Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 323-331. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.2927

RESEARCH ARTICLE OPEN ACCESS

ENHANCING SWELLING PRESSURE PREDICTION THROUGH INTEGRATED MACHINE LEARNING METHODS

Laid Lekouara[1], Fatima Zohra Tebbi[2], Ouassila Bahloul[3] and Rachid Rabehi[*4]

[1]Faculty of Science and Technology, Department of Civil Engineering, University Abbes Laghrour, Khenchela, Algeria.
[2]LRNAT Laboratory, Earth Sciences and Universe Institute. University of Mostepha Ben Boulaid, Batna 2. Algeria.
[3]Civil Engineering Laboratory – Risks and Structures in Interactions,Department of Civil Engineering,Faculty of Technology, University of Batna 2 – Mostefa Ben Boulaïd Batna 05000, Algeria.
[4]Faculty of Civil Engineering, University of Science and Technology Houari Boumediene (USTHB), Bab Ezzouar, 16111, Algeria.

[1]https://orcid.org/0009-0006-6700-1245, [2]https://orcid.org/0000-0003-3321-5609
[3]https://orcid.org/0000-0001-8554-7563, [4]https://orcid.org/0000-0002-9398-7291

Email: lekouara_laid@univ-khenchela.dz, f.tebbi@univ-batna2.dz, o.bahloul@univ-batna2.dz, rachid.rabehi@usthb.edu.dz

ARTICLE INFO

Article History
Received: November 11, 2025
Revised: December 10, 2025
Accepted: January 1, 2026
Published: January 31, 2026

Keywords:
Swelling pressure,
Machine learning,
Feature selection,
Random forest,
Neural network.

ABSTRACT

Swelling soils, characterized by their high capacity to change volume due to variations in water content, pose significant challenges in civil engineering, including structural deformations, cracks in infrastructure, and foundation instabilities. These issues lead to high maintenance costs and safety risks for structures. To address these challenges, this study proposes an integrated approach combining machine learning and feature selection techniques to predict the swelling pressure (SP, kPa) of soils. Using a dataset comprising seven explanatory variables (liquid limit (LL, %), plasticity index (PI, %), shrinkage limit (WS, %), particle size fractions (Ff2mm, Ff80µm, Ff2µm), and methylene blue index (VBS, g/100 g)), four feature selection methods were evaluated: Least Absolute Shrinkage and Selection Operator (Lasso), Random Forest, Gradient Boosting, and the F-statistic test. These methods identified the five most influential variables for training an artificial neural network (ANN). The results show that Random Forest achieved the best predictive performance (MSE = 785.31, R^2 = 0.9706, MAE = 21.39), followed by the F-test (MSE = 1784.99, R^2 = 0.9332, MAE = 30.66), Gradient Boosting (MSE = 2497.38, R^2 = 0.9066, MAE = 40.20), and Lasso CV (MSE = 2605.78, R^2 = 0.9035, MAE = 40.29). Analysis of the variable distributions revealed complex nonlinear relationships between geotechnical parameters and swelling pressure. This multi-method approach demonstrates the effectiveness of machine learning techniques for accurate and cost-effective prediction of geotechnical properties, offering a promising alternative to traditional, time-consuming, and costly oedometer tests.

I. INTRODUCTION

Expansive soils, characterized by their significant volumetric changes in response to moisture content variations, represent one of the most challenging geotechnical phenomena in civil engineering practice [1],[2]. These soils, predominantly composed of clay minerals with high swelling potential, pose substantial threats to infrastructure stability, causing billions of dollars in damage annually through structural deformations, foundation failures, and pavement distress [3],[4]. The swelling pressure generated by these soils upon moisture increase can reach magnitudes sufficient to cause significant uplift forces on foundations and lateral pressures on retaining structures, necessitating accurate prediction methods for safe and economical design [5],[6]. However, traditional approaches to swelling pressure determination rely primarily on laboratory oedometer tests which, while providing reliable results, are time-consuming, expensive, and require specialized equipment and expertise [7],[8]. These limitations have driven the geotechnical engineering community to seek alternative prediction methods based on index properties such as liquid limit (LL), plasticity index (PI), shrinkage limit (WS), particle

size fractions (Ff2mm, Ff80µm, Ff2µm), and methylene blue index (VBS), but their applicability often remains limited to specific soil types [9-11]. In this context, the advent of artificial intelligence and machine learning has revolutionized numerous engineering disciplines, with geotechnical engineering particularly benefiting from these computational advances [12-14]. Artificial neural networks (ANNs), in particular, have emerged as powerful tools for modeling complex, nonlinear relationships inherent in geotechnical problems, due to their ability to capture intricate patterns in data without requiring explicit mathematical formulations [15-19]. Furthermore, recent studies have demonstrated the superior performance of machine learning methods in swelling pressure prediction. According to [20] successfully applied genetic expression programming and artificial neural networks to predict expansive soil properties. In [21] compared multiple machine learning models, demonstrating that neural networks consistently outperformed traditional statistical methods with R^2 values exceeding 0.92.

