Estimation of matric suction in compacted soils using machine learning techniques

Junjie Wang[1], Sai Vanapalli[*,1]

[1]*Department of Civil Engineering, University of Ottawa, Ottawa, Canada*
[*]*Corresponding author's email:* sai.vanapalli@uottawa.ca

Abstract: The hydro-mechanical behaviour of fine-grained compacted soils is significantly influenced by matric suction. However, reliable measurements and continuous monitoring of matric suction are both costly and time-consuming. Due to this reason, there are limitations in implementing the state-of-the-art understanding of the mechanics of unsaturated soils into practice. To address this challenge, in this study, multi-gene genetic programming (MGGP), a powerful machine learning (ML) technique, is employed to develop a model for estimating matric suction in compacted fine-grained soils using an extensive database. The proposed model is capable of reliably estimating matric suction using simple soil properties as inputs, accounting for variations in soil structure induced by initial compaction conditions and changes in water content caused by seasonal or environmental fluctuations. Additionally, a simple equation is derived based on the MGGP model that provides a programming-free method for manual or spreadsheet-based calculation of matric suction using the information of basic soil properties. The proposed approach is valuable for practicing engineers for rationally interpreting and predicting the performance of geotechnical infrastructures constructed with or within fine-grained unsaturated soils.

Introduction

Compacted fine-grained soils are widely used because they offer favorable hydraulic and mechanical properties for the construction of geotechnical infrastructure. Fig. 1 shows several geo-infrastructures that include the embankments, foundations and subgrades of roadways and railways that are constructed with or within compacted soils. In most scenarios, these geo-infrastructures are above the natural groundwater table (GWT), a zone in which soils are typically in an unsaturated state. The hydro-mechanical behavior of unsaturated compacted soils is sensitive to variations in matric suction associated with water content changes [1-4].

