



1st International Conference on Advances in Civil & Environmental Engineering, University of Engineering & Technology Taxila, Pakistan

Conference date 22 & 23 Feb 2022

Prediction of Soil Compaction Parameters using Gene Expression Programming

Hasnain Ayub, Turab Jafri*

NUST Institute of Civil Engineering,
National University of Sciences and Technology, Islamabad, Pakistan.

Mkhan.ms19nice@student.nust.edu.pk; turabjafri@nice.nust.edu.pk

Syed Taseer Abbas Jaffar

Department of Civil Engineering,
University of Management & Technology, Lahore, Pakistan.

taseer.jaffar@umt.edu.pk

ABSTRACT

The determination of soil compaction parameters, maximum dry density (MDD) and optimum moisture content (OMC), in the laboratory requires considerable time and energy. Various efforts have been made to counter the issue by developing models to predict compaction parameters of soil, but they are either applicable to specific soil types or specific compaction energies. A machine learning technique, Gene Expression Programming (GEP), is used in this study to predict the compaction parameters of soil using the index properties of soils collected from literature. In terms of accuracy, the GEP models presented in this study outperformed the previously developed prediction models. The accuracy (R-squared) for training and validation of MDD was found to be 0.8945 and 0.9032, respectively, and that for training and validation of OMC was 0.9165 and 0.9251, respectively.

KEYWORDS: Soil Compaction, Maximum Dry Density, Optimum Moisture Content, Gene Expression Programming, Prediction Models

1 INTRODUCTION

Soil compaction is a dynamic process in earthwork construction in which the soil is compacted through mechanical process. The compaction process reduces the air voids by bringing the soil particle close to each other thus increasing the shear strength and reducing the permeability of soils. The compaction parameters of soils, maximum dry density and optimum moisture content, are determined in the laboratory following standard proctor compression test [1] and modified proctor compression test [2] which require significant amount of time and energy. Thus, various models have been developed in the past to predict the compaction parameters using conventional regression analysis [3],[4],[5],[6],[7],[8],[9],[10],[11]. The accuracy of simple regression analysis techniques reduces as the number of datasets increase. Therefore, machine learning approaches have been adopted to deal with large datasets. So far, the best prediction model developed for compaction parameters found in literature is by using multi expression programming (MEP) for a



total of 226 datasets covering both the fine-grained soil and coarse-grained soils compacted against different compaction energies [12]. Using Gene Expression Programming, this research aims at developing new prediction models with the prediction accuracy better than the conventional regression models and MEP models in the literature.

2 METHODOLOGY

For the development of the new prediction models, the data was used from the literature [12]. The dataset contained 226 records of the index and compaction properties of fine-grained and coarse-grained soils. The statistic of the index and compaction properties of soils is given in the Table 1.

Table 1 Statistics of the soil properties

| Parameters | Maximum | Minimum | Standard deviation | Mean |
|-------------------------------------|---------|---------|--------------------|-------|
| Fine content (CF), % | 100 | 8.6 | 29.9 | 63.1 |
| Sand content (CS), % | 89 | 0 | 23.3 | 29.5 |
| Gravel content (CG), % | 67.1 | 0 | 14.5 | 7.5 |
| Plastic Limit (PL), % | 48.3 | 6.1 | 7.4 | 22 |
| Liquid Limit (LL), % | 608 | 16 | 163.9 | 108.7 |
| Compaction Energy (E), kJ/m^3 | 2755 | 154.5 | 733.9 | 893.8 |
| Maximum Dry Density (MDD), Mg/m^3 | 2.33 | 1.09 | 0.2 | 1.75 |
| Optimum Moisture Content (OMC), % | 43.7 | 5.3 | 6 | 17.5 |

