SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# Machine Learning-Based Comparative Assessment of Nano-$SiO_2$, Nano-$TiO_2$, Nano-$Al_2O_3$, Nano-$Fe_2O_3$, and Nano-$CaCO_3$ in Concrete

Nirmal Kumar Bishwas Chaudhary[1], Md Daniyal[2], Rahat Yezdani[3]

[1]Research Scholar, Department of Civil Engineering, Sandip University, Madhubani, India, E-mail: nirmalkbc555666@gmail.com
[2]Assistant Professor, Department of Civil Engineering, Sandip University, Madhubani, India. E-mail: daniyalzhcet@gmail.com
[3]Assistant Professor, Department of CSE, Purnea College of Engineering, Purnea, India. E-mail: rahatyajdani123@gmail.com

**ABSTRACT**

The incorporation of nanomaterials into concrete offers a promising approach for modifying its mechanical performance; however, the combined effects of nanomaterial type, dosage, mixture characteristics, and curing age make direct prediction of strength development challenging. This study develops a machine-learning-based framework for predicting the compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS) of nanomaterial-modified concrete. A dataset comprising 360 observations was considered, covering five nanomaterial types—Nano-$SiO_2$, Nano-$TiO_2$, Nano-$Al_2O_3$, Nano-$Fe_2O_3$, and Nano-$CaCO_3$—at dosages ranging from 0 to 5%, together with cement content, superplasticizer content, and curing age. Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN) models were developed using an 80:20 training–testing division, with five-fold cross-validation and hyperparameter optimization applied during model development. Model performance was assessed using correlation coefficient, coefficient of determination, RMSE, MAE, MAPE, NSE, AAD, and SI. Among the investigated approaches, XGBoost provided the best testing performance for all three mechanical properties, achieving R2 values of 0.964839, 0.966156, and 0.964994 for CS, STS, and FS, respectively. The corresponding RMSE values were 0.721804, 0.056642, and 0.080656 MPa. Feature-importance analysis indicated that curing age was the dominant predictor, while nanomaterial type and dosage also contributed substantially to the predictions. Explainable-AI analysis using feature importance, permutation importance, and SHAP was employed to improve interpretation of the developed models. The results demonstrate the potential of machine learning for reliable prediction and interpretation of mechanical properties of nanomaterial-modified concrete.

**Keywords:** Nanomaterials; Concrete; Compressive Strength; Splitting Tensile Strength; Flexural Strength; Machine Learning

## 1. Introduction

Concrete is the most extensively used construction material worldwide because of its availability, versatility, relatively low cost, and ability to provide adequate mechanical and durability performance. Nevertheless, the heterogeneous and porous nature of the cementitious matrix, together with microcracking and weaknesses at the interfacial transition zone, can limit the long-term performance of concrete. The development of nanotechnology has therefore attracted considerable attention in cement-based materials because nanoparticles can modify hydration reactions, particle packing, pore structure, and the microstructure of the cementitious matrix. Their extremely small particle size and high specific surface area can provide nucleation and filler effects that are difficult to achieve with conventional mineral additions. The effectiveness of a nanoparticle in concrete depends strongly on its chemical composition, particle characteristics, dosage, dispersion, and interaction with cement hydration products. Appropriate nanoparticle contents can improve matrix densification and mechanical performance, whereas excessive dosages may cause agglomeration, increased water demand, and deterioration of workability and strength. Consequently, the effect of nanomaterials should be evaluated over a range of dosages rather than assuming that increasing nanoparticle content will continuously improve concrete performance [1–4].

Among the different nanoparticles investigated in cementitious materials, Nano-$SiO_2$ has received considerable attention because of its high specific surface area, pozzolanic activity, and ability to act as nucleation sites for hydration products. The incorporation of Nano-$SiO_2$ can accelerate cement hydration, promote additional C-S-H formation, refine the pore structure, and densify the cement paste and interfacial transition zone. These mechanisms can result in enhanced mechanical and durability properties; however, excessive Nano-$SiO_2$ may increase particle agglomeration and water demand, resulting in an optimum dosage rather than a simple linear relationship between dosage and strength [5]. Nano-$TiO_2$ has also been extensively investigated in cementitious composites because of its high surface activity and photocatalytic characteristics. Its incorporation can modify hydration, microstructure, mechanical properties, and durability-related characteristics of concrete. Previous experimental investigations have demonstrated that Nano-$TiO_2$ can influence the properties of cementitious composites at different replacement levels and curing conditions [6]. Further studies have shown that the combined incorporation of Nano-$TiO_2$ and chemical inhibitors can modify fresh, hardened, microstructural, and corrosion-resistance characteristics of cementitious composites [7]. Its application has also been investigated in ferrocement composites under different exposure environments, demonstrating the potential of Nano-$TiO_2$ for multifunctional cement-based applications [8].

Other oxide nanoparticles have received increasing attention. Nano-$Al_2O_3$ can influence cement hydration and microstructural development because of its high surface activity and nanoscale particle size. Experimental investigations have demonstrated that Nano-$Al_2O_3$ can alter the mechanical and microstructural characteristics of concrete through multi-scale effects [9]. Similarly, Nano-$Fe_2O_3$ has been investigated as a nano-modifier capable of influencing the physical, mechanical, and microstructural properties of cementitious composites [10]. These findings indicate that nanoparticles with different chemical compositions can produce substantially different effects even when incorporated at comparable dosages. Nano-$CaCO_3$ provides another important mechanism because it can act primarily through filler and nucleation effects and can influence the early development of hydration products. Consequently, its behaviour may differ from that of highly reactive pozzolanic nanoparticles such as Nano-$SiO_2$. Comparative evaluation of different nanoparticles under similar mixture and curing conditions is therefore important for identifying the relative effectiveness of individual nanomaterials. Recent experimental research has provided an important basis for such comparative assessment. A study examining Nano-$Al_2O_3$, Nano-$SiO_2$, and Nano-$CaCO_3$ at different replacement levels demonstrated that nanoparticle type and dosage significantly influence the properties of cementitious composites [11]. More recent comparative research further investigated the mechanical and durability performance of concrete incorporating Nano-$SiO_2$, Nano-$Al_2O_3$, and Nano-$CaCO_3$ and reported that the response depends strongly on both nanoparticle type and dosage [12]. These recent studies provide valuable experimental evidence, but the comparative assessment of a larger group of nanomaterials over multiple curing ages remains relatively limited.

Curing age is another important factor in evaluating nano-modified concrete. Although 28-day strength is commonly adopted as a reference measure, hydration and microstructural development continue beyond 28 days. Nanoparticles can influence both early-age hydration and later-age matrix development; therefore, evaluating only 28-day performance may not adequately describe their complete effect. In the present study, 28, 56, 90, and 180 days are considered to represent early-, intermediate-, and later-age mechanical development. This multi-age framework provides an opportunity to investigate whether the influence of nanoparticle type and dosage changes as concrete continues to mature. Furthermore, the mechanical behaviour of concrete cannot be adequately represented by compressive strength alone. Compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS) represent different aspects of concrete behaviour and may respond differently to changes in matrix densification, interfacial transition zone characteristics, hydration, and microcrack development. A multi-output analysis incorporating these three properties can therefore provide a more comprehensive assessment of nanomaterial-modified concrete.

The large number of combinations involving nanomaterial type, dosage, cement content, superplasticizer content, and curing age makes extensive experimental investigation costly and time-consuming. Machine learning (ML) provides a promising data-driven approach for modeling nonlinear relationships between concrete mixture parameters and mechanical properties. ML models can identify complex relationships without requiring an explicitly defined mathematical formulation and have increasingly been applied to concrete-strength prediction [13,14]. Different algorithms can capture these relationships through fundamentally different learning mechanisms, making comparative evaluation important for identifying an appropriate predictive framework.

In this study, four ML approaches—Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN)—are considered. RF and XGBoost are tree-based ensemble methods capable of representing nonlinear interactions between input variables, whereas SVR uses kernel-based regression and ANN learns nonlinear relationships through interconnected computational neurons. The comparative use of these algorithms allows the predictive capability of different ML architectures to be assessed using a common dataset and evaluation framework. Prediction accuracy, however, is not sufficient for engineering applications because high-performing ML models can behave as black boxes. Understanding the contribution of individual input variables is important for determining whether the predictions are consistent with engineering knowledge and for identifying the variables that exert the greatest influence on concrete performance. Explainable artificial intelligence (XAI) methods, particularly SHapley Additive exPlanations (SHAP), have increasingly been employed to interpret ML-based concrete-strength prediction models [15–18]. SHAP provides a model-agnostic framework for quantifying the contribution of individual input variables to model predictions and can therefore complement conventional performance indicators such as $R^2$, RMSE, and MAE.

This interpretability is particularly relevant to the present study because the objective is not only to identify the most accurate predictive model but also to determine the relative influence of nanomaterial type, nanomaterial dosage, cement content, superplasticizer content, and curing age on CS, STS, and FS. While RF and XGBoost provide intrinsic tree-based feature-importance measures, SVR and ANN do not have directly equivalent native feature-importance mechanisms. Therefore, model-independent interpretation techniques such as permutation importance and SHAP are useful for obtaining a more consistent interpretation across different ML architectures. Despite the growing application of ML to concrete-strength prediction and the increasing use of XAI, there remains a research gap concerning the comparative prediction of multiple mechanical properties of concrete incorporating different nanomaterial types across multiple curing ages. Existing experimental studies have generally focused on individual nanomaterials or a limited combination of materials, while ML studies have predominantly focused on conventional concrete or individual target properties. The integration of five nanomaterials—Nano-$SiO_2$, Nano-$TiO_2$, Nano-$Al_2O_3$, Nano-$Fe_2O_3$, and Nano-$CaCO_3$—with multiple curing ages and multiple mechanical responses can therefore provide a broader data-driven understanding of nano-modified concrete.

