Evolutionary Prediction Model for Fine-Grained Soils Compression Index Using Gene-Expression Programming

Danial Mohammadzadeh [1], Seyed-Farzan Kazemi [2,3], Amir Mosavi [4,5,6*]

[1] *Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.*
D.Mohammadzadeh.Sh@gmail.com
[2] *Department of Civil Engineering, University of Nevada, Reno, Nevada, United States of America.*
[3] *Michael Baker International, Hamilton, New Jersey, United States of America.*
seyedfarzan.kazemi@mbakerintl.com
[4] *Kalman Kando Faculty of Electrical Engineering, Obuda University, Budapest, Hungary.*
amir.mosavi@kvk.uni-obuda.hu
[5] *School of the Built Environment, Oxford Brookes University, Oxford, UK.*
[6] *Queensland University of Technology (QUT), Centre for Accident Research Road Safety-Queensland (CARRS-Q), Queensland 4059, Australia.*

Abstract

Appropriate estimation of soil settlement is of significant importance since it directly influences the performance of building and infrastructures that are built on soil. In particular, the settlement of fine-grained soils is critical because of low permeability and continuous settlement with time. Coefficient of consolidation (C_c) is a key parameter to estimate settlement of fine-grained soil layers. However, estimation of this parameter is time consuming, needs skilled technicians, and specific equipment. In this study, C_c was estimated using several soil parameters such as liquid limit (LL), plastic limit (PL), and initial void ratio (e_0). Estimating such parameters in laboratory is straight forward and needs substantially less time and cost compared to conventional tests to estimate C_c such as oedometer test. This study presents a novel prediction model for C_c of fine-grained soils using gene-expression programming (GEP). GEP is a biologically inspired technique capable of offering closed-form solution for the optimal solution. A database consisted of 108 different data points was used to develop the model. A closed-form equation solution was derived to estimate C_c based on LL, PL, and e_0. The performance of developed GEP-based model was evaluated through coefficient of determination (R^2), root mean squared error ($RMSE$), and mean average error (MAE). High R^2 and low error values indicated the descent performance of the model. Furthermore, the model was evaluated using the additional performance measures and met all the suggested criteria. Furthermore, the model had a better performance in terms of R^2, $RMSE$, and MAE compared to most of existing models. It is expected that the developed model will decrease the time and cost associate with determining C_c of fine-grained soils.

Keywords: evolutionary model, gene-expression programming (GEP), prediction, soil compression index, estimation, soil engineering, soil informatics, civil engineering, machine learning, data science, big data, soft computing, deep learning, forecasting, subject

classification codes, construction informatics, computational intelligence (CI), artificial intelligence (AI), estimation

Introduction

Soil compressibility is considered as a volume reduction under load of pore water drainage. Precise estimation of this property is critical for calculating settlement of soil layers [1]. This problem has become more critical for fine-grained soils due to their low permeability, resulting in compression index (C_c) to be the most accepted parameter to date to represent soil compressibility [2]. This parameter is often utilized for measuring the individual soil layer settlement. Different empirical equations have been particularly developed to predict C_c [3-9]. These equations were mainly developed based on traditional statistical analyses. Nevertheless, they include a number of drawbacks such as low correlation of input and output parameters [10]. Thus, it is essential to develop a comprehensive model to analyze the complex behaviour of C_c. This model should significantly eliminate the shortcomings of the previous models like practicality and low correlation between input and output parameters.

Soft computing techniques such as artificial neural networks (ANN) are widely accepted and popular along the conventional statistical methods (e.g., regression) [11-21] . These techniques were successfully applied to different geotechnical problems such as C_c prediction [7, 22-27]. However, a major limitation of common soft computing techniques is that no closed-form prediction equation is provided by them. With the introduction of artificial intelligence (AI) techniques and particularly genetic program (GP), researchers in the field of soft computing attempted to solve this issue (i.e., obtaining closed-form solution). AI includes various techniques of ANN, neuro-fuzzy

and support vector machines (SVM) with a great record of successful applications in wide range of problems [28, 29]. Considering AI, a learning mechanism is often embedded the techniques to construct the intelligent structure of the estimation model (i.e., solution of the problem). In this between, ANN is a robust artificial tool which is widely used to predict C_c [7, 22-26]. Though ANN, FIS, and other AI techniques have a good statistical performance in terms of correlation, these techniques are often known as black-box models in soft computing terms, mainly due to lack of the capability to offer close-form estimation formulas. This, however, is reported to be a drawback for AI techniques that limits their practicality [10, 28]. It worth mentioning that the runtime for most of soft computing techniques could be efficiently decreased by using parallel processing methods [30].

Genetic programming (GP) is based on individual computational programming and it is classified as a major family of soft computing techniques. GP can empower and enable the complex and highly nonlinear estimation modelling tasks [31]. While classical GP nominates only a single program, the gene expression programming (GEP) includes several genes of programming for reaching the optimal solutions [32]. Application of GEP is growing significantly compared to GP in the engineering domain mainly due to the accuracy of its predictions [28, 29]. The current study investigates the use of GEP to develop a prediction equation for C_c of fine-grained soils existing in northeast Iran. The objective of this study was developing a GEP-based prediction equation for C_c of fine-grained soil with simple tests such as Atterberg *LL* and *PL*. Since conventional consolidation tests of fine-grained soil (e.g., oedometer test) are time consuming and costly, the application of such a prediction equation will lead to substantial savings for C_c estimation in terms of cost and time.

