



Machine-Learning Prediction and Interpretation of Peak Shear Strength in Fly Ash-Amended, Lignin-Fiber-Reinforced Silty Sand

Masoud Nasiri ¹; Masoud Ahmadi ^{2,*}

1. Faculty of Engineering, Kermanshah University of Technology, Kermanshah, Iran

2. Faculty of Earth Sciences Engineering, Arak University of Technology, Arak, Iran

* Corresponding author: ms.ahmadi@arakut.ac.ir

ARTICLE INFO

Article history:

Received: 26 August 2026

Revised: 26 September 2026

Accepted: 22 October 2026

Keywords:

Peak shear strength;

Machine learning;

Lignin fiber;

Fly ash-amended soil;

Model interpretation.

ABSTRACT

Accurate prediction of peak shear strength can facilitate the mechanical assessment of stabilized and reinforced soils while reducing reliance on extensive laboratory testing. This study develops a data-driven framework for predicting and interpreting the peak shear strength of fly ash-amended, lignin-fiber-reinforced silty sand. Eight machine-learning algorithms were comparatively evaluated using 90 observations obtained from unconsolidated–undrained triaxial tests, with lignin fiber content, curing time, initial moisture content, and confining pressure as input variables. Among the evaluated models, Gradient Boosting exhibited the best generalization performance, achieving a test R^2 of 0.871, an MAE of 92.60 kPa, and an RMSE of 110.10 kPa. Model-based feature importance identified lignin fiber content as the dominant predictor, followed by confining pressure, initial moisture content, and curing time. Partial-dependence and sensitivity analyses further revealed pronounced nonlinear effects of lignin fiber content, whereas confining pressure showed a consistently positive influence on the predicted strength. Residual diagnostics indicated a standard deviation of 113.14 kPa, with no residuals exceeding the adopted $\pm 2\sigma$ threshold. Overall, the findings demonstrate the potential of nonlinear ensemble learning for accurate and interpretable prediction of peak shear strength within the investigated experimental domain, while providing insight into the relative influence of key material and testing parameters.

E-ISSN: 2345-4423

© 2026 The Authors. Journal of Rehabilitation in Civil Engineering published by Semnan University Press.

This is an open access article under the CC-BY 4.0 license. (<https://creativecommons.org/licenses/by/4.0/>)

How to cite this article:

Nasiri, M., Ahmadi, M. (2026). Machine-Learning Prediction and Interpretation of Peak Shear Strength in Fly Ash-Amended, Lignin-Fiber-Reinforced Silty Sand. Journal of Rehabilitation in Civil Engineering, 14 (4), 2756. <https://doi.org/10.22075/jrce.2026.2756>

1. Introduction

Soil shear strength is a crucial mechanical property that affects the stability and function of geotechnical structures, including foundations [1], retaining walls [2], slopes [3], embankments [4], pavements [5], and earthworks [6]. It reflects the soil's capacity to endure shearing deformation. This property is impacted by a complicated interplay of aspects, such as effective stress, particle arrangement, density, gradation, mineral composition, moisture content, confining pressure, and soil structure [7].

The rapid expansion of infrastructure projects has intensified demand for construction materials while exacerbating the environmental challenges related to the generation and disposal of industrial by-products. Among these materials, coal fly ash, a finely powdered residue resulting from the combustion of coal in thermal power stations [8], has attracted significant attention as a promising resource for sustainable ground improvement. The large-scale disposal of fly ash can consume substantial land and pose risks of leaching potentially harmful substances into adjacent soil and water systems. As such, converting fly ash from a waste product into a valuable geotechnical material presents a dual benefit: it mitigates the environmental impact of its disposal and lessens reliance on conventional soil-stabilizing agents. Previous reviews have indicated that fly ash can markedly enhance the physical [9–11] and mechanical [12] features of problematic soils, thereby highlighting its critical role in sustainable soil stabilization and ground engineering applications [13,14].

From a geotechnical perspective, fly ash is appealing due to its distinctive physicochemical properties [15]. The characteristics of fly ash can vary significantly depending on the coal source, combustion conditions, and collection method [16]. Typically comprised of fine, predominantly spherical particles, fly ash contains varying proportions of silica, alumina, iron oxides, calcium-bearing minerals, and unburned carbon. These properties influence its particle-size distribution, specific gravity, surface characteristics, reactivity, and interactions with soil minerals [17]. Fly ash is generally classified by its chemical composition and calcium content, with low-calcium Class F ashes [18] and high-calcium Class C ashes [19] exhibiting notably different stabilization mechanisms. Therefore, the engineering behavior of soil treated with fly ash should not be regarded as separate from the mineralogy and chemistry of the fly ash; rather, it is shaped by the interaction among the soil characteristics, fly ash composition, the quantity of additive utilized, water content, compaction state, and curing conditions [13].

The findings showed that higher fly ash levels in expansive soil reduced plasticity, swelling, and hydraulic conductivity, while increasing dry unit weight and strength. These results suggest that fly ash can effectively address multiple undesirable traits in problematic soils, rather than merely enhancing strength [20]. Additionally, research on silty sand has shown that fly ash stabilization can enhance mechanical properties. Concurrently, alkaline activation of fly ash has been shown to significantly enhance stiffness by forming geopolymer-binding compounds [21]. Recent research suggests that incorporating fly ash may greatly enhance the expansive unconfined compressive strength of soil [22]. In this regard, the unconfined compressive strength increases with both increased curing time and fly ash content [23]. It was also found that fly ash significantly improved the soil's energy absorption [24].

Machine learning has emerged as an effective framework for addressing nonlinear prediction challenges in geotechnical engineering [25]. Unlike traditional analytical methods, soft computing methods [26] and ML algorithms can estimate complex input-output relationships without requiring a predefined functional form [27,28]. Explainable artificial intelligence [29] and gene expression [30] are among the valuable tools in prediction studies. Their application in modeling soil behavior has demonstrated considerable promise in predicting critical parameters such as bearing capacity [31] and shear strength [32], as well as other engineering responses. By analyzing experimental data, Machine learning models can reveal patterns that conventional regression equations may not readily capture. This capability is particularly crucial for fly ash-reinforced soils, as their shear strength is affected by numerous variables that interact

through intricate mechanical and interfacial processes. Employing machine learning to predict the shear strength of fly ash-reinforced soil offers a practical alternative to extensive laboratory testing. Once a sufficiently representative dataset is established, a trained model can rapidly estimate shear strength for various combinations of soil and fly ash characteristics within the training data's parameter range. This predictive capability can facilitate early material selection, optimize fly ash content and geometry, compare different reinforcement options, and conduct sensitivity analyses on design parameters. Additionally, machine learning-based methods can illuminate the relative impact of individual factors, thereby complementing rather than replacing traditional interpretations of soil mechanics.

Chen et al. [33] developed ensemble learning models to predict the shear strength of sly ash-amended, lignin-reinforced soil. This research involved over 270 unconsolidated-undrained triaxial tests. The results highlighted the effectiveness of ensemble learning in capturing the complex strength responses of stabilized soil systems under diverse environmental and material conditions. This study is particularly important because it demonstrates that models that account for interactions among chemical stabilization, reinforcement, environmental factors, and stress states can significantly improve predictions of stabilized soil strength. Wani and Thagunna [34] investigated the prediction of shear strength in fly ash-stabilized clayey soil employing artificial neural networks and support vector regression. Their models incorporated several predictive factors, including soil grain-size characteristics, plasticity indices, moisture content, fly ash content, and curing duration. The results revealed that support vector regression outperformed artificial neural networks on the analyzed dataset, underscoring the ability of machine learning models to capture the nonlinear relationships between stabilization factors and shear strength. This conclusion is important as it highlights that fly ash content should not be considered in isolation; rather, its impact on shear strength must be understood alongside the soil's inherent properties and the specific conditions present during testing or curing [34].

The residual analysis of the machine learning investigation of fly ash-reinforced soil revealed little bias and controlled variance, indicating a high level of generalization for laboratory use [35]. Scholars demonstrate that Random Forest and Gradient Boosting models can predict the strength properties of fly ash-improved soil [36]. The findings indicate that the Gradient Boosting Decision Tree model shows exceptional predictive accuracy, reaching a coefficient of determination as high as 0.99 [37]. In additional studies, the Support Vector Machine utilizing the RBF kernel function was recognized as the best model, capable of forecasting UCS values with sufficient precision [38].

The primary innovation of this paper is the generation of an integrated, interpretable machine learning framework for predicting the peak shear strength of fly ash-reinforced silty sands. This was achieved through a systematic comparison of eight fundamentally different learning algorithms. The research used a database of 90 unconsolidated undrained triaxial experimental observations, with predictors including fiber content, curing time, confining pressure, and initial moisture content, and peak shear strength as the target response. Instead of relying solely on a single performance indicator to assess predictive accuracy, the proposed framework combines various evaluation methods, including comparative model assessment, actual-versus-predicted analysis, residual diagnostics, feature-importance evaluation, partial dependence, one-at-a-time sensitivity analysis, error analysis, scatter-matrix assessment, normalized multi-metric comparison, and formal outlier detection. This integrated approach not only offers robust predictive capabilities but also enhances engineering interpretability, which is crucial for geotechnical applications where understanding the influencing variables is just as significant as achieving accurate predictions.

The study ultimately establishes a comprehensive comparative baseline for predicting peak shear strength in fly ash-reinforced silty sands using machine learning. A notable contrast among the eight algorithms highlights that predictive performance on this problem is heavily algorithm-dependent. Thus, the originality of this work lies not only in the application of machine learning to an established geotechnical dataset but also in the integration of systematic algorithm comparison, generalization-focused evaluation,

interpretable feature analysis, sensitivity assessment, and residual/outlier diagnostics within a single framework. This approach provides a robust foundation for selecting and interpreting machine-learning models for predicting peak shear strength. It offers a reproducible methodological pathway for future data-driven studies in reinforced geomaterials.

2. Research significance

The present study is noteworthy for focusing on predicting the peak shear strength of fly ash-reinforced silty sands. This geotechnical response fundamentally depends on the combined effects of reinforcement content, curing conditions, moisture state, and confinement. Traditionally, determining peak shear strength relies heavily on laboratory testing, which can be time-consuming, especially when assessing numerous combinations of material and stress conditions. In this context, the current work establishes a data-driven framework based on 90 unconsolidated undrained triaxial experimental observations, utilizing fiber content, curing time, initial moisture content, and confining pressure as input variables. The extensive range of experimental conditions in the dataset provides a robust basis for investigating whether machine learning methods can capture the nonlinear response of reinforced silty sands within the examined domain.

A significant scientific contribution of this investigation lies in the systematic comparison of eight fundamentally different machine learning approaches, rather than relying on a single algorithm. The models investigated include Support Vector Regression, Multi-Layer Perceptron, Gradient Boosting, XGBoost, Random Forest, Decision Tree, K-Nearest Neighbors, and Linear Regression. This comparative framework illustrates that the selection of the learning algorithm outstandingly affects predictive reliability. The research is particularly noteworthy as it combines prediction with model diagnostics and interpretability. The analysis encompasses various techniques, including actual-versus-predicted plots, residual plots, residual histograms, feature importance assessments, partial dependence plots, one-at-a-time sensitivity analysis, residual-versus-input analysis, residual boxplots, scatter-matrix analysis, a normalized performance radar chart, and formal outlier detection. This multidimensional evaluation is especially critical in geotechnical engineering, where a model should not be deemed acceptable solely based on a high R^2 value. By integrating predictive accuracy, error behavior, sensitivity, and interpretability, a more robust framework is established for assessing whether the model accurately captures a meaningful and stable relationship within the given experimental domain.

