Comparative Prediction of Unconfined Compressive Strength of Geopolymer-Stabilized Soil Using Interpretable and Black-Box Machine Learning Models

Mahmoud Ghazavi* and Mobina Taslimi Paein Afrakoti

Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran

Abstract: Accurate prediction of unconfined compressive strength (UCS) can reduce the experimental effort required for the design of geopolymer-stabilized soils. However, highly accurate machine-learning models often provide limited engineering transparency. This study addresses this issue by comparing gene expression programming (GEP), support vector regression (SVR), random forest (RF), and gradient boosting regression trees using LSBoost (GBRT). A database of 270 experimental records was used, with seven input variables describing soil plasticity, precursor contents, activator conditions, and chemical ratios. All models were evaluated using the same training–test split and held-out test set. GBRT achieved the highest predictive accuracy, with an independent-test R^2 of 0.9814, RMSE of 0.9771 MPa, and MAE of 0.6080 MPa. RF, SVR, and GEP followed with R^2 values of 0.9661, 0.9416, and 0.9023, respectively. Despite its lower accuracy, GEP was retained as the interpretable model because it provided an explicit mathematical expression involving all seven inputs. Its behavior was further examined using multi-point one-at-a-time (OAT) sensitivity analysis and exact background-averaged SHAP analysis. Both methods identified GGBS as the dominant variable, with normalized important values of 45.43% and 42.40%, respectively. SHAP further ranked PI and FA as the next most influential inputs. Higher GGBS generally increased predicted UCS, whereas higher PI and FA produced negative contributions within the investigated data domain. The proposed framework therefore distinguishes predictive superiority from model interpretability and provides a transparent basis for understanding UCS predictions in geopolymer-stabilized clayey soils.

Keywords: Geopolymer-stabilized soil, Unconfined compressive strength, Gene expression programming, Gradient boosting, SHAP, Sensitivity analysis, Interpretable machine learning.

1. INTRODUCTION

Problematic clayey soils are a major geotechnical concern because of their low bearing capacity, high compressibility, moisture sensitivity, and shrink–swell behavior [1-3]. These characteristics can adversely affect foundations, embankments, and pavements [1,4]. Cement and lime are widely used for soil stabilization, but their production is associated with high energy demand and considerable CO_2 emissions [3-4]. This has increased interest in lower-carbon binders and industrial by-products for ground improvement [2,5,6]. Recent studies have also explored nano-modified and hybrid stabilization systems to improve the mechanical response of problematic soils [7,8].

Geopolymers and alkali-activated binders are promising alternatives for this purpose. They can incorporate aluminosilicate-rich materials such as fly ash (FA) and ground-granulated blast-furnace slag (GGBS) [9-11]. Alkaline activation forms cementitious products that bind soil particles and improve the soil matrix [12-14]. Calcium-rich GGBS can further accelerate strength development [10,14,29]. However, the resulting performance depends on soil properties, precursor composition, activator chemistry, and mixture proportions [12,11,15,46].

Unconfined compressive strength (UCS) is a key index for evaluating stabilized soils. In geopolymer-treated clay, UCS depends on several interacting variables, including soil plasticity, GGBS and FA contents, alkaline molarity, activator-to-binder ratio, and the Na/Al and Si/Al ratios [10,14,11,16]. Their combined effects can be strongly nonlinear. UCS testing also requires specimen preparation, curing, and destructive loading. This becomes costly when many mixtures are evaluated. Data-driven models can reduce this experimental burden and capture relationships that conventional empirical equations may overlook [17-21].

Several machine-learning and soft-computing approaches have therefore been developed for UCS prediction. ANN and SVM were successfully applied by Mozumder and Laskar [17] and Mozumder *et al*. [22]. Explicit relationships were later developed using genetic programming and GEP [18,23,24]. Other studies confirmed the capability of ANN, SVM, random forest, and boosting-based models for stabilized soils [19,25,26,27]. More recent studies have combined prediction with explainability. Ngo *et al*. [28] used SHAP to interpret RF, ANN, XGBoost, and gradient boosting models. Abdullah *et al*. [29] reported strong performance for gradient boosting on a 270-record geopolymer-stabilized clay database. SHAP, sensitivity analysis, and partial dependence methods have also been increasingly used to explain model behavior

*Address correspondence to this author at the Faculty of Civil Engineering, K. N. Toosi University of Technology, Tehran, Iran;
E-mail: ghazavi_ma@kntu.ac.ir

Received on 31-08-2026 Accepted on 28-10-2026 Published on 11-11-2026

E-ISSN: 3083-3590/26

[28-31]. Related ANN and optimization applications in civil and geotechnical engineering have also been reported for pipeline response, seismic amplification, and concrete behavior [32-35].

These developments highlight an accuracy–interpretability trade-off. In this study, SVR, RF, and GBRT are classified as relatively black-box models because their fitted prediction rules are not readily represented by compact, human-readable symbolic equations. In contrast, GEP is classified as an interpretable model because its explicit mathematical expression and expression tree can be examined directly. Post-hoc methods such as SHAP can explain feature contributions, but they do not convert the fitted model into a transparent mathematical equation [36,37,38]. GEP offers a complementary advantage because it can generate an explicit symbolic expression and an expression tree [18, 23,39]. Recent studies using interpretable and soft-computing models have also emphasized explicit predictive formulations, sensitivity assessment, and engineering interpretability [40-44].

Ngo *et al*. [28] developed RF, ANN, XGBoost, and gradient boosting models to predict the 28-day UCS of geopolymer-stabilized cohesive soils. SHAP was used to interpret the selected models and guide feature selection. Abdullah *et al*. [29] evaluated gradient boosting and AdaBoost models for UCS prediction and applied cosine amplitude analysis to assess input importance. Neither study examined an explicit GEP equation through complementary OAT and SHAP analyses. OAT measures the model response to controlled changes in individual inputs, whereas SHAP assigns feature contributions to predictions relative to a defined background dataset. Their application to the same symbolic equation enables a direct comparison of perturbation-based and attribution-based interpretations. The main methodological contribution of the present study therefore lies in integrating comparative predictive evaluation with complementary interpretation of an explicit GEP equation, rather than introducing new machine learning algorithms or explanation techniques.

Accordingly, this study compares four models for UCS prediction: gene expression programming (GEP), support vector regression (SVR), random forest (RF), and gradient boosting regression trees using LSBoost (GBRT). All models are evaluated using the same training–test split and held-out test set. Model performance was further examined through a common five-fold evaluation of the training data. A supplementary nested stratified five-fold cross-validation was also performed to assess predictive performance across partitions stratified by PI levels and within-group UCS quantiles. The most accurate model is identified separately from the model selected for interpretation. GEP is retained for this purpose because it provides an explicit mathematical equation and expression tree. Its behavior is then examined using multi-point one-at-a-time (OAT) sensitivity analysis and exact background-averaged SHAP analysis. This approach provides a combined assessment of predictive accuracy, model stability, and the interpretability of the symbolic model.

2. DATABASE AND PRELIMINARY STATISTICAL ANALYSIS

The database contained 270 experimental records of unconfined compressive strength (UCS) for geopolymer-stabilized clayey soils. Seven variables were used as model inputs: plasticity index (PI), ground-granulated blast-furnace slag content (GGBS), fly ash content (FA), molar concentration of the alkaline solution (M), alkali solution-to-binder ratio (A/B), sodium-to-aluminum atomic ratio (Na/Al), and silicon-to-aluminum atomic ratio (Si/Al). UCS was used as the output variable. GGBS and FA were expressed as percentages by dry weight of soil, M was reported in mol/L, the three ratios were dimensionless, and UCS was expressed in MPa. The database was derived from the experimental dataset reported by Abdullah *et al*. [29]. It was selected because it covers variations in soil plasticity, precursor contents, and alkaline activation conditions. This diversity provides a suitable basis for comparing the predictive performance of the selected models under a common data framework.

To describe the range and dispersion of the variables, the standard deviation, variance, coefficient of variation, minimum, and maximum were calculated. The results are summarized in Table **1**. As seen, the database covers a broad range of soil characteristics and geopolymer mixture conditions. FA exhibited the highest coefficient of variation among the inputs (229.37%), while UCS also showed substantial variability (112.90%). In contrast, M, A/B, and Si/Al varied within relatively narrower ranges. This spread provides sufficient variation for developing and evaluating predictive models.

2.1. Distribution of the Database Variables

The distributions of the seven input variables and UCS are shown in Figure **1**. The histograms were generated directly from the 270 records used in the present modeling database.

Figure **1** shows that several variables do not follow continuous or normal distributions. PI occurs at three discrete levels, while M and A/B also reflect predefined experimental levels. FA is strongly concentrated near

Table 1: Descriptive Statistical Characteristics of the Input Variables and Unconfined Compressive Strength

Variable	SD	Variance	CV (%)	Minimum	Maximum
PI (%)	30.9702	959.155	77.38	14.07	88.46
GGBS (%)	12.9619	168.010	79.90	0	50
FA (%)	4.28154	18.3316	229.37	0	20
M (mol/L)	2.73920	7.50322	22.25	4	15
A/B (-)	0.145038	0.0210359	23.44	0.45	0.85
Na/Al (-)	0.447123	0.199919	38.40	0.24	1.98
Si/Al (-)	0.332089	0.110283	19.80	1.49	2.49
UCS (MPa)	6.46360	41.7781	112.90	0	24.26