Accordingly, the present study develops a comprehensive ML framework for predicting the compressive strength, splitting tensile strength, and flexural strength of concrete incorporating five different nanomaterials. The input variables considered are cement content, nanomaterial type, nanomaterial dosage, superplasticizer content, and curing age, while the mechanical responses are evaluated at 28, 56, 90, and 180 days. RF, XGBoost, SVR, and ANN models are developed and comparatively evaluated using appropriate statistical performance indicators. Feature-importance and SHAP-based analyses are additionally employed to interpret the contribution of the input variables and identify the dominant factors governing the predicted mechanical properties.

## 2. Materials And Methods

### 2.1 Experimental Program and Dataset Development

An experimental investigation was conducted to assess the mechanical performance of nanomaterial-modified concrete under different mixture and curing conditions. The experimental program comprised five nanomaterial types, six dosage levels, and four curing ages, with three replicate observations associated with each experimental condition. The complete dataset comprised 360 observations representing 120 experimental conditions, with three replicate observations for each condition. The investigated curing ages were 28, 56, 90, and 180 days, while the nanomaterial dosage varied from 0% to 5% by weight of cement.

For each observation, the experimental database contained the mixture and curing parameters together with three mechanical responses: compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS). The resulting dataset provided 360 observations for each response variable and was subsequently used for statistical analysis, machine-learning model development, and comparative evaluation.

### 2.2 Materials and Mixture Parameters

Cement was used as the primary binding material, while conventional fine and coarse aggregates constituted the granular skeleton of the concrete. A superplasticizer was incorporated to obtain the required workability of the fresh concrete mixture. Five nanoparticle materials were investigated: Nano-$SiO_2$, Nano-$TiO_2$, Nano-$Al_2O_3$, Nano-$Fe_2O_3$, and Nano-$CaCO_3$. The nanoparticles were incorporated into the cementitious system at controlled proportions relative to the cement content.

The use of different nanomaterial compositions provides a basis for evaluating the influence of particle characteristics and material-specific interactions with the cementitious matrix. Nanoparticle incorporation can modify the microstructure of concrete through mechanisms including nucleation, filler action, pore refinement, and improvement of the cement paste–aggregate interface. These mechanisms can subsequently influence the development of compressive, tensile, and flexural strength. The experimental mixtures were prepared under controlled conditions, while the constituent materials and admixture quantities were maintained according to the defined mixture proportions. The resulting variations in nanomaterial dosage and curing duration provided the principal experimental conditions for the subsequent ML analysis.

### 2.3 Selection of Input and Output Variables

The variables used for machine-learning analysis were selected according to their physical relevance, variability, and predictive significance. The final input and output variables are summarized in Table 1.

**Table 1. Input and output variables used for machine-learning analysis**

| Category | Variable | Symbol | Unit/Representation |
|---|---|---|---|
| Input | Cement content | C | $kg/m^3$ |
| Input | Nanomaterial type | N | Categorical |
| Input | Nanomaterial dosage | D | % |
| Input | Superplasticizer content | SP | % / $kg/m^3$ |
| Input | Curing age | A | days |
| Output | Compressive strength | CS | MPa |
| Output | Splitting tensile strength | STS | MPa |

| | | | |
|---|---|---|---|
| Output | Flexural strength | FS | MPa |

The five selected predictors represent the principal material, compositional, and temporal characteristics expected to influence the mechanical response. Nanomaterial type represents the categorical distinction among the five investigated nanoparticles, whereas nanomaterial dosage represents the quantity incorporated into the cementitious system. Cement content represents the principal binder quantity, and superplasticizer content accounts for the influence of the chemical admixture incorporated into the mixture. Curing age represents the temporal development of the concrete matrix. Variables that did not provide independent predictive information were excluded from the final feature set. In particular, variables exhibiting constant or redundant values were removed to avoid unnecessary dimensionality and potential distortion of the ML models. This feature-selection procedure ensured that the final models were developed using physically interpretable and informative predictors.

### 2.4 Data Preprocessing and Exploratory Analysis

Prior to machine-learning model development, the experimental dataset was systematically examined to ensure data consistency, completeness, and suitability for predictive analysis. The preprocessing procedure was designed to retain the physically meaningful variables identified in Section 2.3 while removing irrelevant information that could introduce redundancy or adversely affect model interpretation.

The dataset was first inspected for missing values, duplicate entries, inconsistent data types, and invalid observations. The numerical variables were converted into appropriate numerical formats, and the categorical variable representing nanomaterial type was encoded into a machine-readable form. Identification variables and other non-predictive attributes were excluded from the final feature matrix. This resulted in a compact feature set consisting of nanomaterial type, nanomaterial dosage, cement content, superplasticizer content, and curing age.

Exploratory data analysis (EDA) was subsequently performed to understand the statistical characteristics and variability of the experimental observations. Descriptive statistical measures, including mean, standard deviation, minimum, maximum, and quartile values, were calculated for the numerical input and output variables. These statistics were used to assess the overall range and dispersion of the experimental data.

The distributions of the principal mechanical responses were further examined using graphical analysis. Distribution plots were used to evaluate the spread of compressive strength, splitting tensile strength, and flexural strength, while comparative plots were used to examine variations associated with nanomaterial type, dosage, and curing age. Such analysis provided an initial understanding of the differences in mechanical performance among the investigated mixtures.

A correlation analysis was also performed to quantify the linear association between the numerical variables and the mechanical responses. The resulting correlation matrix was visualized using a heatmap to facilitate identification of strongly associated variables and potential redundancy among predictors. However, correlation coefficients were not used as the sole basis for feature elimination because the relationships between nanomaterial characteristics and concrete strength may be nonlinear.

The exploratory analysis was particularly important for identifying the variation in mechanical properties with nanomaterial dosage and curing age. The response was not assumed to follow a simple linear trend, allowing the subsequent ML models to capture possible nonlinear interactions between mixture parameters and mechanical performance.

Following preprocessing and EDA, the cleaned dataset was separated into input variables and target variables. The resulting feature matrix was used for the development of the Random Forest (RF), XGBoost, Support Vector Regression (SVR), and Artificial Neural Network (ANN) models. The same processed dataset structure was maintained across the different algorithms to ensure a consistent and unbiased comparison of their predictive performance.

### 2.5 Machine-Learning Models

Four machine-learning (ML) algorithms were employed to predict the mechanical properties of nanomaterial-modified concrete: Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN). These algorithms were selected because they represent different learning strategies, including ensemble tree-based learning, gradient boosting, kernel-based regression, and neural-network-based nonlinear modeling. The models were developed independently for compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS).

The selected input vector can be represented as:

$$X = [N, D, C, SP, A]$$

where N is nanomaterial type, D is nanomaterial dosage, C is cement content, SP is superplasticizer content, and A is curing age. For each target variable, the ML models approximate the nonlinear relationship:

$$\hat{Y} = f(X)$$

where $\hat{Y}$ denotes the predicted mechanical response. i.e., compressive strength (CS), splitting tensile strength (STS), or flexural strength (FS)

### 2.5.1 Random Forest (RF)

Random Forest is an ensemble-based supervised learning algorithm that combines the predictions of multiple decision trees to obtain a robust regression model. Each tree is developed using a randomly selected subset of observations and predictor variables, and the final prediction is obtained by aggregating the predictions generated by the individual trees.

For a regression problem, the RF prediction can be expressed as:

$$\hat{Y}_{RF}(X) = \frac{1}{B}\sum_{b=1}^{B} T_b\,(X)$$

where B is the number of decision trees and Tb(X) represents the prediction of the bth tree for the input vector X.

RF was selected because of its ability to capture nonlinear relationships and interactions among concrete mixture parameters without requiring an explicitly defined functional relationship. It is also relatively insensitive to differences in the scale of numerical predictors and provides an intrinsic measure of feature importance. In the present study, RF was trained separately for CS, STS, and FS, followed by hyperparameter optimization to obtain the best-performing configuration.

### 2.5.2 Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an ensemble learning technique based on gradient-boosted decision trees. Unlike RF, in which trees are generally developed independently, XGBoost constructs trees sequentially, with each subsequent tree attempting to reduce the prediction errors generated by the preceding ensemble.

The general objective function of XGBoost can be expressed as:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where l(yi, ŷi) represents the loss between the observed and predicted values, K is the number of trees, and Ω(fk) represents the regularization term associated with the kth tree.

XGBoost was selected because of its capability to model complex nonlinear relationships and interactions between the input variables while controlling model complexity through regularization. The model was developed independently for each mechanical-property target and optimized using the selected hyperparameter-search procedure.

### 2.5.3 Support Vector Regression (SVR)

Support Vector Regression is a kernel-based regression algorithm that determines a function capable of predicting the target variable while maintaining an acceptable tolerance around the observed responses. SVR seeks a function that minimizes model complexity while limiting prediction errors within a predefined -insensitive region.

The regression function can be represented as:

$$f(X) = \omega^T \phi(X) + b$$

where ω represents the model weight vector, ϕ(x) denotes the transformation of the input variables into feature space, and b is the bias term.

A kernel function is used to represent nonlinear relationships without explicitly transforming the input variables into a higher-dimensional space. In the present study, SVR was included to provide a fundamentally different predictive approach from the tree-based RF and XGBoost models. The input variables were appropriately processed before SVR training because kernel-based models can be sensitive to differences in feature scale. Separate SVR models were developed for CS, STS, and FS, and their hyperparameters were optimized to obtain the most suitable predictive configuration.

### 2.5.4 Artificial Neural Network (ANN)

An Artificial Neural Network (ANN) is a nonlinear computational model composed of interconnected processing units arranged in different layers. In the present investigation, the ANN was used to establish the nonlinear relationship between the selected mixture parameters and the mechanical properties of concrete.

For a neuron, the output can generally be expressed as:

$$z = f\left(\sum_{j=1}^{m} \omega_j x_j + b\right)$$

where xj represents an input variable, ωj is its corresponding weight, b is the bias, and f is the activation function.

