STM JOURNALS

Journal of
Polymer & Composites

ISSN: 2321-2810 (Online)
ISSN: 2321-8525 (Print)
Volume 14, Issue 3, 2026
DOI (Journal): 10.37591/JoPC

https://journals.stmjournals.com/jopc

Research JoPC

Structure–Property Modeling of Cement-Based Multi-Component Composites Using Ensemble Machine Learning and Explainable Feature Attribution

Aseem Kumar[1], Deepak Kumar Tiwari[2*]

Abstract

Accurate prediction of compressive strength is central to structure–property optimization, quality control, and sustainability-driven design in cement-based composite materials. Cementitious systems represent heterogeneous multi-phase composites composed of reactive binder matrices and dispersed aggregate phases, whose macroscopic mechanical performance emerges from complex nonlinear interactions among constituents and curing-dependent microstructural evolution. This study develops a data-driven structure–property modeling framework to quantify the nonlinear dependence of compressive strength on multi-component composite composition and age. An experimental dataset of 1030 mixtures incorporating variations in binder constituents, water content, chemical admixtures, aggregate fractions including polymers waste such as fly ash, plasticizers, and curing duration was used to train and evaluate four machine learning models: artificial neural network (multilayer perceptron), support vector regression (radial basis function kernel), random forest, and extra trees. Model performance was assessed across training, validation, and testing phases using complementary statistical metrics, robustness indicators, error diagnostics, and an engineering tolerance-based reliability check. Ensemble tree-based models demonstrated superior generalization compared with neural and kernel approaches, highlighting their effectiveness in modeling heterogeneous composite systems with complex phase interactions. Among the evaluated algorithms, the extra trees model achieved the most stable predictive accuracy on unseen data. Shapley additive explanations were employed to interpret phase contributions, revealing curing age and water–binder–matrix parameters as dominant drivers of strength development, followed by binder composition and aggregate proportions. The proposed framework provides a transparent and interpretable pathway for data-driven composite design, enabling rapid screening, performance verification, and structure–property-informed optimization of cementitious composite materials.

Keywords: Concrete compressive strength; Fly Ash: Cementitious composites; Machine learning; Ensemble models; Random Forest: Super Plasticizers

***Author for Correspondence**
Deepak Kumar Tiwari
Email ID: deepak.tiwari@gla.ac.in

[1]Research Scholar, Department of Civil engineering, GLA University, Mathura, Uttar Pradesh, India
[2]Assistant Professor, Department of Civil engineering, GLA University, Mathura, Uttar Pradesh, India

Received Date: March 10, 2026
Accepted Date: March 25, 2026
Published Date: May 08, 2026

Citation: Aseem Kumar, Deepak Kumar Tiwari. Structure–Property Modeling of Cement-Based Multi-Component Composites Using Ensemble Machine Learning and Explainable Feature Attribution. Journal of Polymer & Composites. 2026; 14(3): 112–131p.

INTRODUCTION

Concrete remains the most widely used engineered composite material due to its cost-effectiveness, durability, and structural versatility, supporting applications ranging from buildings and transportation systems to hydraulic and energy infrastructure [1]. From a materials science perspective, concrete is a heterogeneous multi-phase composite consisting of a reactive cementitious binder matrix, dispersed aggregate phases, mineral additions, and chemical admixtures. Its macroscopic mechanical performance emerges from complex

physicochemical interactions among these constituents and the evolution of microstructure during hydration and curing. Among its mechanical properties, compressive strength is the most influential parameter because it governs structural safety, serviceability, and long-term performance reliability [2,3]. Reliable estimation of compressive strength is therefore central to structure–property optimization, quality assurance, and performance-driven composite design for both conventional and advanced cement-based composite systems [4,5].

In practice, compressive strength is determined through standardized laboratory testing of specimens at prescribed curing ages (e.g., 7, 28, and 56 days) [6,7]. Although well established, such experimental procedures are time-intensive, resource-demanding, and costly when multiple mix compositions, material substitutions, or curing regimes are investigated [8,9]. The challenge is further intensified by the increasing incorporation of supplementary cementitious materials, chemical admixtures, recycled aggregates, and sustainable binders aimed at improving environmental performance. These additions increase compositional complexity and enhance the nonlinear dependence of strength on matrix chemistry, water–binder interactions, aggregate packing, and curing-driven microstructural evolution [10–12]. Consequently, compressive strength prediction represents a classical structure–property quantification problem within heterogeneous composite materials, where macroscopic performance arises from multi-scale phase interactions.

Several variables, such as cement quantity, water-to-binder ratio, aggregate properties, mineral admixtures such as fly ash, chemical admixtures, and curing age, interact in a highly nonlinear way to control the development of concrete's compressive strength [13–15]. Polymers also affects the durability, adhesion, elasticity[16–18] and strength of concrete when used in partial replacement with coarse and fine aggregate depending upon the size and shape of polymers [19–25]. Polymers also affect the durability, adhesion, and strength of concrete when used as a partial replacement for coarse or fine aggregates. The extent of this influence largely depends on the size, shape, surface texture, and distribution of the polymer particles within the concrete matrix. Polymers with finer particle sizes tend to fill the voids between aggregates, improving the density of the concrete and enhancing its durability by reducing permeability. Conversely, larger or irregularly shaped polymer particles may create weak interfacial zones between the cement paste and aggregates, which can affect the mechanical strength. Additionally, the surface characteristics of polymers influence the bonding behaviour with the cement matrix, thereby impacting the overall adhesion and structural performance of the concrete. Therefore, the proper selection of polymer particle size and morphology plays a crucial role in determining the effectiveness of polymer-modified concrete. Traditional empirical equations and regression-based models, although useful for preliminary estimation, often struggle to capture these nonlinear relationships, particularly when applied to heterogeneous datasets or novel material combinations [26,27]. As a result, such models may exhibit limited predictive accuracy and poor generalization performance beyond the conditions for which they were originally developed. In recent years, machine learning (ML) techniques have emerged as powerful data-driven tools capable of modeling complex nonlinear systems without requiring explicit functional relationships [28,29]. In the context of concrete materials, ML models such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), Decision Trees, Random Forests (RF), and ensemble-based approaches have been increasingly applied for predicting compressive strength [30,31]. These models have demonstrated promising results, often outperforming conventional regression techniques by effectively learning intricate interactions among input variables. The growing availability of experimental databases and computational resources has further accelerated the adoption of ML-based approaches in concrete research.

Several studies have reported successful application of ML techniques for predicting concrete compressive strength using mix design parameters as inputs [32,33]. Neural network–based models have shown strong capability in capturing nonlinear trends [34], while kernel-based methods such as SVR have demonstrated robustness for moderate-sized datasets [35]. Tree-based ensemble models, particularly Random Forest and Extra Trees, have gained attention due to their high predictive accuracy,

resistance to overfitting, and ability to handle multicollinearity among input variables. Despite these encouraging developments, the existing body of literature reveals several important limitations that restrict the broader adoption of ML models in practical concrete engineering applications.

These gaps raise a central question relevant to composite materials science: which machine learning approaches can provide accurate, robust, and interpretable structure–property predictions for heterogeneous cement-based composites? Addressing this question is essential for advancing data-driven modeling from exploratory analysis toward reliable decision-support tools for composite material optimization [36,37]. Motivated by these challenges, the present study develops a systematic machine learning framework for predicting compressive strength as a structure–property outcome of multi-component cementitious composites [38]. Multiple widely used algorithms Artificial Neural Networks, Support Vector Regression, Random Forest, and Extra Trees are evaluated within a unified experimental framework. A rigorous three-way data partitioning strategy (training, validation, testing) is adopted to ensure unbiased generalization assessment.

A distinguishing feature of this study is the adoption of a multi-criteria performance evaluation strategy that extends beyond conventional metrics. In addition to MAE, RMSE, and R^2, the analysis incorporates normalized error indicators, bias metrics, robustness measures, and engineering tolerance–based reliability evaluation. Graphical diagnostics, including predicted-versus-observed comparisons and error distribution analyses, further support comprehensive assessment of model behavior across data partitions. To enhance interpretability and composite-level insight, the study integrates SHAP-based feature attribution for tree-ensemble models. This approach enables quantitative interpretation of phase contributions and structure–property relationships, revealing how matrix parameters, water–binder interactions, and aggregate fractions govern strength development. Such explainable modeling bridges data-driven predictions with established composite material science principles.

The objectives of this study are therefore:

1. to develop multiple machine learning models for predicting compressive strength of multi-component cementitious composites.
2. to rigorously compare model performance across training, validation, and testing datasets using diverse and engineering-relevant metrics.
3. to analyze predictive behavior through comprehensive diagnostic visualization; and
4. to interpret model outputs using SHAP-based attribution to identify dominant phase-level contributors to compressive strength.

By simultaneously addressing predictive accuracy, robustness, and interpretability within a composite-materials framework, this study provides a benchmark reference for data-driven structure–property modeling of heterogeneous cement-based composites and supports informed model selection for advanced composite design and optimization.

DATASET, INPUT VARIABLES, AND PREPROCESSING PROCEDURES

Cement-Based Composite Strength Dataset

The Concrete Compressive Strength dataset, which was acquired from the publicly accessible UCI Machine Learning Repository, was used in this investigation. Yeh was the first to compile this dataset, which is now frequently used as a benchmark dataset for modeling concrete mechanical properties using machine learning. The dataset shows the laboratory-measured compressive strength values of concrete mixtures made with varying constituent material proportions and tested at varied curing ages. It is well known that the material composition and curing age of concrete have a highly nonlinear effect on its compressive strength. The dataset includes 1030 observations, each corresponding to a unique concrete mixture tested under controlled laboratory conditions. All data are provided in raw numerical form, without prior scaling or normalization, ensuring suitability for flexible preprocessing strategies. The experimental compressive strength values were measured in accordance with standard laboratory

procedures and are expressed in megapascals (MPa). No missing values are present in the dataset, making it well-suited for machine learning analysis without the need for imputation.