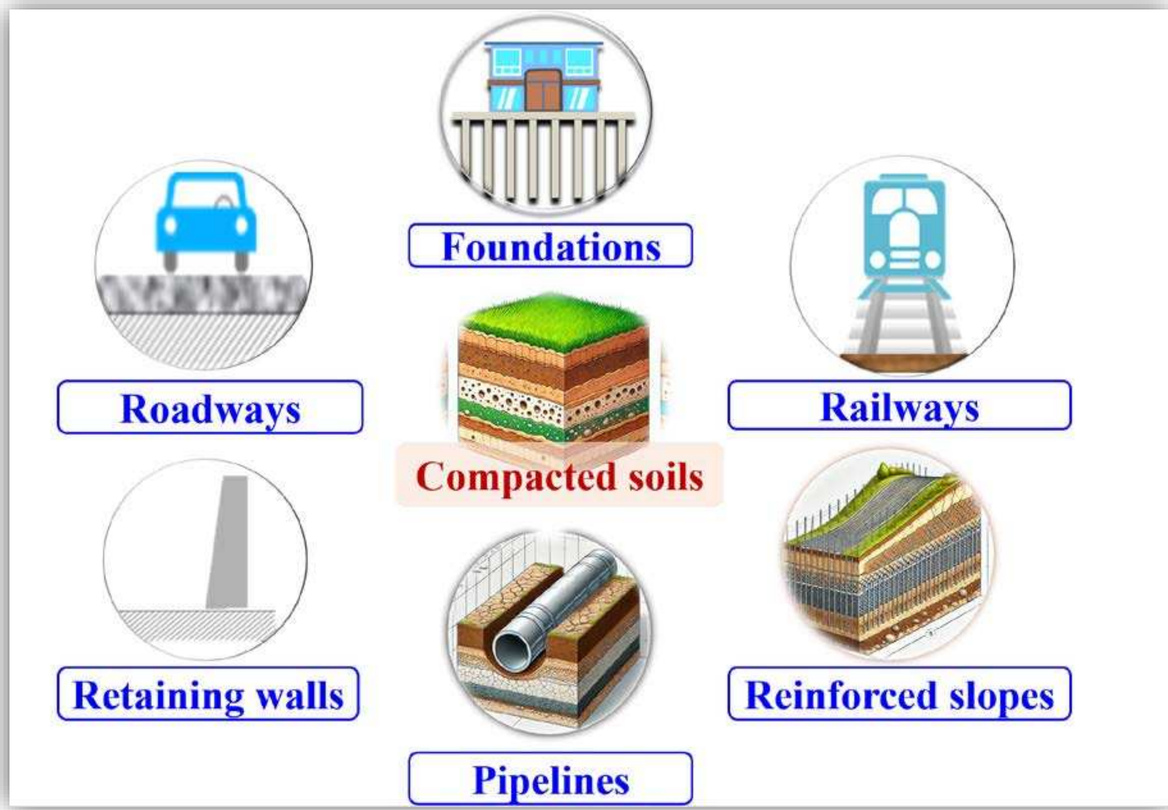


Figure 1: Typical geo-infrastructures constructed with or within compacted soils.

Matric suction, $(u_a - u_w)$ and net normal stress, $(\sigma_n - u_a)$ are two key stress state variables that are required for rational interpretation of the hydro-mechanical behavior of unsaturated soils [5]. While net normal stress can be readily calculated using information on the density and thickness of soils in the field, matric suction determination remains a complex task. To accomplish this task, expensive equipment is often required, which involves time-consuming procedures using the expertise of highly skilled professionals [6-8].

Recent advancements extending highly efficient and reliable ML techniques are revolutionizing applications in unsaturated soils [9-13], offering reduced costs and enhanced efficiency compared to traditional methods. Although some studies have estimated matric suction using ML techniques [14-17], to the best of the author's knowledge, there are no studies that have focused specifically on estimating suction in compacted fine-grained soils. The distinct behavior of compacted fine-grained soils has a strong link to microstructure and aggregates [18-20]. Investigating these characteristics typically necessitates costly and complex methods, such as mercury intrusion porosimetry (MIP) and environmental scanning electron microscopy (ESEM) [21]. Efficient methods are needed to reflect soil structure in suction estimation without adding complexity to engineering practice.

To achieve this objective, a simple equation is derived in this study based on the MGGP model, offering a programming-free method for estimating the matric suction of compacted fine-grained soils using basic soil properties. The proposed equation can be applied either manually or with a spreadsheet tool.

Methodology

Multi-Gene Genetic Programming (MGGP)

Multi-Gene Genetic Programming (MGGP) is an explainable ML technique rooted in the principles of Genetic Programming (GP) inspired by Darwin's theory of evolution [22]. As a non-deterministic algorithm, MGGP does not require prior domain expertise or assumptions about the functional style of the model. This capability, coupled with its exceptional ability to capture non-linear relationships, has made MGGP a valuable tool for estimating the hydro-mechanical properties of unsaturated soils [23-26].

MGGP generates mathematical expressions that elucidate the relationship between inputs and output, as shown in Eq. 1.

$$w_0 + w_1(0.5\ x_1 + 10) + w_2(3x_2\ - \cos(x_3)) + \cdots + w_n(\text{Equation (n)}) \tag{1}$$

where w_i, $i \in [0, n]$ is the weight; x_i is the input.

As shown in Fig. 2, the MGGP model has a tree-based structure with varying depths.

