Estimation of Swelling Index of Expansive Soils Using Machine Learning Techniques

Aolin Zhang[1], and Sai Vanapalli*[1]

[1]*Department of Civil Engineering, University of Ottawa, Ottawa, Canada*

**Corresponding author's email:* sai.vanapalli@uottawa.ca

Abstract: Geo-infrastructures such as the pavements, retaining walls, foundations and slopes constructed with or within expansive soils experience significant swelling behavior. One of the key parameters required for interpreting the behavior of expansive soils is the swelling index, which is determined conventionally from oedometer test results. However, these tests are time-consuming, costly, and require trained professionals to perform them. Therefore, simple yet reliable methods for estimating the swelling index of expansive soils will be valuable for use in geotechnical engineering practice applications. In this study, to address this need, a conventional regression method and two advanced machine learning models; namely, the multilayer perceptron (MLP) and extreme learning machine (ELM) are proposed for estimating the swelling index using only the information of plasticity index. The three equations developed for estimating the swelling index have R^2 values of 0.87 or higher highlighting their good performance. The proposed equations rapidly estimate the swelling index and can be used with confidence in geotechnical engineering practice applications.

Introduction

Practicing geotechnical engineers consider expansive soils as a nightmare because of the challenges associated with them; for this reason, they are widely referred to as problematic soils. These soils are widely distributed across more than 40 countries in the world and are highly sensitive to water content variations [1]. They pose significant risks to lightly loaded geo-infrastructures that include pavements and shallow foundations, as their swelling behavior upon water absorption often leads to ground heave and structural damages. The annual costs associated with addressing issues from expansive soils have been reported worldwide, the losses are staggering and are estimated to be $15 billion in China, $11 billion in the United States, and $3.3 billion in France[2, 3]. Although the total costs associated with expansive soil damage are not available in Canada, an estimated annual repair cost of $1 million for a 850 km length water pipeline in the Regina area of Saskatchewan suggests that the nationwide expenses for addressing expansive soil issues are likely to be substantial [4].

The widespread expansive soils and their significant economic impact encourages development of simple techniques for estimation of their swelling behavior for use in conventional geotechnical engineering practice applications. The swelling index, denoted as (C_s), refers to the slope of the rebound portion of the void ratio versus effective stress relationship, as shown in Figure 1, provides information related to the potential expansiveness of the soil [5, 6]. The consolidation-swell oedometer test is widely regarded as a reliable method for measuring swelling characteristics, including the swelling index [7]. However, oedometer tests require expensive equipment, trained personnel, and significant time typically several weeks making them impractical for budget-constrained or time-sensitive geotechnical projects [8, 9]. To overcome these limitations and enhance efficiency, numerous efforts have been made to empirically or semi-empirically estimate the swelling index using other soil properties. Research has shown that the swelling index is significantly influenced by many factors that include the plastic limit, liquid limit, plasticity index, clay fraction, void ratio, specific gravity, water content, dry density, clay minerals, and specific surface area [9-14]. While empirical methods are simpler and more convenient for use in engineering practice, measurement of multiple soil properties is both expensive and time-consuming. Therefore, a simple method that is both reliable and rapid would be valuable in engineering practice applications for estimating the swelling index. In recent years, machine learning techniques have offered enhanced predictive accuracy and efficiency for complex, nonlinear problems[15] as a transformative approach for use in geotechnical engineering applications. These techniques have been successfully applied in foundation design [16], slope stability analysis [17], and studies on special soils such as expansive [18] and frozen soils [19]. For this reason, in this study, multiple machine learning algorithms and a traditional regression analysis method were employed to estimate the swelling index of expansive soils.