Input and Output Variables

Eight quantitative input variables, which reflect the components of the concrete mix and the curing age, and one quantitative output variable, which represents compressive strength, make up the dataset. The input variables collectively describe the material composition of a concrete mixture per cubic meter, while the output variable reflects the resulting compressive strength.

The input variables are as follows:

- Cement content (kg/m3): This indicates how much Portland cement is used as the main binding agent in the concrete mixture.
- Blast furnace slag (kg/m³): An additional cementitious substance that enhances durability and strength over time.
- Fly ash (kg/m³): This pozzolanic substance affects strength development and workability.
- Water content (kg/m³): A critical parameter governing hydration and strength development.
- Superplasticizer (kg/m³): A chemical admixture used to improve workability without increasing water content.
- Coarse aggregate (kg/m³): Represents the larger aggregate fraction contributing to mechanical interlocking and load transfer.
- Fine aggregate (kg/m³): Represents the smaller aggregate fraction influencing packing density and workability.
- Age (days): The curing age of the concrete specimen, ranging from 1 to 365 days, reflecting strength gain over time.

The concrete specimens experimentally determined compressive strength (MPa) is the output variable. All input variables are continuous and quantitative, making the dataset suitable for regression-based machine learning models. The presence of curing age as an input variable enables modeling of both early-age and long-term strength development.

Data Characteristics and Statistical Properties

The dataset contains a total of 1030 samples and 9 attributes, including the output variable. The wide range of material proportions and curing ages introduces significant variability and nonlinearity into the data. Such characteristics pose challenges for traditional linear regression models but are well suited for advanced machine learning approaches. Since the dataset includes no missing values and no categorical attributes, the preprocessing pipeline focuses primarily on feature scaling and data partitioning rather than data cleaning or transformation. The absence of missing values ensures that all observations can be fully utilized during model training and evaluation.

Data Preprocessing

To guarantee consistency and robustness of the machine learning analysis, a number of preprocessing processes were implemented prior to model building. In order to eliminate superfluous whitespace and inconsistent formatting, column names were first standardized. During model training and evaluation, this phase guarantees dependable feature referencing. Second, a 70:15:15 split ratio was used to separate the dataset into subgroups for testing, validation, and training. The testing set was utilized for the ultimate, objective performance evaluation, the validation set for model tweaking and performance tracking, and the training set for model learning. In addition to reducing overfitting, this three-way data partitioning technique guarantees accurate assessment of the model's capacity for generalization. Third, all input characteristics were transformed into a common range between 0 and 1 by applying feature scaling to the input variables using min–max normalization. Because it keeps variables with wider numerical ranges from controlling the learning process, feature scaling is especially crucial for distance-based and gradient-based models like Support Vector Regression and Artificial Neural Networks. To

ensure consistency across all models, the scaled dataset was also used to train tree-based models. In order to maintain interpretability and engineering relevance, compressive strength values were kept in their original physical units (MPa), hence no modification was made to the output variable.

Rationale for Dataset Selection

The Concrete Compressive Strength dataset was selected for this study due to its engineering relevance, experimental reliability, and widespread adoption in the literature. Its well-documented structure and absence of missing values make it an ideal benchmark for evaluating and comparing machine learning models. Furthermore, the dataset captures realistic nonlinear relationships between material composition, curing age, and compressive strength, allowing meaningful assessment of model accuracy, robustness, and interpretability. By employing this dataset within a rigorous preprocessing and evaluation framework, the present study ensures that the reported results are both reproducible and directly comparable with previous research, while enabling deeper insight into model performance and reliability.

METHODOLOGY

Overall Modeling Framework

Data preprocessing, model training, validation, testing, and performance evaluation make up the entire modeling framework. A 70:15:15 ratio was used to separate the preprocessed dataset into training, validation, and testing groups. The training dataset was used to train all models, while the validation dataset was used to track performance and evaluate parameter stability. To guarantee an objective evaluation of generalization, the final model's performance was only assessed on the testing dataset. The input vector for each model is defined as:

$$\mathrm{X} = [x_1, x_2, x_3, \ldots, x_8] \tag{1}$$

where x_1–x_8 represent cement content, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, fine aggregate, and curing age, respectively. The target variable is the concrete compressive strength y, expressed in MPa.

Artificial Neural Network- Multi-Layer Perceptron (ANN-MLP)

Artificial Neural Networks are biologically inspired computational models capable of approximating complex nonlinear relationships. In this study, a Multi-Layer Perceptron (MLP) architecture was adopted due to its widespread application in concrete strength prediction.

An ANN-MLP consists of an input layer, one or more hidden layers, and an output layer. Each neuron performs a weighted summation of inputs followed by a nonlinear activation function. The output of a neuron is given by:

$$z_j = f\left(\sum_{i=1}^{n} w_{ij}\, x_i + b_j\right) \tag{2}$$

where w_{ij} is the weight connecting the i-th input to the j-th neuron, b_j is the bias term, and $f(\cdot)$ is the activation function. In this study, the Rectified Linear Unit (ReLU) activation function was employed in hidden layers:

$$f(z) = \max(0, z) \tag{3}$$

The output layer uses a linear activation function to predict continuous compressive strength values. Model training was performed using the Adam optimization algorithm, which minimizes the mean squared error loss function through backpropagation.

ANN models can capture highly nonlinear material behavior; however, their performance depends on appropriate network architecture and training stability. Therefore, a moderate network structure was adopted to balance model complexity and generalization.

Support Vector Regression- Radial Basis Function (SVR-RBF)

Support Vector Regression is a kernel-based learning method derived from statistical learning theory. SVR aims to identify a regression function that deviates from observed values by no more than a predefined margin while maintaining maximum flatness.

The SVR regression function is expressed as:

$$f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{4}$$

where x_i are support vectors, $\alpha_i$and α_i^* are Lagrange multipliers, b is the bias term, and $K(\cdot)$ is the kernel function.

In this study, the Radial Basis Function (RBF) kernel was used:

$$K(x_i, x_j) = \exp\left(-\gamma \parallel x_i - x_j \parallel^2 \right) \tag{5}$$

where γ is the kernel width parameter controlling the influence of individual data points. The SVR model also includes a regularization parameter C, which governs the trade-off between model flatness and training error.

SVR-RBF is particularly effective for modeling nonlinear relationships in experimental datasets and offers strong generalization capability, especially when data size is moderate and noise is present.

Random Forest (RF)

An ensemble learning technique called Random Forest is built on the combination of several decision trees. A bootstrap sample of the training data is used to build each tree, and a random subset of the input variables is taken into consideration for splitting at each node. The average of each tree's forecasts yields the final prediction:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} h_t(x) \tag{6}$$

where $h_t(x)$ is the prediction of the t-th decision tree and T is the total number of trees.

The randomization introduced through bootstrap sampling and feature selection enhances model robustness and reduces overfitting. Random Forest models are well suited for tabular engineering datasets, as they can effectively handle nonlinear relationships, multicollinearity among input variables, and experimental uncertainty.

An additional advantage of Random Forest is its ability to quantify feature importance, providing insight into the relative influence of input variables on compressive strength prediction.

Extra Trees (Extremely Randomized Trees)

Extra Trees is an extension of the Random Forest algorithm that introduces additional randomness during tree construction. Unlike Random Forest, which selects optimal split points for each feature, Extra Trees chooses split thresholds randomly. The entire training dataset is typically used for tree construction rather than bootstrap samples.

The prediction mechanism of Extra Trees is like Random Forest and can be expressed as:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} g_t(x) \tag{7}$$

where $g_t(x)$ represents the prediction from the t-th extremely randomized tree.

The increased randomness in Extra Trees can reduce model variance and improve computational efficiency. However, excessive randomness may affect predictive stability, making rigorous evaluation across training, validation, and testing datasets essential.

Performance Evaluation Strategy

To comprehensively assess the predictive capability, robustness, and engineering relevance of the developed machine learning models, a multi-criteria performance evaluation strategy was adopted. Model evaluation was conducted separately for the training, validation, and testing datasets, ensuring unbiased assessment of generalization performance and minimizing the risk of overfitting. The testing dataset was used as the primary basis for model comparison and final model selection. The evaluation metrics were selected to capture different aspects of prediction performance, including error magnitude, goodness-of-fit, relative accuracy, normalization across scales, prediction bias, robustness to outliers, correlation strength, and engineering tolerance. All metrics were computed using the predicted compressive strength values $\hat{y}_i$ and the corresponding observed values y_i, where n denotes the total number of samples.

Error Magnitude Metrics

Mean Absolute Error (MAE)

The Mean Absolute Error quantifies the average absolute deviation between predicted and observed values and is defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{8}$$

MAE provides a direct and interpretable measure of average prediction error in physical units (MPa), making it particularly suitable for engineering applications.

Root Mean Squared Error (RMSE)

The Root Mean Squared Error is expressed as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

RMSE penalizes larger prediction errors more strongly than MAE and is therefore sensitive to extreme deviations. This metric is especially important in structural engineering contexts, where large prediction errors may compromise safety.