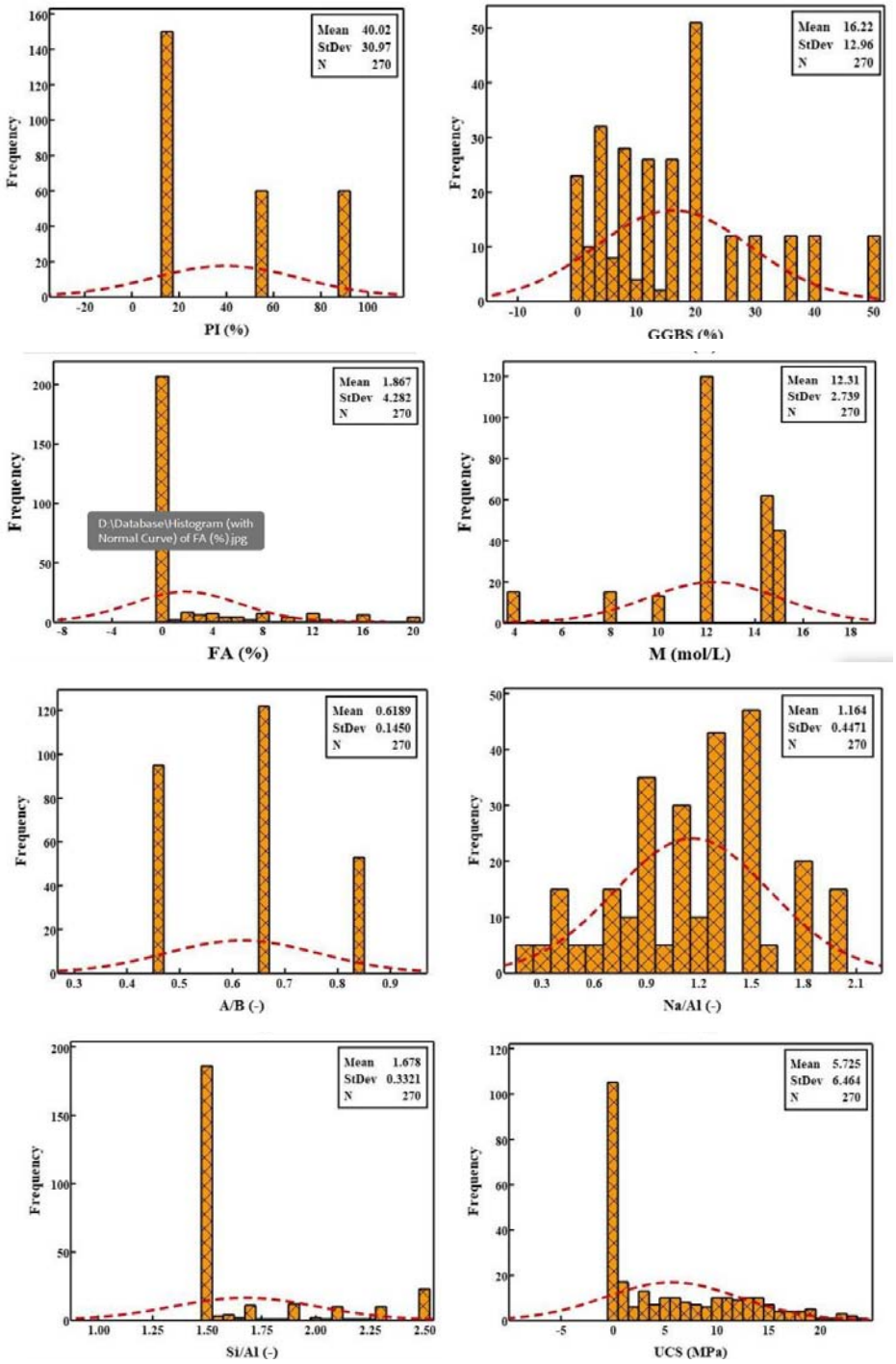


Figure 1: Distributions of the seven input variables and UCS, highlighting discrete experimental levels, the predominance of zero FA content, and the right-skewed UCS distribution.

zero, and UCS is positively skewed toward the lower-strength range. These patterns mainly reflect the structure of the experimental program and the presence of zero-content mixtures. Accordingly, model development should account for non-uniform sampling across the variable ranges.

Pearson correlation analysis was used to examine the pairwise linear relationships among the input variables and UCS. The resulting correlation matrix is presented in Figure **2**.

As shown in Figure **2**, GGBS had the strongest linear relationship with UCS (r=0.79). This indicates that higher GGBS contents were generally associated with higher UCS values within the investigated dataset. The remaining correlations with UCS were weaker: FA (-0.29), Na/Al (0.26), Si/Al (-0.25), PI (-0.20), M (0.16), and A/B (0.05). These small coefficients should not be interpreted as evidence of negligible influence because Pearson correlation captures only pairwise linear dependence.

Among the inputs, the strongest positive correlation was observed between FA and Si/Al (r=0.83). A relatively high correlation was also found between A/B and Na/Al (r=0.71). These relationships indicate that some chemical and mixture parameters were not varied independently in the experimental program and should therefore be considered when interpreting model-based importance measures.

The VIF analysis further assessed multicollinearity among the input variables, particularly because several variable pairs showed relatively high correlations. The calculated VIF values are presented in Figure **3**.

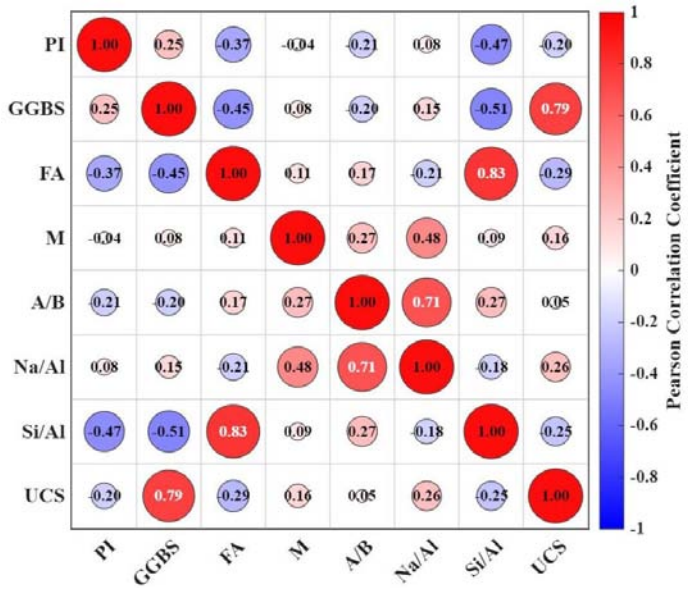


Figure 2: Pearson correlation matrix showing the strongest UCS association with GGBS (r = 0.79) and notable correlations between FA–Si/Al and A/B–Na/Al.

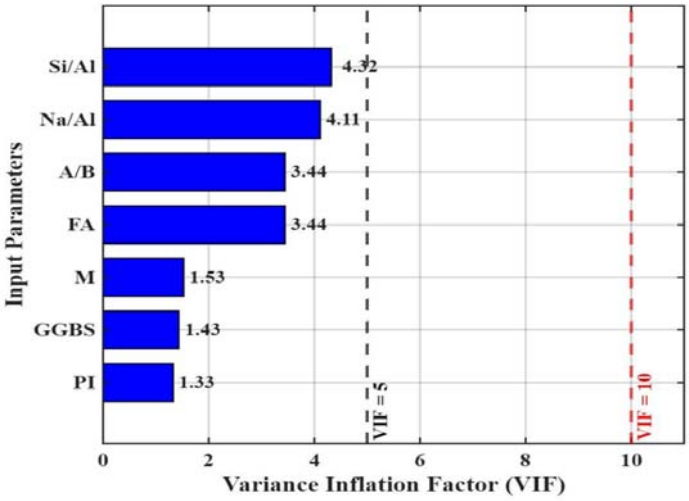


Figure 3: Variance inflation factors of the input variables. All values remain below 5, although Si/Al and Na/Al show relatively higher multicollinearity.

Figure **3** shows that all VIF values were below the conservative threshold of 5. The highest values were obtained for Si/Al (4.32) and Na/Al (4.11), followed by A/B and FA (approximately 3.44). These results do not indicate severe multicollinearity in the input space. However, the VIF values close to 4, together with the Pearson correlations, confirm moderate structural dependence among some variables. Therefore, the inputs should not be considered fully independent. Since no critical multicollinearity was identified, all seven variables were retained for model development.

Overall, the database provides broad variation in soil plasticity, geopolymer constituents, chemical ratios, and UCS. At the same time, the non-uniform distributions and moderate dependence among some inputs should be considered during model validation and interpretation. These characteristics also motivate the use of nonlinear predictive models and model-interpretation tools in the subsequent analyses.

2.2. Model Development and Hyperparameter Optimization

All models were implemented in MATLAB using a common training–test split (216/54 observations) and a random seed of 42. Hyperparameter optimization was restricted to the training data. SVR was standardized, whereas RF and GBRT were trained without input scaling. Bayesian optimization with five-fold cross-validation was used for SVR, RF, and GBRT. The maximum numbers of objective evaluations were 30 per kernel, 15, and 30, respectively. The expected-improvement-plus acquisition function was employed. Linear, Gaussian, and second-order polynomial kernels were examined for SVR. The Gaussian kernel was selected. RF was developed using bootstrap aggregation, whereas GBRT employed LSBoost. The search domains and final hyperparameters are reported in Table **2**.

GEP was implemented as a three-gene evolutionary model with additive gene linking. Each gene had a head length of 8 and a tail length of 9. The population comprised 100 individuals, and 10 independent runs were performed. Each run was limited to 500 generations, with early stopping after 120 generations without improvement. The training data were divided into 172 development and 44 validation observations for model selection. The tenth run was selected and terminated after 305 generations. Numerical constants were subsequently refined using the Nelder–Mead method for up to 300 iterations. The evolutionary operators, mutation and crossover rates, and remaining settings are provided in Table **3**. The held-out test set was excluded from model optimization and selection.

2.3. Supplementary Nested Stratified Cross-Validation

The original 80/20 training–test split was retained to provide a consistent held-out benchmark for all four models. Given the discrete experimental levels and positively skewed UCS distribution, a supplementary nested stratified five-fold cross-validation was performed using the 216 training observations. The outer folds were stratified according to PI levels and approximate UCS quantiles within each PI group. Three-fold inner cross-validation was used for hyperparameter selection. All four models were independently retrained within each outer fold. The original 54-observation held-out test set remained excluded from this procedure.

The supplementary analysis was implemented using a separate hyperparameter optimization protocol. Eight Bayesian optimization evaluations were performed independently for each SVR kernel, while six evaluations were conducted for RF and GBRT, respectively. Two GEP hyperparameter configurations were examined using three-fold inner cross-validation.

