs. Each sub-ET contains nine chromosomes, and the head length is five. It expresses the mathematical equation as:

$$y = \sqrt{d_1 * \left(7 + \ln\left(\frac{d_3}{d_2}\right) \right)} + \cos\left(d_3 + \frac{\sin(d_0 - d_4)}{d_4}\right) \tag{1}$$

where y : dependent variable (output); $d_0, d_1, d_2, d_3,$ and d_4 : independent variables (inputs).

Recently, many studies have shown that the GEP model is more efficient than GP and comparable with the black-box ANN models [24]. Besides, Mousavi [49] developed the GEP model to predict the compressive strength of concrete with more accuracy than traditional models. Leong [44] applied the GEP technique to predict the UCS of fly ash-based geopolymers and evaluate the effect of each parameter on the UCS of geopolymers. Leong [50] then applied the ANN and GEP models to identify the contribution of input variables on the UCS of soil–fly ash geopolymer. In the research of Mohammadzadeh [43], the GEP-based model was developed to predict the coefficient of consolidation for the compression index of fine-grained soils based on the input variables such as the liquid limit, plastic limit, and initial void ratio. The GEP model was applied to estimate the UCS of geopolymer concrete based on ground granulated blast-furnace slag [28]. The GEP technique was also applied to generate predictive models to investigate the soil properties [45, 51, 52]. Abdi [53] developed a GEP-based model for predicting enhanced interaction coefficient based on large-scale direct shear tests conducted on soil–anchored geogrid samples. Johari [54] applied the GEP technique to investigate the collapsible soils treatment using nano-silica in the Sivand Dam region, Iran. Oulapour [55] generated a GEP model for predicting the cracking zones in earthfill dams.

Fig. 1 The GEP solution with two sub-ETs



Furthermore, the GEP model was used to solve many geotechnical engineering problems with high accuracy [26, 56–63]. Due to such advantages, the GEP technique was applied to generate a reliable predictive model for the UCS of clay stabilization.

3.2 GEP modeling procedure

Prior to GEP modeling, the data were divided into training, testing, and verifying subsets. The training data were used to train and select the optimal predicted programs. The selected models then measured their performance by using the testing dataset. Finally, the proposed GEP model was verified by an independent subset (unseen data). In this study, K-fold cross-validation (CV) was applied to split the data. Then, the model for UCS prediction, including the eleven independent variables, was developed based on approximately 1,183 data points. The datasets containing 789 (67.7%) and 197 (16.6%) data points were used for training and testing, respectively, while the remaining 197 (16.6%) unseen data points were used for verifying purposes.

GeneXpro Tools 5.0 software [64] was applied to simulate the GEP model for predicting the UCS of chemical-clay stabilization. It is a powerful and flexible modeling tool designed for regression. It also can process with a large number of variables with high accuracy and generalizability.

In GeneXpro Tools, the user needs to define the number of chromosomes, the number of genes (sub-ETs), the head length of the gene, the fitness function, and the genetic operators. The optimal parameters were determined according to the suggested values in Uysal [29], through

several trial runs and an error approach to ensure sufficient robustness and generalization of the model. In each trial, the value of one parameter was changed, while others were set constant to monitor the results. The fitness function was used to evaluate training and testing subsets at the end of each trial. When the errors of these subsets were small and as close as possible, the value of the parameter was chosen. The most important parameters which impact the complexity and accuracy of the GEP model are the number of genes (sub-ETs) and chromosomes. The head size of the gene is another essential parameter. It decides the number of branches of each sub-ET. The linking function was chosen based on trials to obtain the desired accuracy. Four basic functions (addition, subtraction, multiplication, and division) were tried to monitor the results. The optimal models were chosen when the training and testing processes provide approximately same and high accuracies.

Finally, the GEP-based model was developed with seven sub-ETs and a maximum of 200 chromosomes for each gene. They were linked by the addition function. The maximum head length of 10 was chosen, and the root mean square error (RMSE) was set as the fitness function. Fourteen different mathematical operators, including addition (+), subtraction (−), multiplication (*), division (/), square root ($\sqrt{\quad}$), natural logarithm (ln), power (x^2), inverse ($1/x$), exponential (exp), addition with three inputs ($x_1 + x_2 + x_3$), subtraction with three inputs ($x_1 - x_2 - x_3$), multiplication with three inputs ($x_1 * x_2 * x_3$), and cube root ($\sqrt[3]{\quad}$), were used for GEP modeling. The maximum number of generations was 30,000. In the model, d_0 , d_1 , d_2 , d_3 , d_4 , d_5 , d_6 , d_7 , d_8 , d_9 , and d_{10} represent the plastic index, the percentage of clay, the percentage of silt, the percentage of sand, the total water content, the curing time,

the percentage of lime, the percentage of cement, the percentage of fly ash Type F, the percentage of fly ash Type C, and the percentage of slag, respectively.

3.3 GEP-based model result

Figure 2 presents the tree-like structure of the selected optimal GEP model with seven sub-ETs linked by the addition linking function. The mathematical formula evolved from the GEP model is presented as Eq. (2). It is obvious that the GEP technique yields practical and straightforward predictive formula.

$$\begin{aligned}
 \text{UCS} = & d_0 \cdot (d_{10} - 2) + d_3 + 3 \cdot (d_5 - d_4 - d_{10}) \\
 & - 2 \cdot (d_1 + d_2 - d_7) + d_6 \cdot (d_2 + d_6 + 2) \\
 & + d_5 \cdot d_9 + 2c_{59} - \frac{e^{c_{51} - c_{55} - c_{56}}}{c_{59}} \\
 & - e^{c_{51}} + d_7 \cdot c_{21} \cdot \ln(d_5) + d_8 \cdot (d_4 \cdot d_3)^{1/3} + c_{50} + c_{64} \\
 & + c_{40} \cdot \ln \left[\left(\frac{d_5^{d_{10}}}{d_4} + d_4^2 + e^{d_9} \right)^{-1/2} \right] \\
 & + d_7 \cdot \left(\frac{d_3}{c_{72} + d_4 + c_{78}} \right) \cdot \left(2 \cdot d_2 + \frac{d_4}{c_{78}} \right) + c_{10}
 \end{aligned} \tag{2}$$

where UCS: unconfined compressive strength (kPa); $c_{10} = -9.204$; $c_{21} = 9.910$; $c_{40} = -15.170$; $c_{59} = -9.800$; $c_{51} = 5.616$; $c_{50} = -9.6502$; $c_{55} = 8.759$; $c_{56} = -5.498$; $c_{64} = -9.812$; $c_{72} = -7.635$; $c_{78} = -3.177$; d_0 : PI (%); d_1 : Clay (%); d_2 : Silt (%); d_3 : Sand (%); d_4 : Water (%); d_5 : Age (days); d_6 : Lime (%); d_7 : Cement (%); d_8 : FA F (%); d_9 : FA C (%); d_{10} : Slag (%).

The accuracy of the GEP model was evaluated through the coefficient of correlation (R), root mean square error (RMSE), and mean absolute error (MAE), which are measured as Eqs. (3)–(5). Figure 3 illustrates the performance at the training, testing, and verifying phases and entire datasets. Table 4 shows that the correlation coefficients of all phases are close and fluctuated around 0.95. The root mean square errors of the proposed model are less than 240 kPa, and the mean absolute errors are from 150 to 170 kPa. Besides, it needs to be mentioned that the selected optimal GEP model was developed based on a thousand data points. Even though the data were chosen carefully with the similar making and testing standard, uncertain factors that could affect the final model are unavoidable. Therefore, it could be said that the proposed GEP model is accurate and reliable.

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (P_i - \bar{P})^2}} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \tag{4}$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \tag{5}$$

where N : the number of data points presented to the model; O_i and P_i : the observed and predicted outputs, respectively; and O and P : the mean of observed and predicted outputs, respectively [65].

3.4 Performance analysis

3.4.1 External criteria

Golbraikh [66] provided some criteria as the external validation models for checking the testing dataset. It is suggested that one of the slopes of regression lines (k or k') through the origin should be approximate 1.0. Besides, the m and n indexes (performance indexes) should be lower than 0.1. Moreover, the squared correlation coefficients through the origin between predicted and experimental values (R_o^2 or $R_o'^2$) are recommended to be close to 1.0 [47, 67]. Table 5 presents the results of these criteria. It is clear that the proposed model satisfies the required conditions and shows great performance with a high accurate predictive capability.