The study highlights the importance of integrating machine learning with geotechnical interpretation, rather than treating it merely as a prediction tool. As a result, this study offers both a predictive methodology and an analytical perspective for future research on reinforced geomaterials. Therefore, from a practical geotechnical standpoint, this research presents a potential method to reduce reliance on repeated laboratory testing for preliminary strength estimation. The identified Gradient Boosting model can estimate peak shear strength from combinations of fiber content, curing time, initial moisture, and confining pressure, as outlined in the experimental database. This capability can aid in preliminary material assessments, experimental planning, sensitivity analyses, and the identification of promising reinforcement conditions.

3. Methods

The methodology employed in this study was crafted to create a systematic and reproducible machine learning framework for the prediction of the peak shear strength of fly ash-stabilized silty sands. The overall process consisted of five primary stages: (i) compilation and characterization of the experimental database, (ii) definition of input and output variables, (iii) development and comparative evaluation of eight machine learning models, (iv) assessment of predictive performance using complementary statistical

and graphical criteria, and (v) interpretation of the top-performing model through feature importance, partial dependence, sensitivity, residual, and outlier analyses. The experimental database consisted of 90 observations from unconsolidated undrained triaxial tests, with fiber content, curing time, initial moisture content, and confining pressure selected as predictors, and peak shear strength as the target variable. This workflow was intentionally structured to ensure that model selection considered not only goodness of fit but also generalization capability, error magnitude, residual behavior, and physical interpretability.

The selected input variables collectively represent the key experimental conditions included in the available database. Specifically, fiber content denotes the reinforcement condition, curing time indicates the duration of material treatment, initial moisture content describes the material's initial state, and confining pressure reflects the applied stress conditions. Consequently, the methodology treats peak shear strength as a multivariable response rather than attempting to correlate it with a single experimental parameter. Before developing predictive models, the dataset was statistically characterized to assess the distributional behavior of the variables and the statistical properties of the residuals, both of which are crucial for interpreting the models. These statistics provided an initial assessment of the variability and range of the predictor and target variables.

A comprehensive comparative framework was developed by examining eight machine learning approaches: Support Vector Regression, Multilayer Perceptron, Gradient Boosting, XGBoost, Random Forest, Decision Tree, K-Nearest Neighbors, and Linear Regression. These models represent a range of learning philosophies, including linear regression, neighborhood-based learning, tree-based methods, ensemble learning, neural networks, and kernel-based regression. The simultaneous evaluation of these techniques aimed to determine whether a straightforward linear model could adequately capture the peak shear-strength response, or if nonlinear and ensemble-based algorithms would provide a significant advantage. Including Linear Regression as a benchmark was particularly beneficial, as it offered a baseline against which the performance of nonlinear algorithms could be measured. Additionally, Decision Tree, Random Forest, Gradient Boosting, and XGBoost facilitate a comparative analysis of tree-based and ensemble learning strategies. In contrast, KNN, SVR, and MLP introduce fundamentally different nonlinear modeling approaches. This comparison thus enables a thorough evaluation of the algorithms' predictive capabilities without assuming that any specific model family is inherently superior.

A cross-validated grid-search approach was employed for hyperparameter optimization. The independent testing subset was clearly designated before the optimization process and was not used for parameter selection or model development. Candidate parameter combinations were assessed using k-fold cross-validation on the training and development subsets. The primary optimization criterion was the mean cross-validated coefficient of determination (R^2), while mean absolute error (MAE) and root mean square error (RMSE) served as additional measures of prediction error. After optimization was complete, each algorithm was refitted using the entire training and development subset with its chosen hyperparameter configuration, then evaluated on the independent testing subset.

Figure 1 illustrates the flowchart of the employed methodology for peak shear strength prediction in this paper.

3.1. Experimental database and test conditions

The soil used in the experiment by Chen et al. [33] was silty sand, sourced from a construction site along the Qinghai–Tibet Highway at a depth of approximately one meters below ground surface. This soil was classified as silty sand according to the USCS. The reported index properties include a maximum dry density of 1.83 g/cm^3 and an optimum moisture content of 13.2%. The particle size distribution showed that approximately 16.1% of the material was finer than 0.075 mm, 61.1% was between 0.075 mm and 2 mm, and 22.8% was larger than 2 mm. Additionally, the liquid and plastic limits were reported as 29.6% and 19.6%, respectively, yielding a plasticity index of 10.0%.

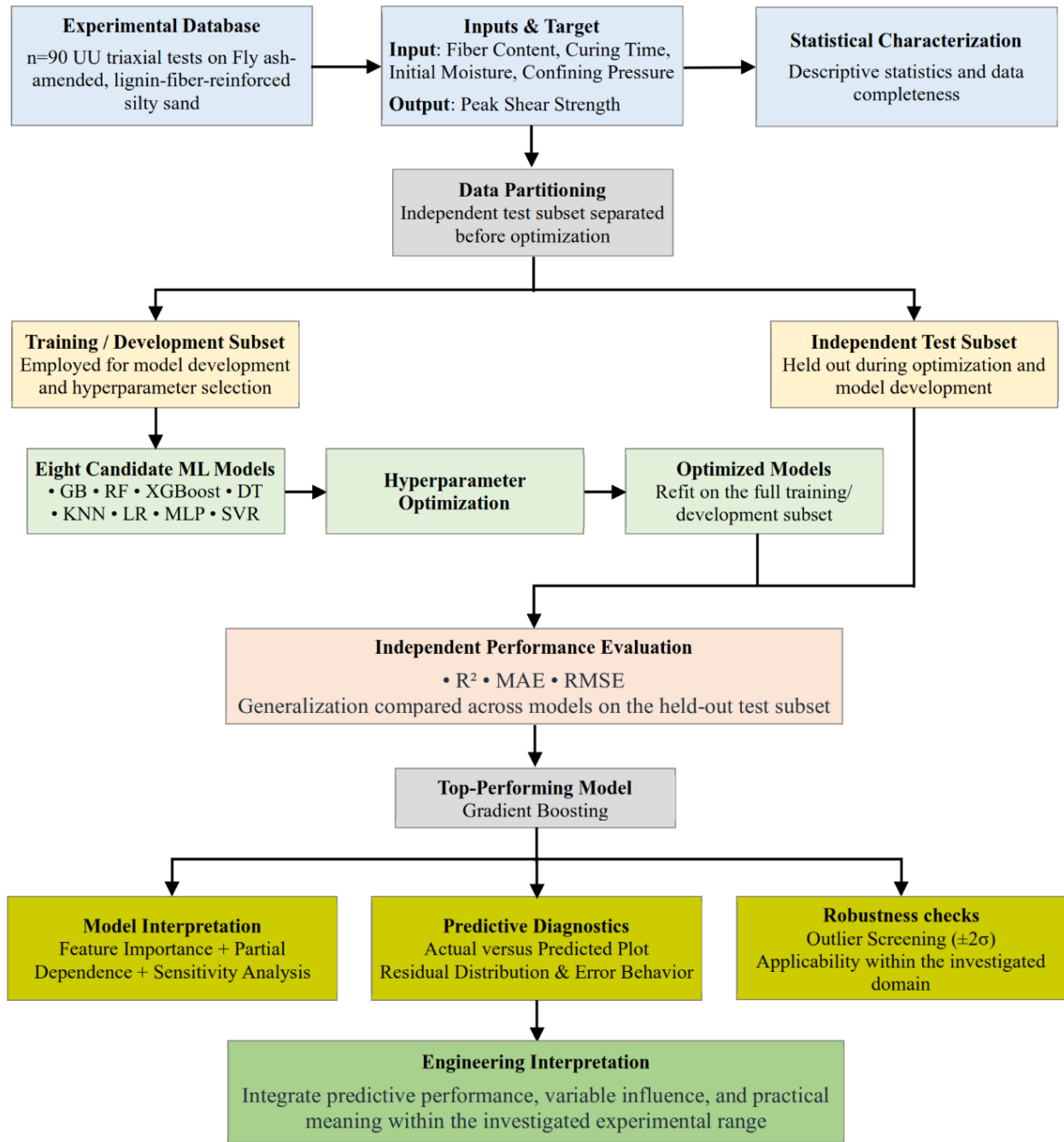


Fig. 1. Flowchart of the investigated methodology for peak shear strength prediction.

The experimental program used 4% fly ash, calculated on the soil's dry mass, as the chemical stabilizing agent. Accordingly, the primary chemical constituents of the fly ash are approximately as follows: $\text{SiO}_2 = 57.0\%$; $\text{Al}_2\text{O}_3 = 28.0\%$; $\text{CaO} = 3.8\%$; $\text{Fe}_2\text{O}_3 = 0.7\%$; and Volatile fraction = 10.5%. The relatively high content of silica and alumina, alongside the comparatively low calcium content, suggests that the fly ash qualifies as a low-calcium, type according to conventional engineering classifications. The reinforcement material consisted of lignin fibers derived from industrial lignin by-products. As reported, these fibers measured less than 1 mm in length, had an ash content of less than 11%, a cellulose content of approximately 95%, a moisture content below 5%, a bulk density of around 27 g/L, and a pH close to 7.0. The lignin content was represented as the mass percentage of dry lignin fibers relative to the dry mass of soil. Five different fiber contents were tested 0, 0.25, 0.50, 0.75, and 1.00%.

3.2. Database pre-analysis

Table 1 outlines the descriptive statistical characteristics of the 90 experimental observations utilized in developing and evaluating the machine learning models aimed at predicting the peak shear strength of fly ash-reinforced silty sands. The dataset was obtained from a previously published research by Chen et al. [33]. As illustrated in Table 1, all five variables have 90 observations, confirming that the dataset is complete for the variables included in the predictive framework and indicating no apparent loss of observations due to missing values. Importantly, the relatively well-defined ranges of the experimental variables suggest that the dataset systematically encompasses combinations of curing conditions, moisture content, fiber content, and confining pressure, thereby providing a solid foundation for analyzing their combined influence on peak shear strength.

Table 1. Dataset statistical analysis.

Statistic	Curing Time (days)	Initial Moisture (%)	Fiber (%)	Confining Pressure (kPa)	Peak Shear Strength (kPa)
count	90	90	90	90	90
mean	7.5	13.2	0.5	200	1764.5667
std	7.542	1.6421	0.3555	82.1071	352.6222
min	0	11.2	0	100	1112.7
25%	0	11.2	0.25	100	1502.75
50%	7.5	13.2	0.5	200	1737.65
75%	15	15.2	0.75	300	2002.5
max	15	15.2	1	300	2621.6

The curing time ranges from 0 to 15 days, with an average of 7.5 days and a standard deviation of 7.542 days. The first and third quartiles are 0 and 15 days, respectively, and the median is also 7.5 days. This distribution suggests that the observations cover the entire specified curing period, with the median positioned approximately halfway between the minimum and maximum values. The relatively large standard deviation relative to the mean indicates wide experimental variation in curing time rather than clustering around a single condition. This variability is particularly significant for predictive modeling, as curing time may impact the interactions and bonding characteristics within the fly ash-reinforced silty sand matrix. Notably, the presence of both uncured and longer-cured specimens allows machine learning models to capture potential nonlinear changes in strength associated with different curing times. Furthermore, the quartile structure indicates that the experimental observations are distributed across the entire curing interval rather than confined to a narrow segment of the domain.