Gene Expression Programming (GEP) was used for the development of new prediction models. GEP is an extension of Genetic programming (GP) [13]. GEP creates computer models in form of complex trees and these trees learn and adapt while changing their structure and composition. The superiority of GEP, a genetic programming-based approach, over other soft computing methods is that it produces mathematical function for the prediction, and it does not assume any prior form of existing relationship [14]. There are five components of GEP that are: the terminal set, the function set, control parameters, fitness function, and termination conditions. GEP is a genotype-phenotype evolutionary algorithm having linear chromosomes of fixed length and the computer program of GEP are encoded in these chromosomes to solve a problem. Genes are sub-programs that are created in GEP. Gene in a GEP chromosome consists of a head and a tail. The head of a gene may consist of a function or terminal symbol, but the tail consists of only terminal symbols. Expression trees (ET's) are generated as a result of these chromosomes [13]. These trees vary in size and shape. Mathematical functions are later derived from these expression trees. A typical expression tree of GEP is shown in Figure 1. In this tree D_0 , D_1 and D_3 are the inputs while C_5 and C_6 are constants. The functions defined for the model are addition (+), subtraction (-), multiplication (\times), and square root (sqrt). GEP is a multi-genetic programming and the genes in a chromosomes are connected via the functions defined by the user, thus creating expression trees, the complexity of which is governed by the number of genes [15]. The equation derived from the expression tree is given in Eq. (1), where Y is the output.

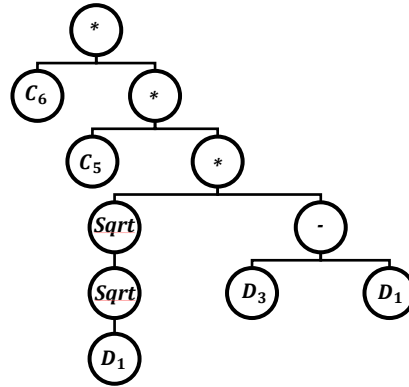


Figure 1: Expression tree (ET)

$$Y = C_6 \times C_5 [\sqrt{\sqrt{D_1} \times (D_3 - D_0)}] \quad (1)$$

To develop GEP prediction models for compaction parameters, GeneXproTool 5.0 was used [16]. The dataset was randomly divided into training (70%) and validation (30%). The optimal parametric setting to develop the GEP algorithm was established on the basis of hit and trial procedure and is provided in Table 2.

Table 2: Optimal parametric setting for GEP algorithm

| GEP parameters | GEP settings for MDD and OMC | GEP parameters | GEP settings for MDD and OMC |
|----------------------------|---------------------------------|------------------------------|------------------------------|
| General | | Genetic operators | |
| Training records | 158 | Mutation | 0.00138 |
| Validation/Testing records | 68 | Inversion rate | 0.00138 |
| Number of chromosomes | 100 | IS transposition rate | 0.00138 |
| Head size | 8 | RIS transposition rate | 0.00138 |
| Number of genes | 5 | One-point recombination rate | 0.00277 |
| Linking function | Addition | Two-point recombination rate | 0.00277 |
| Function set | +, -, ÷, ×, exp, x ² | Gene recombination rate | 0.00277 |
| Numerical constants | | Uniform recombination | 0.00755 |
| Constants per gene | 10 | | |
| Data type | Floating point | | |
| Ephemeral random constant | [-10,10] | | |

It is worth noting that Wang and Yin, 2020 [12] did not consider gravel content (CG) as an input for developing MEP model while CG is included as an input in this research.

3 RESULTS AND DISCUSSION

The expression trees ETs were developed as a result of the modelling. The expression trees were then translated to mathematical functions. The mathematical functions for ETs were then linked together via linking function. The final mathematical equations for MDD and OMC developed using GEP are given in Eq. (2) and (3), respectively.



$$\begin{aligned} \text{MDD (Mg/m}^3\text{)} = & \text{CF} + 91.84 + \frac{-\text{LL}}{e^{11.69}} + (\text{CS}^2 + e^{5.46}) - 7.244 + \frac{0.146}{\text{PL}} + \text{CS} + \\ & \text{CG} - (e^{-8.953} \times \text{PL} \times \text{CF}) + \frac{-0.627 \times \text{CS} \times \text{CF}}{\text{E}} - \text{LL} - 7.599 + \text{LL} - \frac{\text{CG}}{-5.5(\text{CF})^2 + 5.07(\text{CG})} \end{aligned} \quad (2)$$

$$\begin{aligned} \text{OMC (\%)} = & \frac{-34.49(\text{PL} + \text{CS})}{-525 - \text{E}} - \frac{\text{PL}}{3.57(\text{CS} + 1.43)} + \frac{(7.956 \times \text{CG} + \text{CF})(\text{CF})(\text{LL})}{2.06 \times \text{E}(\text{E} + \text{CG})} + \frac{\text{CS}}{(\text{CG} + 3.235)(\text{CF})} + \\ & \text{CF} + 8.4 + \frac{\text{CF}}{\text{PL}} - \frac{\text{LL}}{240.6} - \text{CF} - 4.25 + \frac{\text{CF}(\text{PL})}{\text{CF} + \text{PL} + 38.43} \end{aligned} \quad (3)$$