Gene Expression Programming (GEP)

There are several variants of GP based on optimization technique used by them. Gene expression programming (GEP) is the latest variant of GP which is a powerful tool to approximate optimized solution of a problem in a closed-form format. The conventional GP generates computer models through mimicking the biological evolution of living organisms providing a tree-like form of solution, which leads to closed-form solution for the optimization problem of interest [28, 29, 31-33]. The main objective of GP is obtaining programs that connect inputs to output for each data point creating a population of programs. The population of programs like shape of branches of tree created by GP include functions and terminals which are randomly generated. The final solution of the problem is determined based on the tree-like programs.

Fundamentals of GEP was first developed by Ferreira in 2002 consisting a number of components i.e. terminal set, function set, control parameters, fitness function, and termination function [34]. GEP employs a fixed length of character strings to model the problem, unlike the conventional GP. These characters will further turn into parse trees in various sizes and shapes known as expression trees (ETs). The benefit of GEP over conventional GP is that genetic diversity is represented as genetic operators of chromosome. GEP, in fact, evolves a number of genes (sub-programs)[34] which are individual tree-like programs [10, 34]. Furthermore, GEP has a flexible multi genetic nature suitable for the construction and evolution of complex networks of genes. In the GEP framework, the genes in a chromosome may consist of two types of information stored in either tail or head of genes i.e. information to generate the overall GEP model, and the information of terminals for producing subsequent GEP models. Specific details about GEP can be found elsewhere [10, 31, 32, 34, 35].

Fig 1 presents a sample program illustration of evolving GEP. d_1, d_2, and d_3 are

the model inputs. furthermore the process evolution functions are +, -, ×, /, exponential

function (exp()), natural logarithm function (ln()), and Inv. The presented model is linear

with coefficients of c_0, c_1, and c_2 while utilizing the nonlinear terms [31, 32]. For

obtaining c_0, c_1, and c_2, a simple least square was applied to the training data. A partial

least squares method could also employ for this objective *(18, 22)*. The important GEP

parameters that need to select carefully are the tree depth and the quantity of genes.

However minimizing the tree depth generally results in shorter closed-form equations

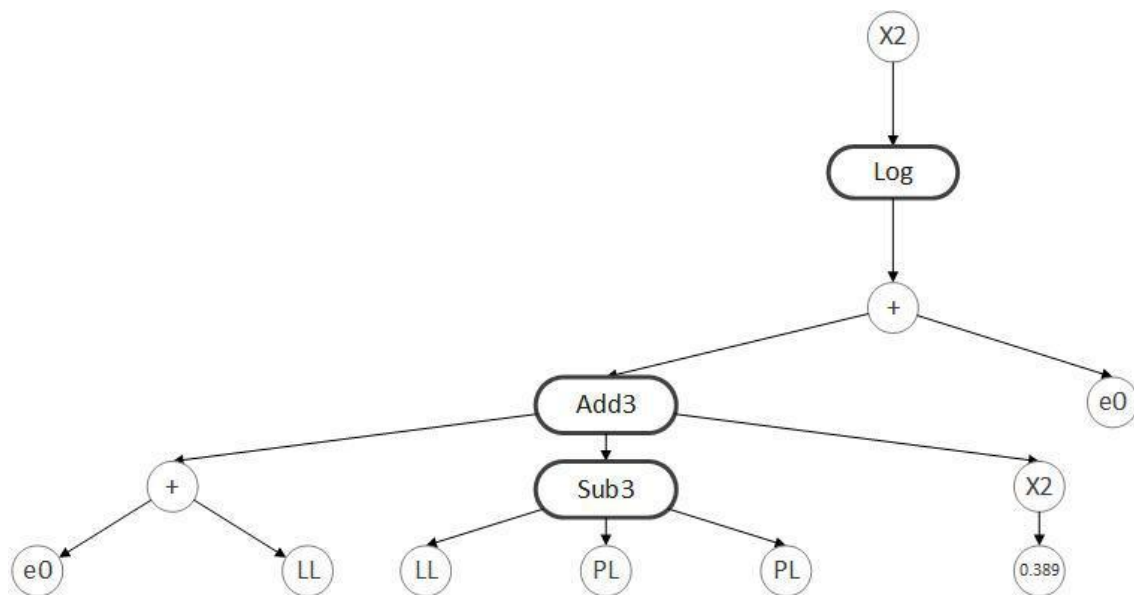with fewer number of terms [29, 34].

Fig 1. Sample GEP Model.