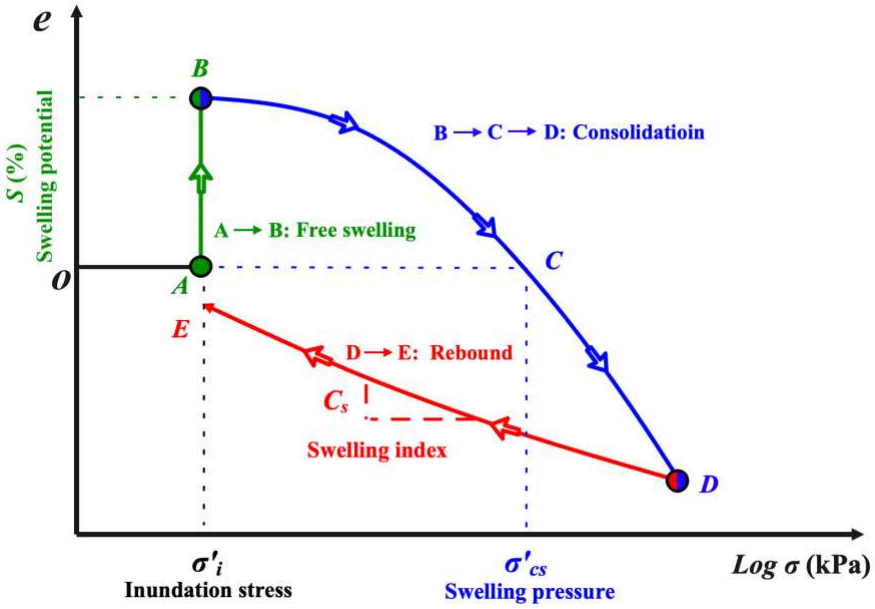

Figure 1: Conventional consolidation-swell oedometer test

Database and Methods

Database establishment

The plasticity index, Ip, defined as the difference between the liquid limit and plastic limit, represents the primary range in which expansive soils undergo water absorption and swelling. The interval between the water contents of plastic and liquid limits is the most significant phase of soil expansion. Numerous researchers have used the plasticity index as a critical parameter for predicting the swelling index [9, 12, 13]. Therefore, a comprehensive database has been gathered to facilitate the estimation of the swelling index, encompassing measured data on the plasticity index and swelling index of several expansive soils[20-31]. The plasticity index within this dataset ranges from 26.4 to 100, while the corresponding swelling index spans from 0.025 to 0.19, effectively covering a wide range of expansive soils that are frequently encountered in geotechnical engineering practice.

Methods

In this study, two machine learning methods; namely, multilayer perceptron (MLP) and extreme learning machine (ELM), are used to estimate the swelling index based on the information related to plasticity index. MLP is a widely used feedforward neural network where signals flow in a single direction, from the input layer to the output layer [32]. The hidden layer plays a crucial role in the MLP algorithm, as it processes the weights and biases assigned to each input feature, passes them through subsequent layers, and applies an activation function to generate predictions. Estimation accuracy can be enhanced by increasing the number of hidden layers as weights and biases are further refined through additional iterations. However, blindly increasing the number of hidden layers can add unnecessary complexity to the model, potentially leading to overfitting, where the model focuses on input details rather than capturing the underlying trends. Fig. 2 illustrates the principle of the MLP algorithm, and the general mathematical expression of MLP is presented in Eq. 1.

$$y = \sum_{j=1}^{M} \left[\omega_j^{OP} f\left(\sum_{i=1}^{N} x_i \omega_{ij}^{IP} + b_j^{IP} \right) \right] + b^{OP} \quad (1)$$

where, x_i is the i^{th} input parameter, y is the target parameter,ω_{ij}^{IP} are the weights assigned to input parameters, ω_j^{OP} are the weights assigned for the output layer, b_j^{IP} is the bias for the j^{th} hidden layer, $b^{OP}$is the bias for the output layer, f is the activation functions, N is the number of input parameters, M is the number of hidden layers.