Similarly, [22] developed constrained XGBoost models achieving remarkable accuracy with R^2 values of 0.9832, while [23] demonstrated the effectiveness of dimensionality reduction combining linear (PCA) and nonlinear (ISOMAP) techniques with neural networks for swelling pressure prediction in Northeast Algeria. Moreover, feature selection plays a crucial role in machine learning applications, particularly in geotechnical engineering where numerous correlated parameters influence soil behavior [24],[25]. Various feature selection techniques have been employed, with particular focus on methods such as Least Absolute Shrinkage and Selection Operator (Lasso), Random Forest importance ranking, Gradient Boosting feature scoring, and statistical F-test analysis. In particular, advanced algorithms such as Boruta have shown promise, as demonstrated by [26] in their model for the Mila region. Lasso regression (LassoCV), known for its automatic feature selection capability through L1 regularization, has been successfully applied in geotechnical applications [27], while Random Forest and Gradient Boosting methods have demonstrated superior performance in capturing nonlinear relationships and feature interactions [28],[29].

Nevertheless, statistical approaches, including F-test based feature selection, continue to provide valuable insights into variable significance, although their linear assumptions may limit effectiveness with highly nonlinear geotechnical data [30],[31]. However, when combined with nonlinear modeling approaches such as ANNs, these methods can contribute to comprehensive predictive frameworks that leverage both statistical rigor and machine learning flexibility [32],[33]. Currently, machine learning applications in geotechnical engineering have expanded rapidly, with artificial neural networks representing approximately 52% of all AI applications in the field (Application of artificial intelligence in geotechnical engineering [34]). The versatility of neural networks in handling complex multidimensional problems has made them particularly attractive for soil behavior modeling [35-37], while support vector machines, random forest, and gradient boosting methods collectively represent nearly 40% of applications (State-of-the-art review on AI-enhanced computational mechanics, [38]). Nevertheless, current challenges include data quality and availability, model interpretability, and generalization to different soil types and environmental conditions [39-41].

Consequently, the integration of multiple machine learning approaches for feature selection and prediction represents a significant advancement in geotechnical engineering practice. Recent research has shown that ensemble methods and hybrid approaches often outperform individual techniques [42-44]. This integrated approach not only improves prediction accuracy through optimal feature subset identification but also provides insights into the relative importance of different geotechnical parameters, contributing to enhanced understanding of expansive soil behavior and the complex relationships between soil index properties and swelling pressure mechanisms. Furthermore, the practical implementation of machine learning-based swelling pressure prediction methods requires careful consideration of model validation using performance metrics such as Mean Squared Error (MSE), coefficient of determination (R^2), and Mean Absolute Error (MAE), along with uncertainty quantification and integration with existing geotechnical design practices [45],[46]. Moreover, recent advances in model interpretability and explainable AI have made machine learning approaches more acceptable to practicing engineers who require understanding of the underlying reasoning behind predictions [47-49]. Ultimately, the significance of this research extends to practical applications in geotechnical engineering design and construction.

Accurate swelling pressure prediction enables engineers to make informed decisions regarding foundation design, ground improvement strategies, and risk mitigation measures, contributing to the development of safer and more economical infrastructure in areas affected by expansive soils. Therefore, the main objective of this research is to develop and systematically compare four distinct feature selection methodologies - Least Absolute Shrinkage and Selection Operator (LassoCV), Random Forest, Gradient Boosting, and statistical F-test - coupled with artificial neural networks to identify the optimal combination for swelling pressure prediction, by rigorously evaluating their effectiveness on a comprehensive dataset of seven explanatory geotechnical variables including Atterberg limits (liquid limit LL and plasticity index PI), shrinkage limit (WS), particle size fractions (Ff2mm, Ff80µm, Ff2µm), and methylene blue index (VBS), using robust performance metrics (MSE, R^2, MAE) and providing evidence-based practical guidelines for method selection in geotechnical engineering applications.

II. MATERIALS AND METHODS

II.1 DATASET

The dataset comprises 160 observations with eight parameters: LL, PI, WS, Ff2mm, Ff80, Ff2, VBS, and SP. Descriptive statistics are provided in Table 1. The target variable, SP, has a mean of 616.82 kPa, a standard deviation of 136.24 kPa, and ranges from 389 to 936 kPa, reflecting significant variability in swelling pressure.

Table 1: Descriptive statistics for all parameters.

Parameter	Mean	Std Dev	Median	Min	Max	Skewness
LL	80.78	7.86	81.71	65.78	95.30	-0.33
PI	51.83	4.97	51.93	41.66	66.97	0.09
WS	8.88	1.79	8.60	5.80	12.60	0.24
Ff2mm	97.08	1.53	97.33	92.54	99.37	-0.85
Ff80	90.06	5.70	91.35	70.80	97.29	-1.30
Ff2	49.63	7.48	50.63	30.10	62.00	-0.70

VBS	12.53	2.09	12.58	8.00	16.00	-0.13
SP	612.82	136.24	582.00	389.00	936.00	0.74

Source: Authors, (2026).