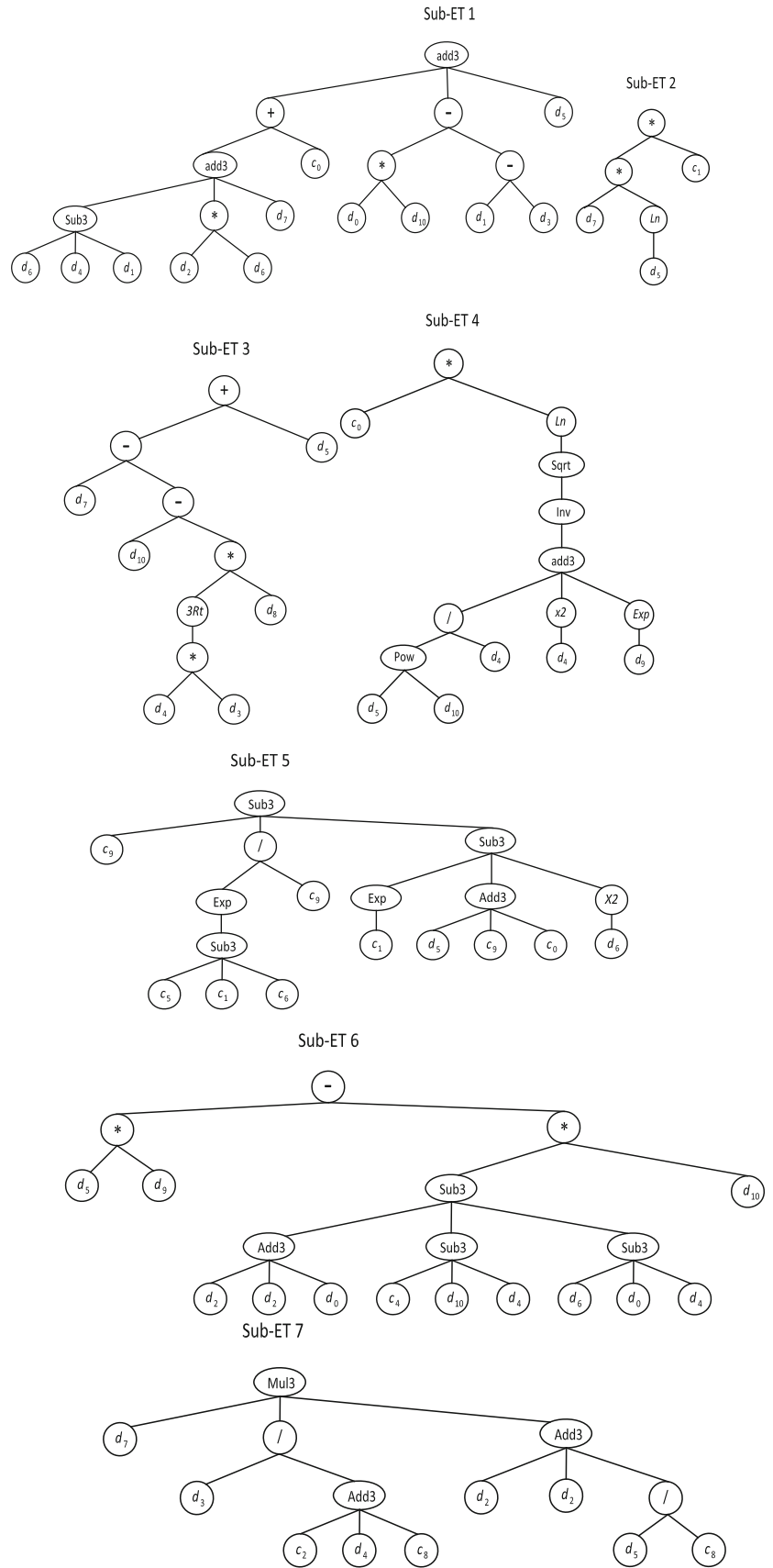
3.4.2 Comparative study

Table 6 illustrates the comparison among several numerical predictive models for UCS of soil stabilization and the proposed GEP-based model in this research. It is obvious that the predictive model in this study was generated based on a large volume of data (1183 data points), which were collected from plenty of studies in different nations. In contrast, other published models were developed based on small group of testing data from specific area. Therefore, that advantage brings distinguished results in comparison with other models for predicting the UCS of stabilized soil.

Furthermore, eleven input variables in the suggested model are the main parameters that affect the UCS result. Especially, the model not only considers the soil properties but also investigates the effect of four chemical binders (cement, lime, fly ash, and slag). As a result, the novel model is unique and different from other previous models.

In addition, the proposed GEP-based model could be applied for a wide range of input parameters. For example, the maximum cement content in the model is 100%, while the normal amount of cement used is around 10–30%. Hence, the applicability of this model is wider than others.

Fig. 2 Tree expression of the proposed GEP model



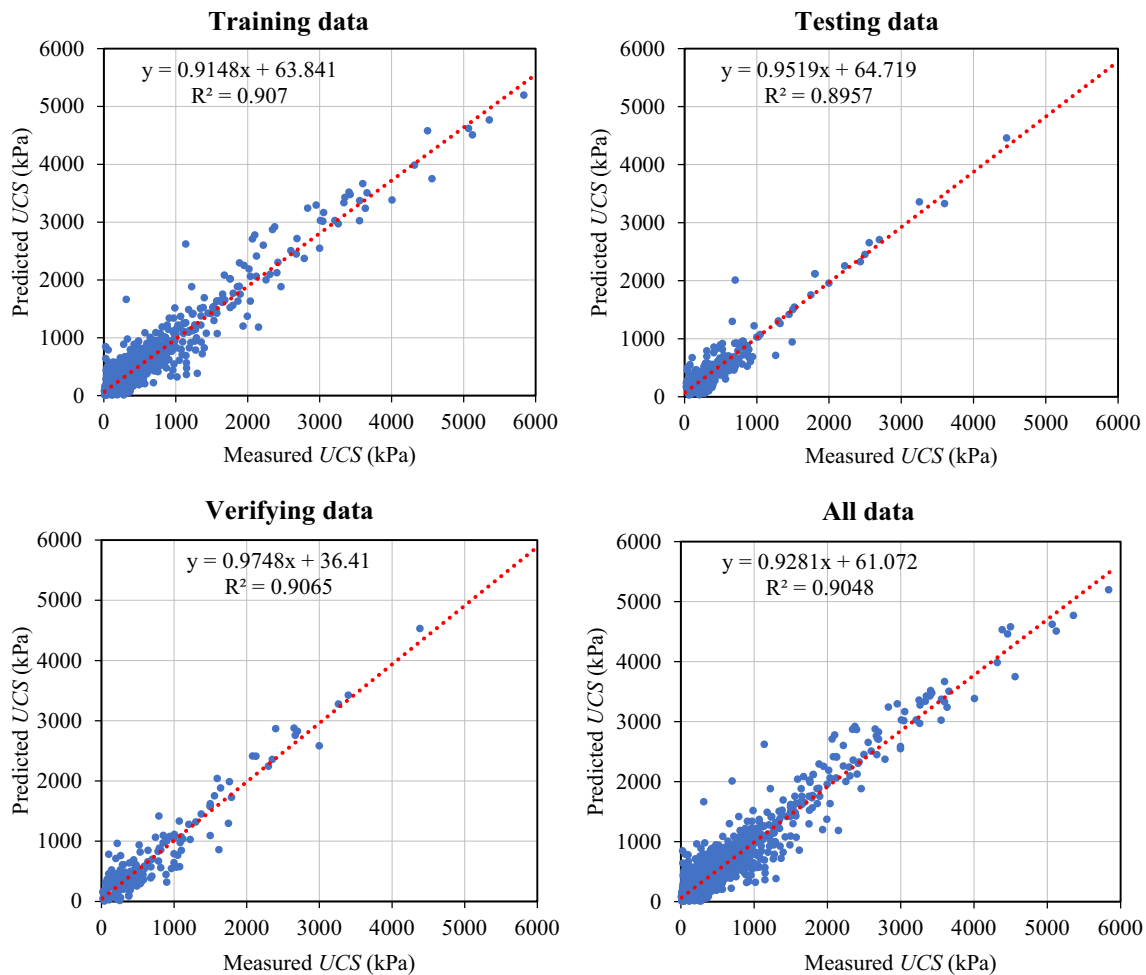


Fig. 3 The performance of the proposed GEP model

Table 4 Performance of the proposed GEP model

| Parameter | Training | Testing | Verifying | All data |
|------------|----------|---------|-----------|----------|
| <i>R</i> | 0.952 | 0.946 | 0.952 | 0.951 |
| RMSE (kPa) | 234.7 | 212.3 | 219.9 | 228.0 |
| MAE (kPa) | 170.2 | 151.5 | 167.9 | 166.3 |