Table 2: Hyperparameter Search Ranges and optimal Settings of SVR, RF, and GBRT Models

Model	Hyperparameter	Search range	Optimal value
SVR	Kernel	Linear, Gaussian, Polynomial	Gaussian
	Box constraint (C)	0.001–1000	27.0788
	Epsilon (ε)	0.00777168–777.168	0.0125
	Kernel scale	0.001–1000	2.9562
RF	Number of trees	80–400	390
	Min. leaf size / Max. splits	1–25 / 2–215	1 / 71
	Predictors per split	1–7	5
GBRT	Learning cycles	50–800	758
	Learning rate	0.01–0.30	0.1417
	Min. leaf size / Max. splits	1–25 / 1–128	3 / 13

Table 3: Evolutionary Settings and Stopping Criteria of the GEP Model

Parameter	Final setting
Genes / Head / Tail	3 / 8 / 9
Population size	100
Function set / Linking	+, −, × / Addition
Tournament size / Elitism	3 / 2
Symbol / Constant mutation rates	0.044 / 0.10
One-point crossover / Gene recombination	0.35 / 0.15
IS transposition rate	0.10
Numerical constants	10 per gene; [−10, 10]
Independent runs	10
Maximum generations / Stall limit	500 / 120
Selected run / Generations	10 / 305
Development / Validation samples	172 / 44
Constant optimization	Nelder–Mead; max. 300 iterations
Output constraint	UCS ≥ 0

Each inner GEP fit involved one evolutionary run with a population of 35 and a maximum of 45 generations. For each outer fold, GEP was independently re-evolved using two evolutionary runs, a population of 55, and a maximum of 110 generations. Model performance was evaluated across the five outer folds and summarized using the mean and standard deviation of R^2, RMSE, and MAE.

3. COMPARATIVE PREDICTIVE PERFORMANCE OF THE MODELS

The predictive performance of GEP, SVR, RF, and GBRT was compared using the same fixed training-test split and the same held-out test set. The independent-test metrics used for this comparison are presented in Figure **4**.

As shown in Figure **4**, GBRT achieved the highest independent-test accuracy, with R^2=0.9814, RMSE=0.9771 MPa, and MAE=0.6080 MPa. RF ranked second, with R^2=0.9661 and RMSE=1.3202 MPa. SVR also showed strong performance, with R^2=0.9416 and RMSE=1.7322 MPa. GEP obtained R^2=0.9023 and RMSE=2.2405 MPa. Thus, when predictive accuracy alone is considered, the overall ranking based on R^2, RMSE, MAE, residual SD, and scatter index (SI) is GBRT, RF, SVR, and GEP, respectively. Abdullah *et al*. [29] reported a test R^2 of 0.980 for gradient boosting using the same 270-record database. Ngo *et al*. [28] obtained an R^2 of 0.9808 using ANN on a related 282-record dataset. These results are consistent with the strong performance of nonlinear models observed here. However, overlapping source data and differences in predictors, model optimization, and data partitioning prevent claims of cross-study superiority.

GBRT also produced the lowest residual dispersion. Its residual SD was 0.9238 MPa, compared with 1.2917, 1.7089, and 2.2605 MPa for RF, SVR, and GEP, respectively. SI followed the same trend, decreasing from 0.3855 for GEP to 0.1681 for GBRT. MAPE was the only metric that did not strictly follow this ranking; RF yielded 130.34%, whereas GBRT yielded 143.97%. Because several measured UCS values are zero or close to zero, percentage-based errors can become unstable when the denominator is small [47]. Therefore, R^2, RMSE, MAE, residual SD, and SI were treated as the primary comparison metrics, while MAPE was considered supplementary.

Model stability was further examined through a common five-fold evaluation performed only on the 216 training samples. The 54 independent-test samples were excluded, and identical fold assignments were used for all four models. The fold-wise results are presented in Figure **5**.

Figure **5** shows that SVR achieved the highest mean five-fold R^2 of 0.9580±0.0159, followed closely by GBRT with 0.9524±0.0255. Their corresponding mean RMSE values were 1.2562±0.2907 and 1.3098±0.2859 MPa. RF reached a mean R^2 of 0.9291±0.0314, whereas GEP obtained 0.8326±0.0293. Hence, GBRT provided the strongest independent-test accuracy, while SVR showed slightly better average fold-wise performance and lower variation in R^2. These results support evaluating the models using both independent-test accuracy and fold-wise stability.

For an additional view of overall agreement, the measured-versus-predicted plots for all 270 observations are presented in Figure **6**. The dashed

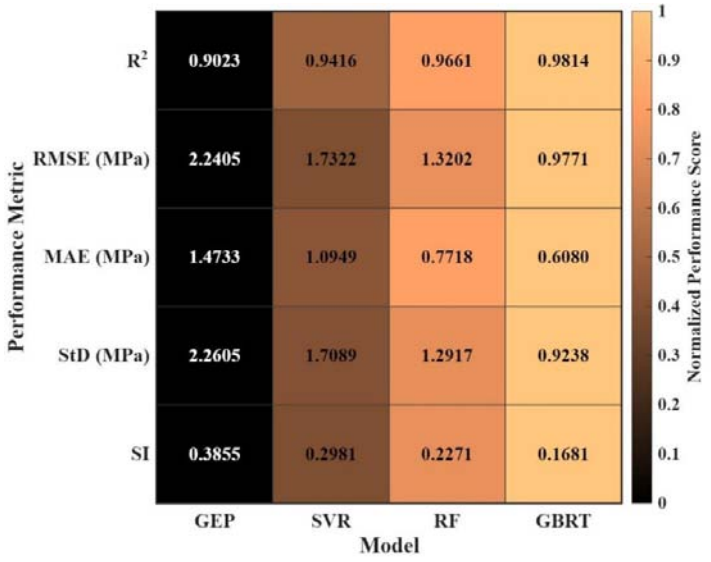


Figure 4: Independent-test performance of the four models. GBRT achieves the highest accuracy and lowest prediction errors, while GEP shows lower accuracy. Cell colors are normalized separately for each metric.

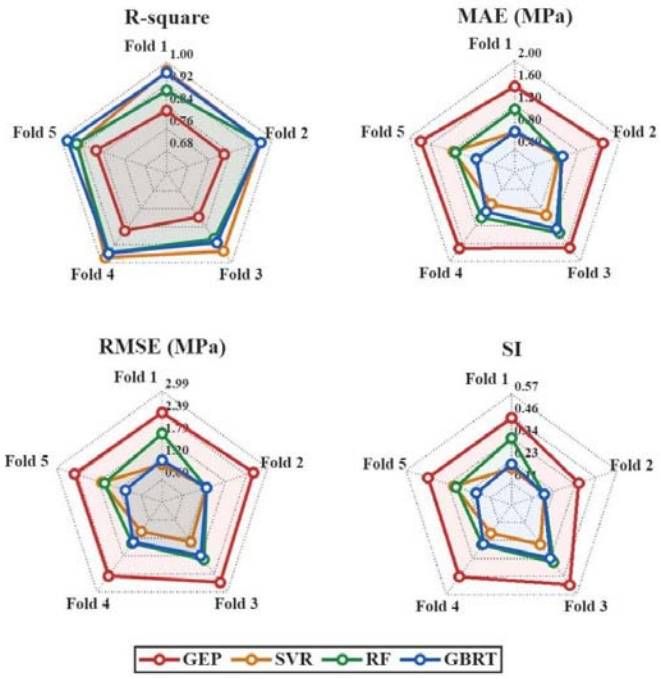


Figure 5: Five-fold performance comparison of GEP, SVR, RF, and GBRT. SVR achieves the highest mean R², closely followed by GBRT, whereas GEP shows lower fold-wise accuracy.

black line denotes perfect 1:1 agreement, and the red line represents the fitted regression line.

Figure **6** shows higher overall agreement for GBRT, RF, and SVR. The all-data R² values were 0.9951, 0.9755, 0.9649, and 0.8561 for GBRT, RF, SVR, and GEP, respectively. However, these plots combine training and test observations. They were therefore used only to visualize overall model fit and were not treated as the primary evidence of generalization. The held-out test set and the common five-fold evaluation were retained as the main bases for model comparison.

3.1. Supplementary Nested Stratified Five-Fold Validation

The supplementary nested stratified five-fold results are summarized in Table **4**. GBRT achieved the highest mean R² (0.9555 ± 0.0169) and the lowest mean RMSE (1.2838 ± 0.2370 MPa). It also achieved the lowest RMSE in all five outer folds. RF and SVR obtained mean R² values of 0.9146 and 0.9002, respectively. GEP exhibited the lowest mean R² (0.8112 ± 0.0894) and the highest mean RMSE (3.0668 ± 0.6653 MPa).

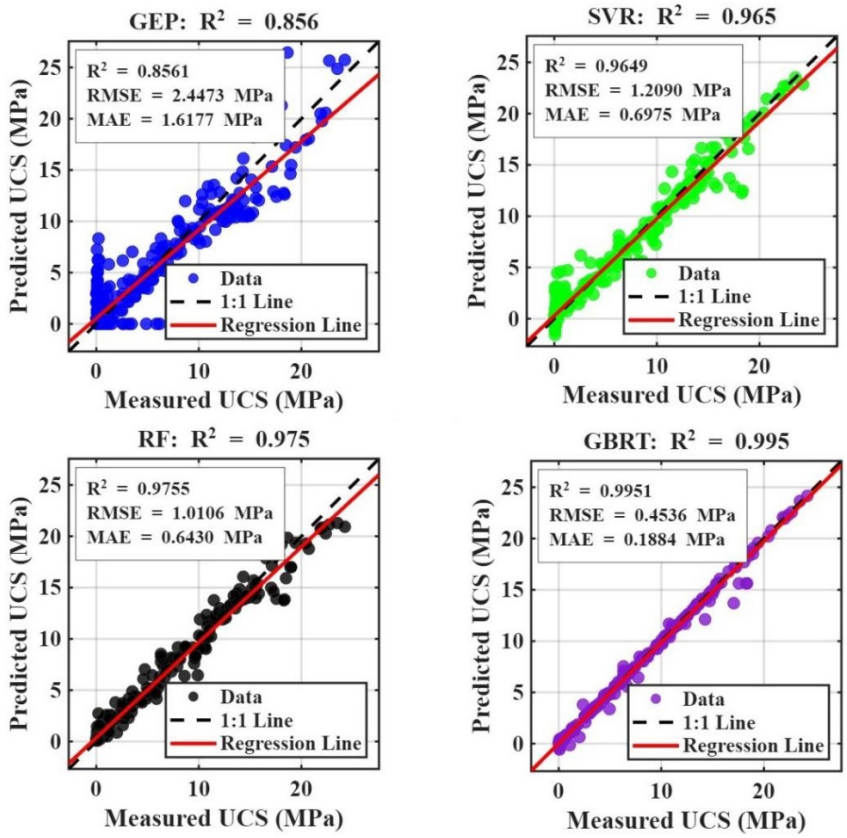


Figure 6: Measured versus predicted UCS for all 270 observations. GBRT shows the closest overall agreement with the 1:1 line. Dashed black lines indicate perfect agreement, and red lines represent fitted regressions.