III. METHODOLOGY

The methodology employed for predicting soil pressure (SP) followed a structured pipeline designed to optimize feature selection and neural network (NN) performance, as depicted in the provided flowchart (Figure 1). The process began with the loading of experimental data, containing 160 samples with seven initial features (LL, PI, WS, Ff2mm, Ff80µm, Ff2µm, VBS) and the target variable SP. All analyses were conducted using Python 3 in Google Colab, providing a cloud-based computational environment with access to necessary machine learning libraries and GPU acceleration capabilities. Data preparation involved handling missing values by imputing feature means, splitting the dataset into training (80%) and testing (20%) sets, and standardizing both features and the target variable using StandardScaler. Four feature selection methods—LassoCV, Random Forest, Gradient Boosting, and Statistical F-test—were applied to identify the top five most relevant features based on their respective criteria: absolute coefficients for Lasso CV, feature importance for Random Forest and Gradient Boosting, and F-scores for the F-test. Each selected feature subset was used to train a dedicated NN model with a multi-layer architecture (128-64-32-16-1 neurons with dropout layers) using the Adam optimizer, incorporating early stopping with a patience of 20 epochs to prevent overfitting. The models were evaluated using Mean Squared Error (MSE), R^2 score, and Mean Absolute Error (MAE), with training conducted over a maximum of 200 epochs and a batch size of 16. Finally, a comprehensive comparison of the models' performance was conducted, including bar plots for R^2, MSE, and feature counts.

III.1 MACHINE LEARNING APPROACHES

Four feature selection methods were applied to identify the five most relevant variables. The algorithms for each method are described below:

III.1.1 Algorithm Lasso Selection

1) Let M0 denote the null model, which contains no predictors and predicts the mean of the target (e.g SP).

2) For k = 0, ……, p – 1.
a. Consider all models that augment the predictors in Mk with a regularization parameter α.
b. Fit a Lasso regression model to minimize the cost function:

$$Cost = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \alpha \sum_{j-1}^{p}|\beta_j|. \quad (1)$$

Where y_i is the observed target, $\hat{y}_i = \beta_0 + \sum_{j=1}^{p} \beta_J \, x_{ij}$ is the prediction, n is the number of samples, p is the number of predictors, α is the regularization parameter, and $\sum_{j-1}^{p}|\beta_j|$ is the L1 penalty.

c. Choose the best model among these by selecting the α that optimizes a cross validated error metric (e.g, MSE).
d. Update $M_{(k+1)}$ with predictors whose coefficients $\beta_j \neq 0$, indicating their selection.

3) Select a single best model from among M_0, ……, M_p using cross validated prediction error (e.g., C_P, AIC, BIC, or adjusted R^2).

4) Output the final model with the selected predictors and their corresponding coefficients.

III.1.2 Algorithm Lasso Selection

1) Let M_0 denote the null model, which contains no predictors and predicts the mean of the target (e.g SP).

2) For k = 0, ……, p – 1.
a. Consider all p – k models that augment the predictors in M_k with a subset of the remaining predictors.
b. Fit a Random Forest model to the training data, using a specified number of trees (e.g., 200) and a Random subset of predictors at each split.
c. Compute the feature importance scores based on the decrease in node impurity across all trees.
d. Choose the best model among these by selecting the top N predictors with the highest importance scores, and update M_{k+1} with these predictors.

3) Select a single best model from among M_0, ……, M_p using cross validated prediction error (e.g., C_P, AIC, BIC, or adjusted R^2).

4) Output the final model with the selected predictors and their importance scores.

III.1.3 Algorithm Gradient Boosting Selection

1) Let M_0 denote the null model, which contains no predictors and predicts the mean of the target (e.g SP).

2) For k = 0, ……, p – 1.
a. Consider all p – k models that augment the predictors in M_k with a subset of the remaining predictors.
b. Fit a Gradient Boosting model to the training data, iteratively adding weak learners (e.g., decision trees) to minimize the loss function by fitting to the residuals.

c. Compute the feature importance scores based on the contribution of each predictor to reducing the loss across all iterations.
d. Choose the best model among these by selecting the top N predictors with the highest importance scores, and update M_{k+1} with these predictors.

3) Select a single best model from among M_0, ……, M_p using cross validated prediction error (e.g., CP, AIC, BIC, or adjusted R^2).

4) Output the final model with the selected predictors and their importance scores.

III.1.4 Algorithm f-Test Selection

1) Let M_0 denote the null model, which contains no predictors and predicts the mean of the target (e.g SP).

2) For k = 0, ……, p – 1.
a. Consider all p – k models that augment the predictors in Mk with one additional predictors.
b. Fit a linear regression model to the training data, for each model and compute the F-statistic for each additional predictor, measuring its significance in explaining the target (e.g SP).
c. Compute the F-scores and associated p-values for each predictor based onthe F-statistic.
d. Choose the best model among these by selecting the top N predictors with the highest importance scores, and update M_{k+1} with these predictors.

3) Select a single best model from among M_0, ……, M_p using cross validated prediction error (e.g., C_P, AIC, BIC, or adjusted R^2).

4) Output the final model with the selected predictors and their importance scores.

III.1.5 Algorithm Neural Network Predictor

1) Let M_0 denote the initial model, which is an untrained neural network with a predefined architecture and no learned weights.

2) For each training epoch t = 1…. T:
a. Consider the current model M_{t-1} and prepare the training data (e.g., $X_{\text{train-selected}}$, $Y_{\text{train-scaled}}$).
b. Forward propagate the input data through the network layers to compute predictions and calculate the loss (e.g MSE) against the target (e.g SP).
c. Backpropagate the error using gradient descent (e.g., Adam optimizer) to update the network weights.
d. Evaluate the validation loss on a held-out subset, and if it does not improve for a specified number of epochs, stop training and restore the best weights.