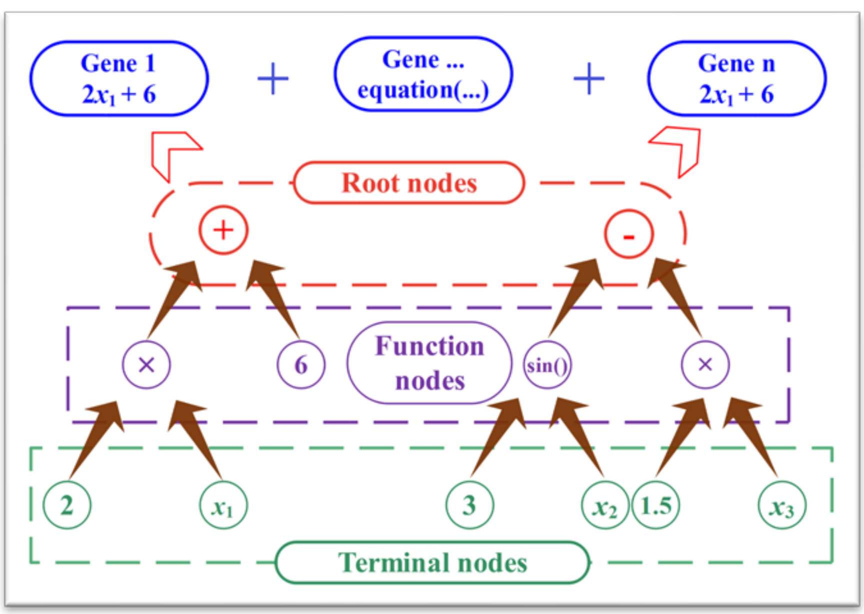


Figure 2: Illustration of a tree structure of the MGGP algorithm.

In the GP model's tree structure, functional nodes represent operations, such as arithmetic or Boolean functions, while terminal nodes, typically at the leaves, consist of independent variables and problem-specific constants. Unlike traditional GP models with a single tree or gene expression, the MGGP algorithm introduces multiple genes, integrating the structural flexibility of GP with the parameter estimation accuracy of classical regression.

Modelling procedure

The AI modeling process undertaken is divided into four stages, as shown in Fig. 3.

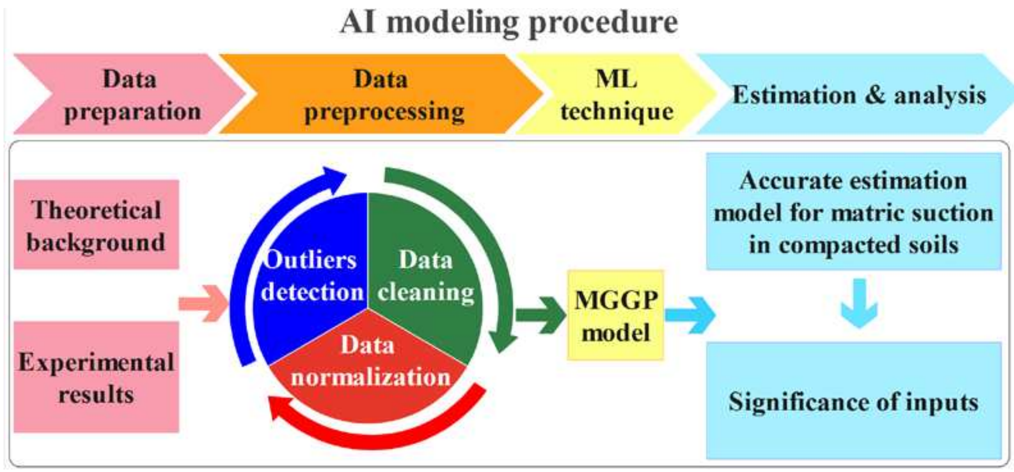


Figure 3: Modeling procedure used in this study

The modeling process begins with the selection of inputs based on theoretical foundations and well-established experimental results. Data are then collected and preprocessed to ensure suitability for modeling, followed by splitting the dataset into training (70%) and testing (30%) subsets. The MGGP model is trained using the training dataset, while its performance is evaluated with the testing dataset. This procedure results in a model capable of estimating the matric suction of compacted fine-grained soils. Furthermore, significant analysis that is undertaken in this study provides more information about the contribution of each input, offering valuable insights into their influence on variations in matric suction due to the influence of soil structure.

Input selection

The key parameters that influence the matric suction in fine-grained compacted soils are selected as inputs and grouped into three categories: basic soil properties, compaction characteristics, and derived parameters that capture the effects of mineralogical composition and plasticity, as outlined in Tables 1 to 3

Table 1: The basic inputs and obtaining methods.

No.	Input	Symbol	Unit	Test method /Relationship*
1	Sand fraction (x_1)	S_d	%	Sieve analysis test
2	Silt fraction (x_2)	M	%	Hydrometer method
3	Clay fraction (x_3)	C	%	Hydrometer method
4	Liquid limit (x_4)	LL	%	Cone penetration test / Casagrande cup
5	Plastic limit (x_5)	PL	%	Cone penetration test / Casagrande cup
6	Plasticity index (x_6)	I_P	%	$I_P = LL\text{-}PL$
7	Specific gravity (x_7)	G_s	-	Pycnometer method
8	Degree of saturation (x_8)	S_r	%	$S_r = wG_s / e_0$

Note: * The testing method for the input listed above adheres to the American Society for Testing and Materials (ASTM) standards.

