Cleaner Engineering and Technology 4 (2021) 100152

Contents lists available at ScienceDirect

Cleaner Engineering and Technology

journal homepage: www.sciencedirect.com/journal/cleaner-engineering-and-technology

Intelligent prediction of coefficients of curvature and uniformity of hybrid cement modified unsaturated soil with NQF inclusion

Kennedy C. Onyelowe [a,b,*], Jamshid Shakeri [c]

[a] Department of Civil Engineering, Michael Okpara University of Agriculture, Umudike, Nigeria
[b] Department of Civil and Mechanical Engineering, Kampala International University, Kampala, Uganda
[c] Department of Mining Engineering, Hamedan University of Technology, Hamadan, Iran

ARTICLE INFO

Keywords:
Intelligent prediction
Coefficient of curvature (C_C)
Coefficient of uniformityC_U
Hybrid cement (HC)
Artificial neural network (ANN)
Gene expression programming (GEP)
Multiple linear regression
Sensitivity analysis
Nanostructured quarry fines (NQF)

ABSTRACT

The cost and sophisticated equipment required to conduct earthwork laboratory experiments have been of concern to the design and performance monitoring of infrastructures in recent times. Lateritic soils especially those under unsaturated conditions are erratic and deserve close attention in terms of laboratory studies. In order to overcome the rigors and time consumed during experimental procedures, soft computing has been used to predict soil parameters for the purpose of design and construction. In this work, the ANN, GEP and LMR were employed to predict the coefficients of curvature and uniformity of lateritic soil treated with multiple binders locally generated, which were hybrid cement (HC) and nanostructured quarry fines (NQF). The effect of the varying dosages of HC and NQF added to the soil were studied and the behavior of clay activity, clay content, frictional angle, coefficients of curvature and uniformity were measure. 121 datasets were generated from the experimental exercise for the selected parameter both for the predictors and for the targets. These datasets were deployed in the ratio of 70 is to 30% for training and testing of the models predictions respectively. The performances of the models were evaluated using error analysis (VAF, RMSE, MAE) and accuracy (R^2) indices and it was observed that the ANN outclassed both GEP and LMR due to its speed and robustness in adopting back-propagation and feed-forward algorithms. Furthermore, the sensitivity analysis showed that F, C, H (HC), NQF and Ac in that order of most influential to least influential influenced the behavior of the Cc model with H (HC) and NQF showing equal effect on the Cc. Also, H (HC), NQF, F, C and Ac in that order of influence from most to least affected the behavior of the Cu predicted model also with HC and NQF having equal effect on the Cu. Generally, the learning techniques showed good performance in predicting the outputs hence are good techniques to be utilized in design and performance evaluation.

1. Introduction

1.1. Background

The changes in soil saturation state between unsaturated to saturated soil conditions and between saturated and unsaturated soil conditions, as the seasons cycle round the year demands critical attention, observation and design were studied by Rahardjo et al. (2019). There have been analytical approaches taken to understudy this soil phases which were taken by Lloret-Cabot et al. (2018) who worked on saturated to unsaturated conditions of soils and vice versa and found the changes that occur during this hydraulic transition of soils and also Rahardjo et al. (2019) who explored the role of unsaturated soil mechanics in geotechnical engineering. In hydraulically bound infrastructures like the pavement subgrade and other substructures subjected to the rise and fall of the water table, it obviously necessary to design analytical methods of handling the erratic behavior of soils under such environments and influence. In recent times, intelligent learning techniques have been employed to predict various quantities in geotechnics for the purpose of design, construction and performance study. Intelligent learning techniques have proven to be sustainable means of predicting earthwork parameters because of the robustness of their application and the reliability of the performance. The learning abilities of LMR, ANN and GEP have been utilized in this work to predict the coefficients of curvature and uniformity of lateritic soil treated with hybrid cement and nanostructured quarry fines. The coefficients of curvature and uniformity are very important factors used in the classification of soils or materials in

* Corresponding author. Department of Civil and Mechanical Engineering, Kampala International University, Kampala, Uganda.
E-mail addresses: kennedychibuzor@kiu.ac.ug, konyelowe@mouau.edu.ng, konyelowe@gmail.com (K.C. Onyelowe).

https://doi.org/10.1016/j.clet.2021.100152
Received 29 January 2021; Received in revised form 19 May 2021; Accepted 4 June 2021
Available online 10 June 2021

Nomenclatures

ANN	artificial neural network
LM	Levenberg– Marquardt
GEP	gene expression programming
LMR	linear multivariate regression
ET	expression tree
GP	genetic programming
GE	genetic expression
VAF	variant allele function
RMSE	root mean square error
MAE	mean average error
R^2	coefficient of determination
HC	hybrid cement
NQF	nanostructured quarry fines
QD	quarry dust
RHA	rice husk ash
XRF	x ray fluorescence
Cc	coefficient of curvature
Cu	coefficient of uniformity
Ac	clay activity
C	clay content
F	angle of friction

terms of their gradation or particle distribution. They are very important in the design of the behavior of soils because soils particle curvature and the distribution within the mass contribute to its behavior when utilized as a construction material. Curvature of the particles of the soil mass contribute also to the friction that takes place at the inter-particle level and conversely contribute to the shear behavior of the soil under loading. So, when soils under unsaturated condition fall below the requirements of these coefficients, treatment process is encouraged to improve these properties. In this present work, the combined effects of HC and NQF were utilized to improve the gradation properties of the soil in order to generate datasets needed for the intelligent prediction work. Previously, there has not been any intelligent prediction of Cc and Cu because earthwork engineers overlook their effects on the behavior of the soils when loaded. Therefore, the present work is focused on employing LMR, GEP and ANN to predict the Cc and Cu and validate and evaluate the performance of these learning techniques in prediction these outputs.

1.2. Linear multivariate regression model

The linear multivariate regression (LMR) is one of the most well-known approaches to generating a multiple equations between one or more independent or input parameters and one dependent or output parameter. This method is extensively used to approximate problems in the fields of rock mechanics, geology, civil engineering and mining sciences, which was studied by Rezaei et al. (2011), who developed a fuzzy model to predict flyrock in surface mining, Manoj and Monjezi, 2013, who proposed a new model based on gene expression programming to estimate air flow in a single rock joint and Armaghani et al. (2016) who by the combined multiple regression analysis and monte carlo simulation of quarry blasting, forecasted risk assessment and prediction of flyrock distance. The linear multivariate regression (LMR) is a generalization of simple linear regression to the case of more than one independent variable, and a special case of general linear models, restricted to one dependent variable as presented by Shakeri et al. (2020). In this paper, the dependent variables which assumed as coefficient of curvature (Cc) and coefficient of uniformity (Cu) may depend on "n" independent variables (x). Equation (2) expresses a LMR with n regression variables from the findings of Montgomery and Peck (1992), who introduced linear regression analysis in their book, Shakeri et al. (2020), with their multiple prediction of blast-induced ground vibration using gene expression programming (GEP), artificial neural networks (ANNs), and linear multivariate regression (LMR) and Shokri et al. (2020) with their findings on predicting silver price by applying a coupled multiple linear regression (MLR) and imperialist competitive algorithm:

$$C = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon \quad (1)$$

Where;

ε: error of the model; $j = 0, 1, \ldots, n$ and β_j are the regression coefficients.

In fact, the prediction process using this model resembles a super plane in an n-dimensional space of the regression variables x_j. On the other hand, one may consider prediction models of more complex structures (nonlinear) than those expressed by Equation (1). For instance, in the following model:

$$C = \beta_0 + \beta_1 x_1 + \beta_2 x_2^3 + \beta_3 e^{x_3} + \beta_4 x_1 x_2 + \varepsilon \quad (2)$$

In order to simplify the analysis of the above equation, which is nonlinear, one can simply substitute its variables with linear variables. Accordingly, taking$z_1 = x_1$, $z_2 = x_2^3$, $z_3 = e^{x_3}$, and $z_4 = x_1 x_2$, Equation (2) will take the following form:

$$C = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \beta_4 z_4 + \varepsilon \quad (3)$$

This is an LMR according to the positions of previous literature by Shakeri et al. (2020), with their multiple prediction of blast-induced ground vibration using gene expression programming (GEP), artificial neural networks (ANNs), and linear multivariate regression (LMR), Montgomery (1992), who introduced linear regression analysis in their book, Armaghani et al. (2016), who by the combined multiple regression analysis and monte carlo simulation of quarry blasting, forecasted risk assessment and prediction of fly rock distance and Shokri et al. (2020), with their findings on predicting silver price by applying a coupled multiple linear regression (MLR) and imperialist competitive algorithm.

1.3. Artificial neural network (ANN)

ANN originally was developed by McCulloch Warren and Pitts, 1943).ANNs are one of the best and most widely used methods for prediction, which is based on the human neural structure is one of the most advanced AI techniques presented by Nguyen and Bui (2019), in predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest, He and Xu (2007), presented in their book process neural networks and Shojaeian and Asadizadeh (2020), who worked on the prediction of surface tension of the binary mixtures containing ionic liquid using heuristic approaches adopting an input parameters investigation. So far, a lot of research has been done using ANNs to predict engineering problems among which were the ANN predictions adopted by Koopialipoor et al. (2018), who applied various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions, Rad et al. (2020), who predicted flyrock in mine blasting employing a new computational intelligence approach, Monjezi et al. (2013), who evaluated and predicted the blast-induced ground vibration at Shur River Dam, Iran, by artificial neural network, Shakeri et al. (2020), with their multiple prediction of blast-induced ground vibration using gene expression programming (GEP), artificial neural networks (ANNs), and linear multivariate regression (LMR), Shojaeian and Asadizadeh (2020), who worked on the

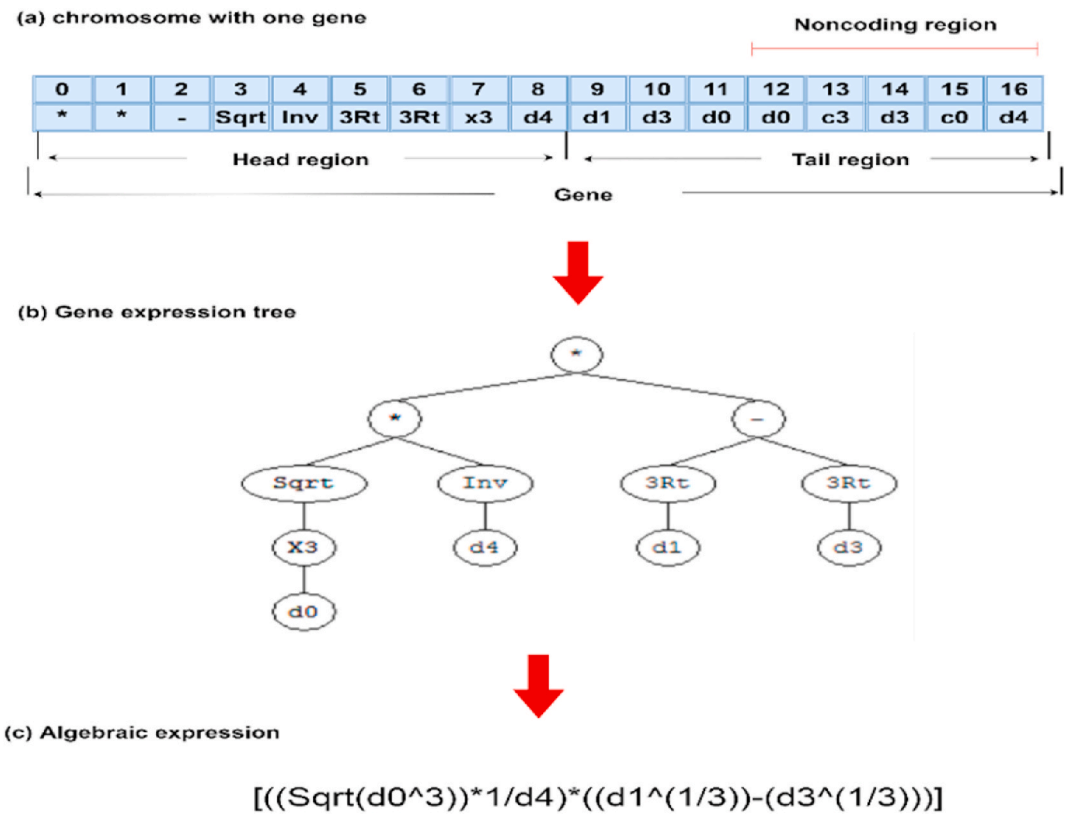

Fig. 1. Different expressions of chromosome, a) K-Expression, b) expression tree, c) mathematical equations.