The initial moisture ranges from 11.2% to 15.2%, with a mean and median of 13.2% and a standard deviation of 1.6421%. The first and third quartiles are 11.2% and 15.2%, respectively. Compared with the curing time, the smaller standard deviation indicates a narrower experimental range. Nonetheless, the dataset encompasses the entire moisture-content range considered in the study, enabling the predictive models to effectively account for variations in peak shear strength associated with different initial moisture states. The equality of the mean and median at 13.2% further indicates a relatively balanced distribution of this variable within the examined experimental range. From a modeling perspective, it is important to consider the relatively narrow variation in initial moisture when interpreting feature importance. While a variable may have a significant physical impact on shear strength, it may demonstrate lower predictive importance simply because its experimental range is more limited than that of other variables. The fiber content in the dataset ranges from 0 to 1%, with a mean and median of 0.5% and a standard deviation of 0.3555%. The first and third quartiles are 0.25% and 0.75%, respectively, and the maximum value is 1%. This statistical distribution indicates that the dataset encompasses the investigated range of fiber content, including specimens without any fiber reinforcement and those with

the maximum fiber content. Consequently, the dataset serves as a robust experimental foundation for assessing the influence of fiber reinforcement on peak shear strength. Confining pressures range from 100 to 300 kPa, with both a mean and median value of 200 kPa, and a standard deviation of 82.1071 kPa. The first and third quartiles are 100 kPa and 300 kPa, respectively. The fact that the mean and median coincide at 200 kPa indicates a balanced distribution around the central experimental condition. Furthermore, the relatively large standard deviation highlights considerable variation in the applied confinement. This range is particularly significant for predicting triaxial shear strength, as confining pressure directly determines the stress environment under which specimens achieve peak shear resistance. Consequently, the broad interval of 100 to 300 kPa allows machine learning models to effectively capture the behavior of reinforced silty sand across markedly different confinement levels.

Overall, Table 1 shows that the dataset provides a coherent and sufficiently variable experimental foundation for developing machine learning models to predict peak shear strength in fly ash-reinforced silty sands. With a complete set of 90 observations, notable variation across the four input parameters, and a wide range of peak shear strength values, this dataset serves as an excellent basis for comparative model assessment. Conversely, the comparatively limited variation in initial moisture and curing time corresponds with a lower estimated importance. Importantly, the findings regarding normality and the absence of extreme residual outliers further validate the dataset's reliability. However, the identified heteroscedasticity underscores a limitation to consider when evaluating prediction errors at different response levels. Overall, the statistical evidence presented in Table 1 lays a transparent quantitative groundwork for the subsequent machine learning analysis, reinforcing the credibility of the reported predictive results by clearly delineating the variability and limitations inherent in the experimental database.

Table 2 presents the statistical tests conducted to evaluate the distributional characteristics and the dataset's variance behavior. The Shapiro–Wilk test yielded a test statistic of 0.95352 and a p-value of 0.48288, indicating that the dataset is normally distributed. This finding suggests that the distribution under analysis does not exhibit a statistically significant departure from normality. Furthermore, this result aligns with the descriptive statistics in Table 1, where the mean peak shear strength is 1764.57 kPa and the median is 1737.65 kPa, indicating they are relatively close. The target variable demonstrates a considerable yet structured range from 1112.7 to 2621.6 kPa. Consequently, the Shapiro–Wilk result provides quantitative support for the conclusion that the dataset does not exhibit pronounced overall skewness or significant non-normality.

Table 2. Dataset statistical tests.

Test	Statistic	p-value	Interpretation
Shapiro-Wilk	0.953520894	0.482883334	Normal
D'Agostino	0.981684725	0.612110558	Normal
Breusch-Pagan	9.744138828	0.044965385	Heteroscedastic

The D'Agostino test independently reinforces the same conclusion. The reported statistic is 0.98168 with a p-value of 0.61211. The alignment between the Shapiro–Wilk and D'Agostino results is particularly noteworthy, as both tests provide independent statistical evidence regarding the distributional assessment. Rather than relying on a single normality test, the agreement among results suggests that the overall dataset is statistically consistent with a normal distribution. This consistency bolsters confidence that the dataset does not display any extreme distributional anomalies that could complicate the interpretation of subsequent statistical and predictive analyses.

In contrast to the two normality tests, the Breusch–Pagan test reveals a statistically significant deviation from constant variance. The reported test statistic is 9.74414, with a p-value of 0.04497. Consequently,

we reject the null hypothesis of homoscedasticity at the 5% significance level and classify the dataset as heteroscedastic in Table 2. This finding suggests that the variability in the response or model residuals is not uniform across the experimental domain. In practical terms, the magnitude of prediction errors may fluctuate depending on the level or combination of the input variables and/or the predicted peak shear strength.

4. Results and discussion

Table 3 provides a summary of the predictive performance of eight machine learning techniques designed for peak shear strength estimation of the of fly ash-reinforced silty sands. Accompanying this summary is a bar chart that visually compares the predictive effectiveness of each model. The techniques evaluated include Support Vector Regression, Multilayer Perceptron, Gradient Boosting, XGBoost, Random Forest, Decision Tree, K-Nearest Neighbors, and Linear Regression. The assessment was conducted using the coefficient of determination (R^2) for the training, all-data, and testing subsets, along with mean absolute error (MAE) and root mean square error (RMSE), which are crucial, as R^2 measures the proportion of variability captured by the model. At the same time, MAE and RMSE assess the magnitude of prediction errors in terms of the original peak shear strength units. Thus, considering all three indicators provides a more comprehensive basis for ranking models than relying solely on a single statistical measure.

Table 3. Model performance metrics.

Model	Train R^2	All R^2	Test R^2	MAE	RMSE
Gradient Boosting	0.98079165	0.9769802	0.871346853	92.59993487	110.0978835
Random Forest	0.974851701	0.979868857	0.755169796	126.6270556	151.8802349
XGBoost	0.999999949	0.999999441	0.746801814	126.7863471	154.4539637
Decision Tree	1	1	0.621574012	156.5222222	188.8248895
KNN	0.725851385	0.721047565	0.525598399	185.3033333	211.4179395
Linear Regression	0.533942407	0.505781485	0.294128396	229.6273895	257.8882876
MLP	0.155766128	0.143412396	0.064488856	234.8720547	296.8884111
SVR	0.012929992	0.011522789	0.002245	262.2686753	306.6060493

The results clearly indicate that Gradient Boosting is the most dependable model on the independent test subset. It achieved a training R^2 of 0.98079, an overall R^2 of 0.97698, and a test R^2 of 0.87135, along with an MAE of 92.60 kPa and an RMSE of 110.10 kPa. The test R^2 of approximately 0.87 indicates that the model captured a significant portion of the variability in peak shear strength in the test data. More importantly, it boasts the lowest MAE and RMSE among the eight models analyzed, indicating that its superiority is not merely due to a favorable R^2 value. The combination of the highest test R^2 and the lowest error magnitudes provides compelling evidence that Gradient Boosting delivers the most balanced predictive performance within the experimental dataset.

The Random Forest algorithm obtained a test R^2 of 0.75517, accompanied by an MAE of 126.63 kPa and an RMSE of 151.88 kPa. While it effectively captures nonlinear interactions, Gradient Boosting outperforms it in predictive accuracy. XGBoost demonstrates nearly perfect training and an overall R^2 of 0.9999; however, it drops to a test R^2 of 0.74680, suggesting a risk of overfitting as it struggles on unseen data. The Decision Tree achieved a training R^2 of 1.0 but dropped to 0.62157 on the test set, suggesting it was memorizing the training data rather than generalizing to new observations. The K-Nearest Neighbors model showed intermediate results, with a training R^2 of 0.72585 and a test R^2 of 0.52560, indicating a lower predictive capability than tree-based methods. Linear Regression performed poorly, yielding a test R^2 of 0.29413 and an MAE of 229.63 kPa, which indicates that linear models inadequately capture the variability in peak shear strength. Finally, the Multilayer Perceptron exhibited limited performance across

all metrics, struggling to learn relationships from the dataset, evidenced by a very low test R^2 of 0.06449. In summary, these results underscore the superiority of nonlinear models, particularly Gradient Boosting, over simpler approaches in predicting peak shear strength.

Figure 2 illustrates the predicted peak shear strength values derived from the top-performing machine learning model, Gradient Boosting. The actual-versus-predicted graph is one of the most insightful diagnostic tools for evaluating the predictive reliability of a regression model, as it enables a direct comparison between laboratory measured peak shear strength and model-predicted values. The strong predictive performance, as illustrated by the actual-predicted plot in Figure 2, aligns well with the numerical results presented in Table 3. The test R^2 value of 0.87135 indicates that Gradient Boosting accounts for a significant portion of the variability in the unseen observations. Notably, this level of performance is accompanied by the lowest MAE and RMSE among all eight models examined. The statistical characteristics of the dataset lend further support to the interpretation of Figure 2. The Shapiro–Wilk and D’Agostino tests provided no significant evidence against normality, with p-values of 0.48288 and 0.61211, respectively. However, the Breusch–Pagan test indicated statistically significant heteroscedasticity, with a p-value of 0.04497. This combination suggests that while the overall data distribution aligns with normality, the variability related to the response or prediction process may not remain constant across the experimental domain.