3) Select the final model MT based on the minimum validation loss achieved during training.

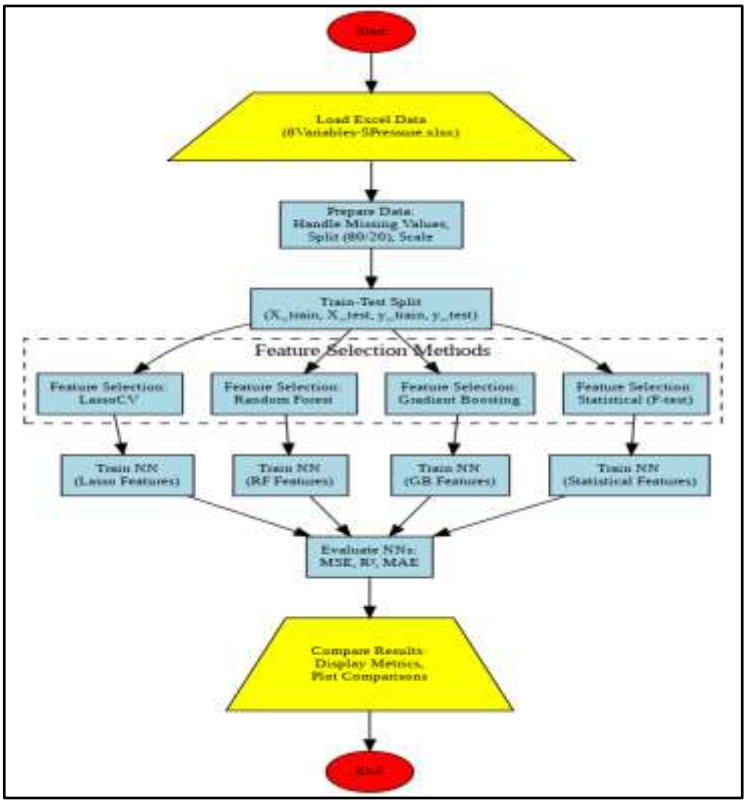

Figure 1: Flowchart of the modeling process.
Source: Authors, (2026).

III.2 PERFORMANCE METRICS

The performance of each model was evaluated using three:

III.2.1 Erreur Quadratique Moyenne (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

y_i : Actual (or observed) value for the i-th observation. This is target value the model aims to predict.
$\hat{y}_i$: Predicted value by the model for the i-th observation. This is the model's output.

III.2.2 Coefficient de Détermination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (3)$$

y_i : Actual value for the i-th observation
$\hat{y}_i$: Predicted value for the i-th observation
$\bar{y}$: Mean of the acyual values.

III.2.3 Erreur Absolue Moyenne (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \quad (4)$$

y_i : Actual value for the i-th observation
$\hat{y}_i$: Predicted value for the i-th observation
$\bar{y}$: Total number of observations, as in MSE.

IV. RESULTS

The comparative evaluation of the four methods reveals a significant performance hierarchy for swelling pressure prediction. All methods select exactly 5 parameters out of 7, but the high-performing methods (Random Forest, Gradient Boosting) prioritize fine fractions (Ff2, Ff80) that are essential to clay behavior, unlike Lasso which selects the less relevant coarse fraction (Ff2mm), explaining its inferior performance (Figure 2). Random Forest demonstrates exceptional superiority with a coefficient of determination of 0.9706 (MSE = 785.31 kPa², MAE = 21.39 kPa), explaining 97.06% of the total variance in swelling pressure. This performance stems from its unique algorithmic capacity to capture highly complex nonlinear relationships between the selected critical geotechnical variables (Ff2, LL, Ff80, PI, VBS). The superiority rests on a sophisticated impurity reduction mechanism based on the Gini criterion that enables remarkable precision in identifying the most discriminating variables, even in the presence of complex multivariate interactions and partial collinearities between geotechnical parameters, through its ensemble approach averaging predictions from hundreds of decision trees trained on different bootstrap subsamples (Figure 3 and 7).

The statistical F-test achieves solid performance in second position with robust metrics (R^2 = 0.9332, MSE = 1784.99 kPa², MAE = 30.66 kPa), explaining 93.32% of the variance. This performance testifies to the persistent effectiveness of classical statistical approaches in modern geotechnical analysis. However, the specific inclusion of WS (shrinkage limit) instead of Ff80 (particle size fraction < 80µm) slightly reduces its predictive accuracy compared to Random Forest, revealing a fundamentally different selection philosophy based on identifying direct linear relationships between individual parameters and swelling pressure (Figure 4 and 7). Gradient Boosting presents a methodological paradox with moderate performance (R^2 = 0.9066, MSE = 2497.38 kPa², MAE = 40.20 kPa) despite feature selection nearly identical to Random Forest (LL, Ff2, Ff80, PI, VBS). This illustrates an algorithmic paradox where its theoretically powerful sequential adaptive learning capability proves more sensitive to overfitting and stochastic variations than Random Forest's more stable ensemble approach.