R: correlation coefficient, RMSE: root mean square error, MAE: mean absolute error

In terms of the model performance, the proposed model achieves a high correlation coefficient ($R = 0.951$) and low errors. According to a logical hypothesis [68], if the model provides a high correlation coefficient ($R > 0.8$) and low error values (e.g., MAE and RMSE), the prediction relationship among input and output variables is accurate and reliable. Thus, it demonstrates an outstanding performance with strong predictive capability of the proposed GEP model.

It is the fact that the ANN-based models could slightly outperform the GEP-based models. However, that technique is considered as a “black-box” programming; hence,

the applicability of the ANN-based formulas is limited in practice.

In conclusion, the evolutionary predictive model in this study could be confidently applied in determining the UCS of chemical-stabilized clayey soil considering the effects of the soil characteristics and the types and the contents of chemical binders. Therefore, the GEP-based formula is a reliable option for a designer and researcher in estimating the UCS value of stabilized clayey soil.

4 Parametric study

4.1 Effect of each variable on the UCS of stabilized clay

The selected optimal GEP-based model (Eq. 2) was used to examine the effect of each input variable on the UCS of stabilized clay. The examined variable was assumed to be varied within the range of its input, while the average values were kept constant for other parameters. The effects

Table 5 Statistical parameters for the external validation of the GEP-based model

| Statistical parameter | Criteria | GEP-based model |
|---|-------------------|-----------------|
| $k = \frac{\sum_{i=1}^n (O_i * P_i)}{\sum_{i=1}^n O_i^2}$ | $0.85 < k < 1.15$ | 0.969 |
| $k' = \frac{\sum_{i=1}^n (O_i * P_i)}{\sum_{i=1}^n P_i^2}$ | $0.85 < k < 1.15$ | 0.974 |
| $R_0^2 = 1 - \frac{\sum_{i=1}^n (P_i - O_i^0)^2}{\sum_{i=1}^n (P_i - \bar{P}_i)^2}$ where $O_i^0 = k' * O_i$ | Close to 1 | 0.900 |
| $R_0'^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i^0)^2}{\sum_{i=1}^n (O_i - \bar{O}_i)^2}$ where $P_i^0 = k * P_i$ | Close to 1 | 0.905 |
| $m = (R^2 - R_0^2)/R^2$ | $m < 0.1$ | 0.0047 |
| $n = (R^2 - R_0'^2)/R^2$ | $n < 0.1$ | 0.0005 |

Table 6 Comparative study

| Study | Objective | Dataset | Variables | Method | R-value |
|--------------------|---|---------|-----------|-------------------------------------|-------------|
| Yao [1] | Strength of marine clay stabilized | 40 | 3 | Linear regression | 0.990 |
| Das [15] | Dry density and unconfined compressive strength of cement stabilized soil | 55 | 7 | ANN | 0.910 |
| Sunitsakul [17] | Unconfined compressive strength of cement stabilized bases | 520 | 3 | Nonlinear multi-variable regression | 0.867 |
| Abbey [18] | Pulverized fuel ash and ground granulated blast slag in cement deep soil mixing | 300 | 2 | Multiple regression analyses | 0.959 |
| Ghorbani [19] | UCS and CBR of micro silica-lime stabilized sulfate silty sand | 90 | 5 | ANN | 0.996 |
| Saadat [20] | Unconfined compressive strength of stabilized soil | 75 | 3 | Nonlinear regression | 0.888 |
| Tsuchida [22] | Compressive strength of cement-treated marine clays | 160 | 4 | Linear regression | 0.800–0.990 |
| Tinoco [23] | Uniaxial compressive strength of jet grouting columns | 175 | 11 | ANN | 0.970 |
| Shahmansouri [28] | Compressive strength of GPC based on ground granulated blast-furnace slag | 351 | 5 | GEP | 0.958 |
| Mohammadzadeh [43] | Compression index of fine-grained soils | 108 | 3 | GEP | 0.912 |
| Leong [44] | Compressive strength of fly ash-based geopolymers | 144 | 7 | GEP | 0.959 |
| Leong [50] | Optimum compressive strength soil-fly ash geopolymer | 64 | 7 | GEP | 0.959 |
| Johari [54] | Collapsible soils treatment using Nano-silica | 24 | 2 | GEP | 0.927 |
| Onyelowe [69] | Unconfined compressive strength of unsaturated lateritic soil treated with a hybridized binder material | 121 | 6 | GP | 0.972 |
| This study | Unconfined compressive strength of chemical stabilized clayey soil | 1183 | 11 | GEP | 0.951 |

ANN: Artificial neural network, GP: genetic programming, GEP: gene-expression programming

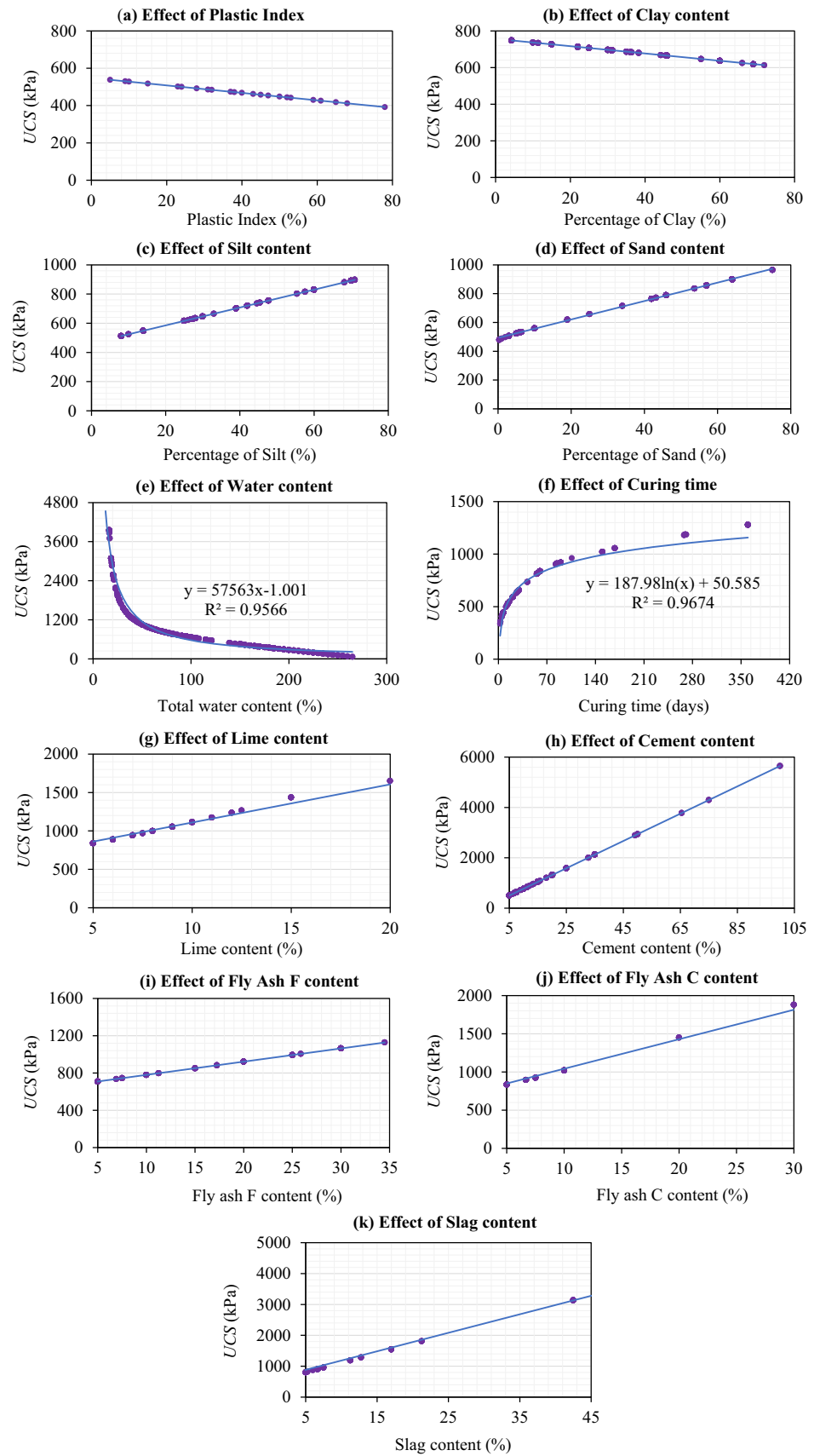
of input parameters on the UCS of stabilized soil based on the parametric study are illustrated in Fig. 4.