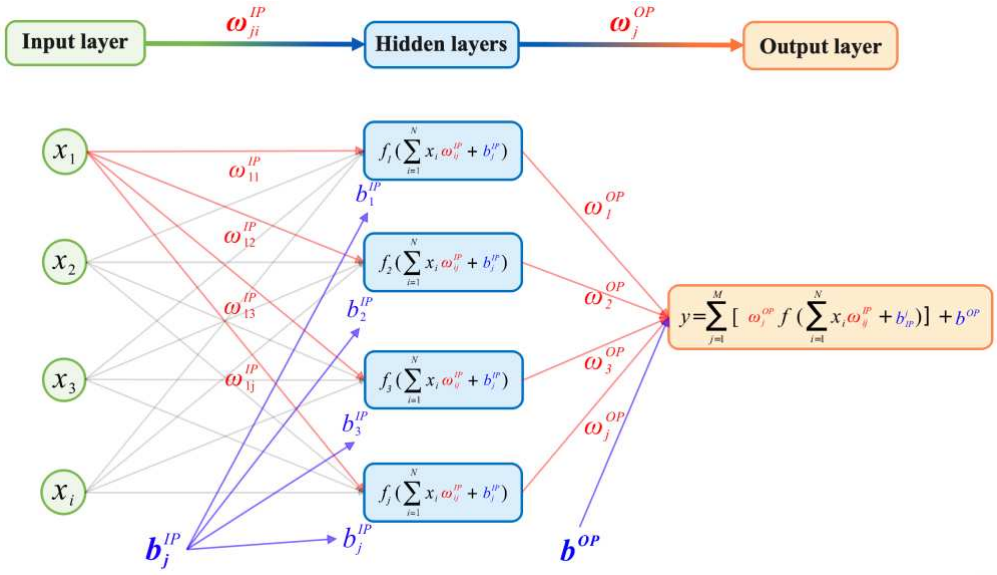

Figure 2: Methodology used in multilayer perceptron (MLP) method

ELM is another type of feedforward neural network, shares the primary structure of MLP but differs in its simplified architecture, featuring only one hidden layer. It is known for achieving excellent performance with exceptionally fast learning speeds. Unlike MLP, ELM randomly assigns hidden node parameters, and keeps them unchanged during the training process. The output weights are computed analytically using the Moore-Penrose pseudoinverse, enabling the model to be optimized by minimizing the error between the predicted and actual outputs [33]. Fig. 3 shows the structure of the ELM algorithm along the mathematical equation shown in Eq. 2.