Coefficient of Determination (R^2)

The coefficient of determination measures the proportion of variance in the observed data explained by the model and is given by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{10}$$

where $\bar{y}$ is the mean of observed compressive strength values. Higher R^2 values indicate stronger agreement between predicted and observed data.

Relative Error Metric: Mean Absolute Percentage Error (MAPE)

MAPE evaluates prediction error relative to observed values and is defined as:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n} | \frac{y_i - \hat{y}_i}{y_i} | \tag{11}$$

This metric expresses prediction error as a percentage, facilitating comparison across different strength ranges. A small numerical constant was introduced in the denominator during computation to avoid division by zero.

Normalized Error Metric: Normalized Root Mean Squared Error (NRMSE)

To enable scale-independent comparison, RMSE was normalized as:

$$\text{NRMSE} = \frac{\text{RMSE}}{y_{\max} - y_{\min}} \tag{12}$$

where $y_{\max}$ and $y_{\min}$ denote the maximum and minimum observed compressive strength values, respectively. NRMSE allows comparison of prediction accuracy across datasets with different value ranges.

Bias Metric: Mean Bias Error (MBE)

The Mean Bias Error quantifies systematic over- or under-prediction and is computed as:

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{13}$$

A positive MBE indicates overall under-prediction, while a negative value indicates over-prediction. This metric is particularly useful for assessing directional bias in engineering predictions.

Robustness Metric: Median Absolute Error (MedAE)

To assess robustness against outliers, the Median Absolute Error was employed:

$$\text{MedAE} = \text{median}(|\, y_i - \hat{y}_i \,|) \tag{14}$$

Unlike MAE and RMSE, MedAE is less sensitive to extreme errors and provides a robust measure of typical prediction performance.

Correlation Metric: Pearson Correlation Coefficient (r)

The linear correlation between predicted and observed values was quantified using the Pearson correlation coefficient:

$$r = \frac{\text{Cov}(y,\hat{y})}{\sigma_y \sigma_{\hat{y}}} \tag{15}$$

where σ_y and $\sigma_{\hat{y}}$ are the standard deviations of observed and predicted values, respectively. Higher values of r indicate stronger linear agreement.

Engineering Tolerance-Based Accuracy: Prediction Accuracy within ±5 MPa

In addition to statistical metrics, an engineering-oriented tolerance-based accuracy was used to assess practical usability. The percentage of predictions falling within ±5 MPa of observed compressive strength was calculated as:

$$Accuracy_{\pm 5} = \frac{1}{n}\sum_{i=1}^{n} I(|\, y_i - \hat{y}_i \,| \leq 5) \times 100 \tag{16}$$

where $I(\cdot)$ is an indicator function that equals 1 if the condition is satisfied and 0 otherwise. This metric directly reflects engineering acceptability and aligns with practical strength assessment requirements.

RESULTS AND DISCUSSION

Phase-Level Feature Attribution and Structure–Property Interpretation

To interpret the predictive behavior of the developed machine learning framework and to elucidate the governing structure–property relationships in the cement-based composite system, SHAP (Shapley Additive Explanations) analysis was performed on the tree-ensemble model demonstrating strong generalization performance. Within heterogeneous composite materials, macroscopic mechanical performance emerges from nonlinear interactions among matrix chemistry, dispersed phases, and curing-driven microstructural evolution. SHAP provides a quantitative approach to attribute model predictions to individual compositional parameters, enabling both global phase-level importance assessment and local contribution analysis for specific mixtures. Figure 1 presents the mean absolute SHAP values, representing the average contribution magnitude of each constituent to compressive strength. The results indicate that cement content is the dominant phase-level parameter, followed by

curing age and water content. This ranking reflects the matrix-controlled nature of compressive strength development in cementitious composites, where mechanical resistance is primarily governed by hydration-derived binder phases rather than aggregate fractions.

Cement content exhibits the highest mean SHAP value, confirming its dominant contribution to the composite's load-bearing matrix. Increased cement content promotes greater formation of calcium silicate hydrate (C–S–H) gel, which constitutes the primary structural binding phase in the composite microstructure. The SHAP results demonstrate that variations in matrix volume fraction significantly influence predicted strength, reinforcing the interpretation that compressive resistance is largely matrix-dominated within the studied compositional range. Curing age is identified as the second most influential parameter. As illustrated in Figure 2, higher curing durations correspond to positive SHAP contributions, indicating a monotonic increase in predicted compressive strength with time. This behavior reflects hydration-driven microstructural densification and progressive refinement of pore structure, characteristic of evolving composite matrices. The observed monotonic trend suggests that the model effectively captures temporal structure–property evolution within the heterogeneous composite system.

Water content exhibits predominantly negative SHAP contributions at higher values. From a composite mechanics perspective, increased water content elevates the effective water–binder ratio, leading to higher porosity and reduced matrix density after hardening. The SHAP analysis confirms that excess water weakens the matrix phase, thereby diminishing overall composite strength. Importantly, the interaction between cement content and water fraction reflects the critical balance between matrix formation and porosity control in composite optimization. Among supplementary cementitious materials, blast furnace slag demonstrates a moderate but positive influence, particularly at extended curing ages. This behavior is consistent with its latent hydraulic reactivity and its contribution to long-term matrix densification and interfacial transition zone improvement. In contrast, fly ash exhibits comparatively lower SHAP magnitude, suggesting a more gradual pozzolanic contribution within the dataset. This reduced influence may be attributed to slower reaction kinetics and partial replacement levels in the studied formulations.

The influence of superplasticizer is moderate and predominantly positive. By improving particle dispersion and enabling lower effective water–binder ratios, superplasticizer enhances matrix compactness without increasing porosity, indirectly strengthening the composite structure. Aggregate-related parameters (fine and coarse aggregates) show comparatively smaller SHAP magnitudes. This indicates that within the analyzed compositional domain, compressive strength is primarily governed by matrix-phase characteristics rather than aggregate volume fraction. While aggregates provide load transfer pathways, geometric confinement, and dimensional stability, their contribution to compressive resistance is secondary to binder microstructure formation. This observation reinforces the matrix-dominated strength behavior typical of cement-based composite systems.

Model Performance Comparison across Training, Validation, and Testing Phases

The predictive performance of the four machine learning models ANN-MLP, SVR-RBF, Random Forest, and Extra Trees was evaluated using a comprehensive set of statistical and engineering-oriented metrics across the training, validation, and testing datasets. The results provide insight into model accuracy, robustness, generalization capability, and potential overfitting behavior.

Training Performance

During the training phase as shown in Table 1, the Extra Trees model exhibited the highest predictive accuracy, achieving the lowest error values (MAE = 0.09 MPa, RMSE = 0.99 MPa) and the highest goodness-of-fit ($R^2 = 0.997$). The very high Pearson correlation coefficient ($r = 0.998$) and tolerance-based accuracy within ±5 MPa exceeding 99% indicate an excellent fit to the training data. Similarly, the Random Forest model demonstrated strong training performance, with RMSE of 2.02 MPa and R^2

of 0.986, suggesting its ability to effectively capture nonlinear relationships in the dataset. In contrast, SVR-RBF and ANN-MLP exhibited comparatively higher training errors, with RMSE values of 5.76 MPa and 6.15 MPa, respectively. Although these models achieved reasonable R^2 values (>0.86), their performance was notably inferior to tree-based ensemble methods. This difference highlights the strong learning capacity of ensemble models for tabular concrete mix data. However, the extremely low training errors observed for the Extra Trees model suggest a strong fit that warrants careful examination of its generalization performance.