4.1.1 Effect of Atterberg limits

Figure 4a shows the effect of the Atterberg limits of the soil on the UCS of stabilized soil. It is obvious that the

UCS is reduced when the plastic index is high. The plastic index is an important parameter that correlates with soil behavior [5]. The stiffness and strength of the soil are decreased when the plastic index is increased [47, 70]. At the high level of the plastic index, the soil is more ductile [1]. As a result, the plastic index has a negative effect on the UCS of stabilized soil.

Fig. 4 Effects of input parameters on the UCS of stabilized soil



4.1.2 Effect of the particle size

The effect of the particle size of the soil on the UCS of stabilized soil is presented in Fig. 4b–d. The high percentage of clay may lead to a negative effect on the UCS of stabilized soil. In contrast, the UCS of stabilized soil increases linearly with the percentage of silt and sand. Increasing the percentage of sand could improve the UCS of stabilized soil. Szymkiewicz [71] concluded that the strength of stabilized soil reaches a higher value if the soil contains a well-graded grain size. Within the scope of this study, the effect of sand particle size and shape and the types of silt were not considered. It could be examined in another research.

4.1.3 Effect of the total water content

Figure 4e illustrates the effect of the total water content in the stabilized soil. Total water includes the natural water content in the soil and the added water used to mix with the binder in the slurry (the wet mixing method). The added water is zero for the dry mixing method. It should be noted that total water content has a negative correlation with the UCS of stabilized soil. The strength of the material markedly reduces when the amount of total water content is high. Besides, it requires more binder content to achieve the designed strength. The correlation relationship between the UCS and the total water content is expressed as:

$$UCS = A \cdot W_c^{-B} \quad (3)$$

where A and B : constant values, and W_c : the total water content.

The finding is well in agreement with the reports in the researches of Horpibulsk [2], Kitazume [10], Naveena [6], and Yao [1]. Their studies showed that water content plays a dominant role in the characterization of stabilized soils, especially on compressibility [2]. It reflects the microfabric of the material. The high amount of water content increases the distance between the soil particles and creates a more porous structure in the stabilized soil. Horpibulsuk [72] showed that the pore size increases remarkably when the water content is increased. As a result, it leads to the growth of capillary pore size and a low level of crystalline structure [6]. Therefore, the strength of stabilized soil is reduced with the increase in total water content [1].

4.1.4 Effect of the curing period

The effect of the curing period on the UCS of stabilized soil is demonstrated in Fig. 4f. It indicates the UCS of stabilized soil could achieve higher values by increasing the curing period. The correlation relationship between the UCS and the curing time is:

$$UCS = a \cdot \ln t + b \quad (4)$$

where a and b : constant values, and t : the curing time (day).

That result is similar to the reports of Horpibulsk [2], Kitazume [10], and Yao [1]. Based on the microstructure analysis, Horpibulsuk [73] found that in a short curing period, the volume of large pores in the soil–cement material increases, and the volume of small pores decreases, leading to low strength. In contrast, in the long-term period, the volume of large pores decreases significantly, while the volume of small pores increases. In addition, during that period, cementitious products are growing. Thus, the strength of stabilized soil increases over time [73].

However, the strength gain rate depends on the types of binder. For cement stabilized soil, the strength increases significantly in the first 1 to 3 months; then, the strength gain is low [74]. For lime and slag, the strength development is continuously for a long period if the water in the soil and the binder are sufficiently for pozzolanic reactions. When lime or slag is mixed with soil, the calcium reacts with the silicates and aluminates to create calcium silicate hydrates and calcium aluminate hydrates [6]. These reactions could occur for a long period if an adequate binder is provided [6].

4.1.5 Effect of the binder type and binder content

The effects of the binder types (lime, cement, fly ash, and slag) and the binder content are illustrated in Fig. 4g–k. Generally, the amount of binder is the essential parameter that has a positive effect on the UCS of stabilized soil. Depending on the types of binders, the strength gain rate is different.

When the substantial binder is provided, the hydration compounds are formed fully and create a hardened skeleton matrix [6]. The hydration products enclose the soil particles. In the long curing period, the alkaline conditions in the soil–binder are increased; the silicate and alumina from clay minerals and amorphous materials on the surface of clay particles are dissolved. They react with calcium ions to form insoluble compounds [6]. Besides, more amount of cementitious products are generated, and they enhance the inter-cluster bonding strength and fill the pore space [73]. Thus, the strength of stabilized soil increases when the binder content is high.

These findings are expected and well agreement with previous studies. Thus, it demonstrates that the proposed GEP-based model is reliable and applicable.

4.2 Effect of the combination of variables

4.2.1 Effect of the particle size of the soil

In order to evaluate the effect of the soil particle size on the UCS of stabilized soil, the percentages of sand, silt, and clay in the soil were changed in different ways. In this section, the amount of cement used was 20%, and other binders were 0%. Firstly, the proportion of sand was kept constant at 10%, 30%, and 50%; the percentage of silt was increased, and the percentage of clay was decreased at the same time (it was assumed that the total amount of sand, silt, and clay in the soil is 100%). Figure 5 shows that increasing the amount of silt (and decreasing the clay content) leads to an increase in the UCS. The level of strength gain is depended on the percentage of sand (10%, 30%, and 50%). If the amount of sand in the soil is high, the UCS significantly improves with the increase in the silt content.

Figure 6 illustrates the effect of sand content in the soil on the UCS of stabilized soil. The percentage of silt is kept constant at different levels (10%, 30%, and 50%), the amount of sand is increased, and the clay content is decreased. It is obvious that with a low percentage of silt, the UCS of stabilized soil remains constant even the amount of sand is changed. However, if the silt content is high, the UCS of stabilized soil increases significantly by increasing the amount of sand.

The effect of the percentage of silt and sand in the soil on the UCS of stabilized soil is demonstrated in Fig. 7. In this case, the UCS is observed by remaining the amount of clay, increasing the percentage of silt, and decreasing the amount of sand at the same time. The UCS of stabilized

soil shows two different trends. In the first part, the UCS grows up following the increase in the silt content. However, if the silt content is still increased (the sand content is decreased), the UCS is then decreased. The peak of the UCS curve could be the suitable ratio of particle size distribution to achieve the reasonable UCS value of the stabilized soil. For example, when the soil contains 10% of clay, the UCS of stabilized soil achieves the highest value if the percentages of silt and sand contents are 50% and 40%, respectively. It is noted that some parameters, such as the type of silt, the size and shape of sand, the water content, and the type and the amount of binder, could cause different strength gain trends. Hence, these findings should be verified by laboratory experiment results with similar input conditions.

The effect of particle size of soil on the UCS of stabilized soil could be explained through the shape and size of the soil particle and microstructure of soil–cement material. The size of soil particles affects the shear strength and characteristics of the soil. Sand and silt particles have irregular in size and shape. The silt particle is smaller than sand. According to the classification in ASTM D3282 [75], the diameter of the sand particle is from 0.075 to 2 mm, the silt particle is from 0.002 to 0.075 mm, and the clay diameter is smaller than 0.002 mm.