prediction of surface tension of the binary mixtures containing ionic liquid using heuristic approaches adopting an input parameters investigation, and Lu et al. (2020), who made use of a novel machine learning approach for prediction of flyrock in mine blasting. These approaches showed remarkable performance. The structure of an ANN model includes input layer, hidden layer(s), and an output layer as presented in the works of Çelik (2020), who worked on the comparative analysis of artificial neural networks and multiplicative decomposition methods in tobacco production prediction in Turkey and Yegnanarayana (2009) in a book, artificial neural networks. The use of only one hidden layer reduces the complexity of the model; however, there is no theoretical limitation in this regard shown in the works of Armaghani et al. (2016) and Sumathi and AuthorAnonymous, 2010, who worked on the paradigms of computational intelligence adopting MATLAB theory and applications. Multi-layer perceptron or feed-forward neural network with backpropagation training algorithm is one of the most useful types of these networks. Each layer is comprised of several neurons among which the number of input and output layer neurons represents the number of input and output parameters of the model, respectively, and the number of neurons in the hidden layer/s is normally determined by trial and error. The neurons of hidden layer(s) are responsible to properly intervene between the network input and output. The most important features of layered networks are performing nontrivial calculations, learning from examples and generalization in the training stage, which was demonstrated in the works of Cigizoglu and Kis (2005), who worked on flow prediction by three back propagation techniques using k-fold partitioning of neural network training data, Rajabi and Vafaee (2020), in a work "prediction of blast-induced ground vibration using empirical models and artificial neural network (Bakhtiari Dam access tunnel, as a case study)", Monjezi et al. (2011(, who worked on "prediction and controlling of flyrock in blasting operation using artificial neural network" and Shojaeian and Asadizadeh (2020), who worked on the prediction of surface tension of the binary mixtures containing ionic liquid using heuristic approaches adopting an input parameters investigation. More details on the neural network can be found in the literature of the works of Monjezi et al. (2011), who worked on "prediction and controlling of flyrock in blasting operation using artificial neural network, Mendez et al. (2019), in a work "the application of artificial neural networks in metabolomics: A historical perspective", and Nguyen and Bui (2019), in predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest.

1.4. Gene expression programming (GEP)

GEP is currently regarded as the most advanced form of traditional genetic programming (GP) techniques. GP is an ML method capable of generating nonlinear prediction models in a highly automated fashion (Majidifard et al., 2020). This algorithm was proposed by Ferreira, 2001 based on the combination of genetic algorithm (GA) and genetic programming (GP). In fact, GEP is an extended version of GP and GA and can fix their shortcomings, such as the problems of applying genetic operators to trees applied by Faradonbeh et al. (2016), who worked on the prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction, Ferreira (2006), in work "gene expression programming: mathematical modeling by an artificial intelligence" and Baykasoğlu et al. (2008) in their works on the employment of GEP in model forecasting. The GP-based techniques have been deployed to help solve a myriad of complicated engineering problems by Shakeri et al. (2020), with their multiple prediction of blast-induced ground vibration using gene expression programming (GEP), artificial neural networks (ANNs), and linear multivariate regression (LMR), Ramesh et al. (2020), in their work "prediction of ground movements in shield-driven tunnels using gene expression programming" and Mahdiyar et al. (2020), a work "practical risk assessment of ground vibrations resulting from blasting, using gene expression programming and Monte Carlo simulation techniques". The obvious differences between these three algorithms are

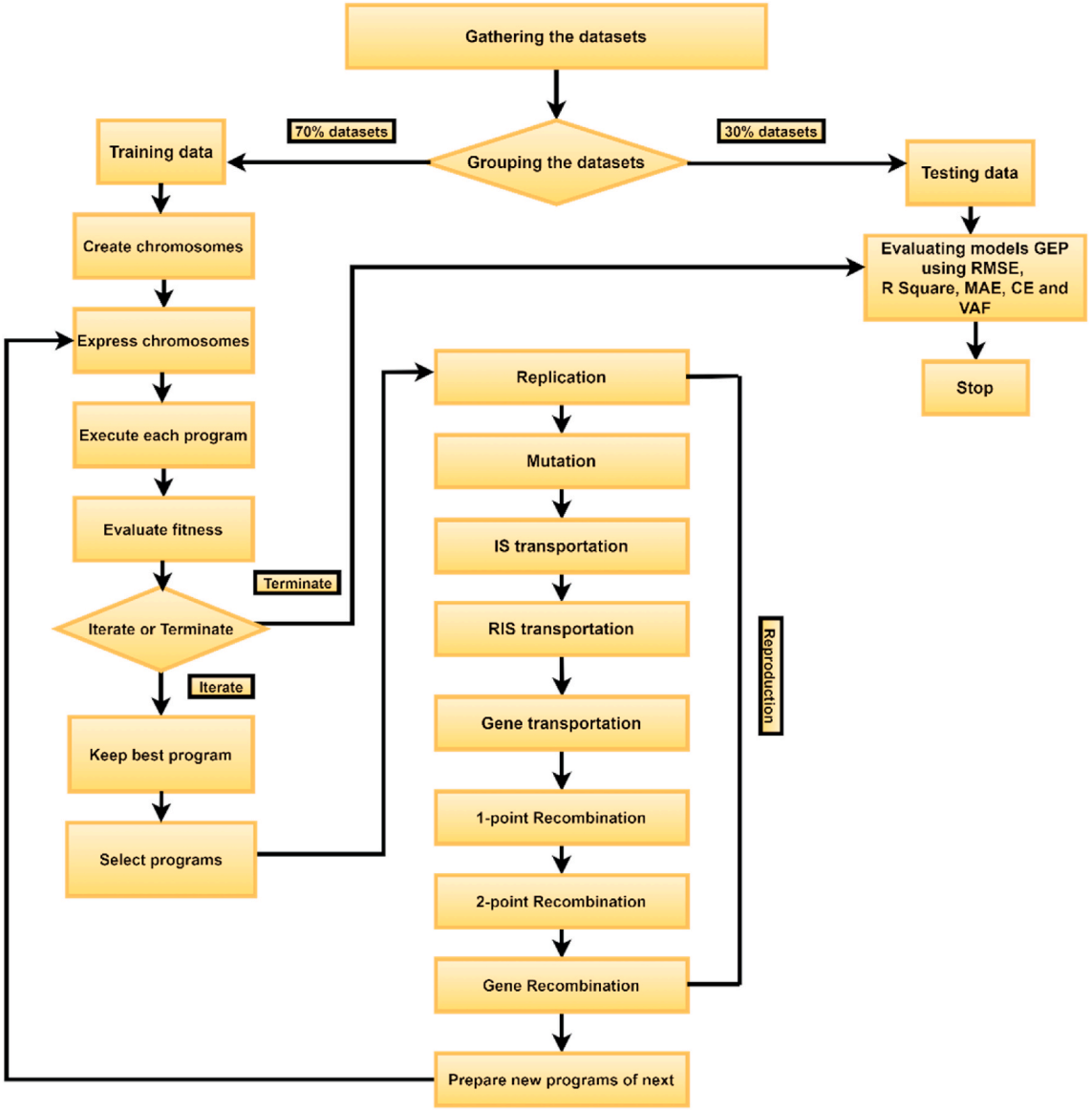

Fig. 2. Flowchart of GEP algorithm.

related to the structure of the individuals. In GA, the solutions are coded in the form of linear binary strings with the fix length which are known as chromosomes while in GP, they are nonlinear computer programs which are able to express as trees with different sizes and shapes (parse trees). In GEP, the individuals' interest from two characteristics of linear strings (chromosomes) and tree structure (which is known as expression tree (ET)). The language that chromosomes follow is called karva (see Fig. 1) as shown by Faradonbeh et al. (2016), Dindarloo (2015), Armaghani et al. (2018), and Ferreira (2001 & 2006).

The simplicity of the gene expression algorithm allows coding to be possible in any part of the program, which makes evolution effective. The two parameters of the expression tree (Expression tree) and the chromosome are the main parameters of this algorithm that the information encoded in the chromosomes are displayed by the expression trees; the display of this information is done by converting information from the chromosome to the expression tree. This code is a one-to-one relationship between chromosomes, functions and terminals. The components of a linear chromosome include Terminals (A, B, C …) and Function (+, -,/, *, sqrt, sin, cos, etc.). A chromosome in GEP can have one or more genes and also a structure based on Fig. 1 that consists of a

Table 1
Statistical indicators of input and output parameters.

Type	Parameter	Symbol	Range	Min	Max	Mean	Std. Deviation	Variance
Input	Hybrid cement (%)	Hc	12.00	0.00	12.00	6.00	0.32	12.30
	Percentage nanostructured quarry fines (%)	Nqf	1.20	0.00	1.20	0.60	0.03	0.12
	Clay content (%)	C	1.05	23.02	24.07	23.49	0.03	0.10
	Clay activity	Ac	1.40	0.60	2.00	1.35	0.04	0.16
	Frictional angle (°)	$F(\varnothing^{\circ})$	6.60	15.00	21.60	18.24	0.19	4.36
Output	Coefficient of curvature	Cc	1.12	0.84	1.96	1.41	0.03	0.09
	Coefficient of uniformity	Cu	3.81	2.05	5.86	3.80	0.12	1.75

Table 2
Pearson's correlation matrix of the input and output variables.

	Hc	Nqf	C	Ac	F	Cc	Cu
Hc	1						
Nqf	1	1					
C	0.996	0.996	1				
Ac	−0.994	−0.994	−0.997	1			
F	0.992	0.992	0.994	−0.991	1		
Cc	0.995	0.995	0.985	−0.983	0.981	1	
Cu	0.983	0.983	0.984	−0.981	0.972	0.966	1

head (h) and a tail (t) (Eq. (4)). The length of the head is determined by the GEP designer, but the length of the tail is calculated using the equation below proposed by Ferreira 2001, & 2002), Khandelwal et al. (2016) and Güllü (2012):

$$t = h(n-1) + 1 \tag{4}$$

Where "t" is the length of the tail, h is the head lengths and n is the maximum argument of the functions. In the GEP algorithm, the length of chromosomes and genes are constant and only the length of the frames (ORF) changes. This makes the GEP endpoint incompatible with the gene endpoint due to the presence of non-coding regions at the end of the gene. These non-coding regions in GEP allow operators to operate without restriction and generate genetic diversity, which is one way to achieve evolution (see Fig. 1) presented by Ferreira (2001) and Ferreira (2006).