The increased sensitivity is explained by the sequential nature of learning that can amplify modeling errors and higher variance due to dependency between successive predictors (Figure 5 and 7). Lasso CV displays the weakest performance with R^2 = 0.9035, MSE = 2605.78 kPa², and MAE = 40.29 kPa, primarily due to the suboptimal inclusion of Ff2mm (fraction < 2mm) instead of Ff80 (fraction < 80µm). This deficient selection reveals the fundamental limitations of L1 regularization which, through its rigid binary selection mechanism and strictly linear mathematical assumptions, cannot effectively capture the complex physicochemical interactions characteristic of clay swelling behavior. The crystalline (interlayer spacing), osmotic (ionic swelling pressure), and capillary (matric suction) processes interact in a highly synergistic, multiplicative, and non-additive manner. These complex mechanisms require sophisticated machine learning approaches capable of transcending the limitations of traditional linear models to exploit the informational richness of nonlinear patterns hidden in geotechnical data (Figure 6 and 7).

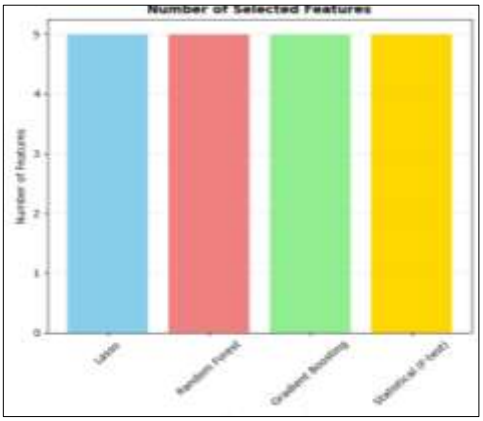

Figure 2: Selected features by each feature selection method.
Source: Authors, (2026).

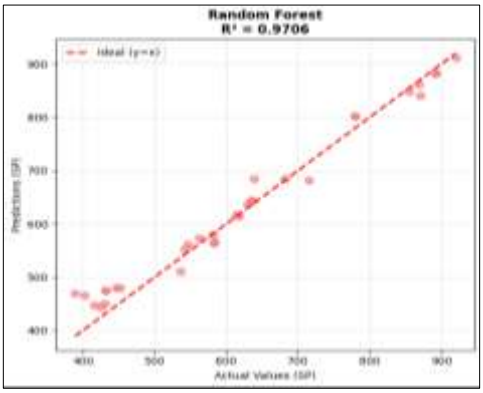

Figure 3: Random Forest performance for swelling pressure prediction.
Source: Authors, (2026).

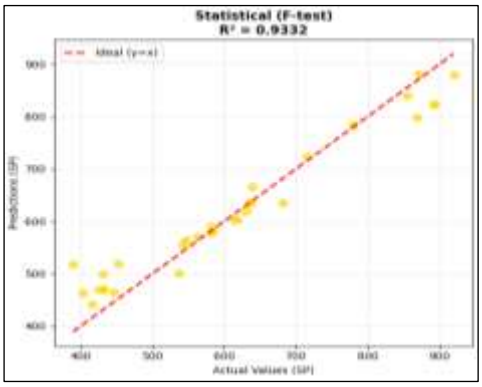

Figure 4: Staitistical (F-test) performance for swelling pressure prediction.
Source: Authors, (2026).

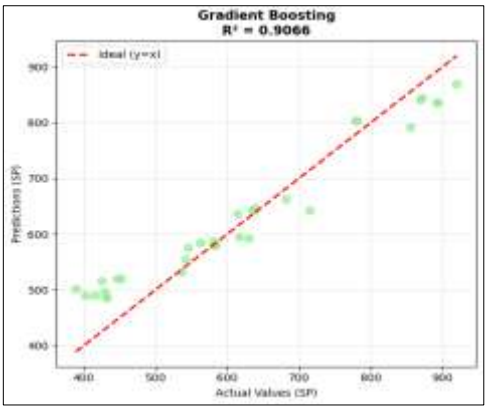

Figure 5: Gradient Boosting performance for swelling pressure prediction.
Source: Authors, (2026).

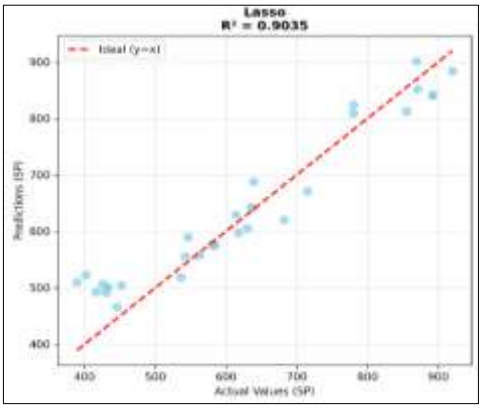

Figure 6: Lasso performance for swelling pressure prediction.
Source: Authors, (2026).