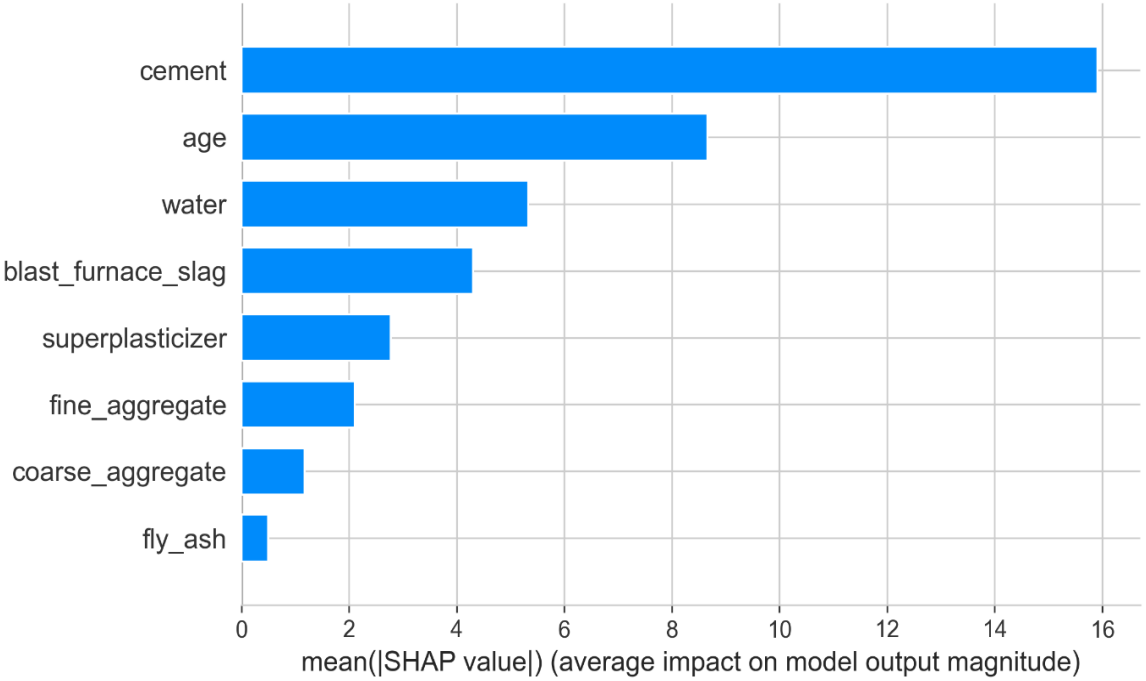

Figure 1. Global feature importance based on SHAP analysis

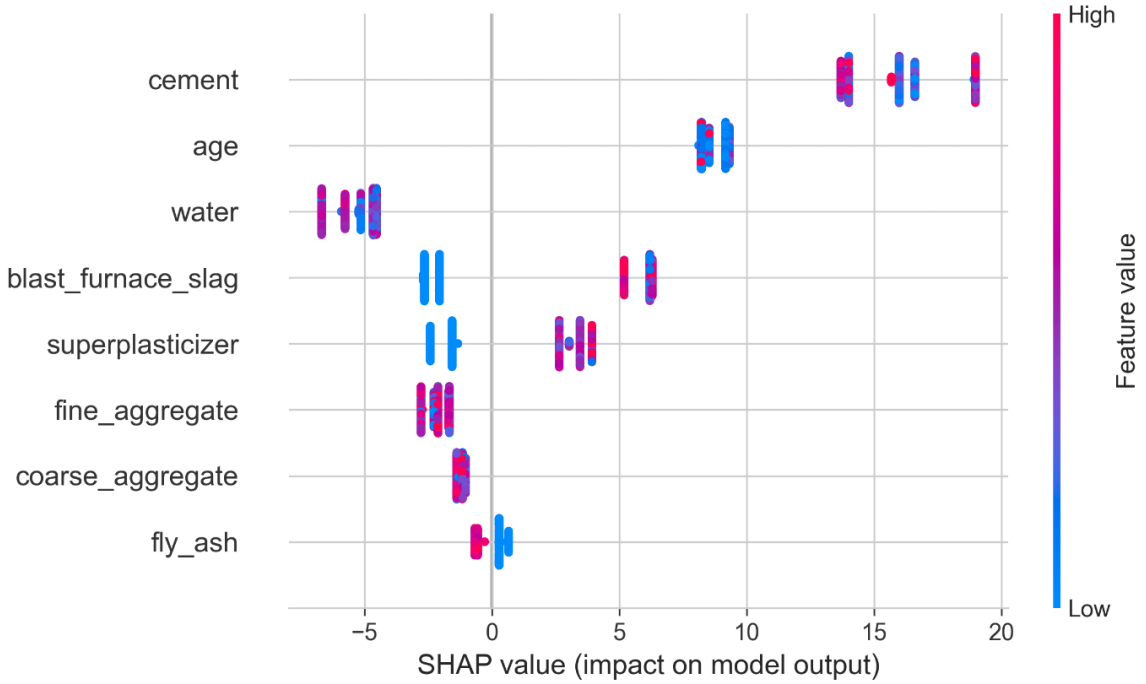

Figure 2. SHAP summary plot showing feature impact and direction

Table 1. Performance of machine learning models during the training phase

Model	MAE (MPa)	RMSE (MPa)	R^2	MAPE (%)	NRMSE	MBE (MPa)	MedAE (MPa)	Pearson r	Acc (±5 MPa) (%)
ANN-MLP	4.62	6.15	0.866	15.45	0.0766	0.12	3.4	0.931	63.25
SVR-RBF	3.71	5.76	0.882	13.74	0.0718	0.21	2.23	0.94	73.09
Random Forest	1.34	2.02	0.986	4.7	0.0251	0.05	0.96	0.993	97.36
Extra Trees	0.09	0.99	0.997	0.26	0.0124	≈0.00	≈0.00	0.998	99.31

Validation Performance

Validation results provide a more realistic assessment of model robustness as shown in Table 2. In this phase, the Extra Trees model continued to outperform other models, yielding the lowest RMSE (4.95 MPa), highest R^2 (0.915), and the highest tolerance-based accuracy (83.1%). The reduction in performance compared to training indicates expected regularization through validation but still reflects strong generalization. The Random Forest model also demonstrated stable validation performance, with RMSE of 5.55 MPa and R^2 of 0.893. Compared to the training phase, the moderate increase in error suggests limited overfitting and good model robustness. The Pearson correlation coefficient ($r \approx 0.95$) further confirms strong agreement between predicted and observed values. The ANN-MLP model showed moderate validation performance, with RMSE of 6.65 MPa and R^2 of 0.847. Although its performance remained consistent with training results, it was less competitive than ensemble models. The SVR-RBF model exhibited the largest degradation in validation performance, with RMSE increasing to 7.21 MPa and R^2 decreasing to 0.820, indicating sensitivity to data partitioning and limited robustness.

Table 2. Performance of machine learning models during the validation phase

Model	MAE (MPa)	RMSE (MPa)	R^2	MAPE (%)	NRMSE	MBE (MPa)	MedAE (MPa)	Pearson r	Acc (±5 MPa) (%)
ANN-MLP	4.84	6.65	0.847	15.42	0.0862	0.8	3.52	0.923	62.99
SVR-RBF	5.22	7.21	0.82	19.26	0.0934	0.01	4.05	0.907	60.39
Random Forest	3.55	5.55	0.893	12.4	0.0719	0.86	2.4	0.95	76.62
Extra Trees	3	4.95	0.915	9.99	0.0641	0.8	1.91	0.959	83.12

Testing Performance and Generalization

Testing performance is the most critical indicator of model reliability as shown in Table 3. On the unseen testing dataset, Random Forest achieved the best overall balance between accuracy and generalization, recording the lowest RMSE (5.41 MPa), high R^2 (0.883), strong Pearson correlation ($r = 0.943$), and tolerance-based accuracy exceeding 70%. These results indicate that Random Forest provides reliable and stable predictions across all data partitions. The Extra Trees model, while still competitive, showed a slight reduction in testing performance compared to Random Forest. Its RMSE increased to 5.51 MPa and R^2 decreased to 0.879, suggesting mild overfitting during training. Nevertheless, its testing performance remained superior to ANN-MLP and SVR-RBF in most metrics, confirming its strong predictive capability. The ANN-MLP model demonstrated moderate generalization, with RMSE of 6.78 MPa and R^2 of 0.816. Although ANN maintained relatively consistent performance across phases, its higher prediction errors and lower tolerance-based accuracy (56.1%) limit its practical applicability for precise strength estimation. The SVR-RBF model exhibited the weakest testing performance among the evaluated models, with the highest RMSE (7.49 MPa) and the lowest R^2 (0.776). Despite reasonable training accuracy, the noticeable decline in testing performance suggests reduced generalization capability, particularly for higher-strength concrete mixtures.

Comparison of Observed and Predicted Compressive Strength

The agreement between observed and predicted concrete compressive strength values was further examined through time-series style plots and prediction error analysis for the training, validation, and

testing datasets. These graphical diagnostics complement the quantitative metrics discussed in Section 4.2 by providing insight into model consistency, peak capturing ability, and error dispersion. Figure 4 illustrates the prediction errors (predicted − observed) for all four models during the training phase. The Random Forest and Extra Trees models exhibit error distributions that are tightly clustered around zero, with comparatively low amplitude and limited dispersion. This behavior indicates strong learning of underlying patterns in the training data and minimal systematic bias. In contrast, ANN-MLP and SVR-RBF display larger error fluctuations, with several instances of both over- and under-prediction exceeding ±15 MPa. These larger deviations suggest reduced learning efficiency for extreme strength values, particularly in mixtures with higher variability. While all models exhibit near-zero mean error, the narrower error band observed for tree-based ensemble models confirms their superior fitting capability during training.

Table 3. Performance of machine learning models during the testing phase

Model	MAE (MPa)	RMSE (MPa)	R^2	MAPE (%)	NRMSE	MBE (MPa)	MedAE (MPa)	Pearson r	Acc (±5 MPa) (%)
ANN-MLP	5.4	6.78	0.816	16.53	0.0969	1.44	4.37	0.909	56.13
SVR-RBF	5.42	7.49	0.776	17.31	0.1071	0.37	3.6	0.883	61.29
Random Forest	3.89	5.41	0.883	13.07	0.0774	0.9	2.82	0.943	70.32
Extra Trees	3.82	5.51	0.879	12.74	0.0788	0.86	2.42	0.939	66.45

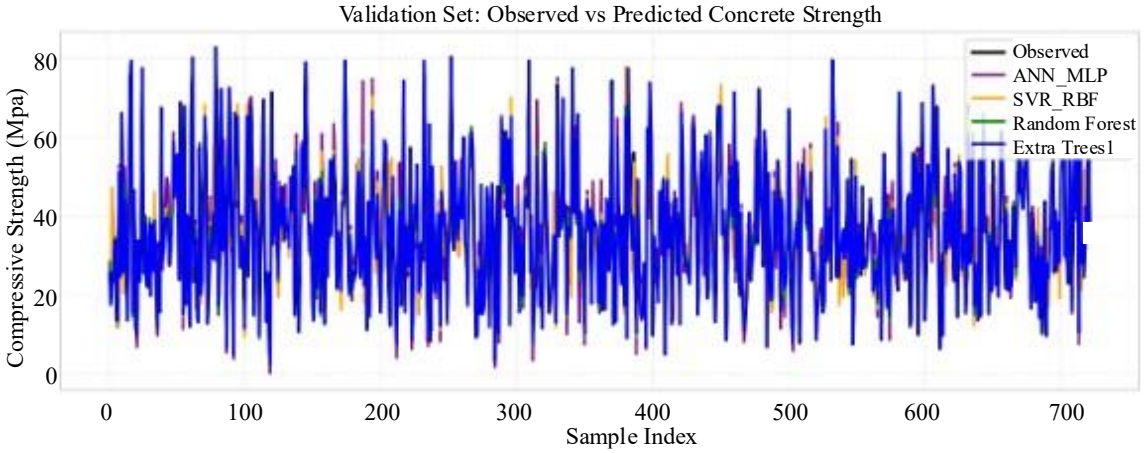

Figure 4. Training-phase prediction errors of concrete compressive strength.