The supplementary results differed from those obtained using the original five-fold evaluation. In the original analysis, SVR achieved a slightly higher mean R^2 than GBRT (0.9580 versus 0.9524). In contrast, GBRT achieved the highest mean R^2 (0.9555 ± 0.0169) in the supplementary stratified analysis. The two procedures differed in fold allocation, model selection, and optimization settings. Their results therefore provide complementary evidence of predictive performance under different validation protocols.

GEP showed lower predictive accuracy in the supplementary analysis than on the original held-out test set. The supplementary GEP models were independently re-evolved within each outer fold using the corresponding training observations and optimization settings. Thus, these models differed from the final GEP equation developed using the original training set. The observed performance differences may reflect variation in data partitioning and model fitting, and their individual contributions cannot be determined from the present comparison.

Overall, GBRT achieved the lowest RMSE in all five supplementary outer folds. This finding is consistent with its leading predictive performance on the original held-out test set. The results provide further evidence of the relative predictive advantage of GBRT across the evaluated data partitions.

3.2. Accuracy–Interpretability Trade-off and Selection of GEP

Although GBRT achieved the highest numerical accuracy, the objective of this study is not limited to minimizing prediction error. A second objective is to obtain an explicit, data-driven predictive equation relating the input variables to UCS. RF, GBRT, and SVR are commonly used to achieve more accurate predictions, although their prediction mechanisms are not readily expressed as a single explicit equation. In contrast, GEP was developed as a symbolic evolutionary approach that can represent the final model as an explicit mathematical expression [39].

Accordingly, GEP was not selected as the most accurate model. It was selected as the interpretable model for the subsequent analysis. This distinction is important: GBRT is identified as the best predictive model, whereas GEP is retained as the selected interpretable model. This strategy is consistent with interpretable machine learning, in which a model is used not only for prediction but also for extracting useful knowledge about relationships learned from data [37].

The independent-test performance of GEP, with R^2=0.9023, indicates that the explicit model still preserves substantial predictive information despite its

Table 4: Supplementary Nested Stratified Five-Fold Cross-Validation Results (Mean ± SD)

Model	R^2	RMSE (MPa)	MAE (MPa)
GEP	0.8112 ± 0.0894	3.0668 ± 0.6653	2.2264 ± 0.5187
SVR	0.9002 ± 0.0663	1.8665 ± 0.5875	1.4547 ± 0.5136
RF	0.9146 ± 0.0340	1.7967 ± 0.4708	1.2593 ± 0.3530
GBRT	0.9555 ± 0.0169	1.2838 ± 0.2370	0.8167 ± 0.1542

lower accuracy relative to the black-box models. The performance gap between GEP and GBRT or RF also illustrates the practical accuracy-interpretability trade-off in the present dataset. GEP is therefore retained as a deliberate choice for engineering interpretation rather than as a replacement for the model with the lowest prediction error.

Based on this accuracy–interpretability trade-off, the final GEP model was retained for the interpretation stage. Its explicit mathematical formulation and expression-tree structure are presented first, followed by sensitivity and SHAP analyses.

3.3. Explicit GEP Formulation and Expression-tree Representation

The selected GEP model consists of three genes linked by addition. A non-negativity constraint was applied after gene summation because negative UCS values are not physically meaningful. The final prediction is written as:

$$\widehat{UCS} = \max(0, G_1 + G_2 + G_3)$$
$$G_1 = (z_{GGBS} - z_{FA}) \times 5.9618 - (z_{GGBS} + 1.9672) \times z_{PI} \quad (1)$$
$$G_2 = z_{A/B} - (z_{PI} - 3.2946) + z_M$$
$$G_3 = z_{Na/Al} \times z_{Si/Al}$$

In Eq. (1), (zi) denotes the standardized value of input (xi), calculated using the training-derived center and scale parameters:

$$z_i = \frac{x_i - Center_i}{Scale_i} \quad (2)$$

The corresponding training-derived center and scale parameters are presented in Table **5**. The original training-derived center and scale values must be retained when the GEP equation is applied to new observations. The equation has not been externally validated using independent experimental datasets. Therefore, its predictive applicability to unseen soil–geopolymer mixtures remains unverified. The GEP equation represents an empirical symbolic regression relationship learned from experimental data rather than a physically derived constitutive model. Its mathematical terms should not be interpreted as direct representations of soil-stabilization mechanisms.

The complete expression-tree representation of the final GEP model is presented in Figure **7** showing how the three genes are combined to obtain the final UCS prediction. Gene 1 contains GGBS, FA, and PI; Gene 2 contains A/B, PI, and M; and Gene 3 contains Na/Al and Si/Al. Thus, all seven input variables participate in the selected symbolic expression. PI appears in two genes, so its contribution is embedded in more than one part of the equation. However, the presence or repeated occurrence of a variable in the expression tree should not be interpreted as a global importance ranking. Its effect also depends on the numerical coefficients, interactions with other variables, and the range of the observed data.

The explicit equation and expression tree describe the internal structure of the selected GEP model, but they do not by themselves quantify how strongly each variable affects predictions over the full database. Sensitivity and SHAP analyses were therefore used as complementary interpretation tools.

3.4. Interpretability Analysis of the Selected GEP Model

The explicit GEP formulation provides a direct mathematical representation of the relationship between the input variables and predicted UCS. However, the presence of a variable in the equation alone does not fully describe its relative influence or the direction of its effect across the input space. Therefore, two complementary interpretation approaches were applied to the selected GEP model. First, a classical sensitivity analysis was used to examine the change in predicted UCS under controlled variations of each input variable. Second, SHAP analysis was employed to quantify the contribution of each variable to the model predictions across the database. Together, these analyses provide both local response information and global feature-level interpretation of the GEP model.

3.4.1. Multi-point one-at-a-time (OAT) Sensitivity Analysis

A multi-point one-at-a-time (OAT) sensitivity analysis was applied to the selected GEP model to

Table 5: Center and Scale Parameters Used to Standardize the Inputs in the Final GEP Model

Variable	Center	Scale
PI (%)	40.7352	31.2825
GGBS (%)	16.5581	13.2317
FA (%)	2.0930	4.4855
M (mol/L)	12.1599	2.7587
A/B (-)	0.6140	0.1462
Na/Al (-)	1.1326	0.4543
Si/Al (-)	1.6797	0.3363

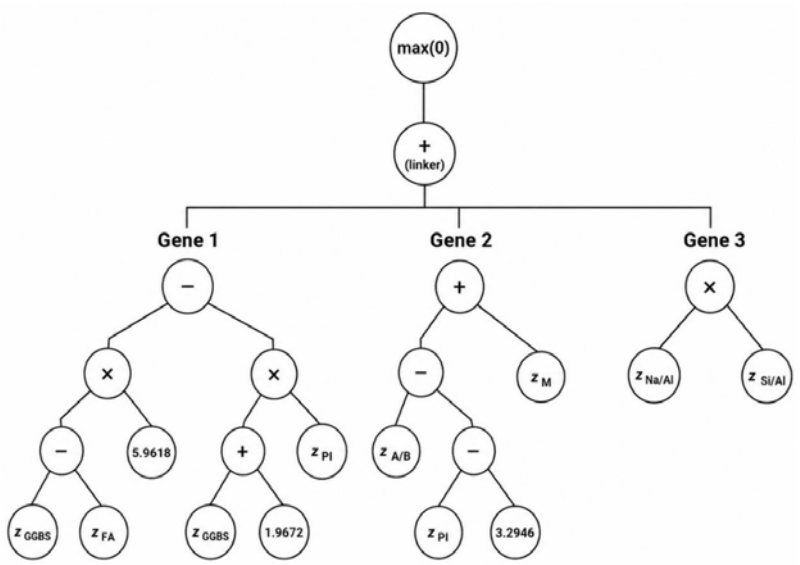


Figure 7: Expression-tree representation of the selected GEP model, showing three linked genes involving all seven inputs and a non-negativity constraint on predicted UCS.

examine how its predicted UCS responds to controlled changes in the input variables. The analysis used only the 216 training observations, while the 54 independent-test observations were excluded. Each training observation was treated as an individual baseline. One input was perturbed at a time by ±1, ±5, ±10, and ±20% of its observed training range, while the remaining six inputs were kept unchanged. Perturbed values were restricted to the training domain to avoid extrapolation beyond the available data.

The mean signed change in predicted UCS for each input is presented in Figure **8**. The shaded regions show ±1 standard deviation across the training observations and indicate how strongly the OAT response depends on the selected baseline sample.

Figure **8** shows that GGBS produced the strongest positive response. Increasing GGBS increased the predicted UCS, whereas reducing GGBS decreased it. FA showed the second strongest response, but in the opposite direction. Higher FA generally reduced the predicted UCS, and the reduction became more pronounced at larger positive perturbations. PI also showed a clear negative response. In contrast, M and A/B produced positive changes of smaller magnitude.