The ANN learns the relationship between input and output variables by iteratively adjusting the network weights and biases during training. The model is capable of representing complex nonlinear interactions among nanomaterial type, dosage, cement content, superplasticizer content, and curing age. Separate ANN models were developed for CS, STS, and FS. The network architecture and training parameters were optimized using the predefined modeling procedure. The trained networks were subsequently evaluated using the same testing data and performance indicators applied to the other ML algorithms, ensuring a consistent comparison among RF, XGBoost, SVR, and ANN.

### 2.6 Model Training and Hyperparameter Optimization

The dataset was divided into 80% training and 20% testing subsets using a fixed random state of 42. The training subset was used for model fitting, five-fold cross-validation, and hyperparameter optimization, whereas the testing subset was retained exclusively for the final evaluation of predictive performance. This procedure provided a consistent basis for comparing the RF, XGBoost, SVR, and ANN models and enabled an independent assessment of their ability to generalize to previously unseen observations.

For the RF, XGBoost, and SVR models, hyperparameter optimization was performed using a systematic search procedure with five-fold cross-validation. The optimal parameter combination was selected based on the cross-validation performance obtained from the training dataset. The optimized models were subsequently refitted using the training data and evaluated on the reserved testing dataset.

For the ANN model, the network parameters and training configuration were optimized using the defined training procedure, while the testing subset remained unseen during model development. The same 80:20 data partition was maintained for all four algorithms to ensure a fair comparison of predictive performance.

The final models were used to predict CS, STS, and FS for the testing observations. The predicted values were then compared with the corresponding experimental values using the statistical indicators.

### 2.7 Model Performance Evaluation

The predictive performance of the developed Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Artificial Neural Network (ANN) models was evaluated using widely accepted statistical performance indicators. These metrics were selected to assess the correlation, goodness of fit, magnitude of prediction errors, relative prediction error, and normalized prediction error between the experimentally observed and model-predicted mechanical properties. The evaluation metrics included the Correlation Coefficient (R), Coefficient of Determination (R2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Nash–Sutcliffe Efficiency (NSE), Average Absolute Deviation (AAD), and Scatter Index (SI).

Let yi, represent the experimentally observed value, ŷi represent the corresponding model-predicted value, ȳ represent the mean experimentally observed value, ŷ represent the mean predicted value, and denote the total number of observations.

**Correlation Coefficient (R)**

The correlation coefficient (R) measures the strength of the linear association between the experimentally observed and predicted corrosion rates. Values approaching unity indicate a strong positive correlation and, consequently, better agreement between the model predictions and experimental observations.

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

**Coefficient of Determination (R2)**

The coefficient of determination (R2) represents the proportion of variation in the experimentally observed corrosion rate explained by the predictive model. Values approaching unity indicate a high degree of agreement between observed and predicted values.

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

**Root Mean Square Error (RMSE)**

RMSE quantifies the magnitude of prediction errors and gives greater weight to larger deviations because the errors are squared before averaging. Lower RMSE values indicate better predictive accuracy.

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

**Mean Absolute Error (MAE)**

MAE represents the average absolute difference between the observed and predicted corrosion rates. Lower MAE values indicate closer agreement between the model predictions and experimental measurements.

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

**Mean Absolute Percentage Error (MAPE)**

MAPE expresses the prediction error relative to the observed corrosion rate in percentage terms. Lower MAPE values indicate better predictive accuracy.

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

**Nash–Sutcliffe Efficiency (NSE)**

The Nash–Sutcliffe Efficiency was used to assess the predictive efficiency of the models relative to the mean observed response:

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

An NSE value approaching 1 indicates high predictive efficiency, whereas lower values indicate poorer agreement between predicted and observed responses.

**Scatter Index (SI)**

The Scatter Index (SI) is a normalized error measure obtained by relating RMSE to the mean observed corrosion rate. Lower SI values indicate better predictive performance.

$$SI = \frac{RMSE}{\bar{y}}$$

**2.8 Feature-Importance and Explainable-AI Analysis**

Feature-importance and explainable-artificial-intelligence (XAI) analyses were performed to investigate the contribution of the selected input variables to the predicted mechanical properties. While predictive accuracy indicates how well a model reproduces the experimental responses, model-interpretation techniques provide additional information regarding which variables control the predictions and how strongly they contribute. The interpretation analysis considered the five input variables: nanomaterial type, nanomaterial dosage, cement content, superplasticizer content, and curing age. The analysis was conducted separately for CS, STS, and FS.

**2.8.1 Feature Importance of RF and XGBoost**

The intrinsic feature-importance measures provided by the tree-based RF and XGBoost models were used to quantify the relative contribution of the input variables. For RF, feature importance was derived from the contribution of individual predictors to the reduction in impurity across the ensemble of decision trees. For XGBoost, feature importance was obtained from the contribution of the predictors to the construction of the boosted trees. The importance values were normalized and expressed as percentages to facilitate comparison among the input variables:

$$I_j(\%) = \frac{I_j}{\sum_{j=1}^{p} I_j} \times 100$$

where Ijrepresents the importance of the jthfeature and pis the total number of input features.

The resulting importance values were used to identify the dominant predictors for CS, STS, and FS and to compare whether the two tree-based algorithms produced consistent rankings.

**2.8.2 SHAP Analysis**

SHAP (SHapley Additive exPlanations) analysis was employed to provide a more detailed interpretation of the relationship between the input variables and model predictions. SHAP assigns each input feature a contribution value for an individual prediction relative to a reference prediction.

The prediction can be represented as:

$$f(x) = \phi_0 + \sum_{j=1}^{p} \phi_j$$

where fxis the model prediction for a given observation, ϕ0is the expected model output, and ϕjrepresents the SHAP contribution of the jthfeature.

The magnitude of the SHAP value indicates the strength of a feature's contribution, while its sign indicates whether the feature increases or decreases the predicted response relative to the reference value. SHAP analysis was used to examine both global and individual prediction behaviour. Global interpretation was obtained from the distribution and mean absolute SHAP values of the input features, whereas individual SHAP values were used to examine the direction and magnitude of feature contributions for specific observations. The SHAP results were subsequently compared with the conventional feature-importance and permutation-importance results to obtain a comprehensive interpretation of the ML models and to identify the parameters most strongly associated with variations in compressive strength, splitting tensile strength, and flexural strength.

## 3. Results And Discussion

**3.1 Descriptive Statistics of the Experimental Dataset**

The descriptive statistics of the 360 experimental observations are presented in Table 2. The dataset covers nanomaterial dosages from 0 to 5%, cement contents from 361 to 380 kg/m$^3$, superplasticizer contents from 1.80 to 1.90 kg/m$^3$, and curing ages from 28 to 180 days. The three mechanical-property responses exhibited considerable variation across the experimental observations. The mean values of CS, STS, and FS were 47.18 MPa, 3.91 MPa, and 5.45 MPa, respectively, with corresponding standard deviations of 4.04 MPa, 0.32 MPa, and 0.45 MPa. The ranges of the three responses were 39.60–58.17 MPa for CS, 3.29–4.70 MPa for STS, and 4.61–6.63 MPa for FS, indicating sufficient variability in the target variables for subsequent machine-learning analysis.

**Table 2. Descriptive statistics of input and output variables**

| Variable | Mean | Std. Dev. | Min. | 25% | Median | 75% | Max. |
|---|---|---|---|---|---|---|---|
| Nano_Dosage (%) | 2.5 | 1.71 | 0 | 1 | 2.5 | 4 | 5 |
| Cement (kg/m$^3$) | 370.5 | 6.499 | 361 | 364.8 | 370.5 | 376.2 | 380 |
| Superplasticizer (kg/m$^3$) | 1.85 | 0.034 | 1.8 | 1.82 | 1.85 | 1.88 | 1.9 |
| Curing Age (days) | 88.5 | 57.288 | 28 | 49 | 73 | 112.5 | 180 |
| Compressive Strength (MPa) | 47.178 | 4.038 | 39.6 | 43.8 | 47.045 | 50.105 | 58.17 |
| Splitting Tensile Strength (MPa) | 3.907 | 0.324 | 3.29 | 3.64 | 3.905 | 4.17 | 4.7 |
| Flexural Strength (MPa) | 5.45 | 0.453 | 4.61 | 5.06 | 5.455 | 5.8 | 6.63 |

### 3.2 Correlation Analysis of Input and Output Variables

The correlation matrix of the numerical variables is presented in Figure 1. The results provide an initial assessment of the linear relationships among the mixture parameters, curing age, and mechanical-property responses.

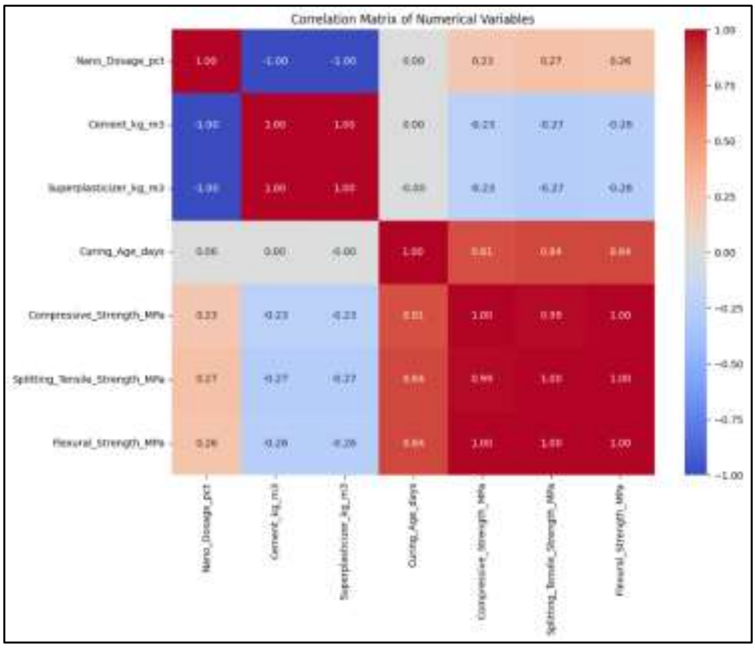


**Figure 1. Correlation matrix of numerical input and output variables.**

A strong positive relationship was observed between curing age and the three mechanical properties, with correlation coefficients of 0.81 for CS, 0.84 for STS, and 0.84 for FS. This indicates that curing age is strongly associated with the development of mechanical performance within the investigated range. The result is consistent across all three response variables and indicates the importance of incorporating curing age into the predictive framework.