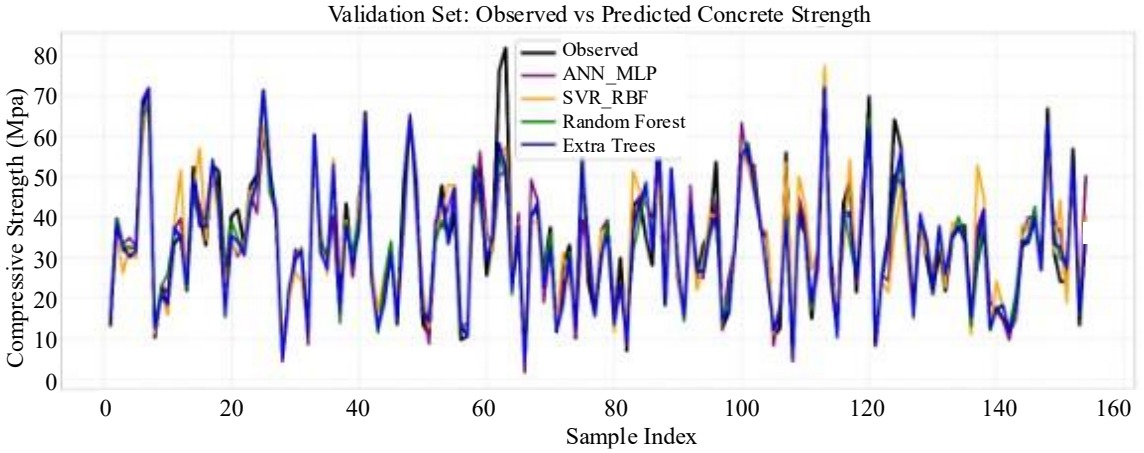

Figure 5. Validation-phase observed and predicted concrete compressive strength.

Figure 5 presents the comparison between observed and predicted compressive strength values for the validation dataset. All models can capture the general trend and variability of concrete strength; however, clear differences in predictive stability are observed. The Random Forest and Extra Trees models demonstrate closer tracking of observed strength values across the entire strength range, including both low- and high-strength samples. Their predicted curves show limited lag and reduced smoothing, indicating effective generalization beyond the training data. The ANN-MLP model follows the overall trend reasonably well but shows noticeable deviations for peak strength values. The SVR-RBF model exhibits occasional over-prediction at higher strength levels, reflecting its sensitivity to nonlinear extremes.

The predictive behavior on unseen data is illustrated in Figure 6, which compares observed and predicted compressive strength values for the testing dataset. The testing results provide the most reliable assessment of model robustness and real-world applicability. The Random Forest model exhibits the closest agreement with observed values across the full-strength spectrum, consistently capturing both sharp peaks and low-strength regions. The Extra Trees model also performs well but shows slightly increased deviation for a small number of high-strength samples, indicating mild overfitting observed earlier in training. The ANN-MLP model demonstrates moderate predictive accuracy but tends to underestimate higher compressive strength values. The SVR-RBF model shows greater scatter around the observed values, particularly at extreme strength levels, consistent with its comparatively lower testing performance metrics. These observations align closely with the statistical results presented in Table 3, reinforcing the conclusion that Random Forest provides the most stable and reliable generalization performance among the evaluated models.

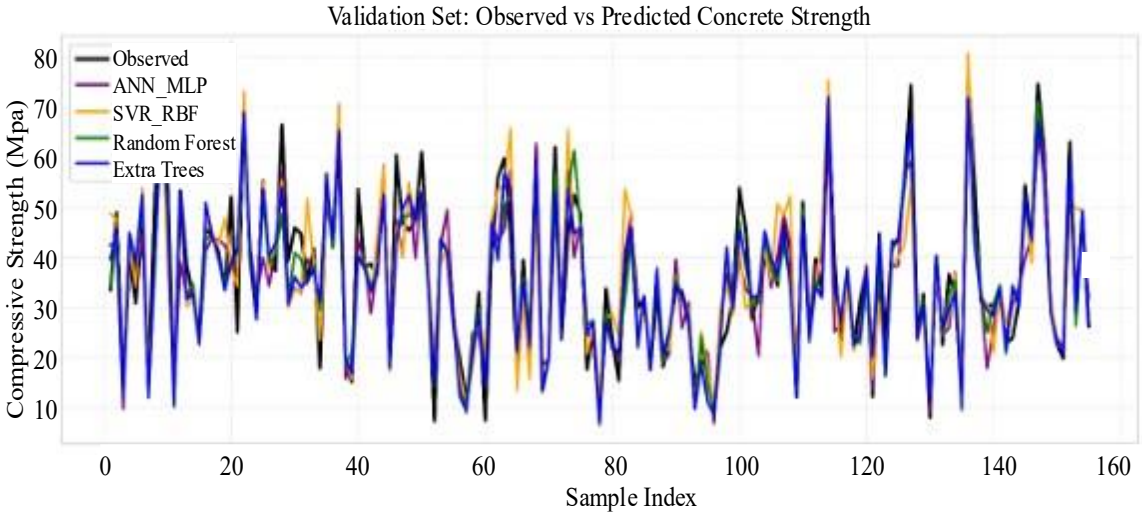

Figure 6. Testing-phase observed and predicted concrete compressive strength.

Error Analysis of Machine Learning Models

The prediction error distributions for the training, validation, and testing phases are illustrated in Figures 7–9, respectively. These error plots provide additional insight into model robustness, bias, and generalization capability by examining the deviation between predicted and observed concrete compressive strength values across different datasets. During the training phase (Figure 7), all models exhibit errors predominantly clustered around zero, indicating effective learning of the underlying nonlinear relationship between concrete constituents and compressive strength. Among the evaluated models, the Extra Trees and Random Forest models show the narrowest error bands with minimal dispersion, reflecting their strong ability to capture complex interactions within the training data. In contrast, the ANN-MLP and SVR-RBF models display comparatively wider fluctuations, suggesting higher sensitivity to local variations in the dataset. Nevertheless, no systematic bias is observed, as positive and negative errors are well balanced around the zero-error reference line.

The validation error patterns presented in Figure 8 demonstrate a moderate increase in error variability for all models, which is expected when models are exposed to unseen data. The SVR-RBF model exhibits occasional large deviations, particularly at higher strength values, indicating sensitivity to boundary cases. The ANN-MLP model shows intermittent oscillations, reflecting its dependence on network initialization and nonlinear activation behaviour. In contrast, the ensemble-based models (Random Forest and Extra Trees) maintain relatively stable error distributions, confirming their superior generalization capability and resistance to overfitting. In the testing phase (Figure 9), error magnitudes further increase slightly, providing a realistic assessment of model performance under completely unseen conditions. Despite this increase, most prediction errors remain within an acceptable engineering tolerance range. The Extra Trees model continues to demonstrate the most consistent performance, with errors closely centered around zero and fewer extreme deviations. The Random Forest model follows a similar trend, whereas ANN-MLP and SVR-RBF show larger sporadic errors, particularly for high-strength concrete mixtures. These observations are consistent with the quantitative performance metrics discussed earlier, reinforcing the reliability of ensemble tree-based approaches for concrete strength prediction.

Classification Performance of Composite Strength Regimes

To further assess the engineering relevance of the developed machine learning framework, continuous compressive strength predictions were categorized into five discrete strength regimes representing distinct performance levels of the cement-based composite system. From a materials perspective, such classification corresponds to identifying composite performance states governed by variations in matrix density, water–binder interactions, supplementary binder reactivity, and aggregate packing effects. The resulting testing-phase classification performance is illustrated using confusion matrices for ANN-MLP, SVR-RBF, Random Forest, and Extra Trees models (Figures 10–13). The confusion matrix of the Random Forest model (Figure 10) demonstrates strong classification capability across all strength regimes, with a clear dominance along the principal diagonal. Both low-strength and high-strength composite classes (Class 1 and Class 5) are identified with high accuracy, indicating the model's ability to differentiate matrix-dominated porous systems from dense, highly hydrated microstructures. Misclassifications are primarily limited to adjacent strength classes, which is expected due to gradual compositional transitions and overlapping microstructural characteristics within intermediate composite formulations. Such near-boundary confusion reflects the continuous nature of structure–property evolution rather than abrupt phase changes.

The SVR-RBF model (Figure 11) exhibits comparatively higher confusion between intermediate strength regimes (Classes 3 and 4). While extreme performance classes are reasonably predicted, the dispersion of samples across neighboring categories indicates reduced robustness in distinguishing subtle variations in composite microstructure and phase interaction intensity. This behavior is consistent with earlier regression findings showing greater sensitivity of SVR-RBF to localized nonlinear fluctuations. The ANN-MLP model (Figure 12) demonstrates moderate classification performance. Low-strength composites are effectively identified; however, increased misclassification is observed among mid- to high-strength regimes. This suggests partial smoothing of nonlinear structure–property boundaries, potentially arising from network parameter sensitivity and generalized approximation of compositional effects. Nevertheless, the absence of strong class bias indicates stable overall learning behavior. The Extra Trees model (Figure 13) exhibits the most consistent and reliable classification performance. Pronounced diagonal dominance is observed, particularly for Classes 1, 2, and 5, indicating accurate identification of distinct composite performance states across a broad strength spectrum. The reduced off-diagonal entries confirm enhanced robustness and improved discrimination of nonlinear compositional regimes. This superior performance can be attributed to the ensemble's ability to capture complex interactions among matrix parameters, curing age, and dispersed phases while mitigating variance through randomized splitting and averaging.