As shown in Fig. 2 in the GEP flowchart of model activities, the GEP process begins with the random production of chromosomes from a certain number of individuals (the initial population). These chromosomes are then expressed as a tree expression (ET), and the degree to which each individual fits into a set of fit items is assessed. The individuals are then selected according to fitness to form a new generation, i.e., the higher the fitness value, the more chance an individual has to be selected. In a cycle, new offspring undergo the same developmental process as before, namely genome expression, exposure to environment selection, fit-based selection, and reproduction with improvement. This process is repeated for a certain number of generations to find a good solution following the iteration sequence presented in the works of Armaghani et al. (2018), who researched on GEP prediction of uniaxial compressive strength prediction, Dindarloo et al. (2015), who worked

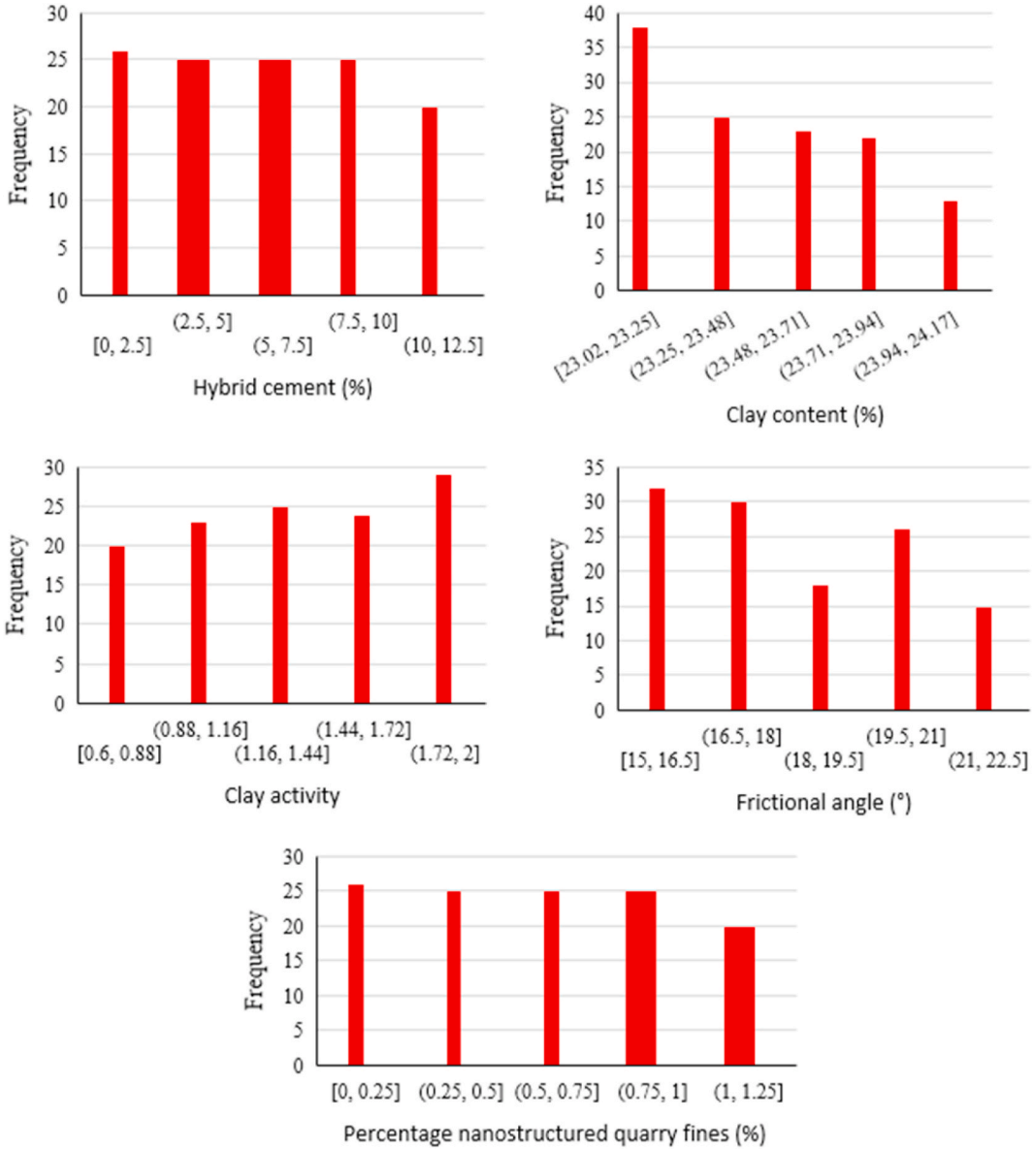

Fig. 3. The frequency histograms for the input data.

Table 3
Performance indicator values for building LMR models for Cc and Cu forecasts.

$$Cc = 21.696 + (1.601 \times NQF) + (-0.876 \times C) + (-0.024 \times F) + (-0.165 \times Ac) \quad (9)$$

$$Cu = -76.441 + (3.599 \times C) + (-0.307 \times F) + (2.187 \times NQF) \quad (10)$$

Methods	Train					Test				
	R^2	RMSE	MAE	VAF	CE	R^2	RMSE	MAE	VAF	CE
Cc	0.9959	0.0232	0.0200	0.9942	0.9959	0.9949	0.0210	0.0176	0.9942	0.9944
Cu	0.9724	0.2169	0.1670	0.9714	0.9724	0.9748	0.2395	0.2014	0.9623	0.9706

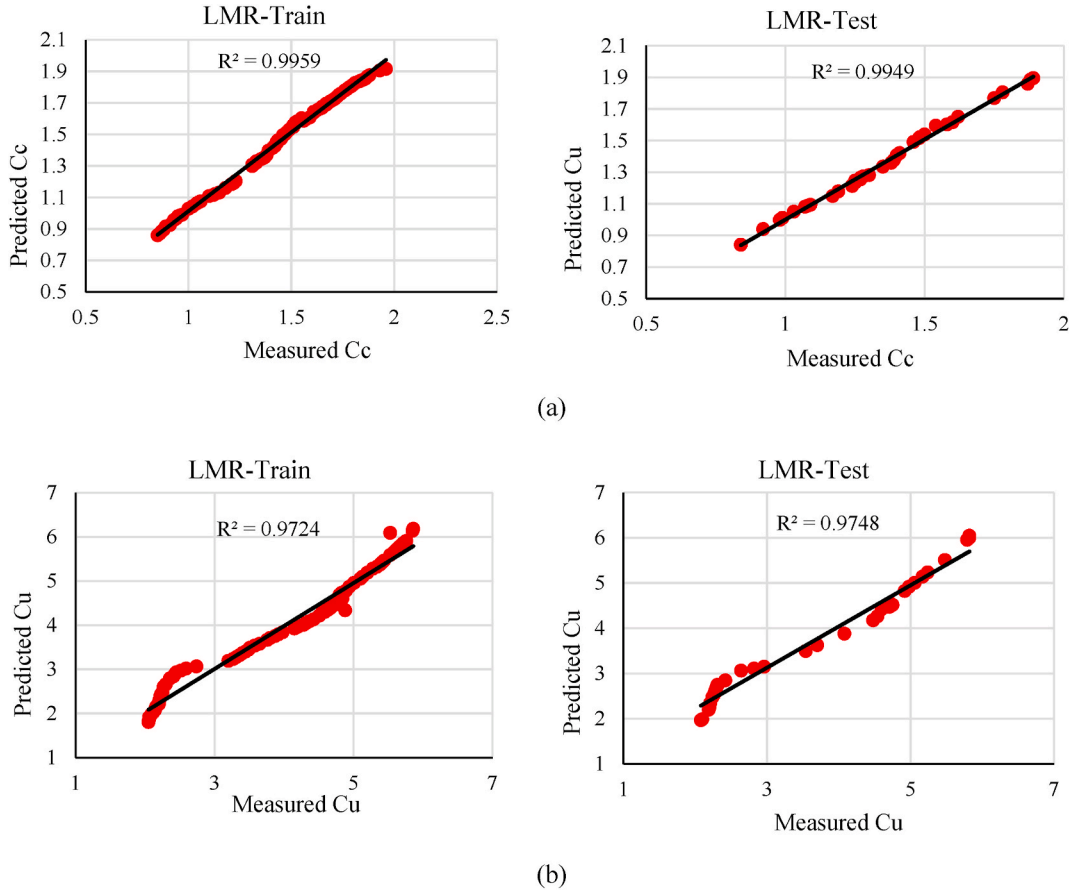

Fig. 4. The values of R-squared for LMR models, a) prediction for Cc and b) for Cu.

on the prediction of blast-induced ground vibrations via genetic programming, Mollahasani et al. (2011), who work on the empirical modeling of plate load test moduli of soil via gene expression programming, Shakeri et al. (2020) and Armaghani et al. (2016) as earlier stated. Reproduction alone cannot show changes, and only through the activity of other actors, genetic changes within the population are shown. These operators randomly select chromosomes to modify; thus, in GEP, one chromosome at a time may be modified or not modified at all by one or more genetic operators, which follow the approaches by Armaghani et al. (2018), Mollahasani et al. (2011), Ferreira (2001, 2002, & 2006) and Faradonbeh et al. (2016).

2. Materials and methods

2.1. Materials

The reference soil was collected from a borrow pit, which has served as a source of construction soil for decades now. The soil was prepared by removing lumps and sundried for three days in open air. Conversely, the rice husk was collected from rice milling factories at Abakaliki, Nigeria where the people's primary occupation is rice farming and the husk a solid waste disposed indiscriminately for lack adequate disposal management program. The husk was combusted in a controlled incinerator to check the emission of CO_2 as a result in order to achieve environmentally friendly procedure. Through this procedure, the rice husk ash was generated. Furthermore, the rice husk ash was activated by blending a 5% dosage of hydrated lime by weight of the ash in order to generate the composite binder called the hybrid cement (HC). Note, "*hydrated lime (Ca (OH)$_2$) is the quicklime combined chemically in water with 33% to 34% magnesium oxide (MgO), 46% to 48% of CaO, and 15% to 17% chemically combined with water. It is a crystal, non-flammable, odorless inorganic powder, which is soluble in water at ambient temperature. It has a melting point of 580°C, a boiling point of 2850°C, and a density of 2.21* g/cm^3. *Its density is less than that of quicklime (3.34* g/cm^3*) due to its more*

aqueous condition that creates pores in the structure of the solid. It is caustic with a pH *of 12.8 and possesses pozzolanic characteristics, which makes it a good supplementary or alternative binder in civil engineering and earthworks"*. Due to the properties outlined, hydrated lime serves as a good activator to ordinary ash materials utilized in soil stabilization. The HC is one of the major locally synthesized binders in this exercise with improved cementing properties. In addition, quarry dust (QD) was also collected from rock blasting site at Amasiri, Nigeria where aggregates are produced for construction works. The QD was further pulverized to fineness and sieved through a 200 nm sieve to generate the nano-structured quarry fines used as a second binder in the stabilization phase of this operation. These binding materials; HC and NQF were observed to meet the standard requirements for materials to be classified as pozzolanas reported in ASTM C618 (1978). Finally, the soil, HC and NQF were made ready and stored for use in the stabilization work.