Modeling of C_c for Fine-Grained Soils

Data Collection

A set of 108 individual consolidation test results obtained from laboratory tests

were used to develop the GEP-based prediction equation. As mentioned earlier, the

objective of this study was to predict C_c using conventional parameters of fine-grained soils, namely *PL*, *LL*, and e_0. 101 out of 108 data points were corresponding to test results conducted on soil samples collected from different locations in Mashhad, Iran. Soil samples were classified as silty-clayey sand (SC–SM), gravelly lean clay with sand (CL) and silty clay with sand (CL–ML) based on unified soil classification system. These samples were cored from a depth of 0.5 m to 1.0 m. *LL*, *PL*, and e_0 were measured for these samples in laboratory based on ASTM D4318-17 and ASTM D854-14 [36, 37]. Furthermore, C_c was measured using oedometer test based on ASTM D2435-11 [38]. In addition, seven consolidation test results conducted by Malih was integrated into the laboratory database to make it more robust [39]. The descriptive statistics of influential input parameters (i.e., *LL*, *PL*, and e_0) and the output parameter (C_c) based on the database utilized for our study is presented in Table 1. Furthermore, Figs 2-5 illustrates the distribution of these parameters using histograms.

Table 1. Descriptive statistics for input and output parameters used in the GEP-based developed model.

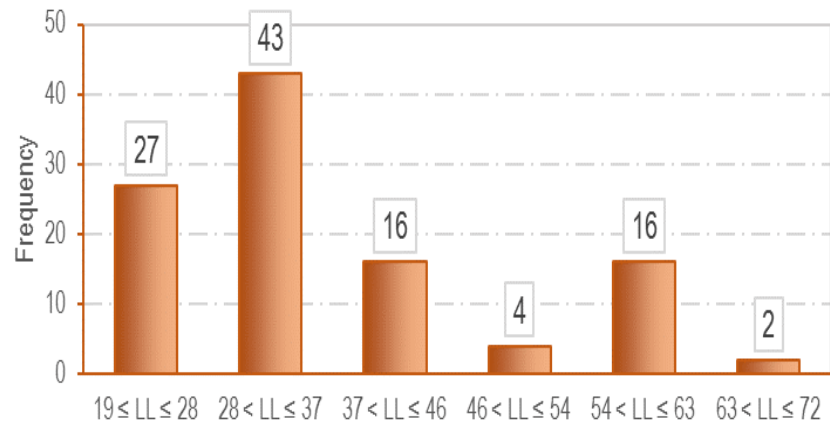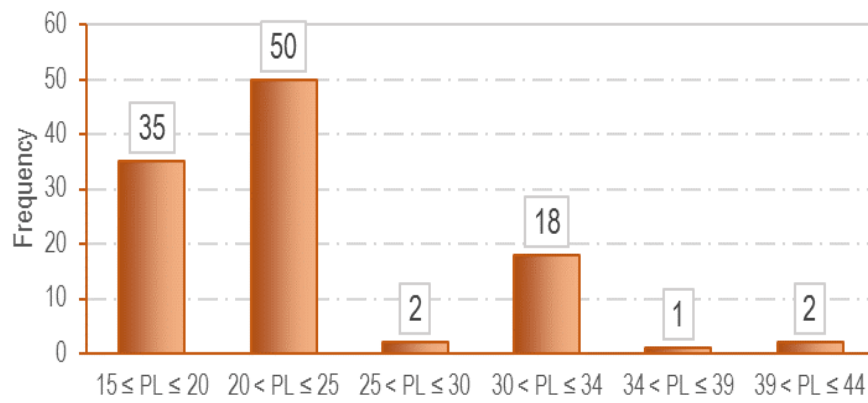| Parameter	*LL* (%)	*PL* (%)	e_0	C_c
Mean	36.16	22.61	0.75	0.17
Standard Deviation	12.79	5.64	0.12	0.05
Minimum	19.40	14.80	0.51	0.08
Maximum	72.00	44.00	1.03	0.025
Range	52.60	29.20	0.52	0.18

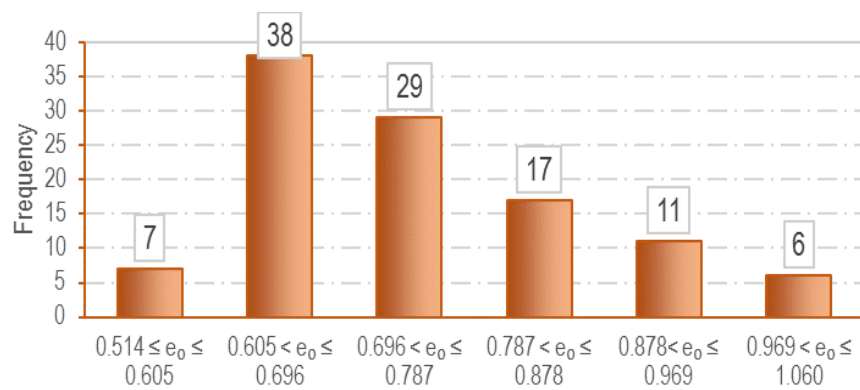Fig 2. Distribution of liquid limit (LL).
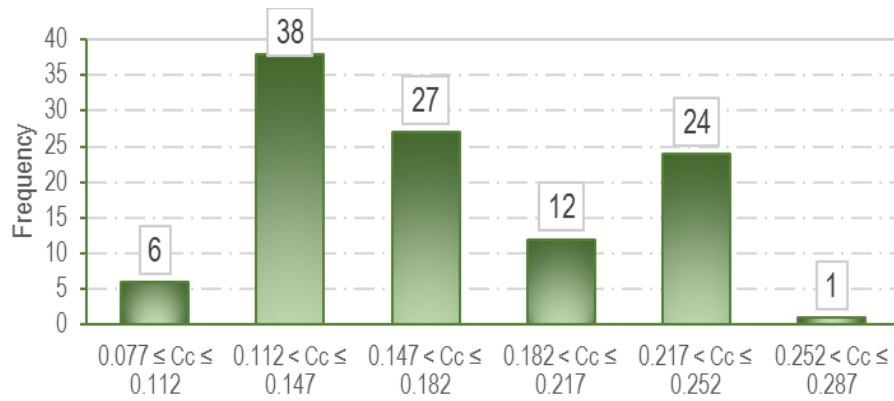