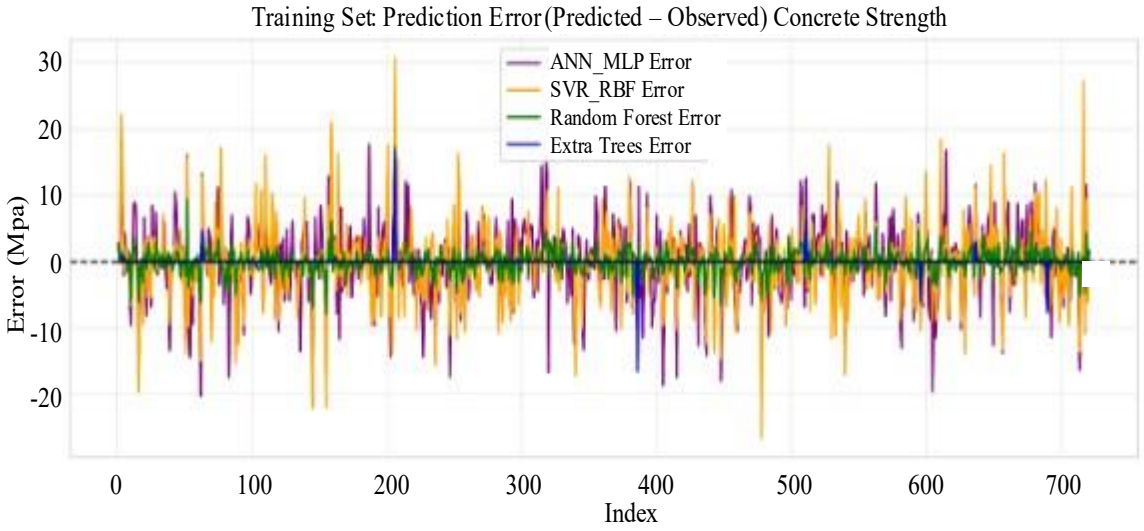

Figure 7. Training-phase prediction errors of concrete compressive strength.

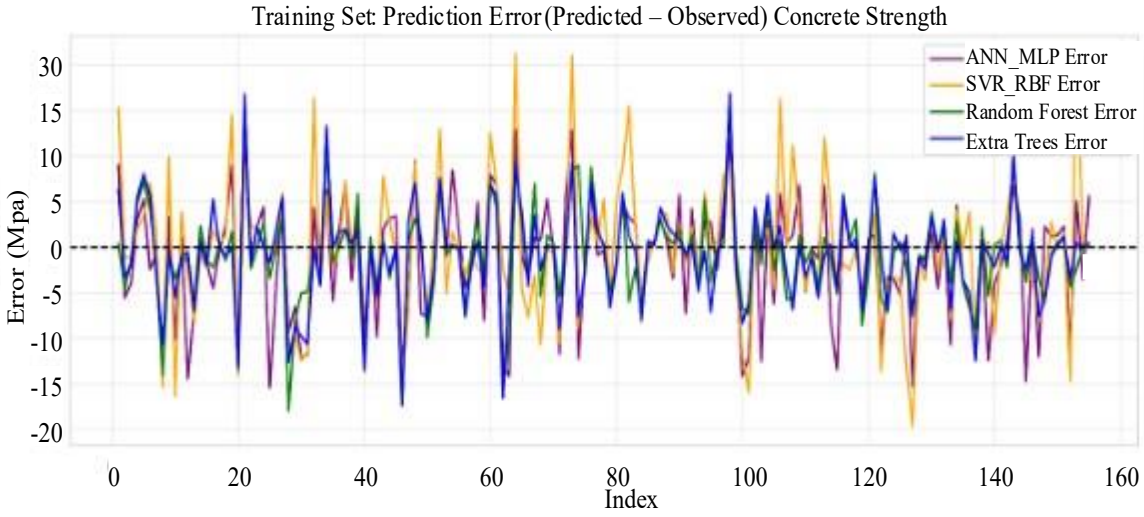

Figure 8. Validation-phase prediction errors of concrete compressive strength.

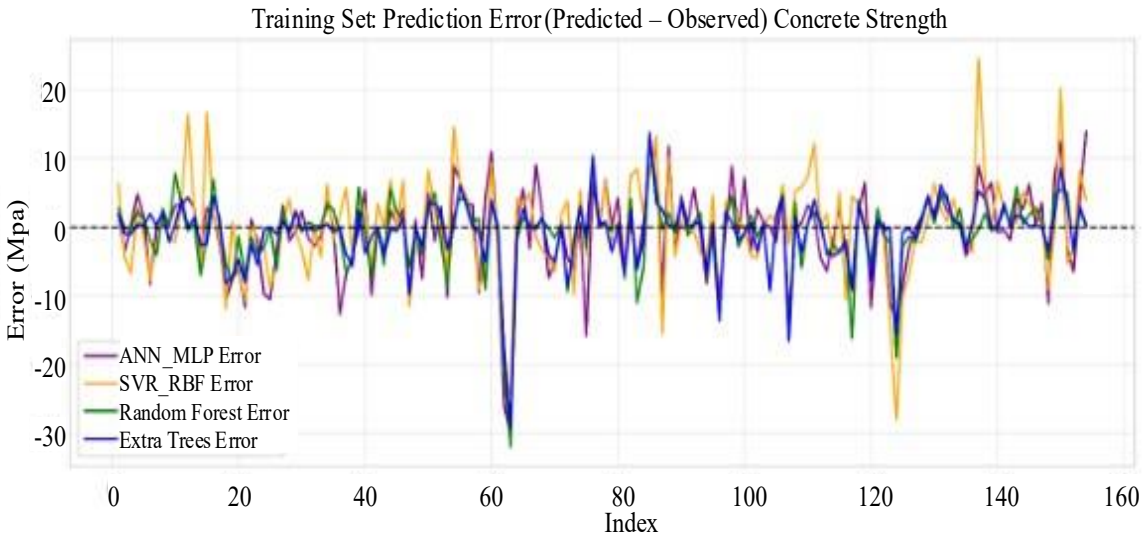

Figure 9. Testing-phase prediction errors of concrete compressive strength.

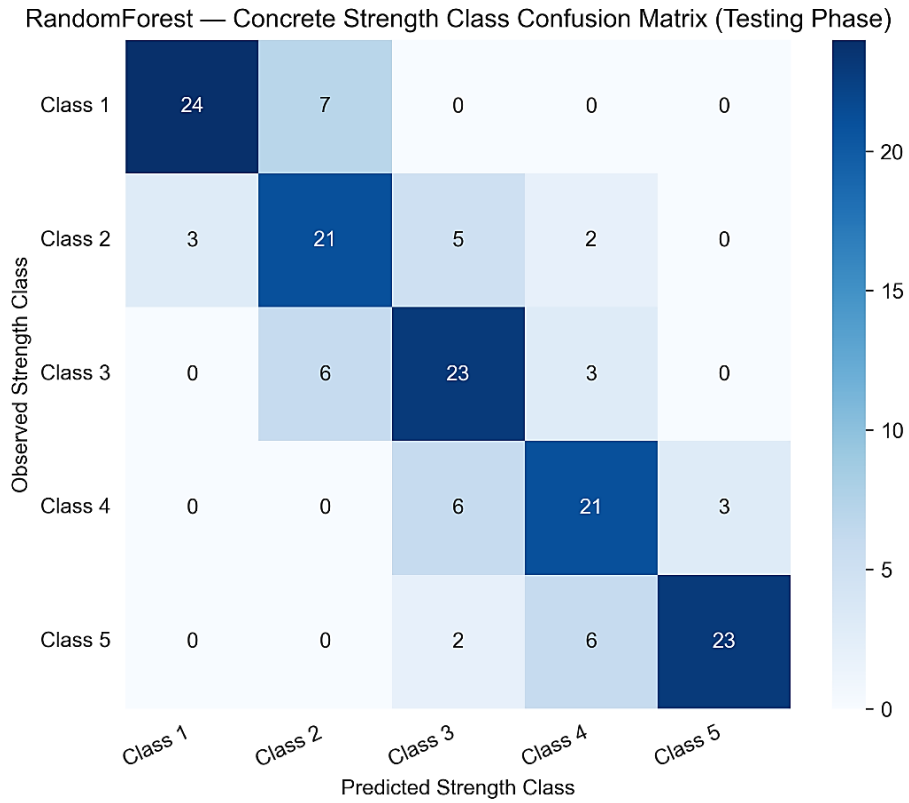

Figure 10. Confusion matrix of concrete strength classes predicted by the Random Forest model during testing.