2.2. Methods

2.2.1. Laboratory methods

The requirements of the British Standard International (BSI) (BS 1377 - 2 3, 1990B,) were observed in conducting general tests on the materials for the purpose of materials characterization and classification. Under the above conditions, the particle size analysis, compaction, Atterberg limits, specific gravity, and the angle of internal friction were conducted. In order to determine the oxides composition of the binder materials (RHA, HC and NQF), the XRF was carried out in accordance with the ASTM E1621-13 (2013), and in order to achieve reliable and precise results, the samples were prepared and mixed to high homogeneity and the results were observed and recorded. Further, the treatment exercise was conducted in accordance with the requirements of the BS 1924 (1990) and multiple data points were generated for varying dosages of HC and NQF between 0% and 12%.

2.2.2. Model development

The multiple experimental exercises gave rise to 121 datasets collected for the 5 independent variables (predictors); HC-hybrid cement, NQF-nanostructured quarry fines, C -clay content, A_C- Clay activity, and $\varnothing^{\circ}$-frictional angle and the 2 dependent variables (targets); C_C-coefficient of curvature, and C_U-coefficient of uniformity and produced a mathematical functional of functions as follows;

$$C_C \& C_U = \mathrm{f}(\mathrm{HC},\ \mathrm{NQF},\ \mathrm{C},\ \mathrm{A_C}, \varphi^{\circ}) \tag{5}$$

The datasets of the input and outputs variables generated from multiple experiments were deployed to the linear multivariate regression (baseline regression), gene expression programming (GEP) and artificial neural network (ANN) to predict the outputs parameters as functions of the predictors.

Table 4
The values of the parameters used for the GEP models.

GEP Parameter	Value	
	Cc	Cu
Fitness function	MSE	RMSE
Inversion rate	0.00546	0.00546
IS transportation rate	0.00546	0.00546
RIS transportation rate	0.00546	0.00546
One-point recombination rate	0.00277	0.00277
Two-point recombination rate	0.00277	0.00277
Gene size	15	19
Head size	7	9
Tail size	8	10
Mutation rate	0.00138	0.00138
Number of Chromosome	30	35
Number of genes	3	3
Gene recombination rate	0.00277	0.00277
Gene transportation rate	0.00277	0.00277
Training	70%	70%
Validation	30%	30%
Number of generation	5000	5000

2.2.2.1. Predicting with linear multivariate regression (LMR). Before predicting the coefficient of curvature and uniformity, using the gene expression algorithm and ANN methods mentioned in this research, the relationship of multivariate regression between input parameters and output parameters was investigated. A representative database has been created for valid statistical analysis, consisting of 121 datasets. The obtained regression databases as well as the relationship between input and output parameters are shown in Tables 1 and 2, respectively. After the studies, in order to construct a predictive model of curvature coefficient and uniformity coefficient, 5 data were used as input parameters and curvature coefficient and uniformity coefficient as output parameters (Table 1). Each of these 5 parameters contains 121 data. The frequency histograms for the input data are also shown in Fig. 3.

In the next step, the data were divided into two parts: model construction data and validation data. The ratio of 70% (85 data) was used for training data and 30% (36 data) was used for test data. Both ratios were randomly selected. The linear dependence of the output parameters on the input parameters was obtained using IBM SPSS v25 software. The results of linear regression analysis for the obtained model data are shown in Table 3. According that, five statistical indicators have been used to select the best models throughout the article.

The VAF (see Eq. (6)) index shows the degree of difference between the variance of the measured and predicted data sets. VAF values close to 1 or 100% indicate low variability and therefore better predictability. The RMSE (see Eq. (11)) index is the square of the average error between the measured and predicted data. The lower the RMSE, the better the model performance. Ideally, the values of RMSE and MAE (see Eq. (13)) should be close to zero and that of CE (see Eq. (7)) should be one. Also the coefficient of determination R^2 (see Eq. (8)), which is the correlation between the measured and predicted data, should be close to 1, which agrees with Asadizadeh and Hossaini (2016). In fact, to find the best relationship, the five statistical parameters mentioned were selected as the best criteria. Table 3 shows the best values for Cc and Cu, and Eqs. (8) and (9) for Cc and Cu, respectively, were obtained using the LMR method. Fig. 4 also shows the relationship between the actual values of Cc and Cu and the values predicted by linear regression for the train and test data.

$$VAF = 100\left(1 - \frac{var(Ximes - Xipred)}{var(Ximes)}\right) \tag{6}$$

$$CE = 1 - \frac{\sum_{k=1}^{N}(Xipred - Ximes)^2}{\sum_{k=1}^{N}\left(Xipred - \overline{Xipred}\right)^2} \tag{7}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{(i=1)}^{N}(\mathrm{Ximes} - \mathrm{Xipred})^2}{\sum_{(i=1)}^{N}\left(\mathrm{Ximes} - \overline{\mathrm{Ximes}}\right)^2} \tag{8}$$

That: Xipred is the predicted values and Ximes are the measured values.

2.2.2.2. Proposed GEP. The main purpose of the GEP designing is function finding for prediction of the coefficient of curvature and uniformity. For this aim, GeneXproTools 5.0 software was used. In this model, hybrid cement (Hc), percentage nanostructured quarry fines (Nqf), clay content (C), clay activity (Ac) and frictional angle (F) are considered the inputs of the model (terminal set) and coefficient of curvature (Cc) and coefficient of uniformity (Cu) are set as the outputs. First, as the linear regression method discussed first, the data are divided into test and training data. The purpose of using some datasets in the training stage is to develop a new model while testing datasets are used for checking the performance prediction of the developed model (Khandelwal et al., 2016). The ratio of 70% (85 data) was used for

training data and 30% (36 data) was used for test data. Both ratios were randomly selected. According to previous studies, GEP optimization consists of five main steps, the first of which is the selection of the fit function. In this study, MSE, RMSE, RRSE and MAE fitness functions were used for prediction. Among these functions, the MSE and RMSE fitness functions yielded the best results for predicting Cc and Cu, respectively. The relationships of each of the fitness functions are listed below (Eqs. (11)–(14)) as presented in the works of Shakeri et al. (2020), Khandelwal et al. (2016), Dehghani (2018) and Armaghani et al. (2018).

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{Xipred} - \mathrm{Ximes})^2} \tag{11}$$

$$\mathbf{Cc} = \mathrm{Nqf} + \left(\mathrm{Hc} \times \left(\frac{\mathrm{Hc}}{\mathrm{Exp(Hc)} + (\mathrm{F} \times 7.815)}\right)\right) + \left(\left(\frac{\mathrm{Ac}}{\mathrm{Ac} \times \mathrm{F}}\right) \times ((-10.053 + \mathrm{C}) + \mathrm{Nqf})\right) \tag{15}$$

$$\mathbf{Cu} = \left(\frac{\mathrm{F}}{(\mathrm{F} - ((6.670 - \mathrm{Ac}) \times \mathrm{Ac})) + \mathrm{Ac}^6}\right) + \left(\sqrt{(((-4.627 + \mathrm{Hc}) + \mathrm{F}) + 0.357) + ((\mathrm{Nqf} \times \mathrm{Hc}) - (\mathrm{Nqf} + 7.937))}\right) + \left(\frac{1}{\left(\left(\frac{\mathrm{C}}{7.019}\right) + (\mathrm{Ac} - \mathrm{Hc})\right) + ((\mathrm{F} - \mathrm{C}) + 0.702)} \times \mathrm{Hc}\right) \tag{16}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{Xipred} - \mathrm{Ximes})^2 \tag{12}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\mathrm{Xipred} - \mathrm{Ximes}| \tag{13}$$

$$\mathrm{RRSE} = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{Xipred} - \mathrm{Ximes})^2}{\sum_{i=1}^{N}\left(\overline{\mathrm{Ximes}} - \mathrm{Ximes}\right)^2}} \tag{14}$$

That: Xipred is the predicted values and Ximes are the measured values. Then a set of terminals and a set of functions were selected to form chromosomes (according to the input and output of the model). In the next step, the most suitable functions were determined to obtain the best final equations in agreement with the approaches of Shakeri et al. (2020), Khandelwal et al. (2016), Dehghani (2018) and Armaghani et al. (2018).

Function set = +, -, *,/, sqrt, x, x∧2, x∧3, x∧(1/3), 1/x, exp(x), ln(x).

The third step involves selecting the chromosomal structure. The fourth step involves selecting the type of link function, where the addition function is used to predict both final models, and finally the final step is to generate a set of genetic operators and rates. According to Table 4, due to the different models built and according to what has been done in previous studies, fixed values of software had better results in building models, which agree with Shakeri et al. (2020), Armaghani et al. (2018), Khandelwal et al. (2016), and Dehghani (2018). Finally, the values of the measurement indices similar to those performed in the linear regression method were calculated for the best models obtained for Cc and Cu, which are listed in Table 5.

According to Table 5, the best forecasting models are based on the values of R^2, RMSE, MAE, VAF and CE. Among the best Cc models, model 2 and for Cu model 1 due to the high value of R2, VAF and CE and low RMSE and MAE values were selected as the best models from all the models mentioned in Table 5. Fig. 5 also shows the relationship between the actual values of Cc and Cu and the values predicted by GEP for the train and test data.

Using the GEP algorithm, the final relation for predicting Cc and Cu is obtained as an expression tree (see Fig. 6), which are plotted based on inputs and fixed coefficients. Each of the sub-ETs in both models is connected using the addition function, which ultimately forms relationships in Equations (15) and (16), to predict the Cc and Cu.

2.2.2.3. ANNs prediction. Possessing sufficient data for network training and utilizing proper training algorithm, the neural network can learn well the complex relationships between the inputs and outputs. In this section, in order to predict Cc and Cu using ANN, as in the previous two methods, LMR and GEP, the data were divided into two categories. So the ratio of 70% (85 data) was used for training data and 30% (36 data) was used for test data.

A feed forward back propagation multilayer perceptron neural network was utilized to predict the Cc and Cu. Hagan and Menhaj (1994) and Jahed Armaghani et al., 2016c stated that the Levenberg– Marquardt (LM) is the most appropriate algorithm for training an ANN. This method was also used in this article. According to the results of training and testing of neural networks with one and two concealed layers and the different number of neurons in each layer, it was found that the network with one hidden layer containing 9 neurons has the lowest error rate and the highest coefficient of determination (R^2) to predict Cc and Cu using the prepared database. Since for each model the number of inputs is 5 and the number of outputs is 1 parameter, so the network with structure 1-9-5 (see Fig. 7) was considered as the optimal network, and the parameters presented for Cc and Cu are listed in Table 6.

Table 7, shows the results obtained from the optimal neural network including R^2, VAF, CE, RMSE and MAE for training and test data. Like the LMR and GEP methods mentioned earlier, the selection criteria for the two final models in Table 7, were based on the highest R^2, VAF, CE and the lowest MAE and RMSE. Fig. 8 also shows the relationship between the actual values of Cc and Cu and the values predicted by ANN for the train and test data.

3. Results and discussions

3.1. Materials characterization

The soil was observed from Fig. 9 to be poorly graded soil with

Table 5
Values of performance indicators for building the GEP models.