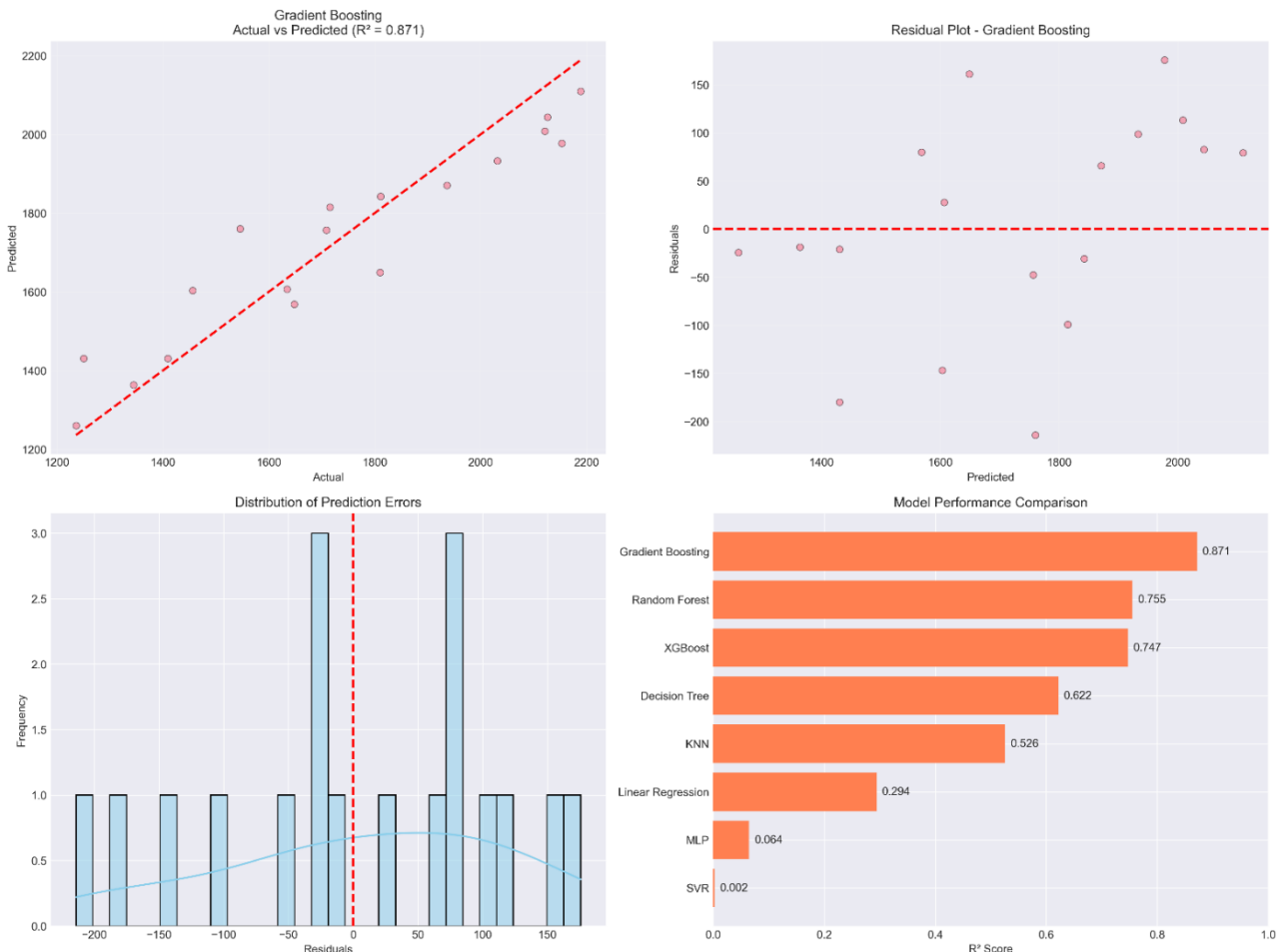


Fig. 2. Basic plots of the top model (Gradient Boosting).

The residual plot derived from the top-performing Gradient Boosting algorithm for forecasting the peak shear strength of fly ash-stabilized silty sands (Figure 2) offers a valuable diagnostic complement to the actual-versus-predicted plots and the numerical performance indicators presented in Table 3. The relatively low residual error aligns with the high coefficient of determination achieved by the Gradient

Boosting model, which explains approximately 87% of the variance in the independent testing data, as evidenced by a test R^2 of 0.87135. This indicates that the model captures most of the experimentally observed variation in peak shear strength, while the residual component accounts for a significantly smaller portion of the variability in the response. Consequently, A robust predictive model should not only achieve a high R^2 but must also produce residuals that do not display any clear systematic structure. The combination of a high test R^2 , low error magnitudes, and thorough residual diagnostics forms a solid foundation for considering Gradient Boosting a reliable predictive model within the examined experimental domain. The absence of detected extreme residual outliers is another important strength of the best model.

The residual histogram of the top-performing model, illustrated in Figure 2, provides a valuable opportunity to assess whether the observed numerical superiority is attributable to a well-controlled error distribution. The magnitude of the residual variability for the best model is relatively small compared with the overall variability of the target variable. Specifically, the residual standard deviation is 113.14 kPa, while the standard deviation of the experimentally measured peak shear strength is 352.62 kPa. This comparison is significant as it suggests that the scale of the prediction errors is considerably smaller than the inherent variation in peak shear strength across the experimental database. Consequently, the model effectively captures a substantial portion of the observed variability in strength, leaving only a minor residual component unexplained. This finding aligns with the test R^2 of 0.87135 and provides quantitative support for interpreting the residual histogram as indicative of a relatively concentrated distribution of prediction errors.

The most noteworthy finding presented in Figure 3 is the significant influence of fiber content, which contributes 0.58741 to the overall feature-importance measure. This value indicates that fiber content is considerably more impactful than any other variable evaluated by the Gradient Boosting model. In relative terms, approximately 58.7% of the model's total feature importance is attributed to fiber content, compared with 28.6% for confining pressure, 10.3% for initial moisture content, and merely 2.4% for curing time. This clear hierarchy suggests that incorporating fiber reinforcement into silty sand is the primary factor driving variations in peak shear strength, as the predictive model captures. The significance of fiber content closely aligns with the study's engineering objectives. The material being examined is a fly ash-reinforced silty sand, and the fiber content directly indicates the degree of reinforcement in the soil matrix. The experimental dataset includes fiber contents ranging from 0% to 1%, with an average of 0.5% and a standard deviation of 0.3555%. The notable importance of this parameter indicates a strong correlation between variations in reinforcement content and changes in peak shear strength within the experimental framework. From a geotechnical perspective, this observation is physically plausible; incorporating discrete fibers can alter the internal load-transfer mechanism, enhance interparticle resistance, and improve the mobilization of shear resistance through interactions between the soil matrix and the reinforcing elements. As a result, the machine learning analysis reveals that fiber content is not only statistically significant but also the primary explanatory factor for the variations in strength observed in the dataset.

The second most influential variable in predicting peak shear strength is confining pressure, which has a feature importance of 0.28550, accounting for approximately 28.6% of the total. This finding is consistent with theories derived from unconsolidated undrained triaxial testing, in which increased confinement elevates normal stress, thereby affecting particle interlocking and frictional resistance. Together, fiber content and confining pressure comprise about 87.3% of the predictive information, highlighting their critical roles in experimental design. Initial moisture content ranks third, with a feature importance of 0.10286, representing 10.3% of the total. Its influence indicates meaningful interactions within the soil-fiber-fly ash system, though it remains less significant than fiber content and confining pressure. The variability of moisture underscores the need to incorporate it into predictive models. Curing time ranks

lowest, with a feature importance of 0.02423 (2.4% of the total importance), suggesting it has lower predictive value within the studied range. However, a low importance score does not imply that curing time lacks physical significance; rather, it indicates a lesser role in predictions compared to the other variables. Overall, these findings underscore the importance of prioritizing specific variables in experimental design to enhance our understanding of peak shear strength.

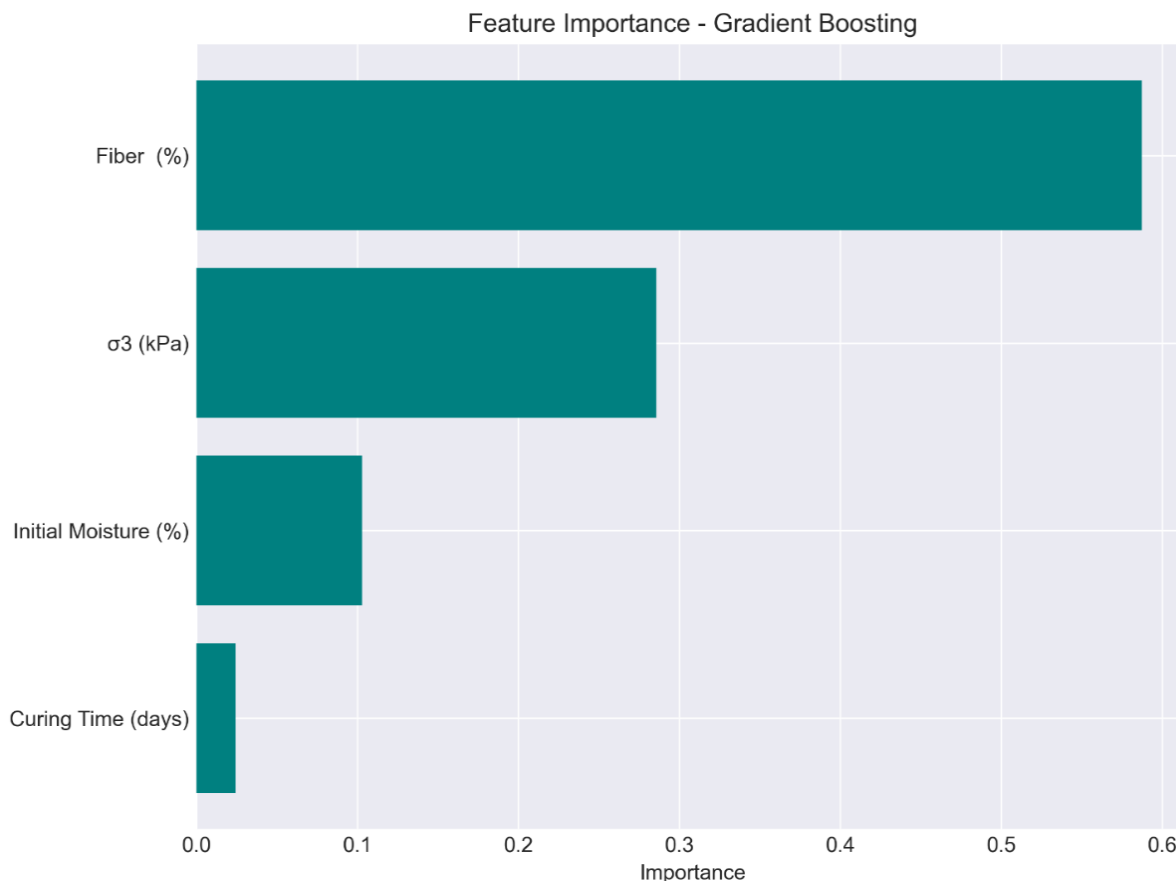


Fig. 3. Feature importance plots of the top model (Gradient Boosting).

Figure 4 illustrates the partial dependence plots for the four input variables in the best-performing Gradient Boosting model, providing an additional layer of interpretation beyond predictive performance metrics and feature-importance analyses. The partial dependence plot for curing time shows a roughly monotonic decrease as model-predicted peak shear strength increases. As the normalized curing time increases, the partial dependence decreases, indicating reduced predictions of peak shear strength. The relationship appears smooth and nearly linear, signifying the absence of any significant threshold in the marginal influence of curing time. Interestingly, despite having the lowest feature-importance value of 0.02423, the partial dependence plot still reveals a clear trend. This distinction is crucial, as feature importance assesses a variable's overall contribution, while the partial dependence plot represents the average response as that variable changes. Consequently, a variable may exhibit a noticeable marginal trend even if it has low overall importance, particularly if its predictive information overlaps with that of stronger predictors. It's important to note that this trend is derived from a limited experimental range of 0–15 days. While the partial dependence plot reflects this specific range, it does not guarantee the same trend beyond it, and additional data would be necessary for extrapolation. Therefore, the conclusion is that Gradient Boosting has identified a decreasing association between curing time and predicted peak shear strength within the tested conditions.

The partial dependence plot for initial moisture content demonstrates a clear decreasing trend: as the normalized initial moisture value increases, the predicted peak shear strength diminishes. This reveals a

consistent negative relationship, indicating that higher initial moisture levels reduce the predicted strength of the soil-fly ash-fiber mixture. This observation is consistent with the physical role of moisture in soil systems, as it influences interparticle interactions and resistance during shearing. The model identifies initial moisture as a significant predictor, with a feature-importance value of 0.10286, underscoring its importance rather than insignificance. The dataset encompasses a narrow range of initial moisture contents, from 11.2% to 15.2%, and this plot illustrates systematic variations in predicted peak shear strength across these moisture levels.