Fig 3. Distribution of PL.

Fig 4. Distribution of e_0.

Fig 5. Distribution of coefficient of consolidation (C_c).

Model Structure and Performance

Prediction equations for C_c developed by previous studies clearly indicated that *LL*, *PL*, and e_0 are three main parameters that influence C_c [3-9]. Thus, these parameters were considered in the current study to develop a simplified prediction equation for C_c. The main motivation of developing such equation was that determination of *LL*, *PL* and e_o is straight forward compared to performing any consolidation tests that directly determines C_c. Therefore, the developed model is anticipated to result in considerable savings in terms of testing time, technician cost, and laboratory equipment. It should be noted that *LL*, *PL*, and e_0 are influenced by natural water content of partially saturated soils, thus making the developed equation applicable to any saturated find-grained soils [28, 40, 41]. Mathematically, the developed equation had the following structure.

$$C_c = f\left(LL, PL, e_0\right) \tag{1}$$

showing that C_c was considered to be a function of *LL*, *PL*, and e_0. In order to develop the GEP-based prediction equation for C_c, a database containing 108 data points was developed. Each data point corresponded to *LL*, *PL*, and e_o, as well as C_c for a particular fine-grained soil sample. The GeneXproTools 5.0 was used to develop the GEP-based prediction equation in MATLAB [42]. The performance of developed GEP models was

evaluated using coefficient of determination (R^2), root mean squared error (*RMSE*), and mean average error (*MAE*) *(21-23)*, applying the following equations:

$$R^2 = \frac{\sum_{i=1}^{n}\left(h_i - \bar{h}_i\right)\left(t_i - \bar{t}_i\right)}{\sqrt{\sum_{i=1}^{n}\left(h_i - \bar{h}_i\right)^2 \cdot \sum_{i=1}^{n}\left(t_i - \bar{t}_i\right)^2}} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(h_i - t_i\right)^2}{n}} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|h_i - t_i\right| \tag{4}$$

In these equations, h_i and t_i are measured and predicted output (C_c) values, respectively, for the *i*-th data point. Furthermore, $\bar{h}_i$ and $\bar{t}_i$ are average of the measured and predicted values, respectively, and *n* is number of samples [28, 29].

Model Development

The database was divided into two subsets in order to avoid the over fitting issue: training subset and validation subset. The GEP-based model was trained using the training subset while validation subset was used for validating purposes and avoiding over fitting [34]. The final model (prediction equation) was selected based on model simplicity and performance of training and validation subsets. Performance criteria was based on highest R^2 and lowest RMSE and MAE, for training and validation subsets. After training, the candidate models were applied on un-seen validation subset to ensure their good performance. The proportion of training and validation subset size with respect to the whole data is commonly selected as 60%-75% and 25%-40%, respectively. In the current study, 75% (81 data points) and 25% (27 data points) of total data points were assigned to training subset and validation subset, respectively.

GEP algorithm was executed several times with a variety combination of influential parameters in order to identify the best model. This process was based on values suggested by previous works [31, 32, 34]. Table 2 includes the parameters of various runs. Reasonably large numbers were considered for size of population and generations to guarantee that optimal models are achieved. In the developed GEP-based model, individuals were identified and transferred into further generation based on the fitness evaluation carried out with roulette wheel sampling considering elitism. Such evaluation can guarantee successful cloning of best individual. Furthermore, the variation in the population was carried out through genetic operators on the chosen chromosomes including crossover, mutation, and rotation [10].

In every GEP-based model, values of setting parameters have significant impact on the model performance. These parameters include the quantity of genes and chromosomes, in addition to gene's head size, and rate of genetic operators. Since minor information was available about GEP parameters in the literature, appropriate settings were selected based on a trial and error scheme (see Table 2).

Table 2. Parameters used for implementation of GEP-based model.