Table 2: The compaction characteristics and inputs.

No.	Input	Symbol	Unit	Test method /Relationship*
1	Initial void ratio (x_9)	e_0	-	$e_0 = wG_s/S_r$
2	Compaction water content (x_{10})	w_c	%	-
3	Dry unit weight (x_{11})	γ_d	kN/m^3	$\gamma_d = \gamma/(1+w)$
4	Degree of compaction (x_{15})	D_c	%	$D_c = \gamma_d / \gamma_{dmax}$

Table 3: The inputs related to the effect of mineralogy and plasticity of soil.

No.	Input	Symbol	Unit	Expressions
1	Fine-grained fraction (x_{12})	F	%	$F = M + C$
2	Weighted plasticity index (x_{13})	I_{wp}	%	$I_{wp} = F \times I_p$
3	Activity index (x_{14})	A	%	$A = I_p / C$

These selected inputs are soil properties that can be readily measured in conventional soil mechanics laboratories performing simple tests.

The dataset for modelling

The dataset consists of 61 different soil samples, as shown in Table 4. These compacted soils contain varying proportions of fine-grained particles and are used to train the MGGP model.

Table 4: Experimental data gathered from the published literature.

ID	Classification*	Testing method	References
1-11	CL	Axis-translation technique	[27]
12-20	CL, CH	Filter paper, pressure plate	[28]
21-27	CH	Pressure plate	[29]

ID	Classification*	Testing method	References
28-33	CL, CH	Pressure membrane test	[1]
34-37	ML, CL, CH	Axis-translation technique	[30]
38	CL	Axis-translation technique	[31]
39-52	ML, CL, CL-ML	Pressure plate	[32]
53-54	MH, CL	Pressure plate	[33]
55	CL	Pressure plate	[34]
56-58	ML	Pressure plate, filter paper	[35]
59	CL	Filter paper	[36]
60-61	CL	Axis-translation technique	[37]

Note: * The soils are classified according to the Unified Soil Classification System (USCS)

The coefficient of determination, R^2, as shown in Eq. 2, can be employed to conduct a preliminary evaluation of the estimation results.

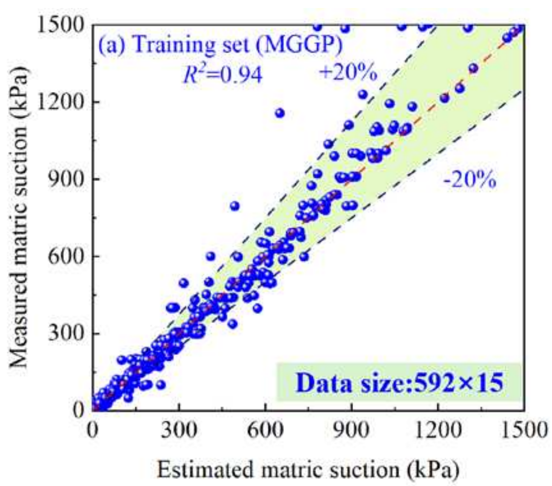
$$R^2 = \left(\frac{\sum_{i=1}^{n} (O_i - \bar{O}_i)(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^{n} (O_i - \bar{O}_i)^2 \sum_{i=1}^{n} (P_i - \bar{P}_i)^2}} \right)^2 \tag{2}$$

where n is the number of data points; O_i is the observed values; P_i is the estimated values.

Estimation result and analysis:

Estimation results

The estimation results are presented in Fig. 4. Overfitting was mitigated through systematic hyperparameter tuning, and the optimal configuration obtained from this process is summarized in Table 5.