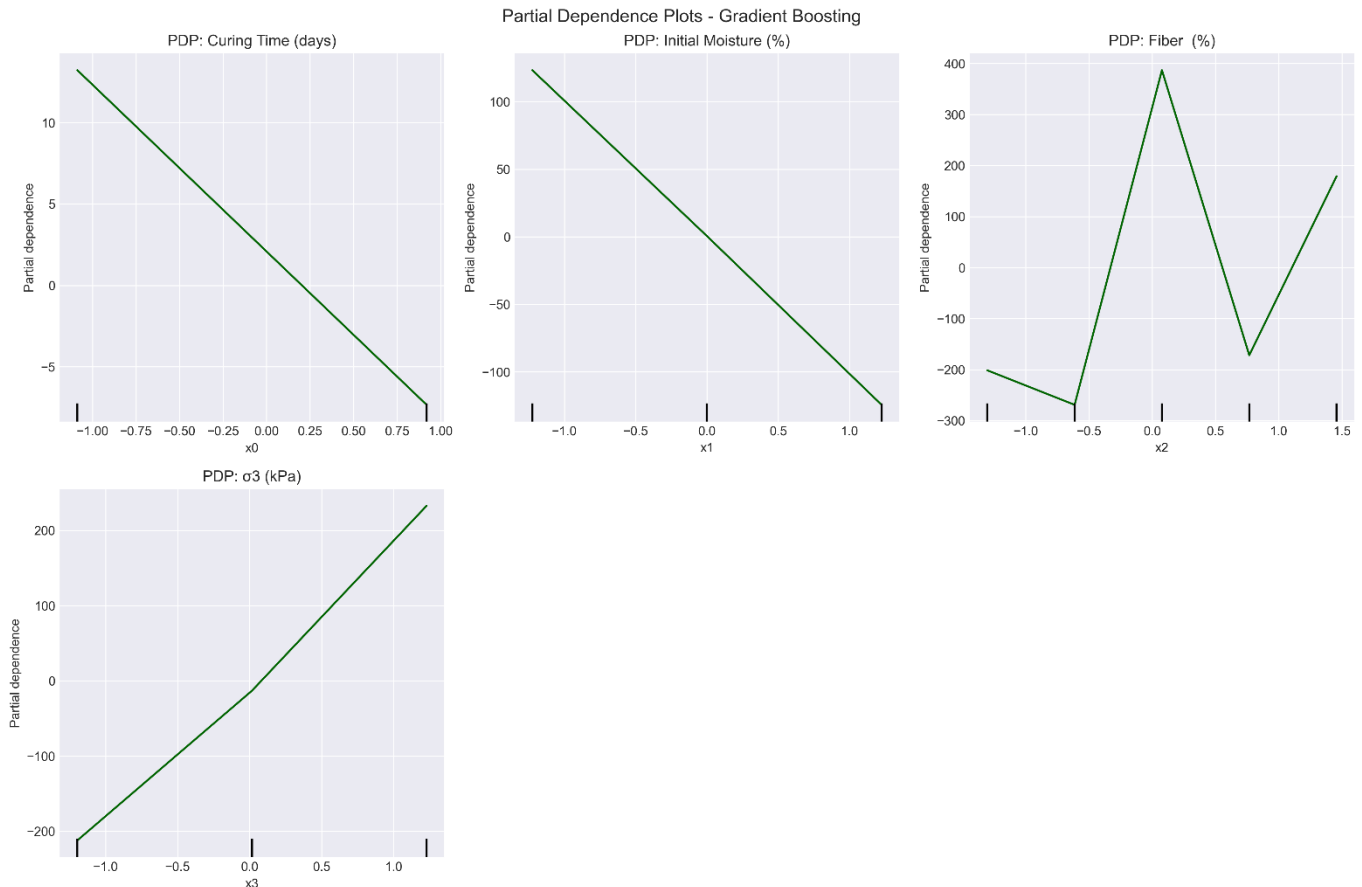


Fig. 4. Partial dependence plots of the top model (Gradient Boosting).

The partial dependence plot for fiber content shows a clearly nonlinear pattern, as illustrated in Figure 4. In contrast to the trends observed for curing time and moisture content, the curve for fiber content starts negative, rises sharply to a peak midway through the range, then decreases before rising again at higher levels. This nonlinear behavior suggests that fiber content significantly influences the predicted peak shear strength, indicating a relationship that goes beyond a simple linear relationship. With a feature importance score of 0.58741, fiber content emerges as the most significant predictor in the model. This explains why nonlinear methods, such as Gradient Boosting, which achieved a test R^2 of 0.87135, outperform linear regression, which attained only 0.29413. Although the experimental range for fiber is narrow (0 to 1%, with a mean of 0.5%), the partial dependence plot reflects pronounced changes, indicating that the model effectively captures distinct responses across different reinforcement levels. This non-monotonic behavior may imply complex interactions with the soil matrix, although it does not definitively confirm any specific micromechanical mechanism. The marked shifts observed in the fiber partial dependence plot are attributable to the tree-based nature of Gradient Boosting, highlighting that the model's responses vary significantly as inputs cross decision boundaries. Therefore, these fluctuations should not be interpreted as physical transitions but rather as changes in model predictions across the fiber-content spectrum, which warrants caution when overinterpreting the results. The partial dependence

plot for confining pressure shows a clear, monotonic increase in predicted peak shear strength as cell pressure increases. This trend reinforces the fundamental role of confining pressure in triaxial testing, where pressures ranging from 100 to 300 kPa are utilized, indicating that greater confinement enhances shear resistance. With a feature importance score of 0.28550, confining pressure ranks as the second most influential predictor, following fiber content, and together they account for approximately 87.3% of the model's feature importance. This positive correlation highlights the importance of both the reinforcement condition and the applied stress state on peak shear strength. Furthermore, the upward trend emphasizes the suitability of nonlinear ensemble learning for this analysis. While the partial dependence plot shows a smooth, monotonic pattern, the model also accounts for additional factors such as fiber content, moisture, and curing time, resulting in a multidimensional response that Gradient Boosting effectively captures and yields robust predictive performance.

Figure 5 illustrates the sensitivity analysis of the Gradient Boosting model, which has been identified as the most effective algorithm for forecasting the peak shear strength of fly ash-reinforced silty sands. The predicted peak shear strength exhibits a step-like response over the curing time range, remaining relatively stable until a significant increase is observed in the middle-to-upper segments. This behavior is indicative of the decision-tree structure inherent in Gradient Boosting. While curing time does influence predictions, its feature importance is notably low (0.02423) compared with fiber content and confining pressure. The curing time ranges from 0 to 15 days, with a median of 7.5 days. The sensitivity analysis suggests that variations in predicted strength with curing time are limited, indicating that extending the curing period is not a primary factor affecting peak shear strength. This finding does not diminish the physical significance of curing time; rather, it suggests that enhancing data on fiber content and confining pressure could provide greater predictive insights. Curing time should continue to be included in future models, particularly for wider curing conditions or varying fly ash compositions.

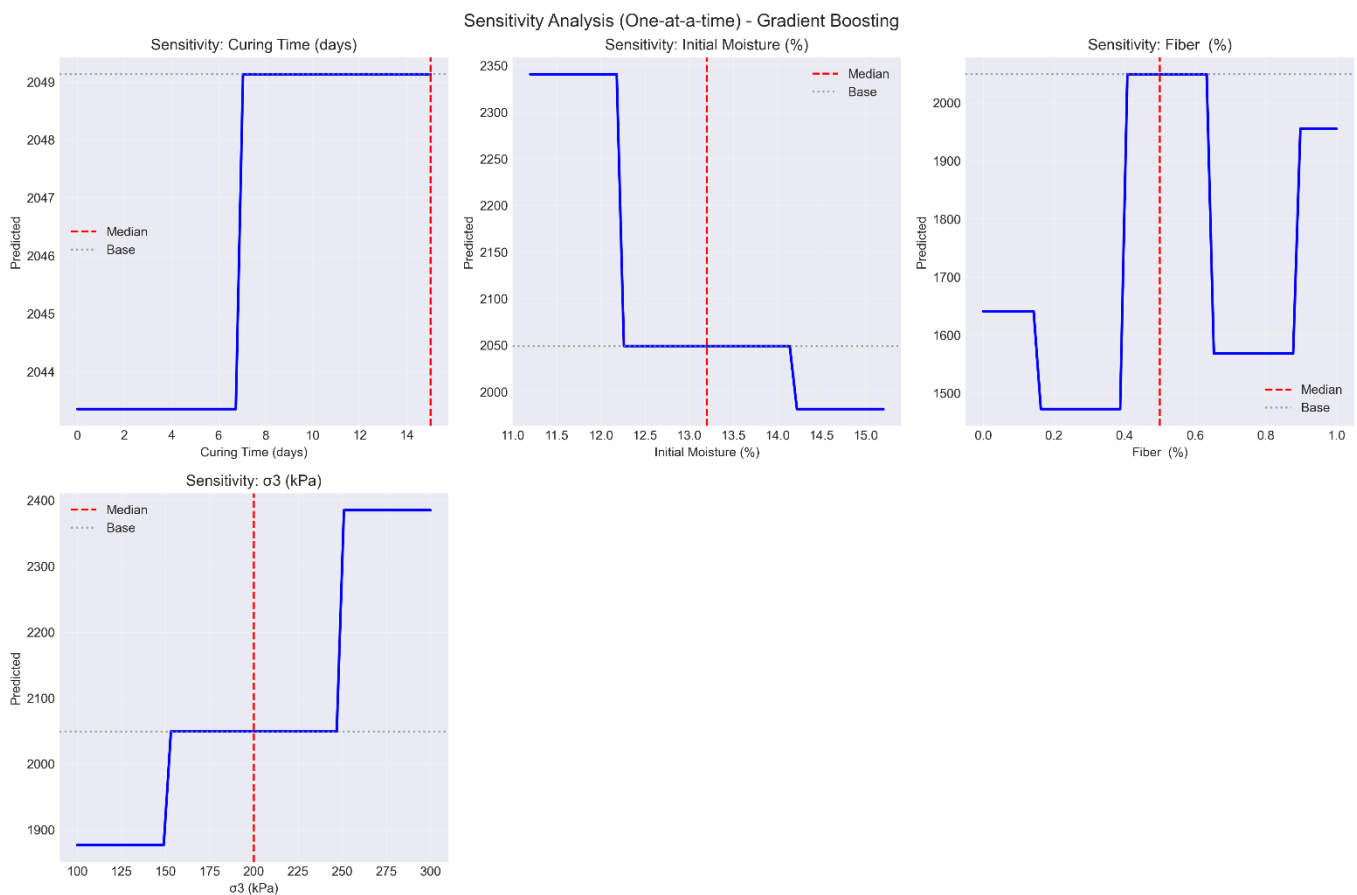


Fig. 5. Sensitivity analysis plots of the top model (Gradient Boosting).