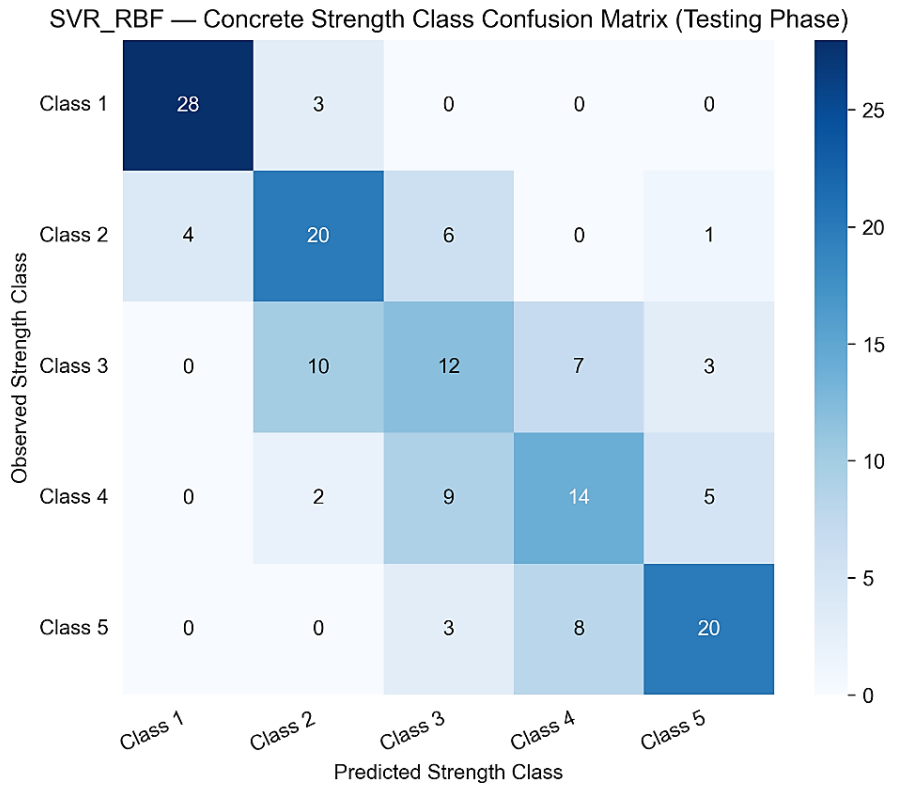

Figure 11. Confusion matrix of concrete strength classes predicted by the SVR-RBF model during testing.

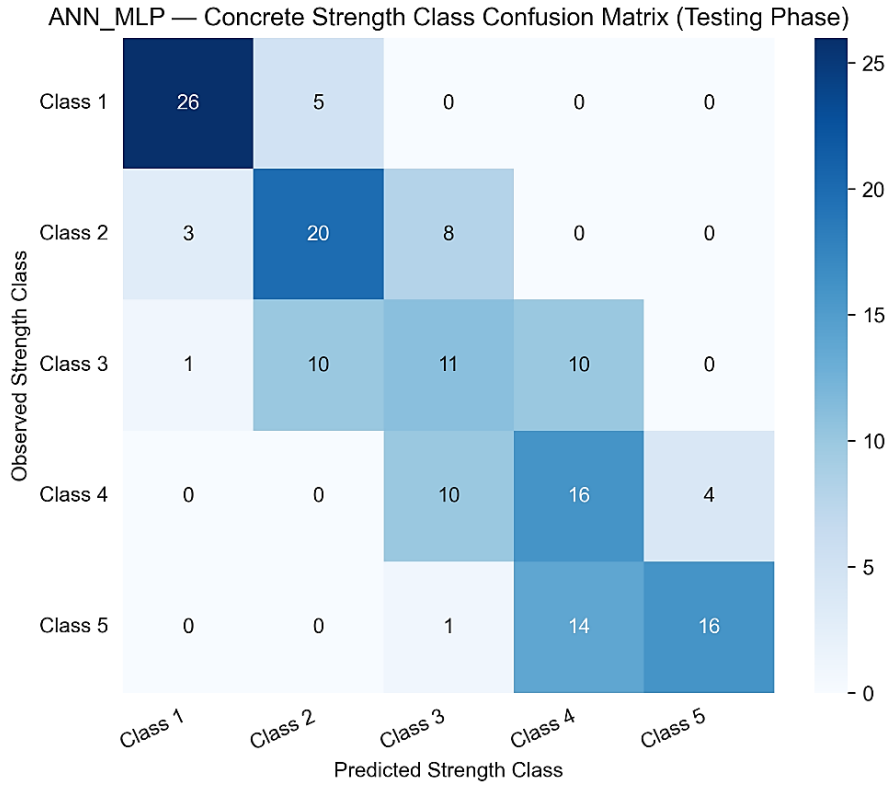

Figure 12. Confusion matrix of concrete strength classes predicted by the ANN-MLP model during testing.

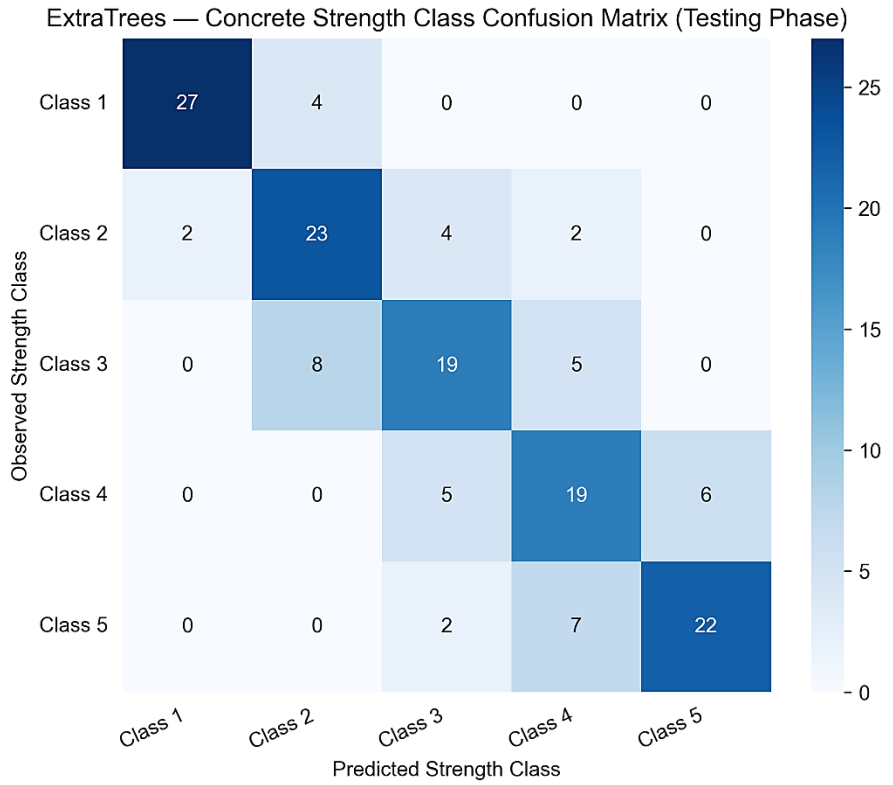

Figure 13. Confusion matrix of concrete strength classes predicted by the Extra Trees model during testing.

CONCLUSIONS

This study developed and systematically evaluated multiple machine learning frameworks for structure–property modeling of heterogeneous cement-based composite systems using compositional parameters and curing age as predictive descriptors. ANN-MLP, SVR-RBF, Random Forest, and Extra Trees models were implemented and assessed across training, validation, and testing phases using comprehensive statistical, robustness-oriented, and classification-based performance measures. The results demonstrate that ensemble tree-based models provide superior capability for modeling nonlinear interactions within multi-component composite systems. In particular, the Extra Trees algorithm achieved the lowest prediction errors and strongest correlation with experimental compressive strength values, while Random Forest exhibited the most stable generalization performance on unseen formulations. Graphical diagnostics and classification-based regime analysis further confirmed the robustness of ensemble approaches in distinguishing composite strength states across heterogeneous compositions. SHAP-based feature attribution enabled quantitative interpretation of phase-level contributions, revealing that matrix-related parameters cement content, curing age, and water–binder interactions dominate compressive strength development, while aggregate fractions exert secondary influence within the investigated compositional domain. These findings reinforce the matrix-controlled nature of strength evolution in cementitious composites and demonstrate that interpretable machine learning can effectively capture underlying structure–property relationships. The proposed framework establishes a transparent and data-driven pathway for rapid screening, performance verification, and optimization of cement-based composite formulations. By combining predictive accuracy with interpretability, the study advances the integration of machine learning into composite materials engineering and supports informed decision-making in mixture design and performance-driven optimization. Future research may extend this approach to durability-related properties, alternative binder chemistries, hybrid and fiber-reinforced composite systems, and sustainability-oriented formulations. Integration with multiscale descriptors and microstructural characterization data could further enhance structure–property modeling capability within advanced composite materials research.