Parameter	Setting
Number of Chromosomes	50 to 1000
Number of Genes	3
Head Size	8
Tail Size	17
Dc Size	17
Gene Size	42
Gene Recombination Rate	0.277
Gene Transportation Rate	0.277
Function Set	+, -, ×, /, *exp*, *ln*, and *Inv*

Furthermore, to facilitate developing GEP-based model, the following closed-form equation has been developed and utilized:

$$C_c = e_0 + \left[\frac{e_0 + 2LL}{e_0 - 6.87}\right] \times \left[-0.35 + LL^2\right] + [\log(2e_0 + 2LL - 2PL + 0.15)]^2 \qquad (5)$$

Figs 6-8 presents the measured values of C_c obtained from laboratory experiments versus predicted values. Figs 6, 7, and 8 present measured values versus predicted values for training subset, validation subset, and entire set, respectively. Furthermore, Table 3 summarizes GEP-based model performance in terms of R^2, *RMSE*, and *MAE* for these sets. Smith states that for a coefficient of determination of |R|> 0.8, a strong correlation exist between measured and predicted values [43]. Based on Table 3, the developed GEP-based model has a high R^2 for training subset, validation subset, and entire data set. In addition, the model exhibited a relatively low *RMSE* and *MAE* for all these sets.
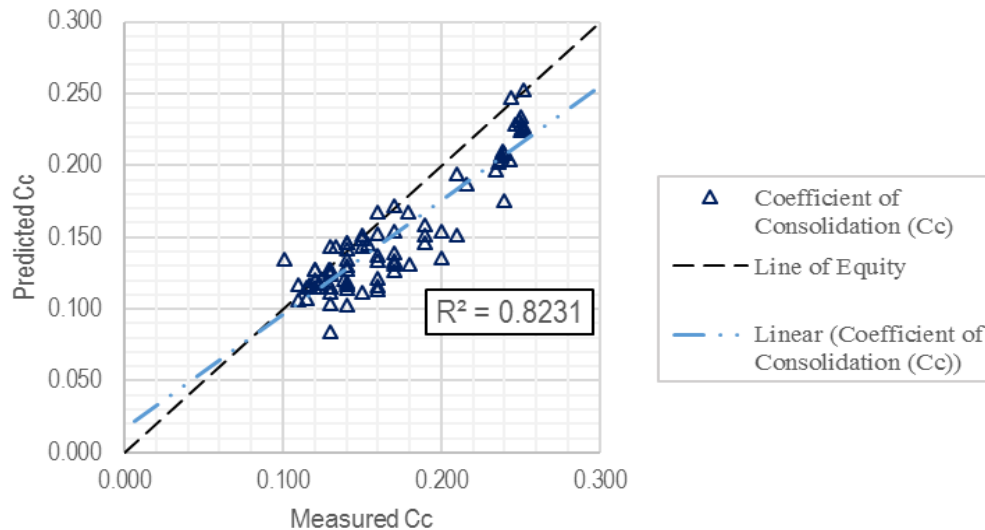

Fig 6. Predicted versus measured C_c for training subset.

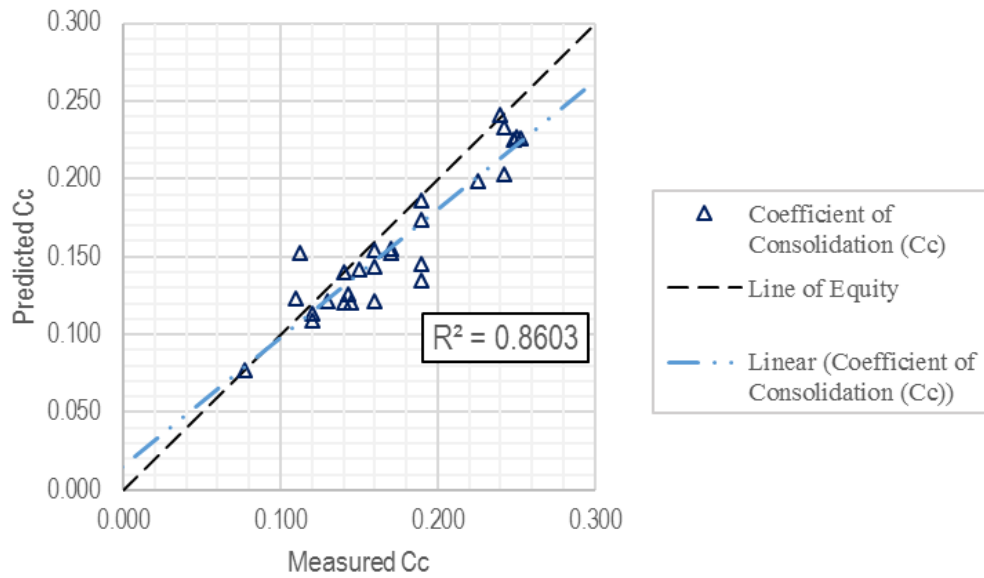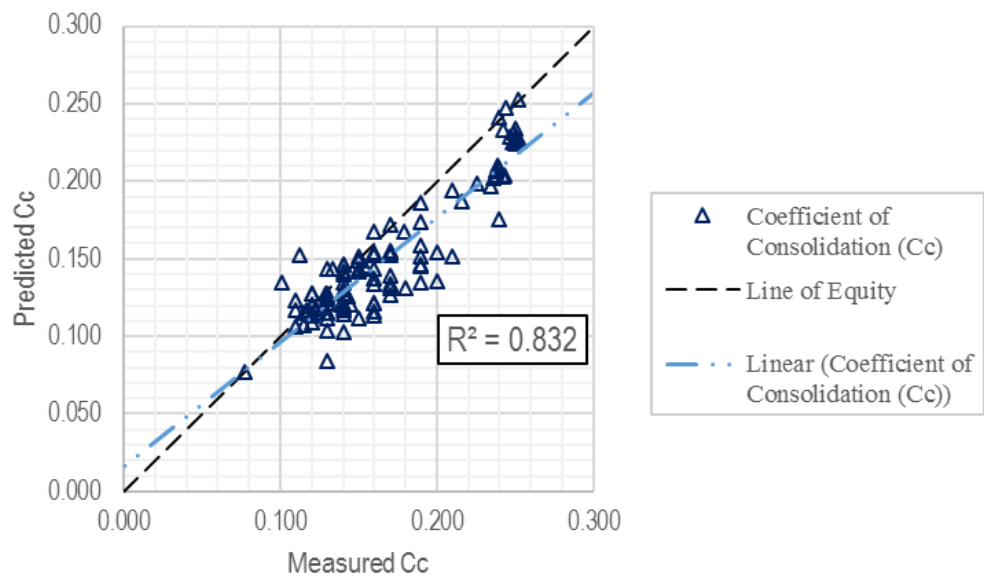Fig 7. Predicted versus measured C_c for validation subset.