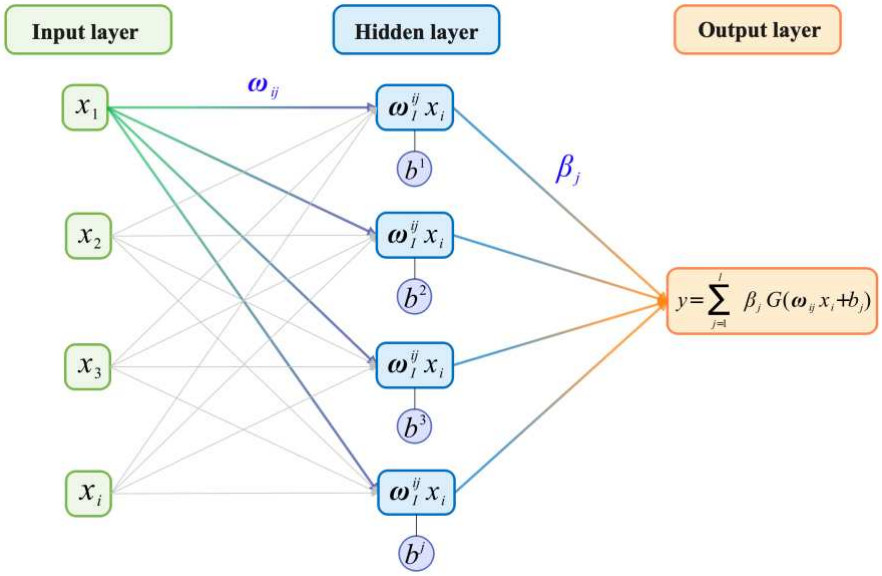

Figure 3: Methodology used in extreme learning machine (ELM) method

$$y = \sum_{j=1}^{l} \beta_j G(\omega_{ij} x_i + b_j) \quad (2)$$

where x_i is the i^{th} input parameter, y is the target parameter, ω_{ij} are the weights of from input layer to hidden layer, β_j are the weights from hidden layer to output layer, b_j is the j^{th} bias of hidden neurons, G is the activation function, l is the number of hidden neurons.

Modelling process

In machine learning model development, data preprocessing is a critical step in establishing a foundation for achieving better model performance. It ensures data quality by addressing noise, normalizing features, and appropriately splitting the dataset, enabling the model to focus on learning meaningful patterns. Proper preprocessing not only enhances model accuracy but also improves its ability to generalize the unseen data. In this study, the Max-Min normalization method is employed to scale the plasticity index and swelling index values to a consistent range between -1 and 1, thereby eliminating the influence of dimensionality and improving predictive performance. Furthermore, to achieve the best performance, 70% of the data is allocated for model training, while the remaining 30% is reserved for testing to evaluate the model's effectiveness [34].

In constructing an MLP model, several critical components must be defined, including the number of hidden layers, the training function, and the activation function. The number of hidden layers determines the model's complexity and predictive capability. The training function governs the optimization and updating of model parameters, while the activation function introduces nonlinearity, enabling the network to represent complex functional relationships and address nonlinear problems effectively. In this study, the MLP model is designed with a single hidden layer, and the Levenberg-Marquardt method (Trainlm) is chosen as the training function, which is well-suited for handling small to medium-sized feedforward neural networks. The Tangent Sigmoid (Tanh) activation function is employed due to its key advantages, including its symmetry (output range of [−1, 1]), which aligns perfectly with the normalized input data range. This alignment accelerates gradient convergence and enhances the model's nonlinear representation capability. Tanh is particularly suitable for shallow networks, as it effectively reduces gradient bias, thereby improving training efficiency and overall model performance.