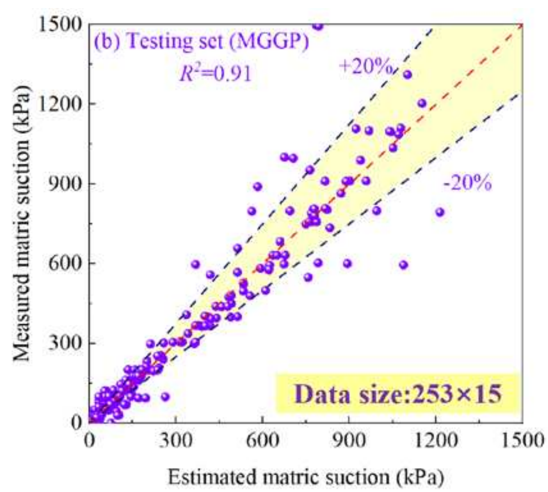


Figure 4: Estimation of matric suction using MGGP models: (a) training set and (b) testing set.

The $\boldsymbol{R}^2$ values for both the training and testing sets exceed 0.80, indicating a strong agreement between the MGGP-estimated and measured values. The comparable performance across datasets demonstrates good generalization capability and confirms that the model does not suffer from overfitting.

Table 5: Parameter settings configuration for the MGGP model.

Setting*	Values/ Names
Size of population	2500
Runs	20
Tournament size	25
MaxGenes	16
Function nodes	times, minus, plus, rdivide, square, tanh, exp, log, mult3, add3, sqrt, cube, negexp, neg, abs
MaxDepth	6

Note: A detailed description of the parameters can be found in [38]

A generic mathematical equation for the MGGP model can be expressed as equation (3). The mathematical expressions for each gene in Eq. 3 are presented in Table 6.

$$(u_a - u_w) = \sum_{i=0}^{n} \mathrm{Gene}(i) \tag{3}$$

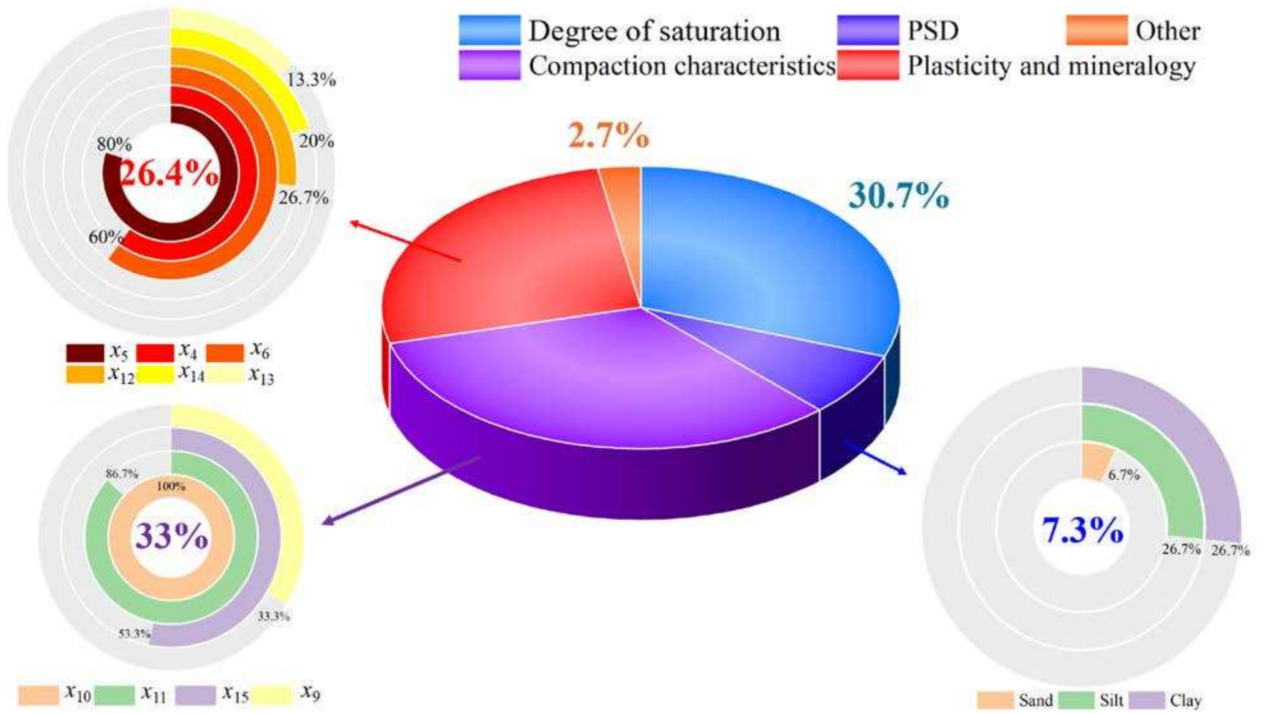
Table 6: Mathematical expressions of Genes in the MGGP Model.