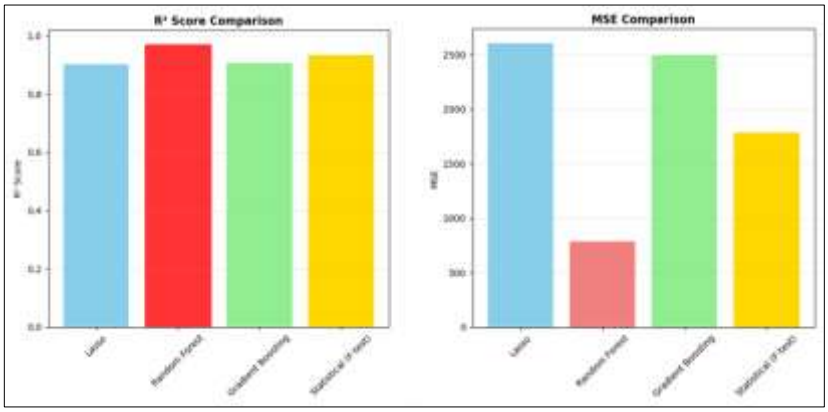

Figure 7: Comparative performance evaluation of feature selection methods using R² and MSE metrics.
Source: Authors, (2026).

IV CONCLUSIONS

The comparative analysis reveals significant performance differences between feature selection methods coupled with neural networks, Random Forest demonstrating clear superiority with an R^2 of 0.9706 using features Ff2, LL, Ff80, PI and VBS, highlighting the crucial importance of selection technique choice for optimizing predictive accuracy. This exceptional performance confirms the remarkable capacity of ensemble methods to capture complex nonlinear relationships inherent in clay swelling behavior, particularly through precise identification of critical fine particle size fractions. The statistical F-test achieves robust results (R^2 = 0.9332), demonstrating that classical statistical approaches retain considerable relevance, although the observed paradox with Gradient Boosting (R^2 = 0.9066) illustrates that optimal feature selection does not automatically guarantee superior performance, revealing the fundamental importance of algorithmic stability. The results reveal that all methods systematically select five parameters among the seven available variables, but high-performing techniques favor fine fractions essential to the complex physicochemical mechanisms governing clay swelling.

Future work should prioritize refinement of selection methods and architectural optimization of models. The Random Forest-neural network integration is recommended for preliminary swelling pressure evaluation, offering optimal balance between predictive accuracy, computational efficiency and result interpretability, facilitating its adoption by geotechnical engineering practitioners as a complement to traditional tests for critical applications. This multi-method approach represents a paradigmatic shift toward data-driven geotechnics, establishing foundations for future research in expansive soil characterization and encouraging broader adoption of artificial intelligence in geotechnics, ultimately contributing to the development of safer and more economical infrastructure by enabling engineers to make informed decisions based on robust scientific evidence, thereby reducing risks and optimizing costs in infrastructure projects affected by expansive soils.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Methodology: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Investigation: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Discussion of results: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Writing – Original Draft: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Writing – Review and Editing: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Resources: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Supervision: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.
Approval of the final text: Laid Lekouara, Fatima Zohra Tebbi, Ouassila Bahloul and Rachid Rabehi.

VI. REFERENCES

[1] V. Cantillo, V. Mercado, C. Pájaro, Empirical correlations for the swelling pressure of expansive clays in the city of Barranquilla, Colombia, Earth Sciences Research Journal 21 (2017) 45–49. https://doi.org/10.15446/esrj.v21n1.60226.

[2] C.I. Christopher, D.N. Chimobi, Emerging trends in expansive soil stabilisation: A review, Journal of Rock Mechanics and Geotechnical Engineering 11 (2019) 423–440. https://doi.org/10.1016/j.jrmge.2018.08.013.

[3] A. Khennouf, M. Baheddi, Heave analysis of shallow foundations founded in swelling clayey soil at N'Gaous city in Algeria, Studia Geotechnica et Mechanica 42 (2020) 210–221. https://doi.org/10.2478/sgem-2020-0009.

[4] J. Du, S. Zhang, H. Li, X. Zhao, L. Wang, Characterization of controlled low-strength materials from waste expansive soils, Construction and Building Materials 411 (2024) 134690. https://doi.org/10.1016/j.conbuildmat.2023.134690.

[5] B. Sharmila, S. Bhuvaneshwari, G. Landlin, Application of lignosulphonate: A sustainable approach towards strength improvement and swell management of expansive soils, Bulletin of Engineering Geology and the Environment 80 (2021) 6395–6413. https://doi.org/10.1007/s10064-021-02332-8.

[6] M.A. Waheed, O.S.B. Al-Amoudi, M.A. Al-Osta, Molecular-level behavior induction in the constitutive modeling of swelling clayey soils: A review, Transportation Geotechnics 39 (2023) 100947. https://doi.org/10.1016/j.trgeo.2023.100947.

[7] Z.A. Erguler, R. Ulusay, A simple test and predictive models for assessing swell potential of Ankara (Turkey) clays, Engineering Geology 67 (2003) 331–352. https://doi.org/10.1016/S0013-7952(02)00205-3.

[8] V.H.R. Barbosa, M.E.S. Marques, A.C.R. Guimarães, C.D. Castro, Characterization of an expansive soil in southwest Brazilian amazon—behavior of an expansive subgrade in a flexible pavement, In Advances in Transportation Geotechnics IV, Springer (2022) 323–335. https://doi.org/10.1007/978-3-030-77234-5_24.