As the moisture content increases within the experimental range, the predicted peak shear strength decreases significantly. This reduction occurs in stepwise declines, with stable intervals interrupted by abrupt drops, ultimately reaching the lowest strength levels at the highest moisture contents. Although moisture content ranks third in feature importance, its impact on predicted strength is meaningful. The range is 11.2-15.2%, with a median of 13.2%. The sensitivity analysis curve shows that transitioning from lower to higher moisture can substantially reduce model predictions. The feature-importance value of 0.10286 indicates a smaller contribution relative to fiber content and confining pressure, but does not imply that moisture variations are insignificant. Increased moisture affects interparticle contacts and effective stress during shearing, reducing resistance in the soil-fly ash-fiber system. This consistent negative response from the model links data-driven analysis with traditional geotechnical insights.

The fiber-content panel in Figure 5 reveals significant findings: the predicted peak shear strength varies substantially and non-monotonically as fiber content changes from 0 to 1%. Initially, the response decreases, then increases at intermediate levels, decreases again, and finally increases at higher fiber contents. This pattern indicates a strong nonlinear relationship between fiber content and peak shear strength, challenging the assumption of a simple monotonic trend. In the feature-importance analysis, fiber content emerges as the most influential variable, with an importance score of 0.58741, underscoring its critical role in driving the model's predictions. With experimental fiber content ranging from 0 to 1% and a median of 0.5%, the model forecasts peak shear responses across this narrow range. From a geotechnical standpoint, the non-monotonic behavior of the fiber response suggests that adding fibers does not necessarily yield a consistent increase in strength. The effectiveness of reinforcement is affected by fiber concentration and its interaction with soil conditions. While certain levels may increase strength, others may decrease it, reflecting the complex interplay among fiber content, soil fabric, confinement, and shear resistance.

The sensitivity analysis of confining pressure reveals that increasing confinement within this range significantly enhances the predicted peak shear strength. This finding aligns with the feature-importance value of 0.28550, positioning confining pressure as the second most influential input variable, following fiber content. Thus, the two key predictors identified by the machine-learning model serve distinct roles: fiber content dictates a highly nonlinear reinforcement response, while confining pressure yields a strong, generally positive response. The increasing sensitivity to confinement is physically justified, as greater confining pressure increases the normal stress state acting on the soil specimen, thereby facilitating the mobilization of shear resistance. In an unconsolidated–undrained triaxial test, the applied confinement is directly related to the stress environment in which peak strength is achieved. The Gradient Boosting model's ability to replicate this expected positive relationship enhances our confidence that it captures physically meaningful patterns rather than merely exploiting statistical correlations.

Figure 6 illustrates the boxplot distribution of prediction residuals for the eight machine learning models developed to forecast the peak shear strength of fly ash-reinforced silty sands. This analysis offers a crucial complementary evaluation of model reliability. While conventional performance indicators such as R^2 , MAE, and RMSE summarize the overall magnitude of prediction error, they do not adequately convey its distribution, central tendency, dispersion, asymmetry, or the existence of potentially extreme errors. The most significant observation from Figure 6 is that Gradient Boosting demonstrates the most favorable overall residual distribution among the eight models examined. Its box is notably compact, with the median residual positioned relatively close to the zero-error reference line, and its overall residual range is considerably narrower than that of the weaker models. This visual evidence aligns perfectly with the quantitative results presented in Table 3, where Gradient Boosting achieved the highest test R^2 of 0.87135 and the lowest MAE and RMSE values of 92.60 and 110.10 kPa, respectively. The consistency between the numerical performance metrics and the residual distribution is especially crucial, as it indicates that

the superior performance of Gradient Boosting is not merely a product of a favorable R^2 value; the model also yields a more controlled and concentrated error distribution.

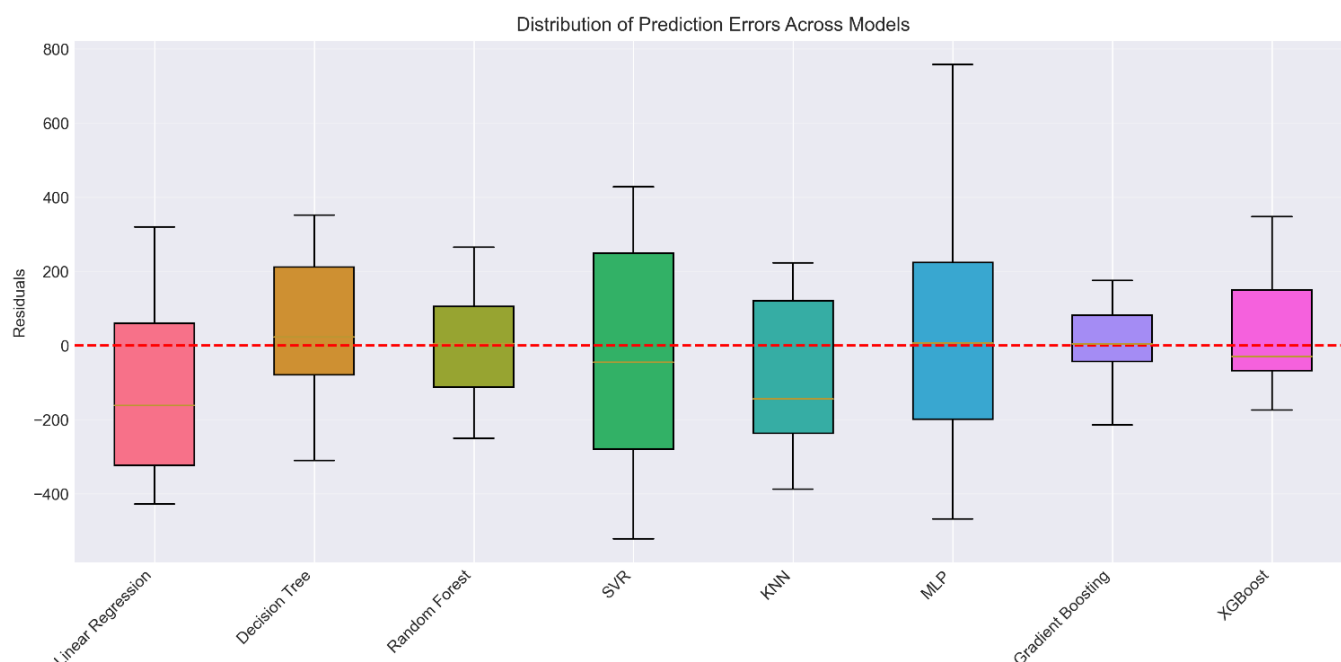


Fig. 6. Error distribution plots of all models.

The boxplot highlights an important distinction between central accuracy and error variability. A predictive model may exhibit a median residual that is reasonably close to zero, yet still show a wide interquartile range or lengthy whiskers. Consequently, evaluating the overall reliability of a predictive model should not be based solely on whether the median residual approaches zero. In this study, Gradient Boosting is preferred due to its combination of a relatively centered residual distribution and lower dispersion. This is more desirable than merely achieving a near-zero median while allowing for considerable positive and negative errors. Moreover, this comparison reinforces the conclusion that ensemble learning effectively balances flexibility and generalization for this dataset. Random Forest, XGBoost, and Gradient Boosting all outperform the individual Decision Tree; however, Gradient Boosting distinguishes itself with the strongest test performance and the most favorable residual distribution. This enhancement is particularly noteworthy, considering that both XGBoost and the Decision Tree attain nearly perfect training R^2 values. Yet, their test performance and residual distributions are inferior to those of Gradient Boosting. This evidence indicates that model selection should be based on performance with unseen data and error behavior, rather than solely on training fit.

Figure 7 presents a radar chart displaying the normalized performance metrics of the eight machine-learning models. The chart clearly highlights Gradient Boosting as the leading model, with its profile extending furthest toward the outer boundaries across the three performance dimensions, thereby forming the largest overall polygon. This visual representation is entirely consistent with the numerical findings detailed in Table 3, where Gradient Boosting achieved a test R^2 of 0.87135, a MAE of 92.60 kPa, and an RMSE of 110.10 kPa. The alignment between the radar chart and the numerical performance indicators provides compelling evidence that the choice of Gradient Boosting as the premier predictive model is not solely reliant on a single statistical metric. Rather, it simultaneously offers the highest explained variance along with the lowest absolute and root-mean-square prediction errors among the eight techniques analyzed.

The most notable characteristic of Figure 7 is the prominent and relatively balanced triangular profile attributed to Gradient Boosting. Its R^2 axis achieves the highest normalized level, while the inverse-MAE

and inverse-RMSE axes also extend further than those of the competing algorithms. This balanced geometry is particularly noteworthy, as a model could theoretically excel in one metric while underperforming in another. However, Gradient Boosting does not demonstrate this trade-off to any significant extent in the current dataset.

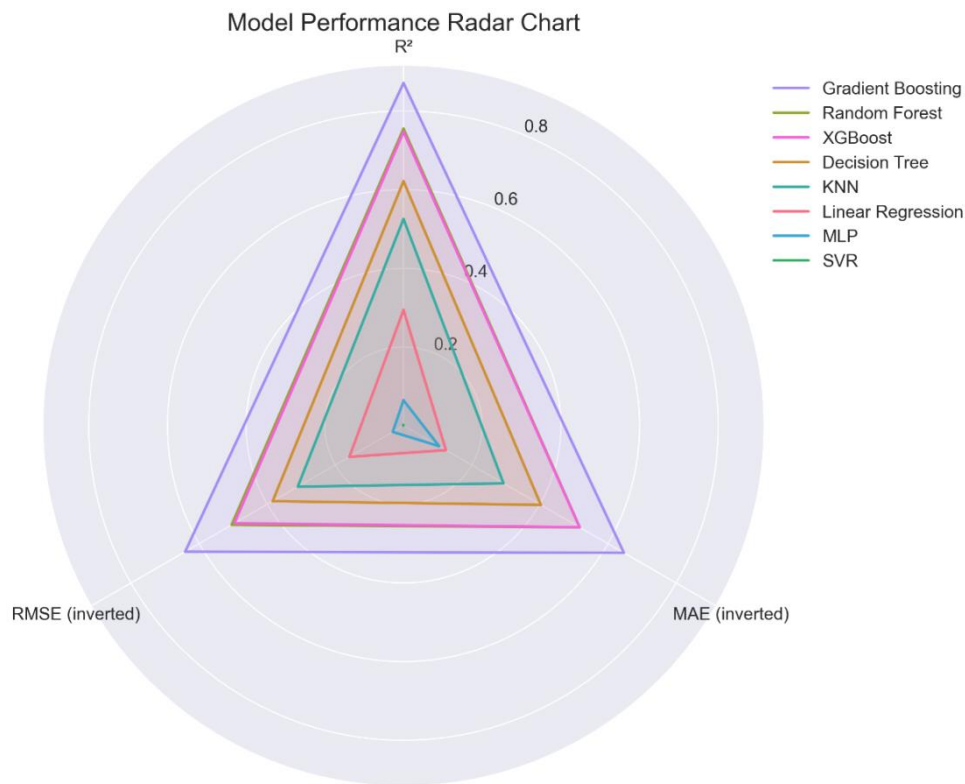


Fig. 7. Radar chart of all models.