The ELM is a feedforward neural network, shares similarities with MLP in model development, including data preprocessing and activation function selection. In this study, the ELM model employs the Sigmoid activation function, which effectively captures complex input feature relationships through nonlinear mapping and smooth output transitions, enhancing the model's

generalization performance. Unlike MLP, ELM features a simplified architecture with only one hidden layer, where multiple neurons control the model's complexity and accuracy. To balance underfitting and overfitting while accommodating the size of the dataset, the number of neurons in the hidden layer was set to 4, as recommended by Sheela and Deepa [35].

Furthermore, an effort has been made to extract the weights and biases from the MLP and ELM models and integrate them with Eq. 1 and Eq. 2 to derive explicit machine learning-based predictive equations for the swelling index. This approach enhances the interpretability of the results, improves the model's practical applicability and versatility for use in engineering practice reducing reliance on specialized software and computational resources. Eq. 3 and Eq. 4 are the proposed equations for estimating swelling index based on the information of plasticity index. Since the data have been normalized, the plasticity index and swelling index in Eq. 3 and Eq. 4 represent normalized values. The actual values can be obtained through inverse normalization using Eq. 5 and the specified data ranges ($26.34 < I_p < 100.04$, $0.025 < C_s < 0.190$). For comparative analysis, this study also employed traditional regression methods to analyze the data, resulting in Eq. 6.

$$C_s = -0.931\left(\frac{2}{1+e^{-2(-1.734\,I_p-0.645)}}-1\right)-0.065 \tag{3}$$

$$\begin{aligned} C_s = {} & 26658.252\frac{1}{1+e^{-(0.594I_p+0.162)}}+10376.544\frac{1}{1+e^{-(-0.722I_p+0.892)}} \\ & -12145.607\frac{1}{1+e^{-(-0.417I_p+0.006)}}-24548.611\frac{1}{1+e^{-(0.644I_p+0.569)}} \end{aligned} \tag{4}$$

$$x = x_{\text{norm}}\,(x_{\max}-x_{\min})+x_{\min} \tag{5}$$

$$C_s = 0.049+\frac{0.13}{\left(1+10^{(52.03-I_p)0.054}\right)} \tag{6}$$

where x is original value, x_{norm} is normalized value, $x_{\min}$, $x_{\max}$ are the minimum and maximum values of the original data.