The typical shape of clay particles could be flaky, needle shape, or elongated particles [76]. Hence, the clay particles and clay clusters in the soil easily slide over each other when sheared, leading to low strength and stiffness [73]. Thus, a high amount of clay content could lead to low strength.

The soil microfabric includes domain (a group of clay particles), cluster (a group of domains), ped (a group of

Fig. 5 Effects of the percentage of silt and clay on the UCS of stabilized soil

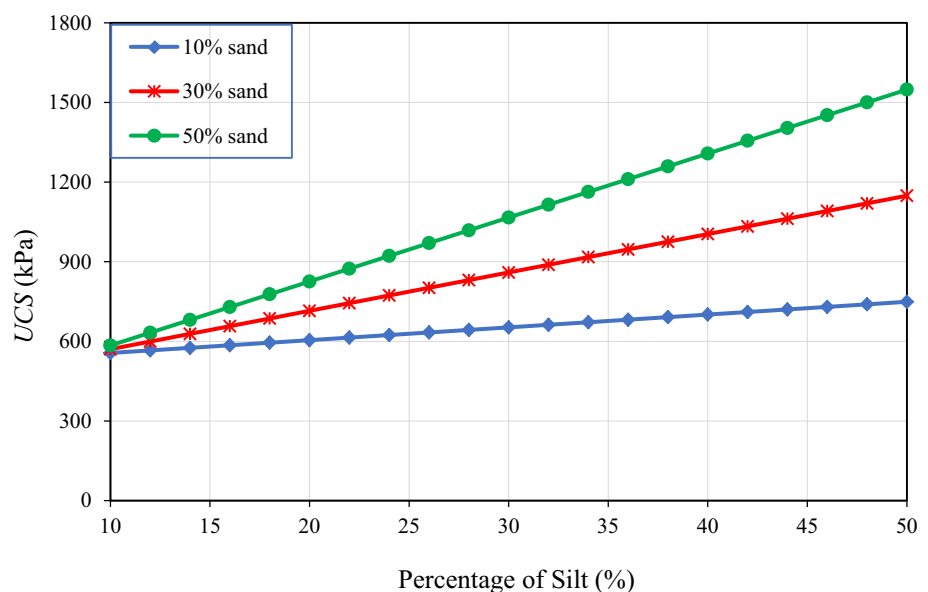


Fig. 6 Effects of the percentage of sand and clay on the UCS of stabilized soil

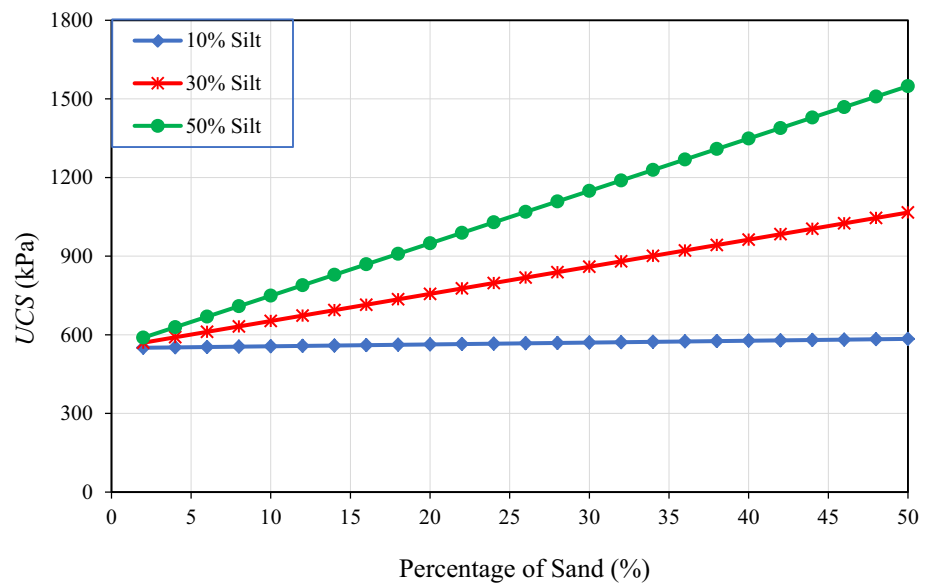
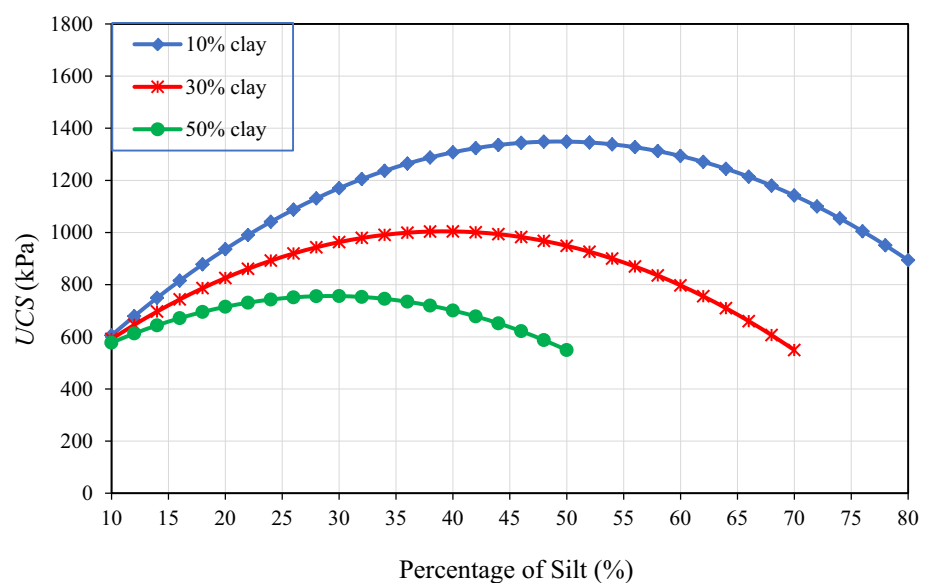


Fig. 7 Effects of the percentage of silt and sand content on the UCS of stabilized soil



clusters), silt or sand grain, micropore, and macropore [77, 78]. If the soil contains sand grain only, the large soil–cement clusters are created, and large pore spaces are developed. It causes a loose structure in the material, which makes the strength of clay stabilization is low [78].

For the soil contains sand, silt, and clay particles, increasing the particle size and the sand content could increase the internal friction angle and shear strength of the soil [79]. Well-graded soil provides the highest shear strength [80]. In this case, the fine particles create the skeleton for the structure, and the clay–cement clusters fill the pore between silt and sand particles. The cementitious products will enclose the particles, connect them together, and fill the micropores to create a rigid structure [73].

4.2.2 Effect of cement in combination with other chemical binders

In the situation that the amount of cement was fixed at 20%, the performance of stabilized soil was examined by adding with other kinds of binder. Figure 8 shows that the strength gain rate is different and depends on the type of added binder (lime, fly ash Types C and F, and slag). Soil stabilization with cement (20%) and fly ash provides lower strength than other binders. Fly ash contributes little to the strength gain from pozzolanic reactions [74]. Furthermore, fly ash Type C shows better performance than fly ash Type F. According to ASTM C618 [81], the CaO content in fly ash Type C is 24%, much higher than the amount of CaO in