ID	**Expression**	**ID**	**Expression**
Gene 1	$1100.0\ x_3$	Gene 9	$2200.0\ (x_8 + x_9)$
Gene 2	$-(3.23\ x_1^2\ x_4\ x_{15}^{1/2})/x_8^{4.78}$	Gene 10	$0.125\ x_{15}$
Gene 3	$-1777.0\ x_8$	Gene 11	$-6888.0\ x_8^{1/2}$
Gene 4	$11.5\ (\cos(x_4^2) - x_{11} - x_{14} - x_3 - \cos(x_3) - \cos(2.7\ x_{12}^3) - (x_{15})^{1/2}) + 89.5$	Gene 12	$-246.0\ \cos(\cos((x_7 - x_{11})/\log(x_8)))$
Gene 5	$1099.0\ x_2 - (1099.0\ (x_7 - x_{11}))/x_8$	Gene 13	$7799.0\ \log(x_8)$
Gene 6	$-\ 50.2\ x_{11} - 50.2\ \log(x_9 + 7.89)$	Gene 14	$-1099.0\ x_{12}$
Gene 7	$-66.4\ \cos(x_4^2)$	Gene 15	$-13.2\ x_8 - 13.2\ \mathrm{abs}(x_5)$
Gene 8	$-889.0\ \mathrm{abs}(\cos((x_7 - x_{11})/x_8))$	Gene 16	$-7.63\ x_{15}^{1/2}$
Bias	-266.0	-	-

Note: variables x_1 to x_{15} are defined in Tables 1–3.

Evaluating inputs importance: a statistical analysis based on model randomness

The stochastic nature of the MGGP model stems from random operations, including the selection of function and terminal nodes, and is further amplified by the mutation operator. This inherent variability ensures that each execution of the MGGP model generates multiple distinct mathematical expressions, which are utilized to evaluate the significance of inputs. Figure 5 summarizes the significance of the inputs derived from 20 distinct mathematical expressions produced in a single operation of the MGGP model.



Note: variables x_1 to x_{15} are defined in Tables 1–3.

Figure 5: Statistical analysis of input importance

As shown in Fig. 5, the inputs are classified into five categories to facilitate the analysis. Among these, the degree of saturation, compaction characteristics, and plasticity-mineralogy properties are identified as the three key categories for reliable estimation of matric suction in compacted fine-grained soils. In contrast, the impact of PSD and specific gravity on matric suction is relatively minor. Within the PSD, the sand fraction exerts a limited effect, as sand primarily acts as an inert filler with minimal physicochemical interactions. For the group of factors related to soil structure variation, the compaction water content is identified as the governing parameter, while soil consistency, which reflects soil plasticity, represents the second most influential factor. These two parameters affect soil aggregation and structural arrangement, contributing towards the development of a flocculated or dispersed structure, which in turn governs the coupled soil-water interaction and influencing the soil suction.

5. Summary and discussion:

The matric suction of compacted fine-grained soils that is influenced by soil structure is indirectly accounted in this study using alternative information derived from compaction characteristics and plasticity-mineralogy properties. The proposed approach for estimating matric suction eliminates the need for extensive testing, thereby enhancing the model's practicality for engineering applications. The main conclusions drawn from the study are as follows:

(1) The proposed MGGP model has been successfully used in the estimation of the matric suction of compacted fine-grained soils, considering the influence of the initial compaction state.

(2) The degree of saturation, plasticity and mineralogy inputs and the compaction characteristics inputs are found to be the key information that's essential for reliable estimation of matric suction in compacted soils.

(3) The equation derived from the MGGP model offers a practical tool for calculating matric suction in a spreadsheet environment, eliminating the need for programming and making it accessible to a broader range of users.