REFERENCES

1. Barbhuiya S, Das BB, Adak D, Kapoor K, Tabish M. Low carbon concrete: advancements, challenges and future directions in sustainable construction. Discover Concrete and Cement. 2025;1:3. https://doi.org/10.1007/s44416-025-00002-y
2. Alexander M, Beushausen H. Durability, service life prediction, and modelling for reinforced concrete structures – review and critique. Cem Concr Res. 2019;122:17–29. https://doi.org/10.1016/j.cemconres.2019.04.018
3. Moccia F, Yu Q, Fernández Ruiz M, Muttoni A. Concrete compressive strength: From material characterization to a structural value. Structural Concrete. 2021;22. https://doi.org/10.1002/suco.202000211
4. DeRousseau MA, Kasprzyk JR, Srubar WV. Computational design optimization of concrete mixtures: A review. Cem Concr Res. 2018;109:42–53. https://doi.org/10.1016/j.cemconres.2018.04.007
5. Ahmed Mohd, Mallick J, AlQadhi S, Ben Kahla N. Development of Concrete Mixture Design Process Using MCDM Approach for Sustainable Concrete Quality Management. Sustainability. 2020;12:8110. https://doi.org/10.3390/su12198110
6. Rahimi MZ, Zhao R, Sadozai S, Zhu F, Ji N, Xu L. Research on the influence of curing strategies on the compressive strength and hardening behaviour of concrete prepared with Ordinary Portland Cement. Case Studies in Construction Materials. 2023;18:e02045. https://doi.org/10.1016/j.cscm.2023.e02045
7. Husem M, Gozutok S. The effects of low temperature curing on the compressive strength of ordinary and high performance concrete. Constr Build Mater. 2005;19:49–53. https://doi.org/10.1016/j.conbuildmat.2004.04.033
8. Simões S. High-Performance Advanced Composites in Multifunctional Material Design: State of the Art, Challenges, and Future Directions. Materials. 2024;17:5997. https://doi.org/10.3390/ma17235997

9. Gao J, Wang C, Chu SH. Mix design of sustainable concrete using generative models. Journal of Building Engineering. 2024;96:110618. https://doi.org/10.1016/j.jobe.2024.110618
10. Farahzadi L, Bozorgmehr Nia S, Shafei B, Kioumarsi M. Sustainability assessment of ultra-high performance concrete made with various supplementary cementitious materials. Cleaner Materials. 2025;15:100301. https://doi.org/10.1016/j.clema.2025.100301
11. Sohail MG, Wang B, Jain A, Kahraman R, Ozerkan NG, Gencturk B, et al. Advancements in Concrete Mix Designs: High-Performance and Ultrahigh-Performance Concretes from 1970 to 2016. Journal of Materials in Civil Engineering. 2018;30. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002144
12. Aziz A, Mehboob SS, Tayyab A, Khan D, Hayyat K, Ali A, et al. Enhancing sustainability in self-compacting concrete by optimizing blended supplementary cementitious materials. Sci Rep. 2024;14:12326. https://doi.org/10.1038/s41598-024-62499-w
13. Mohammed A, Rafiq S, Sihag P, Kurda R, Mahmood W. Soft computing techniques: Systematic multiscale models to predict the compressive strength of HVFA concrete based on mix proportions and curing times. Journal of Building Engineering. 2021;33:101851. https://doi.org/10.1016/j.jobe.2020.101851
14. Ammar MA, Chegenizadeh A, Budihardjo MA, Nikraz H. The Effects of Crystalline Admixtures on Concrete Permeability and Compressive Strength: A Review. Buildings. 2024;14:3000. https://doi.org/10.3390/buildings14093000
15. Ravi Raj V, Madhavan C.S N, Saha BC, Vijay M, D. S, S. S, et al. Graphene-Derived Reinforced Polymer Composites for Self-Healing and Durability Enhancement in Concrete Structures. Journal of Polymer and Composites. 2025;13:77–90.
16. Palanisamy S, Kalimuthu M, Dharmalingam S, Alavudeen A, Nagarajan R, Ismail SO, et al. Effects of fiber loadings and lengths on mechanical properties of Sansevieria Cylindrica fiber reinforced natural rubber biocomposites. Mater Res Express. Institute of Physics; 2023;10. https://doi.org/10.1088/2053-1591/acefb0
17. Palaniappan M, Palanisamy S, Murugesan T mani, Tadepalli S, Khan R, Ataya S, et al. Influence of Washing with Sodium Lauryl Sulphate (SLS) Surfactant on Different Properties of Ramie Fibres. Bioresources. 2024;2:2609–25. https://doi.org/10.15376/biores.19.2.2609-2625
18. Ramasubbu R, Kayambu A, Palanisamy S, Ayrilmis N. Mechanical Properties of Epoxy Composites Reinforced with Areca catechu Fibers Containing Silicon Carbide. Bioresources [Internet]. 2024 [cited 2026 Mar 24];2:2353–70. https://doi.org/10.15376/biores.19.2.2353-2370
19. Yadav M, Srivastava D. Leathercrete-A Sustainable Approach to Fine Aggregate Replacement. Contemp Res in Multi PEER-REVIEWED JOURNAL [Internet]. 4. https://doi.org/10.5281/zenodo.15498179
20. Al-Joulani NMA. Effect of Rubber and Leather Wastes on Concrete Properties. Hebron, West Bank, State of Palestine; 2013 Oct. https://doi.org/https://dsr.ppu.edu/sites/default/files/conference/abstracts/Titl22e.pdf
21. Thakre A, Rajak TK. Utilization of Waste Sole Leather with Fly Ash for Self-Compacting Concrete. IOP Conf Ser Earth Environ Sci. Institute of Physics; 2022. https://doi.org/10.1088/1755-1315/1032/1/012002
22. Litu M, Ali MF, Shabur MA, Taimullah M, Arnob FM. Utilization of chromium tanned leather waste in cementitious and gypsum composites for sustainable construction materials. Discover Sustainability. Springer Nature; 2025;6. https://doi.org/10.1007/s43621-025-01040-z
23. Canhada JCS, Paiva FFG, Shinohara GMM, Okimoto F, Hiranobe CT, Teixeira SR, et al. Production of new concrete with leather tannery waste used as partial replacement of the natural sand. J Mater Cycles Waste Manag. Springer; 2023;25:944–53. https://doi.org/10.1007/s10163-022-01586-4
24. Mohamed S, Elemam H, Seleem MH, Sallam HEDM. Effect of fiber addition on strength and toughness of rubberized concretes. Sci Rep. Nature Research; 2024;14. https://doi.org/10.1038/s41598-024-54763-w
25. Liu Q, Hou F, Ge D, Lv S, Ju Z. Study on Performance and Aging Mechanism of Rubber-Modified Asphalt Under Variable-Intensity UV Aging. Materials. Multidisciplinary Digital Publishing Institute (MDPI); 2025;18. https://doi.org/10.3390/ma18133186

26. Chauhan VS, Sadique MdR, Alam MohdM, Farooqi MohdA. Development of a reliable rock slope stability model utilizing field and analytical data – An integration of FE-ML approaches. Artificial Intelligence in Geosciences. 2025;6:100158. https://doi.org/10.1016/j.aiig.2025.100158
27. Dindar S, Akyurt SE, Sorgun M. A novel nonlinear regression equation for predicting critical flow velocity in slurry transport: A comparative study with advanced machine learning methods and classical empirical correlations. Powder Technol. 2026;467:121604. https://doi.org/10.1016/j.powtec.2025.121604
28. Strielkowski W, Vlasov A, Selivanov K, Muraviev K, Shakhnov V. Prospects and Challenges of the Machine Learning and Data-Driven Methods for the Predictive Analysis of Power Systems: A Review. Energies (Basel). 2023;16:4025. https://doi.org/10.3390/en16104025
29. Tang Y, Kurths J, Lin W, Ott E, Kocarev L. Introduction to Focus Issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2020;30. https://doi.org/10.1063/5.0016505
30. Nafiuzzaman Md, Jakir TI, Aditi IJ, Kabir A, Ahsan KA. Different machine learning approaches to predict the compressive strength of composite cement concrete. Journal of Building Pathology and Rehabilitation. 2025;10:88. https://doi.org/10.1007/s41024-025-00598-5
31. Song Y, Zhao J, Ostrowski KA, Javed MF, Ahmad A, Khan MI, et al. Prediction of Compressive Strength of Fly-Ash-Based Concrete Using Ensemble and Non-Ensemble Supervised Machine-Learning Approaches. Applied Sciences. 2021;12:361. https://doi.org/10.3390/app12010361
32. Shaaban M, Amin M, Selim S, Riad IM. Machine learning approaches for forecasting compressive strength of high-strength concrete. Sci Rep. 2025;15:25567. https://doi.org/10.1038/s41598-025-10342-1
33. Ahmad SA, Ahmed HU, Rafiq SK, Ahmad DA. Machine learning approach for predicting compressive strength in foam concrete under varying mix designs and curing periods. Smart Construction and Sustainable Cities. 2023;1:16. https://doi.org/10.1007/s44268-023-00021-3
34. Ziolkowski P, Niedostatkiewicz M. Machine Learning Techniques in Concrete Mix Design. Materials. 2019;12:1256. https://doi.org/10.3390/ma12081256
35. Du K-L, Jiang B, Lu J, Hua J, Swamy MNS. Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions. Mathematics. 2024;12:3935. https://doi.org/10.3390/math12243935
36. Kanat G, Yıldız Avşar E, Palanisamy S, Ashori A. Utilizing waste manhole covers and fibreboard as reinforcing fillers for thermoplastic composites. Journal of Reinforced Plastics and Composites. 2024;44:1108–18. https://doi.org/10.1177/07316844241238507
37. Manickaraj K, Thirumalaisamy R, Palanisamy S, Ayrilmis N, Massoud EES, Palaniappan M, et al. Value-added utilization of agricultural wastes in biocomposite production: Characteristics and applications. Ann. N. Y. Acad. Sci. John Wiley and Sons Inc; 2025. p. 72–91. https://doi.org/10.1111/nyas.15368
38. Mylsamy B, Aruchamy K, Marudhamuthu Shanmugam SK, Palanisamy S, Ayrilmis N. Improving performance of composites: Natural and synthetic fibre hybridisation techniques in composite materials – A review. Mater Chem Phys [Internet]. 2025;334:130439. https://doi.org/https://doi.org/10.1016/j.matchemphys.2025.130439