Nanomaterial dosage exhibited relatively weak positive correlations with CS (0.23), STS (0.27), and FS (0.26). Although these coefficients are comparatively modest, they should not be interpreted as evidence that nanomaterial dosage is unimportant. The effect of dosage may involve nonlinear responses and interactions with nanomaterial type and curing age, which are not adequately represented by a simple Pearson correlation coefficient.

The correlation between cement content and the three mechanical responses was negative, with coefficients of −0.23 for CS, −0.27 for STS, and −0.26 for FS. Superplasticizer content exhibited the same correlation pattern and magnitude. However, an important characteristic of the dataset is evident from the predictor-to-predictor correlations. Nano dosage and cement content show a correlation of −1.00, whereas cement and superplasticizer content show a correlation of +1.00. This indicates that these variables are mathematically linked within the constructed dataset rather than independently varying experimental factors.

This observation is important for the subsequent ML interpretation. Although cement content was retained as an input because of its physical relevance, the perfect correlations indicate that its individual feature importance should not be interpreted as completely independent of nanomaterial dosage or superplasticizer content. Tree-based and other nonlinear models may distribute predictive contribution among strongly associated variables in a model-dependent manner.

The three target variables also show very strong mutual correlations, with CS–STS = 0.99, CS–FS = 1.00, and STS–FS = 1.00. This indicates that the three mechanical responses vary very similarly across the observations. Since CS, STS, and FS are treated as separate prediction targets rather than input variables, this strong inter-target correlation does not constitute data leakage in the present modeling framework.

### 3.3 Observed versus Predicted Mechanical Properties

The agreement between the observed and predicted values was examined using the testing dataset for the three mechanical-property responses: compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS). Observed-versus-predicted plots provide a direct graphical assessment of model generalization because the testing observations were not used for final model fitting. A 1:1 reference line was used to represent ideal prediction, for which the predicted value is equal to the corresponding observed value.

The statistical results presented in Table 3 indicate that all four models provided a strong correspondence between observed and predicted responses, although the magnitude of agreement varied among the algorithms. XGBoost consistently produced the highest testing R2for all three properties, with values of 0.964839 for CS, 0.966156 for STS, and 0.964994 for FS. The corresponding correlation coefficients were 0.983317, 0.983253, and 0.982719, respectively. These results indicate that the XGBoost predictions closely followed the observed experimental variation.

#### 3.3.1 Compressive Strength

The observed-versus-predicted relationship for compressive strength is presented in Figure 2. The predictions of all four models are distributed predominantly around the 1:1 line, demonstrating a strong agreement between the observed and predicted CS values. Among the investigated models, XGBoost showed the closest agreement, achieving a testing Rof 0.983317 and R2of 0.964839. Its RMSE and MAE were 0.721804 MPa and 0.595595 MPa, respectively. The corresponding MAPE and SI values were 1.243965% and 0.015145. These results demonstrate that XGBoost was able to reproduce the variation in compressive strength with relatively small prediction errors. SVR and RF also demonstrated strong predictive agreement, with testing R2values of 0.954768 and 0.950929, respectively. ANN produced a lower R2of 0.914510, with a comparatively higher RMSE of 1.125497 MPa. Thus, the graphical and statistical assessments consistently identify XGBoost as the most accurate model for CS prediction.

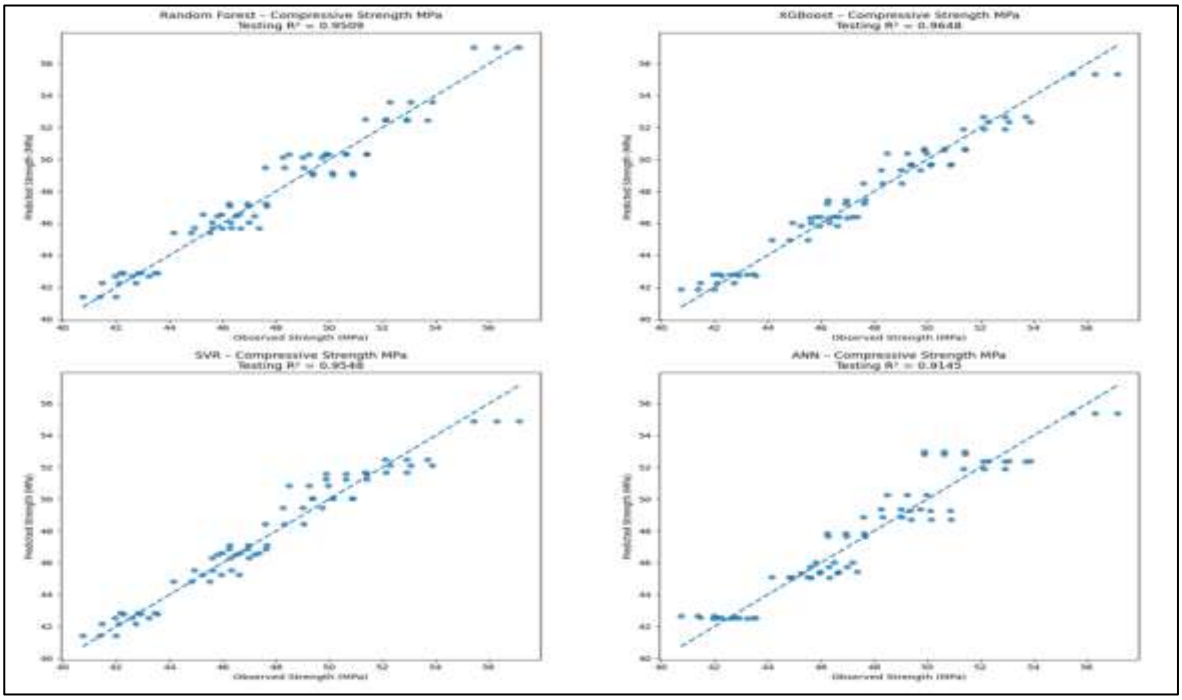

**Figure 2. Observed versus predicted compressive strength for the testing dataset**

#### 3.3.2 Splitting Tensile Strength

The observed and predicted splitting tensile strengths are compared in Figure 3. The predictions are closely distributed around the 1:1 line, indicating a strong agreement between the model outputs and observed STS values. XGBoost produced the strongest testing performance, with R=0.983253and R2=0.966156. The model achieved an RMSE of only 0.056642 MPa and an MAE of 0.047509 MPa. The corresponding MAPE was

1.200576%, while the SI was 0.014334. These results indicate a high degree of predictive accuracy for splitting tensile strength. SVR produced a testing R2of 0.941890, while RF achieved 0.933216. ANN provided an R2of 0.919782. Therefore, although all four approaches captured the overall variation in STS, XGBoost demonstrated the closest agreement between predicted and observed values.