Na/Al and Si/Al showed weaker average responses. Increasing Na/Al slightly reduced the predicted UCS, whereas Si/Al produced a small positive response. These variables enter the final GEP equation through the interaction term zNa/Al × zSi/Al, therefore, their individual effects depend on the value of the paired variable. The wider uncertainty bands for GGBS and FA at larger perturbations also show that response magnitude varies across the training observations.

To compare the overall sensitivity of the seven variables, the mean absolute UCS change was aggregated over all training observations and perturbation levels and then normalized to 100%. The resulting importance ranking is shown in Figure **9**.

As shown in Figure **9**, GGBS had the highest normalized OAT importance at 45.43%, followed by FA at 26.10%. Together, these two variables accounted for 71.53% of the total normalized sensitivity. PI ranked third with 10.84%, followed by M with 6.95%. Na/Al and

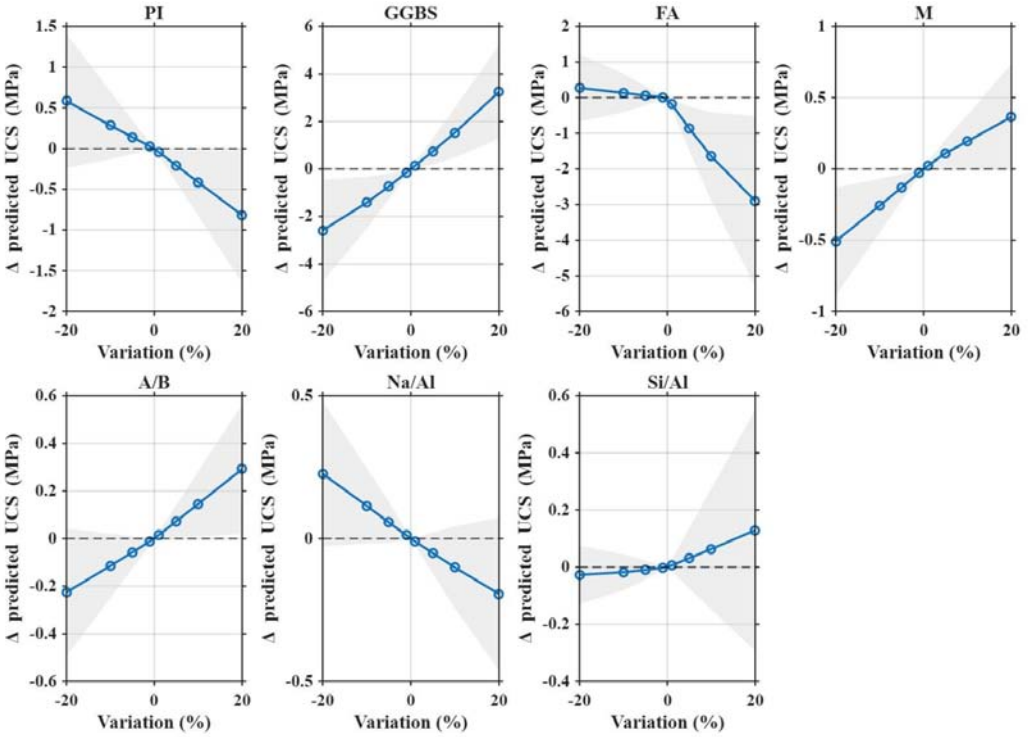


Figure 8: Mean OAT responses of the GEP model across 216 training observations. GGBS produces the strongest positive response, whereas FA and PI show predominantly negative responses.

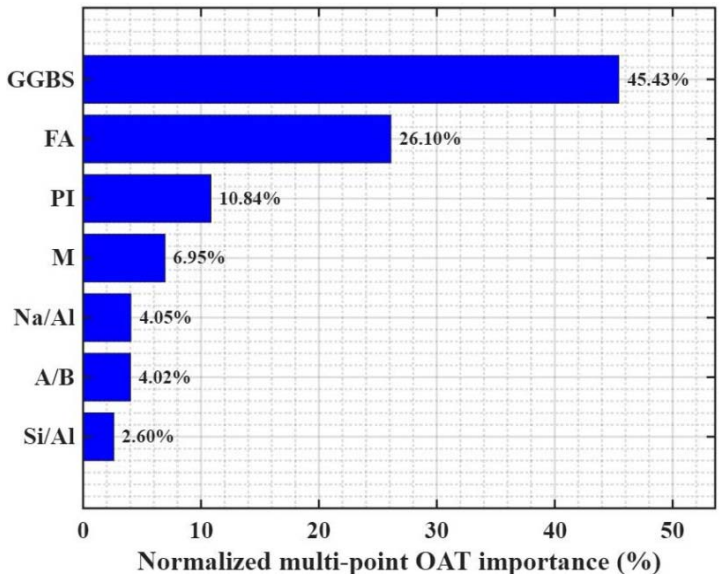


Figure 9: Normalized multi-point OAT importance of the GEP inputs. GGBS ranks first (45.43%), followed by FA (26.10%) and PI (10.84%).

A/B showed similar contributions of 4.05% and 4.02%, respectively, while Si/Al had the lowest importance at 2.60%.

These results indicate that the selected GEP model is most sensitive to GGBS and FA within the investigated training domain. The OAT ranking describes the response of the fitted GEP equation under controlled perturbations and should not be interpreted as physical causality. The subsequent SHAP analysis provides a complementary assessment of feature contributions across individual predictions.

3.4.2. SHAP-based interpretation of the GEP model

SHAP analysis was applied to the final GEP model to quantify the contribution of each input variable to the predicted UCS. The 216 training observations were used as the empirical background distribution, while SHAP values were calculated for all 270 observations. Because the model contains seven inputs, all 128 possible feature coalitions were evaluated. The expected GEP prediction over the background data was 5.411 MPa and was used as the SHAP base value. The maximum reconstruction error was 3.55×10^{-14}

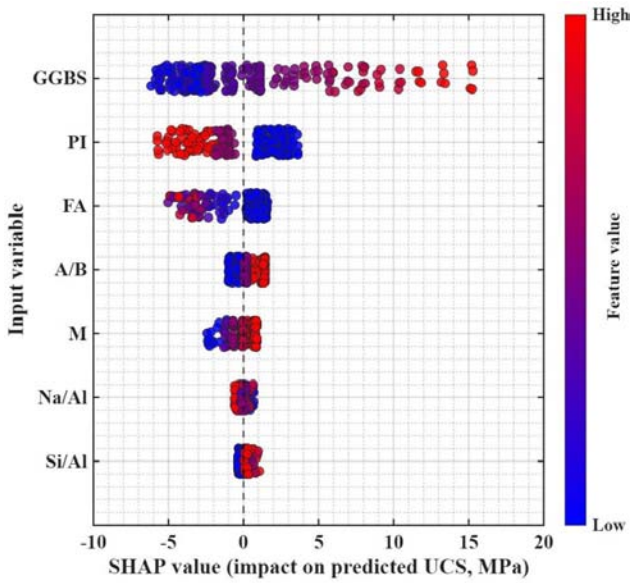


Figure 10: SHAP summary plot of the GEP model. Higher GGBS values generally increase predicted UCS, whereas higher PI and FA values show predominantly negative contributions. Colors indicate feature values.

MPa. This confirms the numerical consistency of the SHAP decomposition.

The overall distribution and direction of the SHAP contributions are shown in Figure **10**. Positive SHAP values increase the predicted UCS relative to the background prediction, whereas negative values reduce it. The color scale represents the corresponding feature value, from low values in blue color to high values in red color.

Figure **10** shows that GGBS has the widest SHAP range and the strongest influence on the model output. Low GGBS contents are mainly associated with negative SHAP values. High GGBS contents shift the prediction toward positive values. This indicates a strong positive contribution of GGBS to the predicted UCS within the investigated domain.

PI shows the opposite pattern. Low PI values generally produce positive contributions, while high PI values are associated with negative SHAP values. Thus, increasing PI tends to reduce the UCS predicted by the GEP model. FA also shows a mainly negative response. Higher FA contents are concentrated on the negative side of the SHAP axis, whereas low FA contents have smaller or positive contributions.

A/B and M have smaller effects than GGBS, PI, and FA. Higher A/B values generally increase the predicted UCS, and M shows a similar positive tendency. The SHAP values of Na/Al and Si/Al remain closer to zero. Their lower individual contributions are consistent with their appearance in the GEP equation through an interaction term. Their effects should therefore be interpreted together rather than as fully independent responses.

The global importance of all seven inputs was quantified using the mean absolute SHAP value. The normalized results are presented in Figure **11**.

Figure **11** illustrates that GGBS is the dominant input, with a normalized SHAP importance of 42.40%. Under the original empirical sampling distribution, PI ranked second at 25.22%, followed by FA at 15.89%. In together, these three variables account for about 83.5% of the total SHAP importance. A/B and M contribute 6.67% and 5.48%, respectively. Na/Al and Si/Al have smaller contributions of 2.42% and 1.92%. The ranking is also broadly consistent between the training and independent-test subsets. For example, the normalized importance of GGBS is 41.01% in the training data and 47.57% in the independent test data. PI remains the second most influential variable in both subsets. This agreement indicates that the main feature hierarchy is not limited to the training observations. Ngo *et al*. [28] also identified GGBFS as the dominant SHAP predictor, followed by liquid limit and A/B. Abdullah *et al*. [29] similarly ranked GGBS first, but their cosine-amplitude analysis assigned greater importance to Na/Al and M. In contrast, the present GEP-based SHAP analysis ranked PI and FA second and third. These differences reflect distinct predictor sets, model structures, interpretation methods, and sampling distributions.

To examine the response patterns in more detail, SHAP dependence plots were generated for the four most influential variables. These results are shown in Figure **12**. The point colors indicate the value of a selected interacting variable.