[9] A.A. Basma, Prediction of expansion degree for natural compacted clays, Geotechnical Testing Journal 16 (1993) 542–549. https://doi.org/10.1520/GTJ10291J.

[10] I. Ashayeri, S. Yasrebi, Free-swell and swelling pressure of unsaturated compacted clays; experiments and neural networks modeling, Geotechnical and Geological Engineering 27 (2009) 137–153. https://doi.org/10.1007/s10706-008-9210-7.

[11] K. Meshram, N. Singh, P. Jain, Estimation of swelling characteristics of expansive soils with influence of clay mineralogy, Applied Clay Science 198 (2021) 105822. https://doi.org/10.1016/j.clay.2020.105822.

[12] M.A. Shahin, State-of-the-art review of some artificial intelligence applications in pile foundations, Geoscience Frontiers 7 (2016) 33–44. https://doi.org/10.1016/j.gsf.2014.10.006.

[13] S. Jong, D. Ong, E. Oh, State-of-the-art review of geotechnical-driven artificial intelligence techniques in underground soil–structure interaction, Tunnelling and Underground Space Technology 113 (2021) 103946. https://doi.org/10.1016/j.tust.2021.103946.

[14] W. Zhang, X. Gu, L. Hong, L. Han, L. Wang, Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges, Applied Soft Computing 136 (2023) 110066. https://doi.org/10.1016/j.asoc.2023.110066.

[15] A.T.C. Goh, Nonlinear modelling in geotechnical engineering using neural networks, Transactions of the Institution of Engineers, Australia. Civil Engineering 36 (1994) 293–297.

[16] Y.M. Najjar, H.E. Ali, CPT-based liquefaction potential assessment: A neuronet approach, Geotechnical Special Publication 75 (1998) 542–553.

[17] M. Lefik, Artificial neural network applications for geotechnical engineering, Applied Sciences 13 (2023) 2988. https://doi.org/10.3390/app13052988.

[18] J.J. Jeremiah, S.J. Abbey, C.A. Booth, A. Kashyap, Results of application of artificial neural networks in predicting geo-mechanical properties of stabilised clays—a review, Geotechnics 1 (2021) 147–171. https://doi.org/10.3390/geotechnics1010008.

[19] Q. Guan, Z. Yang, N. Guo, Z. Hu, Finite element geotechnical analysis incorporating deep learning-based soil model, Computers and Geotechnics 154 (2023) 105120. https://doi.org/10.1016/j.compgeo.2022.105120.

[20] F.E. Jalal, Y. Xu, M. Iqbal, M.F. Javed, B. Jamhiri, Predictive modeling of swell-strength of expansive soils using artificial intelligence approaches: ANN, ANFIS and GEP, Journal of Environmental Management 289 (2021) 112420. https://doi.org/10.1016/j.jenvman.2021.112420.

[21] S. Gahlot, R. Mangal, A. Arya, B.P. Sethy, K.C. Sethi, Prediction of swelling pressure of expansive soil using machine learning methods, Asian Journal of Civil Engineering 25 (2024) 3279–3292. https://doi.org/10.1007/s42107-024-01205-7.

[22] M. Shehab, R. Taherdangkoo, C. Butscher, Towards reliable barrier systems: A constrained XGBoost model coupled with gray wolf optimization for maximum swelling pressure of bentonite, Computers and Geotechnics 168 (2024) 106132. https://doi.org/10.1016/j.compgeo.2024.106132.

[23] O. Bahloul, F.Z. Tebbi, L. Lekouara, H. Bekhouche, Neural networks based linear (PCA) and nonlinear (ISOMAP) feature extraction for soil swelling pressure prediction (North East Algeria), Heliyon 9 (2023) e18673. https://doi.org/10.1016/j.heliyon.2023.e18673.

[24] I.G. Tejada, P. Antolin, Use of machine learning for unraveling hidden correlations between particle size distributions and the mechanical behavior of granular materials, Acta Geotechnica 17 (2022) 1443–1461. https://doi.org/10.1007/s11440-021-01327-x.

[25] N. Tiwari, N. Satyam, Coupling effect of pond ash and polypropylene fiber on strength and durability of expansive soil subgrades: An integrated experimental and machine learning approach, Journal of Rock Mechanics and Geotechnical Engineering 13 (2021) 1101–1112. https://doi.org/10.1016/j.jrmge.2021.03.010.

[26] L. Lekouara, O. Bahloul, F.Z. Tebbi, R. Rabehi, A swelling pressure prediction model based on boruta feature selection mila region, Studies in Engineering and Exact Sciences 5 (2024) e11613. https://doi.org/10.54021/seesv5n2-633.

[27] S.D. Mohammadzadeh, S.-F. Kazemi, A. Mosavi, E. Nasseralshariati, J.H.M. Tah, Prediction of compression index of fine-grained soils using a gene expression programming model, Infrastructures 4 (2019) 26. https://doi.org/10.3390/infrastructures4020026.

[28] M. Kumar, P. Samui, D.R. Kumar, P.G. Asteris, State-of-the-art XGBoost, RF and DNN based soft-computing models for PGPN piles, Geomechanics and Geoengineering 19 (2024) 320–335. https://doi.org/10.1080/17486025.2024.2307746.