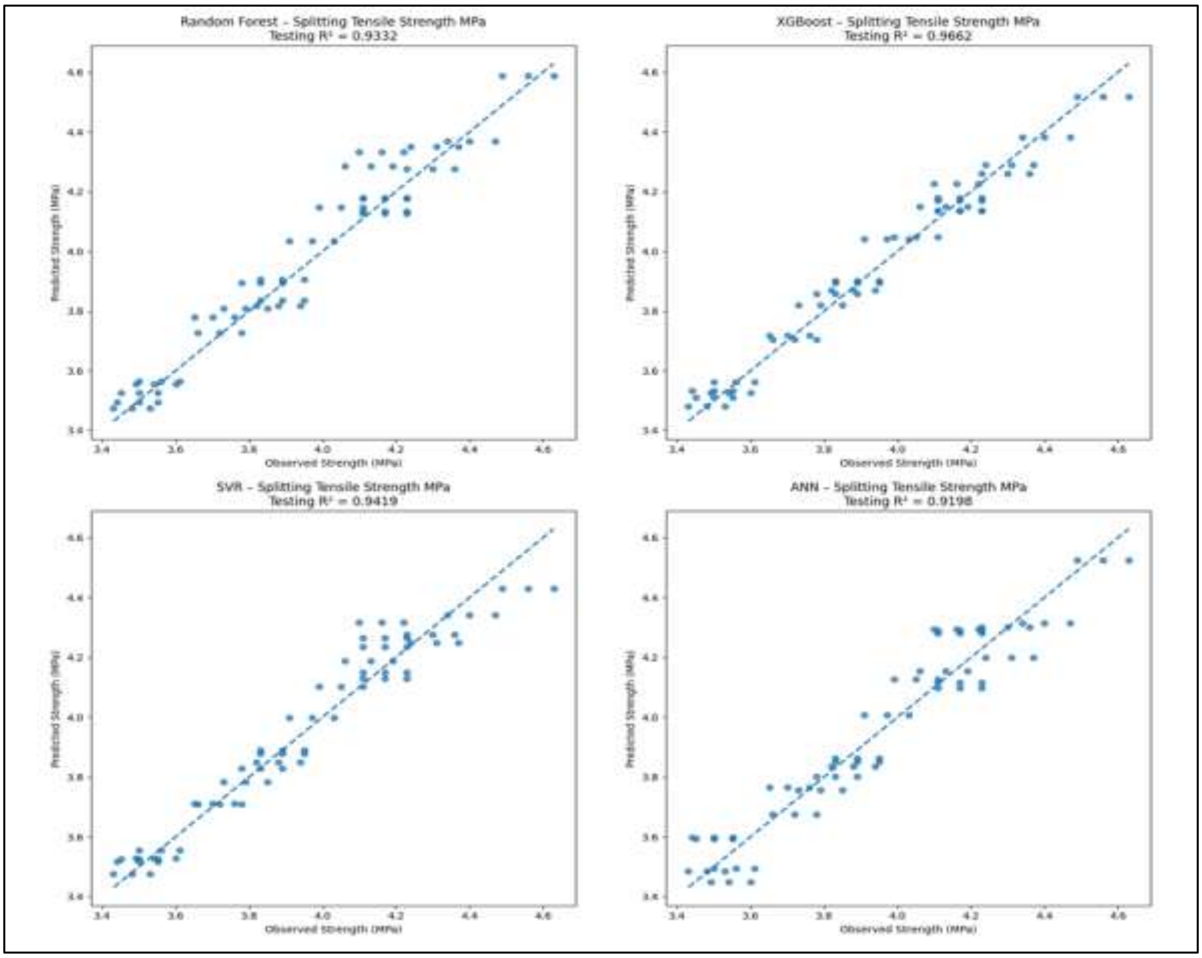


**Figure 3. Observed versus predicted splitting tensile strength for the testing dataset**

### 3.3.3 Flexural Strength

The observed-versus-predicted results for flexural strength are presented in Figure 4. The predicted values of the tree-based and kernel-based models generally follow the 1:1 reference line, indicating good predictive capability. XGBoost again provided the best testing performance, obtaining R=0.982719and R2=0.964994. The corresponding RMSE and MAE were 0.080656 MPa and 0.066620 MPa, respectively. The model also produced a MAPE of 1.205818% and an SI of 0.014632. SVR achieved a testing R2of 0.952726, followed by RF with 0.941405. ANN showed the lowest predictive performance for FS, with R2=0.864152, RMSE of 0.158889 MPa, and MAE of 0.130079 MPa.

The observed-versus-predicted analysis therefore confirms the findings from the quantitative model evaluation. XGBoost consistently provided the closest correspondence between observed and predicted CS, STS, and FS values, while SVR and RF also demonstrated strong predictive capability. ANN produced acceptable predictions but showed comparatively greater dispersion, particularly for flexural strength.

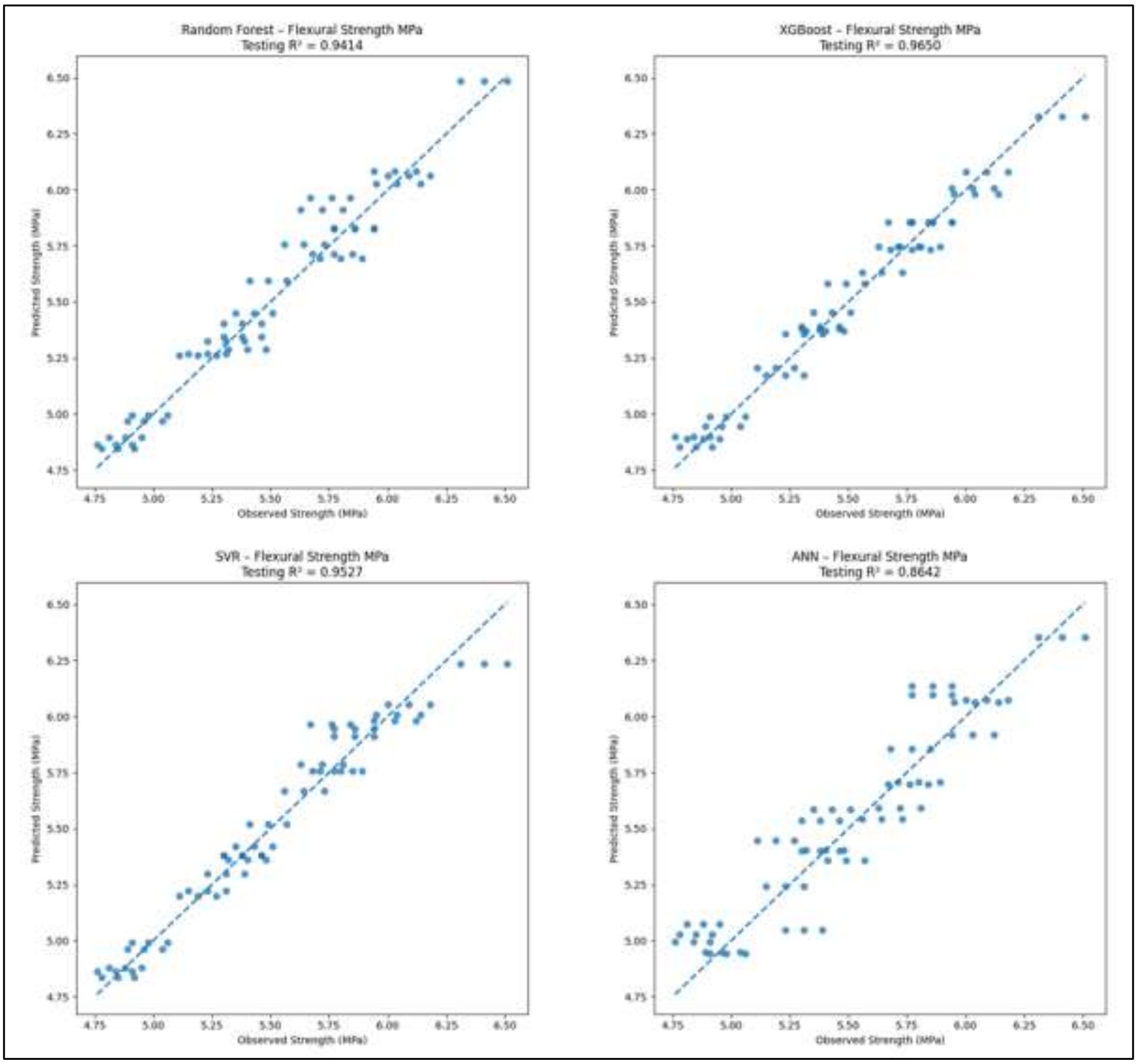


**Figure 4. Observed versus predicted flexural strength for the testing dataset**

### 3.4 Residual Distribution Analysis

Residual analysis was performed to further assess the predictive reliability of the developed machine-learning models. For each observation, the residual was calculated as the difference between the experimentally observed value and the corresponding model-predicted value. The residual distributions were examined separately for compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS) using the testing dataset. A well-performing model is expected to produce residuals that are generally centered around zero, with a relatively narrow spread and without a systematic pattern. A balanced distribution around zero indicates that the model does not consistently overpredict or underpredict the response.

#### 3.4.1 Residual Distribution for Compressive Strength

The residual distributions for CS should be presented in Figure 5. The distributions should be compared among RF, XGBoost, SVR, and ANN.Based on the testing performance reported in Table 3, XGBoost provides the lowest RMSE (0.721804 MPa) and MAE (0.595595 MPa) among the four models for CS. Therefore, its residuals are expected to exhibit a comparatively narrow distribution. RF and SVR also demonstrate relatively low prediction errors, whereas ANN exhibits a wider residual spread, consistent with its higher RMSE (1.125497 MPa).

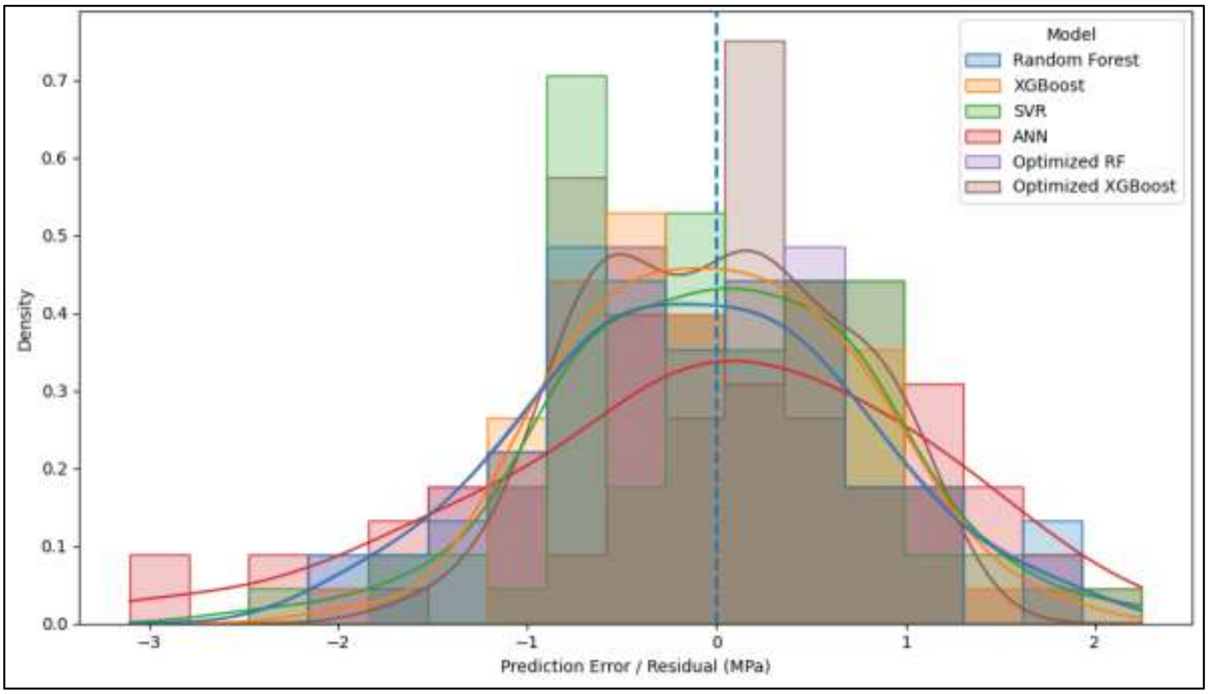


**Figure 5. Residual distributions of RF, XGBoost, SVR, and ANN for compressive strength.**

### 3.4.2 Residual Distribution for Splitting Tensile Strength

The residual distributions for STS are presented in Figure 6. XGBoost achieved the lowest testing RMSE (0.056642 MPa) and MAE (0.047509 MPa), indicating a relatively concentrated residual distribution. SVR and RF also produced low prediction errors, while ANN showed a comparatively greater residual spread.