Methods		Train					Test				
		R^2	RMSE	MAE	VAF	CE	R^2	RMSE	MAE	VAF	CE
Cc	RMSE	0.9922	0.0269	0.0224	0.9922	0.9921	0.9895	0.0301	0.0250	0.9892	0.9879
	MSE	0.9975	0.0176	0.0151	0.9967	0.9964	0.9976	0.0140	0.0113	0.9974	0.9973
	MAE	0.9512	0.0673	0.0509	0.9512	0.9479	0.9612	0.0542	0.0396	0.9611	0.9574
	RRSE	0.9901	0.0304	0.0248	0.9901	0.9900	0.9865	0.0344	0.0291	0.9855	0.9847
Cu	RMSE	0.9864	0.1517	0.1158	0.9864	0.9862	0.9872	0.1707	0.1397	0.9860	0.9823
	MSE	0.9662	0.2344	0.1762	0.9662	0.9648	0.9710	0.2398	0.1823	0.9708	0.9708
	MAE	0.9792	0.1916	0.1502	0.9780	0.9758	0.9772	0.2179	0.1725	0.9722	0.9669
	RRSE	0.9659	0.2396	0.1864	0.9655	0.9628	0.9746	0.2271	0.1681	0.9731	0.9698

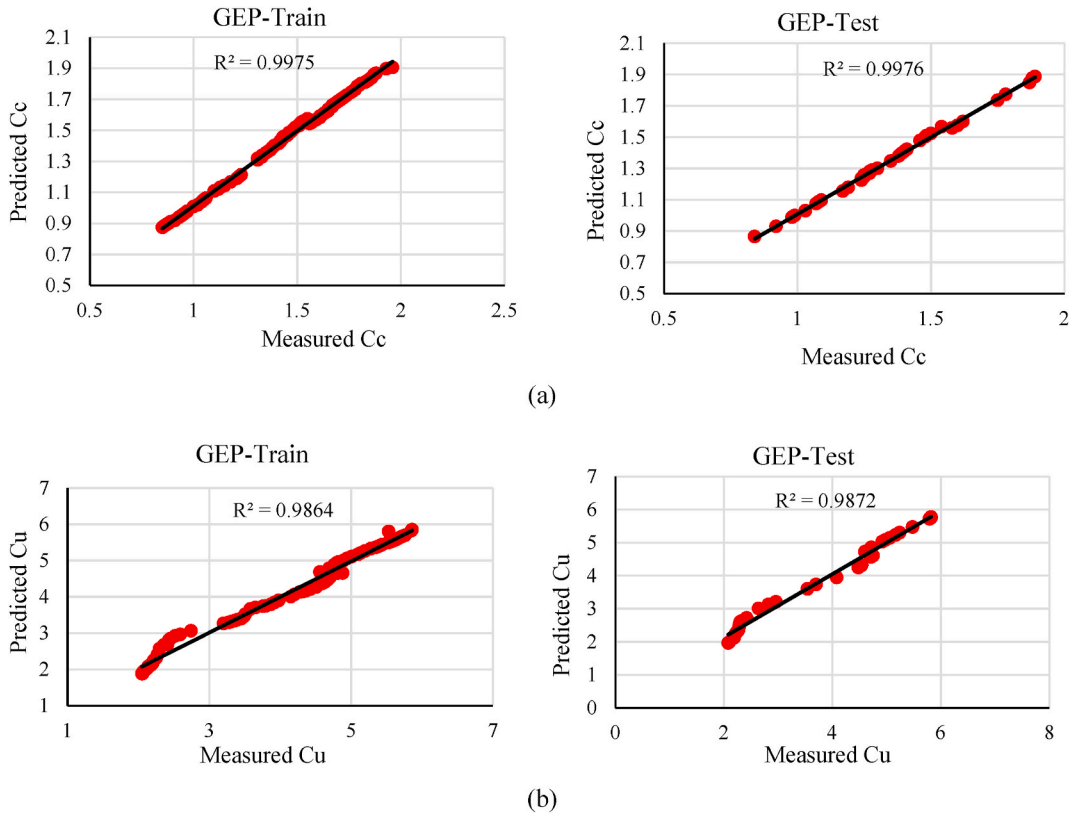

Fig. 5. The values of R-squared for GEP models, a) prediction for Cc and b) for Cu.

coefficients of curvature and uniformity as 0.84 and 2.05 respectively and classified as an A-7-6 group according to AASHTO classification method. The soil was discovered to be high plastic with high clay content. The angle of internal friction of the soil was observed to be 15°, with clay content and clay activity of 23.02% and 2 respectively (see Appendix A). It can be observed from Table 8 that the RHA, HC and NQF showed characteristics of pozzolanas with the combined oxide compositions of SiO_2, Al_2O_3 and Fe_2O_3 as more than 70% (ASTM C618, 1978). The behavior was important in the stabilization operation where these materials in single and composite forms were blended with soil to trigger hydration, pozzolanic, calcination, and cation exchange reactions culminating to the behavioral changes observed in the treated soil with respected to the measured parameters.

3.2. Predicted results validation

As mentioned, the main aim of this paper was, predicting the coefficient of curvature (Cc) and coefficient of uniformity (Cu) based on the parameters including the hybrid cement (Hc), Percentage nanostructured quarry fines (Nqf), Clay content (C), clay activity (Ac) and frictional angle (F). To propose the relationships, firstly, the data collected from the multiple experiments were used to establish a database between the input and output data. Then, the best relationships between the outputs and each input were investigated by comprehensive statistical analyses. Subsequently, the GEP, ANN and LMR algorithms were applied to predicting the Cc and Cu. The obtained results of these three methods were then compared based on the highest R^2, VAF, CE and the lowest MAE and RMSE values for training and test stages. The values obtained for all three methods are listed in Table 9 based on the statistical indicators.

According to Table 9, all three methods GEP, ANN and LMR have high accuracy in predicting Cc and Cu values due to the high correlation of the collected data, and if we consider the best models based on the desired indicators, we can say ANN, GEP and LMR had the best performance in predicting Cc and Cu, respectively (see Fig. 10). But what makes the GEP method superior to ANN and LMR and other algorithms is the high ability to predict based on a strong statistical relationship (Shakeri et al., 2020), which is also shown in this study (Eqs (15) and

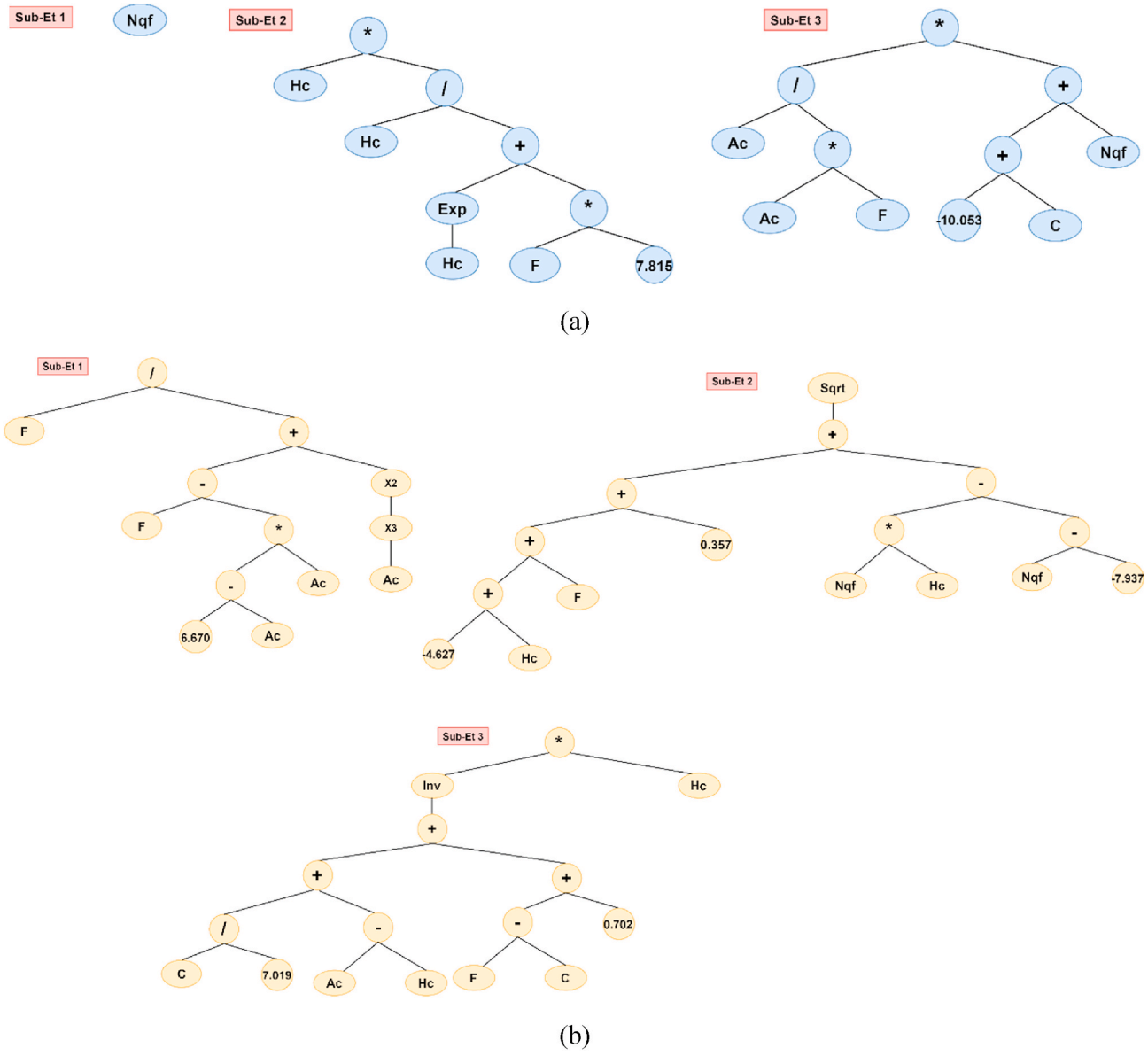

Fig. 6. The expression tree corresponding to, a) Cc and b) Cu.

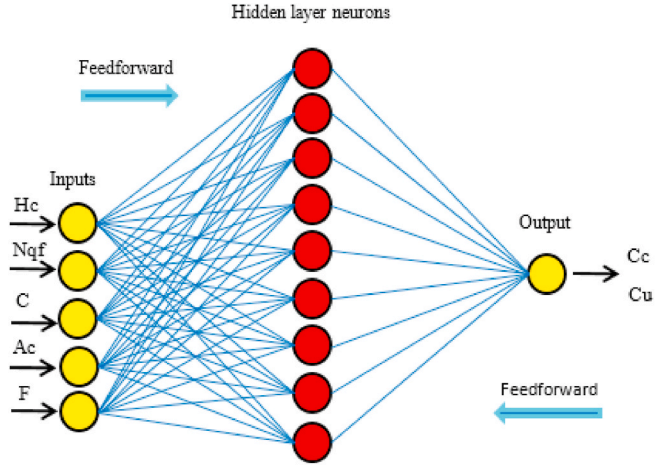

Fig. 7. Network with structure 1-5-9 and Logsig membership function for hidden layer and Pureline for output layer.

Table 6
The optimal neural network parameters for Cc and Cu prediction.