Figure 8 presents a crucial analysis for detecting outliers, serving as the final diagnostic assessment of the reliability and stability of the Gradient Boosting model, which has been identified as the most effective algorithm for fly ash-reinforced silty sands. Unlike traditional performance metrics such as R^2 , MAE, and RMSE that measure overall predictive accuracy, outlier analysis specifically investigates whether the perceived predictive performance is impacted by a small number of significantly large errors. This distinction is particularly vital for geotechnical datasets, as experimental measurements may encompass observations of rare combinations of material properties or testing conditions. In this study, the outlier analysis was conducted using the residual standard deviation and a 2σ threshold. Table 5 indicates a residual standard deviation of 113.1400 kPa, with a corresponding 2σ threshold of 226.2800 kPa, revealing zero detected outliers and an outlier ratio of 0.00%. These findings strongly suggest that the chosen Gradient Boosting model's predictive behavior is stable and unaffected by a small subset of anomalous predictions.

The residual standard deviation of 113.14 kPa reflects the overall dispersion of prediction errors around their central tendency. With the target peak shear strength ranging from 1112.7 to 2621.6 kPa, and a mean of 1764.57 kPa coupled with a standard deviation of 352.62 kPa, it is evident that the residual dispersion is significantly smaller than the intrinsic variability of the target variable. This finding is crucial for demonstrating predictive effectiveness.

While the experimental peak shear strengths exhibit considerable variability, the Gradient Boosting model captures most of this variation and yields substantially less residual variability. The established threshold of $2\sigma = 226.28$ kPa provides an objective criterion for identifying unusually large prediction errors. As indicated in Table 5, no residuals surpass this threshold, meaning that zero observations are classified as

outliers. This conclusion is also supported by the plot, which shows that all residual points fall within the $\pm 2\sigma$ bounds, with the dashed horizontal line representing the zero-residual reference. As a result, both the graphical and numerical outlier analyses are consistent. This outcome is significant because it demonstrates that the favorable performance of Gradient Boosting is not due to accurately fitting most observations while allowing a small number of observations to have extremely large errors. Instead, the errors consistently remain within the predefined acceptable statistical range across the observations analyzed.

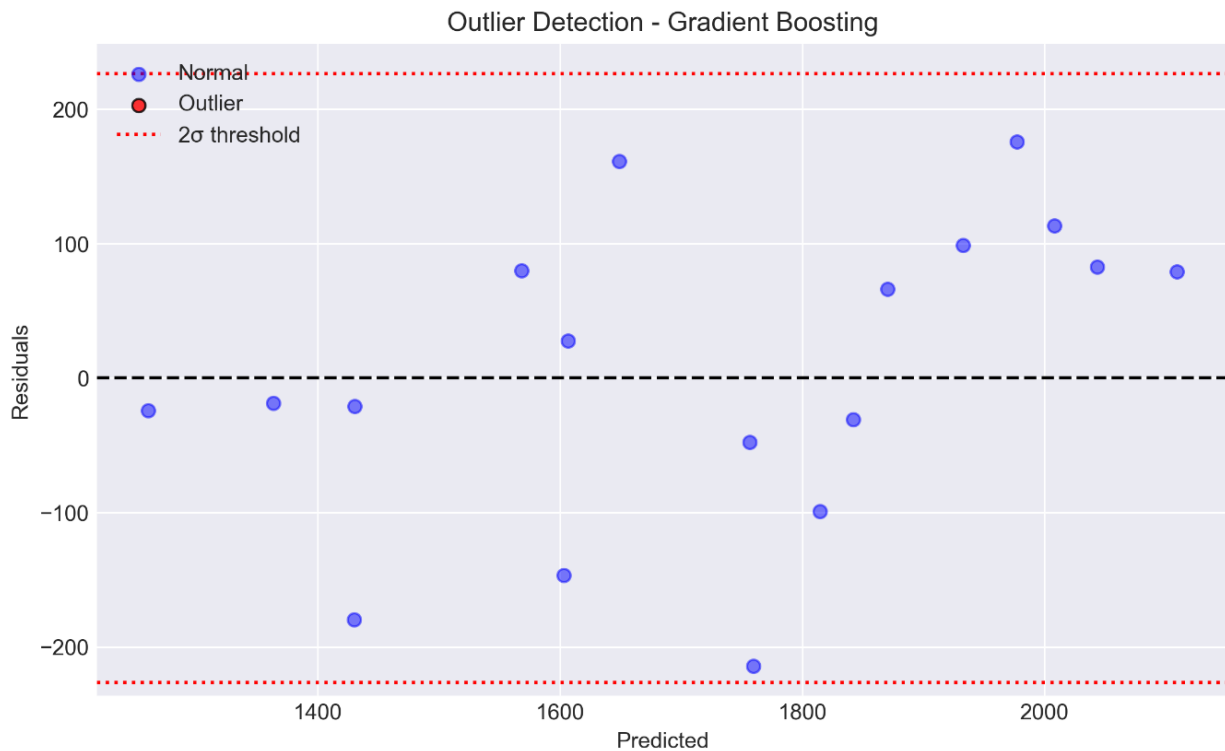


Fig. 8. Outlier detection plot of the top model (Gradient Boosting).

The most significant finding is the complete absence of outliers. With a residual standard deviation of 113.14 kPa and a 2σ threshold of 226.28 kPa, none of the predictions were classified as outliers. Consequently, the outlier ratio of 0.00% strongly indicates consistency in prediction errors within the established criteria. This outcome is particularly noteworthy when considered alongside the model's performance metrics. Gradient Boosting achieved a test R^2 of 0.87135, an MAE of 92.60 kPa, and a RMSE of 110.10 kPa, marking it as the top performer among the eight algorithms analyzed. Thus, the model not only exhibits high predictive accuracy but also shows no statistically significant extreme residuals. The connection between RMSE and the outlier analysis is also important to highlight. RMSE is particularly sensitive to large prediction errors because residuals are squared before averaging. Therefore, the relatively low RMSE of 110.10 kPa, coupled with the absence of residuals exceeding the 2σ threshold, provides further corroboration that the model does not exhibit a small number of exceptionally large prediction errors.

The findings from the outlier detection are well aligned with the previously presented residual box plot. The boxplot illustrates a relatively concentrated distribution of Gradient Boosting residuals, and Table 5 quantitatively affirms that none of the residuals exceed the established 2σ criterion. This consistency between two independent visual and statistical diagnostics significantly enhances confidence in the model. Additionally, the residual standard deviation of 113.14 kPa corresponds closely with the model's RMSE of 110.10 kPa. Although these figures are not mathematically identical, their similar magnitudes suggest that both the overall variability of the residuals and the magnitude of the prediction error are on a

comparable scale. Furthermore, the absence of observations beyond 2σ indicates that extreme values have little influence on the residual distribution.

The analysis highlights a notable disparity between the training and testing performances of tree-based algorithms. The Decision Tree and XGBoost models attained high training R^2 values (1.0 and 0.999, respectively), yet exhibited considerable declines in testing R^2 (0.622 and 0.747), suggesting overfitting due to their over-reliance on limited training data. The Decision Tree's tendency to overfit stems from its flexible structure, which can excessively divide the predictor space. Although XGBoost addresses some of these concerns through boosting techniques, it still poses a risk of overfitting when its complexity surpasses the dataset size. In contrast, Gradient Boosting produced a lower training R^2 (0.981) but achieved a significantly higher testing R^2 (0.871), coupled with the lowest MAE and RMSE. This outcome illustrates a more favorable balance between flexibility and generalization. The study underscores the critical role of regularization, validation, and cross-validation in model development, particularly when working with small datasets. Ultimately, the selection of algorithms should prioritize out-of-sample predictive performance, positioning Gradient Boosting as the preferred model for this dataset.

5. Conclusions

This study explored the effectiveness of eight machine learning approaches, Support Vector Regression, Multilayer Perceptron, Gradient Boosting, XGBoost, Random Forest, Decision Tree, K-Nearest Neighbors, and Linear Regression, in predicting the shear strength of fly ash-reinforced silty sands. The analysis used 90 experimental observations from unconsolidated undrained triaxial tests, with fiber content, curing time, initial moisture content, and confining pressure as input variables, and peak shear strength as the target for prediction. The experimental database encompassed a wide range of conditions, including fiber content from 0% to 1%, curing times from 0 to 15 days, initial moisture content from 11.2% to 15.2%, and confining pressures from 100 to 300 kPa. Concurrently, the measured peak shear strength varied from 1112.7 to 2621.6 kPa. This combination of variables allowed the machine-learning models to effectively capture the effects of reinforcement, material conditions, and confinement on the observed shear-strength response.

From a geotechnical engineering standpoint, the findings highlight the potential of machine learning as a valuable supplementary tool for estimating peak shear strength, particularly when direct experimental determination is prohibitively expensive, time-consuming, or challenging across all conceivable combinations of materials and testing conditions. The recognized significance of fiber content and confining pressure is especially pertinent to the development of predictive frameworks for reinforced silty sands. This model can aid in preliminary assessments and experimental planning by providing strength estimates within the parameter ranges represented in the database. However, it is important to note that these results should not be seen as a substitute for laboratory testing or a comprehensive understanding of material behavior. Instead, the machine learning framework should be viewed as a data-driven predictive complement to traditional geotechnical characterization methods.

The principal conclusions of this study, derived from comprehensive statistical, predictive, graphical, interpretability, and error analyses, can be summarized as follows:

- Gradient Boosting revealed as the top algorithm among the eight techniques evaluated, achieving a test R^2 of 0.87135, an MAE of 92.60 kPa, and a RMSE of 110.10 kPa. This performance highlights its superior predictive capability for peak shear strength.
- Overall, the nonlinear ensemble approaches consistently outperformed simpler linear models and less effective alternatives. Random Forest and XGBoost followed closely as the next most competitive models, with test R^2 values of 0.75517 and 0.74680, respectively. In contrast, Linear

Regression, Multilayer Perceptron, and Support Vector Regression demonstrated significantly poorer generalization.

- Training performance alone proved to be an unreliable indicator of predictive capability. Both the Decision Tree and XGBoost models attained exceptionally high R^2 values on the training and all data sets, yet showed significantly lower performance on the test data. This underscores the necessity of assessing model generalization using previously unseen observations.
- Among the model inputs, fiber content emerged as the most influential factor, with a feature importance of 0.58741. It was followed by confining pressure (0.28550), initial moisture (0.10286), and curing time (0.02423). Therefore, the predictive framework developed using Gradient Boosting was primarily driven by fiber content and confining pressure within the examined ranges.
- Error diagnostics further affirmed the robustness of the chosen model. The residual standard deviation was computed at 113.14 kPa, with a 2σ threshold of 226.28 kPa. Importantly, no residuals were identified as outliers. This indicates that the presence of extreme errors did not significantly affect the model's strong predictive performance.
- The statistical analysis demonstrated characteristics of both normality and heteroscedasticity. The Shapiro–Wilk and D'Agostino tests supported normality, while the Breusch–Pagan test indicated heteroscedasticity ($p=0.04497$). This finding underscores the necessity of employing multiple complementary diagnostic approaches when assessing predictive models for geotechnical experimental data.
- The overall results validate the appropriateness of machine learning for predicting peak shear strength in fly ash-reinforced silty sands, particularly when nonlinear ensemble methods are utilized. The consistency among model performance metrics, actual versus predicted behavior, residual diagnostics, feature importance, sensitivity analysis, radar chart comparisons, and outlier detection offers a more comprehensive foundation for model selection than any single evaluation metric.