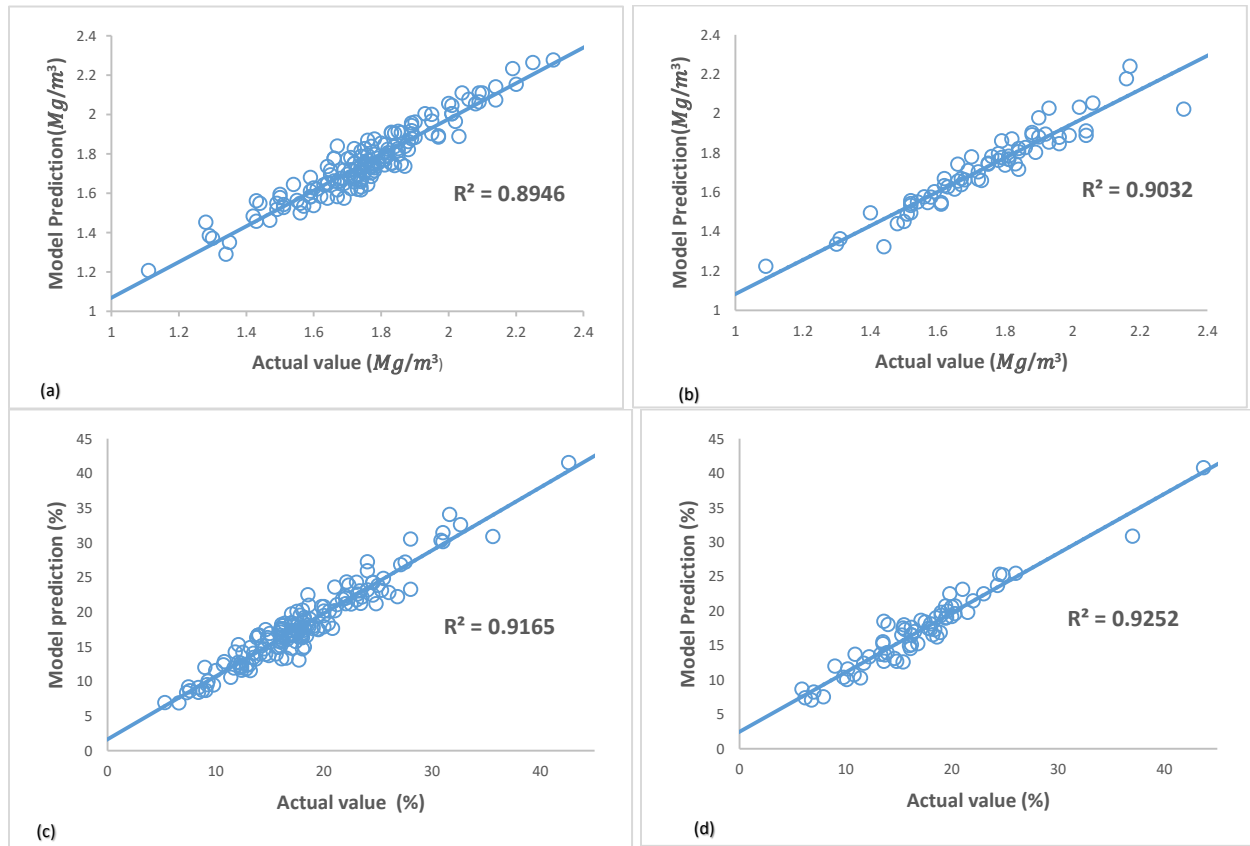


Figure 2: GEP model prediction for output vs actual output for (a) MDD Training, (b) MDD Validation, (c) OMC Training and (d) OMC Validation.