Fig 8. Predicted versus measured C_c for entire data set (training + validation).

Table 3. Model performance.

Set	Number of Data Points	R^2	RMSE	MAE
Training Subset	81	0.8231	0.0269	0.0213
Validation Subset	27	0.8603	0.0237	0.0189
Entire Data Set	108	0.8320	0.0262	0.0207

Additional evaluation of model performance

In this section, performance of the developed GEP-based model is evaluated based on various statistical parameters found in the literature. These statistical parameters along with their acceptance criteria is presented in Table 4. Parameters used in this table are all as previously defined. Furthermore, the developed model was evaluated based on these statistical parameters and results are presented in this tale. As can be seen in Table 4, the developed model met all the criteria for additional statistical parameters revealing the descent performance of the developed model.

Table 4. Evaluating the developed GEP-Based model using additional statistical parameters.

Statistical Parameter	Source	Criteria	Evaluation for GEP-based Model
$k = \dfrac{\sum_{i=1}^{n}\left(h_i \times t_i\right)}{h_i^2}$	Golbraikh and Tropsha [44]	$0.85 < k < 1.15$	1.001
$k' = \dfrac{\sum_{i=1}^{n}\left(h_i \times t_i\right)}{t_i^2}$	Roy and Roy [45]	$0.85 < k' < 1.15$	0.989
$R_m = R^2 \times \left(1 - \sqrt{R^2 - Ro^2}\right)$	Roy and Roy [45]	$0.5 < R_m$	0.503
$Ro^2 = 1 - \dfrac{\sum_{i=1}^{n}\left(t_i - h_i^o\right)^2}{\sum_{i=1}^{n}\left(t_i - \bar{t}_i\right)^2}, h_i^o = k \times t_i$	Roy and Roy [45]	Should be close to 1.0.	1.000
$Ro'^2 = 1 - \dfrac{\sum_{i=1}^{n}\left(t_i - t_i^o\right)^2}{\sum_{i=1}^{n}\left(h_i - \bar{h}_i\right)^2}$	Roy and Roy [45]	Should be close to 1.0.	0.998

Table 5 presents the comparison of the developed GEP-based model with previous models found in the literature. The previous model consisted of either regression-based equations or robust AI methods such as multi-expression programming (MEP), artificial

neural network (ANN), and multi-gene genetic programming (MGGP). It worth mentioning that these AI methods do not provide any closed-form solution. AI methods had relatively high R^2 mainly due to their black-box nature of connecting inputs and outputs. Nevertheless, the developed GEP-based model had a higher R^2 compared to existing AI methods. However, MEP, ANN, and MGGP had a lower error in terms of *RMSE* and *MAE*.

Based on Table 5, the developed GEP-based model outperforms regression models since regression models considers only a small quantity of base functions. Therefore, such models cannot be used to complex interactions of soil parameters (i.e., *LL*, *PL*, and e_0) and C_c. However, the developed GEP-based model considers a variety of base functions and their combination in order to achieve a closed-form equation with high performance. The developed GEP-based model directly considers the experimental data with no prior assumptions. In other words, contrary to traditional regression models, GEP does not assume any pre-defined shape for the solution equation. High values of R^2 presented in Table 5 indicates that the developed GEP-based model was very successful in fitting the measured C_c to the input parameters of *LL*, *PL*, and e_0.

Table 5. Performance comparison of current developed GEP-based model with existing models.