The GGBS dependence plot shows a clear increase in SHAP contribution as GGBS content increases. At

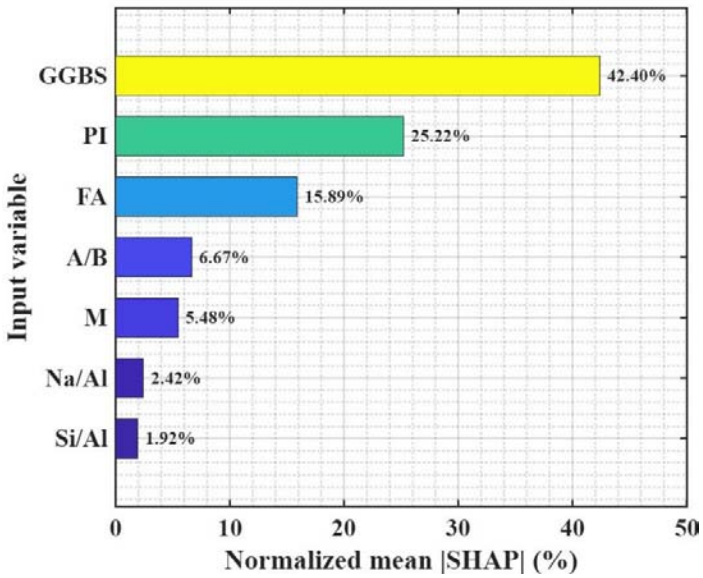


Figure 11: Normalized global SHAP importance of the GEP inputs. GGBS dominates (42.40%), followed by PI (25.22%) and FA (15.89%).

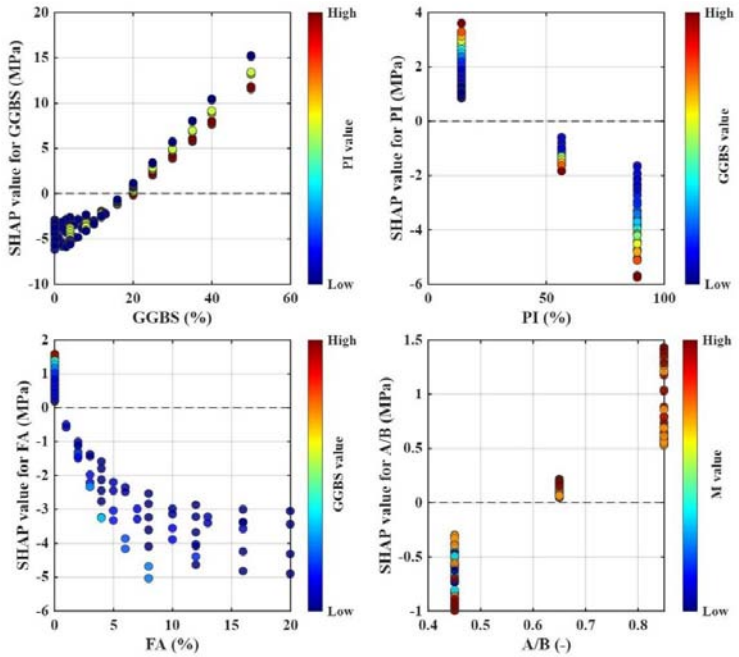


Figure 12: SHAP dependence plots showing increasing model contributions with GGBS, predominantly negative responses to higher PI and FA, and positive contributions at higher A/B levels. Colors indicate interacting feature values.

low contents, the contributions are predominantly negative. They become positive at intermediate levels and rise strongly toward the upper end of the observed range. The color pattern also indicates that PI modifies this response, which is consistent with the GGBS-PI interaction in the explicit GEP equation.

PI exhibits a clear inverse response. The lowest PI level is associated with positive SHAP values, whereas intermediate and high PI values produce increasingly negative contributions. The color distribution again indicates interaction with GGBS. FA shows a predominantly negative response as its content increases. In contrast, A/B displays a positive stepwise response. The clustered A/B pattern mainly reflects the discrete experimental levels available in the database.

The SHAP ranking was finally compared with the multi-point OAT sensitivity results, as shown in Figure **13**.

Figure **13** displays that both methods identify GGBS as the most influential input. Its normalized importance is 45.43% in the OAT analysis and 42.40% in SHAP. The main difference appears for PI and FA. OAT assigns 26.10% importance to FA and 10.84% to PI, whereas SHAP assigns 15.89% to FA and 25.22% to PI. This difference does not indicate a contradiction

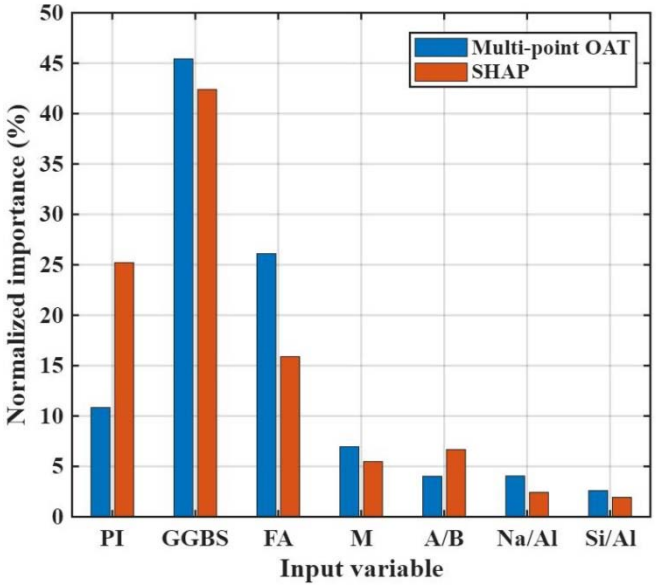


Figure 13: Comparison of OAT and SHAP importance for the GEP model. GGBS dominates both methods (45.43% and 42.40%), reflecting its strong perturbation response and prediction contribution. FA ranks second in OAT, whereas PI ranks second in SHAP.

between the two methods. Multi-point OAT measures the response of the model to controlled changes in one variable while the remaining inputs are held fixed. SHAP distributes each prediction among the input variables over the observed data space and reflects the model structure and feature combinations. The two approaches therefore describe different aspects of the GEP response. Overall, both analyses consistently identify GGBS as the dominant input in the selected GEP model. They also show that PI and FA have important but directionally different roles. Higher GGBS generally increases the predicted UCS, whereas higher PI and FA tend to reduce it. A/B and M have smaller positive contributions, while Na/Al and Si/Al show weaker global effects. These findings explain the behavior learned by the GEP model and should not be interpreted as direct evidence of physical causality.

3.4.3. Sensitivity of Model Interpretations to Sampling Distribution and Feature Dependence

The sensitivity of the GEP interpretation results to non-uniform sampling and correlated input variables was further examined. Among the 216 training observations, 165 (76.4%) had zero FA content. Under the original training distribution, SHAP assigned importance values of 25.82% to PI and 16.46% to FA. When the two FA strata (FA = 0 and FA > 0) were assigned equal weights in the aggregation of feature importance, these values changed to 22.02% and 22.82%, respectively. The corresponding OAT importance values changed from 10.84% and 26.10% to 9.90% and 28.19%. GGBS remained the most influential variable under both weighting schemes. Figure **14** compares the normalized SHAP and OAT importance values obtained under the original empirical sampling distribution and the FA-stratified balanced weighting scheme. The balanced weighting represents a sensitivity scenario rather than the original experimental sampling distribution.

The discrepancy between the PI and FA rankings was particularly evident under the original training distribution. SHAP ranked PI above FA (25.82% versus 16.46%), whereas OAT ranked FA above PI (26.10% versus 10.84%). After FA-stratified balancing, the SHAP ranking was reversed, although the importance values of FA (22.82%) and PI (22.02%) became nearly equal. In contrast, the OAT ranking remained unchanged. These findings demonstrate that the disagreement between the two methods is sensitive to sample weighting. However, this sensitivity does not fully explain the discrepancy, because SHAP and OAT quantify different aspects of the fitted model. Their importance rankings should therefore not be interpreted as definitive measures of physical variable influence.

Feature dependence was further examined using grouped SHAP analysis. The FA–Si/Al and A/B–Na/Al pairs exhibited correlations of 0.83 and 0.71, respectively, in the original dataset. The corresponding VIF values for Si/Al and Na/Al were 4.32 and 4.11, indicating moderate structural dependence. The FA–Si/Al and A/B–Na/Al groups accounted for 17.59% and 5.32% of the normalized group-level importance, respectively. These values represent joint group attributions rather than the sums of individual SHAP importance values. Grouping correlated predictors allows their joint contributions to be examined but does not fully eliminate the interpretive limitations associated with feature dependence.

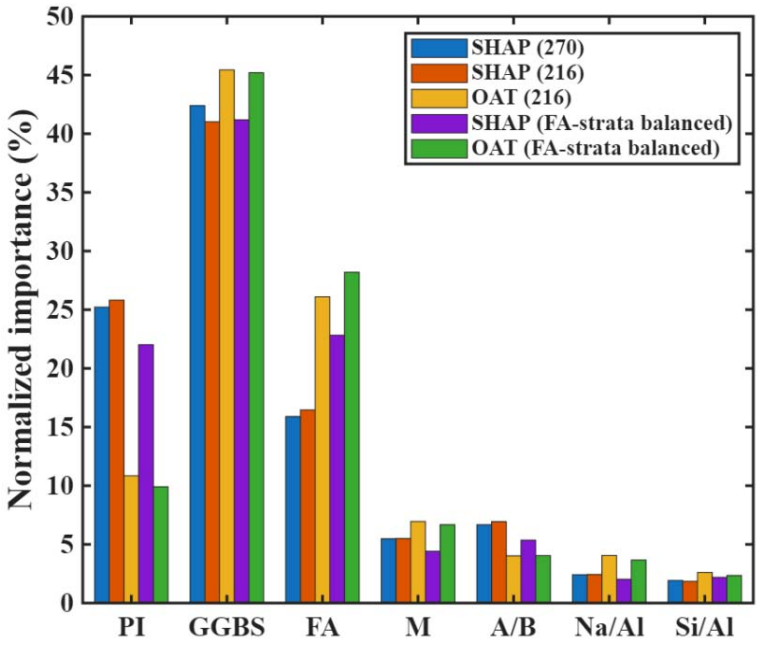


Figure 14: Sensitivity of GEP feature importance to FA-stratified sample weighting. GGBS remains dominant, whereas balancing the FA groups changes the SHAP ranking of PI and FA but preserves their OAT ranking.