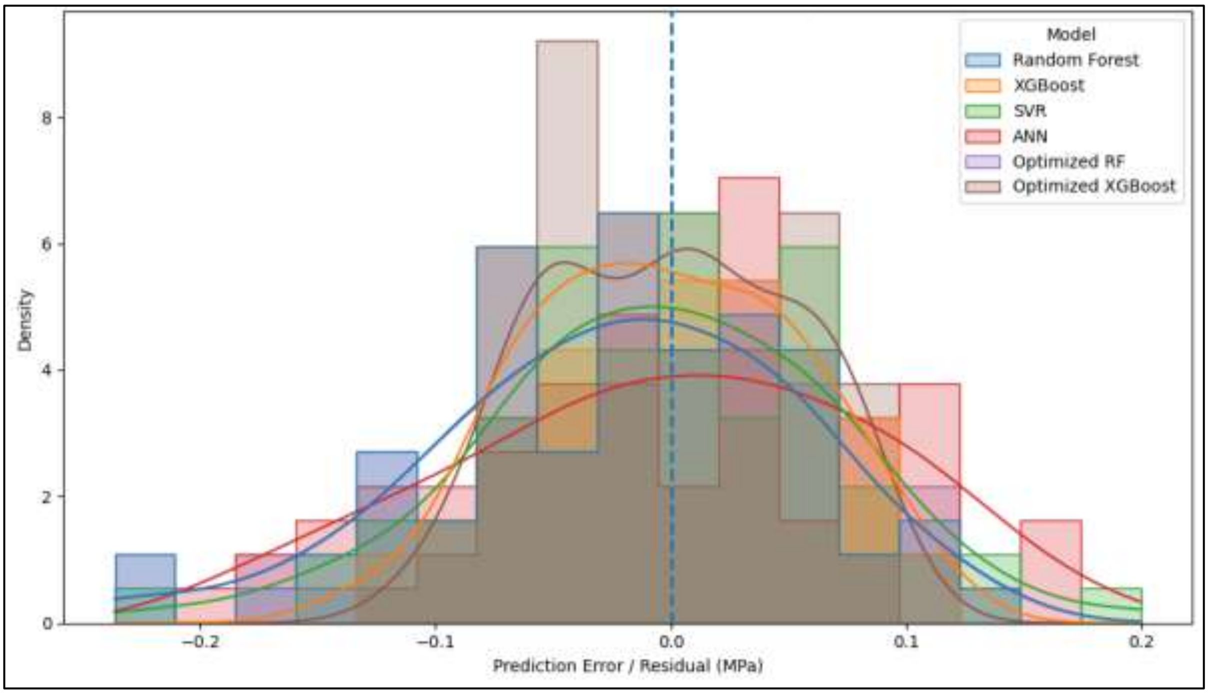


**Figure 6. Residual distributions of RF, XGBoost, SVR, and ANN for splitting tensile strength.**

### 3.4.3 Residual Distribution for Flexural Strength

The residual distributions for FS are presented in Figure 7. XGBoost produced the lowest testing RMSE (0.080656 MPa) and MAE (0.066620 MPa), followed by SVR and RF. ANN exhibited the largest RMSE (0.158889 MPa) among the models, indicating comparatively greater dispersion of prediction errors.

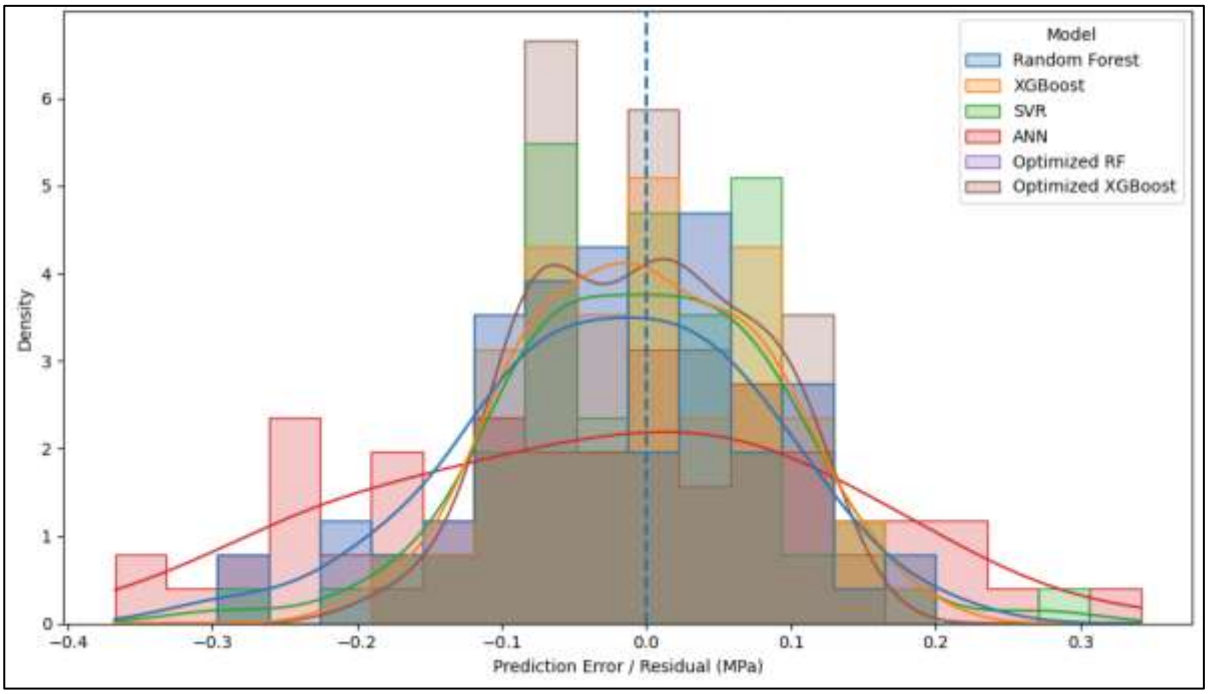


**Figure 7. Residual distributions of RF, XGBoost, SVR, and ANN for flexural strength**

### 3.5 Feature-Importance Analysis of RF and XGBoost

Feature-importance analysis was performed using the intrinsic importance measures of the Random Forest (RF) and Extreme Gradient Boosting (XGBoost) models. The analysis was conducted independently for compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS). The normalized importance values were converted into percentages to facilitate comparison among the input variables.

**Table 3. Feature importance obtained from RF and XGBoost models**

| Feature | RF-CS (%) | XGB-CS (%) | RF-STS (%) | XGB-STS (%) | RF-FS (%) | XGB-FS (%) |
|---|---|---|---|---|---|---|
| Curing Age (days) | 76.432 | 57.314 | 80.536 | 64.844 | 80.472 | 64.509 |
| Nano-$SiO_2$ | 9.096 | 17.615 | 2.812 | 6.848 | 4.514 | 9.627 |
| Nano-$CaCO_3$ | 4.983 | 12.874 | 4.042 | 9.58 | 3.834 | 9.159 |
| Nano Dosage (%) | 3.193 | 3.705 | 4.601 | 4.784 | 3.886 | 4.355 |
| Cement (kg/$m^3$) | 2.751 | 5.645 | 3.353 | 10.429 | 3.093 | 9.539 |
| Superplasticizer (kg/$m^3$) | 2.793 | 0 | 3.433 | 0 | 3.36 | 0 |
| Nano-$Fe_2O_3$ | 0.348 | 1.323 | 0.845 | 2.096 | 0.461 | 1.427 |
| Nano-$TiO_2$ | 0.251 | 0.962 | 0.191 | 0.794 | 0.201 | 0.707 |
| Nano-$Al_2O_3$ | 0.154 | 0.563 | 0.187 | 0.626 | 0.18 | 0.677 |

The RF results demonstrate a highly consistent importance pattern across the three mechanical properties. Curing age was the dominant feature, accounting for 76.43% of the importance for CS, 80.54% for STS, and 80.47% for FS. For CS, Nano-$SiO_2$ was the second-most influential feature (9.10%), followed by Nano-$CaCO_3$ (4.98%) and nanomaterial dosage (3.19%). For STS, nanomaterial dosage (4.60%) and Nano-$CaCO_3$ (4.04%) were the next most important predictors. For FS, Nano-$SiO_2$ (4.51%), nanomaterial dosage (3.89%), and Nano-$CaCO_3$ (3.83%) followed curing age. The importance of Nano-$Fe_2O_3$, Nano-$TiO_2$, and Nano-$Al_2O_3$ remained comparatively small, each contributing less than 1% for all three target properties.

The XGBoost results also identified curing age as the most influential predictor for all three mechanical properties. However, its relative contribution was lower than that obtained from RF, while the importance assigned to specific nanomaterial types was generally higher. For CS, curing age contributed 57.31%, followed by Nano-$SiO_2$ (17.61%) and Nano-$CaCO_3$ (12.87%). Cement contributed 5.64%, while nanomaterial dosage contributed 3.71%. The importance of superplasticizer was 0% in the XGBoost model. For STS, curing age accounted for 64.84%, followed by cement (10.43%), Nano-$CaCO_3$ (9.58%), and Nano-$SiO_2$ (6.85%). Nanomaterial dosage contributed 4.78%. For FS, curing age again dominated with 64.51%, followed by Nano-$SiO_2$ (9.63%), cement (9.54%), and Nano-$CaCO_3$ (9.16%). Nanomaterial dosage contributed 4.36%.

Interestingly, superplasticizer received zero importance in all three XGBoost models. This result does not necessarily mean that superplasticizer has no physical influence on concrete performance. Rather, within the present dataset, its predictive information may be represented by other strongly associated variables.