Fig. 8 Effects of the added binder types on the UCS of stabilized soil

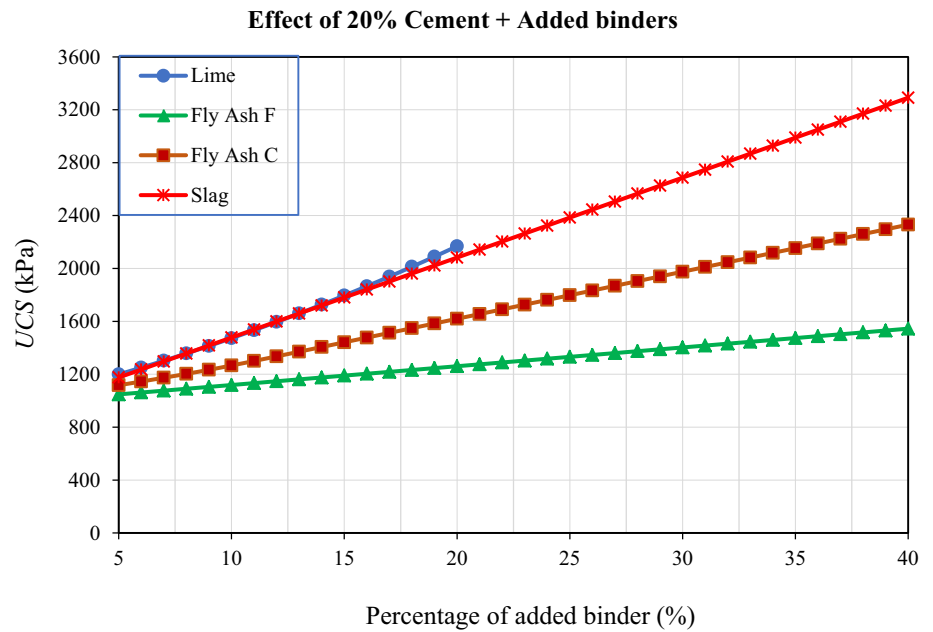
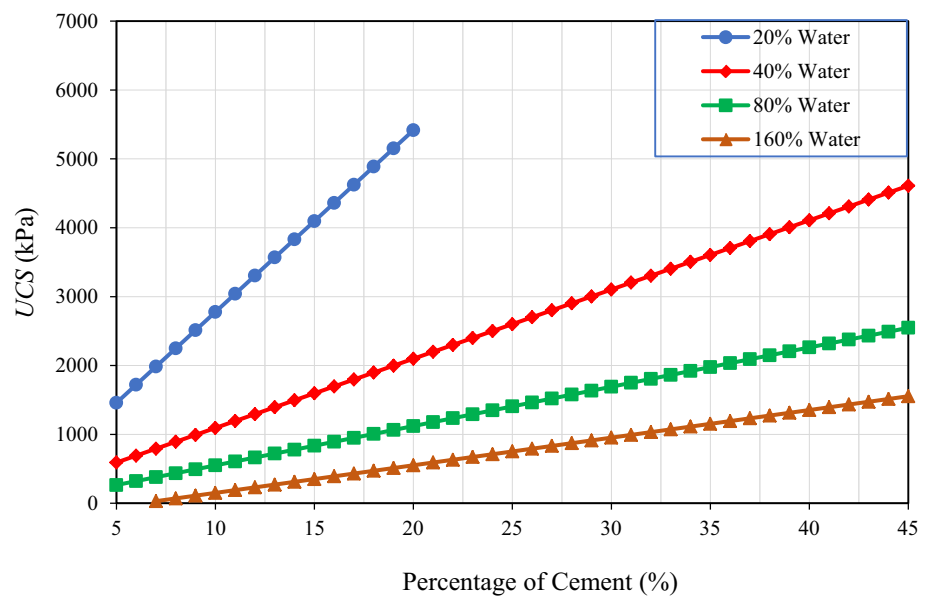


Fig. 9 Effects of the cement content and total water content on the UCS of stabilized soil



fly ash Type F (8%). Hence, more hydration productions are created by fly ash Type C than fly ash Type F.

Figure 8 demonstrates that stabilized soil with cement in combination with slag could achieve high strength. The blast-furnace slag, a type of steel slag, contains the main components of Portland cement, such as CaO, SiO₂, and Al₂O₃. Especially, the amount of SiO₂ and Al₂O₃ in the slag is much higher than in Portland cement. Therefore, it could create more hydration productions in the long-term period through hydration reactions [31, 74].

Lime could be used to combine with cement to enhance the strength of stabilized soil. By combining 20% cement with lime (in this study, the lime content is up to 20%), the UCS of stabilized soil is improved significantly. The high percentage of CaO content in lime (93%) provides an essential mineral for the pozzolanic reactions and generates long-term strength for the stabilized soil [6].

4.2.3 Effect of the total water content and cement content

Figure 9 shows the effect of total water content and the binder content on the UCS of stabilized soil. The total water content was kept constant at different levels, such as 20%, 40%, 80%, and 160%, while the cement content was increased linearly. As mentioned above, the cement content has a positive effect on the UCS. However, the strength gain rate depends on the amount of total water content. Besides, it requires more cement content to achieve the desired strength if the water content is high. For example, if the total water content is 20%, the strength of stabilized soil could be 2000 kPa by mixing with 7% cement content. On the other hand, if the total water content is 40% or 80%, the cement content needs to be increased to 19% and 35% to each the same strength value. Thus, this finding could help the consultancy decide the suitable binders for soil stabilization depending on the water content. Finally, these results support engineers in choosing reasonable methods to enhance the UCS of stabilized clay.

5 Conclusions

The study gathered over thousand comprehensive data points on the UCS of clay stabilization with common cementitious binders, such as lime, cement, fly ash, and slag. The GEP technique was applied to generate the predictive model. Eleven independent variables including the plastic index (PI), the percentage of clay (Clay), the percentage of silt (Silt), the percentage of sand (Sand), the total water content (Total water), the curing period (Age), the lime content (Lime), the cement content (Cement), the fly ash Type F content (FA F), the fly ash Type C content (FA C), and the slag content (Slag) were considered in the model. The results show that the proposed predictive model performs well with a high correlation coefficient ($R = 0.951$) and low errors (e.g., RMSE and MAE). Furthermore, the selected optimal model satisfied well with all external criteria. The comparative study demonstrates that the GEP-based model in this study was generated based on a large volume of data (1183 data points), while other studies just used a small dataset. As a result, the selected optimal GEP model could be confidently applied for different clay conditions in mixing with common chemical binders. The research results show distinguished accuracy and reliability in comparison with previous models. Besides, the GEP model generates transparent and practical mathematical equations which could be ready to use in practice. As such advantages, the proposed GEP-based model could help engineers in estimating the UCS of clay stabilization with different binders. However, the model was developed based on the laboratory test results; thus, it

needs to be modified when applying in the in situ test as there are plenty of uncertain variables on the site.

The parametric study was conducted to examine the effects of the essential parameters on the UCS of stabilized soil. Most findings from the parametric analysis are expected and well in agreement with other experimental results. It confirms that the proposed model is reliable and accurate. Thus, it helps consultants to understand the ways to optimize the ultimate strength and choose suitable binders for clay stabilization. The parametric study results indicate that:

- The plastic index, the percentage of clay, and the total water content have a negative effect on the UCS of stabilized soil. The correlation relationship between the UCS and total water content is expressed as $UCS = A \cdot W_c^{-B}$.
- The percentage of silt and sand in the soil, the binder type, the binder content, and the curing time show a positive effect on the UCS of stabilized soil. Increasing these parameters could enhance the strength of stabilized clay. The correlation relationship between the UCS and the curing time is $UCS = a \cdot \ln t + b$.
- Increasing the sand and silt content in the soil could improve the strength of stabilized clay. Well-graded soil provides the highest strength when the clay is mixed with binders. However, the effect of the type of silt, the amount of silt content, and the size and shape of sand particles should be deeply studied by laboratory tests.
- The stabilized clay with cement in combination with slag could achieve excessive strength than other adding binders. Lime and fly ash could also be used to enhance the UCS of stabilized soil. Moreover, fly ash Type C shows better performance than fly ash Type F in stabilizing clay.
- The stabilized soil could reach higher strength by reducing the natural water content in the soil and the water to binder ratio. Moreover, it requires more cement content to achieve the desired strength if the water content is high.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Data availability Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration

Conflict of interest The authors declare that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Yao K, Pan Y, Jia L, Yi JT, Hu J, Wu C (2020) Strength evaluation of marine clay stabilized by cementitious binder. *Mar Georesources Geotechnol* 38(6):730–743. <https://doi.org/10.1080/1064119X.2019.1615583>
2. Horpibulsk S, Rachan R, Suddeepong A, Chinkulkijniwat A (2011) Strength development in cement admixed Bangkok clay: laboratory and field investigations. *Soils Found* 51(2):239–251. <https://doi.org/10.3208/sandf.51.239>
3. Cong M, Longzhu C, Bing C (2014) Analysis of strength development in soft clay stabilized with cement-based stabilizer. *Construct Build Mater* 71:354–362. <https://doi.org/10.1016/j.conbuildmat.2014.08.087>
4. Brand EW, Brenner RP (1981) *Soft clay engineering*. Elsevier, Amsterdam
5. Narendra B, Sivapullaiah P, Suresh S, Omkar S (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>
6. Naveena P, Mamatha K, Dinesh S (2013) Prediction of strength development in stabilized sandy clay at high water contents. *Int J Geol* 7(1):9–23
7. Eskisar T (2015) Influence of cement treatment on unconfined compressive strength and compressibility of lean clay with medium plasticity. *Arab J Sci Eng* 40(3):763–772. <https://doi.org/10.1007/s13369-015-1579-z>
8. Donald B (2000) An introduction to the deep soil mixing methods as used in geotechnical applications. FHWA-RD-99-138
9. Holm G (2003) State of practice in dry deep mixing methods. In: *Proceedings of the 3th international conference on grout ground treat*. New Orleans, Louisiana, United States. [https://doi.org/10.1061/40663\(2003\)5](https://doi.org/10.1061/40663(2003)5)
10. Kitazume M, Terashi M (2013) *The deep mixing method*. Taylor and Francis, London
11. Westerberg B, Pousette K, Rogbeck Y, Andersson M, & Norrman T (2005) Laboratory tests of stabilised sulphide soil from Northern Sweden. In: *Proceedings of the international conference on DM2005*. Stockholm
12. Tokunaga S, Miura H, Otake T (2005) Laboratory tests on effect of cement content on permeability of cement treated soils. In: *Proceedings of the international conference on DM 2005*. Stockholm
13. Kosmatka SH, Panarese WC, Association PC (2002) *Design and control of concrete mixtures*, Portland Cement Association
14. Rodriguez R, Uribe R (2002) Importance of using the natural pozzolans on concrete durability. *Cem Concr Res* 32(12):1851–1858. [https://doi.org/10.1016/S0008-8846\(01\)00714-1](https://doi.org/10.1016/S0008-8846(01)00714-1)
15. Das SK, Samui P, Sabat AK (2011) Application of artificial intelligence to maximum dry density and unconfined compressive strength of cement stabilized soil. *Geotech Geol Eng* 29(3):329–342. <https://doi.org/10.1007/s10706-010-9379-4>
16. Tinoco J, Alberto A, da Venda P, Correia AG, Lemos L (2019) A novel approach based on soft computing techniques for unconfined compression strength prediction of soil cement mixtures. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-019-04399-z>
17. Sunitsakul J, Sawatparnich A, Sawangsuriya A (2012) Prediction of unconfined compressive strength of soil–cement at 7 days. *Geotech Geol Eng* 30(1):263–268. <https://doi.org/10.1007/s10706-011-9460-7>
18. Abbey S, Ngambi S, Ganjian E (2017) Development of strength models for prediction of unconfined compressive strength of cement/by product material improved soils. *Geotech Tes J* 40(6):928–935. <https://doi.org/10.1520/GTJ20160138>
19. Ghorbani A, Hasanzadehshooiili H (2018) Prediction of UCS and CBR of microsilica-lime stabilized sulfate silty sand using ANN and EPR models; application to the deep soil mixing. *Soils Found* 58(1):34–49. <https://doi.org/10.1016/j.sandf.2017.11.002>
20. Saadat M, Bayat M (2019) Prediction of the unconfined compressive strength of stabilised soil by adaptive neuro fuzzy inference system (ANFIS) and non-linear regression (NLR). *Geomech Geoeng*. <https://doi.org/10.1080/17486025.2019.1699668>
21. Liu S, Zhang D, Liu Z, Deng Y (2008) Assessment of unconfined compressive strength of cement stabilized marine clay. *Mar Georesour Geotechnol* 26(1):19–35. <https://doi.org/10.1080/10641190801937916>
22. Tsuchida T, Tang YX (2015) Estimation of compressive strength of cement-treated marine clays with different initial water contents. *Soils Found* 55(2):359–374. <https://doi.org/10.1016/j.sandf.2015.02.011>
23. Tinoco J, Correia AG, Cortez P (2011) Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. *Construct Build Mater* 25(3):1257–1262. <https://doi.org/10.1016/j.conbuildmat.2010.09.027>
24. Gandomi AH, Roke DA, Sett K (2013) Genetic programming for moment capacity modeling of ferrocement members. *Eng Struct* 57:169–176. <https://doi.org/10.1016/j.engstruct.2013.09.022>
25. Koza JR (1992) *Genetic programming: on the programming of computers by means of natural selection*. MIT press, Massachusetts
26. Johari A, Nejad AH (2015) Prediction of soil-water characteristic curve using gene expression programming. *Iran J Sci Tech Trans Civ Eng* 39(C1):143. <https://doi.org/10.22099/IJSTC.2015.2763>
27. Yang X-S, Gandomi AH, Talatahari S, Alavi AH (2013) *Meta-heuristics in water, geotechnical and transport engineering*. Elsevier, Amsterdam
28. Shahmansouri AA, Bengar HA, Ghanbari S (2020) Compressive strength prediction of eco-efficient GGBS-based geopolymer concrete using GEP method. *J Build Eng*. <https://doi.org/10.1016/j.jobe.2020.101326>
29. Uysal F (2020) Prediction of collapse potential of soils using gene expression programming and parametric study. *Arab J Geosci* 13(19):1–13. <https://doi.org/10.1007/s12517-020-06050-x>
30. Bolton M (2014) *Soil improvement using optimised cementitious materials design* [Doctoral thesis]: Griffith University, Australia
31. Ge L, Wang C-C, Hung C-W, Liao W-C, Zhao H (2018) Assessment of strength development of slag cement stabilized kaolinite. *Construct Build Mater* 184:492–501. <https://doi.org/10.1016/j.conbuildmat.2018.06.236>