Additional OAT diagnostics showed that 39.35% of FA perturbations were clipped to the marginal training bounds, while 56.02% remained within the observed pairwise conditional range associated with Si/Al. This indicates that some controlled perturbations had limited support in the observed joint input space, even when individual variable bounds were respected.

These findings indicate that GGBS importance was comparatively stable across the examined scenarios, whereas the relative importance of PI and FA was sensitive to sample weighting. Correlated input variables and limited empirical support for some OAT perturbations complicate the interpretation of individual importance rankings. Therefore, the reported importance values should be interpreted as model- and distribution-dependent measures rather than direct evidence of physical causality. The differences from previous studies should also be considered in relation to soil characteristics, binder chemistry, and activation conditions. In particular, the high proportion of zero-FA mixtures and the correlation between FA and Si/Al may affect the importance assigned to FA. Therefore, the negative FA response observed in the present GEP model should not be generalized to other geopolymer formulations.

3.5. Prediction Error Diagnostics of the GEP Model

To further assess predictive reliability, additional error diagnostics were performed using the 54 held-out test observations. Particular attention was given to 26 low-strength specimens with measured UCS ≤ 0.5 MPa. This threshold was used for exploratory subgroup analysis rather than as an engineering acceptance criterion. For these specimens, GEP yielded an MAE of 0.976 MPa, an RMSE of 1.776 MPa, and a mean bias error (MBE) of +0.875 MPa. In contrast, the MBE for the 28 higher-strength specimens was −0.940 MPa. These opposing biases resulted in a near-zero overall MBE of −0.066 MPa. Thus, the overall bias concealed systematic differences between strength ranges. Although the explicit GEP equation ensured non-negative predictions, its positive mean bias in the low-strength subgroup indicates that additional experimental verification is necessary before applying the equation to engineering mixture selection.

3.6. Engineering Application and Model Applicability Domain

The proposed GEP equation provides a practical tool for preliminary screening of geopolymer-stabilized soil mixtures. For a given soil and activator system, engineers can compare alternative GGBS/FA proportions using predicted 28-day UCS and prioritize promising mixtures for laboratory verification. This decision-oriented application is consistent with the mix-design approaches investigated by Maheepala *et al*. [46] and Zhang *et al*. [20].

To evaluate this application, 13 experimentally tested mixtures from the 54 held-out observations were organized into four groups. Within each group, PI, M, A/B, Na/Al, and Si/Al remained constant, while precursor contents varied. The GEP equation ranked the mixtures without using their measured UCS values during selection.

GEP identified the mixture with the highest measured UCS in three groups. In Case 4, all

Table 6: Prediction Error Diagnostics for the 26 Low-Strength Test Specimens (UCS ≤ 0.5 MPa)

Model	MAE (MPa)	RMSE (MPa)	MBE (MPa)
GEP	0.9762	1.7764	+0.8749
SVR	0.7260	1.0708	+0.1985
RF	0.1920	0.2614	+0.1910
GBRT	0.1756	0.2434	−0.0367

Table 7: Retrospective GEP-Based Screening of Experimentally Tested Mixtures

Case	PI (%)	Mixtures	Maximum measured UCS (MPa)	GEP selection
1	14.07	4	22.380	Correct
2	56.46	3	15.690	Correct
3	88.46	3	6.285	Correct
4	14.07	3	0.2107	No unique selection

candidates received zero predictions, preventing a unique selection. Thus, the equation can assist in prioritizing binder formulations for confirmatory testing but shows limited discrimination among very-low-strength mixtures.

The training domain comprised PI = 14.07–88.46%, GGBS = 0–50%, FA = 0–20%, M = 4–15 mol/L, A/B = 0.45–0.85, Na/Al = 0.24–1.98, and Si/Al = 1.49–2.49. All 54 test observations satisfied these individual limits, although two exceeded the descriptive 95th-percentile training nearest-neighbor distance. Therefore, applications should consider both the input ranges and the support for their combined values. The equation is intended for preliminary screening under comparable experimental conditions, not direct design approval. Laboratory UCS confirmation remains essential, and reductions in testing effort or cost have not been experimentally quantified.

3.7. Limitations and Future Research

The present study has several limitations. First, all 270 experimental records were obtained from a previously published database. Although a held-out test set and five-fold evaluation were used, both assessments relied on data from the same experimental source. Therefore, the reported performance reflects internal validation rather than external generalization. Consequently, the results provide additional evidence of predictive performance across stratified partitions but do not establish the full-budget stability of the final optimized models.

Second, the database includes a limited number of soil types and predefined mixture conditions. The input variables also exhibit non-uniform distributions and moderate correlations. These characteristics may limit the reliability of predictions beyond the investigated data domain. Third, the OAT and SHAP analyses describe the behavior of the fitted GEP model. Their results should not be interpreted as direct evidence of physical causality.

Future studies should assess the developed models using independent experimental datasets. Different soil types, precursor compositions, and curing conditions should be considered. Group-based validation, including leave-one-soil-type-out evaluation, could also be applied where sufficient data are available. Further research should examine prediction uncertainty and model robustness under different experimental conditions. These investigations would provide stronger evidence of model transferability and support the practical application of the proposed framework in geotechnical engineering.

4. CONCLUSIONS

The present study compared GEP, SVR, RF, and GBRT for predicting the UCS of geopolymer-stabilized clayey soils. A common database, training–test split, and held-out test set were used to provide a consistent comparison. The main conclusions are:

GBRT provided the highest independent-test accuracy. It achieved R^2=0.9814, RMSE=0.9771 MPa, and MAE=0.6080 MPa. RF ranked second, followed by SVR and GEP. The results confirm the strong predictive capability of ensemble tree-based models for the investigated problem.

The original five-fold evaluation showed a slightly higher mean R^2 for SVR (0.9580 ± 0.0159) than for GBRT (0.9524 ± 0.0255). In the supplementary nested stratified five-fold analysis, GBRT achieved the highest mean R^2 (0.9555 ± 0.0169) and the lowest mean RMSE (1.2838 ± 0.2370 MPa). These findings indicate

that the relative model performance depends partly on the validation protocol, while GBRT consistently achieved strong predictive accuracy.

GEP provided a useful compromise between predictive capability and engineering transparency. Although its independent-test R^2 of 0.9023 was lower than those of the black-box models, it generated an explicit equation involving all seven input variables. GEP was therefore retained as the interpretable model rather than being presented as the most accurate predictor.

The multi-point OAT analysis identified GGBS as the most influential variable, with a normalized importance of 45.43%. FA ranked second at 26.10%, followed by PI at 10.84%. The GEP model showed a positive response to GGBS and predominantly negative responses to PI and FA within the investigated data domain. These trends depend on the sampled mixture compositions and activation conditions and should not be generalized.

The SHAP analysis independently confirmed the dominant role of GGBS. Its normalized SHAP importance was 42.40%, followed by PI at 25.22% and FA at 15.89%. These three variables together accounted for approximately 83.5% of the total SHAP importance. A/B and M showed smaller contributions, while Na/Al and Si/Al had relatively weak global effects.

OAT and SHAP provided complementary interpretations rather than identical rankings. Both methods identified GGBS as the dominant input, but OAT assigned greater importance to FA, whereas SHAP gave a larger contribution to PI. This difference arises because OAT measures model response under controlled perturbations, while SHAP distributes individual predictions among the input variables over the observed data space.

Overall, the results show that predictive accuracy and interpretability should be treated as related but distinct objectives. GBRT is preferred when predictive accuracy is the primary objective, whereas GEP offers an interpretable, data-driven equation for examining UCS predictions. The combined use of GEP, OAT, and SHAP also provides a clearer description of the learned relationships than any single interpretation method alone. The identified effects should, however, be interpreted within the range and structure of the present database and should not be considered direct evidence of physical causality.

ACKNOWLEDGEMENT

The first author is partially supported by the Iran National Science Foundation (INSF).

AUTHOR CONTRIBUTIONS

Mahmoud Ghazavi and Mobina Taslimi Paein Afrakoti contributed to the conceptualization and design of the study, methodology development, software implementation, data analysis, interpretation of results, and preparation of the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Firoozi AA, Guney Olgun C, Firoozi AA, Baghini MS. Fundamentals of soil stabilization. International Journal of Geo-Engineering 2017; 8(1): 26. https://doi.org/10.1186/s40703-017-0064-9

[2] Ikeagwuani CC, Nwonu DC. Emerging trends in expansive soil stabilisation: A review. Journal of Rock Mechanics and Geotechnical Engineering 2019; 11(2): 423-440. https://doi.org/10.1016/j.jrmge.2018.08.013

[3] Barman D, Dash SK. Stabilization of expansive soils using chemical additives: A review. Journal of Rock Mechanics and Geotechnical Engineering 2022; 14(4): 1319-1342. https://doi.org/10.1016/j.jrmge.2022.02.011

[4] Mohammed AA, Nahazanan H, Mohd Nasir NA, Huseien GF, Saad AH. Calcium-based binders in concrete or soil stabilization: Challenges, problems, and calcined clay as partial replacement to produce low-carbon cement. Materials 2023; 16(5): 2020. https://doi.org/10.3390/ma16052020

[5] Hov S, Paniagua P, Sætre C, Long M, Cornelissen G, Ritter S. Stabilisation of soft clay, quick clay and peat by industrial by-products and biochars. Applied Sciences 2023; 13(16): 9048. https://doi.org/10.3390/app13169048

[6] Varsha B, Moghal AAB, Rehman A, Chittoori BCS. Shear, consolidation characteristics and carbon footprint analysis of clayey soil blended with calcium lignosulphonate and granite sand for earthen dam application. Sustainability 2023; 15(7): 6117. https://doi.org/10.3390/su15076117

[7] Barimani M, Motaghedi H, Soleimani Kutanaei S, Ebrahimi A. Stabilizing collapsible soils using nano calcium carbonate to enhance mechanical properties. Scientific Reports 2026; 16(1): 9353. https://doi.org/10.1038/s41598-026-39716-9