Machine learning results and discussion

The performance of the three models investigated in this study using MLP, ELM, and regression analysis is illustrated in Fig. 4 based on their training and testing datasets. The MLP model demonstrates balanced accuracy with R^2 values of 0.883 and 0.876 for training and testing, respectively, and low RMSE values of 0.017 and 0.018, indicating consistent predictive capability. The ELM model achieves higher R^2 values of 0.900 and 0.888, suggesting improved fitting capacity, but its higher RMSE values (0.190 and 0.201) indicate larger deviations for certain predictions. Regression analysis, while achieving comparable RMSE (0.018), has a lower R^2 of 0.870, highlighting its limitations in capturing complex relationships. All three models demonstrate reasonable predictive performance, with most data points falling within the ±20%

error margins. However, MLP provides a more balanced trade-off between accuracy and error, while ELM exhibits a stronger fitting capability but with slightly larger prediction variances. Regression analysis serves as a simpler baseline model with less predictive power but comparable error levels.

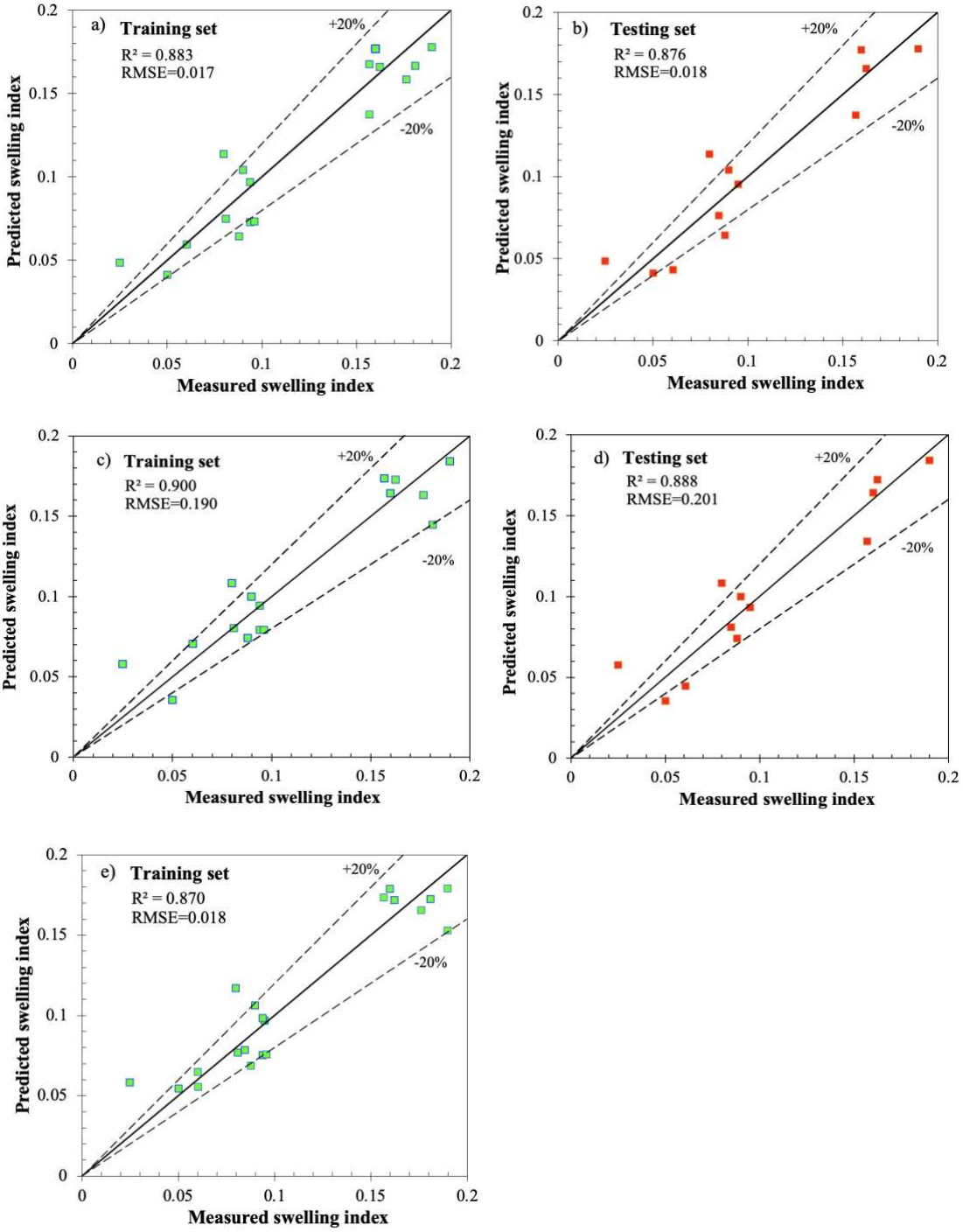

Figure 4: Prediction results of swelling index from a and b) MLP Model, c and d) ELM Model, e) Regression Model.

To further assess the prediction consistency of the three models for the swelling index as a continuous function of the plasticity index, Fig. 5 presents the relationships between these variables as represented by three predicted curves: empirical (green), MLP (red), and ELM (blue), overlaid with the measured data. Among the models, the ELM curve achieves the highest R^2 value of 0.888, effectively capturing both the overall trend and localized variations, particularly in regions with higher I_p. However, the ELM curve exhibits slight overfitting, especially near the peaks, where it deviates from the measured data. The MLP curve (R^2 = 0.876) follows the measured data closely, offering a smoother and more consistent trend. This balance between accuracy and generalization makes it robust, though it occasionally underestimates rapid localized changes. In contrast, the empirical model (R^2 = 0.870) generates a more linear and less flexible trend, adequately capturing the overall pattern but failing to represent nonlinear behaviors, particularly in regions with rapid changes at mid-range and higher I_p values.