No. of Input layer neurons	5
No. of hidden layers	1
No. of hidden layer neurons	9 neurons
No. of Output layer neurons	1 neuron
Learning algorithm	Levenberg-Marquradt
Hidden layer transfer function	Log-Sigmoid (logsig)
Output layer transfer function	Pureline
No. of total data	121
No. of train data	85
No. of test data	36
No. of epochs	300
Learning rate	0.05

(16)).

Fig. 11 shows a cross plot of the measured data versus predicted values of Cc and Cu using the GEP, the ANN and the LMR algorithms. As mentioned, based on this figure, all three methods have high accuracy in predicting Cc and Cu, but it can be said that the ANN method with a very

Table 7
Values of performance indicators for building the ANN models.

Methods	Train					Test				
	R^2	RMSE	MAE	VAF	CE	R^2	RMSE	MAE	VAF	CE
Cc	0.9999	0.0029	0.0023	0.9999	0.9999	0.9995	0.0064	0.0046	0.9995	0.9995
Cu	0.9996	0.0267	0.0137	0.9996	0.9996	0.9974	0.0747	0.0392	0.9969	0.9967

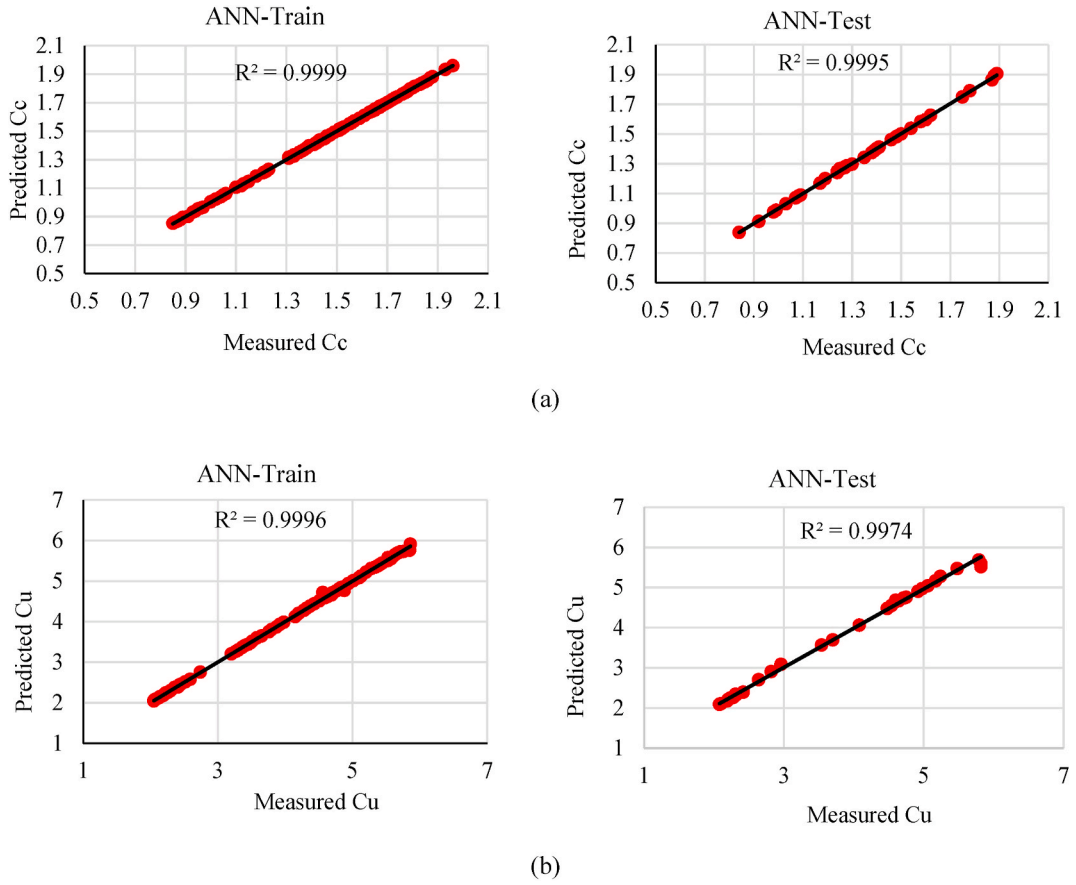

Fig. 8. The values of R-squared for ANN models, a) prediction for Cc and b) for Cu.

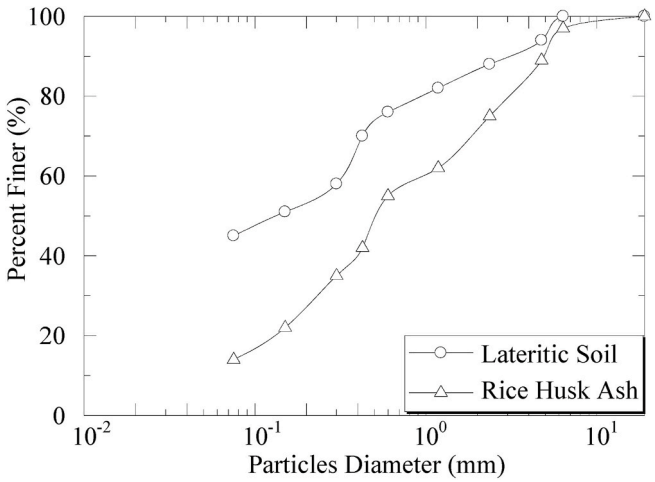

Fig. 9. Particle size distribution curve of clayey soil and rice husk ash.

Table 8
Chemical oxide composition of the additive materials.

Materials	Oxides composition (content by weight, %)												
	SiO_2	Al_2O_3	CaO	Fe_2O_3	MgO	K_2O	Na_2O	TiO_2	LOI	P_2O_5	SO_3	IR	free CaO
clay soil	12.45	18.09	2.30	10.66	4.89	12.10	34.33	0.07	–	5.11	–	–	–
rice husk ash	56.48	22.72	5.56	3.77	4.65	2.76	0.01	3.17	0.88	–	–	–	–
HC	59.12	25.3	6.3	4.23	2.5	1.21	–	1.34	–	–	–	–	–
NQF	62.48	18.72	4.83	6.54	2.56	3.18	–	0.29	1.01	–	–	–	–

*IR is insoluble residue; LOI is loss on ignition.

Table 9
All predicted values of R2, VAF, CE, RMSE and MAE for forecasting Cc and Cu.

Methods											
		Train					Test				
		R^2	RMSE	MAE	VAF	CE	R^2	RMSE	MAE	VAF	CE
Cc	LMR	0.9959	0.0232	0.0200	0.9942	0.9959	0.9949	0.0210	0.0176	0.9942	0.9944
	GEP	0.9975	0.0176	0.0151	0.9967	0.9964	0.9976	0.0140	0.0113	0.9974	0.9973
	ANN	0.9999	0.0029	0.0023	0.9999	0.9999	0.9995	0.0064	0.0046	0.9995	0.9995
Cu	LMR	0.9724	0.2169	0.1670	0.9714	0.9724	0.9748	0.2395	0.2014	0.9623	0.9706
	GEP	0.9864	0.1517	0.1158	0.9864	0.9862	0.9872	0.1707	0.1397	0.9860	0.9823
	ANN	0.9996	0.0267	0.0137	0.9996	0.9996	0.9974	0.0747	0.0392	0.9969	0.9967

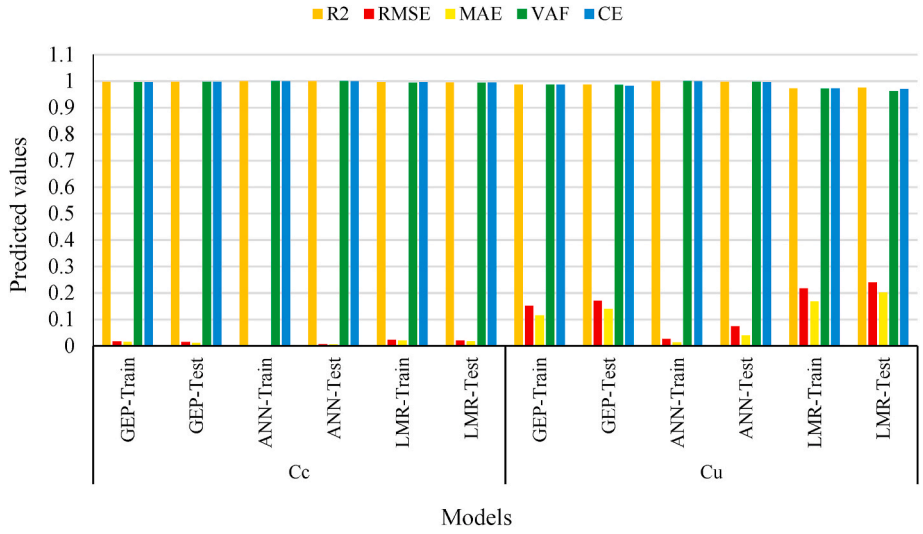

Fig. 10. Comparison of predicted values of R^2, VAF, CE, RMSE and MAE for forecasting Cc and Cu.

small difference than the other two methods has better accuracy for forecasting.

3.3. Sensitivity analysis

A useful concept has been proposed to identify the significance of each "cause" factors (the input data) on the "effect factor" (the output). This enables us to hierarchically recognize the most sensitive factors affecting the Cc and Cu. For achieving this aim, sensitivity analysis was carried out to recognize the relative influence of each parameter in the network system by the cosine amplitude method, which agrees with Yang and Zhang (1997). To apply this method, all of the data pairs were expressed in common X-space. To undertake this technique, all data pairs should be utilized to build a data array X presented by Faradonbeh et al. (2016), Khandelwal et al. (2016), and Sayevand et al. (2018) as follows:

$$X = \{x_1, x_2, x_3, \ldots, x_i, \ldots, x_n\} \tag{17}$$

Each of the elements, xi, in the data array X is a vector of lengths of m, that is:

$$X = \{x_1 1, x_2 2, x_3 3, \ldots, x_{im}\} \tag{18}$$

The strength of the relation between the dataset, xi and xj, is presented as follows:

$$r_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{ik}^2}} \tag{19}$$

Fig. 12 shows the strengths of the relations (rij values) between the model inputs and outputs. The results showed that, F, C, H (HC), Nqf and Ac have the most and the least effect on Cc, respectively (Fig. 11a). While the highest and lowest effects on Cu were obtained as H (HC), Nqf, F, C and Ac (Fig. 11b).