While the proposed model performs well, it has certain limitations in its implicit representation of soil structure. Future research should aim to explicitly incorporate structural features, with particular attention to dispersion, flocculation, and aggregation in compacted fine-grained soils, to further improve prediction accuracy and interpretability.

References

[1] Miller, C.J., N. Yesiller, K. Yaldo, and S. Merayyan, *Impact of soil type and compaction conditions on soil water characteristic.* Journal of Geotechnical and Geoenvironmental Engineering, 128(9): p. 733-742, 2002.

[2] Marinho, F.A., O.M.d. Oliveira, H. Adem, and S. Vanapalli, *Shear strength behavior of compacted unsaturated residual soil.* International Journal of Geotechnical Engineering, 7(1): p. 1-9, 2013.

[3] Chen, R. and C.W.W. Ng, *Impact of wetting–drying cycles on hydro-mechanical behavior of an unsaturated compacted clay.* Applied Clay Science, 86: p. 38-46, 2013.

[4] Gens, A., B. Vállejan, M. Sánchez, C. Imbert, M.V. Villar, and M. Van Geet, *Hydromechanical behaviour of a heterogeneous compacted soil: experimental observations and modelling.* Géotechnique, 61(5): p. 367-386, 2011.

[5] Fredlund, D.G. and H. Rahardjo, *Soil mechanics for unsaturated soils*. 1993: John Wiley & Sons.

[6] Deka, A., S. Vadlamudi, and S. Sekharan, *Effect of High Suction Measurement Technique on the SWCC of Expansive Soil*, in *Transportation, Water and Environmental Geotechnics*. 2021, Springer. p. 109-118.

[7] Sreedeep, S. and D.N. Singh, *Critical review of the methodologies employed for soil suction measurement.* International Journal of Geomechanics, 11(2): p. 99-104, 2011.

[8] Leong, E., S. Widiastuti, C. Lee, and H. Rahardjo, *Accuracy of suction measurement.* Geotechnique, 57(6): p. 547-556, 2007.
[9] Achieng, K.O., *Modelling of soil moisture retention curve using machine learning techniques: Artificial and deep neural networks vs support vector regression models.* Computers & Geosciences, 133, 2019.
[10] Li, Y., H. Rahardjo, A. Satyanaga, S. Rangarajan, and D.T.-T. Lee, *Soil database development with the application of machine learning methods in soil properties prediction.* Engineering Geology, 306: p. 106769, 2022.
[11] Zhang, Z. and X. Song, *Nanoscale soil-water retention mechanism of unsaturated clay via MD and machine learning.* Computers and Geotechnics, 163: p. 105678, 2023.
[12] Li, K.-Q., Z.-Y. Yin, N. Zhang, and J. Li, *A PINN-based modelling approach for hydromechanical behaviour of unsaturated expansive soils.* Computers and Geotechnics, 169: p. 106174, 2024.
[13] Ren, J., X. Fan, X. Yu, S.K. Vanapalli, and S. Zhang, *Use of an artificial neural network model for estimation of unfrozen water content in frozen soils.* 2022.
[14] Cheng, Z.-L., K.P.M. Kannangara, L.-J. Su, W.-H. Zhou, and C. Tian, *Physics-guided genetic programming for predicting field-monitored suction variation with effects of vegetation and atmosphere.* Engineering Geology, 315: p. 107031, 2023.
[15] Nobahar, M. and M. Khan, *Prediction of Matric Suction of Highway Slopes Using Autoregression Artificial Neural Network (ANN) Model*, in *Geo-Extreme 2021*. 2021. p. 40-50.
[16] Davar, S., M. Nobahar, M.S. Khan, and F. Amini, *The development of PSO-ANN and BOA-ANN models for predicting matric suction in expansive clay soil.* Mathematics, 10(16): p. 2825, 2022.
[17] Wang, J. and S. Vanapalli, *A framework for estimating the matric suction in unsaturated soils using multiple artificial intelligence techniques.* Int J Numer Anal Methods Geomech, 1-26, 2024.
[18] Vanapalli, S.K., D.G. Fredlund, and D.E. Pufahl, *The influence of soil structure and stress history on the soil–water characteristics of a compacted till.* Geotechnique, 49(2): p. 143-159, 1999.
[19] Romero, E., G. Della Vecchia, and C. Jommi, *An insight into the water retention properties of compacted clayey soils.* Géotechnique, 61(4): p. 313-328, 2011.
[20] Yin, P. and S.K. Vanapalli, *Model for predicting evolution of microstructural void ratio in compacted clayey soils.* Canadian Geotechnical Journal, 59(9): p. 1602-1621, 2022.
[21] Romero, E. and P.H. Simms, *Microstructure investigation in unsaturated soils: a review with special attention to contribution of mercury intrusion porosimetry and environmental scanning electron microscopy.* Geotechnical and Geological engineering, 26: p. 705-727, 2008.
[22] Koza, J.R., *Genetic programming as a means for programming computers by natural selection.* Statistics and computing, 4: p. 87-112, 1994.