The GEP models performed better for MDD in terms of R^2 (0.8946 for training and 0.9032 for validation) while for OMC, the prediction accuracy (R^2) was almost the same as that of MEP model (0.9156 for training and 0.9252 for validation) as shown in Figure 2.

4 CONCLUSION

In this study, a comprehensive dataset of soil was used to develop prediction models for soil compaction parameters using Gene Expression Programming. Out of the 226 records, 158 were



used for the training of the model while 68 records were used for the validation of the trained models. GEP models were derived for MDD, and OMC and the accuracy of the newly developed models was evaluated. It is concluded that the presented GEP models are more accurate than both the conventional regression models and the MEP model in the literature.

5 REFERENCES

- [1] A. C. D.-18 on S. and Rock, *Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort (12 400 Ft-lbf/ft³ (600 KN-m/m³))* 1. ASTM international, 2007.
- [2] D. Astm, "Standard Test Method for Laboratory Compaction Characteristics of Soil Using Modified Effort (D1557-12)," *ASTM, West Conshohocken, Pennsylvania, USA*, 2012.
- [3] A. N. Al-Khafaji, "Estimation of soil compaction parameters by means of Atterberg limits," *Q. J. Eng. Geol. Hydrogeol.*, vol. 26, no. 4, pp. 359–368, 1993.
- [4] L. R. Blotz, C. H. Benson, and G. P. Boutwell, "Estimating optimum water content and maximum dry unit weight for compacted clays," *J. Geotech. Geoenvironmental Eng.*, vol. 124, no. 9, pp. 907–912, 1998.
- [5] M. Omar, A. Shanableh, A. Basma, and S. Barakat, "Compaction characteristics of granular soils in United Arab Emirates," *Geotech. Geol. Eng.*, vol. 21, no. 3, pp. 283–295, 2003.
- [6] A. Sridharan and H. B. Nagaraj, "Plastic limit and compaction characteristics of finegrained soils," *Proc. Inst. Civ. Eng. Improv.*, vol. 9, no. 1, pp. 17–22, 2005.
- [7] O. Günaydın, "Estimation of soil compaction parameters by using statistical analyses and artificial neural networks," *Environ. Geol.*, vol. 57, no. 1, p. 203, 2009.
- [8] L. Di Matteo, F. Bigotti, and R. Ricco, "Best-fit models to estimate modified proctor properties of compacted soil," *J. Geotech. geoenvironmental Eng.*, vol. 135, no. 7, pp. 992–996, 2009.
- [9] C. Patra, N. Sivakugan, and B. Das, "Relative density and median grain-size correlation from laboratory compaction tests on granular soil," *Int. J. Geotech. Eng.*, vol. 4, no. 1, pp. 55–62, 2010.
- [10] H. Mujtaba, K. Farooq, N. Sivakugan, and B. M. Das, "Correlation between gradational parameters and compaction characteristics of sandy soils," *Int. J. Geotech. Eng.*, vol. 7, no. 4, pp. 395–401, 2013.
- [11] A. Saikia, D. Baruah, K. Das, H. J. Rabha, A. Dutta, and A. Saharia, "Predicting compaction characteristics of fine-grained soils in terms of Atterberg limits," *Int. J. Geosynth. Gr. Eng.*, vol. 3, no. 2, pp. 1–9, 2017.
- [12] H.-L. Wang and Z.-Y. Yin, "High performance prediction of soil compaction parameters using multi expression programming," *Eng. Geol.*, vol. 276, p. 105758, 2020.
- [13] C. Ferreira, "Gene expression programming in problem solving," in *Soft computing and industry*, Springer, 2002, pp. 635–653.
- [14] A. Johari, G. Habibagahi, and A. Ghahramani, "Prediction of soil–water characteristic curve using genetic programming," *J. Geotech. Geoenvironmental Eng.*, vol. 132, no. 5, pp. 661–665, 2006.
- [15] Z.-L. Cheng, W.-H. Zhou, and A. Garg, "Genetic programming model for estimating soil suction in shallow soil layers in the vicinity of a tree," *Eng. Geol.*, vol. 268, p. 105506, 2020.
- [16] C. Ferreira, "Gene expression programming: a new adaptive algorithm for solving problems," *arXiv Prepr. cs/0102027*, 2001.

DECLARATION

The authors declare that there is no conflict of interest.