32. JGS (2000) Japanese Geotechnical Society Standard “Practice for Making and Curing Stabilized Soil Specimens Without Compaction” Tokyo.
33. Xiao H, Shen W, Lee FH (2017) Engineering properties of marine clay admixed with Portland cement and blended cement with siliceous fly ash. *J Mater Civ Eng* 29(10):04017177. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0002014](https://doi.org/10.1061/(ASCE)MT.1943-5533.0002014)
34. Asgari M, Dezfuli AB, Bayat M (2015) Experimental study on stabilization of a low plasticity clayey soil with cement/lime. *Arab J Geosci* 8(3):1439–1452. <https://doi.org/10.1007/s12517-013-1173-1>
35. Correia A, Venda Oliveira P, Lemos L (2013) Prediction of the unconfined compressive strength in soft soil chemically stabilized. In: Proceedings of the 18th international conference on soil mechanical geotechnic engineering, pp 2457–60
36. Oh E (2006) Geotechnical and ground improvement aspects of motorway embankments in soft clay, Southeast Queensland [Doctoral thesis]: Griffith University, Australia
37. Tastan EO, Edil TB, Benson CH, Aydilek AH (2011) Stabilization of organic soils with fly ash. *J Geotech Geoenviron Eng* 137(9):819–833. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000502](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000502)
38. Kwan PS, Bouazza A, Fletcher P, Ranjith P, Oh EY, Shuttlewood K, et al. (2005) Behaviour of cement treated Melbourne and Southeast Queensland soft clays in deep stabilization works. In: Proceedings of the international conference on DM2005. Stockholm. vol 1(1), pp 101–10
39. Consoli NC, Prietto PDM, da Silva Lopes Jr L, Winter D (2014) Control factors for the long term compressive strength of lime treated sandy clay soil. *Transp Geotech* 1(3):129–36. <https://doi.org/10.1016/j.trgeo.2014.07.005>
40. Kassim KA, Chern KK (2004) Lime stabilized Malaysian cohesive soils. *Malays J Civ Eng* 16(1):13–23
41. Ferreira C (2006) Gene expression programming: mathematical modeling by an artificial intelligence. Springer, Berlin
42. Ferreira C (2002) Gene expression programming in problem solving. *Soft computing and industry*. Springer, Berlin, pp 635–653
43. Mohammadzadeh S, Kazemi S-F, Mosavi A, Nasserlshariati E, Tah JH (2019) Prediction of compression index of fine-grained soils using a gene expression programming model. *Infrastructures* 4(2):26. <https://doi.org/10.3390/infrastructures4020026>
44. Leong H, Ong D, Sanjayan J, Nazari A (2015) A genetic programming predictive model for parametric study of factors affecting strength of geopolymers. *RSC Adv* 5(104):85630–85639. <https://doi.org/10.1039/C5RA16286F>
45. Tenpe AR, Patel A (2020) Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Mater Pavement Des* 21(5):1183–1200. <https://doi.org/10.1080/14680629.2018.1544924>
46. Sadrossadat E, Ghorbani B, Hamooni M, Sheikhanloo MHM (2018) Numerical formulation of confined compressive strength and strain of circular reinforced concrete columns using gene expression programming approach. *Struct Concr* 19(3):783–794. <https://doi.org/10.1002/suco.201700131>
47. Soleimani S, Rajaei S, Jiao P, Sabz A, Soheilinia S (2018) New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming. *Measurement* 113:99–107. <https://doi.org/10.1016/j.measurement.2017.08.043>
48. Tenpe AR, Patel A (2018) Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Mater Pav Des*. <https://doi.org/10.1080/14680629.2018.1544924>
49. Mousavi SM, Aminian P, Gandomi AH, Alavi AH, Bolandi H (2012) A new predictive model for compressive strength of HPC using gene expression programming. *Adv Eng Softw* 45(1):105–114. <https://doi.org/10.1016/j.advengsoft.2011.09.014>
50. Leong HY, Ong DEL, Sanjayan JG, Nazari A, Kueh SM (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 Civ Eng* 30(7):04018129. <https://ascelibrary.org/doi/10.1061>
51. Jalal FE, Xu Y, Iqbal M, Jamhiri B, Javed MF (2021) Predicting the compaction characteristics of expansive soils using two genetic programming-based algorithms. *Transp Geotech* 30:100608. <https://doi.org/10.1016/j.trgeo.2021.100608>
52. Jalal FE, Xu Y, Iqbal M, Javed MF, Jamhiri B (2021) Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP. *J Environ Manag* 289:112420. <https://doi.org/10.1016/j.jenvman.2021.112420>
53. Abdi MR, Nakhaei P, Gonbad MSS (2021) Prediction of enhanced soil-anchored geogrid interactions in direct shear mode using gene expression programming. *Geotech Geol Eng* 39(2):957–972. <https://doi.org/10.1007/s10706-020-01537-6>
54. Johari A, Golkarfard H, Davoudi F, Fazeli A (2021) Experimental investigation of collapsible soils treatment using nano-silica in the sivand dam region. *Iran Iran J Sci Tech Trans Civ Eng*. <https://doi.org/10.1007/s40996-021-00675-y>
55. Oulapour M, Adib A, Gholamzadeh S (2021) GEP prediction of the cracking zones in earthfill dams. *Arab J Geosci* 14(7):1–11. <https://doi.org/10.1007/s12517-021-06933-7>
56. Alavi AH, Aminian P, Gandomi AH, Esmaeili MA (2011) Genetic-based modeling of uplift capacity of suction caissons. *Exp Syst Appl* 38(10):12608–12618. <https://doi.org/10.1016/j.eswa.2011.04.049>
57. Alavi AH, Gandomi AH (2011) A robust data mining approach for formulation of geotechnical engineering systems. *Eng Comput*. <https://doi.org/10.1007/s12040-011-0119-9>
58. Alkroosh I, Nikraz H (2012) Predicting axial capacity of driven piles in cohesive soils using intelligent computing. *Eng Appl Artif Intell* 25(3):618–627. <https://doi.org/10.1016/j.engappai.2011.08.009>
59. Alkroosh I, Nikraz H (2014) Predicting pile dynamic capacity via application of an evolutionary algorithm. *Soils Found* 54(2):233–242. <https://doi.org/10.1016/j.sandf.2014.02.013>
60. Tarawneh B, Imam R (2014) Regression versus artificial neural networks: predicting pile setup from empirical data. *KSCE J Civ Eng* 18(4):1018–1027. <https://doi.org/10.1007/s12205-014-0072-7>
61. Tarawneh B (2017) Predicting standard penetration test N-value from cone penetration test data using artificial neural networks. *Geosci Fron* 8(1):199–204. <https://doi.org/10.1016/j.gsf.2016.02.003>
62. Ozbek A, Unsal M, Dikec A (2013) Estimating uniaxial compressive strength of rocks using genetic expression programming. *J Rock Mech Geotech Eng* 5(4):325–329. <https://doi.org/10.1016/j.jrmge.2013.05.006>
63. Reza Abdi M, Rashed H (2021) Shear strength enhancement prediction of sand-fiber mixtures using genetic expression programming. *J Mater Civ Eng* 33(11):04021323. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0003954](https://doi.org/10.1061/(ASCE)MT.1943-5533.0003954)
64. GEP/Soft (2014) GeneXproTools Version 5.0
65. Shahin MA (2013) Artificial intelligence in geotechnical engineering: applications, modeling aspects, and future directions. *Metaheuristics in water, geotechnical and transport engineering*. Elsevier, Amsterdam, pp 169–204
66. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Modell* 20(4):269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)

67. Alavi AH, Gandomi AH, Mollahasani A, Bazaz JB, Talatahari S (2013) Linear and tree-based genetic programming for solving geotechnical engineering problems. *Metaheuristics in water, geotechnical and transport engineering*, Elsevier, Amsterdam, pp 289–310
68. Smith GN (1986) *Probability and statistics in civil engineering*, Collins professional technical books
69. Onyelowe KC, Ebid AM, Onyia ME, Nwobia LI (2021) Predicting nanocomposite binder improved unsaturated soil UCS using genetic programming. *Nanotechnol Environ Eng* 6(2):1–12
70. Naeini SA, Naderinia B, Izadi E (2012) Unconfined compressive strength of clayey soils stabilized with waterborne polymer. *KSCE J Civ Eng* 16(6):943–949. <https://doi.org/10.1007/s12205-012-1388-9>
71. Szymkiewicz F, Guimond-Barrett A, Kouby AL, Reiffsteck P (2012) Influence of grain size distribution and cement content on the strength and aging of treated sandy soils. *Eur J Environ Civ Eng* 16(7):882–902. <https://doi.org/10.1080/19648189.2012.676362>
72. Horpibulsuk S, Miura N, Nagaraj T (2005) Clay–water/cement ratio identity for cement admixed soft clays. *J Geotech Geoenviron Eng* 131(2):187–192. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2005\)131:2\(187\)](https://doi.org/10.1061/(ASCE)1090-0241(2005)131:2(187))
73. Horpibulsuk S, Rachan R, Chinkulkijniwat A, Raksachon Y, Suddepong A (2010) Analysis of strength development in cement-stabilized silty clay from microstructural considerations. *Construct Build Mater* 24(10):2011–2021. <https://doi.org/10.1016/j.conbuildmat.2010.03.011>
74. Åhnberg H, Johansson S-E, Pihl H, Carlsson T (2003) Stabilising effects of different binders in some Swedish soils. *Proc Inst Civ Eng: Ground Improv* 7(1):9–23. <https://doi.org/10.1680/grim.2003.7.1.9>
75. ASTM (2015) *Standard practice for classification of soils and soil-aggregate mixtures for highway construction purposes (ASTM D3282-15)*. West Conshohocken, PA
76. Ural N (2018) *The importance of clay in geotechnical engineering*, IntechOpen
77. Young R (2012) *Soil properties and behaviour*. Elsevier, Amsterdam
78. Ekinçi A (2019) Effect of preparation methods on strength and microstructural properties of cemented marine clay. *Construct Build Mater* 227:116690. <https://doi.org/10.1016/j.conbuildmat.2019.116690>
79. Li Y (2013) Effects of particle shape and size distribution on the shear strength behavior of composite soils. *Bull Eng Geol Environ* 72(3–4):371–381. <https://doi.org/10.1007/s10064-013-0482-7>
80. Islam MN, Siddika A, Hossain MB, Rahman A, Asad MA (2011) Effect of particle size on the shear strength behavior of sands. *Aus Geomech J* 46(3):75–85
81. ASTM (2019) *Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in Concrete (ASTM C618-19)*. West Conshohocken, PA: www.astm.org

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.