Authors contribution statement

Masoud Nasiri: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft.

Masoud Ahmadi: Conceptualization, Data curation, Investigation, Methodology, Resources, Software, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies

During the preparation of this work, the authors used ChatGPT and Grammarly to improve the translation from Persian to English, as well as the writing and readability of some parts of the manuscript. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the article's content.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts 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] Shen Z, Pan Q, Chian SC, Gourvenec S, Tian Y. Probabilistic failure envelopes of strip foundations on soils with non-stationary characteristics of undrained shear strength. *Géotechnique* 2023;73:716–35. <https://doi.org/10.1680/jgeot.21.00169>.
- [2] Guler E, Hamderi M, Demirkan MM. Numerical analysis of reinforced soil-retaining wall structures with cohesive and granular backfills. *Geosynth Int* 2007;14:330–45. <https://doi.org/10.1680/gein.2007.14.6.330>.
- [3] Nasiri M, Hajiazizi M. An experimental and numerical investigation of reinforced slope using geotextile encased stone column. *Int J Geotech Eng* 2021;15:543–52. <https://doi.org/10.1080/19386362.2019.1651029>.
- [4] Hatami K, Esmaili D, Chan EC, Miller GA. Moisture Reduction Factors for Shear Strength of Unsaturated Reinforced Embankments. *Int J Geomech* 2016;16:D4016001. [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0000624](https://doi.org/10.1061/(ASCE)GM.1943-5622.0000624).
- [5] Nian T, Li J, Li P, Liu Z, Guo R, Ge J, et al. Method to predict the interlayer shear strength of asphalt pavement based on improved back propagation neural network. *Constr Build Mater* 2022;351:128969. <https://doi.org/10.1016/j.conbuildmat.2022.128969>.
- [6] Mortazavi A. Anchorage and shear strength properties for composite tendons used in earthwork support systems. *Constr Build Mater* 2007;21:109–17. <https://doi.org/10.1016/j.conbuildmat.2005.06.034>.
- [7] Fallah M, Mazaheri A, Lajvardi H, Nasiri M. Shear Strength Behavior of Oil-Contaminated Bushehr Carbonate Sand. *J Rehabil Civ Eng* 2025;13:177–89. <https://doi.org/10.22075/jrce.2025.35520.2180>.
- [8] Ahmaruzzaman M. A review on the utilization of fly ash. *Prog Energy Combust Sci* 2010;36:327–63. <https://doi.org/10.1016/j.pecs.2009.11.003>.
- [9] Mir BA, Sridharan A. Physical and Compaction Behaviour of Clay Soil–Fly Ash Mixtures. *Geotech Geol Eng* 2013;31:1059–72. <https://doi.org/10.1007/s10706-013-9632-8>.
- [10] Siddiqua S, Barreto PNM. Chemical stabilization of rammed earth using calcium carbide residue and fly ash. *Constr Build Mater* 2018;169:364–71. <https://doi.org/10.1016/j.conbuildmat.2018.02.209>.
- [11] Gangloff WJ, Ghodrati M, Sims JT, Vasilas BL. Impact of Fly Ash Amendment and Incorporation Method on Hydraulic Properties of a Sandy Soil. *Water Air Soil Pollut* 2000;119:231–45. <https://doi.org/10.1023/A:1005150807037>.
- [12] Qiu Z, Zhang L, Likos WJ, Yin J. Mechanical behavior and microstructural insights of bagasse fiber and fly ash-treated dredged clay. *Mar Georesources Geotechnol* 2026;44:2261–76. <https://doi.org/10.1080/1064119X.2026.2620582>.
- [13] Shaheen SM, Hooda PS, Tsadilas CD. Opportunities and challenges in the use of coal fly ash for soil improvements – A review. *J Environ Manage* 2014;145:249–67. <https://doi.org/10.1016/j.jenvman.2014.07.005>.
- [14] Yao Z, Ji X, Sarker PK, Tang JH, Ge LQ, Xia MS, et al. A comprehensive review on the applications of coal fly ash. *Earth-Science Rev* 2015;141:105–21. <https://doi.org/10.1016/j.earscirev.2014.11.016>.
- [15] Wang Y, Acarturk BC, Burris L, Hooton RD, Shearer CR, Suraneni P. Physicochemical characterization of unconventional fly ashes. *Fuel* 2022;316:123318. <https://doi.org/10.1016/j.fuel.2022.123318>.
- [16] Xu G, Shi X. Characteristics and applications of fly ash as a sustainable construction material: A state-of-the-art review. *Resour Conserv Recycl* 2018;136:95–109. <https://doi.org/10.1016/j.resconrec.2018.04.010>.
- [17] Vassilev S V, Vassileva CG. A new approach for the classification of coal fly ashes based on their origin, composition, properties, and behaviour. *Fuel* 2007;86:1490–512. <https://doi.org/10.1016/j.fuel.2006.11.020>.
- [18] Chancey RT, Stutzman PE, Juenger MCG, Fowler DW. Comprehensive phase characterization of crystalline and amorphous phases of a Class F fly ash. *Cem Concr Res* 2010;40:146–56. <https://doi.org/10.1016/j.cemconres.2009.08.029>.
- [19] Guo X, Shi H, Dick WA. Compressive strength and microstructural characteristics of class C fly ash geopolymer. *Cem Concr Compos* 2010;32:142–7. <https://doi.org/10.1016/j.cemconcomp.2009.11.003>.
- [20] Phani Kumar BR, Sharma RS. Effect of Fly Ash on Engineering Properties of Expansive Soils. *J Geotech Geoenvironmental Eng* 2004;130:764–7. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2004\)130:7\(764\)](https://doi.org/10.1061/(ASCE)1090-0241(2004)130:7(764)).

- [21] Rios S, Cristelo N, Viana da Fonseca A, Ferreira C. Stiffness Behavior of Soil Stabilized with Alkali-Activated Fly Ash from Small to Large Strains. *Int J Geomech* 2017;17:4016087. [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0000783](https://doi.org/10.1061/(ASCE)GM.1943-5622.0000783).
- [22] Li LH, Ning T, Zhan ZQ, Gui YL, Zhang C. Study on mechanical properties of fly ash and coir fiber improved expansive soil subgrade. *Acta Geotech* 2026;21:745–65. <https://doi.org/10.1007/s11440-025-02784-8>.
- [23] Wang X, Zhu S, Han Z, Zhang Y, Kai W, Yu X, et al. Volumetric, strength and microstructural characteristics of expansive soil modified by lignin fibers and fly ash. *J Rock Mech Geotech Eng* 2026;18:6634–50. <https://doi.org/10.1016/j.jrmge.2026.05.003>.
- [24] Yachang O, Kalita A, Taku M. Synergistic Effect of Cement-Fly Ash Stabilization and Fibre Reinforcement on Soil Strength. In: Cameselle C, Gouveia S, Urréjola S, Reddy KR, Agnihotri AK, editors. *Sustain. Energy Infrastruct. Keynote Vol. 6th EGRWSE 2025*, vol. 802, Cham: Springer; 2026, p. 457–67. https://doi.org/10.1007/978-3-032-15286-2_41.
- [25] Nasiri M, Amiri E. Stability Prediction of Excavations Reinforced with Different Geosynthetics—An Integrated Numerical and Machine Learning Approach. *Appl Comput Geosci* 2026:100351.
- [26] Ebadi-Jamkhaneh M, Ahmadi M. Comprehensive Investigations of the Effect of Bolt Tightness on Axial Behavior of a MERO Joint System: Experimental, FEM, and Soft Computing Approaches. *J Struct Eng* 2021;147. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0003168](https://doi.org/10.1061/(ASCE)ST.1943-541X.0003168).
- [27] Nasiri M, Amiri E, Ashrafi AH. Slope Stability Analysis Using Machine Learning Algorithms Based on Real Cases Datasets. *Transp Infrastruct Geotechnol* 2026;13:31. <https://doi.org/10.1007/s40515-026-00816-8>.
- [28] Fakharian P. Compressive Strength Prediction of 3D Printed Concrete with Machine Learning Approaches. *Comput Eng Phys Model* 2025;8:1385. <https://doi.org/10.22115/cepm.2026.1385>.
- [29] Nouri Y, Ghanbari MA, Fakharian P. Flexural behavior of hybrid GFRP-steel reinforced concrete beam: Experimental and explainable artificial intelligence. *Eng Struct* 2025;345:121565. <https://doi.org/10.1016/j.engstruct.2025.121565>.
- [30] Fakharian P, Nouri Y, Asaad Samani A, Rezazadeh Eidgahee D, Torabi MR, Hoseini Vaez SR. Explainable machine learning model and gene expression programming for predicting reinforced concrete beams moment capacity exposed to fire. *Sci Rep* 2026;16:2309. <https://doi.org/10.1038/s41598-025-32100-z>.
- [31] Nasiri M, Amiri E, Fadaei E. Ensemble Machine Learning Models for Estimating Stone Column Load-Bearing Capacity Using Combined Experimental and Field Test Data. *Transp Infrastruct Geotechnol* 2025;12:246. <https://doi.org/10.1007/s40515-025-00710-9>.
- [32] Nasiri M, Amiri E. ANN Insight in Forecasting Slope Stability by Analyzing the Influence of Strength Parameters in 2D and 3D Scenarios. *Transp Infrastruct Geotechnol* 2025;12:126. <https://doi.org/10.1007/s40515-025-00584-x>.
- [33] Chen W, Qu S, Lin L, Luo Q, Wang T. Ensemble Learning Methods for Shear Strength Prediction of Fly Ash-Amended Soils with Lignin Reinforcement. *J Mater Civ Eng* 2023;35:4023022. [https://doi.org/10.1061/\(ASCE\)MT.1943-5533.0004676](https://doi.org/10.1061/(ASCE)MT.1943-5533.0004676).
- [34] Wani NM, Thagunna P. Predictive modeling of shear strength in fly ash-stabilized clayey soils using artificial neural networks and support vector regression. *Asian J Civ Eng* 2024;25:6131–46. <https://doi.org/10.1007/s42107-024-01167-w>.
- [35] Aggarwal V, Abhi BS, Gupta S, Kaur R, Saini R, Sharma P, et al. Prediction of unconfined compressive strength of fly ash and waste paper sludge stabilized clayey soil using explainable machine learning. *Sci Rep* 2026;16:16376. <https://doi.org/10.1038/s41598-026-47096-3>.
- [36] Awoyera PO, Fadugba OG, Vasić M V. Optimized machine learning models for predictive performance assessment of fly ash-integrated clay bricks. *Sci Rep* 2026. <https://doi.org/10.1038/s41598-026-63535-7>.
- [37] Hu J, Li Z, Qin X, Dong S, Zhao T, Li H. Machine learning-based optimization of geopolymer-stabilized soil formulation using blast furnace slag and fly ash. *Mater Today Commun* 2025;48:113500. <https://doi.org/10.1016/j.mtcomm.2025.113500>.
- [38] Kumar A, Sinha S. Support vector machine-based prediction of unconfined compressive strength of Multi-Walled Carbon nanotube doped soil-fly ash mixes. *Multiscale Multidiscip Model Exp Des* 2024;7:5365–86. <https://doi.org/10.1007/s41939-024-00524-1>.