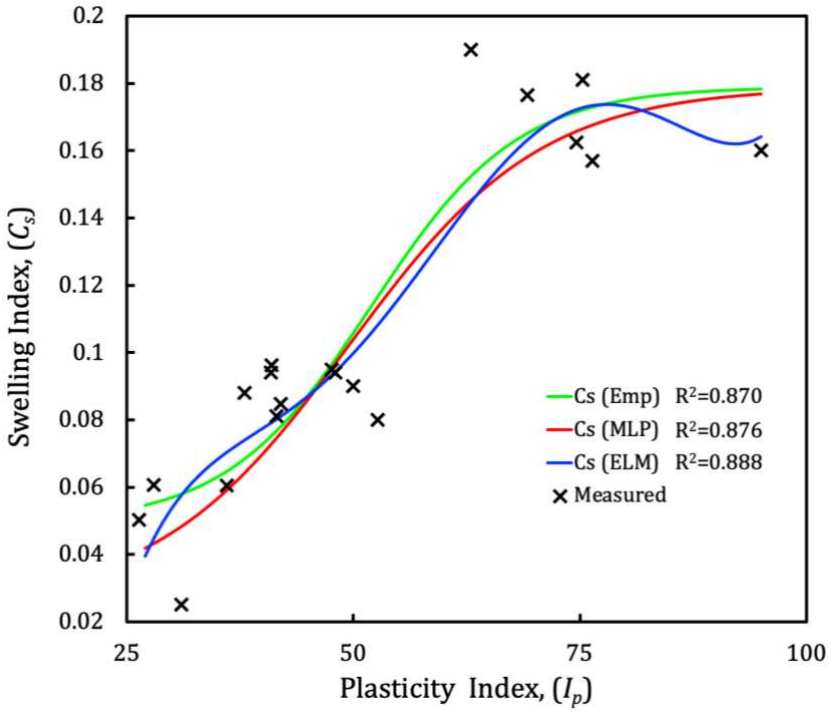

Figure 5: Comparison of empirical and machine learning models for predicting swelling index based on plasticity index

Despite the models' ability to capture the positive correlation between the plasticity index and the swelling index, certain limitations persist. One limitation is the lack of soil samples

information with plasticity indices within the ranges ($55 < I_p < 70$) and ($80 < I_p < 90$), leaving the models' validity within these intervals uncertain. Future improvements could include expanding the data sample size to enhance the models' accuracy and reliability.

Conclusions

In this study, two machine learning models were developed to predict the swelling index of expansive soils, using the plasticity index as the sole input parameter. The proposed approach offers a practical solution for use in geotechnical engineering applications alleviating the need for complex experimental procedures that are both time consuming and expensive. Machine learning techniques: namely, MLP and ELM, alongside traditional regression analysis that were employed to establish predictive equations for the swelling index, achieving R^2 values of 0.87 or higher demonstrating high accuracy. The results of this study highlight the potential of integrating machine learning in geotechnical engineering applications to streamline predictive modeling that are promising and can facilitate more cost-effective design practices for expansive soils, in future.

References

[1] Y Liu, et al. Soil-water characteristic curve of expansive soils considering cumulative damage effects of wetting and drying cycles. *Engineering Geology*, 339:107642, 2024.

[2] A Abden, M Al-Shamrani, and M Dafalla. Evaluating the feasibility of inverted T foundation on expansive soils. *Journal of Building Engineering*, 97:110788, 2024.

[3] LD Johnson. *Influence of suction on heave of expansive soils.* 1973.

[4] S Azam, Y Hu, and R Chowdhury. Sensor installation for field monitoring in expansive soils. in Proceedings of 62nd Canadian geotechnical conference, Halifax. 2009.

[5] A Rao, B Phanikumar, and RS Sharma. Prediction of swelling characteristics of remoulded and compacted expansive soils using free swell index. *Quarterly Journal of Engineering Geology and Hydrogeology*, 37(3):217-226, 2004.

[6] LD Jones and I Jefferson. *Expansive soils*. 2012.

[7] A Soltani, et al. Swell–shrink–consolidation behavior of rubber–reinforced expansive soils. *Geotechnical Testing Journal*, 42(3):761-788, 2019.

[8] AA Fondjo, E Theron, and RP Ray. Semi-empirical model for predicting the swelling stress of compacted, unsaturated expansive soils. *Civil Engineering and Architecture*, 9(1):225-239, 2021.

[9] H Tu and S K Vanapalli. Prediction of the variation of swelling pressure and one-dimensional heave of expansive soils with respect to suction using the soil-water retention curve as a tool. *Canadian Geotechnical Journal*, 53(8):1213–1234, 2016.

[10] T Nagaraj and B Srinivasa Murthy. A critical reappraisal of compression index equations. *Geotechnique*, 36(1):27-32, 1986.

[11] NS Işık. Estimation of swell index of fine grained soils using regression equations and artificial neural networks. *Scientific Research and Essays*, 4(10):1047-1056, 2009.

[12] S Vanapalli and L Lu. A state-of-the art review of 1-D heave prediction methods for expansive soils. *International Journal of Geotechnical Engineering*, 6(1):15-41, 2012.