[29] J. Khatti, K.S. Grover, Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models, Journal of Rock Mechanics and Geotechnical Engineering 15 (2023) 3010–3038. https://doi.org/10.1016/j.jrmge.2023.02.016.

[30] M. Kumor, Selected geotechnical problems of expansive clays in the area of Poland, Architecture Civil Engineering Environment 1 (2008) 75–92.

[31] R. Bag, A. Rabbani, Effect of temperature on swelling pressure and compressibility characteristics of soil, Applied Clay Science 136 (2017) 1–7. https://doi.org/10.1016/j.clay.2016.11.014.

[32] L. Yunlong, S. Vanapalli, Pile behavior modeling in unsaturated expansive soils, In Modeling in Geotechnical Engineering, Elsevier (2021) 393–427. https://doi.org/10.1016/B978-0-12-821205-9.00018-7.

[33] X. Chen, X. Shi, J. Zhou, X. Du, Q. Chen, X. Qiu, Effect of overflow tailings properties on cemented paste backfill, Journal of Environmental Management 235 (2019) 133–144. https://doi.org/10.1016/j.jenvman.2019.01.040.

[34] Application of artificial intelligence in geotechnical engineering, Geotechnical Engineering Research 40 (2022) 152–168.

[35] Y.M. Najjar, H.E. Ali, I.A. Basheer, On the use of neurons for simulating the stress-strain behavior of soils, Proceedings of the 7th International Conference on Neural Information Processing, Boulder, Colorado (1999).

[36] D. Penumadu, R. Zhao, Triaxial compression behavior of sand and gravel using artificial neural networks (ANN), Computers and Geotechnics 24 (1999) 207–230. https://doi.org/10.1016/S0266-352X(99)00002-4.

[37] G.W. Ellis, C. Yao, R. Zhao, D. Penumadu, Stress-strain modelling of sands using artificial neural networks, Journal of Geotechnical Engineering, ASCE 121 (1995) 429–435. https://doi.org/10.1061/(ASCE)0733-9410(1995)121:5(429).

[38] State-of-the-art review on AI-enhanced computational mechanics, Computational Mechanics and Engineering 45 (2024) 89–112.

[39] R. Taherdangkoo, A. Tatomir, T. Anighoro, M. Sauter, Modeling of hydraulic fracturing process in unconventional reservoirs: A review, Environmental Earth Sciences 82 (2023) 1–23. https://doi.org/10.1007/s12665-023-10803-1.

[40] P. Narmandakh, R. Taherdangkoo, M. Javadi, C. Butscher, Neural network model to determine maximum swelling pressure of clayey soils, Computers and Geotechnics 154 (2023) 105157. https://doi.org/10.1016/j.compgeo.2022.105157.

[41] M. Ozsagir, C. Erden, E. Bol, S. Sert, A. Özocak, Machine learning approaches for prediction of fine-grained soils liquefaction, Computers and Geotechnics 152 (2022) 105014. https://doi.org/10.1016/j.compgeo.2022.105014.

[42] M. Habib, B. Bashir, A. Alsalman, H. Bachir, Prediction and parametric assessment of soil one-dimensional vertical free swelling potential using ensemble machine learning models, Advanced Modeling and Simulation in Engineering Sciences 11 (2024) 26. https://doi.org/10.1186/s40323-024-00270-w.

[43] B. Alibrahim, A.H.B. Garoushi, E. Uygar, The role of calcium-based additives in bentonite stabilization: a comparative evaluation, Arabian Journal of Science and Engineering 49 (2024) 7751–7764. https://doi.org/10.1007/s13369-024-09281-2.

[44] A.A. Sabtan, Geotechnical properties of expansive clay shale in Tabuk, Saudi Arabia, Journal of Asian Earth Sciences 25 (2005) 747–757. https://doi.org/10.1016/j.jseaes.2004.07.003.

[45] X. Guo, H. Shi, W.A. Dick, L. Ping, X. Chen, Modeling the influence of lime on the unconfined compressive strength of reconstituted graded soil using advanced machine learning approaches for subgrade and liner applications, PLoS ONE 19 (2024) e0301075. https://doi.org/10.1371/journal.pone.0301075.

[46] S. Hanandeh, A. Ardah, M. Abu-Farsakh, Using artificial neural network and genetics algorithm to estimate the resilient modulus for stabilized subgrade and propose new empirical formula, Transportation Geotechnics 24 (2020) 100358. https://doi.org/10.1016/j.trgeo.2020.100358.

[47] M.R.H. Hossain, M.A. Kabir, Machine learning techniques for estimating soil moisture from smartphone captured images, Agriculture 13 (2023) 1–22. https://doi.org/10.3390/agriculture13030536.

[48] J. Khatti, K.S. Grover, Assessment of uniaxial strength of rocks: A critical comparison between evolutionary and swarm optimized relevance vector machine models, Structures 59 (2024) 107159. https://doi.org/10.1016/j.istruc.2024.107159.

[49] M. Kumar, V. Kumar, B.G. Rajagopal, P. Samui, A. Burman, State of art soft computing based simulation models for bearing capacity of pile foundation: A comparative study of hybrid ANNs and conventional models, Modeling Earth Systems and Environment 9 (2023) 2533–2551. https://doi.org/10.1007/s40808-022-01595-3.