[23] Garg, A., A. Garg, W.-H. Zhou, K. Tai, and M. Deo, *A new simulation approach of genetic programming in modelling of soil water retention property of unsaturated soil.* Engineering Computations, 32(3): p. 914-930, 2015.

[24] Garg, A., A. Garg, and K. Tai, *A multi-gene genetic programming model for estimating stress-dependent soil water retention curves.* Computational Geosciences, 18: p. 45-56, 2014.

[25] Nikooee, E. and K. Kamran Nejad, *Determining the effective stress parameter in drying path by MGGP method.* Amirkabir Journal of Civil Engineering, 55(5): p. 1039-1056, 2023.

[26] Gandomi, A.H. and A.H. Alavi, *A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems.* Neural Computing and Applications, 21: p. 189-201, 2012.

[27] Vanapalli, S.K., *Simple test procedures and their interpretation in evaluating the shear strength of an unsaturated soil*. 1996: (Ph.D. dissertation) University of Saskatchewan.

[28] Tinjum, J.M., C.H. Benson, and L.R. Blotz, *Soil-water characteristic curves for compacted clays.* Journal of geotechnical and geoenvironmental engineering, 123(11): p. 1060-1069, 1997.

[29] Ng, C.W. and Y.W. Pang, *Experimental investigations of the soil-water characteristics of a volcanic soil.* Canadian Geotechnical Journal, 37(6): p. 1252-1264, 2000.

[30] Sawangsuriya, A., *Stiffness-suction-moisture relationship for compacted soils*. 2006: The University of Wisconsin-Madison.

[31] Khalili, N. and S. Zargarbashi, *Influence of hydraulic hysteresis on effective stress in unsaturated soils.* Géotechnique, 60(9): p. 729-734, 2010.

[32] Johari, A., G. Habibagahi, and A. Ghahramani, *Prediction of SWCC using artificial intelligent systems: A comparative study.* Scientia Iranica, 18(5): p. 1002-1008, 2011.

[33] Goh, S., H. Rahardjo, and E. Leong, *Shear strength of unsaturated soils under multiple drying-wetting cycles.* Journal of Geotechnical and Geoenvironmental Engineering, 140(2): p. 06013001, 2014.

[34] Hoyos, L.R., C.L. Velosa, and A.J. Puppala, *Residual shear strength of unsaturated soils via suction-controlled ring shear testing.* Engineering Geology, 172: p. 1-11, 2014.

[35] Ye, Y.-x., W.-l. Zou, Z. Han, and X.-w. Liu, *Predicting the entire soil-water characteristic curve using measurements within low suction range.* Journal of Mountain Science, 16(5): p. 1198-1214, 2019.

[36] Chen, H., J. Li, X. Yuan, B. Shi, and W. Shan, *Influence of water and salt on suction characteristics of unsaturated clay: variation, mechanism, and fitting.* Bulletin of Engineering Geology and the Environment, 80: p. 8535-8551, 2021.

[37] Tripathy, S., S. Al-Khyat, and P.J. Cleall, *Impact of single and multiple specimen suction control oedometer testing on the measurement of the soil–water characteristic curve.* Geotechnical Testing Journal, 44(6), 2021.

[38] Searson, D.P., *GPTIPS 2: an open-source software platform for symbolic data mining*, in *Handbook of genetic programming applications*. 2015, Springer. p. 551-573.