Source	Model Description	Performance Measure		
		R^2	RMSE	MAE
Skempton [8]	Regression Equation	0.367	0.072	0.056
Nishida [6]	Regression Equation	0.752	0.301	0.285
Cozzolino [4]	Regression Equation	0.752	0.105	0.103
Terzaghi and Peck [9]	Regression Equation	0.367	0.110	0.077
Azzouz et al. [3]	Regression Equation	0.752	0.036	0.032
Mayhe [5]	Regression Equation	0.367	0.102	0.073
Park and Lee [7]	ANN	0.752	0.089	0.085
Mohammadzade et al. [28]	MEP	0.811	0.019	0.016
Mohammadzade et al. [29]	ANN	0.859	0.017	0.014
Current Study	**GEP**	**0.832**	**0.026**	**0.021**

Summary and conclusion

Coefficient of consolidation (C_c) is a significant parameter for determining settlement of fine-grained soil layers subjected to loads such as that of buildings, vehicles, and infrastructures. If C_c is not estimated accurately, soil settlement is under- or over-estimated, which both cases are not desirable from technical prospective. Thus, determining C_c is of significance importance in settlement calculations. However, measuring C_c using the traditional oedometer test method is time-consuming, needs skilled technicians, and requires special laboratory equipment. Therefore, estimation of C_c using other parameters of fine-grained soils such as *LL*, *PL*, and e_0 would eliminate the time and costliness associated with oedometer test. In this study, gene-expression programming (GEP) was employed to develop a model for estimating C_c using *LL*, *PL*, and e_0. 108 data points containing C_c, *LL*, *PL*, and e_0 were used to train and validate the model. The model was developed based on tuned calibration parameters using trial and error. A closed-form solution was derived from the developed GEP-based model which is anticipated to aid geotechnical researchers in determining C_c with considerable saving in associated time and costs. This closed-form equation for predicting C_c was employed to develop surface charts to predict C_c based on *LL* and *PL* for a certain e_0.

The performance of developed GEP-based model was evaluated using coefficient of determination (R^2) and two error measures, namely root mean square error (*RMSE*) and mean average error (*MAE*). R^2 values were 0.8231, 0.8603, and 0.8320 for training subset, validation subset, and entire data set, respectively. In addition, *RMSE* was 0.0269, 0.0237, and 0.0262 for training subset, validation subset, and entire data set, respectively. High R^2 and low error indicates the highly acceptable performance of the GEP-based model. Additional performance measures found in the literature were employed to further

evaluate the performance of developed GEP-based model. This evaluation revealed that the model has a descent performance based on additional performance measures.

　　　　Contrary to the classical models for estimating C_c such as regression models, the developed GEP-based model reveals a highly nonlinear behavior and includes complex combination of influential input parameters (i.e., LL, PL, and e_0,). In general, C_c was positively correlated with e_0. Furthermore, LL and e_0 had a higher influence on estimation of C_c compared to PL. Comparison of the developed model with previous models in the literature revealed its good performance, which guarantees the use of GEP-based model in practical applications.

References

1.　　Tiwari, B. and B. Ajmera, *New correlation equations for compression index of remolded clays.* Journal of Geotechnical and Geoenvironmental Engineering, 2011. **138**(6): p. 757-762.
2.　　Carter, M. and S.P. Bentley, *Correlations of soil properties*. 1991: Pentech press publishers.
3.　　Azzouz, A.S., R.J. KRIZEK, and R.B. CoRoTis, *Regression analysis of soil compressibility.* Soils and Foundations, 1976. **16**(2): p. 19-29.
4.　　Cozzolino, V. *Statistical forecasting of compression index*. in *Proceedings of the fifth international conference on soil mechanics and foundation engineering, Paris*. 1961.
5.　　Mayhe, P., *CAM-CLAYS PREDICTIONS OF UNDRAINED STRENGTH*. Journal of Geotechnical and Geoenvironmental Engineering, 1980. **106**(ASCE 15816).
6.　　Nishida, Y., *A brief note on compression index of soil.* Journal of the Soil Mechanics and Foundations Division, 1956. **82**(3): p. 1-14.
7.　　Park, H.I. and S.R. Lee, *Evaluation of the compression index of soils using an artificial neural network.* Computers and Geotechnics, 2011. **38**(4): p. 472-481.
8.　　Skempton, A.W. and O. Jones, *Notes on the compressibility of clays.* Quarterly Journal of the Geological Society, 1944. **100**(1-4): p. 119-135.
9.　　Terzaghi, K., R.B. Peck, and G. Mesri, *Soil mechanics in engineering practice.* 1996: John Wiley & Sons.
10.　　Hossein Alavi, A. and A. Hossein Gandomi, *A robust data mining approach for formulation of geotechnical engineering systems.* Engineering Computations, 2011. **28**(3): p. 242-274.
11.　　Choubin, B., et al., *An Ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines.* Science of the Total Environment, 2019. **651**: p. 2087-2096.

12.   Rezakazemi, M., A. Mosavi, and S. Shirazian, *ANFIS pattern for molecular membranes separation optimization.* Journal of Molecular Liquids, 2019. **274**: p. 470-476.

13.   Mosavi, A., P. Ozturk, and K.-w. Chau, *Flood prediction using machine learning models: Literature review.* Water, 2018. **10**(11): p. 1536.

14.   Mosavi, A., A. Lopez, and A.R. Varkonyi-Koczy. *Industrial applications of big data: state of the art survey*. in *International Conference on Global Research and Education*. 2017. Springer.

15.   Mosavi, A., Y. Bathla, and A. Varkonyi-Koczy. *Predicting the future using web knowledge: state of the art survey*. in *International conference on global research and education*. 2017. Springer.