**3.6 SHAP-Based Explainable-AI Analysis**

In the present investigation, SHAP analysis was used to examine the three response variables, namely compressive strength (CS), splitting tensile strength (STS), and flexural strength (FS). Particular attention was given to the variables identified as important by the RF and XGBoost analyses, including curing age, nanomaterial type, nanomaterial dosage, cement content, and superplasticizer content.

The SHAP summary plot for compressive strength is presented in Figure 8. The distribution of SHAP values illustrates the contribution of the input variables to individual CS predictions. Features located toward the upper portion of the ranking have a greater overall influence on the model output. The analysis should be interpreted in conjunction with the feature-importance results presented previously. In particular, curing age was identified as the dominant predictor by both RF and XGBoost, with importance values of approximately 76.43% and 57.31%, respectively. The SHAP analysis provides an additional perspective by indicating whether increasing curing age generally contributes positively or negatively to the predicted compressive strength. The SHAP distributions associated with the different nanomaterial categories can further reveal whether particular nanomaterial types systematically shift the predicted CS. The effects of Nano-$SiO_2$ and Nano-$CaCO_3$, which exhibited comparatively high importance in the tree-based models, should therefore receive particular attention. Nanomaterial dosage can also be examined to determine whether its contribution remains consistently positive or varies across the investigated dosage range.

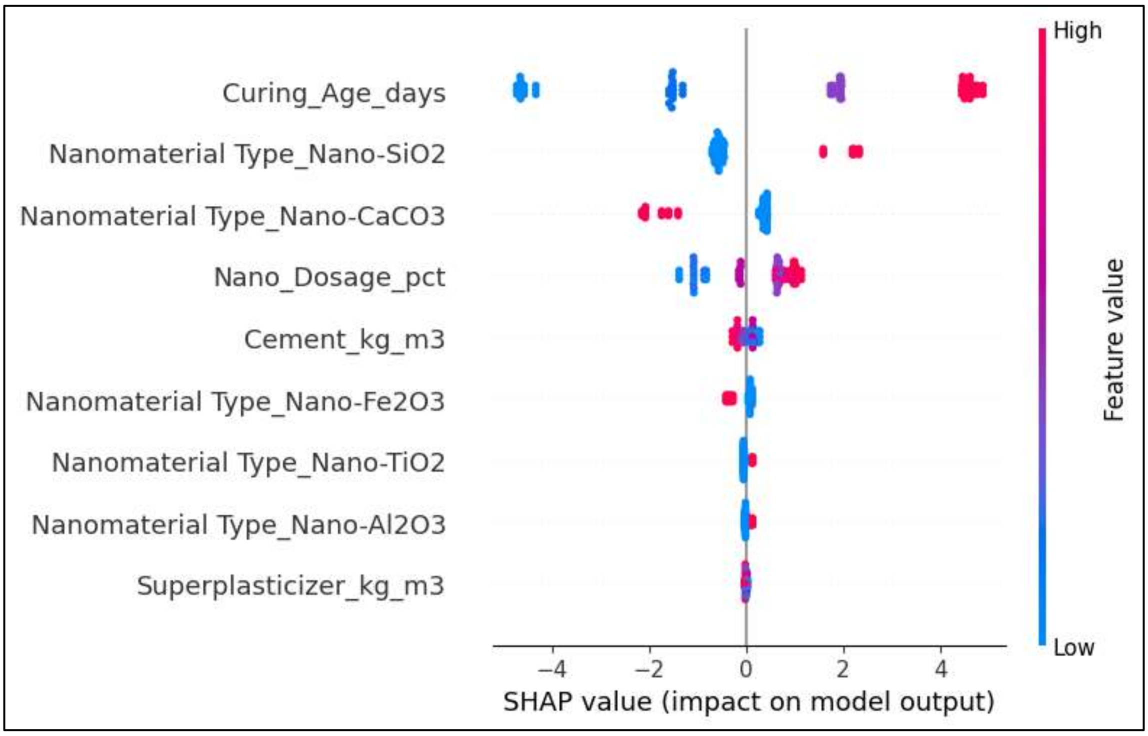


**Figure 8. SHAP summary plot for compressive strength.**

The SHAP summary plot for STS is presented in Figure 9. The analysis provides a feature-level explanation of the predicted splitting tensile strength and allows the direction of individual predictor effects to be evaluated. The tree-based feature-importance analysis identified curing age as the principal predictor of STS, with importance values of 80.54% for RF and 64.84% for XGBoost. The SHAP results can therefore be used to determine how changes in curing age affect individual predictions rather than merely reporting its overall importance. The contributions of nanomaterial dosage and nanomaterial type are also important because their effects may depend on the specific nanoparticle incorporated. The SHAP distributions can reveal whether higher values of dosage tend to increase the predicted STS or whether the effect changes across the investigated range. Similarly, the categorical nanomaterial variables can be interpreted according to their positive or negative SHAP contributions.

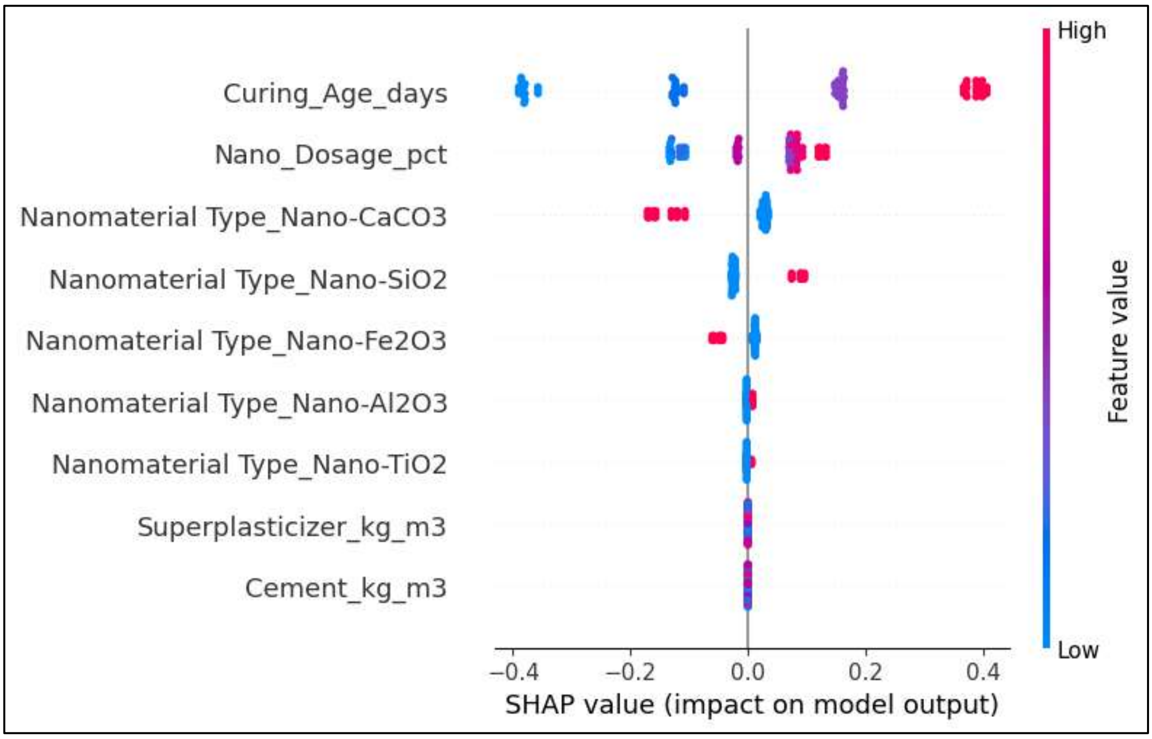


**Figure 9. SHAP summary plot for splitting tensile strength.**

The SHAP analysis for FS is presented in Figure 10. Similar to the other mechanical responses, curing age was found to be the dominant predictor in the tree-based feature-importance analysis, accounting for approximately 80.47% in RF and 64.51% in XGBoost. The SHAP results provide additional information regarding the direction and magnitude of the contribution of curing age. The analysis also facilitates examination of the individual effects of nanomaterial type, nanomaterial dosage, cement content, and superplasticizer content on flexural strength. The relatively high importance of Nano-$SiO_2$ and Nano-$CaCO_3$ identified in the XGBoost model can be further investigated through their SHAP distributions. Such interpretation helps distinguish a feature that is merely statistically important from one that has a consistent directional influence on the predicted response.

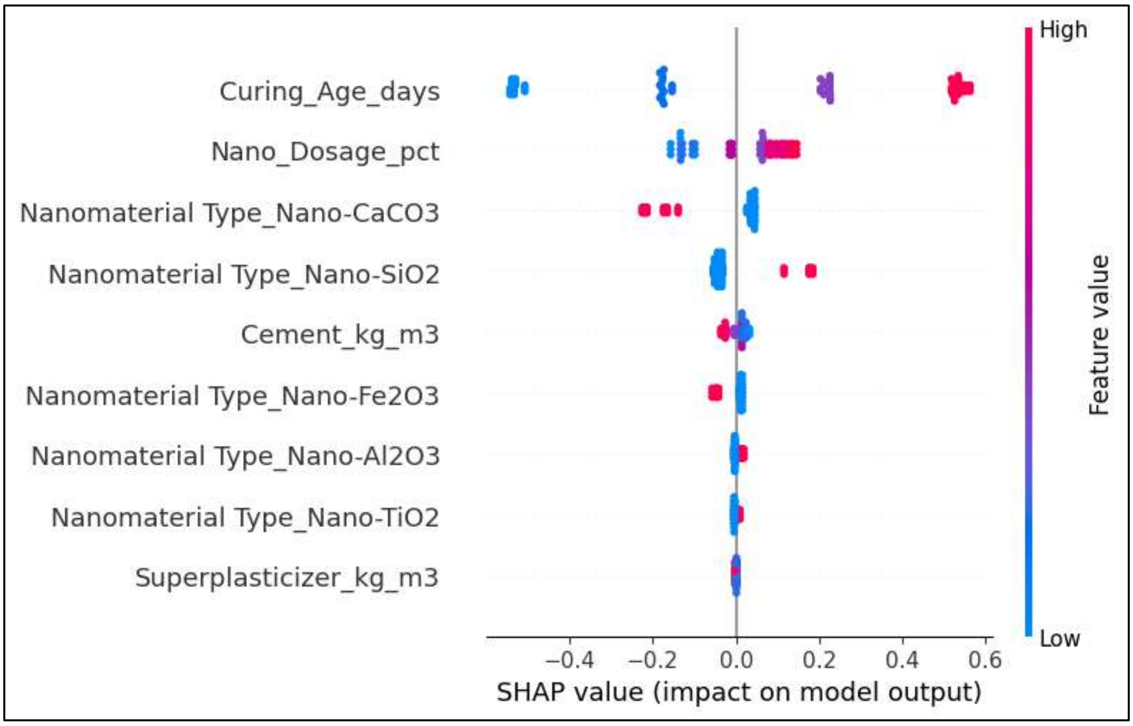


**Figure 10. SHAP summary plot for flexural strength.**

## 4. Conclusions

This study investigated the applicability of machine-learning techniques for predicting the mechanical performance of nanomaterial-modified concrete using 360 observations covering different nanomaterial types, dosages, mixture parameters, and curing ages. Based on the results obtained, the following conclusions can be drawn:

i. Machine learning demonstrated strong predictive capability for all three investigated mechanical properties, namely compressive strength, splitting tensile strength, and flexural strength. The four evaluated models—RF, XGBoost, SVR, and ANN—generally provided good agreement between observed and predicted values on the independent testing dataset.

ii. XGBoost was the best-performing model for all three target properties. Its testing R2values were 0.964839 for CS, 0.966156 for STS, and 0.964994 for FS, confirming its ability to capture the nonlinear relationships between the input variables and mechanical responses.

iii. For compressive strength, XGBoost achieved an RMSE of 0.721804 MPa and MAE of 0.595595 MPa, while for STS the corresponding values were 0.056642 MPa and 0.047509 MPa. For FS, XGBoost achieved an RMSE of 0.080656 MPa and MAE of 0.066620 MPa. These relatively low errors demonstrate satisfactory prediction accuracy within the investigated data range.

iv. The observed-versus-predicted analysis showed that the XGBoost predictions were closely distributed around the 1:1 reference line for CS, STS, and FS. Residual analysis further provided evidence of comparatively limited prediction errors for the best-performing model.

v. (v)Curing age was identified as the most influential input variable in both RF and XGBoost analyses. For RF, its feature importance was approximately 76.43% for CS, 80.54% for STS, and 80.47% for FS. In XGBoost, the corresponding importance values were 57.31%, 64.84%, and 64.51%, respectively. This consistent dominance indicates the importance of curing-age-dependent strength development within the investigated range.

vi. Nanomaterial type and dosage were also meaningful predictors. Nano-$SiO_2$ and Nano-$CaCO_3$ generally exhibited higher importance than the other nanomaterial categories in the tree-based models, although their relative contributions varied according to the target mechanical property and algorithm.

vii. Cement content and superplasticizer content contributed comparatively less than curing age and the major nanomaterial variables, particularly in the RF and XGBoost models. Nevertheless, retaining cement content in the predictive framework was appropriate because it represents a physically meaningful mixture parameter.

viii. The comparison of RF, XGBoost, SVR, and ANN indicates that model selection can influence prediction quality considerably. Although SVR provided competitive results, ANN generally exhibited comparatively greater prediction errors, particularly for flexural strength. XGBoost offered the most consistent performance across all three responses.

ix. The application of explainable-AI techniques provided an additional interpretation of the predictive models. Conventional feature importance identified the variables contributing most strongly to model decisions, while permutation importance for SVR and ANN can provide model-specific interpretation independent of their internal mathematical structure. SHAP analysis further enables examination of the magnitude and direction of individual feature contributions.

**Declarations**

**Availability of Data and Materials**

The data generated and/or analysed during the current study are available from the corresponding author on reasonable request.

**Competing Interests**

The authors declare that they have no competing interests.

**Funding**

The authors received no specific financial support or funding for the research, authorship, or publication of this article.

**Authors' Contributions**

Nirmal Kumar Bishwas Chaudhary: Experimental investigation, data collection, machine-learning modelling, data analysis, visualization, and manuscript preparation. Md Daniyal: Conceptualization, research supervision, methodology, interpretation of results. Rahat Yezdani: Critical review, and revision of the manuscript.

**Acknowledgements**

The authors gratefully acknowledge the Department of Civil Engineering, Sandip University, Madhubani, India, for providing the necessary academic and laboratory facilities to conduct this research.

## References


1. Bautista-Gutierrez, K. P., Herrera-May, A. L., Santamaría-López, J. M., Honorato-Moreno, A., & Zamora-Castro, S. A. (2019). Recent progress in nanomaterials for modern concrete infrastructure: Advantages and challenges. Materials, 12(21), 3546. https://doi.org/10.3390/ma12213548
2. Sikora, P., Elrahman, M. A., & Stephan, D. (2018). The influence of nanomaterials on the thermal resistance of cement-based composites—A review. Nanomaterials, 8(7), 465. https://doi.org/10.3390/nano8070465
3. Kawashima, S., Hou, P., Corr, D. J., & Shah, S. P. (2013). Modification of cement-based materials with nanoparticles. Cement and Concrete Composites, 36, 8–15. https://doi.org/10.1016/j.cemconcomp.2012.06.012
4. Norhasri, M. S. M., Hamidah, M. S., & Fadzil, A. M. (2017). Applications of using nano material in concrete: A review. Construction and Building Materials, 133, 91–97. https://doi.org/10.1016/j.conbuildmat.2016.12.005
5. Abhilash, P. P., Nayak, D. K., Sangoju, B., Kumar, R., & Kumar, V. (2021). Effect of nano-silica in concrete: A review. Construction and Building Materials, 278, 122347. https://doi.org/10.1016/j.conbuildmat.2021.122347
6. Daniyal, M., Akhtar, S., & Azam, A. (2019). Effect of nano-$TiO_2$ on the properties of cementitious composites under different exposure environments. Journal of Materials Research and Technology, 8(6), 6158–6172. https://doi.org/10.1016/j.jmrt.2019.10.010
7. Daniyal, M., Akhtar, S., Azam, A., & Islam, S. U. (2020). Influence of ethanolamine and nano-$TiO_2$ on the fresh, hardened, microstructural and corrosion resistance properties of cementitious composites. Arabian Journal for Science and Engineering, 45, 4369–4385. https://doi.org/10.1007/s13369-020-04469-8
8. Daniyal, M., Akhtar, S., Azam, A., & Islam, S. U. (2020). Effect of ethanolamine and nano-$TiO_2$ on the properties of ferrocement composites under different exposure environments. Sādhanā, 45, Article 86. https://doi.org/10.1007/s12046-020-01324-2
9. Li, C., Li, G., Chen, D., Gao, K., Mao, Y., Fan, S., Tang, L., & Jia, H. (2023). Influencing mechanism of nano-$Al_2O_3$ on concrete performance based on multi-scale experiments. Construction and Building Materials, 384, 131402. https://doi.org/10.1016/j.conbuildmat.2023.131402
10. Najafi Kani, E., Rafiean, A. H., Alishah, A., Hojjati Astani, S., & Ghaffar, S. H. (2021). The effects of Nano-$Fe_2O_3$ on the mechanical, physical and microstructure of cementitious composites. Construction and Building Materials, 266, 121137. https://doi.org/10.1016/j.conbuildmat.2020.121137
11. Verma, S. K., Daniyal, M., & Goldar, D. (2025). Effect of Nano-$Al_2O_3$, Nano-$SiO_2$, and Nano-$CaCO_3$ on the properties of cementitious composites. Procedia Structural Integrity, 70, 327–334. https://doi.org/10.1016/j.prostr.2025.07.060
12. Verma, S. K., Daniyal, M., & Goldar, D. (2026). Comparative study on the mechanical and durability performance of concrete incorporating nanomaterials. Next Nanotechnology, 9, 100361. https://doi.org/10.1016/j.nxnano.2025.100361
13. Wan, Z., Xu, Y., & Šavija, B. (2021). On the use of machine learning models for prediction of compressive strength of concrete: Influence of dimensionality reduction on the model performance. Materials, 14(4), 713. https://doi.org/10.3390/ma14040713
14. Nguyen, H., Vu, T., Vo, T. P., & Thai, H.-T. (2021). Efficient machine learning models for prediction of concrete strengths. Construction and Building Materials, 266, 120950. https://doi.org/10.1016/j.conbuildmat.2020.120950
15. Ekanayake, I. U., Meddage, D. P. P., & Rathnayake, U. (2022). A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). Case Studies in Construction Materials, 16, e01059. https://doi.org/10.1016/j.cscm.2022.e01059
16. Wu, Y., & Zhou, Y. (2022). Hybrid machine learning model and Shapley additive explanations for compressive strength of sustainable concrete. Construction and Building Materials, 330, 127298. https://doi.org/10.1016/j.conbuildmat.2022.127298
17. Wang, Z., Liu, H., Amin, M. N., Khan, K., Qadir, M. T., & Khan, S. A. (2024). Optimizing machine learning techniques and SHapley Additive exPlanations (SHAP) analysis for the compressive property of self-compacting concrete. Materials Today Communications, 39, 108804. https://doi.org/10.1016/j.mtcomm.2024.108804
18. Kashem, A., Karim, R., Malo, S. C., Das, P., Datta, S. D., & Alharthai, M. (2024). Hybrid data-driven approaches to predicting the compressive strength of ultra-high-performance concrete using SHAP and PDP analyses. Case Studies in Construction Materials, 20, e02991. https://doi.org/10.1016/j.cscm.2024.e02991