[13] B Tiwari and B Ajmera. Consolidation and swelling behavior of major clay minerals and their mixtures. *Applied Clay Science*, 54(3-4):264-273, 2011.

[14] G Ross. Relationships of specific surface area and clay content to shrink-swell potential of soils having different clay mineralogical compositions. *Canadian Journal of Soil Science*, 58(2):159-166, 1978.

[15] M Elahi, et al. A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment. *Discover Artificial Intelligence*, 3(1):43, 2023.

[16] J Zhang, et al. Bayesian network based machine learning for design of pile foundations. *Automation in Construction*, 118:103295, 2020.

[17] S Suman, et al. Slope stability analysis using artificial intelligence techniques. *Natural Hazards*, 84:727-748, 2016.

[18] S Gahlot, et al. Prediction of swelling pressure of expansive soil using machine learning methods. *Asian Journal of Civil Engineering*,1-16, 2024.

[19] X Song, SK Vanapalli, and J Ren. Prediction of thermal conductivity of frozen soils from basic soil properties using ensemble learning methods. *Geoderma*, 450:117053, 2024.

[20] R Yoshida, DG Fredlund, and J Hamilton. The prediction of total heave of a slab-on-grade floor on Regina clay. *Canadian Geotechnical Journal*, 20(1):69-81, 1983.

[21] DG Fredlund. Consolidometer test procedural factors affecting swell properties. In *Proceedings of the 2nd International Conference on Expansive Clay Soils*, pages 435–456, 1969.

[22] F Shuai.Simulation of swelling pressure measurements on expansive soils.University of Saskatchewan,1998.

[23] M Ito and S Azam. Engineering characteristics of a glacio-lacustrine clay deposit in a semi-arid climate. *Bulletin of engineering geology and the environment*, 68(4):551-557, 2009.

[24] AW Clifton, RT Yoshida, DG Fredlund, and RW Chursinoff. Performance of dark hall, Regina, Canada, constructed on a highly swelling clay. In *Proceedings of the 5th International Conference on Expansive Soils*, pages 197–201, 1984.

[25] S Abduljauwad, et al. Laboratory and field studies of response of structures to heave of expansive clay. *Geotechnique*, 48(1):103-121, 1998.

[26] RKH Ching and DG Fredlund. A small Saskatchewan town copes with swelling clay problems. In *Proceedings of the 5th International Conference on Expansive Soils*, 1984.

[27] S Azam and GW Wilson. Volume change behavior of a fissured expansive clay containing anhydrous calcium sulfate. In *Unsaturated Soils 2006*, pages 906–915, 2006.

[28] S Azam. Study on the swelling behaviour of blended clay–sand soils. *Geotechnical and Geological Engineering*, 25(3):369-381, 2006.

[29] J Nelson and DJ Miller.*Expansive soils: problems and practice in foundation and pavement engineering*.John Wiley & Sons,1997.

[30] HL Yao, P Cheng, and W P Wu. Prediction of ground heaves in expansive soils based on consolidation tests. *Chinese Journal of Rock Mechanics and Engineering*, 24(21):3911–3915, 2005.

[31] TL Zhan, CW Ng, and DG Fredlund. Field study of rainfall infiltration into a grassed unsaturated expansive soil slope. *Canadian Geotechnical Journal*, 44(4):392-408, 2007.

[32] MC Popescu, VE Balas, L Perescu-Popescu, and N Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588, 2009.

[33] GB Huang, QY Zhu, and CK Siew. Extreme learning machine: theory and applications. *Neurocomputing*, 70(1–3):489–501, 2006.

[34] B Vrigazova. The proportion for splitting data into training and test set for the bootstrap in classification problems. Business Systems Research: *International Journal of the Society for Advancing Innovation and Research in Economy*, 12(1):228-242, 2021.

[35] KG Sheela and SN Deepa. Review on methods to fix number of hidden neurons in neural networks. *Mathematical problems in engineering*, 2013(1):425740, 2013.