16.   Vargas, R., A. Mosavi, and R. Ruiz, *Deep learning: a review.* 2018.

17.   Mosavi, A. and T. Rabczuk. *Learning and intelligent optimization for material design innovation*. in *International Conference on Learning and Intelligent Optimization*. 2017. Springer.

18.   Mosavi, A., T. Rabczuk, and A.R. Varkonyi-Koczy. *Reviewing the novel machine learning tools for materials design*. in *International Conference on Global Research and Education*. 2017. Springer.

19.   Mosavi, A. and A.R. Varkonyi-Koczy, *Integration of machine learning and optimization for robot learning*, in *Recent Global Research and Education: Technological Challenges*. 2017, Springer. p. 349-355.

20.   Mosavi, A. and M. Edalatifar. *A Hybrid Neuro-Fuzzy Algorithm for Prediction of Reference Evapotranspiration*. in *International Conference on Global Research and Education*. 2018. Springer.

21.   Torabi, M., et al. *A hybrid machine learning approach for daily prediction of solar radiation*. in *International Conference on Global Research and Education*. 2018. Springer.

22.   Das, S.K. and P.K. Basudhar, *Prediction of residual friction angle of clays using artificial neural network.* Engineering Geology, 2008. **100**(3): p. 142-145.

23.   Das, S.K., et al., *Classification of slopes and prediction of factor of safety using differential evolution neural networks.* Environmental Earth Sciences, 2011. **64**(1): p. 201-210.

24.   Daryaei, M., et al., *Modeling the compression index of fine soils using artificial neural network and comparison with the other empirical equations.* 2010.

25.   Farkhonde, S. and J. Bolouri. *Estimation of compression index of clayey soils using artificial neural network*. in *Fifth national conference on civil engineering, Mashhad, Iran, Paper ID*. 2010.

26.   Kumar, V.P. and C.S. Rani, *Prediction of compression index of soils using artificial neural networks (ANNs).* Int J Eng Res Appl, 2011. **1**(4): p. 1554-1558.

27.   Talaei-Khoei, A., J.M. Wilson, and S.-F. Kazemi, *Period of Measurement in Time-Series Predictions of Disease Counts from 2007 to 2017 in Northern Nevada: Analytics Experiment.* JMIR public health and surveillance, 2019. **5**(1): p. e11357.

28.   Mohammadzadeh, D., et al., *Deriving an intelligent model for soil compression index utilizing multi-gene genetic programming.* Environmental Earth Sciences, 2016. **75**(3): p. 262.

29.   Mohammadzadeh, D., J.B. Bazaz, and A.H. Alavi, *An evolutionary computational approach for formulation of compression index of fine-grained soils.* Engineering Applications of Artificial Intelligence, 2014. **33**: p. 58-68.

30.   Kazemi, S.F. and Y. Shafahi. *An Integrated Model Of Parallel Processing And PSO Algorithm For Solving Optimum Highway Alignment Problem*. in *ECMS*. 2013.

31.   Ferreira, C., *Gene expression programming: mathematical modeling by an artificial intelligence*. Vol. 21. 2006: Springer.

32.   Ferreira, C., *Gene expression programming and the evolution of computer programs.* Recent developments in biologically inspired computing, 2004: p. 82-103.

33.   Koza, J.R., *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. 1992: MIT press.

34.   Ferreira, C., *Gene expression programming in problem solving*, in *Soft computing and industry*. 2002, Springer. p. 635-653.

35.   Batioja-Alvarez, D.D., et al., *Probabilistic Mechanistic-Based Pavement Damage Costs for Multitrip Overweight Vehicles.* Journal of Transportation Engineering, Part B: Pavements, 2018. **144**(2): p. 04018004.

36.   *Standard Test Methods for Liquid Limit, Plastic Limit, and Plasticity Index of Soils*. 2017, ASTM International.

37.   ASTM, *Standard Test Method for Bulk Specific Gravity and Density of Non-Absorptive Compacted Asphalt Mixtures*. 2017, ASTM International.

38.   *Standard Test Methods for One-Dimensional Consolidation Properties of Soils Using Incremental Loading*. 2011, ASTM International.

39.   Malih, S.K., *Fuzzy Model for Predicting Fine-Grained Soil Consolidation Test Parameters*. 2006, University of Guilan, Iran.

40.   Gandomi, A.H., et al., *Linear genetic programming for shear strength prediction of reinforced concrete beams without stirrups.* Applied Soft Computing, 2014. **19**: p. 112-120.

41.   Ziaee, S.A., et al., *Explicit formulation of bearing capacity of shallow foundations on rock masses using artificial neural networks: application and supplementary studies.* Environmental earth sciences, 2015. **73**(7): p. 3417-3431.

42.   GEPSOFT, G., *Version 5.0*. 2014.

43.   Smith, G.N., *Probability and statistics in civil engineering.* Collins Professional and Technical Books, 1986. **244**.

44.   Golbraikh, A. and A. Tropsha, *Beware of q 2!* Journal of molecular graphics and modelling, 2002. **20**(4): p. 269-276.

45.   Roy, P.P. and K. Roy, *On some aspects of variable selection for partial least squares regression models.* Molecular Informatics, 2008. **27**(3): p. 302-313.