4. Conclusions

ANN, GEP and baseline regression, the LMR have been used to predict the coefficients of curvature (Cc) and uniformity (Cu) of lateritic soil with a saturation degree of 60% treated with the dual effects of hybrid

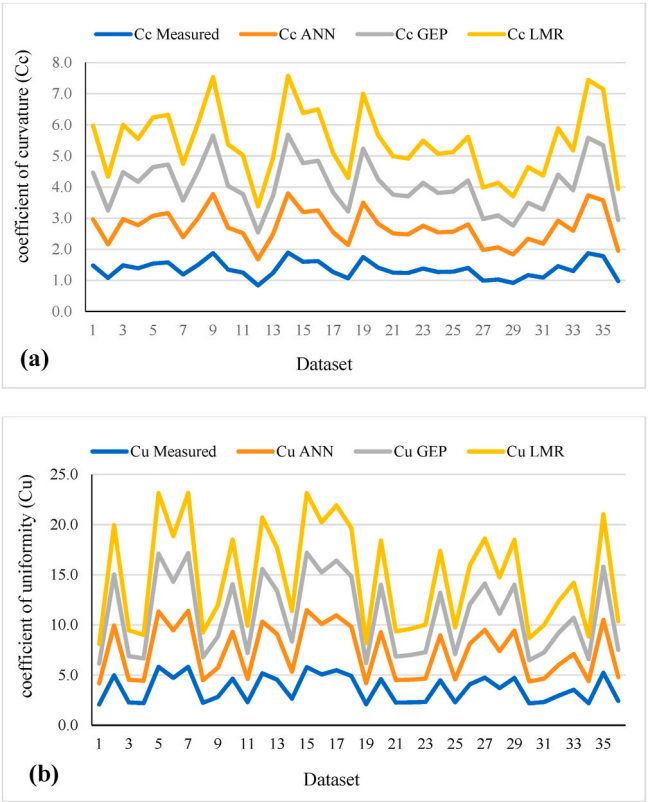

Fig. 11. Comparison of measured and predicted values using GEP, ANN and LMR models for a) coefficient of curvature (Cc) and b) coefficient of uniformity (Cu).

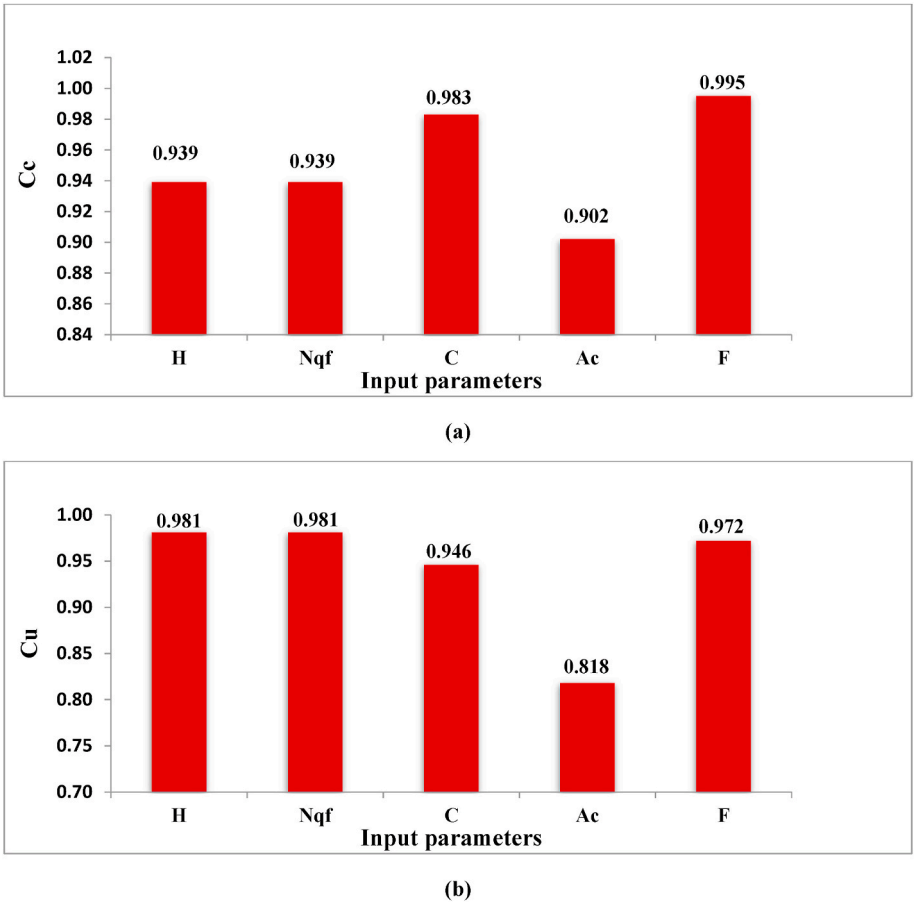

Fig. 12. Sensitivity analysis to determine the impact of each data on the output for a) Cc and b) Cu.

cement (HC or H) and nanostructured quarry fines (NQF). Multiple experiments were conducted to generate datasets for the input and output variables. The selected predictor variables were HC, NQF, C, Ac and frictional angle designated as F while the targets were Cc and Cu. From the foregoing, the following conclusions can be drawn;

- The LMR showed a linear and closer correlation between the input and outputs variables as this showed R of more than 0.95
- The LMR, GEP and ANN predicted the outputs with flexibility and accuracy and closer agreement
- From the predictions performance, ANN outclassed the LMR and GEP, however, what makes the GEP method superior to ANN and LMR and other algorithms is the high ability to predict based on a strong statistical relationship
- Generally, the three techniques performed well in the prediction of the gradation properties of the treated soil.

Declaration of competing interest

The authors declare no competing or conflicting interests in this work. We further declare that this work is a review and efforts have been made to avoid any unethical issues.

Appendix A

Treated and untreated soil with five (5) predictor parameters, two targets and 121 datasets

Predictor parameters					C_C	C_U
HC (%)	NQF (%)	C (%)	A_C	$\varnothing^{\circ}$		
0	0	23.02	2.0	15	0.84	2.05
0.1	0.01	23.02	1.98	15.01	0.85	2.05
0.2	0.02	23.03	1.96	15.04	0.86	2.06
0.3	0.03	23.03	1.96	15.09	0.87	2.06
0.4	0.04	23.04	1.93	15.14	0.88	2.06
0.5	0.05	23.04	1.9	15.19	0.89	2.07
0.6	0.06	23.05	1.88	15.25	0.91	2.08
0.7	0.07	23.05	1.88	15.29	0.92	2.09
0.8	0.08	23.05	1.87	15.32	0.93	2.1
0.9	0.09	23.06	1.85	15.38	0.94	2.12
1	0.1	23.07	1.8	15.44	0.95	2.14
1.1	0.11	23.08	1.8	15.5	0.96	2.15
1.2	0.12	23.09	1.81	15.55	0.97	2.16
1.3	0.13	23.10	1.8	15.6	0.98	2.18
1.4	0.14	23.10	1.81	15.65	0.99	2.19
1.5	0.15	23.10	1.8	15.7	1	2.19
1.6	0.16	23.10	1.8	15.74	1.02	2.2
1.7	0.17	23.11	1.79	15.8	1.03	2.2
1.8	0.18	23.11	1.81	15.85	1.04	2.2
1.9	0.19	23.12	1.8	15.9	1.05	2.21
2	0.2	23.13	1.8	15.95	1.06	2.21
2.1	0.21	23.14	1.8	16	1.07	2.22
2.2	0.22	23.15	1.8	16.05	1.08	2.23
2.3	0.23	23.16	1.8	16.1	1.09	2.24
2.4	0.24	23.16	1.8	16.15	1.1	2.25
2.5	0.25	23.17	1.8	16.2	1.12	2.26
2.6	0.26	23.18	1.79	16.25	1.13	2.27
2.7	0.27	23.19	1.77	16.3	1.15	2.27
2.8	0.28	23.19	1.75	16.35	1.17	2.28
2.9	0.29	23.2	1.72	16.4	1.18	2.29
3	0.3	23.2	1.7	16.45	1.19	2.3
3.1	0.31	23.21	1.7	16.5	1.21	2.3
3.2	0.32	23.22	1.7	16.55	1.22	2.31
3.3	0.33	23.22	1.71	16.6	1.23	2.32
3.4	0.34	23.23	1.69	16.65	1.24	2.36
3.5	0.35	23.24	1.7	16.7	1.24	2.41
3.6	0.36	23.24	1.69	16.75	1.25	2.42
3.7	0.37	23.25	1.65	16.8	1.25	2.43
3.8	0.38	23.26	1.64	16.85	1.27	2.45
3.9	0.39	23.27	1.61	16.9	1.27	2.52
4	0.4	23.28	1.6	16.95	1.28	2.59
4.1	0.41	23.29	1.59	16.99	1.30	2.64
4.2	0.42	23.29	1.57	17.05	1.31	2.74
4.3	0.43	23.3	1.55	17.1	1.31	2.82
4.4	0.44	23.31	1.52	17.15	1.33	2.96
4.5	0.45	23.32	1.5	17.2	1.33	3.2
4.6	0.46	23.33	1.5	17.25	1.35	3.27
4.7	0.47	23.34	1.5	17.31	1.35	3.31
4.8	0.48	23.35	1.5	17.35	1.36	3.36
4.9	0.49	23.36	1.5	17.39	1.37	3.4
5	0.5	23.37	1.5	17.45	1.38	3.45
5.1	0.51	23.38	1.48	17.5	1.38	3.49
5.2	0.52	23.39	1.46	17.55	1.39	3.51

(continued on next page)

(*continued*)

Treated and untreated soil with five (5) predictor parameters, two targets and 121 datasets						
Predictor parameters					C_C	C_U
HC (%)	NQF (%)	C (%)	A_C	$\varnothing^{\circ}$		
5.3	0.53	23.39	1.43	17.6	1.39	3.54
5.4	0.54	23.4	1.41	17.65	1.4	3.58
5.5	0.55	23.41	1.4	17.7	1.41	3.65
5.6	0.56	23.42	1.4	17.75	1.41	3.7
5.7	0.57	23.43	1.41	17.8	1.42	3.76
5.8	0.58	23.44	1.42	17.85	1.42	3.8
5.9	0.59	23.45	1.41	17.9	1.425	3.87
6	0.6	23.46	1.4	17.95	1.43	3.92
6.1	0.61	23.47	1.38	17.99	1.43	3.98
6.2	0.62	23.48	1.37	18.05	1.435	4.08
6.3	0.63	23.49	1.35	18.09	1.45	4.15
6.4	0.64	23.5	1.33	18.15	1.455	4.2
6.5	0.65	23.51	1.3	18.2	1.46	4.28
6.6	0.66	23.52	1.31	18.25	1.46	4.32
6.7	0.67	23.53	1.31	18.3	1.47	4.37
6.8	0.68	23.54	1.3	18.35	1.475	4.43
6.9	0.69	23.55	1.3	18.4	1.48	4.48
7	0.7	23.56	1.3	18.45	1.484	4.51
7.1	0.71	23.57	1.29	18.5	1.49	4.54
7.2	0.72	23.58	1.27	18.55	1.5	4.59
7.3	0.73	23.59	1.26	18.6	1.51	4.63
7.4	0.74	23.6	1.23	18.65	1.51	4.67
7.5	0.75	23.61	1.2	18.7	1.52	4.7
7.6	0.76	23.62	1.19	18.75	1.52	4.71
7.7	0.77	23.63	1.18	18.8	1.53	4.75
7.8	0.78	23.64	1.16	18.84	1.54	4.8
7.9	0.79	23.65	1.14	18.89	1.55	4.84
8	0.8	23.66	1.13	19.95	1.56	4.88
8.1	0.81	23.67	1.12	20	1.57	4.56
8.2	0.82	23.68	1.11	20.05	1.58	4.6
8.3	0.83	23.69	1.11	20.1	1.59	4.64
8.4	0.84	23.7	1.1	20.15	1.60	4.69
8.5	0.85	23.71	1.0	20.2	1.61	4.72
8.6	0.86	23.72	1.0	20.25	1.61	4.76
8.7	0.87	23.73	1.0	20.28	1.62	4.79
8.8	0.88	23.74	1.0	20.32	1.63	4.8
8.9	0.89	23.75	1.0	20.36	1.64	4.83
9	0.9	23.76	1.0	20.4	1.65	4.89
9.1	0.91	23.77	1.0	20.44	1.65	4.92
9.2	0.92	23.78	1.0	20.48	1.66	4.94
9.3	0.93	23.79	1.0	20.52	1.67	4.98
9.4	0.94	23.8	1.0	20.56	1.67	5.01
9.5	0.95	23.81	1.0	20.6	1.68	5.06
9.6	0.96	23.82	0.99	20.64	1.69	5.1
9.7	0.97	23.83	0.98	20.68	1.7	5.13
9.8	0.98	23.84	0.97	20.72	1.71	5.17
9.9	0.99	23.85	0.94	20.76	1.72	5.2
10	1	23.86	0.9	20.8	1.73	5.24
10.1	1.01	23.87	0.88	20.84	1.74	5.28
10.2	1.02	23.88	0.86	20.88	1.75	5.34
10.3	1.03	23.89	0.84	20.92	1.76	5.38
10.4	1.04	23.9	0.82	20.96	1.77	5.41
10.5	1.05	23.91	0.8	21	1.78	5.44
10.6	1.06	23.92	0.79	21.04	1.78	5.48
10.7	1.07	23.94	0.78	21.08	1.79	5.53
10.8	1.08	23.95	0.75	21.12	1.8	5.57
10.9	1.09	23.96	0.72	21.16	1.81	5.6
11	1.1	23.97	0.7	21.2	1.83	5.63
11.1	1.11	23.98	0.7	21.24	1.84	5.66
11.2	1.12	23.99	0.71	21.28	1.85	5.69
11.3	1.13	24	0.71	21.32	1.86	5.72
11.4	1.14	24.01	0.71	21.36	1.87	5.76
11.5	1.15	24.02	0.7	21.4	1.87	5.79
11.6	1.16	24.03	0.69	21.44	1.88	5.82
11.7	1.17	24.04	0.67	21.48	1.88	5.82
11.8	1.18	24.05	0.65	21.52	1.89	5.53
11.9	1.19	24.06	0.62	21.56	1.93	5.85
12	1.2	24.07	0.6	21.6	1.96	5.86