[8] Amiri AA, Ranjbar Malidarreh N, Soleimani Kutanaei S, Seyedkazemi A, Davoudi-Ki A. Enhancing the mechanical and shear behavior of clay soil using lime, Nano-MgO, and recycled PET fibers: experimental and UPV-based assessment. Scientific Reports 2026; 16(1): 7548. https://doi.org/10.1038/s41598-026-38956-z

[9] Cristelo N, Glendinning S, Fernandes L, Pinto AT. Effects of alkaline-activated fly ash and Portland cement on soft soil stabilisation. Acta Geotechnica 2013; 8(4): 395-405. https://doi.org/10.1007/s11440-012-0200-9

[10] Singhi B, Laskar AI, Ahmed MA. Investigation on soil-geopolymer with slag, fly ash and their blending. Arabian Journal for Science and Engineering 2016; 41(2): 393-400. https://doi.org/10.1007/s13369-015-1677-y

[11] Abdullah HH, Shahin MA, Walske ML, Karrech A. Systematic approach to assessing the applicability of fly-ash-based geopolymer for clay stabilization. Canadian Geotechnical Journal 2020; 57(9): 1356-1368. https://doi.org/10.1139/cgj-2019-0215

[12] Sukmak P, Horpibulsuk S, Shen S-L. Strength development in clay-fly ash geopolymer. Construction and Building Materials 2013; 40: 566-574. https://doi.org/10.1016/j.conbuildmat.2012.11.015

[13] Phetchuay C, Horpibulsuk S, Arulrajah A, Suksiripattanapong C, Udomchai A. Strength development in soft marine clay stabilized by fly ash and calcium carbide residue based geopolymer. Applied Clay Science 2016; 127: 134-142. https://doi.org/10.1016/j.clay.2016.04.005

[14] Abdullah HH, Shahin MA, Walske ML. Geo-mechanical behavior of clay soils stabilized at ambient temperature with fly-ash geopolymer-incorporated granulated slag. Soils and Foundations 2019; 59(6): 1906-1920. https://doi.org/10.1016/j.sandf.2019.08.005

[15] Nagaraju TV, Mantena S, Gobinath R, Bonthu S, Alisha SS. Geopolymer-stabilized soils: Influencing factors, strength development mechanism and sustainability. Journal of Taibah University for Science 2023; 17(1): 2248651. https://doi.org/10.1080/16583655.2023.2248651

[16] Luo Z, Luo B, Zhao Y, Li X, Su Y, Huang H, Wang Q. Experimental investigation of unconfined compression strength and microstructure characteristics of slag and fly ash-based geopolymer stabilized riverside soft soil. Polymers 2022; 14(2): 307. https://doi.org/10.3390/polym14020307

[17] Mozumder RA, Laskar AI. Prediction of unconfined compressive strength of geopolymer stabilized clayey soil using artificial neural network. Computers and Geotechnics 2015; 69: 291-300. https://doi.org/10.1016/j.compgeo.2015.05.021

[18] Pham VN, Oh E, Ong DE. Effects of binder types and other significant variables on the unconfined compressive strength of chemical-stabilized clayey soil using gene-expression programming. Neural Computing and Applications 2022; 34(11): 9103-9121. https://doi.org/10.1007/s00521-022-06931-0

[19] Tran VQ. Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction. Journal of Cleaner Production 2022; 355: 131683. https://doi.org/10.1016/j.jclepro.2022.131683

[20] Zhang J, Choi CE, Liang Z, Li R. A generic framework for mix design of geopolymer for soil stabilization: Composition-informed machine learning model. Computers and Geotechnics 2024; 170: 106322. https://doi.org/10.1016/j.compgeo.2024.106322

[21] Afrakoti MTP, Ghazavi M. Unified model for prediction of the strength of soils stabilized with xanthan gum using white-box machine learning. Scientific Reports 2026; 16(1): 22136. https://doi.org/10.1038/s41598-026-51877-1

[22] Mozumder R A, Laskar AI, Hussain M. Empirical approach for strength prediction of geopolymer stabilized clayey soil using support vector machines. Construction and Building Materials 2017; 132: 412-424. https://doi.org/10.1016/j.conbuildmat.2016.12.012

[23] Soleimani S, Rajaei S, Jiao P, Sabz A, Soheilinia S. New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming. Measurement 2018; 113: 99-107. https://doi.org/10.1016/j.measurement.2017.08.043

[24] Ghazavi M, Afrakoti MTP. Unconfined compressive strength prediction of soils improved with biopolymers: machine learning approach. Transportation Infrastructure Geotechnology 2025; 12(1): 14. https://doi.org/10.1007/s40515-024-00456-w

[25] Javdanian H, Lee S. Evaluating unconfined compressive strength of cohesive soils stabilized with geopolymer: a computational intelligence approach. Engineering with Computers 2019; 35(1): 191-199. https://doi.org/10.1007/s00366-018-0592-8

[26] Tabarsa A, Latifi N, Osouli A, Bagheri Y. Unconfined compressive strength prediction of soils stabilized using artificial neural networks and support vector machines. Frontiers of Structural and Civil Engineering 2021; 15(2): 520-536. https://doi.org/10.1007/s11709-021-0689-9

[27] Zeini HA, Al-Jeznawi D, Imran H, Bernardo LFA, Al-Khafaji Z, Ostrowski KA. Random forest algorithm for the strength prediction of geopolymer stabilized clayey soil. Sustainability 2023; 15(2): 1408. https://doi.org/10.3390/su15021408

[28] Ngo AQ, Nguyen LQ, Tran VQ. Developing interpretable machine learning-Shapley additive explanations model for unconfined compressive strength of cohesive soils stabilized with geopolymer. Plos One 2023; 18(6): e0286950. https://doi.org/10.1371/journal.pone.0286950

[29] Abdullah GM, Ahmad M, Babur M, Badshah MU, Al-Mansob RA, Gamil Y, Fawad M. Boosting-based ensemble machine learning models for predicting unconfined compressive strength of geopolymer stabilized clayey soil. Scientific Reports 2024; 14(1): 2323. https://doi.org/10.1038/s41598-024-52825-7

[30] Chen Q, Hu G, Wu J. Comparative study on the prediction of the unconfined compressive strength of the one-part geopolymer stabilized soil by using different hybrid machine learning models. Case Studies in Construction Materials 2024; 21: e03439. https://doi.org/10.1016/j.cscm.2024.e03439

[31] Murtaza RS, Sultana N, Debanath OC, Rahman MA. Estimating ML-driven compressive strength of geo-polymer stabilized soils. Discover Civil Engineering 2026; 3(1): 121. https://doi.org/10.1007/s44290-026-00524-9

[32] Choobbasti AJ, Tavakoli H, Kutanaei SS. Modeling and optimization of a trench layer location around a pipeline using artificial neural networks and particle swarm optimization algorithm. Tunnelling and Underground Space Technology 2014; 40: 192-202. https://doi.org/10.1016/j.tust.2013.10.003

[33] Tavakoli HR, Omran OL, Kutanaei SS. Prediction of energy absorption capability in fiber reinforced self-compacting concrete containing nano-silica particles using artificial neural network. Latin American Journal of Solids and Structures 2014; 11: 966-979. https://doi.org/10.1590/S1679-78252014000600004

[34] Tavakoli H, Soleimani Kutanaei S. Evaluation of effect of soil characteristics on the seismic amplification factor using the neural network and reliability concept. Arabian Journal of Geosciences 2015; 8(6): 3881-3891. https://doi.org/10.1007/s12517-014-1458-z

[35] Kutanaei SS, Choobbasti AJ. Prediction of liquefaction potential of sandy soil around a submarine pipeline under earthquake loading. Journal of Pipeline Systems Engineering and Practice 2019; 10(2): 04019002. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000349

[36] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 2017; 30: 4765-4774.

[37] Murdoch WJ, Singh C, Kumbier K, Abbasi-Asl R, Yu B. Definitions, methods, and applications in interpretable machine learning. Proceedings of the National Academy of Sciences 2019; 116(44): 22071-22080. https://doi.org/10.1073/pnas.1900654116

[38] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 2019; 1(5): 206-215. https://doi.org/10.1038/s42256-019-0048-x

[39] Ferreira C. Gene expression programming: A new adaptive algorithm for solving problems. Complex Systems 2001; 13(2): 87-129.

[40] Fereidouni P, Seyedkazemi A, Kutanaei SS, Davoudi-Kia A. Modeling and sensitivity analysis of the compressive strength of recycled brick aggregates concrete using GMDH, GEP and RSM methods. Results in Engineering 2025; 26: 104891. https://doi.org/10.1016/j.rineng.2025.104891

[41] Ghorbani V, Seyedkazemi A, Kutanaei SS. Predicting compressive strength of mortars containing recycled CRT

glass using GMDH and GEP methods. Scientific Reports 2026; 16(1): 6655. https://doi.org/10.1038/s41598-026-36553-8

[42] Amiri G, Esmaeilnia Amiri M, Ghasemnejad H, Seyedkazemi A, Soleimani Kutanaei S. White-box modeling and reliability assessment of roller-compacted concrete for dam and pavement engineering applications. Scientific Reports 2026; 16. https://doi.org/10.1038/s41598-026-63838-9

[43] Ghamari AK, Seyedkazemi A, Jahangir S, Soleimani Kutanaei S. Predicting the compressive strength of sustainable concrete with recycled foundry sand using advanced soft computing methods. Scientific Reports 2026. https://doi.org/10.1038/s41598-026-59765-4

[44] Tavakkolian R, Ghasemnejad H, Soleimani Kutanaei S. White-box soft computing models for predicting the strength of sustainable pozzolanic concretes. Scientific Reports 2026; 16(1): 25088. https://doi.org/10.1038/s41598-025-29980-6

[45] Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting 2006; 22(4): 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[46] Maheepala MMALN, Nasvi MCM, Robert DJ, Gunasekara C, Kurukulasuriya LC. Mix design development for geopolymer treated expansive subgrades using artificial neural network. Computers and Geotechnics 2023; 161: 105534. https://doi.org/10.1016/j.compgeo.2023.105534

https://doi.org/10.65904/3083-3590.2026.02.12

© 2026 Ghazavi and Afrakoti.
This is an open access article licensed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution and reproduction in any medium, provided the work is properly cited.