HC = hybrid cement, NQF = nanostructured quarry fines, C = clay content, A_C = Clay activity, $\varnothing^{\circ}$ = Frictional angle, C_C = Coefficient of curvature, C_U = Coefficient of uniformity.

References

Armaghani, D.J., Mahdiyar, A., Hasanipanah, M., Faradonbeh, R.S., Khandelwal, M., Amnieh, H.B., 2016. Risk assessment and prediction of flyrock distance by combined multiple regression analysis and Monte Carlo simulation of quarry blasting. Rock Mech. Rock Eng. 49 (9), 3631–3641.

Armaghani, A., DJ, Faradonbeh, R.S., Rezaei, H., Rashid, A.S.A., Amnieh, H.B., 2018. Settlement prediction of the rock-socketed piles through a new technique based on gene expression programming. Neural Comput. Appl. 29 (11), 1115–1125.

Armaghani, D.J., Safari, V., Fahimifar, A., Monjezi, M., Mohammadi, M.A., 2018. Uniaxial compressive strength prediction through a new technique based on gene expression programming. Neural Comput. Appl. 30 (11), 3523–3532.

Asadizadeh, M., Hossaini, M.F., 2016. Predicting rock mass deformation modulus by artificial intelligence approach based on dilatometer tests. Arabian Journal of Geosciences 9 (2), 96.

American Standard for Testing and Materials (ASTM) C618, 1978. Specification for Pozzolanas. ASTM International, Philadelphia, USA.

American Standard for Testing and Materials (ASTM) E1621-13, 2013. Standard Guide for Elemental Analysis by Wavelength Dispersion X-Ray Fluorescence Spectrometry. ASTM International, West Conshohocken, PA. https://doi.org/10.1520/E1621-13.

Baykasoğlu, A., Güllü, H., Çanakçı, H., Özbakır, L., 2008. Prediction of compressive and tensile strength of limestone via genetic programming. Expert Syst. Appl. 35 (1–2), 111–123.

BS 1377 - 2, 3, 1990. Methods of Testing Soils for Civil Engineering Purposes. British Standard Institute, London.

BS 1924, 1990. Methods of Tests for Stabilized Soil. British Standard Institute, London.

Çelik, Ş., 2020. Estimation modelling of tobacco production in Turkey: comparative analysis of artificial neural networks and multiplicative decomposition methods. International Journal of Trend in Research and Development 7 (4), 154–187.

Cigizoglu, H.K., Kis, O.Z.R., 2005. Flow prediction by three back propagation techniques using k-fold partitioning of neural network training data. Nord. Hydrol 36, 49–64.

Dehghani, H., 2018. Forecasting copper price using gene expression programming. Journal of Mining and Environment 9 (2), 349–360 ().

Dindarloo, S.R., 2015. Prediction of blast-induced ground vibrations via genetic programming. International Journal of Mining Science and Technology 25 (6), 1011–1015.

Faradonbeh, R.S., Armaghani, D.J., Abd Majid, M.Z., Tahir, M.M., Murlidhar, B.R., Monjezi, M., Wong, H.M., 2016. Prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction. Int. J. Environ. Sci. Technol. 13 (6), 1453–1464.

Ferreira, C., 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems. arXiv preprint cs/0102027.

Ferreira, C., 2002. Gene expression programming in problem solving. In: Roy, R., Köppen, M., Ovaska, S., Furuhashi, T., Hoffmann, F. (Eds.), Soft Computing and Industry: Recent Applications. Springer, pp. 635–654.

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

Güllü, H., 2012. Prediction of peak ground acceleration by genetic expression programming and regression: a comparison using likelihood-based measure. Eng. Geol. 141, 92–113.

Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Network. 5 (6), 989–993.

He, X., Xu, S., 2007. Process Neural Networks Theory and Applications. Springer.

Jahed Armaghani, D., Hasanipanah, M., Mohamad, E.T., 2016c. A combination of the ICA-ANN model to predict air- overpressure resulting from blasting. Eng. Comput. 32 (1), 155–171.

Khandelwal, M., Armaghani, D.J., Faradonbeh, R.S., Ranjith, P.G., Ghoraba, S., 2016. A new model based on gene expression programming to estimate air flow in a single rock joint. Environmental Earth Sciences 75 (9), 739.

Koopialipoor, M., Armaghani, D.J., Hedayat, A., et al., 2018. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. Soft Comput.

Lloret-Cabot, M., Wheeler, S.J., Pineda, J.A., et al., 2018. From saturated to unsaturated conditions and vice versa. Acta Geotech 13, 15–37. https://doi.org/10.1007/s11440-017-0577-6, 2018.

Lu, X., Hasanipanah, M., Brindhadevi, K., Amnieh, H.B., Khalafi, S., 2020. ORELM: a novel machine learning approach for prediction of flyrock in mine blasting. Nat. Resour. Res. 29 (2), 641–654.

Mahdiyar, A., Jahed Armaghani, D., Koopialipoor, M., Hedayat, A., Abdullah, A., Yahya, K., 2020. Practical risk assessment of ground vibrations resulting from blasting, using gene expression programming and Monte Carlo simulation techniques. Appl. Sci. 10 (2), 472.

Majidifard, H., Jahangiri, B., Rath, P., Contreras, L.U., Buttlar, W.G., Alavi, A.H., 2020. Developing a prediction model for rutting depth of asphalt mixtures using gene expression programming. Construct. Build. Mater. 267, 120543.

Manoj, K., Monjezi, M., 2013. Prediction of flyrock in open pit blasting operation using machine learning method. International Journal of Mining Science and Technology 23 (3), 313–316.

McCulloch Warren, S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.

Mendez, K.M., Broadhurst, D.I., Reinke, S.N., 2019. The application of artificial neural networks in metabolomics: a historical perspective. Metabolomics 15 (11), 142.

Mollahaşani, A., Alavi, A.H., Gandomi, A.H., 2011. Empirical modeling of plate load test moduli of soil via gene expression programming. Comput. Geotech. 38 (2), 281–286.

Monjezi, M., Bahrami, A., Varjani, A.Y., Sayadi, A.R., 2011. Prediction and controlling of flyrock in blasting operation using artificial neural network. Arabian Journal of Geosciences 4 (3–4), 421–425.

Monjezi, M., Hasanipanah, M., Khandelwal, M., 2013. Evaluation and prediction of blast-induced ground vibration at Shur River Dam, Iran, by artificial neural network. Neural Comput. Appl. 22 (7–8), 1637–1643.

Montgomery, D.C., Peck, E.A., 1992. Introduction to Linear Regression Analysis. Wiley, New York, USA.

Nguyen, H., Bui, X.N., 2019. Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. Nat. Resour. Res. 28 (3), 893–907.

Rad, H.N., Bakhshayeshi, I., Jusoh, W.A.W., Tahir, M.M., Foong, L.K., 2020. Prediction of flyrock in mine blasting: a new computational intelligence approach. Nat. Resour. Res. 29 (2), 609–623.

Rahardjo, H., Kim, Y., Satyanaga, 2019. A. Role of unsaturated soil mechanics in geotechnical engineering. Geo-Engineering 10, 8. https://doi.org/10.1186/s40703-019-0104-8, 2019.

Rajabi, A.M., Vafaee, A., 2020. Prediction of blast-induced ground vibration using empirical models and artificial neural network (Bakhtiari Dam access tunnel, as a case study). J. Vib. Contr. 26 (7–8), 520–531.

Ramesh, A., Hajihassani, M., Rashiddel, A., 2020. Ground movements' prediction in shield-driven tunnels using gene expression programming. Open Construct. Build Technol. J. 14 (1).

Rezaei, M., Monjezi, M., Varjani, A.Y., 2011. Development of a fuzzy model to predict flyrock in surface mining. Saf. Sci. 49 (2), 298–305.

Sayevand, K., Arab, H., Golzar, S.B., 2018. Development of imperialist competitive algorithm in predicting the particle size distribution after mine blasting. Eng. Comput. 34 (2), 329–338.

Shakeri, J., Shokri, B.J., Dehghani, H., 2020. Prediction of blast-induced ground vibration using gene expression programming (GEP), artificial neural networks (ANNs), and linear multivariate regression (LMR). Arch. Min. Sci. 317–355.

Shojaeian, A., Asadizadeh, M., 2020. Prediction of surface tension of the binary mixtures containing ionic liquid using heuristic approaches; an input parameters investigation. J. Mol. Liq. 298, 111976.

Shokri, B.J., Dehghani, H., Shamsi, R., 2020. Predicting silver price by applying a coupled multiple linear regression (MLR) and imperialist competitive algorithm. ICA). 1 (1), 101, 1.

Sumathi, S., Paneerselvam, S., 2010. Computational Intelligence Paradigms: Theory & Applications Using MATLAB. Taylor and Francis Group, p. 851pp.

Yang, Y., Zhang, Q., 1997. A hierarchical analysis for rock engineering using artificial neural networks. Rock Mech. Rock Eng. 30 (4), 207–222.

Yegnanarayana, B., 2009. Artificial Neural Networks. PHI Learning Pvt. Ltd, New DelhI.